ANALYTIC COMBINATORICS

—

(Chapters I, II, III, IV, V, VI, VII, VIII, IX*)

PHILIPPE FLAJOLET & ROBERT SEDGEWICK

Algorithms Project
INRIA Rocquencourt
78153 Le Chesnay
France

Department of Computer Science
Princeton University
Princeton, NJ 08540
USA

Zeroth Edition. Seventh Printing: April 13, 2006—Paskebryg—(last corrected April 19, 2006).
(This temporary version will expire on July 14th, 2006)

♡ ♡ ♡ ♡ ♡ ♡

In this edition of April 19, 2006:

- Chapters 0, 1, 2, 3, 4, 5, 6 are in quasi-semi-final form.
- Chapter 9 is still in preliminary form
- Chapter 10, in preparation is not included.

Others chapter are in intermediate form, waiting to be revised.

Also, both long and short forms of construction names are currently used, with one destined to become the standard eventually. For unlabelled classes, the dictionary is

$$\mathfrak{S} \equiv \textsc{Seq}, \quad \mathfrak{M} \equiv \textsc{MSet}, \quad \mathfrak{P} \equiv \textsc{PSet}, \quad \mathfrak{C} \equiv \textsc{Cyc},$$

while for labelled classes,

$$\mathfrak{S} \equiv \textsc{Seq}, \quad \mathfrak{P} \equiv \textsc{Set}, \quad \mathfrak{C} \equiv \textsc{Cyc}\,.$$

PREFACE

ANALYTIC COMBINATORICS aims at predicting precisely the properties of large structured combinatorial configurations, through an approach based extensively on analytic methods. Generating functions are the central objects of the theory.

Analytic combinatorics starts from an exact enumerative description of combinatorial structures by means of generating functions, which make their first appearance as purely formal algebraic objects. Next, generating functions are interpreted as analytic objects, that is, as mappings of the complex plane into itself. Singularities determine a function's coefficients in asymptotic form and lead to precise estimates for counting sequences. This chain applies to a large number of problems of discrete mathematics relative to words, trees, permutations, graphs, and so on. A suitable adaptation of the methods also opens the way to the quantitative analysis of characteristic parameters of large random structures, via a perturbational approach.

Analytic combinatorics can accordingly be organized based on three components:

> *Symbolic Methods* develops systematic relations between some of the major constructions of discrete mathematics and operations on generating functions which exactly encode counting sequences.
>
> *Complex Asymptotics* elaborates a collection of methods by which one can extract asymptotic counting information from generating functions, once these are viewed as analytic transformations of the complex domain. Singularities then appear to be a key determinant of asymptotic behaviour.
>
> *Random Structures* concerns itself with probabilistic properties of large random structures. Which properties hold with high probability? Which laws govern randomness in large objects? In the context of analytic combinatorics, these questions are treated by a deformation (adding auxiliary variables) and a perturbation (examining the effect of small variations of such auxiliary variables) of the standard enumerative theory.

THE APPROACH to quantitative problems of discrete mathematics provided by analytic combinatorics can be viewed as an *operational calculus* for combinatorics. The present book exposes this view by means of a very large number of examples concerning classical combinatorial structures—most notably, words, trees, permutations, and graphs. The eventual goal is an effective way of quantifying metric properties of large random structures.

Given its capacity of quantifying properties of large discrete structures, *Analytic Combinatorics* is susceptible to many applications, within combinatorics itself, but, perhaps more importantly, within other areas of science where discrete probabilistic models recurrently surface, like statistical physics, computational biology, or electrical engineering. Last but not least, the analysis of algorithms and data structures in computer science has served and still serves as an important motivation in the development of the theory.

⋆ ⋆ ⋆ ⋆ ⋆⋆

Part A: Symbolic Methods. This part specifically exposes *Symbolic Combinatorics*, which is a unified algebraic theory dedicated to setting up functional relations between counting generating functions. As it turns out, a collection of general (and simple) theorems provide a systematic translation mechanism between combinatorial constructions and operations on generating functions. This translation process is a purely formal one. Precisely, as regards basic counting, two parallel frameworks coexist—one for unlabelled structures and ordinary generating functions, the other for labelled structures and exponential generating functions. Furthermore, within the theory, parameters of combinatorial configurations can be easily taken into account by adding supplementary variables. Three chapters then compose this part: Chapter I deals with unlabelled objects; Chapter II develops in a parallel way labelled objects; Chapter III treats multivariate aspects of the theory suitable for the analysis of parameters of combinatorial structures.

⋆ ⋆ ⋆ ⋆ ⋆⋆

Part B: Complex asymptotics. This part specifically exposes *Complex Asymptotics*, which is a unified analytic theory dedicated to the process of extracting asymptotic information from counting generating functions. A collection of general (and simple) theorems provide a systematic translation mechanism between generating functions and asymptotic forms of coefficients. Four chapters compose this part. Chapter IV serves as an *introduction to complex-analytic methods* and proceeds with the treatment of *meromorphic functions*, that is, functions whose singularities are poles, *rational functions* being the simplest case. Chapter V develops *applications of rational and meromorphic asymptotics of generating functions*, with numerous applications related to words and languages, walks and graphs, as well as permutations. Chapter VI develops a general theory of *singularity analysis* that applies to a wide variety of singularity types, such as square-root or logarithmic, and has applications to trees as well as to other recursively defined combinatorial classes. Chapter VII presents *applications of singularity analysis* to 2-regular graphs and polynomials, trees of various sorts, mappings, context-free languages, walks, and maps. It contains in particular a discussion of the analysis of coefficients of algebraic functions. Chapter VIII explores *saddle point methods*, which are instrumental in analysing functions with a violent growth at a singularity, as well as many functions with only a singularity at infinity (i.e., entire functions).

⋆ ⋆ ⋆ ⋆ ⋆⋆

Part C: Random Structures. This part includes Chapter IX dedicated to the analysis of multivariate generating functions viewed as deformation and perturbation of simple (univariate) functions. As a consequence, many important characteristics of classical combinatorial structures can be precisely quantified in distribution. Chapter **??** is an epilogue, which offers a brief recapitulation of the major asymptotic properties of discrete structures developed in earlier chapters.

⋆ ⋆ ⋆ ⋆ ⋆⋆

Part D: Appendices. Appendix A summarizes some key elementary concepts of combinatorics and asymptotics, with entries relative to asymptotic expansions, languages, and trees, amongst others. Appendix B recapitulates the necessary background in complex analysis. It may be viewed as a self-contained minicourse on the subject, with entries relative to analytic functions, the Gamma function, the implicit function theorem, and Mellin transforms. Appendix C recalls some of the basic notions of probability theory that are useful in analytic combinatorics.

⋆ ⋆ ⋆ ⋆ ⋆⋆

THIS BOOK is meant to be reader-friendly. Each major method is abundantly illustrated by means of concrete examples[1] treated in detail—there are scores of them, spanning from a fraction of a page to several pages—offering a complete treatment of a specific problem. These are borrowed not only from combinatorics itself but also from neighbouring areas of science. With a view of addressing not only mathematicians of varied profiles but also scientists of other disciplines, *Analytic Combinatorics* is self-contained, including ample appendices that recapitulate the necessary background in combinatorics and complex function theory. A rich set of short Notes—there are more than 250 of them—are inserted in the text[2] and can provide exercises meant for self-study or for students' practice, as well as introductions to the vast body of literature that is available. We have also made every effort to focus on *core ideas* rather than technical details, supposing a certain amount of mathematical maturity but only basic prerequisites on the part of our gentle readers. The book is also meant to be strongly problem-oriented, and indeed it can be regarded as a manual, or even a huge algorithm, guiding the reader to the solution of a very large variety of problems regarding discrete mathematical models of varied origins. In this spirit, many of our developments connect nicely with computer algebra and symbolic manipulation systems.

COURSES can be (and indeed have been) based on the book in various ways. Chapters I–III on *Symbolic Methods* serve as a systematic yet accessible introduction to the formal side of combinatorial enumeration. As such it organizes transparently some of the rich material found in treatises[3] like those of Bergeron-Labelle-Leroux, Comtet, Goulden-Jackson, and Stanley. Chapters IV–VIII relative to *Complex Asymptotics* provide a large set of concrete examples illustrating the power of classical complex analysis and of asymptotic analysis outside of their traditional range of applications. This material can thus be used in courses of either pure or applied mathematics, providing a wealth of nonclassical examples. In addition, the quiet but ubiquitous presence of symbolic manipulation systems provides a number of illustrations of the power of these systems while making it possible to test and concretely experiment with a great many combinatorial models. Symbolic systems allow for instance for fast random generation, close examination of non-asymptotic regimes, efficient experimentation with analytic expansions and singularities, and so on.

---

[1] Examples are marked by "EXAMPLE $\cdots$ $\square$".

[2] Notes are indicated by $\triangleright \cdots \triangleleft$.

[3] References are to be found in the bibliography section at the end of the book.

Our initial motivation when starting this project was to build a coherent set of methods useful in the analysis of algorithms, a domain of computer science now well-developed and presented in books by Knuth, Hofri, and Szpankowski, in the survey by Vitter–Flajolet, as well as in our earlier *Introduction to the Analysis of Algorithms* published in 1996. This book can then be used as a systematic presentations of methods that have proved immensely useful in this area; see in particular the *Art of Computer Programming* by Knuth for background. Studies in statistical physics (van Rensburg, and others), statistics (e.g., David and Barton) and probability theory (e.g., Billingsley, Feller), mathematical logic (Burris' book), analytic number theory (e.g., Tenenbaum), computational biology (Waterman's textbook), as well as information theory (e.g., the books by Cover–Thomas, MacKay, and Szpankowski) point to many startling connections with yet other areas of science. The book may thus be useful as a supplementary reference on methods and applications in courses on statistics, probability theory, statistical physics, finite model theory, analytic number theory, information theory, computer algebra, complex analysis, or analysis of algorithms.

Acknowledgements. This book would be substantially different and much less informative without Neil Sloane's *Encyclopedia of Integer Sequences*, Steve Finch's *Mathematical Constants*, Eric Weisstein's *MathWorld*, and the *MacTutor History of Mathematics* site hosted at St Andrews. All are (or at least have been at some stage) freely available on the Internet. Bruno Salvy and Paul Zimmermann have developed algorithms and libraries for combinatorial structures and generating functions that are based on the MAPLE system for symbolic computations and have proven to be extremely useful. We are deeply grateful to the authors of the free software Unix, Linux, Emacs, X11, TEX and LATEX as well as to the designers of the symbolic manipulation system MAPLE for creating an environment that has proved invaluable to us. We also thank students in courses at Barcelona, Berkeley (MSRI), Bordeaux, Caen, Paris (École Polytechnique, École Normale, University), Princeton, Santiago de Chile, Udine, and Vienna whose feedback has greatly helped us prepare a better book. Thanks finally to numerous colleagues for their feedback. In particular, we wish to acknowledge the support, help, and interaction provided at an incredibly high level by members of the *Analysis of Algorithms (AofA)* community, with a special mention for Hsien-Kuei Hwang, Svante Janson, Don Knuth, Guy Louchard, Andrew Odlyzko, Daniel Panario, Helmut Prodinger, Bruno Salvy, Michèle Soria, Wojtek Szpankowski, Brigitte Vallée, and Mark Wilson. Stan Burris, Svante Janson, Loïc Turban, and Brigitte Vallée especially have provided insightful suggestions and detailed feedback that have led us to revise the presentation of several sections of this book and correct many errors. Finally, support of our home institutions (INRIA and Princeton University) as well as various grants (French government, European Union and the ALCOM Project, NSF) have contributed to making our collaboration possible.

PHILIPPE FLAJOLET & ROBERT SEDGEWICK
April 19, 2006

Contents

An invitation to Analytic Combinatorics

διὸ δὴ συμμειγνύμενα αὐτά τε πρὸς αὑτὰ καὶ πρὸς ἄλληλα τὴν ποικιλίαν ἐστὶν ἄπειρα· ἧς δὴ δεῖ θεωροὺς γίγνεσθαι τοὺς μέλλοντας περὶ φύσεως εἰκότι λόγῳ

— PLATO, The Timaeus[1]

ANALYTIC COMBINATORICS is primarily a book about *Combinatorics*, that is, the study of finite structures built according to a finite set of rules. *Analytic* in the title means that we concern ourselves with methods from mathematical analysis, in particular complex and asymptotic analysis. The two fields, combinatorial enumeration and complex asymptotics, are organized into a coherent set of methods for the first time in this book. Our broad objective is to discover how the continuous may help us to understand the discrete and to *quantify* its properties.

COMBINATORICS is as told by its name the science of combinations. Given basic rules for assembling simple components, what are the properties of the resulting objects? Here, our goal is to develop methods dedicated to *quantitative* properties of combinatorial structures. In other words, we want to measure things. Say that we have n different items like cards or balls of different colours. In how many ways can we lay them on a table, all in one row? You certainly recognize this counting problem—finding the number of *permutations* of n elements. The answer is of course the factorial number, $n\,! = 1 \cdot 2 \cdots n$. This is a good start, and, equipped with patience or a calculator, we soon determine that if $n = 31$, say, then the number is the rather large[2]

$$31\,! = 8222838654177922817725562880000000 \doteq 0.8222838654 \cdot 10^{34}.$$

The factorials solve an enumerative problem, one that took mankind some time to sort out, because the sense of the '$\cdots$' in the formula is not that easily grasped. In his book

[1]"*So their combinations with themselves and with each other give rise to endless complexities, which anyone who is to give a likely account of reality must survey.*" Plato speaks of Platonic solids viewed as idealized primary constituents of the physical universe.

[2] We use '$\alpha \doteq \mathsf{d}$ to represent a numerical estimation of the real α by the decimal d, with the last digit being at most ± 1 from its actual value.

The Art of Computer Programming (vol III, p. 23), Donald Knuth traces the discovery to the Hebrew *Book of Creation* (c. A.D. 400) and the Indian classic *Anuyogadvārasutra* (c. A.D. 500).

Here is another more subtle problem. Assume that you are interested in permutations such that the first element is smaller than the second, the second is larger than the third, itself smaller than the fourth, and so on. The permutations go up and down and they are diversely known as up-and-down or zigzag permutations, the more dignified name being *alternating* permutations. Say that $n = 2m + 1$ is odd. An example is for $n = 9$:

$$\begin{array}{ccccccccccccccccc} & & 8 & & & & 7 & & & & 9 & & & & 3 & & \\ & \nearrow & & \searrow & & \nearrow & & \searrow & & \nearrow & & \searrow & & \nearrow & & \searrow & \\ 4 & & & & 6 & & & & 5 & & & & 1 & & & & 2 \end{array}$$

The number of alternating permutations for $n = 1, 3, 5, \ldots$ turns out to be

$$1, 2, 16, 272, 7936, 353792, 22368256, \ldots .$$

What are these numbers and how do they relate to the total number of permutations of corresponding size? A glance at the corresponding figures, that is, $1!, 3!, 5!, \ldots$ or

$$1, 6, 120, 5040, 362880, 39916800, 6227020800, \ldots$$

suggests that the factorials grow somewhat faster—just compare the lengths of the last two displayed lines. But how and by how much? This is the prototypical question we are addressing in this book.

Let us now examine the counting of alternating permutations. In 1881, the French mathematician Désiré André made a startling discovery. Look at the first terms of the Taylor expansion of the trigonometric function $\tan(z)$:

$$\tan z = 1\frac{z}{1!} + 2\frac{z^3}{3!} + 16\frac{z^5}{5!} + 272\frac{z^7}{7!} + 7936\frac{z^9}{9!} + 353792\frac{z^{11}}{11!} + \cdots .$$

The counting sequence for alternating permutations curiously surfaces. We say that the function on the left is a *generating function* for the numerical sequence (precisely, a generating function of the *exponential* type due to the presence of factorials in the denominators).

André's derivation may nowadays be viewed very simply as reflecting of the construction of permutations by means of certain binary trees: Given a permutation σ a tree can be obtained once σ has been decomposed as a triple $\langle \sigma_L, \max, \sigma_R \rangle$, by taking the maximum element as the root, and appending, as left and right subtrees, the trees recursively constructed from σ_L and σ_R. Part A of this book develops at length *symbolic methods* by which the construction of the class $\mathcal{T}$ of all such trees,

$$\mathcal{T} \quad = \quad \mathbf{1} \quad + \quad (\mathcal{T}, \max, \mathcal{T})$$

translates into an equation relating generating functions,

$$T(z) \quad = \quad z \quad + \quad \int_0^z T(w)^2 \, dw.$$

In this equation, $T(z) := \sum_n T_n z^n / n!$ is the exponential generating function of the sequence (T_n), where T_n is the number of alternating permutations of (odd) length n. There is a compelling formal analogy between the combinatorial *specification* and the world of generating functions: Unions ($\cup$) give rise to sums ($+$), max-placement gives an integral ($\int$), forming a pair of trees means taking a square ($[\cdot]^2$).

At this stage, we know that $T(z)$ must solve the differential equation

$$\frac{d}{dz}T(z) = 1 + T(z)^2, \qquad T(0) = 0,$$

which, by classical manipulations, yields $T(z) \equiv \tan z$. The generating function then provides a simple *algorithm* to compute recurrently the coefficients, since the formula,

$$\tan z = \frac{\sin z}{\cos z} = \frac{z - \frac{z^3}{3!} + \frac{z^5}{5!} - \cdots}{1 - \frac{z^2}{2!} + \frac{z^4}{4!} - \cdots},$$

implies (n odd)

$$T_n - \binom{n}{2}T_{n-2} + \binom{n}{4}T_{n-4} - \cdots = (-1)^{(n-1)/2}, \quad \text{where} \quad \binom{a}{b} = \frac{a!}{b!(a-b)!}$$

is the conventional notation for binomial coefficients. At this stage, the exact enumerative problem may be regarded as solved since a very simple recurrent algorithm is available for determining the counting sequence, while the generating function admits an explicit expression in terms of a well known function.

ANALYSIS, by which we mean mathematical analysis, is often described as the art and science of *approximation*. How fast do the factorial and the tangent number sequences grow? What about *comparing* their growths? These are typical problems of analysis.

First, consider the number of permutations, $n!$. Quantifying the growth of these numbers as n gets large takes us to the realm of *asymptotic analysis*. The way to express factorial numbers in terms of elementary functions is known as Stirling's formula,

$$n! \sim n^n e^{-n}\sqrt{2\pi n},$$

where the $\sim$ sign means "approximately equal" (in fact, in the precise sense that the ratio of both terms tends to 1 as n gets large). This beautiful formula, associated with the name of the eighteenth century Scottish mathematician James Stirling, curiously involves both the basis e of natural logarithms and the perimeter 2π of the circle. Certainly, you cannot get such a thing without analysis. As a first step, there is an estimate for

$$\log n! = \sum_{j=1}^{n} \log j \sim \int_1^n \log x \, dx \sim n \log(\frac{n}{e}),$$

explaining at least the $n^n e^{-n}$ term, but already requiring some amount of elementary calculus. (Stirling's formula precisely came a few decades after the fundamental bases of calculus had been laid by Newton and Leibniz.) Note the usefulness of Stirling's formula: it tells us almost instantly that $100!$ has 158 digits, while $1000!$ borders the astronomical 10^{2568}.

We are now left with estimating the growth of the sequence of tangent numbers, T_n. The analysis leading to the derivation of the generating function $\tan(z)$ has been so far essentially algebraic or "formal". Well, we can plot the graph of the tangent function, for real values of its argument and see that the function becomes infinite

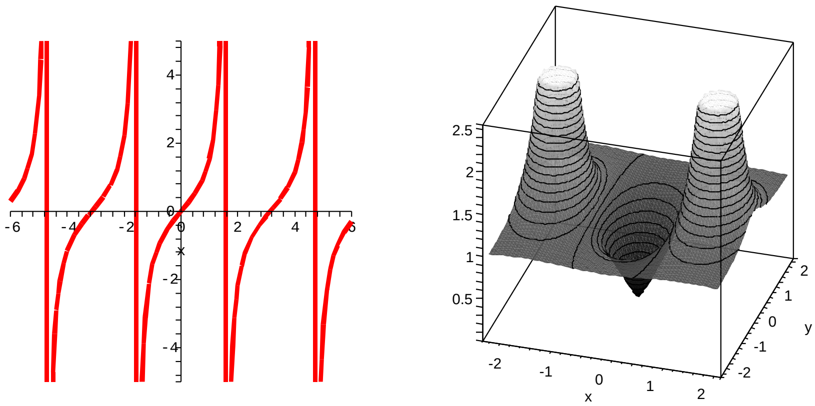


FIGURE 1. Two views of the function $z \mapsto \tan z$: (left) a plot for real values of $z \in [-5..5]$; (right) the modulus $|\tan z|$ when z is assigned complex values in the square $\pm 2.25 \pm 2.25\sqrt{-1}$.

at the points $\pm\frac{\pi}{2}$, $\pm 3\frac{\pi}{2}$, and so on (Figure 1). Such points where a function ceases to be smooth are called *singularities*. By methods amply developed in this book, it is the local nature of a generating function at its "dominant" singularities (i.e., the ones closest to the origin) that determines the asymptotic growth of the sequence of coefficients. In this perspective, the basic fact that $\tan z$ has dominant singularities at $\pm\frac{\pi}{2}$ enables us to reason as follows: first approximate the generating function $\tan z$ near its two dominant singularities, namely,

$$\tan(z) \underset{z \to \pm\pi/2}{\sim} \frac{8z}{\pi^2 - 4z^2};$$

then extract coefficients of this approximation; finally, get in this way a valid approximation of coefficients:

$$\frac{T_n}{n!} \underset{n\to\infty}{\sim} 2 \cdot \left(\frac{2}{\pi}\right)^{n+1} \qquad (n \text{ odd}).$$

With present day technology, we also have available *symbolic manipulation* systems (also called "computer algebra" systems) and it is not difficult to verify the accuracy of our estimates. Here is a small pyramid for $n = 3, 5, \ldots, 21$,

2	***1***
16	1***5***
272	27***1***
7936	793***5***
353792	35379***1***
22368256	2236825***1***
1903757312	1903757***267***
209865342976	20986534***2434***
29088885112832	290888851***04489***
4951498053124096	495149805***2966307***
(T_n)	$(T_n^\star)$

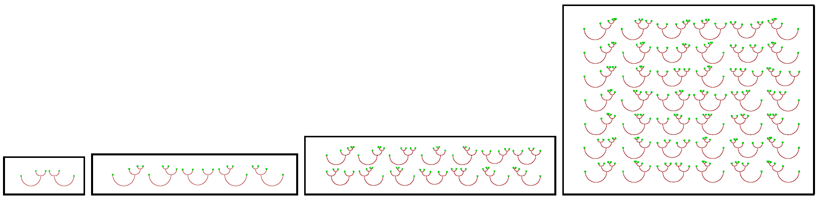


FIGURE 2. The collection of all binary trees for sizes $n = 2, 3, 4, 5$ with respective cardinalities $2, 5, 14, 42$.

comparing the exact values of T_n against the approximations $T_n^\star$, where (n odd)

$$T_n^\star := \left\lfloor 2 \cdot n! \, \left(\frac{2}{\pi}\right)^{n+1} \right\rfloor,$$

and discrepant digits of the approximation are displayed in bold. For $n = 21$, the error is only of the order of one in a billion. Asymptotic analysis is in this case wonderfully accurate.

In the foregoing discussion, we have played down a fact, and an important one. When investigating generating functions from an analytic standpoint, one should generally assign *complex* values to arguments not just real ones. It is singularities in the complex plane that matter and complex analysis is needed in drawing conclusions regarding the asymptotic form of coefficients of a generating function. Thus, a large portion of this book relies on a *complex analysis* technology, which starts to be developed in Part B of the book titled *Complex Asymptotics*. This approach to combinatorial enumeration parallels what happened in the nineteenth century, when Riemann first recognized the deep relation between complex-analytic properties of the $zeta$ function, $\zeta(s) := \sum 1/n^s$, and the distribution of primes, eventually leading to the long-sought proof of the Prime Number Theorem by Hadamard and de la Vallée-Poussin in 1896. Fortunately, relatively elementary complex analysis suffices for our purposes, and we can include in this book a complete treatment of the fragment of the theory needed to develop the bases of analytic combinatorics.

Here is yet another example illustrating the close interplay between combinatorics and analysis. When discussing alternating permutations, we have enumerated binary trees bearing distinct integer labels that satisfy a constraint—to increase along branches. What about the simpler problem of determining the number of possible *shapes* of binary trees? Let C_n be the number of binary trees that have n binary branching nodes, hence $n + 1$ "external nodes". It is not hard to come up with an exhaustive listing for small values of n; see Figure 2, from which we determine that

$$C_0 = 1, C_1 = 1,\ C_2 = 2,\ C_3 = 5,\ C_4 = 14,\ C_5 = 42.$$

These numbers are probably the most famous ones of elementary combinatorics. They have come to be known as the *Catalan numbers* as a tribute to the Belgian French mathematician Eugène Charles Catalan (1814-1894), but they already appear in works of Euler and Segner in the second half of the eighteenth century. In his reference treatise on *Enumerative Combinatorics*, Stanley lists over twenty pages a collection of some 66 different types of combinatorial structures that are enumerated by the Catalan numbers.

First, one can write a combinatorial equation, very much in the style of what has been done earlier, but without labels:

$$\mathcal{C} \quad = \quad \Box \quad + \quad (\mathcal{C}, \bullet\,, \mathcal{C})\,.$$

With symbolic methods, it is easy to see that the *ordinary generating function* of the Catalan numbers defined as

$$C(z) := \sum_{n \geq 0} C_n z^n,$$

satisfies an equation that is a direct reflection of the combinatorial definition, namely,

$$C(z) \quad = \quad 1 \quad + \quad z\, C(z)^2.$$

This is a quadratic equation whose solution is

$$C(z) = \frac{1 - \sqrt{1-4z}}{2z}.$$

Then, by means of Newton's theorem relative to the expansion of $(1+x)^\alpha$, one finds easily ($x = -4z$, $\alpha = \frac{1}{2}$) the *closed form* expression

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Regarding asymptotic approximation, Stirling's formula comes to the rescue: it implies

$$C_n \sim C_n^\star \qquad \text{where} \quad C_n^\star := \frac{4^n}{\sqrt{\pi n^3}}.$$

This approximation is quite usable: it predicts $C_1^\star \doteq 2.25$ (whereas $C_1 = 1$), which is off by a factor of 2, but the error drops to 10% already for $n = 10$, and it appears to be less than 1% for any $n \geq 100$.

A plot of the generating function $C(z)$ in Figure 3 illustrates the fact that $C(z)$ has a *singularity* at $z = \frac{1}{4}$ as it ceases to be differentiable (its derivative becomes infinite). That singularity is quite different from a pole and for natural reasons it is known a square-root singularity. As we shall see repeatedly, under suitable conditions in the complex plane, a square root singularity for a function at a point ρ invariably entails an asymptotic form $\rho^{-n} n^{-3/2}$ for its coefficients. More generally, it suffices to estimate a generating function near a singularity in order to deduce an asymptotic approximation of its coefficients. This correspondence is a major theme of the book, one that motivates the four central chapters.

A consequence of the complex-analytic vision of combinatorics is the detection of *universality phenomena* in large random structures. (The term is originally borrowed from statistical physics and is nowadays finding increasing use in areas of mathematics like probability theory.) By universality is meant here that many quantitative properties of combinatorial structures only depend on a few global features of their definitions, not on details. For instance a growth in the counting sequence of the form

$$C \cdot A^n n^{-3/2},$$

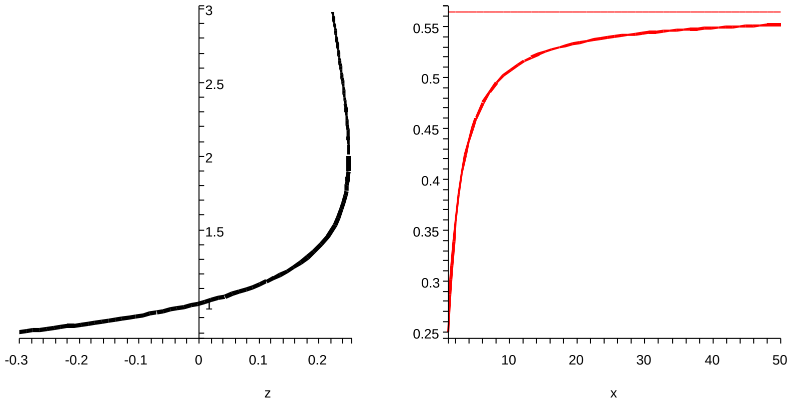


FIGURE 3. Left: the real values of the Catalan generating function, which has a square-root singularity at $z = \frac{1}{4}$. Right: the ratio $C^n/(4^n n^{-3/2})$ plotted together with its asymptote at $1/\sqrt{\pi} \doteq 0.56418$.

arising from a square-root singularity, will be shown to be universal across *all* varieties of trees determined by a finite set of allowed node degrees—this includes unary-binary trees, ternary trees, 0–11–13 trees, as well as many variations like nonplane trees and labelled trees. Even though generating functions may become arbitrarily complicated—like an algebraic function of a very high degree or even the solution to an infinite functional equation—it is still possible to extract with relative ease *global asymptotic laws* governing *counting sequences*.

RANDOMNESS is another ingredient in our story. How useful is it to determine, exactly or approximately, counts that may be so large as to require hundreds if not thousands of digits in order to be written down? Take again the example of alternating permutations. When estimating their number, we have indeed quantified the proportion of these amongst all permutations. In other words, we have been predicting the *probability* that a random permutation of some size n is alternating. Results of this sort are of interest in all branches of science. For instance, biologists routinely deal with genomic sequences of length 10^5, and the interpretation of data requires developing enumerative or probabilistic models where the number of possibilities is of the order of 4^{10^5}. The language of probability theory then proves a great convenience when discussing characteristic parameters of discrete structures, as we can interpret exact or asymptotic enumeration results as saying something concrete about the likeliness of values that such parameters assume. Equally important of course are results from several areas of probability theory: as demonstrated in the later sections of this book, such results merge extremely well with the analytic-combinatorial framework.

Say we are now interested in runs in permutations. These are the longest fragments of a permutation that already appear in (increasing) sorted order. Here is a permutation where runs have been separated by vertical bars:

$$2\,5\,8\,|\,3\,9\,|\,1\,4\,7\,|\,6\,|\,.$$

Runs naturally present in a permutation are for instance exploited by a sorting algorithm called "natural list mergesort", which builds longer and longer runs, starting from the original ones and merging them until the permutation is eventually sorted. For our understanding of this algorithm, it is then of obvious interest to quantify how many runs a permutation is likely to have.

Let $A_{n,k}$ be the number of permutations of size n having k runs. Then, the problem is once more best approached by generating functions and one finds that the coefficient of $u^k z^n$ inside the *bivariate* generating function,

$$\frac{1-u}{1-ue^{z(1-u)}} = 1 + zu + \frac{z^2}{2!}(u+1) + \frac{z^3}{3!}(u^2+4u+1) + \cdots\ ,$$

gives the sought numbers $A_{n,k}/n!$. (A simple way of establishing this formula bases itself on the tree decomposition of permutations and on the symbolic method.) From there, we can easily determine effectively the mean, variance, and even the higher moments of the number of runs that a random permutation has: it suffices to expand blindly, or even better with the help of a computer, the bivariate generating function above as $u \to 1$:

$$\frac{1}{1-z} + \frac{1}{2}\,\frac{z\,(2-z)}{(1-z)^2}\,(u-1) + \frac{1}{2}\,\frac{z^2\left(6-4\,z+z^2\right)}{(1-z)^3}\,(u-1)^2 + \cdots\ .$$

When $u = 1$, we just enumerate all permutations: this is the constant term $1/(1-z)$ equal to the exponential generating function of all permutations. The coefficient of $u-1$ gives the generating function of the *mean* number of runs, the next one gives access to the second moment, and so on. In this way, we discover that the expectation and standard deviation of the number of runs in a permutation of size n evaluate to

$$\mu_n = \frac{n+1}{2}, \qquad \sigma_n = \sqrt{\frac{n+1}{12}}.$$

Then by easy analytic-probabilistic inequalities (Chebyshev inequalities) that otherwise form the basis of what is known as the second moment method, we learn that the distribution of the number of runs is concentrated around its mean: in all likelihood, if one takes a random permutation, the number of its runs is going to be very close to its mean. The effects of such quantitative laws are quite tangible. It suffices to draw a *sample of one element* for $n = 30$ to get something like

$$13, 22, 29|12, 15, 23|8, 28|18|6, 26|4, 10, 16|1, 27|3, 14, 17, 20|2, 21, 30|25|11, 19|9|7, 24.$$

For $n = 30$, the mean is $15\frac{1}{2}$, and this sample comes rather close as it has 13 runs. We shall furthermore see in Chapter IX that even for moderately large permutations of size 10,000 and beyond, the probability for the number of observed runs to deviate by more than 10% from the mean is less than 10^{-65}. As witnessed by this example, much regularity accompanies properties of large combinatorial structures.

More refined methods combine the observation of singularities with analytic resuts from probability theory (e.g., continuity theorems for characteristic functions). In the case of runs in permutations, the quantity $F(z,u)$ viewed as a function of z when u

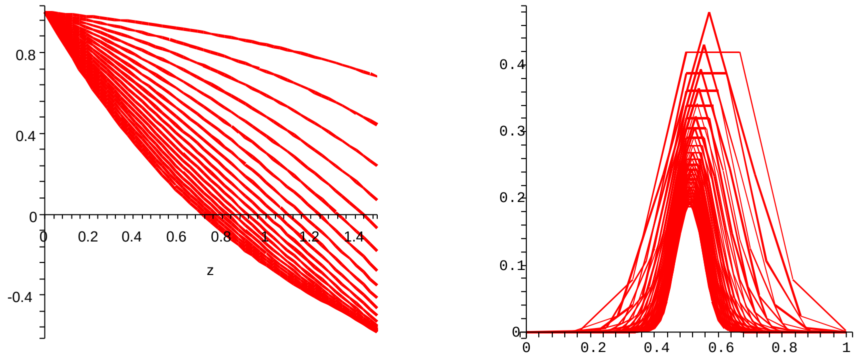


FIGURE 4. Left: A partial plot of the real values of the inverse $1/F(z,u)$ for $u = 0.1\,..\,2$, with F the bivariate generating function of Eulerian numbers, illustrates the presence of a movable pole for F. Right: A diagram showing the distribution of the number of runs in permutations for $n = 6\,..\,60$.

is fixed appears to have a pole: this fact is apparent on Figure 4 [left] since $1/F$ has a zero at some $z = \rho(u)$ where $\rho(1) = 1$. Then we are confronted with a fairly regular *deformation* of the generating function of all permutations. A parameterized version (with parameter u) of singularity analysis then gives access to a description of the asymptotic behaviour of the Eulerian numbers $A_{n,k}$. This enables us to describe very precisely what goes on: In a random permutation of large size n, once centred by its mean and scaled by its standard deviation, *the distribution of the number of runs is asymptotically gaussian*; see Figure 4 [right].

A somewhat similar type of situation prevails for binary trees, despite the fact that the counting sequences and the counting generating functions look rather different from their permutation counterparts. Say we are interested in leaves (also sometimes known as "cherries") in trees: these are binary nodes that are attached to two external nodes ($\Box$). Let $C_{n,k}$ be the number of trees of size n having k leaves. The bivariate generating function $C(z,u) := \sum_{n,k} C_{n,k} z^n u^k$ encodes all the information relative to leaf statistics in random binary trees. A modification of previously seen symbolic arguments shows that $C(z,u)$ still satisfies a quadratic equation resulting in the explicit form,

$$C(z,u) = \frac{1 - \sqrt{1 - 4z + 4z^2(1-u)}}{2z}.$$

This reduces to $C(z)$ for $u = 1$, as it should, and the bivariate generating function $C(z,u)$ is a deformation of $C(z)$ as u varies. In fact, the network of curves of Figure 5 for several fixed values of u shows that there is a smoothly varying square-root singularity. It is possible to analyse the *perturbation* induced by varying values

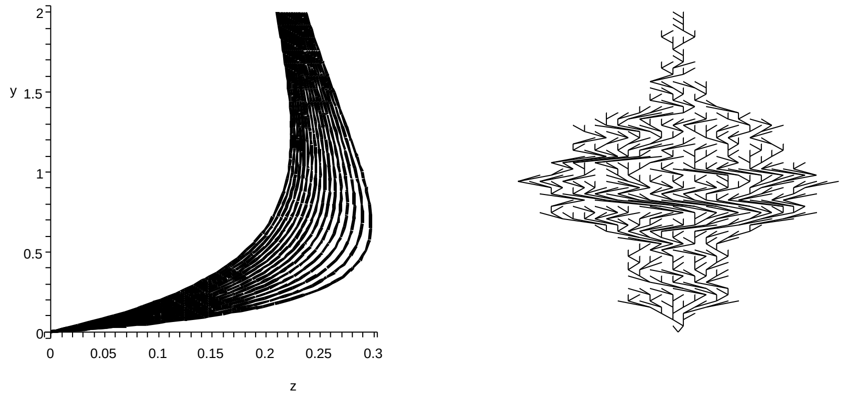


FIGURE 5. Left: The bivariate generating function $C(z,u)$ enumerating binary trees by size and number of leaves exhibits consistently a square-root singularity as function of z for several values of u. Right: a binary tree of size 300 drawn uniformly at random has 69 leaves or "cherries".

of u, to the effect that $C(z,u)$ is of the global analytic type

$$\lambda(u) \cdot \sqrt{1 - \frac{z}{\rho(u)}},$$

for some analytic $\lambda(u)$ and $\rho(u)$. The already evoked process of singularity analysis then shows that the probability generating function of the number of leaves in a tree of size n satisfies an approximation of the form

$$\left(\frac{\lambda(u)}{\lambda(1)}\right) \cdot \left(\frac{\rho(1)}{\rho(u)}\right)^n (1 + o(1)).$$

This "quasi-powers" approximation thus resembles very much the probability generating function of a sum of n independent random variables, a situation that resorts to the classical Central Limit Theorem of probability theory. Accordingly, *the limit distribution of the number of leaves in a large tree is Gaussian*. In abstract terms, the deformation induced by the secondary parameter (here, the number of leaves, previously, the number of runs) is susceptible to a *perturbation analysis*, to the effect that a singularity gets smoothly displaced without changing its nature (here, a square root singularity, earlier a pole) and a limit law systematically results. Again some of the conclusions can be verified even by very small samples: the single tree of size 300 drawn at random and displayed in Figure 5 has 69 cherries while the expected value of this number is $\doteq 75.375$ and the standard deviation is a little over 4. In a large number of cases of which this one is typical, we find *metric laws* of combinatorial structures that govern large structures with high probability and eventually make them highly predictable.

Such randomness properties form the subject of Part C of this book dedicated to *random structures*. As our earlier description implies, there is an extreme degree of

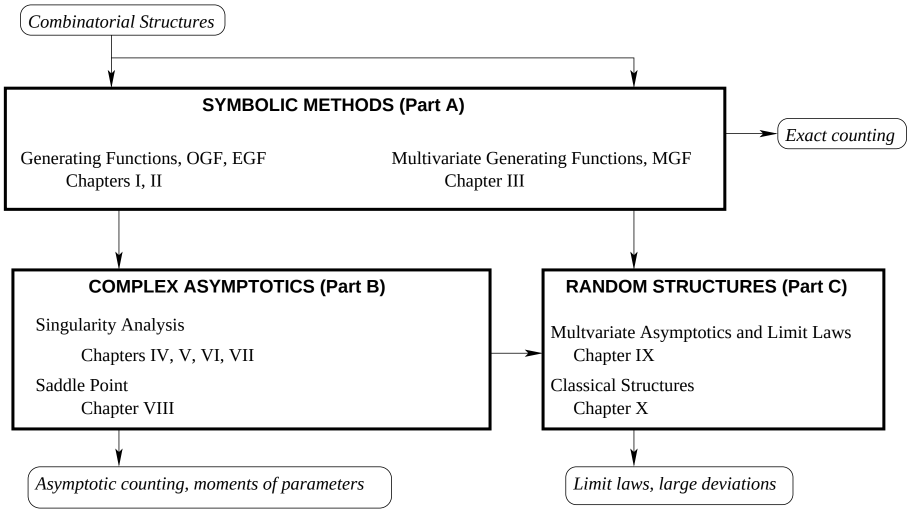


FIGURE 6. The logical structure of *Analytic Combinatorics*.

generality in this analytic approach to combinatorial parameters, and after reading this book, the reader will be able to recognize by herself dozens of such cases at sight, and effortlessly establish the corresponding theorems.

A RATHER ABSTRACT VIEW of combinatorics emerges from the previous discussion; see Figure 6. A combinatorial class, as regards its enumerative properties, can be viewed as a *surface in four-dimensional real space*: this is the graph of its generating function, considered as a function from the set $\mathbb{C} \cong \mathbb{R}^2$ of complex numbers to itself, and is otherwise known as a Riemann surface. This surface has "cracks", that is, *singularities*, which determine the asymptotic behaviour of the counting sequence. A combinatorial construction (like forming freely sequences, sets, and so on) can then be examined based on the effect it has on singularities. In this way, seemingly different types of combinatorial structures appear to be subject to *common laws* governing not only counting but also finer characteristics of combinatorial structures. For the already discussed case of universality in tree enumerations, additional universal laws valid across many tree varieties constrain for instance height (which, with high probability, is proportional to the square-root of size) and the number of leaves (which is invariably normal in the asymptotic limit).

Next, the probabilistic behaviour of a parameter of a combinatorial class is fully determined by a bivariate generating function, which is a deformation of the basic counting generating function of the class. (In the sense that setting the secondary variable u to 1 erases the information relative to the parameter and leads back to the univariate counting generating function). Then, the *asymptotic distribution* of a parameter of interest is characterized by a collection of surfaces, each having its own singularities. The way the singularities' locations move or their nature changes under deformation encodes all the necessary information regarding the distribution of the parameter under consideration. Limit laws for combinatorial parameters can then be obtained and the corresponding phenomena can be organized into broad categories, called *schemas*. It would not be conceivable to attain such a far-reaching classification of metric properties of combinatorial structures by elementary real analysis alone.

OBJECTS to which we are going to inflict the treatments just described include many of the most important ones of discrete mathematics, also the ones that surface recurrently in several branches of the applied sciences. We shall thus encounter words and sequences, trees and lattice paths, graphs of various sorts, mappings, allocations, permutations, integer partitions and compositions, and planar maps, to name a few. In most cases, their principal characteristics will be finely quantified by the methods of analytic combinatorics; see our concluding Chapter **??** for a summary. This book indeed develops a coherent theory of random combinatorial structures based on a powerful analytic methodology. Literally dozens of quite diverse combinatorial types can then be treated by a logically transparent chain. You will not find ready-made answers to all questions in this book, but, hopefully, *methods* that can be successfully used to address a great *many* of them.

Part A

SYMBOLIC METHODS

I

Combinatorial Structures and Ordinary Generating Functions

Laplace discovered the remarkable correspondence between set theoretic operations and operations on formal power series and put it to great use to solve a variety of combinatorial problems.
— GIAN–CARLO ROTA [**365**]

Contents

This chapter and the next are devoted to enumeration, where the problem is to determine the number of combinatorial configurations described by finite rules, and do so for all possible sizes. For instance, how many different words are there of length 17? of length n, for general n? These questions are easy, but what if some constraints are imposed, e.g., no four identical elements in a row? The counting sequences are exactly encoded by *generating functions*, and, as we shall see, *generating functions are the central mathematical object* of combinatorial analysis. We examine here a framework that, contrary to traditional treatments based on recurrences, explains the surprising efficiency of generating functions in the solution of combinatorial enumeration problems.

This chapter serves to introduce the *symbolic* approach to combinatorial enumerations. The principle is that many general set-theoretic *constructions* admit a direct translation as operations over generating functions. This principle is made concrete by means of a dictionary that includes a collection of core constructions, namely the operations of union, cartesian product, sequence, set, multiset, and cycle. Supplementary operations like pointing and substitution can be also be similarly translated. In this way, a *language* describing elementary combinatorial classes is defined. The problem of enumerating a class of combinatorial structures then simply reduces to finding a proper *specification*, a sort of program for the class expressed in terms of the basic constructions. The translation into generating functions then becomes a purely mechanical symbolic process.

We show here how to describe integer partitions and compositions in such a context, as well as several basic string and tree enumeration problems. A parallel approach, developed in Chapter II, applies to labelled objects and exponential generating functions—in contrast the plain structures considered in this chapter are called *unlabelled*. The methodology is susceptible to multivariate extensions with which many characteristic parameters of combinatorial objects can also be analysed in a unified manner: this is to be examined in Chapter III. The symbolic method also has the great merit of connecting nicely with complex asymptotic methods that exploit analyticity properties and singularities, to the effect that precise asymptotic estimates are usually available whenever the symbolic method applies—a systematic treatment of these aspects forms the basis of Part B of this book *Complex Asymptotics* (Chapters IV–VIII).

I. 1. Symbolic enumeration methods

First and foremost, combinatorics deals with *discrete objects*, that is, objects that can be finitely described by construction rules. Examples are words, trees, graphs, permutations, allocations, functions from a finite set into itself, topological configurations, and so on. A major question is to *enumerate* such objects according to some characteristic parameter(s).

DEFINITION I.1. *A* combinatorial class, *or simply a* class, *is a finite or denumerable set on which a* size *function is defined, satisfying the following conditions:*

(i) *the size of an element is a nonnegative integer;*
(ii) *the number of elements of any given size is finite.*

If $\mathcal{A}$ is a class, the size of an element $\alpha \in \mathcal{A}$ is denoted by $|\alpha|$, or $|\alpha|_{\mathcal{A}}$ in the few cases where the underlying class needs to be made explicit. Given a class $\mathcal{A}$, we consistently let $\mathcal{A}_n$ be the set of objects in $\mathcal{A}$ that have size n and use the same group of letters for the counts $A_n = \mathrm{card}(\mathcal{A}_n)$ (alternatively, also $a_n = \mathrm{card}(\mathcal{A}_n)$). An axiomatic presentation is then as follows: a combinatorial class is a pair $(\mathcal{A}, |\cdot|)$ where $\mathcal{A}$ is at most denumerable and the mapping $|\cdot| \in (\mathcal{A} \mapsto \mathbb{N})$ is such that the inverse image of any integer is finite.

DEFINITION I.2. *The* counting sequence *of a combinatorial class $\mathcal{A}$ is the sequence of integers $(A_n)_{n\geq 0}$ where $A_n = \mathrm{card}(\mathcal{A}_n)$ is the number of objects in class $\mathcal{A}$ that have size n.*

EXAMPLE 1. *Binary words.* Consider first the set $\mathcal{W}$ of binary words, which are words over the binary alphabet $\mathcal{A} = \{\mathsf{0},\mathsf{1}\}$,

$$\mathcal{W} := \{\varepsilon, \mathsf{0}, \mathsf{1}, \mathsf{00}, \mathsf{01}, \mathsf{10}, \mathsf{11}, \mathsf{000}, \mathsf{001}, \mathsf{010}, \dots, \mathsf{1001101}, \dots\},$$

with ε the empty word. Define size to be the number of letters a word comprises. There are two possibilities for each letter and possibilities multiply, so that the counting sequence (W_n) satisfies

$$W_n = 2^n.$$

(This sequence has a well-known legend associated with the invention of the game of chess: the inventor was promised by his king one grain of rice for the first square of the chessboard, two for the second, four for the third, and so on. The king naturally could not deliver the promised 2^{63} grains!) .... END OF EXAMPLE 1. $\square$

EXAMPLE 2. *Permutations.* The set $\mathcal{P}$ of permutations is

$$\mathcal{P} = \{\ldots 12, 21, 123, 132, 213, 231, 312, 321, 1234, \ldots, 532614, \ldots\},$$

since a permutation of $I_n := [1\,..\,n]$ is a bijective mapping that is representable by an an array,

$$\begin{pmatrix} 1 & 2 & & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix},$$

or equivalently by the sequence $\sigma_1\sigma_2\cdots\sigma_n$ of distinct elements from $\mathcal{I}_n$. Let us define the size of a permutation to be its length, n. For a permutation written as a sequence of n distinct numbers, there are n places where one can accommodate n, then $n-1$ remaining places for $n-1$, and so on. Therefore, the number P_n of permutations of size n satisfies

$$P_n = n! = 1 \cdot 2 \cdots n\,.$$

As indicated in our Invitation chapter, this formula has been known for a long time: Knuth [**269**, p. 23] refers to the Hebrew *Book of Creation* (c. A.D.. 400), and to the *Anuyogadvārasutra* (India, c. A.D. 500) for its discovery. .... END OF EXAMPLE 2. □

EXAMPLE 3. *Triangulations.* The class $\mathcal{T}$ of triangulations comprises triangulations of convex polygonal domains which are decompositions into non-overlapping triangles (taken up to continuous deformations of the plane). Let us define the size of a triangulation to be the number of triangles it is composed of. For the purpose of the present discussion, the reader may content herself with what is suggested by Figure 1; the formal specification of triangulations appears on p. 33. It is a nontrivial combinatorial result due to Euler and Segner around 1750 that the number T_n of triangulations is

$$T_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!}. \tag{1}$$

Following Euler, the counting of triangulations (T_n) is best approached by generating functions: the modified binomial coefficients so obtained are known as Catalan numbers (see the discussion p. 33) and are central in combinatorial analysis (Section I. 5.3). END OF EXAMPLE 3. □

Although the previous three examples are simple enough, it is generally a good idea, when confronted with a combinatorial enumeration problem, to determine the initial values of counting sequences, either by hand or better with the help of a computer, somehow. Here, we find:

(2)

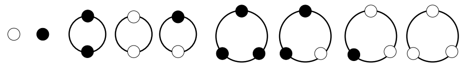
n	0	1	2	3	4	5	6	7	8	9	10
W_n	1	2	4	8	16	32	64	128	256	512	1024
P_n	1	1	2	6	24	120	720	5040	40320	362880	3628800
T_n	1	1	2	5	14	42	132	429	1430	4862	16796

Such an experimental approach may greatly help identify sequences. For instance, had we not known the formula (1) for triangulations, observing an unusual factorization like

$$T_{40} = 2^2 \cdot 5 \cdot 7^2 \cdot 11 \cdot 23 \cdot 43 \cdot 47 \cdot 53 \cdot 59 \cdot 61 \cdot 67 \cdot 71 \cdot 73 \cdot 79,$$

which contains all prime numbers from 43 to 79, would quickly put us on the tracks of the right formula. There even exists nowadays a huge *Encyclopedia of Integer Sequences* due to Sloane that is available in electronic form [**384**] (see also an earlier book by Sloane and Plouffe [**385**]). Indeed, the three sequences (W_n), (P_n), and (T_n) are respectively identified[1] as *EIS* **A000079**, *EIS* **A000142**, and *EIS* **A000108**.

$\triangleright$ **1.** *Necklaces.* How many different types of necklace designs can you form with n beads, each having one of two colours, $\circ$ and $\bullet$? Here are the possibilities for $n = 1, 2, 3$,



and it is postulated that orientation matters. This is equivalent to enumerating circular arrangements of two letters and an exhaustive listing program can be based on the smallest lexicographical represent of each word, as suggested by (17) below. The counting sequence starts as $2, 3, 4, 6, 8, 14, 20, 36, 60, 108, 188, 352$ and constitutes *EIS* **A000031**. [An explicit formula appears later in this chapter (p. 60).] What if two necklace designs that are mirror image of one another are identified? $\triangleleft$

$\triangleright$ **2.** *Unimodal permutations.* Such a permutation has exactly one local maximum. In other words it is of the form $\sigma_1 \cdots \sigma_n$ with $\sigma_1 < \sigma_2 < \cdots < \sigma_k = n$ and $\sigma_k = n > \sigma_{k+1} > \cdots > \sigma_n$, for some $k \geq 1$. How many such permutations are there of size n? For $n = 5$, the number is 16: the permutations are 12345, 12354, 12453, 12543, 13452, 13542, 14532 and 15432 and their reversals. [Due to Jon Perry, see *EIS* **A000079**.] $\triangleleft$

It is also of interest to note that words and permutations could be enumerated using the most elementary counting principles, namely, for finite sets $\mathcal{B}$ and $\mathcal{C}$

$$(3)\qquad \begin{cases} \operatorname{card}(\mathcal{B} \cup \mathcal{C}) = \operatorname{card}(B) + \operatorname{card}(C) & \text{(provided } \mathcal{B} \cap \mathcal{C} = \emptyset) \\ \operatorname{card}(\mathcal{B} \times \mathcal{C}) = \operatorname{card}(B) \cdot \operatorname{card}(C), & \end{cases}$$

We shall see soon that these principles, which lie at the basis of our very concept of number, admit a powerful generalization (Equation (16) below).

Next, for combinatorial enumeration purposes, it proves convenient to identify combinatorial classes that are merely variant of one another.

DEFINITION I.3. *Two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$ are said to be (combinatorially)* isomorphic, *which is written $\mathcal{A} \cong \mathcal{B}$, iff their counting sequences are identical. This condition is equivalent to the existence of a bijection from $\mathcal{A}$ to $\mathcal{B}$ that preserves size, and one also says that $\mathcal{A}$ and $\mathcal{B}$ are* bijectively equivalent.

We normally identify isomorphic classes and accordingly employ a plain equality sign ($\mathcal{A} = \mathcal{B}$). We then confine the notation $\mathcal{A} \cong \mathcal{B}$ to stress cases where combinatorial isomorphism results some nontrivial transformation.

DEFINITION I.4. *The* ordinary generating function *(OGF) of a sequence (A_n) is the formal power series*

$$(4)\qquad A(z) = \sum_{n=0}^{\infty} A_n z^n.$$

The ordinary generating function *(OGF) of a combinatorial class $\mathcal{A}$ is the generating function of the numbers $A_n = \operatorname{card}(\mathcal{A}_n)$. Equivalently, the OGF of class $\mathcal{A}$ admits*

[1]Throughout this book, a reference like *EIS* **Axxx** points to Sloane's *Encyclopedia of Integer Sequences* [**384**]. The data base contains more than 100,000 entries.

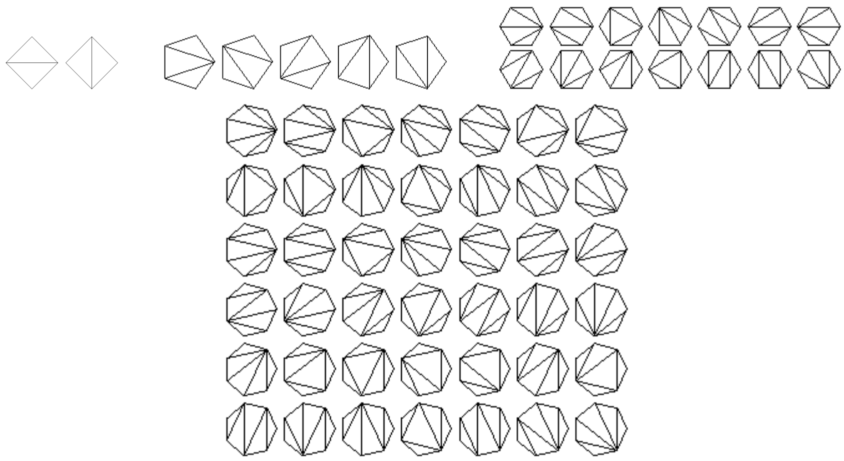


FIGURE 1. The class $\mathcal{T}$ of all triangulations of regular polygons (with size defined as the number of triangles) is a combinatorial class. The counting sequence starts as

$$T_0 = 1,\ T_1 = 1,\ T_2 = 2,\ T_3 = 5,\ T_4 = 14,\ T_5 = 42.$$

Euler determined the OGF $T(z) = \sum_n T_n z^n$ as $T(z) = \dfrac{1-\sqrt{1-4z}}{2z}$, from which there results that $T_n = \frac{1}{n+1}\binom{2n}{n}$. These numbers are known as the *Catalan numbers* (p. 33).

the combinatorial form

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|}. \tag{5}$$

It is also said that the variable z marks *size in the generating function.*

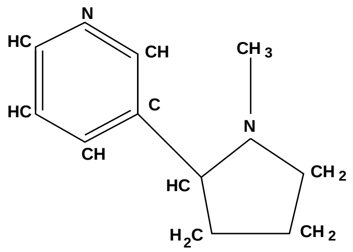
The combinatorial form of an OGF in (5) results straightforwardly from observing that the term z^n occurs as many times as there are objects in $\mathcal{A}$ having size n.

Naming convention. We adhere to a systematic *naming convention*: classes, their counting sequences, and their generating functions are systematically denoted by the same groups of letters: for instance, $\mathcal{A}$ for a class, $\{A_n\}$ (or $\{a_n\}$) for the counting sequence, and $A(z)$ (or $a(z)$) for its OGF.

Coefficient extraction. We let generally $[z^n]f(z)$ denote the operation of extracting the coefficient of z^n in the formal power series $f(z) = \sum f_n z^n$, so that

$$[z^n]\left(\sum_{n\geq 0} f_n z^n\right) = f_n. \tag{6}$$

(The coefficient extractor $[z^n]f(z)$ reads as "coefficient of z^n in $f(z)$".)



$$\Longrightarrow \quad C_{10}H_{14}N_2 \quad \leadsto \quad z^{26}$$

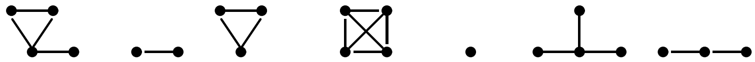
FIGURE 2. A molecule, methylpyrrolidinyl-pyridine (nicotine), is a complex assembly whose description can be reduced to a single formula corresponding here to a total of 26 atoms.

The OGFs corresponding to our three examples $\mathcal{W}, \mathcal{P}, \mathcal{T}$ are then

$$(7) \qquad \begin{cases} W(z) &= \displaystyle\sum_{n=0}^{\infty} 2^n z^n &= \dfrac{1}{1-2z} \\ P(z) &= \displaystyle\sum_{n=0}^{\infty} n!\, z^n & \\ T(z) &= \displaystyle\sum_{n=0}^{\infty} \frac{1}{n+1}\binom{2n}{n} z^n &= \dfrac{1-\sqrt{1-4z}}{2z}. \end{cases}$$

The first expression relative to $W(z)$ is immediate as it is the sum of a geometric progression; The second generating function $P(z)$ is not related to simple functions of analysis. (Note that the expression makes sense within the strict framewok of formal power series; see APPENDIX A: *Formal power series*, p. 648.) The third expression relative to $T(z)$ is equivalent to the explicit form of T_n via Netwon's expansion of $(1+x)^{1/2}$ (p. 33). The OGFs $W(z)$ and $T(z)$ can then also be interpreted as standard analytic objects, upon assigning to the formal variable z values in the complex domain $\mathbb{C}$. In effect, the series $W(z)$ and $T(z)$ converge in a neighbourhood of 0 and represent complex functions that are well defined near the origin, namely when $|z| < \frac{1}{2}$ for $W(z)$ and $|z| < \frac{1}{4}$ for $T(z)$. The OGF $P(z)$ is a purely formal power series (its radius of convergence is 0) that can nonetheless be subjected to the usual algebraic operations of power series. As a matter of fact, with very few exceptions, permutation enumeration is most conveniently approached by exponential generating functions developed in Chapter II.

Combinatorial form of GFs. The combinatorial form (5) shows that generating functions are nothing but a reduced representation of the combinatorial class, where internal structures are destroyed and elements contributing to size (atoms) are replaced by the variable z. In a sense, this is analogous to what chemists do by writing linear reduced formulæ for complex molecules (Figure 2). Great use of this observation was made by Schützenberger as early as the 1950's and 1960's. It explains in many ways why so many formal similarities are to be found between combinatorial structures and generating functions.

$\mathcal{H}$ = 

$$\begin{array}{ccccccc} zzzz & zz & zzz & zzzz & z & zzzz & zzz \\ +z^4 & +z^2 & +z^3 & +z^4 & +z & +z^4 & +z^3 \end{array}$$

$$H(z) = z + z^2 + 2z^3 + 3z^4$$

FIGURE 3. A finite family of graphs and its eventual reduction to a generating function.

Figure 3 provides a combinatorial illustration: start with a (finite) family of graphs $\mathcal{H}$, with size taken as the number of vertices. Each vertex in each graph is replaced by the variable z and the graph structure is “forgotten”; then the monomials corresponding to each graph are formed and the generating function is finally obtained by gathering all the monomials. For instance, there are 3 graphs of size 4 in $\mathcal{H}$, in agreement with the fact that $[z^4]H(z) = 3$. If size had been instead defined by number of edges, another generating function would have resulted, namely, with y marking the new size: $1 + y + y^2 + 2y^3 + y^4 + y^6$. If both number of vertices and number of edges are of interest, then a bivariate generating function, $H(z, y) = z + z^2y + z^3y^2 + z^3y^3 + z^4y^3 + z^4y^4 + z^4y^6$; such multivariate generating functions are developed systematically in Chapter III.

A path often taken in the literature is to decompose the structures to be enumerated into smaller structures either of the same type or of simpler types, and then extract from such a decomposition *recurrence relations* satisfied by the $\{A_n\}$. In this context, the recurrence relations are either solved directly—whenever they are simple enough—or by means of *ad hoc* generating functions, introduced as a mere technical artifice.

By contrast, in the framework to be described, classes of combinatorial structures are built *directly* in terms of simpler classes by means of a collection of elementary combinatorial *constructions*. (This closely resembles the description of formal languages by means of grammars, as well as the construction of structured data types in programming languages.) The approach developed here has been termed *symbolic*, as it relies on a formal specification language for combinatorial structures. Specifically, it is based on so–called *admissible constructions* that admit direct translations into generating functions.

DEFINITION I.5. *Assume that Φ is a construction that associates to a finite collection of classes $\mathcal{B}, \mathcal{C}, \cdots$ a new class*

$$\mathcal{A} := \Phi[\mathcal{B}, \mathcal{C}, \ldots],$$

in a finitary way: each $\mathcal{A}_n$ depends on finitely many of the $\{\mathcal{B}_j\}, \{\mathcal{C}_j\}, \ldots$. Then Φ is admissible *iff the counting sequence $\{A_n\}$ of $\mathcal{A}$ only depends on the counting sequences $\{B_j\}, \{C_j\}, \ldots$ of $\mathcal{B}, \mathcal{C}, \ldots$, and for some operator Ξ on sequences:*

$$\{A_n\} = \Xi[\{B_j\}, \{C_j\}, \ldots].$$

In that case, since generating functions are determined by their coefficient sequences, there exists a well defined operator Ψ translating Ξ on the associated ordinary generating functions

$$A(z) = \Psi[B(z), C(z), \ldots].$$

As an introductory example, take the construction of cartesian product.

DEFINITION I.6. *The cartesian product construction of two classes $\mathcal{A}$ and $\mathcal{B}$ forms ordered pairs,*

$$\mathcal{A} = \mathcal{B} \times \mathcal{C} \quad \textit{iff} \quad \mathcal{A} = \{\, \alpha = (\beta, \gamma) \mid \beta \in \mathcal{B},\ \gamma \in \mathcal{C}\,\}, \tag{8}$$

with the size of a pair $\alpha = (\beta, \gamma)$ being defined by

$$|\alpha|_{\mathcal{A}} = |\beta|_{\mathcal{B}} + |\gamma|_{\mathcal{C}}. \tag{9}$$

By considering all possibilities, it is immediately seen that the counting sequences corresponding to $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are related by the convolution relation

$$A_n = \sum_{k=0}^{n} B_k C_{n-k}. \tag{10}$$

We recognize here the formula for a product of two power series. Therefore,

$$A(z) = B(z) \cdot C(z). \tag{11}$$

Thus, the cartesian product is admissible: *A cartesian product translates as a product of OGFs.*

Similarly, let $\mathcal{A}, \mathcal{B}, \mathcal{C}$ be combinatorial classes satisfying

$$\mathcal{A} = \mathcal{B} \cup \mathcal{C}, \qquad \text{with} \quad \mathcal{B} \cap \mathcal{C} = \emptyset, \tag{12}$$

with size defined in a consistent manner: for $\alpha \in \mathcal{A}$,

$$|\omega|_{\mathcal{A}} = \begin{cases} |\omega|_{\mathcal{B}} & \text{if } \omega \in \mathcal{B} \\ |\omega|_{\mathcal{C}} & \text{if } \omega \in \mathcal{C}. \end{cases} \tag{13}$$

One has

$$A_n = B_n + C_n, \tag{14}$$

which, at generating function level, means

$$A(z) = B(z) + C(z). \tag{15}$$

Thus, *a union of sets translates as a sum of generating functions provided the sets are* disjoint.

The correspondences provided by (8)–(11) and (12)–(15) are summarized by the dictionary

$$\begin{cases} \mathcal{A} = \mathcal{B} \cup \mathcal{C} & \Longrightarrow \quad A(z) = B(z) + C(z) \qquad \text{(provided } \mathcal{B} \cap \mathcal{C} = \emptyset\text{)} \\ \mathcal{A} = \mathcal{B} \times \mathcal{C} & \Longrightarrow \quad A(z) = B(z) \cdot C(z) \end{cases} \tag{16}$$

(Compare with the plain arithmetic case of (3).) Their merit is that they can be stated as general-purpose translation rules that only need to be established once and for all.

As soon as the problem of counting elements of a union of disjoint sets or a cartesian product is recognized, it becomes possible to dispense altogether with the intermediate stages of writing explicitly coefficient relations or recurrences like in (10) or (14). This is the spirit of the symbolic method for combinatorial enumerations. Its interest lies in the fact that several powerful set-theoretic constructions are amenable to such a treatment.

I. 2. Admissible constructions and specifications

The main goal of this section is to introduce formally the basic *constructions* that constitute the core of a specification language for combinatorial structures. This core is based on disjoint unions, also known as combinatorial sums, and on Cartesian products that we have just discussed. We shall augment it by the constructions of sequence, cycle, multiset, and powerset. A class is *constructible* or *specifiable* if it can be defined from primal elements by means of these constructions. The generating function of any such class satisfies functional equations that can be transcribed systematically from a specification; see Theorems I.1 and I.2, as well as Figure 14 at the end of this chapter for a summary.

I. 2.1. Basic constructions. First, we assume given a class $\mathcal{E}$ called the *neutral class* that consists of a single object of size 0; any such an object of size 0 is called a *neutral object*. and is usually denoted by symbols like ϵ or $\mathbf{1}$. The reason for this terminology becomes clear if one considers the combinatorial isomorphism

$$\mathcal{A} \cong \mathcal{E} \times \mathcal{A} \cong \mathcal{A} \times \mathcal{E}.$$

We also assume as given an *atomic class* $\mathcal{Z}$ comprising a single element of size 1; any such element is called an atom; an atom may be used to describe a generic node in a tree or graph, in which case it may be represented by a circle ($\bullet$ or $\circ$), but also a generic letter in a word, in which case it may be instantiated as $a, b, c, \ldots$. Distinct copies of the neutral or atomic class may also be subscripted by indices in various ways. Thus, for instance we use the classes $\mathcal{Z}_a = \{a\}$, $\mathcal{Z}_b = \{b\}$ (with a, b of size 1) to build up binary words over the alphabet $\{a, b\}$, or $\mathcal{Z}_\bullet = \{\bullet\}$, $\mathcal{Z}_\circ = \{\circ\}$ (with $\bullet, \circ$ taken to be of size 1) to build trees with nodes of two coulurs. Similarly, we introduce $\mathcal{E}_\square, \mathcal{E}_1, \mathcal{E}_2$ to denote a class comprising the neutral objects $\square, \epsilon_1, \epsilon_2$ respectively.

Clearly, the generating functions of a neutral class $\mathcal{E}$ and an atomic class $\mathcal{Z}$ are

$$E(z) = 1, \qquad Z(z) = z,$$

corresponding to the unit 1, and the variable z, of generating functions.

Combinatorial sum (disjoint union). First consider *combinatorial sum* also known as *disjoint union*. The intent is to capture the union of disjoint sets, but without the constraint of any extraneous condition of disjointness. We formalize the (combinatorial) sum of two classes $\mathcal{B}$ and $\mathcal{C}$ as the union (in the standard set–theoretic sense) of two *disjoint* copies, say $\mathcal{B}^\square$ and $\mathcal{C}^\diamondsuit$, of $\mathcal{B}$ and $\mathcal{C}$. A picturesque way to view the construction is as follows: first choose two distinct colours and repaint the elements of $\mathcal{B}$ with the $\square$-colour and the elements of $\mathcal{C}$ with the $\diamondsuit$-colour. This is made precise by introducing two distinct "markers" $\square$ and $\diamond$, each a neutral object (*i.e.*, of size zero); the disjoint union $\mathcal{B}+\mathcal{C}$ of $\mathcal{B},\mathcal{C}$ is then defined as the standard set-theoretic union,

$$\mathcal{B}+\mathcal{C} := (\{\square\}\times\mathcal{B}) \cup (\{\diamond\}\times\mathcal{C}).$$

The size of an object in a disjoint union $\mathcal{A}=\mathcal{B}+\mathcal{C}$ is by definition inherited from its size in its class of origin, like in Equation (13). One good reason behind the definition adopted here is that the combinatorial sum of two classes is *always* well-defined. Furthermore, disjoint union is equivalent to a standard union whenever it is applied to disjoint sets.

Because of disjointness, one has the implication

$$\mathcal{A}=\mathcal{B}+\mathcal{C} \quad\Longrightarrow\quad A_n=B_n+C_n \quad\Longrightarrow\quad A(z)=B(z)+C(z),$$

so that disjoint union is admissible. Note that, in contrast, standard set-theoretic union is not an admissible construction since

$$\operatorname{card}(\mathcal{B}_n\cup\mathcal{C}_n)=\operatorname{card}(\mathcal{B}_n)+\operatorname{card}(\mathcal{C}_n)-\operatorname{card}(\mathcal{B}_n\cap\mathcal{C}_n),$$

and information on the internal structure of $\mathcal{B}$ and $\mathcal{C}$ (i.e., the nature of this intersection) is needed in order to be able to enumerate the elements of their union.

Cartesian product. This construction $\mathcal{A}=\mathcal{B}\times\mathcal{C}$ forms all possible ordered pairs in accordance with Definition I.6. The size of a pair is obtained additively from the size of components in accordance with (9).

Next, we introduce a few fundamental constructions that build upon set-theoretic union and product, and form sequences, sets, and cycles. These powerful constructions suffice to define a broad variety of combinatorial structures.

Sequence construction. If $\mathcal{C}$ is a class then the *sequence* class $\mathrm{SEQ}(\mathcal{C})$ is defined as the infinite sum

$$\mathrm{SEQ}(\mathcal{C})=\{\epsilon\}+\mathcal{C}+(\mathcal{C}\times\mathcal{C})+(\mathcal{C}\times\mathcal{C}\times\mathcal{C})+\cdots$$

with ϵ being a neutral structure (of size 0). (The neutral structure in this context plays a rôle similar to that of the "empty" word in formal language theory, while the sequence construction is somewhat analogous to the Kleene star operation ('$\star$'); see APPENDIX A: *Regular languages*, p. 650.) It is then readily checked that the construction $\mathcal{A}=\mathrm{SEQ}(\mathcal{C})$ defines a proper class satisfying the finiteness condition for sizes if and only if $\mathcal{C}$ contains no object of size 0. From the definition of size for sums and products, there results that the size of a sequence is to be taken as the sum of the sizes of its components:

$$\gamma=(\alpha_1,\ldots,\alpha_\ell) \qquad\Longrightarrow\qquad |\gamma|=|\alpha_1|+\cdots+|\alpha_\ell|.$$

Cycle construction. Sequences taken up to a circular shift of their components define cycles, the notation being $\mathrm{CYC}(\mathcal{B})$. Precisely, one has

$$\mathrm{CYC}(\mathcal{B}) := \mathrm{SEQ}(\mathcal{B})/\mathbf{S},$$

where $\mathbf{S}$ is the equivalence relation between sequences defined by

$$(\alpha_1,\ldots,\alpha_r)\,\mathbf{S}\,(\beta_1,\ldots,\beta_r)$$

iff there exists some *circular shift* τ of $[1\,..\,n]$ such that for all j, $\beta_j=\alpha_{\tau(j)}$; in other words, for some d, one has $\beta_j=\alpha_{1+(j+d) \bmod n}$. Here is for instance a depiction of the cycles formed from the 8 and 16 sequences of lengths 3 and 4 over two types of objects (a, b): the number of cycles is 4 (for $n = 3$) and 6 (for $n = 4$). Sequences are grouped into equivalence classes according to the relation $\mathbf{S}$.

$$
\begin{array}{c}
aaa \\
aab\ aba\ baa \\
abb\ bba\ bab \\
bbb
\end{array}
\qquad
\begin{array}{c}
aaaa \\
aaab\ aaba\ abaa\ baaa \\
aabb\ abba\ bbaa\ baab \\
abab\ baba \\
abbb\ bbba\ bbab\ babb \\
bbbb
\end{array}
\tag{17}
$$

According to the definition, this construction corresponds to the formation of directed cycles. We make only a limited use of it for unlabelled objects; however, its counterpart plays a rather important rôle in the context of labelled structures and exponential generating functions.

Multiset construction. Following common mathematical terminology, *multisets* are like finite sets (that is the order between element does not count), but arbitrary repetitions of elements are allowed. The notation is $\mathcal{A} = \text{MSET}(\mathcal{B})$ when $\mathcal{A}$ is obtained by forming all *finite* multisets of elements from $\mathcal{B}$. The precise way of defining $\text{MSET}(\mathcal{B})$ is as a quotient:

$$\text{MSET}(\mathcal{B}) := \text{SEQ}(\mathcal{B})/\mathbf{R} \quad \text{with} \quad \mathbf{R},$$

the equivalence relation between sequences being defined by $(\alpha_1, \ldots, \alpha_r)\,\mathbf{R}\,(\beta_1, \ldots, \beta_r)$ iff there exists some *arbitrary permutation* σ of $[1 \mathinner{.\,.} n]$ such that for all j, $\beta_j = \alpha_{\sigma(j)}$.

Powerset construction. The *powerset* class (or set class) $\mathcal{A} = \text{SET}(\mathcal{B})$ is defined as the class consisting of all *finite* subsets of class $\mathcal{B}$, or equivalently, as the class $\text{PSET}(\mathcal{B}) \subset \text{MSET}(\mathcal{B})$ formed of multisets that involve no repetitions.

We again need to make explicit the way the size function is defined when such constructions are performed: like for products and sequences, the size of a composite object—set, multiset, or cycle—is defined as the sum of the sizes of its components.

$\triangleright$ **3.** *The semi-ring of combinatorial classes.* Under the convention of identifying isomorphic classes, sum and product acquire pleasant algebraic properties: combinatorial sums and cartesian products become commutative and associative operations, e.g.,

$$(\mathcal{A} + \mathcal{B}) + \mathcal{C} = \mathcal{A} + (\mathcal{B} + \mathcal{C}), \qquad \mathcal{A} \times (\mathcal{B} \times \mathcal{C}) = (\mathcal{A} \times \mathcal{B}) \times \mathcal{C},$$

while distributivity holds, $(\mathcal{A}+\mathcal{B})\times\mathcal{C} = (\mathcal{A}\times\mathcal{C})+(\mathcal{B}\times\mathcal{C})$. The proofs are simple verifications from the definitions. $\triangleleft$

$\triangleright$ **4.** *Natural numbers.* Let $\mathcal{Z} := \{\bullet\}$ with $\bullet$ an atom (of size 1). Then $\mathcal{I} = \text{SEQ}(Z) \setminus \{\epsilon\}$ is a way of describing natural integers in unary notation: $\mathcal{I} = \{\bullet,\ \bullet\bullet,\ \bullet\bullet\bullet, \ldots\}$. The corresponding OGF is $I(z) = z/(1-z) = z + z^2 + z^3 + \cdots$. $\triangleleft$

$\triangleright$ **5.** *Interval coverings.* Let $\mathcal{Z} := \{\bullet\}$ be as before. Then $\mathcal{A} = \mathcal{Z} + (\mathcal{Z} \times \mathcal{Z})$ is a set of two elements, $\bullet$ and $(\bullet, \bullet)$, which we choose to draw as $\{\bullet,\ \bullet\!\!-\!\!\bullet\}$. Then $\mathcal{C} = \text{SEQ}(\mathcal{A})$ contains elements like

$$\bullet,\ \bullet\bullet,\ \bullet\!\!-\!\!\bullet,\ \bullet\bullet\!\!-\!\!\bullet,\ \bullet\!\!-\!\!\bullet\bullet,\ \bullet\!\!-\!\!\bullet\bullet\!\!-\!\!\bullet,\ \bullet\ \bullet\ \bullet\ \bullet\ .$$

With the notion of size adopted, the objects of size n in $\mathcal{C} = \text{SEQ}(\mathcal{Z} + (\mathcal{Z} \times \mathcal{Z}))$ are (isomorphic to) the *coverings* of the interval $[0, n]$ by intervals (matches) of length either 1 or 2. The generating function,

$$C(z) = 1 + z + 2\,z^2 + 3\,z^3 + 5\,z^4 + 8\,z^5 + 13\,z^6 + 21\,z^7 + 34\,z^8 + 55\,z^9 + \cdots,$$

is, as we shall see shortly (p. 40), the OGF of Fibonacci numbers. $\triangleleft$

I. 2.2. The admissibility theorem for ordinary generating functions. This section is a formal treatment of admissibility proofs for the constructions we have considered. The final implication is that any specification of a constructible class translates directly into generating function equations. The cycle construction involves the Euler totient function $\varphi(k)$ defined as the number of integers in $[1, k]$ that are relatively prime to k (APPENDIX A: *Arithmetical functions*, p. 639).

THEOREM I.1 (Admissible unlabelled constructions). *The constructions of union, cartesian product, sequence, multiset, powerset, and cycle are all admissible. The associated operators are*

$$
\begin{array}{llcl}
\textit{Sum:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(z) = B(z) + C(z) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \times \mathcal{C} & \Longrightarrow & A(z) = B(z) \cdot C(z) \\
\textit{Sequence:} & \mathcal{A} = \mathrm{SEQ}(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{1 - B(z)} \\
\textit{Cycle:} & \mathcal{A} = \mathrm{CYC}(\mathcal{B}) & \Longrightarrow & A(z) = \displaystyle\sum_{k=1}^{\infty} \frac{\varphi(k)}{k} \log \frac{1}{1 - B(z^k)}. \\
\textit{Multiset:} & \mathcal{A} = \mathrm{MSET}(\mathcal{B}) & \Longrightarrow & A(z) = \begin{cases} \displaystyle\prod_{n \geq 1} (1 - z^n)^{-B_n} \\ \exp\left(\displaystyle\sum_{k=1}^{\infty} \frac{1}{k} B(z^k) \right) \end{cases} \\
\textit{Powerset:} & \mathcal{A} = \mathrm{PSET}(\mathcal{B}) & \Longrightarrow & A(z) = \begin{cases} \displaystyle\prod_{n \geq 1} (1 + z^n)^{B_n} \\ \exp\left(\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} B(z^k) \right) \end{cases}
\end{array}
$$

The sequence, cycle, and set translations necessitate that $\mathcal{B}_0 = \emptyset$.

The class $\mathcal{E} = \{\epsilon\}$ consisting of the neutral object only, and the class $\mathcal{Z}$ consisting of a single "atomic" object (node, letter) of size 1 have OGFs

$$E(z) = 1 \qquad \text{and} \qquad Z(z) = z.$$

PROOF. The proof proceeds by cases, building upon what we have just seen regarding unions and products.

Combinatorial sum (disjoint union). Let $\mathcal{A} = \mathcal{B} + \mathcal{C}$. Since the union is *disjoint*, and the size of an $\mathcal{A}$–element coincides with its size in $\mathcal{B}$ or $\mathcal{C}$, one has $A_n = B_n + C_n$ and $A(z) = B(z) + C(z)$, as discussed earlier. The rule also follows directly from the combinatorial form of generating functions as expressed by (5):

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|} = \sum_{\alpha \in \mathcal{B}} z^{|\alpha|} + \sum_{\alpha \in \mathcal{C}} z^{|\alpha|} = B(z) + C(z).$$

Cartesian Product. The admissibility result for $\mathcal{A} = \mathcal{B} \times \mathcal{C}$ was considered as an example for Definition I.6, the convolution equation (10) leading to the relation $A(z) = B(z) \cdot C(z)$. We can offer a direct derivation based on the combinatorial form of generating functions (5),

$$A(z) = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|} = \sum_{(\beta,\gamma)\in(\mathcal{B}\times\mathcal{C})} z^{|\beta|+|\gamma|} = \left(\sum_{\beta\in\mathcal{B}} z^{|\beta|}\right)\times\left(\sum_{\gamma\in\mathcal{C}} z^{|\gamma|}\right) = B(z)\cdot C(z),$$

as follows from distributing products over sums. This derivation readily extends to an arbitrary number of factors.

Sequence construction. Admissibility for $\mathcal{A} = \text{SEQ}(\mathcal{B})$ (with $\mathcal{B}_0 = \emptyset$) follows from the union and product relations. One has

$$\mathcal{A} = \{\epsilon\} + \mathcal{B} + (\mathcal{B}\times\mathcal{B}) + (\mathcal{B}\times\mathcal{B}\times\mathcal{B}) + \cdots,$$

so that

$$A(z) = 1 + B(z) + B(z)^2 + B(z)^3 + \cdots = \frac{1}{1-B(z)},$$

where the geometric sum converges in the sense of formal power series since $[z^0]B(z) = 0$, by assumption.

Powerset construction. Let $\mathcal{A} = \text{PSET}(\mathcal{B})$ and first take $\mathcal{B}$ to be finite. Then, the class $\mathcal{A}$ of all the finite subsets of $\mathcal{B}$ is isomorphic to a product,

$$\text{PSET}(\mathcal{B}) \cong \prod_{\beta\in\mathcal{B}} (\{\epsilon\} + \{\beta\})$$

with ϵ a neutral structure of size 0. Indeed, distributing the products in all possible ways forms all the possible combinations, i.e., sets, of elements of $\mathcal{B}$ with no repetition allowed, by reasoning similar to what leads to such an identity as

$$(1+a)(1+b)(1+c) = 1 + [a+b+c] + [ab+bc+ac] + abc,$$

where all combinations of variables appear. Then, directly from the combinatorial form of generating functions and the sum and product rules, we find

$$A(z) = \prod_{\beta\in\mathcal{B}} (1+z^{|\beta|}) = \prod_n (1+z^n)^{B_n}.$$

The *exp-log transformation* $A(z) = \exp(\log A(z))$ then yields

$$\begin{aligned} A(z) &= \exp\left(\sum_{n=1}^{\infty} B_n \log(1+z^n)\right) \\ &= \exp\left(\sum_{n=1}^{\infty} B_n \cdot \sum_{k=1}^{\infty} (-1)^{k-1}\frac{z^{nk}}{k}\right) \\ &= \exp\left(\frac{B(z)}{1} - \frac{B(z^2)}{2} + \frac{B(z^3)}{3} - \cdots\right), \end{aligned} \tag{18}$$

where the second line results from expanding the logarithm,

$$\log(1+u) = \frac{u}{1} - \frac{u^2}{2} + \frac{u^3}{3} - \cdots,$$

and the third line results from exchanging the order of summation.

The proof finally extends to the case of $\mathcal{B}$ being infinite by noting that each $\mathcal{A}_n$ depends only on those $\mathcal{B}_j$ for which $j \le n$, to which the relations given above for the finite case apply. Precisely, let $\mathcal{B}^{(\leq m)} = \sum_{k=1}^{m} \mathcal{B}_j$ and $\mathcal{A}^{(\leq m)} = \text{SET}(\mathcal{B}^{(\leq m)})$. Then, with $O(z^{m+1})$ denoting any series that has no term of degree $\leq m$, one has

$$A(z) = A^{(\leq m)}(z) + O(z^{m+1}) \qquad \text{and} \qquad B(z) = B^{(\leq m)}(z) + O(z^{m+1}).$$

On the other hand, $A^{(\leq m)}(z)$ and $B^{(\leq m)}(z)$ are connected by the fundamental exponential relation (18) , since $\mathcal{B}^{(\leq m)}$ is finite. Letting m tend to infinity, there follows in the limit

$$A(z) = \exp\left(\frac{B(z)}{1} - \frac{B(z^2)}{2} + \frac{B(z^3)}{3} - \cdots\right).$$

(See APPENDIX A: *Formal power series*, p. 648 for the definition of formal convergence.)

Multiset construction. First for finite $\mathcal{B}$ (with $\mathcal{B}_0 = \emptyset$), the multiset class $\mathcal{A} = \text{MSET}(\mathcal{B})$ is definable by

$$\text{MSET}(\mathcal{B}) \cong \prod_{\beta \in \mathcal{B}} \text{SEQ}(\beta).$$

In words, any multiset can be sorted, in which case it can be viewed as formed of a sequence of repeated elements β_1, followed by a sequence of repeated elements β_2, where $\beta_1, \beta_2, \ldots$ is a canonical listing of the elements of $\mathcal{B}$. The relation translates into generating functions by the product and sequence rules,

$$\begin{aligned} A(z) &= \prod_{\beta \in \mathcal{B}} (1 - z^{|\beta|})^{-1} = \prod_{n=1}^{\infty} (1 - z^n)^{-B_n} \\ &= \exp\left(\sum_{n=1}^{\infty} B_n \log(1 - z^n)^{-1}\right) \\ &= \exp\left(\frac{B(z)}{1} + \frac{B(z^2)}{2} + \frac{B(z^3)}{3} + \cdots\right), \end{aligned}$$

where the exponential form results from the exp-log transformation. The case of an infinite class $\mathcal{B}$ follows by a continuity argument analogous the one used for powersets.

Cycle construction. The translation of the cycle relation $\mathcal{A} = \text{CYC}(\mathcal{B})$ turns out to be

$$A(z) = \sum_{k=1}^{\infty} \frac{\varphi(k)}{k} \log \frac{1}{1 - B(z^k)},$$

where $\varphi(k)$ is the Euler totient function. The first terms, with $L_k(z) := \log(1 - B(z^k))^{-1}$ are

$$A(z) = \frac{1}{1}L_1(z) + \frac{1}{2}L_2(z) + \frac{2}{3}L_3(z) + \frac{2}{4}L_4(z) + \frac{4}{5}L_5(z) + \frac{2}{6}L_6(z) + \cdots.$$

We defer the proof to APPENDIX A: *Cycle construction*, p. 646, since it relies in part on multivariate generating functions to be officially introduced in Chapter III. □

The results for sets, multisets, and cycles are particular cases of the well known *Pólya theory* that deals more generally with the enumeration of objects under group symmetry actions [**347, 349**]. This theory is exposed in many textbooks, see for instance [**82, 223**]. The approach adopted here consists in considering simultaneously all possible values of the number of components by means of bivariate generating functions. Powerful generalizations within the theory of species are presented in the book by Bergeron, Labelle, and Leroux [**37**].

▷ **6.** *Vallée's identity.* Let $\mathcal{M} = \text{MSET}(\mathcal{C}), \mathcal{P} = \text{PSET}(\mathcal{C})$. Separating elements of $\mathcal{C}$ according to the parity of the number of times they appear in a multiset gives rise to the identity

$$M(z) = P(z)M(z^2).$$

(Hint: a multiset contains elements of either odd or even multiplicity.) Accordingly, one can deduce the translation of powersets from the formula for multisets. Iterating the relation above yields $M(z) = P(z)P(z^2)P(z^4)P(z^8)\cdots$, that is closely related to the binary representation of numbers and to Euler's identity (p. 46). It is used for instance in Note 52 p. 83. ◁

Restricted constructions. In order to increase the descriptive power of the framework of constructions, we also want to allow restrictions on the number of components in sequences, sets, multisets, and cycles. Let $\mathfrak{K}$ be a metasymbol representing any of $\text{SEQ}, \text{CYC}, \text{MSET}, \text{PSET}$ and let Ω be a predicate over the integers, then $\mathfrak{K}_\Omega(\mathcal{A})$ will represent the class of objects constructed by $\mathfrak{K}$ but with a number of components constrained to satisfy Ω. Then, the notations

$$\text{SEQ}_{=k} \text{ (or simply } \text{SEQ}_k), \ \text{SEQ}_{>k}, \ \text{SEQ}_{1\,..\,k}$$

refer to sequences whose number of components are exactly k, larger than k, or in the interval $1\,..\,k$ respectively and the same holds for other constructions. In particular,

$$\text{SEQ}_k(\mathcal{B}) := \overbrace{\mathcal{B} \times \cdots \mathcal{B}}^{k \text{ times}} \equiv \mathcal{B}^k, \qquad \text{SEQ}_{\geq k}(\mathcal{B}) = \sum_{j \geq k} \mathcal{B}^j \cong \mathcal{B}^k \times \text{SEQ}(\mathcal{B}),$$
$$\text{MSET}_k(\mathcal{B}) := \text{SEQ}_k(\mathcal{B})/\mathbf{R}.$$

Similarly, $\text{SEQ}_{\text{odd}}, \text{SEQ}_{\text{even}}$ will denote sequences with an odd or even number of components, and so on.

Translations for such restricted constructions are available, as shown generally in Subsection I. 6.1. Suffice it to note for the moment that the construction $\mathcal{A} = \text{SEQ}_k(\mathcal{B})$ is really an abbreviation for a k-fold product, hence it admits the translation into OGFs

$$(19) \qquad \mathcal{A} = \text{SEQ}_k(\mathcal{B}) \qquad \Longrightarrow \qquad A(z) = B(z)^k.$$

I. 2.3. Constructibility and combinatorial specifications. By composing basic constructions, we can build compact descriptions (specifications) of a broad variety of combinatorial classes. Since we restrict attention to *admissible* constructions, we can immediately derive OGFs for these classes. Put differently, the task of enumerating a combinatorial class is reduced to *programming* a specification for it in the language of admissible constructions. In this subsection, we first discuss the expressive power of the language of constructions, then summarize the symbolic method (for unlabelled classes and OGFs) by Theorem I.2.

First, in the framework just introduced, the class of all binary words is described by

$$\mathcal{W} = \text{SEQ}(\mathcal{A}) \qquad \text{where} \qquad \mathcal{A} = \{a, b\} \cong \mathcal{Z} + \mathcal{Z},$$

the ground alphabet, comprises two elements (letters) of size 1. The size of a binary word then coincides with its length (the number of letters it contains). In other words, we start from basic atomic elements and build up words by forming freely all the objects determined by the sequence construction. Such a combinatorial description of a class that only involves a composition of basic constructions applied to initial classes $\mathcal{E}, \mathcal{Z}$ is said to be an *iterative* (or *nonrecursive*) *specification*. Other examples already encountered include binary necklaces (Note 1, p. 18) and the natural integers (Note 4, p. 25) respectively defined by

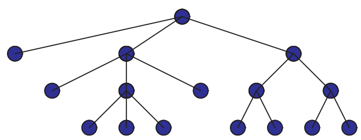
$$\mathcal{N} = \text{CYC}(\mathcal{Z} + \mathcal{Z}) \qquad \text{and} \qquad \mathcal{I} = \text{SEQ}_{\geq 1}(\mathcal{Z}).$$

From there, one can construct ever more complicated objects. For instance,

$$\mathcal{P} = \text{MSET}(\mathcal{I}) \equiv \text{MSET}(\text{SEQ}_{\geq 1}(\mathcal{Z}))$$

means the class of multisets of natural integers, which is isomorphic to the class of integer partitions (see Section I. 3 below for a detailed discussion). As such examples demonstrate, a specification that is iterative can be represented as a single term built on $\mathcal{E}, \mathcal{Z}$ and the constructions $+, \times$, SEQ, CYC, MSET, PSET. An iterative specification can be equivalently listed by naming some of the subterms (for instance partitions in terms of natural integers themselves defined as sequences of atoms).

Semantics of recursion. We next turn our attention to recursive specifications, starting with trees (cf also APPENDIX A: *Tree concepts*, p. 653 for basic definitions). In graph theory, a tree is classically defined as an undirected graph that is connected and acyclic. Additionally, a tree is *rooted* if a particular vertex is distinguished to be the root. Computer scientists commonly make use of trees called *plane* that are rooted but also embedded in the plane, so that the ordering of subtrees attached to any node matters. Here, we will give the name of *general plane trees* to such rooted plane trees and call $\mathcal{G}$ their class, where size is the number of vertices; see, e.g., [**382**]. (The term "general" refers to the fact that all nodes degrees are allowed.) For instance, a general tree of size 16, drawn with the root on top, is:

$\tau =$ 

As a consequence of the definition, if one interchanges, say, the second and third root subtrees, then a different tree results—the original tree and its variant are not homeomorphically equivalent. (General trees are thus comparable to graphical renderings of genealogies where children are ordered by age.). Although we have introduced plane trees as 2-dimensional diagrams, it is obvious that any tree also admits a linear representation: a tree τ with root ζ and root subtrees $\tau_1, \ldots, \tau_r$ (in that order) can be seen as the object $\zeta\,\boxed{\tau_1, \ldots, \tau_r}$, where the box encloses similar representations of subtrees. Typographically, a box $\boxed{\cdot}$ may be reduced to a matching pair of parentheses, '$(\cdot)$', and one gets in this way a linear description that illustrates the correspondence between trees viewed as plane diagrams and functional terms of mathematical logic and computer science.

Trees are best described recursively. A tree is a root to which is attached a (possibly empty) sequence of trees. In other words, the class $\mathcal{G}$ of general trees is definable by the recursive equation

$$\mathcal{G} = \mathcal{Z} \times \text{SEQ}(\mathcal{G}), \tag{20}$$

where $\mathcal{Z}$ comprises a single atom written "•" and denoting a generic node.

Although such recursive definitions are familiar to computer scientists, the specification (20) may look dangerously circular to some. One way of making good sense of it is via an adaptation of the numerical technique of iteration. Start with $\mathcal{G}^{[0]} = \emptyset$, the empty set, and define successively the classes

$$\mathcal{G}^{[j+1]} = \mathcal{Z} \times \text{SEQ}(\mathcal{G}^{[j]}).$$

For instance, $\mathcal{G}^{[1]} = \mathcal{Z} \times \text{SEQ}(\emptyset) = \{(\bullet, \epsilon)\} \cong \{\bullet\}$ describes the tree of size 1, and

$$\begin{aligned}
\mathcal{G}^{[2]} &= \left\{\bullet,\ \bullet\boxed{\bullet},\ \bullet\boxed{\bullet\bullet},\ \bullet\boxed{\bullet\bullet\bullet},\ \ldots,\right\}\\
\mathcal{G}^{[3]} &= \Big\{\bullet,\ \bullet\boxed{\bullet},\ \bullet\boxed{\bullet\bullet},\ \bullet\boxed{\bullet\bullet\bullet},\ \ldots\\
&\qquad \bullet\boxed{\bullet\boxed{\bullet}},\ \bullet\boxed{\bullet\boxed{\bullet\bullet}},\ \bullet\boxed{\boxed{\bullet\bullet}\bullet},\ \bullet\boxed{\bullet\boxed{\bullet\bullet}\boxed{\bullet\bullet}}, \ldots\Big\}.
\end{aligned}$$

First, each $\mathcal{G}^{[j]}$ is well-defined since it corresponds to a purely iterative specification. Next, we have the inclusion $\mathcal{G}^{[j]} \subset \mathcal{G}^{[j+1]}$, ($\mathcal{G}^{[j]}$ admits of a simple interpretation as the class of all trees of height $< j$). We can therefore regard the complete class $\mathcal{G}$ as defined by the limit of the $\mathcal{G}^{[j]}$, that is, $\mathcal{G} := \bigcup_j \mathcal{G}^{[j]}$.

$\triangleright$ **7.** *Limes superior of classes.* Let $\{\mathcal{A}^{[j]}\}$ be any increasing sequence of combinatorial classes, in the sense that $\mathcal{A}^{[j]} \subset \mathcal{A}^{[j+1]}$. If $\mathcal{A}^{[\infty]} = \bigcup_j \mathcal{A}^{[j]}$ is a combinatorial class, then the corresponding OGFs satisfy $A^{[\infty]}(z) = \lim_{j\to\infty} A^{[j]}(z)$ in the formal topology (APPENDIX A: *Formal power series*, p. 648). $\triangleleft$

DEFINITION I.7. *A* specification *for an r–tuple $\vec{\mathcal{A}} = (\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})$ of classes is a collection of r equations,*

$$\left\{\begin{array}{lcl}
\mathcal{A}^{(1)} &=& \Xi_1(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})\\
\mathcal{A}^{(2)} &=& \Xi_2(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})\\
 & \cdots & \\
\mathcal{A}^{(r)} &=& \Xi_r(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})
\end{array}\right. \tag{21}$$

where each Ξ_i denotes a term built from the $\mathcal{A}$'s using the constructions of disjoint union, cartesian product, sequence, set, multiset, and cycle, as well as the initial structures $\mathcal{E}$ and $\mathcal{Z}$.

We also say that the system is a specification of $\mathcal{A}^{(1)}$. A specification for a class of combinatorial structures is thus a sort of formal grammar defining that class. Formally, the system (21) is an *iterative* specification if it is strictly upper-triangular, that is, $\mathcal{A}^{(r)}$ is defined solely in terms of initial classes $\mathcal{Z}, \mathcal{E}$; the definition of $\mathcal{A}^{(r-1)}$ only involves $\mathcal{A}^{(r)}$, etc, so that $\mathcal{A}^{(1)}$ can be equivalently described by a single term. Otherwise, the system is said to be *recursive*. In the latter case, the semantics of recursion is identical to the one introduced in the case of trees: start with the "empty" vector of classes, $\vec{\mathcal{A}}^{[0]} := (\emptyset, \ldots, \emptyset)$, iterate $\vec{\mathcal{A}}^{[j+1]} = \vec{\Xi}\big[\vec{\mathcal{A}}^{[j]}\big]$, and finally take the limit.

DEFINITION I.8. *A class of combinatorial structures is said to be* constructible *or* specifiable *iff it admits a (possibly recursive) specification in terms of sum, product, sequence, set, multiset, and cycle constructions.*

At this stage, we have therefore defined a specification language for combinatorial structures which is some fragment of set theory with recursion added. Each constructible class has by virtue of Theorem I.1 an ordinary generating function for which defining equations can be produced systematically. In fact, it is even possible to use computer algebra systems in order to compute it *automatically*! See the article of Flajolet, Salvy, and Zimmermann [**173**] for the description of such a system.

THEOREM I.2 (Symbolic method, unlabelled case). *The generating function of a constructible class is a component of a system of generating function equations whose terms are built from*

$$1,\ z,\ +\ ,\ \times\ ,\ Q\ ,\ \mathrm{Exp}\ ,\ \overline{\mathrm{Exp}}\ ,\mathrm{Log},$$

where

$$\begin{cases} Q[f] &=& \dfrac{1}{1-f}, & \mathrm{Log}[f] &=& \displaystyle\sum_{k=1}^{\infty}\frac{\varphi(k)}{k}\log\frac{1}{1-f(z^k)}, \\ \mathrm{Exp} &=& \exp\left(\displaystyle\sum_{k=1}^{\infty}\frac{f(z^k)}{k}\right), & \overline{\mathrm{Exp}}[f] &=& \exp\left(\displaystyle\sum_{k=1}^{\infty}(-1)^{k-1}\frac{f(z^k)}{k}\right). \end{cases}$$

The operator Q translating sequences (SEQ) is known as the quasi-inverse. The operator Exp (multisets, MSET) is called the Pólya exponential and $\overline{\mathrm{Exp}}$ (powersets, PSET) is the modified Pólya exponential. The operator Log is the Pólya logarithm. They are named after Pólya who first developed the general enumerative theory of objects under permutation groups [**37, 347, 349**].

The statement of Corollary I.2 signifies that iterative classes have explicit generating functions involving compositions of the basic operators only, while recursive structures have OGFs that are accessible indirectly via systems of functional equations. As we see at various places in this chapter, the following classes are constructible: binary words, binary trees, general trees, integer partitions, integer compositions, nonplane trees, polynomials over finite fields, necklaces, and wheels. We conclude this section with a few examples.

Binary words. The OGF of binary words, as seen already, can be obtained directly from the iterative specification,

$$\mathcal{W} = \mathrm{SEQ}(\mathcal{Z}+\mathcal{Z}) \qquad \Longrightarrow \qquad W(z) = \frac{1}{1-2z},$$

whence the expected result, $W_n = 2^n$.

General trees. The recursive specification of general trees leads to an implicit definition of their OGF,

$$\mathcal{G} = \mathcal{Z} \times \mathrm{SEQ}(\mathcal{G}) \qquad \Longrightarrow \qquad G(z) = \frac{z}{1-G(z)}.$$

From this point on, basic algebra does the rest. First the original equation is equivalent (in the ring of formal power series) to $G - G^2 - z = 0$. Next, the quadratic equation is solvable by radicals, and one finds

$$\begin{aligned} G(z) &= \tfrac{1}{2}\left(1-\sqrt{1-4z}\right) \\ &= z + z^2 + 2\,z^3 + 5\,z^4 + 14\,z^5 + 42\,z^6 + 132\,z^7 + 429\,z^8 + \cdots \\ &= \sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1} z^n. \end{aligned}$$

(The conjugate root is to be discarded since it involves a term z^{-1} as well as negative coefficients.) The expansion then results from Newton's binomial expansion,

$$(1+x)^\alpha = 1 + \frac{\alpha}{1}x + \frac{\alpha(\alpha-1)}{2!}x^2 + \cdots ,$$

applied with $\alpha = \frac{1}{2}$ and $x = -4z$.

The numbers

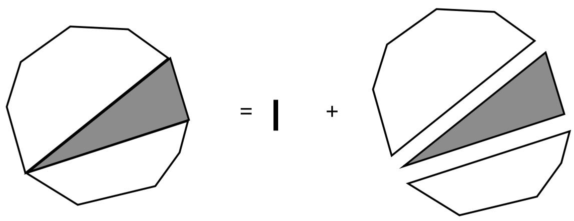
$$(22) \qquad \mathrm{C}_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!} \qquad \text{with OGF} \qquad \mathrm{C}(z) = \frac{1-\sqrt{1-4z}}{2z}$$

are known as the Catalan numbers (*EIS* **A000108**) in the honour of Eugène Catalan (1814-1894), a French and Belgian mathematician who developed many of their properties. These numbers are so common in combinatorics that we have decided to use a roman font for denoting them (like "$\log$", "$\sin$", and so on). In summary, *general trees are enumerated by Catalan numbers:*

$$G_n = \mathrm{C}_{n-1} \equiv \frac{1}{n}\binom{2n-2}{n-1}.$$

For this reason the term *Catalan tree* is often employed as synonymous to "general (rooted unlabelled plane) tree".

Triangulations. Fix $n+2$ points arranged in anticlockwise order on a circle and conventionally numbered from 0 to $n+1$ (for instance the $n+2$nd roots of unity). A triangulation is defined as a maximal decomposition of the convex $n+2$-gon defined by the points into n triangles. Triangulations are taken here as abstract topological configurations defined up to continuous deformations of the plane. The size of the triangulation is the number of triangles, that is, n. Given a triangulation, we define its "root" as a triangle chosen in some conventional and unambiguous manner (e.g., at the start, the triangle that contains the two smallest labels). Then, a triangulation decomposes into its root triangle and two subtriangulations (that may well be "empty") appearing on the left and right sides of the root triangle; the decomposition is illustrated by the following diagram:



The class $\mathcal{T}$ of all triangulations can be specified recursively as

$$\mathcal{T} \quad = \quad \{\epsilon\} \quad + \quad (\mathcal{T} \times \nabla \times \mathcal{T}),$$

provided that we consider a 2-gon (a diameter) as giving rise to an empty triangulation. Consequently, the OGF satisfies the equation $T = 1 + zT^2$ and

$$T(z) = \frac{1}{2z}\left(1 - \sqrt{1-4z}\right).$$

As a result, *triangulations are enumerated by Catalan numbers*:

$$T_n = \mathrm{C}_n \equiv \frac{1}{n+1}\binom{2n}{n}.$$

This particular result goes back to Euler and Segner (1753), a century before Catalan; see Figure 1 for first values and p. 70 for related bijections.

$\triangleright$ **8.** *A bijection.* Since both general trees and triangulations are enumerated by Catalan numbers, there must exist a size-preserving bijection between the two classes. Find one such bijection. [Hint: the construction of triangulations is evocative of binary trees, and binary trees are themselves in bijective correspondence with general trees; see APPENDIX A: *Tree concepts*, p. 653]. $\triangleleft$

$\triangleright$ **9.** *A variant specification of triangulations.* Consider the class $\mathcal{U}$ of "nonempty" triangulations of the n-gon, that is, we exclude the 2-gon and the corresponding "empty" triangulation of size 0. Then, $\mathcal{U} = \mathcal{T} \setminus \{\epsilon\}$ admits the specification

$$\mathcal{U} = \nabla + (\nabla \times \mathcal{U}) + (\mathcal{U} \times \nabla) + (\mathcal{U} \times \nabla \times \mathcal{U})$$

which also leads to the Catalan numbers via $U = z(1+U)^2$ and $U(z) = (1 - 2z - \sqrt{1-4z})/(2z)$, so that $U(z) = T(z) - 1$. $\triangleleft$

I. 2.4. Exploiting generating functions and counting sequences. In this book we are going to see altogether more than a hundred applications of the symbolic method. Before engaging in technical developments, it is worth inserting a few comments on the way generating functions and counting sequences can be put to good use in order to solve combinatorial problems.

Explicit enumeration formulæ. In a number of situations, generating functions are explicit and can be expanded in such a way that explicit formulae result for their coefficients. A prime example is the counting of general trees and of triangulations above, where the quadratic equation satisfied by an OGF is amenable to an explicit solution—the resulting OGF could then be expanded by means of Newton's binomial theorem. Similarly, we derive later in this Chapter an explicit form for the number of integer compositions by means of the symbolic method and OGFs (the answer turns out to be simply 2^{n-1}) and derive many explicit specializations. In this book, we assume as known the elementary techniques from basic calculus by which the Taylor expansion of an explicitly given function can be obtained. (Good references on such elementary aspects are Wilf's *Generatingfunctionology* [**437**], Graham, Knuth, and Patashnik's *Concrete Mathematics* [**212**], and our book [**382**].)

Implicit enumeration formulæ. In a number of cases, the generating functions obtained by the symbolic method are still in a sense explicit, but their form is such that their coefficients are not clearly reducible to a closed form. It is then still possible to obtain initial values of the corresponding counting sequence by means of a symbolic manipulation system. Also, from generating functions, it is possible to derive systematically recurrences[2] that lead to a procedure for computing an arbitrary number of terms of the counting sequence in a reasonably efficient manner. A typical example of this situation is the OGF of integer partitions,

$$P(z) = \prod_{m=1}^{\infty} \frac{1}{1-z^m},$$

for which recurrences obtained from the OGF and associated to fast algorithms are given in Note 12 (p. 39) and Note 17 (p. 46).

Asymptotic formulæ. Such forms are our eventual goal as they allow for an easy interpretation and comparison of counting sequences. From a quick glance at the table of initial values of W_n, P_n, T_n given in Eq. (2), it is apparent that W_n grows more slowly than T_n, which itself grows more slowly than P_n. The classification of growth rates of counting sequences belongs properly to the asymptotic theory of combinatorial structures which neatly relates to the symbolic method via complex analysis. A thorough treatment of this part of the theory is presented in Chapters IV–VIII. Given the methods exposed there, it becomes possible to estimate asymptotically the coefficients of virtually any generating function, however complicated[3], that is provided by the symbolic method.

Here, we content ourselves with a few remarks based on elementary real analysis. (The basic notations are described in APPENDIX A: *Asymptotic Notation*, p. 640.) The sequence $W_n = 2^n$ grows exponentially and, in such an extreme simple case, the exact form coincides with the asymptotic form. The sequence $P_n = n!$ must grow at a faster asymptotic regime. But how fast? The answer is provided by Stirling's formula, an approximation to the factorial numbers due to the Scottish mathematician James Stirling (1692–1770):

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(1 + O(\frac{1}{n})\right) \qquad (n \to +\infty). \tag{23}$$

The ratios of the exact values to Stirling's approximations

n:	1	2	5	10	100	1,000
$\dfrac{n!}{n^n e^{-n}\sqrt{2\pi n}}$:	1.084437	1.042207	1.016783	1.008365	1.000833	1.000083

show an *excellent quality* of the asymptotic estimate: the error is only 8% for $n = 1$, less than 1% for $n = 10$, and less than 1 per thousand for any n greater than 100.

[2]See [**173, 183, 326**] for such systematic approaches.

[3]In a number of cases, asymptotic analysis even applies to situations where the generating function itself is not even explicit, but only accessible through a functional equation of sorts.

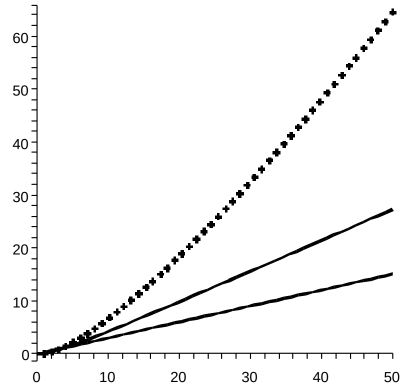


FIGURE 4. The growth regimes of three sequences $f(n) = 2^n, T_n, n!$ (from bottom to top) rendered by a plot of $\log_{10} f(n)$ versus n.

Stirling's formula in turn gives access to the asymptotic form of the Catalan numbers, by means of a simple calculation:

$$\mathrm{C}_n = \frac{1}{n+1}\frac{(2n)!}{(n!)^2} \sim \frac{1}{n}\frac{(2n)^{2n}e^{-2n}\sqrt{4\pi n}}{n^{2n}e^{-2n}2\pi n},$$

which simplifies to

$$\mathrm{C}_n \sim \frac{4^n}{\sqrt{\pi n^3}}. \tag{24}$$

Thus, the growth of Catalan numbers is roughly comparable to an exponential, 4^n, modulated by a subexponential factor, here $1/\sqrt{\pi n^3}$. A surprising consequence of this asymptotic estimate to the area of boolean function complexity appears in Example 16 below.

Altogether, the asymptotic number of general trees and triangulations is well summarized by a simple formula. Approximations become more and more accurate as n becomes large. Figure 4 illlustrates the different growth regimes of our three reference sequences while Figure 5 exemplifies the quality of the approximation with subtler phenomena also apparent on the figures and well explained by asymptotic theory. Such asymptotic formulæ then make comparison between the growth rates of sequences easy.

$\triangleright$ **10.** *The complexity of coding.* A company specialized in computer aided design has sold to you a scheme that (they claim) can encode any triangulation of size $n \geq 100$ using at most $1.5n$ bits of storage. After reading these pages, what do you do? [Hint: sue them!] See also Note 22 for related coding arguments. $\triangleleft$

$\triangleright$ **11.** *Experimental asymptotics.* From the data of Figure 5, guess the value of $\mathrm{C}^\star_{10^7} \,/\, \mathrm{C}_{10^7}$ and of $\mathrm{C}^\star_{5\cdot 10^6} \,/\, \mathrm{C}_{5\cdot 10^6}$ to 25D. (See, e.g., [**275**] for related asymptotic expansions and [**60**] for similar properties.) $\triangleleft$

The interplay between combinatorial structure and asymptotic structure is indeed the principal theme of this book. We shall see that a vast majority of the generating functions provided by the symbolic method, however complicated, eventually lead to similarly simple asymptotic estimates.

n	C_n	$C_n^\star$	$C_n^\star/C_n$
1	1	2.25	2.25675 83341 91025 14779 23178
10	16796	18707.89	1.11383 05127 5244589437 89064
100	$0.89651 \cdot 10^{57}$	$0.90661 \cdot 10^{57}$	1.01126 32841 24540 52257 13957
1000	$0.20461 \cdot 10^{598}$	$0.20484 \cdot 10^{598}$	1.00112 51328 1542 41647 01282
10000	$0.22453 \cdot 10^{6015}$	$0.22456 \cdot 10^{6015}$	1.00011 25013 28127 92913 51406
100000	$0.17805 \cdot 10^{60199}$	$0.17805 \cdot 10^{60199}$	1.00001 12500 13281 25292 96322
1000000	$0.55303 \cdot 10^{602051}$	$0.55303 \cdot 10^{602051}$	1.00000 11250 00132 81250 29296

FIGURE 5. The Catalan numbers C_n, their Stirling approximation $\mathrm{C}_n^\star = 4^n/\sqrt{\pi n^3}$, and the ratio $\mathrm{C}_n^\star \,/\, \mathrm{C}_n$.

I. 3. Integer compositions and partitions

This section and the next ones provide examples of counting via specifications in classical combinatorial domains. They illustrate the benefits of the symbolic method: generating functions are obtained with hardly any computation, and at the same time, many counting refinements follow from a basic combinatorial construction. The most direct applications described here relate to the additive decomposition of integers into summands with the classical combinatorial-arithmetic structures of partitions and compositions. The specifications are iterative and simply combine two levels of constructions of type SEQ, MSET, CYC, PSET.

I. 3.1. Compositions and partitions. Our first examples have to do with decomposing integers into sums.

DEFINITION I.9. *A* composition *of an integer n is a sequence $(x_1, x_2, \ldots, x_k)$ of integers (for some k) such that*

$$n = x_1 + x_2 + \cdots + x_k, \qquad x_j \geq 1.$$

A partition *of an integer n is a sequence $(x_1, x_2, \ldots, x_k)$ of integers (for some k) such that*

$$n = x_1 + x_2 + \cdots + x_k \qquad \textit{and} \qquad x_1 \geq x_2 \geq \cdots \geq x_k.$$

In both cases, the x_i's are called the summands or the parts and the quantity n is called the size of the composition or the partition.

By representing summands in unary using small discs ("•"), we can render graphically a composition by drawing bars between some of the balls; if we arrange summands vertically, compositions appear as ragged-landscapes. In contrast, partitions appear as staircases also known as Ferrers diagrams [**82**, p. 100]; see Figure 6. We let $\mathcal{C}$ and $\mathcal{P}$ denote the class of pacement all compositions and all partitions. Since a set can always be presented in sorted order, the difference between compositions and partitions lies in the fact that the order of summands *does* or *does not* matter. This is reflected by the use of a sequence construction (for $\mathcal{C}$) against a multiset construction

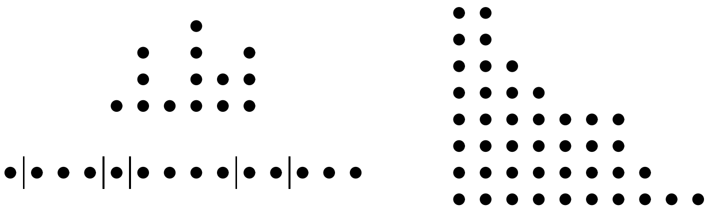


FIGURE 6. Graphical representations of compositions and partitions: (left) the composition $1+3+1+4+2+3=14$ with its "ragged-landscape" and "balls-and-bars" models; (right) the partition $8+8+6+5+4+4+4+2+1+1=43$ with its staircase (Ferrers diagram) model.

(for $\mathcal{P}$). In this perspective, it proves convenient to regard 0 as obtained by the empty sequence of summands ($k=0$), and we shall do so from now on.

First, let $\mathcal{I}=\{1,2,\ldots\}$ denote the combinatorial class of all integers at least 1 (the summands), and let the size of each integer be its value. Then, the OGF of $\mathcal{I}$ is, as we know,

$$I(z)=\sum_{n\geq 1} z^n=\frac{z}{1-z}, \tag{25}$$

since $I_n=1$ for $n\geq 1$, corresponding to the fact that there is exactly one object in $\mathcal{I}$ for each size $n\geq 1$. If integers are represented in unary, say by small balls, one has,

$$\mathcal{I}=\{1,\ 2,\ 3,\ \ldots\}=\{\bullet,\ \bullet\bullet,\ \bullet\bullet\bullet,\ \ldots\}\cong \textsc{Seq}_{\geq 1}\{\bullet\}, \tag{26}$$

which is another way to view the equality $I(z)=z/(1-z)$.

Compositions. First, the specification of compositions as sequences admits, by Theorem I.1, a direct translation into OGF:

$$\mathcal{C}=\textsc{Seq}(\mathcal{I}) \qquad \Longrightarrow \qquad C(z)=\frac{1}{1-I(z)}. \tag{27}$$

The collection of equations (25), (27) thus fully determines $C(z)$:

$$\begin{aligned} C(z) &= \frac{1}{1-\frac{z}{1-z}}=\frac{1-z}{1-2z} \\ &= 1+z+2z^2+4z^3+8z^4+16z^5+32z^6+\cdots. \end{aligned}$$

From there, the counting problem for compositions is solved by a straightforward expansion of the OGF: one has

$$C(z)=\left(\sum_{n\geq 0} 2^n z^n\right)-\left(\sum_{n\geq 0} 2^n z^{n+1}\right),$$

implying

$$C_n=2^{n-1},\ \ n\geq 1; \qquad C_0=1.$$

This agrees with basic combinatorics since a composition of n can be viewed as the placement of separation bars at a subset of the $n-1$ existing places inbetween n

0	1	1
10	512	42
20	524288	627
30	536870912	5604
40	549755813888	37338
50	562949953421312	204226
60	576460752303423488	966467
70	590295810358705651712	4087968
80	604462909807314587353088	15796476
90	618970019642690137449562112	56634173
100	633825300114114700748351602688	190569292
110	649037107316853453566312041152512	607163746
120	664613997892457936451903530140172288	1844349560
130	680564733841876926926749214863536422912	5371315400
140	696898287454081973172991196020261297061888	15065878135
150	713623846352979940529142984724747568191373312	40853235313
160	730750818665451459101842416358141509827966271488	107438159466
170	748288838313422294120286634350736906063837462003712	274768617130
180	766247770432944429179173513575154591809369561091801088	684957390936
190	784637716923335095479473677900958302012794430558004314112	1667727404093
200	803469022129495137770981046170581301261101496891396417650688	3972999029388
210	822752278660603021077484591278675252491367932816789931674304512	9275102575355
220	842498333348457493583344221469363458551160763204392890034487820288	21248279009367
230	862718293348820473429344482784628181556388621521298319395315527974912	47826239745920
240	883423532389192164791648750371459257913741948437809479060803100646309888	105882246722733
250	904625697166532776746648320380374280103671755200316906558262375061821325312	230793554364681

FIGURE 7. For $n = 0, 10, 20, \ldots, 250$ (left), the number of compositions C_n (middle) and the number of partitions (right). The figure illustrates the difference in growth between $C_n = 2^{n-1}$ and $P_n = e^{O(\sqrt{n})}$.

aligned balls (the "balls and bars" model of Figure 6), of which there are clearly 2^{n-1} possibilities.

Partitions. For partitions specified as multisets, the general translation mechanism provides

$$\mathcal{P} = \text{MSET}(\mathcal{I}) \quad \Longrightarrow \quad P(z) = \exp\left(I(z) + \frac{1}{2}I(z^2) + \frac{1}{3}I(z^3) + \cdots\right), \tag{28}$$

with product form

$$\begin{aligned} P(z) &= \prod_{m=1}^{\infty} \frac{1}{1-z^m} \\ &= \left(1 + z + z^2 + \cdots\right)\left(1 + z^2 + z^4 + \cdots\right)\left(1 + z^3 + z^6 + \cdots\right)\cdots \\ &= 1 + z + 2z^2 + 3z^3 + 5z^4 + 7z^5 + 11z^6 + 15z^7 + 22z^8 + \cdots . \end{aligned} \tag{29}$$

where the counting sequence is *EIS* **A000041**. Contrary to compositions that are counted by the explicit formula 2^{n-1}, no simple form exists for P_n. Asymptotic analysis of the OGF (28) based on the saddle point method (Chapter VIII) shows that $P_n = e^{O(\sqrt{n})}$. In fact a very famous theorem of Hardy and Ramanujan later improved by Rademacher (see Andrew's book [**10**] and Chapter VIII) provides a full expansion of which the asymptotically dominant term is

$$P_n \sim \frac{1}{4n\sqrt{3}} \exp\left(\pi\sqrt{\frac{2n}{3}}\right).$$

There are consequently appreciably fewer partitions than compositions (Figure 7).

▷ **12.** *A recurrence for the partition numbers.* Logarithmic differentiation gives

$$z\frac{P'(z)}{P(z)} = \sum_{n=1}^{\infty} \frac{nz^n}{1-z^n} \qquad \text{implying} \quad nP_n = \sum_{j=1}^{n-1} \sigma(j) P_{n-j},$$

where $\sigma(n)$ is the sum of the divisors of n (e.g., $\sigma(6) = 1 + 2 + 3 + 6 = 12$). Consequently, $P_1, \ldots, P_N$ can be computed in $O(N^2)$ integer-arithmetic operations. (The technique is generally applicable to powersets and multisets; see Note 40 for another application. Note 17 further lowers the bound in the case of partitions to $O(N\sqrt{N})$.) $\triangleleft$

By varying (27) and (28), we can use the symbolic method to derive a number of counting results in a straightforward manner. First, we state:

PROPOSITION I.1. *Let $\mathcal{T} \subseteq \mathcal{I}$ be a subset of the positive integers. The OGF of the classes $\mathcal{C}^{\mathcal{T}} := \text{SEQ}(\text{SEQ}_{\mathcal{T}}(\mathcal{Z}))$ and $\mathcal{P}^{\mathcal{T}} := \text{MSET}(\text{SEQ}_{\mathcal{T}}(\mathcal{Z}))$ of compositions and partitions having summands restricted to $\mathcal{T}$ is given by*

$$C^{\mathcal{T}}(z) = \frac{1}{1 - \sum_{n \in T} z^n} = \frac{1}{1 - T(z)}, \qquad P^{\mathcal{T}}(z) = \prod_{n \in \mathcal{T}} \frac{1}{1 - z^n}.$$

PROOF. The statement results directly from Theorem I.1. $\square$

This proposition permits us to enumerate compositions and partitions with restricted summands, as well as with a fixed number of parts.

EXAMPLE 4. *Compositions with restricted summands.* In order to enumerate the class $\mathcal{C}^{\{1,2\}}$ of compositions of n whose parts are only allowed to be taken from the set $\{1, 2\}$, simply write

$$\mathcal{C}^{\{1,2\}} = \text{SEQ}(\mathcal{I}^{\{1,2\}}) \qquad \text{with } \mathcal{I}^{\{1,2\}} = \{1, 2\}.$$

Thus, in terms of generating functions, one has

$$C^{\{1,2\}}(z) = \frac{1}{1 - I^{\{1,2\}}(z)} \qquad \text{with} \quad I^{\{1,2\}}(z) = z + z^2.$$

This formula implies

$$C^{\{1,2\}}(z) = \frac{1}{1 - z - z^2} = 1 + z + 2z^2 + 3z^3 + 5z^4 + +8z^5 + 13z^6 + \cdots,$$

and the number of compositions of n in this class is expressed by a Fibonacci number,

$$C_n^{\{1,2\}} = \mathrm{F}_{n+1} \text{ where } \mathrm{F}_n = \frac{1}{\sqrt{5}} \left[\left(\frac{1+\sqrt{5}}{2} \right)^n - \left(\frac{1-\sqrt{5}}{2} \right)^n \right].$$

In particular, the rate of growth is of the exponential type φ^n, where $\varphi := \frac{1+\sqrt{5}}{2}$ is the golden ratio.

Similarly, compositions such that all their summands lie in the set $\{1, 2, \ldots, r\}$ have generating function

$$C^{\{1,\ldots,r\}}(z) = \frac{1}{1 - z - z^2 - \cdots z^r} = \frac{1}{1 - z\frac{1-z^r}{1-z}} = \frac{1-z}{1 - 2z + z^{r+1}},$$

and the corresponding counts are given by generalized Fibonacci numbers. A double combinatorial sum expresses these counts

$$(30) \qquad C_n^{\{1,\ldots,r\}} = [z^n] \sum_j \left(\frac{z(1-z^r)}{(1-z)} \right)^j = \sum_{j,k} (-1)^k \binom{j}{k} \binom{n - rk - 1}{j - 1}.$$

This result is perhaps not too useful for grasping the rate of growth of the sequence when n gets large, so that asymptotic analysis is called for. Asymptotically, for any fixed $r \geq 2$, there is a unique root ρ_r of the denominator $1 - 2z + z^{r+1}$ in $(\frac{1}{2}, 1)$, this root dominates all the other roots and is simple. Methods amply developed in Chapter IV, imply that, for some constant $c_r > 0$,

$$C_n^{\{1,\dots,r\}} \sim c_r \rho_r^{-n} \quad \text{for fixed } r \text{ as } n \to \infty. \tag{31}$$

The quantity ρ_r plays a rôle similar to that of the golden ratio when $r = 2$. END OF EXAMPLE 4. □

▷ **13.** *Compositions into primes.* The additive decomposition of integers into primes is still surrounded with mystery. For instance, it is not known whether every even number is the sum of two primes (Goldbach's conjecture). However, the number of compositions of n into prime summands (*any* number of summands is permitted) is $B_n = [z^n]B(z)$ where

$$\begin{aligned} B(z) &= \left(1 - \sum_{p \text{ prime}} z^p\right)^{-1} = \left(1 - z^2 - z^3 - z^5 - z^7 - z^{11} - \cdots\right)^{-1} \\ &= 1 + z^2 + z^3 + z^4 + 3\,z^5 + 2\,z^6 + 6\,z^7 + 6\,z^8 + 10\,z^9 + 16\,z^{10} + \cdots \end{aligned}$$

(*EIS* **A023360**) and complex asymptotic method make it *easy* from there to determine the asymptotic form $B_n \sim 0.30365 \cdot 1.47622^n$; see Chapter IV. ◁

EXAMPLE 5. *Partitions with restricted summands (denumerants).* Whenever summands are restricted to a finite set, the special partitions that result are called denumerants. A popular denumerant problem consists in finding the number of ways of giving change of 99 cents using coins that are pennies (1 ¢), nickels (5 ¢), dimes (10 ¢) and quarters (25 ¢). (The order in which the coins are taken does not matter and repetitions are allowed.) For the case of a finite $\mathcal{T}$, we predict from Proposition I.1 that $P^{\mathcal{T}}(z)$ is always a *rational* function with poles that are at roots of unity; also the $P_n^{\mathcal{T}}$ satisfy a linear recurrence related to the structure of $\mathcal{T}$. The solution to the original coin change problem is found to be

$$[z^{99}]\,\frac{1}{(1-z)(1-z^5)(1-z^{10})(1-z^{25})} = 213.$$

In the same vein, one proves that

$$P_n^{\{1,2\}} = \left\lceil \frac{2n+3}{4} \right\rfloor \qquad P_n^{\{1,2,3\}} = \left\lceil \frac{(n+3)^2}{12} \right\rfloor.$$

There $\lceil x \rfloor \equiv \lfloor x + \frac{1}{2} \rfloor$ denotes the integer closest to the real number x. Such results are typically obtained by the two step process: (i) decompose the rational generating function into simple fractions; (ii) compute the coefficients of each simple fraction and combine them to get the final result [**82**, p. 108].

The general argument also gives the generating function of partitions whose summands lie in the set $\{1, 2, \dots, r\}$ as

$$P^{\{1,\dots,r\}}(z) = \prod_{m=1}^{r} \frac{1}{1 - z^m}. \tag{32}$$

In other words, we are enumerating partitions according to the value of the largest summand. One then finds by looking at the poles (Chapter IV):

$$P_n^{\{1,\dots,r\}} \sim c_r n^{r-1} \qquad \text{with} \quad c_r = \frac{1}{r!(r-1)!}. \tag{33}$$

A similar argument provides the asymptotic form of $P_n^{\mathcal{T}}$ when $\mathcal{T}$ is an arbitrary finite set:

$$P_n^{\mathcal{T}} \sim \frac{1}{\tau} \frac{n^{r-1}}{(r-1)!} \qquad \text{with } \tau := \prod_{n \in \mathcal{T}} n, \quad r := \operatorname{card}(\mathcal{T}).$$

This result originally due to Schur is discussed in Chapter IV. . . . . . . END OF EXAMPLE 5. □

We next examine compositions and partitions with a fixed number of summands.

EXAMPLE 6. *Compositions with a fixed number of parts.* Let $\mathcal{C}^{(k)}$ denote the class of compositions made of k summands, k a fixed integer ≥ 1. One has

$$\mathcal{C}^{(k)} = \text{SEQ}_k(\mathcal{I}) \equiv \mathcal{I} \times \mathcal{I} \times \cdots \times \mathcal{I},$$

where the number of terms in the cartesian product is k. From there, the corresponding generating function is found to be

$$C^{(k)} = \big(I(z)\big)^k \qquad \text{with} \qquad I(z) = \frac{z}{1-z}.$$

The number of compositions of n having k parts is thus

$$C_n^{(k)} = [z^n]\frac{z^k}{(1-z)^k} = \binom{n-1}{k-1},$$

a result which constitutes a combinatorial refinement of $C_n = 2^{n-1}$. (Note that the formula $C_n^{(k)} = \binom{n-1}{k-1}$ also results easily from the balls-and-bars model of compositions (Figure 6)). In such a case, the asymptotic estimate $C_n^{(k)} \sim n^{k-1}/(k-1)!$ results immediately from the polynomial form of the binomial coefficient $\binom{n-1}{k-1}$. . . . . END OF EXAMPLE 6. □

EXAMPLE 7. *Partitions with a fixed number of parts.* Let $\mathcal{P}^{(\leq k)}$ be the class of integer partitions with at most k summands. With our notation for restricted constructions (p. 29), this class is specified as

$$\mathcal{P}^{(\leq k)} = \text{MSET}_{\leq k}(\mathcal{I}).$$

It would be possible to appeal to the admissibility of such restricted compositions as developed in Section I. 6.1, but the following direct argument suffices.

Geometrically, partitions, are represented as collections of points: this is the staircase model of Figure 6). A symmetry around the main diagonal (also known in the specialized literature as conjugation) exchanges number of summands and value of largest summand: one has (with previous notations)

$$\mathcal{P}^{(\leq k)} \cong \mathcal{P}^{\{1,\,..\,k\}} \qquad \Longrightarrow \qquad P^{(\leq k)}(z) = P^{\{1,\,..\,k\}}(z),$$

so that, by (32),

$$P^{(\leq k)}(z) \equiv P^{\{1,\ldots,k\}} = \prod_{m=1}^{k} \frac{1}{1-z^m}. \tag{34}$$

As a consequence, the OGF of partitions with *exactly* k summands, $P^{(k)}(z) = P^{(\leq k)}(z) - P^{(\leq k-1)}(z)$, evaluates to

$$P^{(k)}(z) = \frac{z^k}{(1-z)(1-z^2)\cdots(1-z^k)}.$$

Given the equivalence between number of parts and largest part in partitions, the asymptotic estimate (33) applies verbatim here. . . . . END OF EXAMPLE 7. □

▷ **14.** *Compositions with summands bounded in number and size.* The number of compositions of size n with k summands each at most r is

$$[z^n]\left(z\frac{1-z^r}{1-z}\right)^k,$$

and is expressible as a simple binomial convolution. ◁

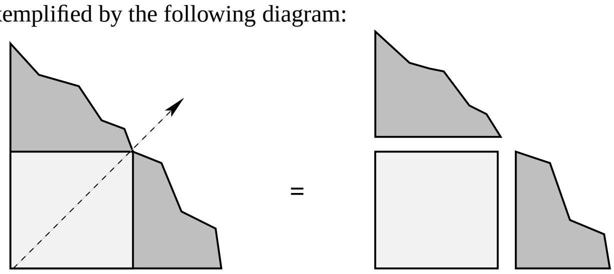
▷ **15.** *Partitions with summands bounded in number and size.* The number of partitions of size n with at most k summands each at most ℓ is

$$[z^n]\frac{(1-z)(1-z^2)\cdots(1-z^{k+\ell})}{((1-z)(1-z^2)\cdots(1-z^k))\cdot((1-z)(1-z^2)\cdots(1-z^\ell))}.$$

(The verification by recurrence is easy.) The GF reduces to the binomial coefficient $\binom{k+\ell}{k}$ as $z\to 1$; it is known as a Gaussian binomial coefficient, denoted $\binom{k+\ell}{k}_z$, or a "q–analogue" of the binomial coefficient [**10, 82**]. ◁

The last example of this section illustrates the close interplay between combinatorial decompositions and special function identities, which constitutes a recurrent theme of classical combinatorial analysis.

EXAMPLE 8. *The Durfee square of partitions and stack polyominoes.* The diagram of any partition contains a uniquely determined square (known as the Durfee square) that is maximal, as exemplified by the following diagram:



This decomposition is expressed in terms of partition GFs as

$$\mathcal{P}\cong\bigcup_{k\ge 0}\left(\mathcal{Z}^{k^2}\times\mathcal{P}^{(\le k)}\times\mathcal{P}^{\{1,\dots,k\}}\right),$$

It gives, via (32) and (34), the non-trivial identity

$$\prod_{n=1}^{\infty}\frac{1}{1-z^n}=\sum_{k\ge 0}\frac{z^{k^2}}{((1-z)\cdots(1-z^k))^2}$$

(k is the size of the Durfee square), which is nothing but a formal rewriting of the geometric decomposition.

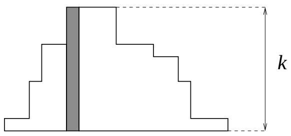
Here is a similar case illustrating the direct correspondence between geometric diagrams and generating functions, as afforded by the symbolic method.

Stack polyominoes are diagrams of compositions such that for some j,ℓ, one has $1\le x_1\le x_2\le\cdots\le x_j\ge x_{j+1}\ge\cdots\ge x_\ell\ge 1$ (see [**391**, §2.5] for further properties). The diagram

	Spec.	OGF	coeff.	asympt.
Composition	$\textsc{Seq}(\textsc{Seq}_{\geq 1}(\mathcal{Z}))$	$\frac{1-z}{1-2z}$	2^{n-1}	$\frac{1}{2}2^n$
—, sum. $\leq r$	$\textsc{Seq}(\textsc{Seq}_{1\,..\,r}(\mathcal{Z}))$	$\frac{1-z}{1-2z+z^{r+2}}$	Eq. (30)	$c_r\rho_r^{-n}$
—, k sum.	$\textsc{Seq}_k(\textsc{Seq}_{\geq 1}(\mathcal{Z}))$	$\frac{z^k}{(1-z)^k}$	$\binom{n-1}{k-1}$	$\frac{n^{k-1}}{(k-1)!}$
Partitions	$\textsc{MSet}(\textsc{Seq}_{\geq 1}(\mathcal{Z}))$	$\prod_{m=1}^{\infty}(1-z^m)^{-1}$	—	$\frac{1}{4n\sqrt{3}}e^{\pi\sqrt{\frac{2n}{3}}}$
—, sum. $\leq r$	$\textsc{MSet}(\textsc{Seq}_{1\,..\,r}(\mathcal{Z}))$	$\prod_{m=1}^{r}(1-z^m)^{-1}$	—	$\frac{n^{r-1}}{r!(r-1)!}$
—, $\leq k$ sum.	$\cong \textsc{MSet}(\textsc{Seq}_{1\,..\,k}(\mathcal{Z}))$	$\prod_{m=1}^{k}(1-z^m)^{-1}$	—	$\frac{n^{k-1}}{k!(k-1)!}$
Cyclic comp.	$\textsc{Cyc}(\textsc{Seq}_{\geq 1}(\mathcal{Z}))$	Eq. (35)	Eq. (36)	$\frac{2^n}{n}$
Part., distinct sum.	$\textsc{PSet}(\textsc{Seq}_{\geq 1}(\mathcal{Z}))$	$\prod_{m=1}^{\infty}(1+z^m)$	—	$\frac{3^{3/4}}{12n^{3/4}}e^{\pi\sqrt{\frac{n}{3}}}$

FIGURE 8. Partitions and compositions: specifications, generating functions, counting sequences, and asymptotic approximation.

representation of stack polyominoes,



$$\longleftrightarrow \quad \mathcal{P}^{\{1,\ldots,k-1\}} \times \mathcal{Z}^k \times P^{\{1,\ldots,k\}}$$

translates immediately into the OGF

$$S(z) = \sum_{k\geq 1} \frac{z^k}{1-z^k} \frac{1}{((1-z)(1-z^2)\cdots(1-z^{k-1}))^2},$$

once use is made of the partition GFs $P^{\{1,\ldots,k}(z)$ of (32). This last relation provides a *bona fide* algorithm for computing the initial values of the number of stack polyominoes (*EIS* **A001523**):

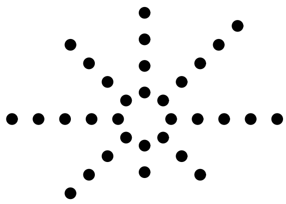
$$S(z) = z + 2\,z^2 + 4\,z^3 + 8\,z^4 + 15\,z^5 + 27\,z^6 + 47\,z^7 + 79\,z^8 + \cdots.$$

The book of van Rensburg [**423**] describes many such constructions and their relation to certain models of statistical physics. .... END OF EXAMPLE 8. □

Figure 8 summarizes what has been learnt regarding compositions and partitions. The way several combinatorial problems are solved effortlessly by the symbolic method is worth noting.

I. 3.2. Related constructions. It is also natural to consider the two constructions of cycle and powerset that we have not yet applied to $\mathcal{I}$.

Cyclic compositions (wheels). The class $\mathcal{D} = \text{CYC}(I)$ comprises compositions defined up to circular shift of the summands; so, for instance $2 + 3 + 1 + 2 + 5$, $3 + 1 + 2 + 5 + 2$, etc, are identified. Alternatively, we may view elements of $\mathcal{D}$ as "wheels" composed of circular arrangements of rows of balls (taken up to circular symmetry).

A "wheel" (cyclic composition): 

By the cycle construction, the OGF is

$$(35)\qquad \begin{aligned} D(z) &= \sum_{k=1}^{\infty} \frac{\varphi(k)}{k} \log\left(1 - \frac{z^k}{1-z^k}\right)^{-1} \\ &= z + 2\,z^2 + 3\,z^3 + 5\,z^4 + 7\,z^5 + 13\,z^6 + 19\,z^7 + 35\,z^8 + \cdots . \end{aligned}$$

The coefficients are thus (*EIS* **A008965**)

$$(36)\qquad D_n = \frac{1}{n} \sum_{k \mid n} \varphi(k)(2^{n/k} - 1) \equiv -1 + \frac{1}{n} \sum_{k \mid n} \varphi(k) 2^{n/k} \sim \frac{2^n}{n}.$$

Notice that D_n is of the same asymptotic order as $\frac{1}{n} C_n$, which is suggested by circular symmetry of wheels, but $D_n \sim 2C_n/n$.

Partitions into distinct summands. The class $\mathcal{Q} = \text{PSET}(\mathcal{I})$ is the subclass of $\mathcal{P} = \text{MSET}(\mathcal{I})$ corresponding to partitions determined like in Definition I.9, but with the strict inequalities $x_k > \cdots > x_1$, so that the OGF is

$$Q(z) = \prod_{n \geq 1} (1 + z^n) = 1 + z + z^2 + 2z^3 + 2z^4 + 3z^5 + 4z^6 + 5z^7 + 6z^8 + \cdots .$$

The coefficients (*EIS* **A000009**) are not amenable to closed from. However the saddle point method (Chapter VIII) yields the approximation:

$$(37)\qquad Q_n \sim \frac{3^{3/4}}{12 n^{3/4}} \exp\left(\pi \sqrt{\frac{n}{3}}\right),$$

which has a shape similar to that of P_n.

$\triangleright$ **16.** *Odd versus distinct summands.* The partitions of n into odd summands $(\mathcal{O}_n)$ and into distinct summands $(\mathcal{Q}_n)$ are equinumerous. Indeed, one has

$$Q(z) = \prod_{m=1}^{\infty} (1 + z^m), \qquad O(z) = \prod_{j=0}^{\infty} (1 - z^{2j+1})^{-1}.$$

Equality results from substituting $(1 + a) = (1 - a^2)/(1 - a)$ with $a = z^m$,

$$Q(z) = \frac{\mathbf{1 - z^2}}{1 - z} \frac{\mathbf{1 - z^4}}{\mathbf{1 - z^2}} \frac{\mathbf{1 - z^6}}{1 - z^3} \frac{\mathbf{1 - z^8}}{\mathbf{1 - z^4}} \frac{\mathbf{1 - z^{10}}}{1 - z^5} \cdots = \frac{1}{1 - z} \frac{1}{1 - z^3} \frac{1}{1 - z^5} \cdots ,$$

and simplification of the numerators with half of the denominators (in boldface). $\triangleleft$

Partitions into powers. Let $\mathcal{I}^{\rm pow} = \{1, 2, 4, 8, \ldots\}$ be the set of powers of 2. The corresponding $\mathcal{P}$ and $\mathcal{Q}$ partitions have OGFs

$$\begin{aligned} P^{\rm pow}(z) &= \prod_{j=0}^{\infty} \frac{1}{1-z^{2^j}} \\ &= 1 + z + 2z^2 + 2z^3 + 4z^4 + 4z^5 + 6z^6 + 6z^7 + 10z^8 + + \cdots \\ Q^{\rm pow}(z) &= \prod_{j=0}^{\infty}(1+z^{2^j}) \\ &= 1 + z + z^2 + z^3 + z^4 + z^5 + \cdots . \end{aligned}$$

The first sequence $1, 1, 2, 2, \ldots$ is the "binary partition sequence" (*EIS* **A018819**); the difficult asymptotic analysis was performed by de Bruijn [**92**] who obtained an estimate that involves subtle fluctuations and is of the global form $e^{O(\log^2 n)}$. The function $Q^{\rm pow}(z)$ reduces to $(1-z)^{-1}$ since every number has a unique additive decomposition into powers of 2. Accordingly, the identity

$$\frac{1}{1-z} = \prod_{j=0}^{\infty}(1+z^{2^j})$$

first observed by Euler is sometimes nicknamed the "computer scientist's identity" as it expresses the fact that every number admits a unique binary representation.

There exists a rich set of identities satisfied by partition generating functions—this fact owes to deep connections with elliptic functions, modular forms, and q–analogues of special functions on the one hand, basic combinatorics and number theory on the other hand. See [**10, 82**] for an introduction to this fascinating subject.

▷ **17.** *Euler's pentagonal number theorem.* This famous identity expresses $1/P(z)$ as

$$\prod_{n\ge 1}(1-z^n) = \sum_{k\in\mathbb{Z}}(-1)^k z^{k(3k+1)/2}.$$

It is proved formally and combinatorially in [**82**, p. 105]. As a consequence, the numbers $\{P_j\}_{j=0}^{N}$ can be determined in $O(N\sqrt{N})$ arithmetic operations. ◁

▷ **18.** *A digital surprise.* Define the constant

$$\varphi := \frac{9}{10}\,\frac{99}{100}\,\frac{999}{1000}\,\frac{9999}{10000}\cdots .$$

Is it a surprise that it evaluates numerically to

$$\varphi \doteq 0.89001009999899990000001000099999999899999000000000010\cdots ,$$

that is, its decimal representation involves only the digits $0, 1, 8, 9$? [This is suggested by a note of S. Ramanujan, "Some definite integrals", *Messenger of Math.* XLIV, 1915, pp. 10–18.] ◁

▷ **19.** *Lattice points.* The number of lattice points with integer coordinates that belong to the closed ball of radius n in d–dimensional Euclidean space is

$$[z^{n^2}]\frac{1}{1-z}(\Theta(z))^d \qquad \text{where} \qquad \Theta(z) = 1 + 2\sum_{n=1}^{\infty} z^{n^2}.$$

Such OGFs are useful in cryptography [**283**]. Estimates may be obtained from the saddle point method; see Chapter VIII. ◁

I. 4. Words and regular languages

Fix a finite *alphabet* $\mathcal{A}$ whose elements are called *letters*. Each letter is taken to have size 1, *i.e.*, it is an atom. A *word* is then any finite sequence of letters, usually written without separators. So, for us, with the choice of the latin alphabet ($\mathcal{A} = \{\mathtt{a},\ldots,\mathtt{z}\}$), sequences written as `ygololihp`, `philology`, `zgrmblglps` are words. We denote the set of all words (often written as $\mathcal{A}^\star$ in formal linguistics) by $\mathcal{W}$. Following a well-established tradition in theoretical computer science and formal linguistics, any subset of $\mathcal{W}$ is called a *language* (or formal language, when the distinction with natural languages has to be made).

From the definition of the set of words $\mathcal{W}$, one has

$$\mathcal{W} \cong \textsc{Seq}(\mathcal{A}) \qquad \Longrightarrow \qquad W(z) = \frac{1}{1-mz}, \tag{38}$$

where m is the cardinality of the alphabet, *i.e.*, the number of letters. The generating function gives us the counting result

$$W_n = m^n.$$

This result is elementary, but, as is usual with symbolic methods, many enumerative consequences result from a given construction. It is precisely the purpose of this section to examine some of them.

We shall introduce separately two frameworks that each have great expressive power to describe languages. The first one is iterative (*i.e.*, nonrecursive) and it bases itself on "regular specifications" that only involve sums, products, and sequences; the other one that is recursive (but of a very simple form) is best conceived of in terms of finite automata and is equivalent to linear systems of equations. Both frameworks turn out to be logically equivalent in the sense that they determine the same family of languages, the *regular languages*, though the equivalence[4] is nontrivial and each particular problem usually admits a preferred representation. The resulting OGFs are invariably rational functions, a fact to be systematically exploited from an asymptotic standpoint in Chapters IV and V.

I. 4.1. Regular specifications. Consider words (or strings) over the binary alphabet $\mathcal{A} = \{a, b\}$. There is an alternative way to construct binary strings. It is based on the observation that (with a minor adjustment at the beginning) a string decomposes into a succession of "blocks" each formed with a single b followed by an arbitrary (possibly empty) sequence of a's. For instance $aaabaababaabbabbaaa$ decomposes as

$$aaa \,||\, baa \mid ba \mid baa \mid b \mid ba \mid b \mid baaa.$$

Omitting redundant[5] symbols, we have the alternative decomposition:

$$\mathcal{W} \cong \textsc{Seq}(a) \times \textsc{Seq}(b\ \textsc{Seq}(a)) \quad \Longrightarrow \quad W(z) = \frac{1}{1-z}\frac{1}{1-z\frac{1}{1-z}}. \tag{39}$$

[4] Appendix A: *Regular languages*, p. 650 provides a basis for this equivalence.

[5] As usual, when dealing with words, we freely omit redundant braces '$\{,\}$' and cartesian products '$\times$'. For instance, $\textsc{Seq}(a+b)$ and $a\,b$ are shorthand notations for $\textsc{Seq}(\{a\}+\{b\})$ and $\{\{a\}\times\{b\}\}$.

This last expression reduces to $(1-2z)^{-1}$ as it should.

Longest runs. The interest of the construction just seen is to take into account various meaningful properties, for example longest runs. Denote by $a^{<k} := \mathrm{SEQ}_{<k}(a)$ the collection of all words formed with the letter a only and whose length is between 0 and $k-1$; the corresponding OGF is $1+z+\cdots+z^{k-1} = (1-z^k)/(1-z)$. The collection $\mathcal{W}^{\langle k\rangle}$ of words which do not have k consecutive a's is described by an amended form of (39), and

$$\mathcal{W}^{\langle k\rangle} = a^{<k}\,\mathrm{SEQ}(ba^{<k}) \implies W^{\langle k\rangle}(z) = \frac{1-z^k}{1-z}\cdot\frac{1}{1-z\frac{1-z^k}{1-z}} = \frac{1-z^k}{1-2z+z^{k+1}}.$$

The OGF is in principle amenable to expansion, but the resulting coefficients expressions are complicated and, in such a case, asymptotic estimates tend to be more usable. From an analysis developed in Chapter V, it can indeed be deduced that the longest run of a's in a random binary string of length n is asymptotic to $\log_2 n$.

$\triangleright$ **20.** *Runs in arbitrary alphabets.* For an alphabet of cardinality m, the quantity

$$\frac{1-z^k}{1-mz+(m-1)z^{k+1}}$$

is the OGF of words without k consecutive occurrences of a designated letter. $\triangleleft$

The case of longest runs exemplifies the usefulness of nested constructions involving sequences. We set:

DEFINITION I.10. *An iterative specification that only involves atoms (e.g., letters of a finite alphabet $\mathcal{A}$) together with combinatorial sums, cartesian products, and sequence constructions is said to be a* regular specification.

*A language $\mathcal{L}$ is said to be S-*regular *(specification-regular) if there exists a class M described by a regular specification $\mathcal{R}$ such that $\mathcal{L}$ and $\mathcal{M}$ are combinatorially isomorphic: $\mathcal{L} \cong \mathcal{M}$.*

An equivalent way of expressing the definition is as follows: a language is S-regular if it can be described *unambiguously* by a regular expression (APPENDIX A: *Regular languages*, p. 650). The definition of a regular specification and the basic admissibility theorem imply immediately:

PROPOSITION I.2. *Any S-regular language has an OGF that is a* rational function. *This OGF is obtained from a regular specification of the language by translating each letter into the variable z, disjoint unions into sums, cartesian products into products, and sequences into quasi-inverses,* $(1-\cdot)^{-1}$.

This result is technically shallow but its importance derives from the fact that regular languages have great expressive power devolving from their rich closure properties (APPENDIX A: *Regular languages*, p. 650) as well as their relation to finite automata discussed in the next subsection. Examples 9 and 10 make use of Proposition I.2 and treat two problems closely related to longest runs.

EXAMPLE 9. *Combinations and spacings.* A regular specification describes the set $\mathcal{L}$ of words that contain exactly k occurrences of the letter b, from which the OGF automatically derives:

$$\mathcal{L} = \mathrm{SEQ}(a)\,(b\,\mathrm{SEQ}(a))^k \qquad\implies\qquad L(z) = z^k/(1-z)^{k+1}. \tag{40}$$

Hence the number of words in the language satisfies $L_n = \binom{n}{k}$. This is otherwise combinatorially evident, since each word of length n is characterized by the positions of its letters b, that is, the choice of k positions amongst n possible ones. Symbolic methods thus give us back the well-known count of combinations by binomial coefficients.

Let $\binom{n}{k}_{<d}$ be the number of combinations of k elements amongst $[1, n]$ with constrained spacings: no element can be at distance d or more from its successor. The refinement of (40)

$$\mathcal{L}^{[d]} = \text{SEQ}(a)\,(b\,\text{SEQ}_{<d}(a))^{k-1}\,(b\,\text{SEQ}(a)) \quad \Longrightarrow \quad \sum_{n\geq 0} \binom{n}{k}_{<d} z^n = \frac{z^k(1-z^d)^{k-1}}{(1-z)^{k+1}},$$

leads to a binomial convolution expression,

$$\binom{n}{k}_{<d} = \sum_j (-1)^j \binom{k-1}{j}\binom{n-dj}{k}.$$

(This problem is analogous to compositions with bounded summands.) What we have just analysed in the *largest* spacing (constrained to be $< d$) in subsets; a parallel analysis yields information regarding the *smallest* spacing. . . . . END OF EXAMPLE 9. □

EXAMPLE 10. *Double run statistics.* By forming maximal groups of equal letters in words, one finds easily that, for a binary alphabet,

$$\mathcal{W} = \text{SEQ}(b)\ \text{SEQ}(a\,\text{SEQ}(a)\,b\,\text{SEQ}(b))\ \text{SEQ}(a).$$

Let $\mathcal{W}^{\langle\alpha,\beta\rangle}$ be the class of all words that have at most α consecutive a's and at most β consecutive b's. The specification of $\mathcal{W}$ produces a specification of $\mathcal{W}^{\langle\alpha,\beta\rangle}$, upon replacing $\text{SEQ}(a), \text{SEQ}(b)$ by $\text{SEQ}_{<\alpha}(a), \text{SEQ}_{<\beta}(b)$ internally, and by $\text{SEQ}_{\leq\alpha}(a), \text{SEQ}_{\leq\beta}(b)$ externally. In particular, the OGF of binary words that never have more than r consecutive equal letters is found to be (set $\alpha = \beta = r$)

$$W^{\langle r,r\rangle} = \frac{1-z^{r+1}}{1-2z+z^{r+1}} = \frac{1+z+\cdots+z^r}{1-z-\cdots-z^r}, \tag{41}$$

after simplification.

Révész in [**362**] tells the following amusing story attributed to T. Varga: " A class of high school children is divided into two sections. In one of the sections, each child is given a coin which he throws two hundred times, recording the resulting head and tail sequence on a piece of paper. In the other section, the children do not receive coins, but are told instead that they should try to write down a 'random' head and tail sequence of length two hundred. Collecting these slips of paper, [a statistician] then tries to subdivide them into their original groups. Most of the time, he succeeds quite well."

The statistician's secret is to determine the probability distribution of the maximum length of runs of consecutive letters in a random binary word of length n (here $n = 200$). The probability of this parameter to equal k is

$$\frac{1}{2^n}\left(W_n^{\langle k,k\rangle} - W_n^{\langle k-1,k-1\rangle}\right)$$

and is fully determined by (41). The probabilities are then easily computed using any symbolic package: For $n = 200$, the values found are

k	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}$:	$6.54\,10^{-8}$	$7.07\,10^{-4}$	0.0339	0.1660	0.2574	0.2235	0.1459	0.0829	0.0440	0.0226

Thus, in a randomly produced sequence of length 200, there are usually runs of length 7 or more: the probability of the event turns out to be close to 80% (and there is still a probability of about 8% to have a run of length 11 or more). On the other hand most children (and adults) are usually afraid of writing down runs longer than 4 or 5 as this is felt as strongly "non-random". The statistician simply selects the slips that contain runs of length 6 or more. Et voilà! .... END OF EXAMPLE 10. □

▷ **21.** *Alice and Bob.* Alice wants to communicate n bits of information to Bob over a channel (a wire, an optic fiber) that transmits 0,1–bits but is such that any occurrence of 11 terminates the transmission. Thus, she can only send on the channel an encoded version of her message (where the code is of some length $\ell \geq n$) that does not contain the pattern 11.

Here is a first coding scheme: given the message $m = m_1 m_2 \cdots m_n$, where $m_j \in \{0, 1\}$, apply the substitution: $0 \mapsto 00$ and $1 \mapsto 10$; terminate the transmission by sending 11. This scheme has $\ell = 2n + O(1)$, and we say its rate is 2. Can one design codes with rate arbitrarily close to 1, asymptotically?

Let $\mathcal{C}$ be the class of allowed code words. A code of length at most L is achievable only if there is a one-to-one mapping from $\{0,1\}^n$ into $\bigcup_{j=0}^{L} \mathcal{C}_j$, i.e., $2^n \leq \sum_{j=0}^{L} C_j$. Working out the OGF of $\mathcal{C}$, one finds that necessarily

$$L \geq \lambda n + O(1), \qquad \lambda = \frac{1}{\log_2 \varphi} \doteq 1.440420, \quad \varphi = \frac{1+\sqrt{5}}{2}.$$

Thus no code can achieve a rate better than 1.44; i.e., a loss of at least 44% is unavoidable. (For this and the next note, see, e.g., MacKay [**305**, Ch. 17].) ◁

▷ **22.** *Coding without long runs.* Because of hysteresis in magnetic heads, certain storage devices cannot store binary sequences that have more than 4 consecutive 0's or more than 4 consecutive 1's. A coding scheme that transforms an arbitrary binary string into a string obeying this constraint is sought.

From the OGF, one finds $[z^{11}]W^{\langle 4,4\rangle}(z) = 1546 > 2^{10} = 1024$. Consequently, a substitution can be built that translates an original 10 bit block into an 11 bit block without five consecutive equal letters. When substituted blocks are concatenated, this may give rise to unwanted sequences of consecutive letters that are longer than acceptable. It then suffices to use "separators" and replace a substituted block of the form $\alpha \cdots \beta$ by the longer block $\overline{\alpha}\alpha \cdots \beta\overline{\beta}$, where $\overline{0} = 1$ and $\overline{1} = 0$. The resulting code has rate $\frac{13}{10}$.

Extensions of this method show that the rate 1.057 is achievable (theoretically). On the other hand, by the previous note, any acceptable code must use asymptotically at least 1.056n bits to encode strings of n bits. (Hint: let α be the root near $\frac{1}{2}$ of $1 - 2\alpha + \alpha^5 = 0$, which is a pole of $W^{\langle 4,4\rangle}$. One has $1/\log_2(1/\alpha) = 1.05621$.) ◁

Patterns. There are many situations in the sciences where it is of interest to determine whether the appearance of a certain *pattern* in long sequences of observations is significant. In a genomic sequence of length 100,000 (the alphabet is A,G,C,T), is it or not meaninful to detect three occurrences of the pattern TAGATAA, where the letters appear consecutively and in the prescribed order? In computer network security, certain attacks can be detected by some well defined alarming sequences of events, though these events may be separated by perfectly legitimate actions. On an other register, data mining aims at broadly categorizing electronic documents in an automatic way, and in this context the observation of well chosen patterns can provide highly discriminating criteria. These various applications require determining which patterns are, with high probability, bound to occur (these are *not* significant) and which are very unlikely to arise, so that actually observing them carries useful information.

Quantifying the corresponding probabilistic phenomena reduces to an enumerative problem—the case of double runs in Example 10 is in this respect typical.

The notion of pattern can be formalized in several ways. In this book, we shall consider two of them:

(a) *Subsequence pattern*: such a pattern is defined by the fact that its letter must appear in the right order, but not necessarily contiguously [**182**]. Subsequence patterns are also known as "hidden patterns".

(b) *Factor pattern*: such a pattern is defined by the fact that its letter must appear in the right order *and* contiguously [**218, 401**]. Factor patterns are also called "block patterns" or simply "patterns" when the context is clear.

For a given notion of pattern, there are then two related categories of problems. First, one may aim at determining the probability that a random word contains (or dually, exclude) a pattern; this problem is equivalently formulated as an existence problem—enumerate all words in which the pattern exists (i.e., occurs) independently of the number of occurrences. Second, one may aim at determining the expectation (or even the distribution) of the number of occurrences of a pattern in a random text; this problem involves enumerating enriched words, each with one occurrence of the pattern distinguished.

Such questions are amenable to methods of analytic combinatorics and in particular to the theory of regular specifications and automata: see Example 11 below for a first analysis of hidden patterns (to be continued in Chapter V) and Example 12 for an analysis of factor patterns (to be further extended in Chapters III, IV, and IX).

EXAMPLE 11. *Subsequence (hidden) patterns in a text.* A sequence of letters that occurs in the right order, but not necessarily contiguously in a text is said to be a "hidden pattern". For instance the pattern "*combinatorics*" is to be found hidden in Shakespeare's Hamlet (Act I, Scene 1)

Dared to the $\boxed{\text{comb}}$at; $\boxed{\text{in}}$ which our v$\boxed{\text{a}}$lian$\boxed{\text{t}}$ Hamlet–
F$\boxed{\text{or}}$ so th$\boxed{\text{i}}$s side of our known world esteem'd him–
Did slay this Fortinbras; who by a seal'd compact,
Well ratified by law and heraldry,
Did forfeit, with his life, all those his lands
Whi$\boxed{\text{c}}$h he $\boxed{\text{s}}$tood seized of, to the conqueror...

Take a fixed finite alphabet $\mathcal{A}$ comprising m letters ($m = 26$ for English). First, let us examine the language $\mathcal{L}$ of all words, also called "texts", that contain a given word $\mathfrak{p} = p_1 p_2 \cdots p_k$ of length k as a subsequence. These words can be described unambigously as starting with a sequence of letters not containing p_1 followed by the letter p_1 followed by a sequence not containing p_2, and so on:

$$\mathcal{L} = \mathrm{SEQ}(\mathcal{A} \setminus p_1)p_1 \, \mathrm{SEQ}(\mathcal{A} \setminus p_2)p_2 \cdots \mathrm{SEQ}(A \setminus p_k)p_k \, \mathrm{SEQ}(\mathcal{A}).$$

This is in a sense equivalent to parsing words unambiguously according to the leftmost occurrence of $\mathfrak{p}$ as a subsequence. The OGF is accordingly

$$L(z) = \frac{z^k}{(1-(m-1)z)^k} \frac{1}{1-mz}.$$

An easy analysis of the dominant simple pole at $z = 1/m$ shows that

$$L(z) \underset{z\to 1/m}{\sim} \frac{1}{1-mz}, \qquad \text{so that} \quad L_n \underset{n\to\infty}{\sim} m^n.$$

Thus, a proportion tending to 1 of all the words of length n *do* contain $\mathfrak{p}$ as a subsequecne.

▷ **23.** *A refined analysis.* Further consideration of the subdominant pole at $z = 1/(m-1)$ yields, by the methods of Chapter IV, the refined estimate:

$$1 - \frac{L_n}{n^m} = O\left(n^{k-1}\left(1-\frac{1}{m}\right)^n\right).$$

Thus, the probability of *not* containing a given subsequence pattern is exponentialy small. ◁

A census (Note 24) shows that there are in fact $1.63\,10^{39}$ occurrences of "*combinatorics*" as a subsequence hidden somewhere in the text of Hamlet, whose length is 120,057 (this is the number of letters that constitute the text). Is this the sign of a secret encouragement passed to us by the author of Hamlet?

Here is an analysis of the expected number of hidden patterns based on enumerating enriched words, where an enriched word is a word together with a distinguished occurrence of the pattern as a subsequence. Consider the regular specification

$$\mathcal{O} = \textsc{Seq}(\mathcal{A})\, p_1\ \textsc{Seq}(\mathcal{A})\, p_2\ \textsc{Seq}(\mathcal{A}) \cdots \textsc{Seq}(\mathcal{A})\, p_{k-1}\ \textsc{Seq}(\mathcal{A})\, p_k\ \textsc{Seq}(\mathcal{A}).$$

An element of $\mathcal{O}$ is a $(2k+1)$-tuple whose first component is an arbitrary word, whose second component is the letter p_1, and so on, with letters of the pattern and free blocks alternating . In other terms, any $\omega \in \mathcal{O}$ represents precisely one possible occurrence of the hidden pattern $\mathfrak{p}$ in a text built over the alphabet $\mathcal{A}$. The associated OGF is simply

$$O(z) = \frac{z^k}{(1-mz)^{k+1}}.$$

The ratio between the number of occurrences and the number of words of length n then equals

$$\Omega_n = \frac{[z^n]O(z)}{m^n} = m^{-k}\binom{n}{k}, \tag{42}$$

and this quantity represents the expected number of occurrences of the hidden pattern in a random word of length n, assuming all such words to be equally likely. For the parameters corresponding to the text of Hamlet ($n = 120,057$) and the pattern "*combinatorics*" ($k = 13$), the quantity Ω_n evaluates to $6.96\,10^{37}$. The number of hidden occurrences observed is thus 23 times higher than what the uniform model predicts! However, similar methods make it possible to take into account nonuniform letter probabilities (see Chapter III): based on the frequencies of letters in the English text itself, the expected number of occurrences is found to be $1.71\,10^{39}$—this is now only within 5% of what is observed. Thus, Shakespeare did not (probably) conceal in his text any message relative to combinatorics. END OF EXAMPLE 11. □

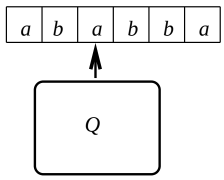
▷ **24.** *Dynamic programming.* The number of occurrences of a subsequence pattern in a text can be determined efficiently by scanning the text from left to right and maintaining a running count of the number of occurrences of the pattern as well as all its prefixes. ◁

I. 4.2. Finite automata. We begin with a simple device, the *finite automaton*, that is widely used in models of computation [**123**] and has wide descriptive power as regards structural properties of words.

DEFINITION I.11. *A* finite automaton *is a directed multigraph whose edges are labelled by letters of the alphabet $\mathcal{A}$. It is customary to refer to vertices as* states *and to denote by Q the set of states. An initial state $q_0 \in Q$ and a set of final states $Q_f \subseteq Q$ are designated.*

The automaton is said to be deterministic *if for each pair (q, α) with $q \in Q$ and $\alpha \in A$ there exists at most one edge (one also says a transition) starting from q that is labelled by the letter α.*

A finite automaton is able to process words, as we now explain. A word $w = w_1 \dots w_n$ is *accepted* by the automaton if there exists a path in the multigraph connecting the initial state q_0 to one of the final states of Q_f and whose sequence of edge labels is precisely $w_1, \dots, w_n$. For a deterministic finite automaton, it suffices to start from the initial state q_0, scan the letters of the word from left to right, and follow at each stage the only transition permitted; the word is accepted if the state reached in this way after scanning the last letter of w is a final state. Schematically:



A finite automaton thus keeps only a finite memory of the past (hence its name) and is in a sense a combinatorial counterpart of the notion of Markov chain in probability theory. In this book, we shall only consider deterministic automata.

As an illustration, consider the class $\mathcal{L}$ of all words w that contain the pattern abb as a factor (the letters of the pattern should appear contiguously). Such words are recognized by a finite automaton with 4 states, q_0, q_1, q_2, q_3. The construction is classical: state q_j is interpreted as meaning "*the first j characters of the pattern have just been scanned*", and the corresponding automaton appears in Figure 9. The initial state is q_0, and there is a unique final state q_3.

DEFINITION I.12. *A language is said to be A–*regular *(automaton regular) if it coincides with the set of words accepted by a deterministic finite automaton. A class $\mathcal{M}$ is A–regular if for some regular language $\mathcal{L}$, one has $\mathcal{M} \cong \mathcal{L}$.*

$\triangleright$ **25.** *Congruence languages.* The language of binary representations of numbers that are congruent to 2 to modulo 7 is A-regular. A similar property holds for any numeration base and any boolean combination of basic congruence conditions. $\triangleleft$

$\triangleright$ **26.** *Binary representation of primes.* The language of binary representations of prime numbers is neither A-regular nor S-regular. [Hint: this requires the Prime Number Theorem and asymptotic methods of Chapter IV.] $\triangleleft$

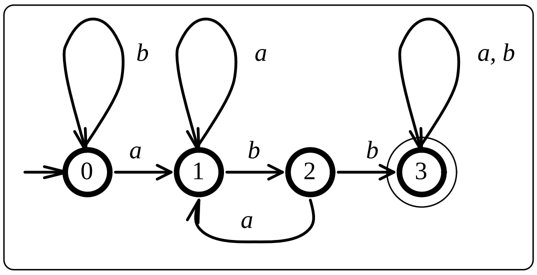


FIGURE 9. Words that contain the pattern abb are recognized by a 4–state automaton with initial state q_0 and final state q_3.

The following equivalence theorem is briefly discussed in the Appendix (see APPENDIX A: *Regular languages*, p. 650).

Equivalence theorem (Kleene–Rabin–Scott). *A language is S–regular (specification regular) if and only if it is A–regular (automaton regular).*

These two equivalent notions also coincide with the notion of regularity in formal language theory (defined there by means of regular expressions and nondeterministic finite automata [**3, 123**]). As already pointed out, the equivalence is non-trivial: it is given by an algorithm that transforms one formalism into the other, but does not transparently preserve combinatorial structure (e.g., in some cases, an exponential blow up in the size of descriptions is involved). For this reason, we have opted to develop both notions of S-regularity and A-regularity in an independent way.

We next examine the way generating functions can be obtained from a deterministic automaton. The process was first discovered in the late 1950's by Chomsky and Schützenberger [**80**].

PROPOSITION I.3. *Let G be a deterministic finite automaton with state set $Q = \{q_0, \dots, q_s\}$, initial state q_0, and set of final states $\overline{Q} = \{q_{i_1}, \dots, q_{i_f}\}$. The generating function of the language $\mathcal{L}$ of all words accepted by the automaton is a rational function that is determined under matrix form as*

$$L(z) = \mathrm{u}(I - zT)^{-1}\mathrm{v}.$$

There the transition matrix T is defined by

$$T_{i,j} = \operatorname{card}\{\alpha \in \mathcal{A} \text{ such that an edge } (q_i, q_j) \text{ is labelled by } \alpha\}\,;$$

the row vector u *is the vector* $(1, 0, 0, \dots, 0)$ *and the column vector* $\mathrm{v} = (v_0, \dots, v_s)^t$ *is such that*[6] $v_j = [\![q_j \in \overline{Q}]\!]$.

In particular, by Cramer's rule, the OGF of a regular language is the quotient of two sparse determinants whose structure directly reflects the automaton transitions.

PROOF. For $j \in \{0, \dots, s\}$, introduce the class (language) $\mathcal{L}_j$ of all words w such that the automaton, when started in state q_j, terminates in one of the final states after

[6]It proves convenient at this stage to introduce Iverson's bracket notation: for a predicate P, the variable $[\![P]\!]$ has value 1 if P is true and 0 otherwise.

having read w. The following relation holds for any j:

$$\mathcal{L}_j \cong \Delta_j + \left(\sum_{\alpha \in \mathcal{A}} \{\alpha\} \mathcal{L}_{(q_j \circ \alpha)} \right); \tag{43}$$

there Δ_j is the class $\{\epsilon\}$ formed of the word of length 0 if q_j is final and the empty set ($\emptyset$) otherwise; the notation $(q_j \circ \alpha)$ designates the state reached in one step from state q_j upon reading letter α. The justification is simple: a language $\mathcal{L}_j$ contains the word of length 0 only if the corresponding state q_j is final; a word of length ≥ 1 that is accepted starting from state q_j has a first letter α followed by a word that must lead to an accepting state when starting from state $q_j \circ \alpha$.

The translation of (43) is then immediate:

$$L_j(z) = [\![q_j \in \overline{Q}]\!] + z \sum_{\alpha \in \mathcal{A}} L_{(q_j \circ \alpha)}(z). \tag{44}$$

The collection of all the equations as j varies forms a linear system: with $\mathrm{L}(z)$ the column vector $(L_0(z), \ldots, L_s(z))$, one has

$$\mathrm{L}(z) = \mathrm{v} + zT\,\mathrm{L}(z),$$

where v and T are as described in the statement. The result follows by matrix inversion upon observing that $L(z) \equiv L_0(z)$. $\square$

The pattern abb***.*** Consider the automaton recognizing the pattern abb as given in Figure 9. The languages $\mathcal{L}_j$ (where L_j is the set of accepted words when starting from state q_j) are connected by the system of equations

$$\begin{array}{lllll} \mathcal{L}_0 & = & a\mathcal{L}_1 & + b\mathcal{L}_0 & \\ \mathcal{L}_1 & = & a\mathcal{L}_1 & + b\mathcal{L}_2 & \\ \mathcal{L}_2 & = & a\mathcal{L}_1 & + b\mathcal{L}_3 & \\ \mathcal{L}_3 & = & a\mathcal{L}_3 & + b\mathcal{L}_3 & + \epsilon, \end{array}$$

which directly reflects the graph structure of the automaton. This gives rise to a set of equations for the associated OGFs

$$\begin{array}{lllll} L_0 & = & zL_1 & + zL_0 & \\ L_1 & = & zL_1 & + zL_2 & \\ L_2 & = & zL_1 & + zL_3 & \\ L_3 & = & zL_3 & + zL_3 & + 1. \end{array}$$

Solving the system, we find the OGF of all words containing the pattern abb: it is $L_0(z)$ since the initial state of the automaton is q_0, and

$$L_0(z) = \frac{z^3}{(1-z)(1-2z)(1-z-z^2)}. \tag{45}$$

The partial fraction decomposition

$$L_0(z) = \frac{1}{1-2z} - \frac{2+z}{1-z-z^2} + \frac{1}{1-z},$$

then yields

$$L_{0,n} = 2^n - \mathrm{F}_{n+3} + 1,$$

with F_n a Fibonacci number. In particular the number of words of length n that do *not* contain abb is $\mathrm{F}_{n+3} - 1$, a quantity that grows at an exponential rate of φ^n, with

$\varphi = (1+\sqrt{5})/2$ the golden ratio. Thus, all but an exponentially vanishing proportion of the strings of length n contain the given pattern abb, a fact that was otherwise to be expected on probabilistic grounds. (For instance, from Note 29, a random word contains a large number, about $\sim n/8$, of occurrences of the pattern abb.)

▷ **27.** *Regular specification for pattern abb.* The pattern abb is simple enough that one can come up with an equivalent regular expression describing $\mathcal{L}_0$, whose existence is otherwise predicted by the Kleene-Rabin-Scott Theorem. An accepting path in the automaton of Figure 9 loops around state 0 with a sequence of b, then reads an a, loops around state 1 with a sequence of a's and moves to state 2 upon reading a b; then there should be letters making the automaton passs through states 1-2-1-2-$\cdots$-1-2 and finally a b followed by an arbitrary sequence of a's and b's at state 3. This corresponds to the specification

$$\begin{aligned}&\mathcal{L}_0 = \mathrm{SEQ}(b)\, a\, \mathrm{SEQ}(a) b\ \mathrm{SEQ}(a\, \mathrm{SEQ}(a) b)\, b\, \mathrm{SEQ}(a+b)\\ &\implies \quad L_0(z) = \frac{z^3}{(1-z)^2(1-\frac{z^2}{1-z})(1-2z)},\end{aligned}$$

which gives back a form equivalent to (45). ◁

EXAMPLE 12. *Words containing or excluding a pattern.* Fix an arbitrary pattern $\mathfrak{p} = p_1 p_2 \cdots p_k$ and let $\mathcal{L}$ be the language of words containing *at least* one occurrence of $\mathfrak{p}$ as a factor. Automata theory implies that the set of words containing a pattern as a factor is A–regular, hence admits a rational generating function. Indeed, the construction given for $\mathfrak{p} = abb$ generalizes in an easy manner: there exists a deterministic finite automaton with $k+1$ states that recognizes $\mathcal{L}$, the states memorizing at each stage the largest prefix of the pattern $\mathfrak{p}$ just seen. As a consequence: *The OGF of the language of words containing a given factor pattern of length k is a rational function of degree at most $k+1$.* (The corresponding automaton is in fact known as a Knuth–Morris–Pratt automaton [**272**].) The automaton construction however provides the OGF $L(z)$ in determinantal form, so that the relation between this rational form and the structure of the pattern is not transparent.

Autocorrelations. An explicit construction due to Guibas and Odlyzko [**217**] nicely circumvents this problem. It is based on an "equational" specification that yields an alternative linear system. The fundamental notion is that of an *autocorrelation vector*. For a given $\mathfrak{p}$, this vector of bits $c = (c_0, \dots, c_{k-1})$ is most conveniently defined in terms of Iverson's bracket as

$$c_i = [\![p_{i+1} p_{i+2} \cdots p_k = p_1 p_2 \cdots p_{k-i}]\!].$$

In other words, the bit c_i is determined by shifting $\mathfrak{p}$ right by i positions and putting a 1 if the remaining letters match the original $\mathfrak{p}$. Graphically, $c_i = 1$ if the two framed factors of $\mathfrak{p}$ coincide in

$$\begin{array}{llll}\mathfrak{p} \equiv & p_1 \cdots p_i & \boxed{p_{i+1} \cdots p_k} & \\ & & \boxed{p_1 \cdots p_{k-i}} & p_{k-i+1} \cdots p_k \quad \equiv \mathfrak{p}.\end{array}$$

For instance, with $\mathfrak{p} = aabbaa$, one has

```
a a b b a a
------------------------
a a b b a a               1
  a a b b a a             0
    a a b b a a           0
      a a b b a a         0
        a a b b a a       1
          a a b b a a     1
```

The autocorrelation is then $c = (1,0,0,0,1,1)$. The *autocorrelation polynomial* is defined as

$$c(z) := \sum_{j=0}^{k-1} c_j z^j.$$

For the example pattern, this gives $c(z) = 1 + z^4 + z^5$.

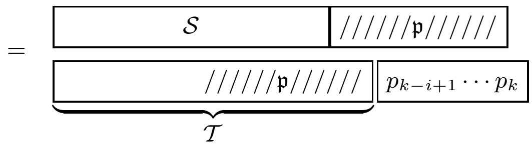
Let $\mathcal{S}$ be the language of words with *no* occurrence of $\mathfrak{p}$ and $\mathcal{T}$ the language of words that end with $\mathfrak{p}$ but have no other occurrence of $\mathfrak{p}$. First, by appending a letter to a word of $\mathcal{S}$, one finds a nonempty word either in $\mathcal{S}$ or $\mathcal{T}$, so that

$$\mathcal{S} + \mathcal{T} = \{\epsilon\} + \mathcal{S} \times \mathcal{A}. \tag{46}$$

Next, appending a copy of the word $\mathfrak{p}$ to a word in $\mathcal{S}$ may only give words that contain $\mathfrak{p}$ at or "near" the end. Precisely, the decomposition based on the leftmost occurrence of $\mathfrak{p}$ in $\mathcal{S}\mathfrak{p}$ is

$$\mathcal{S} \times \{\mathfrak{p}\} = \mathcal{T} \times \sum_{c_i \neq 0} \{p_{k-i+1} p_{k-i+2} \cdots p_k\}, \tag{47}$$

corresponding to the configurations



The translation of the system (46), (47) into OGFs then gives a system of two equations in the two unknown S, T,

$$S + T = 1 + mzS, \qquad S \cdot z^k = Tc(z),$$

which is then readily solved.

PROPOSITION I.4. *The OGF of words* not *containing the pattern* $\mathfrak{p}$ *as a factor is*

$$S(z) = \frac{c(z)}{z^k + (1 - mz)c(z)}, \tag{48}$$

where m is the alphabet cardinality, $k = |\mathfrak{p}|$ the pattern length, and $c(z)$ the autocorrelation polynomial of $\mathfrak{p}$.

A bivariate generating function based on the autocorrelation polynomial is derived in Chapter III, from which is deduced the existence of a limiting Gaussian law for the number of occurrences of any pattern in Chapter IX. .... END OF EXAMPLE 12. □

▷ **28.** *At least once.* The GFs of words containing at least once the pattern (anywhere) and containing it only once at the end are

$$L(z) = \frac{z^k}{(1 - mz)(z^k + (1 - mz)c(z))}, \quad T(z) = \frac{z^k}{z^k + (1 - mz)c(z)},$$

respectively. ◁

▷ **29.** *Expected number of occurrences of a pattern.* For the *mean* number of occurrences of a factor pattern, calculations similar to those employed for the number of occurrences of a subsequence (even simpler) can be based on regular specifications. All the occurrences $\mathfrak{p} = p_1 p_2 \cdots p_k$ as a factor are described by

$$\widehat{\mathcal{O}} = \text{SEQ}(\mathcal{A})\,(p_1 p_2 \cdots p_k)\ \text{SEQ}(\mathcal{A}), \quad \Longrightarrow \quad \widehat{O}(z) = \frac{z^k}{(1 - mz)^2}.$$

Consequently, the expected number of such contiguous occurrences satisfies

$$\widehat{\Omega}_n = m^{-k}(n - k + 1) \sim \frac{n}{m^k}. \tag{49}$$

Thus, the mean number of occurrences is proportional to n. ◁

$\triangleright$ **30.** *Waiting times in strings.* Let $\mathcal{L} \subset \mathrm{SEQ}\{a, b\}$ be a language and $S = \{a, b\}^\infty$ be the set of infinite strings with the product probability induced by $\mathbb{P}(a) = \mathbb{P}(b) = \frac{1}{2}$. The probability that a random string $\omega \in S$ starts with a word of $\mathcal{L}$ is $\widehat{L}(1/2)$, where $\widehat{L}(z)$ is the OGF of the "prefix language" of $\mathcal{L}$, that is, the set of words $w \in \mathcal{L}$ that have no strict prefix belonging to $\mathcal{L}$. The GF $\widehat{L}(z)$ serves to express the expected time at which a word in $\mathcal{L}$ is first encountered: this is $\frac{1}{2}\widehat{L}'(\frac{1}{2})$. For a regular language, this quantity must be a rational number. $\triangleleft$

$\triangleright$ **31.** *A probabilistic paradox on strings.* In a random infinite sequence, a pattern $\mathfrak{p}$ of length k first occurs on average at time $2^k c(1/2)$, where $c(z)$ is the correlation polynomial. For instance, the pattern $\mathfrak{p} = abb$ tends to occur "sooner" (at average position 8) than $\mathfrak{p}' = aaa$ (at average position 14). See [**217**] for a thorough discussion. Here are for instance the epochs at which $\mathfrak{p}$ and $\mathfrak{p}'$ are first found in a sample of 20 runs

$$\begin{array}{ll} \mathfrak{p}: & 3, 4, 5, 5, 6, 6, 7, 8, 8, 8, 8, 9, 9, 10, 11, 14, 15, 15, 16, 21 \\ \mathfrak{p}': & 3, 4, 8, 8, 9, 10, 11, 11, 11, 12, 17, 22, 23, 27, 27, 27, 44, 47, 52, 52. \end{array}$$

On the other hand, patterns of the same length have the same expected number of occurrences, which is puzzling. The catch is that, due to overlaps of $\mathfrak{p}'$ with itself, occurrences of $\mathfrak{p}'$ tend to occur in clusters, but, then, clusters tend to be separated by wider gaps than for $\mathfrak{p}$; eventually, there is no contradiction. $\triangleleft$

$\triangleright$ **32.** *Borges's Theorem.* Take any fixed set Π of finite patterns. A random text of length n contains all the patterns of the set Π (as factors) with probability tending to 1 exponentially fast as $n \to \infty$. (Reason: the rational functions $S(z/2)$ with $S(z)$ as in (48) have no pole in $|z| \le 1$; see also Chapters IV, V.)

Note: similar properties hold for many random combinatorial structures They are sometimes called "Borges's Theorem" as a tribute to the famous Argentinian writer Jorge Luis Borges (1899–1986) who, in his essay *"The Library of Babel"*, describes a library so huge as to contain: "Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogues of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, the demonstration of the fallacy of the true catalogue, the Gnostic gospel of Basilides, the commentary on that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book in all languages, the interpolations of every book in all books." $\triangleleft$

In general, automata are useful in establishing *a priori* the rational character of generating functions. They are also surrounded by interesting analytic properties (e.g., Perron-Frobenius theory, Chapter IV, that characterizes the dominant poles) and by asymptotic probability distributions of associated parameters that are normally Gaussian. They are most conveniently used for proving existence theorems, then supplemented when possible by regular specifications, which are likely to lead to more tractable expressions.

$\triangleright$ **33.** *Variable length codes.* A finite set $\mathcal{F} \subset \mathcal{W}$, where $\mathcal{W} = \mathrm{SEQ}(\mathcal{A})$ is called a *code* if any word of $\mathcal{W}$ decomposes in at most one manner into factors that belong to $\mathcal{F}$ (with repetitions allowed). For instance $\mathcal{F} = \{a, ab, bb\}$ is a code and $aaabbb = a|a|ab|bb$ has a unique decomposition; $\mathcal{F}' = \{a, aa, b\}$ is not a code since $aaa = a|aa = aa|a = a|a|a$. The OGF of the set $\mathcal{S}_\mathcal{F}$ of all words that admit a decomposition into factors all in $\mathcal{F}$ is a computable rational function, irrespective of whether $\mathcal{F}$ is a code. (Hint: use an "Aho–Corasick" automaton [**4**].) A finite set $\mathcal{F}$ is a code iff $S_\mathcal{F}(z) = (1 - F(z))^{-1}$. Consequently, the property of being a code can be decided in polynomial time using linear algebra. The book by Berstel and Perrin [**44**] develops systematically the theory of such variable-length codes. $\triangleleft$

$$\boxed{\alpha\,\beta\,\gamma\,\delta} \qquad \begin{array}{c} \boxed{\alpha\,\beta\,|\,\gamma\,\delta} \\ \boxed{\alpha\,\gamma\,|\,\beta\,\delta} \\ \boxed{\alpha\,\delta\,|\,\beta\,\gamma} \\ \boxed{\alpha\,|\,\beta\,\gamma\,\delta} \\ \boxed{\beta\,|\,\alpha\,\gamma\,\delta} \\ \boxed{\gamma\,|\,\alpha\,\beta\,\delta} \\ \boxed{\delta\,|\,\alpha\,\beta\,\gamma} \end{array} \qquad \begin{array}{c} \boxed{\alpha\,|\,\beta\,|\,\gamma\,\delta} \\ \boxed{\alpha\,|\,\gamma\,|\,\beta\,\delta} \\ \boxed{\alpha\,|\delta\,|\,\beta\,\gamma} \\ \boxed{\beta\,|\,\gamma\,|\,\alpha\,\delta} \\ \boxed{\beta\,|\,\delta\,|\,\alpha\,\gamma} \\ \boxed{\gamma\,|\,\delta\,|\,\alpha\,\beta} \end{array} \qquad \boxed{\alpha\,|\,\beta\,|\,\gamma\,|\,\delta}$$

FIGURE 10. The 15 ways of partitioning a four-element domain into blocks correspond to $S_4^{(1)} = 1, \quad S_4^{(2)} = 7, \quad S_4^{(3)} = 6, \quad S_4^{(4)} = 1.$

I. 4.3. Related constructions. Words can, at least in principle, encode any combinatorial structure. We detail here one example that demonstrates the usefulness of such encodings: it is relative to set partitions and Stirling numbers. The point to be made is that some amount of "combinatorial preprocessing" is sometimes necessary in order to bring combinatorial structures into the framework of symbolic methods.

Set partitions and Stirling partition numbers. A *set partition* is a partition of a finite domain into a certain number of nonempty sets, also called blocks. For instance, if the domain is $\mathcal{D} = \{\alpha, \beta, \gamma, \delta\}$, there are 15 ways to partition it (Figure 10). Let $\mathcal{S}_n^{(r)}$ denote the collection of all partitions of the set $[1\,..\,n]$ into r non–empty blocks and $S_n^{(r)} = \mathrm{card}(\mathcal{S}_n^{(r)})$ the corresponding cardinality. The basic object under consideration here is a *set partition* (not to be confused with integer partitions considered earlier).

It is possible to find an encoding of partitions in $\mathcal{S}_n^{(r)}$ of an n–set into r blocks by words over a r letter alphabet, $\mathcal{B} = \{b_1, b_2, \ldots, b_r\}$ as follows. Consider a set partition ϖ that is formed of r blocks. Identify each block by its smallest element called the block *leader*; then sort the block leaders into increasing order. Define the index of a block as the rank of its leader amongst all the r leaders, with ranks conventionally starting at 1. Scan the elements 1 to n in order and produce sequentially n letters from the alphabet $\mathcal{B}$: for an element belonging to the block of index j, produce the letter b_j.

For instance to $n = 6$, $r = 3$, the set partition $\varpi = \{\{6, 4\}, \{5, 1, 2\}, \{3, 7, 8\}\}$, is reorganized by putting leaders in first position of the blocks and sorting them,

$$\varpi = \{\overbrace{\{\mathbf{1}, 2, 5\}}^{b_1}, \overbrace{\{\mathbf{3}, 7, 8\}}^{b_2}, \overbrace{\{\mathbf{4}, 6\}}^{b_3}\},$$

so that the encoding is

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ b_1 & b_1 & b_2 & b_3 & b_1 & b_3 & b_2 & b_2 \end{pmatrix}.$$

In this way, a partition is encoded as a word of length n over $\mathcal{B}$ with the additional properties that: (i) all r letters occur; (ii) the first occurrence of b_1 precedes the first occurrence of b_2 which itself precedes the first occurrence of b_3, etc. Thus $\mathcal{S}_n^{(r)}$ is mapped into words of length n in the language

$$b_1\ \mathrm{SEQ}(b_1)\cdot b_2\ \mathrm{SEQ}(b_1+b_2)\cdot b_3\ \mathrm{SEQ}(b_1+b_2+b_3)\ \cdots\ b_r\ \mathrm{SEQ}(b_1+b_2+\cdots+b_r).$$

Graphically, this correspondence can be rendered by an "irregular staircase" representation, like

$$\begin{array}{cccccccc} & & & \mathbf{4} & - & 6 & - & - \\ & & \mathbf{3} & - & - & - & 7 & 8 \\ \mathbf{1} & 2 & - & - & 5 & - & - & - \end{array}$$

where the staircase has length n and height r, each column contains exactly one element, each row corresponds to a class in the partition.

The language specification immediately gives the OGF

$$S^{(r)}(z)=\frac{z^k}{(1-z)(1-2z)(1-3z)\cdots(1-rz)}.$$

The partial fraction expansion of $S^{(r)}(z)$ is readily computed,

$$S^{(r)}(z)=\frac{1}{r!}\sum_{j=0}^{r}\binom{r}{j}\frac{(-1)^{r-j}}{1-jz},\quad \text{so that}\quad S_n^{(r)}=\frac{1}{r!}\sum_{j=1}^{r}(-1)^{r-j}\binom{r}{j}j^n.$$

In particular, one has

$$S_n^{(1)}=1;\ S_n^{(2)}=\frac{1}{2!}(2^n-2);\ S_n^{(3)}=\frac{1}{3!}(3^n-3\cdot 2^n+3).$$

These numbers are known as the Stirling numbers of the second kind, or better, as the Stirling partition numbers, and the $S_n^{(r)}$ are nowadays usually denoted by $\left\{ {n \atop r} \right\}$; see APPENDIX A: *Stirling numbers*, p. 652.

The counting of set partitions could eventually be done successfully thanks an encoding into words, and the corresponding language forms a constructible class of combinatorial structures (actually a regular language). In the next chapter, we shall examine another approach to the counting of set partitions that is based on labelled structures and exponential generating functions.

Circular words (necklaces). Let $\mathcal{A}$ be a binary alphabet, viewed as comprised of beads of two distinct colours. The class of *circular words* or *necklaces* (p. 18 and Equation (17)) is defined by a CYC composition:

$$\mathcal{N}=\mathrm{CYC}(\mathcal{A})\qquad\Longrightarrow\qquad N(z)=\sum_{k=1}^{\infty}\frac{\varphi(k)}{k}\log\frac{1}{1-2z^k}.$$

The series starts as (*EIS* **A000031**)

$$N(z)=2z+3z^2+4z^3+6z^4+8z^5+14z^6+20z^7+36z^8+60z^9+\cdots,$$

and the OGF can be expanded:

$$N_n=\frac{1}{n}\sum_{k\,|\,n}\varphi(k)2^{n/k}. \tag{50}$$

It turns out that $N_n=D_n+1$ where D_n is the wheel count, p. 45. [The connection is easily explained combinatorially: start from a wheel and repaint in white all the nodes that are not on the basic circle; then fold them onto the circle.] The same argument proves that the number of necklaces over an m-ary alphabet is obtained by replacing 2 by m in (50).

$\triangleright$ **34.** *Finite languages.* Viewed as a combinatorial object, a *finite language* λ is a set of distinct words, with size being the total number of letters of all words in λ. For a binary alphabet, the class of all finite languages is thus

$$\mathcal{FL} = \mathrm{PSET}(\mathrm{SEQ}_{\geq 1}(\mathcal{A})) \qquad \Longrightarrow \qquad FL(z) = \exp\left(\sum_{k\geq 1} \frac{(-1)^{k-1}}{k} \frac{2z^k}{1-2z^k}\right).$$

The series starts as (*EIS* **A102866**) $1 + 2z + 5z^2 + 16z^3 + 42z^4 + 116z^5 + 310z^6 + \cdots$. $\triangleleft$

I. 5. Tree structures

This section is concerned with basic tree enumerations. Trees are, as we saw already, the prototypical recursive structure. There, recursive specifications normally lead to nonlinear equations (and systems of such equations) over generating functions. The Lagrange inversion theorem is useful in solving the simplest category of problems. The functional equations furnished by the symbolic method are then conveniently exploited by the asymptotic theory of Chapters VI and VII. A certain type of analytic behaviour appears to be universal in trees, namely a $\sqrt{\ }$–singularity; accordingly, as we shall see, most trees families occurring in the combinatorial world have counting sequences obeying an asymptotic form $C\,A^n n^{-3/2}$ that widely extends what we know already for Catalan numbers (p. 36).

I. 5.1. Plane trees. Trees are commonly defined as undirected acyclic connected graphs. In additions, the trees considered in this book are, unless specified otherwise, rooted. In this subsection, we focus attention on *plane trees*, also sometimes called ordered trees, where subtrees dangling from a node are ordered between themselves. Alternatively, these trees may be viewed as abstract graph structures accompanied by an embedding into the plane (see APPENDIX A: *Tree concepts*, p. 653 and[**268**, §2.3]). They are precisely described in terms of a sequence construction.

First, consider the class $\mathcal{G}$ of general plane trees where all node degrees are allowed (this repeats p. 33): we have

$$\mathcal{G} = \mathcal{Z} \times \mathrm{SEQ}(\mathcal{G}) \qquad \Longrightarrow \qquad G(z) = \frac{z}{1-G(z)}, \tag{51}$$

and, accordingly, $G(z) = \dfrac{1-\sqrt{1-4z}}{2}$, so that the number of general trees of size n is a Catalan number:

$$G_n = \mathrm{C}_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = \frac{(2n-2)!}{n!\,(n-1)!}.$$

Many classes of trees defined by all sorts of constraints on properties of nodes appear to be of interest in combinatorics and in related areas like logic and computer science. Let Ω be a subset of the integers that contains 0. Define the class $\mathcal{T}^{\Omega}$ of Ω-restricted trees as formed of trees such that the outdegrees of nodes are constrained to lie in Ω. In what follows, an essential rôle is played by a characteristic function that encapsulates Ω,

$$\phi(u) := \sum_{\omega\in\Omega} u^\omega.$$

Thus, $\Omega = \{0, 2\}$ determines binary trees, where each node has either 0 or 2 descendants, and $\phi(u) = 1 + u^2$; the choices $\Omega = \{0, 1, 2\}$ and $\Omega = \{0, 3\}$ determine respectively unary-binary trees ($\phi(u) = 1+u+u^2$) and ternary trees ($\phi(u) = 1+u^3$); the case of general trees corresponds to $\Omega = \mathbb{Z}_{\ge 0}$ and $\phi(u) = (1 - u)^{-1}$.

PROPOSITION I.5. *The ordinary generating function $T^\Omega(z)$ of the class $\mathcal{T}^\Omega$ of Ω-restricted trees is determined implicitly by the equation*

$$T^\Omega(z) = z\,\phi(T^\Omega(z)),$$

where ϕ is the characteristic of Ω, namely $\phi(u) := \sum_{\omega\in\Omega} u^\omega$. The tree counts are given by

$$T_n^\Omega \equiv [z^n]T^\Omega(z) = \frac{1}{n}\,[u^{n-1}]\phi(u)^n. \tag{52}$$

PROOF. Clearly, for Ω-restricted sequences, we have

$$\mathcal{A} = \textsc{Seq}_\Omega(\mathcal{B}) \qquad A(z) = \phi(B(z)),$$

so

$$\mathcal{T}^\Omega = \mathcal{Z} \times \textsc{Seq}_\Omega(\mathcal{T}^\Omega) \qquad \Longrightarrow \qquad T(z) = z\phi(T^\Omega(z)).$$

This shows that $T \equiv T^\Omega$ is related to z by functional inversion:

$$z = \frac{T}{\phi(T)}.$$

The Lagrange Inversion Theorem precisely provides expressions for such a case (see APPENDIX A: *Lagrange Inversion*, p. 649):

> **Lagrange Inversion Theorem.** *The coefficients of an inverse function and of all its powers are determined by coefficients of powers of the direct function: if $z = T/\phi(T)$, then*
>
> $$[z^n]T(z) = \frac{1}{n}[w^{n-1}]\phi(w)^n, \qquad [z^n]T(z)^k = \frac{k}{n}[w^{n-k}]\phi(w)^n.$$

The theorem immediately implies (52). $\square$

The statement extends trivially to the case where Ω is a multiset of integers, that is, a set of integers with repetitions allowed. For instance, $\Omega = \{0, 1, 1, 3\}$ corresponds to unary-ternary trees with two types of unary nodes, say, having one of two colours; in this case, the characteristic is $\phi(u) = u^0 + 2u^1 + u^3$. The theorem gives back the enumeration of general trees, where $\phi(u) = (1 - u)^{-1}$, by way of the binomial theorem applied to $(1-u)^{-n}$. In general, it implies that, whenever Ω comprises r elements, $\Omega = \{\omega_1, \ldots, \omega_r\}$, the tree counts are expressed as an $(r - 1)$-fold summation of binomial coefficients (use the multinomial expansion). An important special case detailed below is when Ω has only two elements.

▷ **35.** *Forests.* Consider ordered k-forests of trees defined by $\mathcal{F} = \text{SEQ}_k\{\mathcal{T}\}$. The Bürmann form of Lagrange inversion implies

$$[z^n]F(z) \equiv [z^n]T(z)^k = \frac{k}{n}[u^{n-k}]\,\phi(u)^n.$$

In particular, one has for forests of general trees ($\phi(u) = (1-u)^{-1}$):

$$[z^n]\left(\frac{1-\sqrt{1-4z}}{2}\right)^k = \frac{k}{n}\binom{2n-k-1}{n-1};$$

the coefficients are also known as "ballot numbers". ◁

EXAMPLE 13. *"Regular" (t-ary) trees.* A tree is said to be t-regular or t-ary if Ω consists only of the elements $\{0, t\}$. In other words, all internal nodes have degree t exactly, hence the name (Figure 11). Let $\mathcal{A} := \mathcal{T}^{\{0,t\}}$. In an element of $\mathcal{A}$, a node is either terminal or it has exactly t children. In this case, the characteristic is $\phi(u) = 1 + u^t$ and the binomial theorem combined with the Lagrange inversion formula gives

$$\begin{aligned} A_n &= \frac{1}{n}[u^{n-1}]\,(1+u^t)^n \\ &= \frac{1}{n}\binom{n}{\frac{n-1}{t}} \quad \text{provided } n \equiv 1 \bmod t. \end{aligned}$$

As the formula shows, only trees of total size of the form $n = t\nu + 1$ exist (a well-known fact otherwise easily checked by induction), and

$$A_{t\nu+1} = \frac{1}{t\nu+1}\binom{t\nu+1}{\nu} = \frac{1}{(t-1)\nu+1}\binom{t\nu}{\nu}. \tag{53}$$

A particular rôle is played by 2-regular trees known as *binary trees*. Then a form equivalent to (53) reads:

> *The number of plane binary trees having a total of* $2\nu+1$ *nodes (i.e.,* ν *binary nodes and* $\nu+1$ *external nodes) is the Catalan number* $\mathrm{C}_\nu = \frac{1}{\nu+1}\binom{2\nu}{\nu}$.

In this book, we shall use $\mathcal{B}$ to denote the class of binary trees. Size will be freely measured, depending on context and convenience, by recording internal, external, or all nodes.

There is a variant of the determination of (53) that avoids congruence restrictions. Let $\mathcal{A}$ be the class of t-ary trees and define the class $\widehat{\mathcal{A}}$ of "pruned" trees as trees of $\mathcal{A}$ deprived of all their external nodes. The trees in $\widehat{\mathcal{A}}$ now have nodes that are of degree at most t. In order to make $\widehat{\mathcal{A}}$ bijectively equivalent to $\mathcal{A}$, it suffices to regard trees of $\widehat{\mathcal{A}}$ as having $\binom{t}{j}$ possible types of nodes of degree j for any $j \in [0, t]$: each node type in $\widehat{\mathcal{A}}$ plainly encodes which of the original $t-j$ subtrees have been pruned. The equations above immediately generalize to the case of an Ω with multiplicities. One finds $\widehat{\phi}(u) = (1+u)^t$ and $\widehat{A}(z) = z\widehat{\phi}(\widehat{A}(z))$, so that, by Lagrange inversion,

$$\widehat{A}_\nu = \frac{1}{\nu}\binom{t\nu}{\nu-1},$$

yet another equivalent form of (53), since, by basic combinatorics, $\widehat{A}_\nu = A_{t\nu+1}$. END OF EXAMPLE 13. □

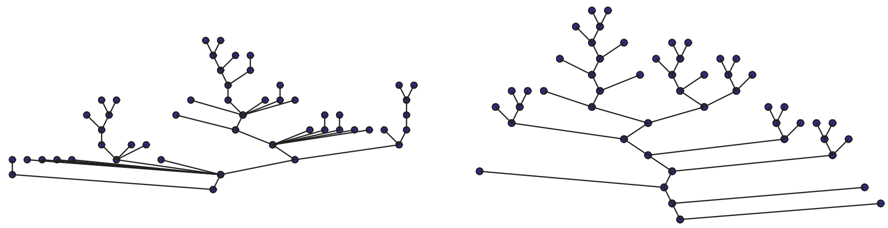


FIGURE 11. A general tree of $\mathcal{G}_{51}$ (left) and a binary tree of $\mathcal{T}_{51}^{\{0,2\}}$ (right) drawn uniformly at random amongst the C_{50} and C_{25} possible trees respectively, with $\mathrm{C}_n = \frac{1}{n+1}\binom{2n}{n}$ the nth Catalan number.

▷ **36.** *Motzkin numbers.* Let $M(z)$ be the generating function for unary-binary trees ($\Omega = \{0,1,2\}$):

$$M(z) = z(1 + M(z) + M(z)^2) \qquad \Longrightarrow \qquad M(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z}.$$

One has $M(z) = z + z^2 + 2\,z^3 + 4\,z^4 + 9\,z^5 + 21\,z^6 + 51\,z^7 + \cdots$. The coefficients $M_n = [z^n]M(z)$ are given in Lagrange form as

$$M_n = \frac{1}{n}\sum_k \binom{n}{k}\binom{n-k}{k-1},$$

and called Motzkin numbers (*EIS* **A001006**). ◁

▷ **37.** *Yet another variant of t-ary trees.* Let $\widetilde{\mathcal{A}}$ be the class of t-ary trees, but with size now defined as the number of external nodes (leaves). Then, one has

$$\widetilde{\mathcal{A}} = \mathcal{Z} + \mathrm{SEQ}_t(\widetilde{\mathcal{A}}).$$

The binomial form of $\widetilde{A}_n$ follows from Lagrange inversion, since $\widetilde{A} = z/(1 - \widetilde{A}^{t-1})$. ◁

EXAMPLE 14. *Hipparchus of Rhodes and Schröder.* In 1870, the German mathematician Ernst Schröder (1841–1902) published a paper entitled *Vier combinatorische Probleme*. The paper had to do with the number of terms that can be built out of n variables using nonassociative operations. In particular, the second of his four problems asks for the number of ways a string of n identical letters, say x, can be "bracketted". The rule is best stated recursively: x itself is a bracketting and if $\sigma_1, \sigma_2, \ldots, \sigma_k$ with $k \geq 2$ are bracketted expressions, then the k-ary product $(\sigma_1)(\sigma_2)\cdots(\sigma_k)$ is a bracketting.

Let $\mathcal{S}$ denote the class of all brackettings, where size is the number of variables. Then, the recursive definition is readily translated into the formal specification

$$\mathcal{S} = \mathcal{Z} + \mathrm{SEQ}_{\geq 2}(\mathcal{S}), \quad \mathcal{Z} = \{x\}. \tag{54}$$

To each bracketting of size n is associated a tree whose external nodes contain the variable x (and determine size), with internal nodes corresponding to brackettings and having degree at least 2 (while not contributing to size). The functional equation satisfied by the OGF is then

$$S(z) = z + \frac{S(z)^2}{1 - S(z)}. \tag{55}$$

$$(x_1) \wedge (x_2 \vee (x_3 \wedge x_4 \wedge x_5) \vee x_6) \wedge ((x_7 \wedge x_8) \vee (x_9 \wedge x_{10}))$$

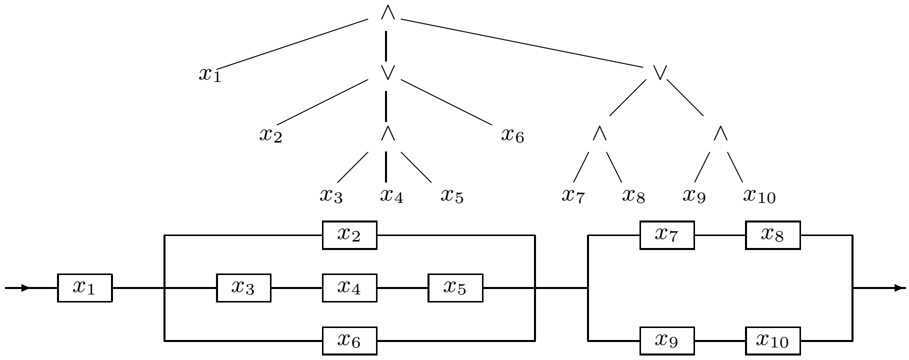


FIGURE 12. An and–or positive proposition of the conjunctive type (top), its associated tree (middle), and an equivalent planar series-parallel network of the serial type (bottom).

This is not *a priori* of the type corresponding to Proposition I.5 because *not* all nodes contribute to size in this particular application. However, the quadratic equation induced by (55) can be solved, giving

$$\begin{aligned} S(z) &= \frac{1}{4}\left(1 + z - \sqrt{1 - 6z + z^2}\right) \\ &= z + z^2 + 3z^3 + 11z^4 + 45z^5 + 197z^6 + 903z^7 + 4279z^8 + 20793z^9 \\ &\qquad + 103049z^{10} + 518859z^{11} + \cdots , \end{aligned}$$

where the coefficients are *EIS* **A001003**. (These numbers also count series-parallel networks of a specified type (e.g., serial in Figure 12, bottom), where placement in the plane matters.)

In an instructive paper, Stanley [**392**] discusses a page of Plutarch's *Moralia* where there appears the following statement:

> "Chrysippus says that the number of compound propositions that can be made from only ten simple propositions exceeds a million. (Hipparchus, to be sure, refuted this by showing that on the affirmative side there are 103,049 compound statements, and on the negative side 310,952.)"

It is notable that the tenth number of Hipparchus of Rhodes[7] (*c.* 190–120B.C.) is precisely $S_{10} = 103,049$. This is, for instance, the number of logical formulæ that can be formed from ten boolean variables $x_1, \ldots, x_{10}$ (used once each and in this order) using and–or connectives in alternation (no "negation"), upon starting from the top in some conventional fashion (e.g, with an and-clause); see Figure 12[8]. Hipparchus was naturally not cognizant of generating functions,

[7]This was first observed by David Hough in 1994; see [**392**]. In [**220**], Habsieger *et al.* further note that $\frac{1}{2}(S_{10} + S_{11}) = 310,954$, and suggest a related interpretation (based on negated variables) for the other count given by Hipparchus.

[8]Any functional term admits a unique tree representation. Here, as soon as the root type has been fixed (e.g., an $\wedge$ connective), the others are determined by level parity. The constraint of node degrees ≥ 2 in the tree means that no superfluous connectives are used. Finally, any monotone boolean expression can be represented by a series-parallel network: the x_j are viewed as switches with the *true* and *false* values being associated with closed and open circuits, respectively.

Tree variety		1	2	3	4	5	6	7	8	n	$+\infty$
Plane gen.	$\mathcal{G} = \mathcal{Z} \times \text{SEQ}(\mathcal{G})$	1	1	2	5	14	42	132	429	$\frac{1}{n}\binom{2n-2}{n-1}$	$\sim 4^{n-1}/\sqrt{\pi n^3}$
Plane bin.	$\mathcal{T} = \mathcal{Z} + \text{SEQ}_2(\mathcal{T})$	1	1	2	5	14	42	132	429	$\frac{1}{n}\binom{2n-2}{n-1}$	$\sim 4^{n-1}/\sqrt{\pi n^3}$
Unord. gen.	$\mathcal{H} = \mathcal{Z} \times \text{MSET}(\mathcal{H})$	1	1	2	4	9	20	48	115	–	$\sim \lambda \cdot \beta^n / n^{3/2}$
Unord. bin.	$\mathcal{U} = \mathcal{Z} + \text{MSET}_2(\mathcal{U})$	1	1	1	2	3	6	11	23	–	$\lambda_2 \cdot \beta_2^n / n^{3/2}$

FIGURE 13. The number of rooted trees of type plane/unordered and general/binary for $n = 1\,..\,8$ and the corresponding asymptotic forms. There, $\lambda \doteq 0.43992$, $\beta \doteq 2.95576$ for unordered general (*EIS* **A000081**); $\lambda_2 \doteq 0.31877$, $\beta_2 \doteq 2.48325$ for unordered binary. For binary trees (*EIS* **A001190**), size is by, convention here, the number of external nodes.

but with the technology of the time (and a rather remarkable mind!), he would still be able to discover a recurrence equivalent to (55),

$$S_n = [\![n \geq 2]\!] \left(\sum_{n_1+\cdots+n_k=n} S_{n_1} S_{n_2} \cdots S_{n_k} \right) + [\![n = 1]\!], \tag{56}$$

where the sum has only 42 essentially different terms for $n = 10$ (see [**392**] for a discussion), and finally determine S_{10}. . . . . END OF EXAMPLE 14. □

▷ **38.** *The Lagrangean form of Schröder's GF.* The generating function $S(z)$ admits the form

$$S(z) = z\phi(S(z)) \quad \text{where} \quad \phi(y) = \frac{1-y}{1-2y}$$

is the OGF of compositions. Consequently, one has

$$\begin{aligned} S_n &= \frac{1}{n}[u^{n-1}] \left(\frac{1-u}{1-2u} \right)^n \\ &= \frac{(-1)^{n-1}}{n} \sum_k (-2)^k \binom{n}{k+1} \binom{n+k-1}{k} \\ &= \frac{1}{n} \sum_{k=0}^{n-2} \binom{2n-k-2}{n-1} \binom{n-2}{k}. \end{aligned}$$

Is there a direct combinatorial relation to compositions? ◁

▷ **39.** *Faster determination of Schröder numbers.* By forming a differential equation satisfied by $S(z)$ and extracting coefficients, one obtains a recurrence

$$(n+2)S_{n+2} - 3(2n+1)S_{n+1} + (n-1)S_n = 0, \qquad n \geq 1,$$

that entails a fast determination (in linear time) of the S_n. In contrast, Hipparchus's recurrence implies an algorithm of complexity $e^{O(\sqrt{n})}$ in the number of arithmetic operations involved. ◁

I. 5.2. Nonplane tree. An *unordered tree*, also called *nonplane* tree, is a tree in the general graph–theoretic sense, so that there is no order distinction between subtrees emanating from a common node. The unordered trees considered here are furthermore rooted, meaning that one of the nodes is distinguished as the root. Accordingly, in the language of constructible structures, a rooted *unordered* tree is a root node linked to a *multiset* of trees. Thus, the class $\mathcal{H}$ of all unordered trees, admits the recursive specification:

$$\mathcal{H} = \mathcal{Z} \times \text{MSET}(\mathcal{H}) \quad \Longrightarrow \quad \begin{cases} H(z) = z \displaystyle\prod_{m=1}^{\infty} (1 - z^m)^{-H_m} \\ \quad = z \exp\left(H(z) + \frac{1}{2}H(z^2) + \frac{1}{3}H(z^3) + \cdots\right). \end{cases}$$

The first form of the OGF was given by Cayley in 1857 [**52**, p. 43]; it does not admit a closed form solution, though the equation permits one to determine all the H_n recurrently (*EIS* **A000081**)

$$H(z) = z + z^2 + 2z^3 + 4z^4 + 9z^5 + 20z^6 + 48z^7 + 115z^8 + 286z^9 + \cdots .$$

In addition, the local analysis of the singularities of $H(z)$ yields a *bona fide* asymptotic expansion for H_n, a fact first discovered by Pólya [**349**] who proved that

$$H_n \sim \lambda \cdot \frac{\beta^n}{n^{3/2}}, \tag{57}$$

for some positive constants $\lambda \doteq 0.43992$ and $\beta \doteq 2.95576$ (see Chapter VII).

$\triangleright$ **40.** *Fast determination of the Cayley–Pólya numbers.* Logarithmic differentiation of the equation satisfied by $H(z)$ provides for the H_n a recurrence that permits one to compute H_n in time polynomial in n. (Note: a similar technique applies to the partition numbers P_n; see p. 39.) $\triangleleft$

The enumeration of the class of trees defined by an arbitrary set Ω of nodes degree immediately results from the translation of sets of fixed cardinality.

PROPOSITION I.6. *Let $\Omega \subset \mathbb{N}$ be a finite set of integers containing 0. The OGF $U(z)$ of nonplane trees with degrees constrained to lie in Ω satisfies a functional equation of the form*

$$U(z) = z\Phi(U(z), U(z^2), U(z^3), \ldots), \tag{58}$$

for some computable polynomial Φ.

PROOF. The class of trees satisfies the combinatorial equation,

$$\mathcal{U} = \mathcal{Z} \times \text{MSET}_{\Omega}(\mathcal{U}) \qquad \left(\text{MSET}_{\Omega}(\mathcal{U}) \equiv \sum_{\omega \in \Omega} \text{MSET}_{\omega}(\mathcal{U})\right),$$

where the multiset construction reflects non-planarity, since subtrees stemming from a node can be freely rearranged between themselves and may appear repeated. Theorem I.3 (p. 77) provides the translation of $\text{MSET}_k(\mathcal{U})$:

$$\Phi(U(z), U(z^2), U(z^3), \ldots) = \sum_{\omega \in \Omega} [u^{\omega}] \exp\left(\frac{u}{1}U(z) + \frac{u^2}{2}U(z^2) + \cdots\right).$$

The result follows. $\square$

Once more, there are no explicit formulæ but only functional equations implicitly determining the generating functions. However, as we shall see in Chapter VII, the equations may be used to analyse the dominant singularity of $U(z)$. It is found that a "universal" law governs the singularities of simple tree generating functions that are of the type $\sqrt{1-z/\rho}$, corresponding to a general asymptotic scheme (see Figure 13),

$$U_n^{\Omega} \sim \lambda_{\Omega} \frac{(\beta_{\Omega})^n}{\sqrt{n^3}}. \tag{59}$$

Many of these questions have their origin in combinatorial chemistry, starting with Cayley in the 19th century [**52**, Ch. 4]. Pólya reexamined these questions, and in his important paper published in 1937 [**347**] he developed at the same time a general theory of combinatorial enumerations under group actions and of asymptotics methods giving rise to estimates like (59). See the book by Harary and Palmer [**223**] for more on this topic or Read's edition of Pólya's paper [**349**].

▷ **41.** *Binary nonplane trees.* Unordered binary trees with size measured by the number of external nodes are described by the equation $\mathcal{U} = \mathcal{Z} + \text{MSET}_2(\mathcal{U})$. The functional equation determining $U(z)$ is

$$U(z) = z + \frac{1}{2}U(z)^2 + \frac{1}{2}U(z^2); \quad U(z) = z + z^2 + z^3 + 2z^4 + 3z^5 + \cdots . \tag{60}$$

The asymptotic analysis of the coefficients (*EIS* **A001190**) was carried out by Otter [**335**] who established an estimate of type (59). (The values of the constants are summarized in Figure 13.) The quantity U_n is also the number of structurally distinct products of n elements under a commutative nonassociative binary operation. ◁

▷ **42.** *Hierarchies.* Define the class $\mathcal{K}$ of hierarchies to be trees without nodes of outdegree 1 and size determined by the number of external nodes. The corresponding OGF satisfies (Cayley 1857, see [**52**, p.43])

$$K(z) = \frac{1}{2}z + \frac{1}{2}\left[\exp(K(z) + \frac{1}{2}K(z^2) + \cdots) - 1\right],$$

from which the first values are found (*EIS* **A000669**)

$$K(z) = z + z^2 + 2z^3 + 5z^4 + 12z^5 + 33z^6 + 90z^7 + 261z^8 + 766z^9 + 2312z^{10} + \cdots .$$

These numbers also enumerate hierarchies in statistical classification theory [**417**]. They are the non-planar analogues of the Hipparchus–Schröder's numbers on p. 64. ◁

▷ **43.** *Nonplane series-parallel networks.* Consider the class $\mathcal{SP}$ of series-parallel networks as previously considered in relation to Hipparchus of Rhodes' example, p. 65, but ignoring planar embeddings. Thus, all parallel arrangements of the (serial) networks $s_1, \ldots, s_k$ are considered equivalent, while the linear arrangement in each serial network matters. For instance, for $n = 2, 3$:

```
-o--o-   |-o-|     -o--o--o-   |-o-|   |-o-o-|   -o--|-o-|-  -|-o-|--o-
         |-o-|                 |-o-|   |--o--|       |-o-|     |-o-|
                               |-o-|
```

Thus, $SP_2 = 2$ and $SP_3 = 5$. This is modelled by the grammar:

$$\mathcal{S} = \mathcal{Z} + \text{SEQ}_{\geq 2}(\mathcal{P}), \qquad \mathcal{P} = \mathcal{Z} + \text{MSET}_{\geq 2}(\mathcal{S}),$$

and, avoiding to count networks of one element twice,

$$SP(z) = S(z) + P(z) - z = z + 2z^2 + 5z^3 + 15z^4 + 48z^5 + 167z^6 + 602z^7 + 2256z^8 + \cdots .$$

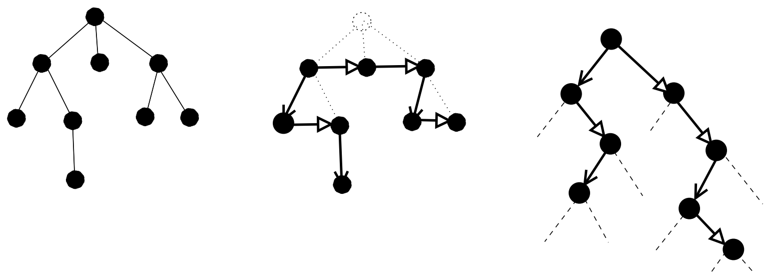
This is *EIS* **A003430**. The objects are usually described as networks of electric resistors. ◁

I. 5.3. Related constructions. Trees underlie recursive structures of all sorts. A first illustration is provided by the fact that the Catalan numbers, $\mathrm{C}_n = \frac{1}{n+1}\binom{2n}{n}$ count general trees ($\mathcal{G}$) of size $n+1$, binary trees ($\mathcal{B}$) of size n (if size is defined as the number of internal nodes), as well as triangulations ($\mathcal{T}$) comprised of n triangles. The combinatorialist John Riordan even coined the name *Catalan domain* for the area within combinatorics that deals with objects enumerated by Catalan numbers, and Stanley's book contains an exercise [**393**, Ex. 6.19] whose statement alone spans ten full pages, with a lists of 66 types of objects(!) belonging to the Catalan domain. We shall illustrate the importance of Catalan numbers by describing a few fundamental correspondences that explain the occurrence of Catalan numbers in several areas of combinatorics.

Rotation of trees. The combinatorial isomorphism relating $\mathcal{G}$ and $\mathcal{B}$ (albeit with a shift in size) coincides with a classical technique of computer science [**268**, §2.3.2]. To wit, a general tree can be represented in such a way that every node has two types of links, one pointing to the leftmost child, the other to the next sibling in left-to-right order. Under this representation, if the root of the general tree is left aside, then every node is linked to two other (possibly empty) subtrees. In other words, general trees with n nodes are equinumerous with pruned binary trees with $n-1$ nodes:

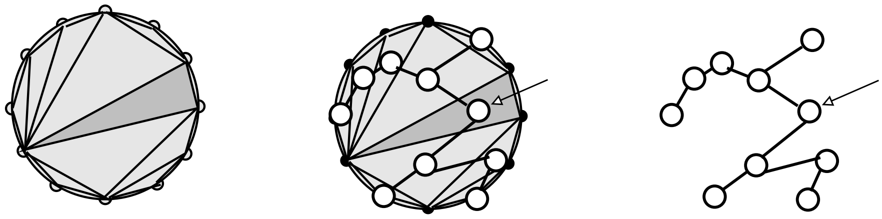
$$\mathcal{G}_n \cong \mathcal{B}_{n-1}.$$

Graphically, this is illustrated as follows:



The rightmost tree is a binary tree drawn in a conventional manner, following a 45° tilt. This justifies the name of "rotation correspondence" often given to this transformation.

Tree decomposition of triangulations. The relation betwen binary trees $\mathcal{B}$ and triangulations $\mathcal{T}$ is equally simple: draw a triangulation; define the root triangle as the one that contains the edge connecting two designated vertices (for instance, the vertices numbered 0 and 1); associate to the root triangle the root of a binary tree; next, associate recursively to the subtriangulation on the left of the root triangle a left subtree; do similarly for the right subtriangulation giving rise to a right subtree.

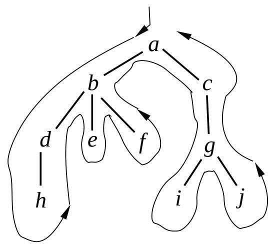


Under this correspondence, tree nodes correspond to triangle faces, while edges connect adjacent triangles. What this correspondence proves is the combinatorial isomorphism

$$\mathcal{T}_n \cong \mathcal{B}_n.$$

We turn next to another type of objects that are in correspondence with trees. These can be interpreted as words encoding tree traversals and, geometrically, as paths in the discrete plane $\mathbb{Z} \times \mathbb{Z}$.

Tree codes and Łukasiewicz words. . Any tree can be traversed starting from the root, proceeding depth-first (and left-to-right), and backtracking upwards once a subtree has been completely traversed. For instance, in the tree

(61) $\tau =$ 

the first visits to nodes take place in the following order

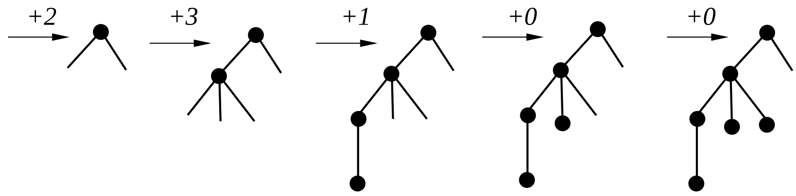
$$a, \quad b, \quad d, \quad h, \quad e, \quad f, \quad c, \quad g, \quad i, \quad j\,.$$

(Note: the tags $a, b, \ldots$ added for convenience in order to distinguish nodes have no special meaning; only the abstract tree shape matters here.) This order is known as *preorder* or *prefix order* since a node is preferentially visited before its children.

Given a tree, the listing of the outdegrees of nodes in prefix order will be called the preorder degree sequence. For the tree of (61), this is

$$\sigma = (2, 3, 1, 0, 0, 0, 1, 2, 0, 0).$$

It is a fact that the degree sequence determines the tree unambiguously. Indeed, given the degree sequence, the tree is reconstructed step by step, adding nodes one after the other at the leftmost available place. For σ, the first steps are then



Next, if one represents degree j by a "symbol" f_j, then the degree sequence becomes a *word* over the infinite alphabet $\mathcal{F} = \{f_0, f_1, \ldots\}$, for instance,

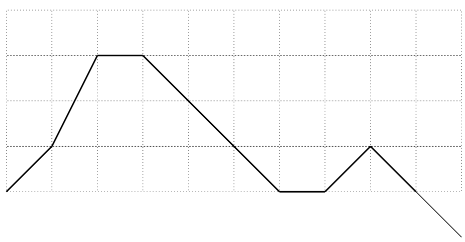
$$\sigma \rightsquigarrow f_2 f_3 f_1 f_0 f_0 f_0 f_1 f_2 f_0 f_0.$$

This can be interpreted in logical language as a denotation for a functional term built out symbols from $\mathcal{F}$, where f_j represents a function of degree (or "arity") j. The correspondence even becomes obvious if superfluous parentheses are added at appropriate place to delimitate scope:

$$\sigma \rightsquigarrow f_2(f_3(f_1(f_0), f_0, f_0), f_1(f_2(f_0, f_0))).$$

Such codes are known as Łukasiewicz codes[9], in recognition of the work of the Polish logician with that name. Jan Łukasiewicz (1878–1956) introduced them in order to completely specify the *syntax* of terms in various logical calculi; they prove nowadays basic in the development of parsers and compilers in computer science.

Finally, a tree code can be rendered as a walk over the discrete lattice $\mathbb{Z} \times \mathbb{Z}$. Associate to any f_j (i.e., any node of outdegree j) the displacement $(1, j-1) \in \mathbb{Z} \times \mathbb{Z}$, and plot the sequence of moves starting from the origin. On the example one finds:



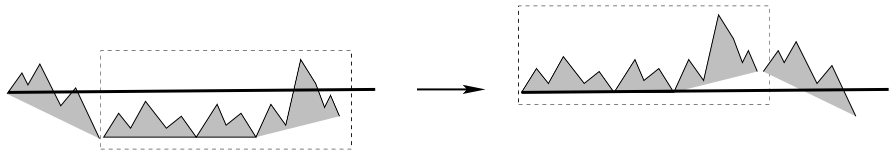
f_2	f_3	f_1	f_0	f_0	f_0	f_1	f_2	f_0	f_0
1	2	0	-1	-1	-1	0	1	-1	-1

There, the last line represents the vertical displacements. The resulting paths are known as Łukasiewicz paths. Such a walk is then characterized by two conditions: the vertical displacements are in the set $\{-1, 0, 1, 2, \ldots\}$; all its points, except for the very last step, lie in the upper half-plane.

By this correspondence, the number of Łukasiewicz paths with n steps is the shifted Catalan number, $\frac{1}{n}\binom{2n-2}{n-1}$.

[9]A less dignified name is "Polish prefix notation". The "reverse Polish notation" is a variant based on postorder that has been used in some calculators since the 1970's.

$\triangleright$ **44.** *Conjugacy principle and cycle lemma.* Let $\mathcal{L}$ be the class of all Łukasiewicz paths. Define a "relaxed" path as one that starts at level 0, ends at level -1 but is otherwise allowed arbitrary negative steps; let $\mathcal{M}$ be the corresponding class. Then, each relaxed path can be cut-and-pasted uniquely after its leftmost minimum as described here:



This associates to every relaxed path of length ν a unique standard path. A bit of combinatorial reasoning shows that correspondence is 1-to-ν (each element of $\mathcal{L}$ has *exactly* ν preimages.) One thus has $M_\nu = \nu L_\nu$. This correspondence preserves the number of steps of each type $(f_0, f_1, \ldots)$, so that the number of Łukasiewicz paths with ν_j steps of type f_j is

$$\frac{1}{\nu}[x^{-1}u_0^{\nu_0}u_1^{\nu_1}\cdots]\left(x^{-1}u_0 + u_1 + xu_2 + x^2u_3 + \cdots\right)^\nu = \frac{1}{\nu}\binom{\nu}{\nu_0, \nu_1, \ldots},$$

under the necessary condition $(-1)\nu_0 + 0\nu_1 + 1\nu_2 + 2\nu_3 + \cdots = -1$. This combinatorial way of obtaining refined Catalan statistics is known as the *conjugacy principle* [**359**] or the *cycle lemma* [**82, 102, 120**]. Raney has derived from it a purely combinatorial proof of the Lagrange inversion formula [**359**] while Dvoretzky & Motzkin [**120**] have employed this technique to solve a number of counting problems related to circular arrangements. $\triangleleft$

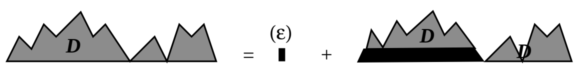
EXAMPLE 15. *Binary tree codes and Dyck paths.* Walks associated with binary trees have a very special form since the vertical displacements can only be $+1$ or -1. The resulting paths of Łukasiewicz type are then equivalently characterized as sequences of numbers $x = (x_0, x_1, \ldots, x_{2n}, x_{2n+1})$ satisfying the conditions

(62) $$x_0 = 0; \qquad x_j \geq 0 \quad \text{for } 1 \leq j \leq 2n; \qquad |x_{j+1} - x_j| = 1; \qquad x_{2n+1} = -1.$$

These coincide with "gambler ruin sequences", a familiar object from probability theory: a player plays head and tails. He starts with no capital ($x_0 = 0$) at time 0; his total gain is x_j at time j; he is allowed no credit ($x_j \geq 0$) and loses at the very end of the game $x_{2n+1} = -1$; his gains are ± 1 depending on the outcome of the coin tosses ($|x_{j+1} - x_j| = 1$).

It is customary to drop the final step and consider "excursions' that take place in the upper half-plane. The resulting objects defined as sequences $(x_0 = 0, x_1, \ldots, x_{2n} = 0)$ satisfying the first three conditions of (62) are known in combinatorics as *Dyck paths*[10]. By construction, Dyck paths of length $2n$ correspond bijectively to binary trees with n internal nodes and are consequently enumerated by Catalan numbers. Let $\mathcal{D}$ be the combinatorial class of Dyck paths, with size defined as length. This property can also be checked directly: the quadratic decomposition

(63)



$$\begin{array}{rcl} \mathcal{D} & = & \{\epsilon\} + (\nearrow \mathcal{D} \searrow) \times \mathcal{D} \\ \Longrightarrow \quad D(z) & = & 1 + (zD(z)z)\, D(z). \end{array}$$

[10]Dyck paths are closely associated with free groups on one generator and are named after the German mathematician Walther (von) Dyck (1856–1934) who introduced free groups around 1880.

From this OGF, the Catalan number are found (as expected): $D_{2n} = \frac{1}{n+1}\binom{2n}{n}$. The decomposition (63) is known as the "first passage" decomposition as it is based on the first time the cumulated gains in the coin-tossing game pass through the value zero.

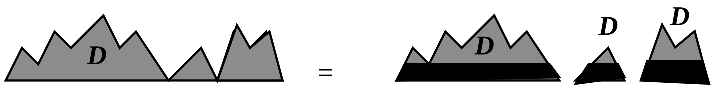
Dyck paths also arise in connection will well-parenthetized expressions. These are recognized by keeping a counter that records at each stage the excess of the number of opening brackets '(' over closing brackets ')'. Finally, one of the origins of Dyck path is the famous *ballot problem*, which goes back to the nineteenth century [**303**]: there are two candidates A and B that stand for election, $2n$ voters, and the election eventually results in a tie; what is the probability that A is always ahead of or tied with B when the ballots are counted? The answer is

$$\frac{D_{2n}}{\binom{2n}{n}} = \frac{1}{n+1},$$

since there are $\binom{2n}{n}$ possibilities in total, of which the number of favorable cases is D_{2n}, a Catalan number. The central rôle of Dyck paths and Catalan numbers in problems coming from such diverse areas is quite remarkable. Chapter V will present refined counting results regarding lattice paths. .... END OF EXAMPLE 15. □

▷ **45.** *Dyck paths, parenthesis systems, and general trees.* The class of Dyck paths admits an alternative sequence decomposition

(64)



$$\mathcal{D} \quad = \quad \text{SEQ}(\mathcal{Z} \times \mathcal{D} \times \mathcal{Z}),$$

which again leads to the Catalan GF. The decomposition (64) is known as the "arch decomposition" (see Subsection V. 3.1 for more). It can also be directly related to traversal sequences of general trees, but with the directions of *edge* traversals being recorded (instead of traversals based on node degrees): fo a general tree τ, define its encoding $\kappa(\tau)$ over the binary alphabet $\{\nearrow, \searrow\}$ recursively by the rules:

$$\kappa(\tau) = \epsilon, \qquad \kappa(\bullet(\tau_1, \ldots, \tau_r)) = \nearrow \kappa(\tau_1) \cdots \kappa(\tau_r) \searrow .$$

This is the classical representation of trees by a parenthesis system (interpret '$\nearrow$' and '$\searrow$' as '(' and ')', respectively), which associates to a tree of n nodes a path of length $2n - 2$. ◁

▷ **46.** *Random generation of Dyck paths.* Dyck paths of length $2n$ can be generated uniformly at random in time linear in n. (Hint: By the conjugacy principle of Note 44, it suffices to generate uniformly a sequence of n a's and $n + 1$ b's, then reorganize it according to the conjugacy principle. ◁

▷ **47.** *Motzkin paths and unary-binary trees.* Motzkin paths are defined by changing the third condition of (62) defining Dyck paths into $|x_{j+1} - x_j| \leq 1$. They appear as codes for unary-binary trees and are enumerated by the Motzkin numbers of Note 36. ◁

EXAMPLE 16. *The complexity of boolean functions.* Complexity theory provides many surprising applications of enumerative combinatorics and asymptotic estimates. In general, one starts with a finite set of mathematical objects Ω and a combinatorial class $\mathcal{D}$ of *descriptions*. By assumption, to every object of $\delta \in \mathcal{D}$ is associated an element $\mu(\delta) \in \Omega$, its "meaning";

conversely any object of Ω admits at least one description in $\mathcal{D}$, that is, the function μ is surjective. It is then of interest to quantify properties of the shortest description function defined for $\omega \in \Omega$ as

$$\sigma(\omega) := \min \left\{ |\delta|_{\mathcal{D}} \;\middle|\; \mu(\delta) = \omega \right\},$$

and called the *complexity* of element of Ω (with respect to $\mathcal{D}$).

We take here Ω to be the class of all boolean functions on m variables. Their number is $\|\Omega\| = 2^{2^m}$. As descriptions, we adopt the class of logical expressions involving the logical connectives $\vee, \wedge$ and pure or negated variables. Equivalently, $\mathcal{D}$ is the class of binary trees, where internal nodes are tagged by a logical disjunction ('$\vee$') or a conjunction ('$\wedge$'), and each external node is tagged by either a boolean variable of $\{x_1, \ldots, x_m\}$ or a negated variable of $\{\neg x_1, \ldots, \neg x_m\}$. Define the size of a tree description as the number of internal nodes, that is, the number of logical operators. Then, one has

$$D_n = \left(\frac{1}{n+1} \binom{2n}{n} \right) \cdot 2^n \cdot (2m)^{n+1}, \tag{65}$$

as seen by counting tree shapes and possibilities for internal as well as external node tags.

The crux of the matter is that if the inequality

$$\sum_{j=0}^{\nu} D_j < \|\Omega\|, \tag{66}$$

holds, then there are not enough descriptions of size $\leq \nu$ to exhaust Ω. In other terms, there must exist at least one object in Ω whose complexity exceeds ν. If the left side of (66) is much smaller than the right side, then, it must even be the case that "most" Ω-objects have a complexity that exceeds ν.

In the case of boolean functions and tree descriptions, the asymptotic form (24) is available. There results from (65) that, for n, ν getting large, one has

$$D_n = O(16^n m^n n^{-3/2}), \qquad \sum_{j=0}^{\nu} D_j = O(16^\nu m^\nu \nu^{-3/2}).$$

Choose ν such that the second expression is $o(\|\Omega\|)$. This is ensured for instance by taking for ν the value

$$\nu(m) := \frac{2^m}{4 + \log_2 m},$$

as verified by a simple asymptotic calculation. With this choice, one has the following suggestive statement:

> *A fraction tending to 1 (as $m \to \infty$) of boolean functions in m variables have tree complexity at least $2^m / \log_2 m$.*

Regarding upper bounds on boolean function complexity, a function always has a tree complexity that is at most $2^{m+1} - 3$. To see it, note that for $m = 1$, the 4 functions are

$$0 \equiv (x_1 \wedge \neg x_1), \quad 1 \equiv (x_1 \vee \neg x_1), \quad x_1, \quad \neg x_1.$$

Next, a function of m variables is representable by a technique known as the binary decision tree (BDT),

$$f(x_1, \ldots, x_{m-1}, x_m) = (\neg x_m \wedge f(x_1, \ldots, x_{m-1}, 0)) \vee (x_m \wedge f(x_1, \ldots, x_{m-1}, 1)),$$

which provides the basis of the induction as it reduces the representation of an m-ary function to the representation of two $(m-1)$-ary functions, consuming on the way three logical connectives.

Altogether, basic counting arguments have shown that "most" boolean functions have a tree-complexity that is "close" to the maximum possible, namely, $O(2^m)$. A similar result has been established by Shannon for the measure called circuit complexity: circuits are more powerful than trees, but Shannon's result states that *almost all boolean functions of m variables have circuit complexity $O(2^m/m)$*. See [**422**], especially the chapter by Li and Vitányi, for a discussion of such counting techniques within the framework of complexity theory. . . . . END OF EXAMPLE 16. $\square$

We finally conclude with a vast generalization of the previous examples.

DEFINITION I.13. *A class $\mathcal{T}$ of trees is said to be a* context-free variety of trees *if it coincides with the first component of a system of equations ($\mathcal{T} = \mathcal{S}_1$) of a recursive system*

$$\left\{\begin{array}{lcl} \mathcal{S}_1 & = & \Phi_1(\mathcal{Z}, \mathcal{S}_1, \ldots, \mathcal{S}_r) \\ \vdots & \vdots & \vdots \\ \mathcal{S}_r & = & \Phi_r(\mathcal{Z}, \mathcal{S}_1, \ldots, \mathcal{S}_r), \end{array}\right. \tag{67}$$

where each Φ_j is a constructor that only involves the operations of combinatorial sum ($+$) and cartesian product ($\times$), as well possibly as the neutral class, $\mathcal{E} = \{\epsilon\}$.

A combinatorial class $\mathcal{C}$ is said to be context-free *if it is combinatorially isomorphic to a context-free variety of trees: $\mathcal{C} \cong \mathcal{T}$.*

The classes of general trees ($\mathcal{G}$) and binary trees ($\mathcal{B}$) are context-free varieties of trees since they are specifiable as

$$\left\{\begin{array}{lcl} \mathcal{G} & = & \mathcal{Z} \times \mathcal{F} \\ \mathcal{F} & = & \{\epsilon\} + (\mathcal{G} \times \mathcal{F}) \end{array}\right. , \qquad \mathcal{B} = \mathcal{Z} + (\mathcal{B} \times \mathcal{B}).$$

($\mathcal{F}$ designates ordered forests of general trees.) The Łukasiewicz language and the set of Dyck paths are context-free classes since they are bijectively equivalent to $\mathcal{G}$ and $\mathcal{T}$.

This terminology is an extension of the concept of context-free language in the theory of formal languages; there, one defines a context-free language as the language formed with words that are obtained as sequences of leaf tags (read in left-to-right order) of a context-free variety of trees. In formal linguistics, the one-to-one mapping between trees and words is not generally imposed; when it is satisfied, the context-free language is said to be *unambiguous*, since words and trees determine each other uniquely.

An immediate consequence of admissibility theorems is the following proposition first encountered by Chomsky and Schützenberger [**80**] in the course of their research relating formal languages and formal power series:

PROPOSITION I.7. *A combinatorial class $\mathcal{C}$ that is* context-free *admits an OGF that is an* algebraic function. *In other words, there exists a (non-null) bivariate polynomial $P(z, y) \in \mathbb{C}[z, y]$ such that*

$$P(z, C(z)) = 0.$$

PROOF. The context-free system (67) translates into a system

$$\begin{cases} S_1(z) & = & \Psi_1(z, S_1(z), \ldots, S_r(z)) \\ \vdots & \vdots & \vdots \\ S_r(z) & = & \Psi_r(z, S_1(z), \ldots, S_r(z)), \end{cases}$$

where the Ψ_j are polynomials. This follows by the basic sum and product rules.

It is then well-known that algebraic elimination is possible in polynomials systems. Here, it is possible to eliminate the auxiliary variables $S_2, \ldots, S_r$, one by one, preserving the polynomial character of the system at each stage. The end result is then a single polynomial equation satisfied by $C(z) \equiv S_1(z)$. (Methods for effectively performing polynomial elimination include a repeated use of resultants and Groebner basis algorithms; see APPENDIX B: *Algebraic elimination*, p. 657 for a brief discussion and references.) □

Proposition I. 5.3 justifies the importance of algebraic functions in enumerative theory and it will be put to use in later chapters of this book; see especially Chapter VII for examples accompanied by asymptotic analyses. It constitutes a counterpart of Proposition I.3 which asserts that rational generating functions arise from finite state devices.

I. 6. Additional constructions

This section is devoted to the the constructions of sequences, sets, and cycles in the presence of restrictions on the number of components as well as to mechanisms that enrich the framework of core constructions, namely, pointing, substitution, and the use of implicit combinatorial definitions.

I. 6.1. Restricted constructions. An immediate formula for OGFs is that of the *diagonal* Δ of a cartesian product $\mathcal{B} \times \mathcal{B}$ defined as

$$\mathcal{A} \equiv \Delta(\mathcal{B} \times \mathcal{B}) := \{(\beta, \beta) \mid \beta \in \mathcal{B}\}.$$

Then, clearly $A_{2n} = B_n$ so that

$$A(z) = B(z^2).$$

The diagonal construction permits us to access the class of all unordered pairs of (distinct) elements of $\mathcal{B}$, which is $\mathcal{A} = \mathrm{PSET}_2(\mathcal{B})$. A direct argument then runs as follows: the unordered pair $\{\alpha, \beta\}$ is associated to the two ordered pairs (α, β) and (β, α) except when $\alpha = \beta$, where an element of the diagonal is obtained. In other words, one has the combinatorial isomorphism,

$$\mathrm{PSET}_2(\mathcal{B}) + \mathrm{PSET}_2(\mathcal{B}) + \Delta(B \times B) \cong B \times B,$$

meaning that

$$2A(z) + B(z^2) = B(z)^2.$$

The resulting translation into OGFs is thus

$$\mathcal{A} = \mathrm{PSET}_2(\mathcal{B}) \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 - \frac{1}{2}B(z^2).$$

Similarly, for multisets, we find

$$\mathcal{A} = \text{MSET}_2(\mathcal{B}) \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 + \frac{1}{2}B(z^2),$$

while for cycles one has $\text{CYC}_2 \cong \text{MSET}_2$, and

$$\mathcal{A} = \text{CYC}_2(\mathcal{B}) \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 + \frac{1}{2}B(z^2).$$

This type of direct reasoning could be extended to treat triples, and so on, but the computations (if not the reasoning) tend to grow out of control. An approach based on multivariate generating functions generates *simultaneously* all cardinality restricted constructions.

THEOREM I.3 (Component-restricted constructions). *The OGF of sequences with k components $\mathcal{A} = \text{SEQ}_k(\mathcal{B})$ satisfies*

$$A(z) = B(z)^k.$$

The OGF of sets, $\mathcal{A} = \text{PSET}_k(\mathcal{B})$, is a polynomial in the quantities $B(z), \dots, B(z^k)$,

$$A(z) = [u^k] \exp\left(\frac{u}{1}B(z) - \frac{u^2}{2}B(z^2) + \frac{u^3}{3}B(z^3) - \cdots\right).$$

The OGF of multisets, $\mathcal{A} = \text{MSET}_k(\mathcal{B})$, is

$$A(z) = [u^k] \exp\left(\frac{u}{1}B(z) + \frac{u^2}{2}B(z^2) + \frac{u^3}{3}B(z^3) + \cdots\right).$$

The OGF of cycles, $\mathcal{A} = \text{CYC}_k(\mathcal{B})$, is

$$A(z) = [u^k] \sum_{\ell=1}^{\infty} \frac{\varphi(\ell)}{\ell} \log \frac{1}{1 - u^\ell B(z^\ell)}.$$

The explicit forms for small values of k are summarized in Figure 14.

PROOF. The result for sequences is obvious since $\text{SEQ}_k(\mathcal{B})$ means $\mathcal{B} \times \cdots \times \mathcal{B}$ (k times). For the other constructions, the proof makes use of the techniques of Theorem I.1, but it is best based on bivariate generating functions that are otherwise developed fully in Chapter III to which we refer for details. The idea consists in describing all composite objects and introducing a supplementary marking variable to keep track of the number of components.

Take $\mathfrak{K}$ to be a construction amongst $\text{SEQ}, \text{CYC}, \text{MSET}, \text{PSET}$, set $\mathcal{A} = \mathfrak{K}(\mathcal{B})$, and let $\chi(\alpha)$ for $\alpha \in \mathcal{A}$ be the parameter "number of $\mathcal{B}$–components". Define the multivariate quantities

$$\begin{array}{lcl} A_{n,k} & := & \text{card}\left\{\alpha \in \mathcal{A} \;\middle|\; |\alpha| = n,\, \chi(\alpha) = k\right\} \\ A(z,u) & := & \displaystyle\sum_{n,k} A_{n,k} u^k z^n = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|} u^{\chi(\alpha)}. \end{array}$$

For instance, a direct calculation shows that, for sequences, there holds

$$\begin{aligned} A(z,u) &= \sum_{k\geq 0} u^k B(z)^k \\ &= \frac{1}{1-uB(z)}. \end{aligned}$$

For multisets and powersets, a simple adaptation of the already seen argument gives $A(z,u)$ as

$$A(z,u) = \prod_n (1-uz^n)^{-B_n}, \qquad A(z,u) = \prod_n (1+uz^n)^{B_n},$$

respectively. The result follows from there by the exp-log transformation upon extracting $[u^k]A(z,u)$. The case of cycles results from the bivariate generating function for cycles derived in APPENDIX A: *Cycle construction*, p. 646. □

▷ **48.** *Sets with distinct component sizes.* Let $\mathcal{A}$ be the class of the finite sets of elements from $\mathcal{B}$, with the additional constraint that no two elements in a set have the same size. One has

$$A(z) = \prod_{n=1}^{\infty} (1 + B_n z^n).$$

Similar identities serve in the analysis of polynomial factorization algorithms [**155**]. ◁

▷ **49.** *Sequences without repeated components.* The generating function is formally:

$$\int_0^\infty \exp\left(\sum_{k\geq 1} (-1)^{j-1} \frac{u^j}{j} A(z^j)\right) e^{-u}\, du.$$

(This form is based on the Eulerian integral: $k! = \int_0^\infty e^{-u} u^k \, du$.) ◁

I. 6.2. Pointing and substitution. Two more constructions, namely pointing and substitution, translate agreeably into generating functions. Combinatorial structures are viewed here as formed of “atoms” (words are composed of letters, graphs of nodes, etc) which determine their sizes. In this context, pointing means “pointing at a distinguished atom”; substitution, written $\mathcal{B} \circ \mathcal{C}$ or $\mathcal{B}[\mathcal{C}]$, means “substitute elements of $\mathcal{C}$ for atoms of $\mathcal{B}$”.

DEFINITION I.14. *Let $\{\epsilon_1, \epsilon_2, \ldots\}$ be a fixed collection of distinct neutral objects of size 0. The* pointing *of a class $\mathcal{B}$, noted $\mathcal{A} = \Theta\mathcal{B}$, is formally defined by*

$$\Theta\mathcal{B} := \sum_{n\geq 0} \mathcal{B}_n \times \{\epsilon_1, \ldots, \epsilon_n\}.$$

The substitution *of $\mathcal{C}$ into $\mathcal{B}$ (also known as composition of $\mathcal{B}$ and $\mathcal{C}$), noted $\mathcal{B} \circ \mathcal{C}$ or $\mathcal{B}[\mathcal{C}]$, is formally defined as*

$$\mathcal{B} \circ \mathcal{C} \equiv \mathcal{B}[\mathcal{C}] := \sum_{k\geq 0} \mathcal{B}_k \times \mathrm{SEQ}_k(\mathcal{C}).$$

If B_n is the number of $\mathcal{B}$ structures of size n, then nB_n can be interpreted as counting pointed structures where *one* of the n atoms composing a $\mathcal{B}$-structure has been distinguished (here by a special "pointer" of size 0 attached to it). Elements of $\mathcal{B} \circ \mathcal{C}$ may also be viewed as obtained by selecting in all possible ways an element $\beta \in \mathcal{B}$ and replacing each of its atoms by an arbitrary element of $\mathcal{C}$.

The interpretations above rely (silently) on the fact that atoms in an object can be eventually distinguished from each other. This can be obtained by "canonicalizing"[11] the representations of objects: first define inductively the lexicographic ordering for products and sequences; next represent powersets and multisets as increasing sequences with the induced lexicographic ordering (more complicated rules can also canonicalize cycles). In this way, any constructible object admits a unique "rigid" representation in which each particular atom is determined by its place. Such a canonicalization thus reconciles the abstract definition, Definition I.14, and the intuitive interpretation of pointing and substitution.

THEOREM I.4 (Pointing and substitution). *The constructions of pointing and substitution are admissible*[12]*:*

$$\begin{aligned} \mathcal{A} = \Theta\mathcal{B} \quad &\Longrightarrow \quad A(z) = z\partial_z B(z) \quad \partial_z := \frac{d}{dz} \\ \mathcal{A} = \mathcal{B} \circ \mathcal{C} \quad &\Longrightarrow \quad A(z) = B(C(z)) \end{aligned}$$

PROOF. By the definition of pointing, one has

$$A_n = n \cdot B_n \qquad \text{and} \qquad A(z) = z\frac{d}{dz}B(z).$$

From the definition of substitution, $\mathcal{A} = \mathcal{B}[\mathcal{C}]$ implies, by the sum and product rules,

$$A(z) = \sum_{k \geq 0} B_k \cdot (C(z))^k = B(C(z)),$$

and the proof is completed. $\square$

Permutations as pointed objects. As an example of pointing, consider the class $\mathcal{P}$ of all permutations written as words over integers starting from 1. One can go from a permutation of size $n-1$ to a permutation of size n by selecting a "gap" and inserting the value n. When this is done in all possible ways, it gives rise to the combinatorial relation

$$\mathcal{P} = \mathcal{E} + \Theta(\mathcal{Z} \times \mathcal{P}), \qquad \mathcal{E} = \{\epsilon\}, \qquad \Longrightarrow \qquad P(z) = 1 + z\frac{d}{dz}(zP(z)).$$

This means that the OGF satisfies an ordinary differential equation whose formal solution is $P(z) = \sum_{n \geq 0} n! z^n$.

[11] Such canonicalization techniques also serve to develop fast algorithms for the exhaustive listing of objects of a given size as well as for the range of problems known as "ranking" and "unranking", with implications in fast random generation. See, e.g., [**308, 326, 436**] for the general theory as well as [**357, 447**] for particular cases like necklaces and trees.

[12] In this book, we borrow from differential algebra the convenient notation $\partial_z := \frac{d}{dz}$ to represent derivatives.

Unary-binary trees as substituted objects. As an example of substitution, consider the class $\mathcal{B}$ of (plane rooted) binary trees, where all nodes contribute to size. If at each node there is substituted a linear chain of nodes (linked by edges placed on top of the node), one forms an element of the class $\mathcal{M}$ of unary-binary trees; in symbols:

$$\mathcal{M} = \mathcal{B} \circ \text{SEQ}_{\geq 1}(\mathcal{Z}) \qquad \Longrightarrow \qquad M(z) = B\left(\frac{z}{1-z}\right).$$

Thus from the known OGF, $B(z) = (1 - \sqrt{1-4z^2})/(2z)$, one derives

$$M(z) = \frac{1 - \sqrt{1 - 4z^2(1-z)^{-2}}}{2z(1-z)^{-1}} = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z},$$

which matches the direct derivation on p. 64 (Motzkin numbers).

$\triangleright$ **50.** *Combinatorics of derivatives.* The combinatorial operation $\mathbf{D}$ of "eraser–pointing" points to an atom in an object and replaces it by a neutral object, otherwise preserving the overall structure of the object. The translation of $\mathbf{D}$ on OGFs is then simply $\partial \equiv \partial_z$. Classical identities of analysis then receive simple combinatorial interpretations, for instance,

$$\partial(A \times B) = (A \times \partial B) + (\partial A) \times B);$$

Leibniz's identity, $\partial^m (f \cdot g) = \sum_j \binom{m}{j} (\partial^j f) \cdot (\partial^{m-j} g)$, also follows from basic combinatorics. Similarly, for the "chain rule" $\partial(f \circ g) = ((\partial f) \circ g) \cdot \partial g$. $\triangleleft$

I. 6.3. Implicit structures. There are many cases where a combinatorial class $\mathcal{X}$ is determined by a relation $\mathcal{A} = \mathcal{B} + \mathcal{X}$, where $\mathcal{A}$ and $\mathcal{B}$ are known. In terms of generating functions, one has $A(z) = B(z) + X(z)$, so that

$$\mathcal{A} = \mathcal{B} + \mathcal{X} \quad \Longrightarrow \quad X(z) = A(z) - B(z).$$

For instance, the autocorrelation technique of Section I. 4.2 makes it possible to describe the class $\mathcal{S}$ of all words in $\mathcal{W}$ that do *not* contain a given pattern $\mathfrak{p}$, whereas the language of words containing the pattern is determined as the solution in $\mathcal{X}$ of the equation $\mathcal{W} = \mathcal{S} + \mathcal{X}$; see p. 56. Similarly, for products, basic algebra gives

$$\mathcal{A} = \mathcal{B} \times \mathcal{X} \quad \Longrightarrow \quad X(z) = \frac{A(z)}{B(z)}.$$

Here are the corresponding solutions for two of the composite constructions.

THEOREM I.5 (Implicit specifications). *The generating functions associated to the implicit equations in $\mathcal{X}$*

$$\mathcal{A} = \text{SEQ}(\mathcal{X}), \quad \mathcal{A} = \text{MSET}(\mathcal{X})$$

are respectively

$$X(z) = 1 - \frac{1}{A(z)}, \quad X(z) = \sum_{k \geq 1} \frac{\mu(k)}{k} \log A(z^k),$$

where $\mu(k)$ is the Möbius function.

PROOF. For sequences, the relation $A(z) = (1 - X(z))^{-1}$ is readily inverted. For multisets, start from the fundamental relation of Theorem I.1 and take logarithms:

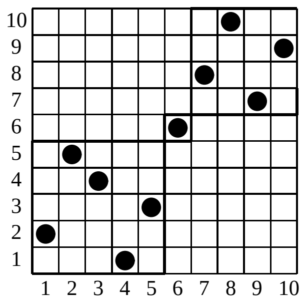
$$\log(A(z)) = \sum_{k=1}^{\infty} \frac{1}{k} X(z^k).$$

Let $L = \log A$ and $L_n = [z^n]L(z)$. One has

$$nL_n = \sum_{d \mid n} (dX_d),$$

to which it suffices to apply Möbius inversion; see APPENDIX A: *Arithmetical functions*, p. 639. □

EXAMPLE 17. *Indecomposable permutations.* A permutation $\sigma = \sigma_1 \cdots \sigma_n$ (written here as a word of distinct letters) is said to be *decomposable* if, for some $k < n$, $\sigma_1 \cdots \sigma_k$ is a permutation of $\{\sigma_1, \ldots, \sigma_k\}$, *i.e.*, a strict prefix of the permutation is itself a permutation. Any permutation decomposes uniquely as a catenation of indecomposable permutations; for instance, here is the decomposition of $\sigma = 2\,5\,4\,1\,3\,6\,8\,7\,10\,9$:



$$\sigma = \boxed{2\,5\,4\,1\,3}\ \boxed{6}\ \boxed{8\,10\,7\,9}$$

Thus the class $\mathcal{P}$ of all permutations and the class $\mathcal{I}$ of indecomposable ones are related by

$$\mathcal{P} = \text{SEQ}(\mathcal{I}).$$

This determines $I(z)$ implicitly, and Theorem I.5 gives:

$$I(z) = 1 - \frac{1}{P(z)} \quad \text{where} \quad P(z) = \sum_{n \geq 0} n!\, z^n .$$

This example illustrates the implicit structure theorem, but also the possibility of *bona fide* algebraic calculations with power series even in cases where they are divergent (APPENDIX A: *Formal power series*, p. 648). One finds

$$I(z) = z + z^2 + 3\,z^3 + 13\,z^4 + 71\,z^5 + 461\,z^6 + +3447\,z^7 + \cdots ,$$

where the coefficients are *EIS* **A003319** and

$$I_n = n! - \sum_{\substack{n_1+n_2=n \\ n_1,n_2 \geq 1}} (n_1! n_2!) + \sum_{\substack{n_1+n_2+n_3=n \\ n_1,n_2,n_3 \geq 1}} (n_1! n_2! n_3!) - \cdots .$$

From there, simple majorizations of the terms imply that $I_n \sim n!$, so that *almost all permutations are indecomposable*; see [**82**, p. 262]. . . . . END OF EXAMPLE 17. □

$\triangleright$ **51.** *2-dimensional wanderings.* A drunkard starts from the origin in the $\mathbb{Z} \times \mathbb{Z}$ plane and, at each second, he makes a step in either one of the four directions, NW, NE, SW, SE. The steps are thus $\nwarrow, \nearrow, \swarrow, \searrow$. Consider the class $\mathcal{L}$ of "primitive loops" defined as walks that start and end at the origin, but do not otherwise touch the origin. The GF of $\mathcal{L}$ is (*EIS* **A002894**)

$$L(z) = 1 - \frac{1}{\sum_{n=0}^{\infty} \binom{2n}{n}^2 z^{2n}} = 4\,z^2 + 20\,z^4 + 176\,z^6 + 1876\,z^8 + \cdots .$$

(Hint: a walk is determined by its projections on the horizontal and vertical axes; 1-dimensional walks that return to the origin in $2n$ steps are enumerated by $\binom{2n}{n}$.) In particular $[z^n]L(z/4)$ is the probability that the random walk first returns to the origin in n steps.

Such problems largely originate with Pólya and the implicit structure technique above was most likely known to him [**348**]. See [**64**] for similar multidimensional extensions. The first return analysis is given in Chapter VII, based on singularity analysis and Hadamard closure properties. $\triangleleft$

EXAMPLE 18. *Irreducible polynomials over finite fields.* Objects not apparently combinatorial can sometimes be enumerated by symbolic methods. Here is an indirect construction relative to polynomials over finite fields. We fix a prime number p and consider the base field $\mathbb{F}_p$ of integers taken modulo p. The polynomial ring $\mathbb{F}_p[X]$ is the ring of polynomials in X with coefficients taken in $\mathbb{F}_p$. For all practical purposes, one may restrict attention to polynomials that are monic, that is, whose leading coefficient is 1.

First, let $\mathcal{P}$ be the class of all monic polynomials, with the size of a polynomial being its degree. Since a monic polynomial of degree n is described by a choice of n coefficients, one has

$$P \cong \textsc{Seq}(\mathbb{F}_p) \qquad \Longrightarrow \qquad P(z) = \frac{1}{1-pz} \quad \text{and} \quad P_n = p^n.$$

A polynomial is said to be *irreducible* if it does not decompose as a product of two polynomials of smaller degrees. By unique factorization, each monic polynomial decomposes uniquely into a product (with repetitions being possible) of monic irreducible polynomials. For instance, over $\mathbb{F}_3$, one has

$$X^{10} + X^8 + 1 = (X+1)^2(X+2)^2(X^6 + 2X^2 + 1).$$

Let I be the set of monic irreducible polynomials. The combinatorial isomorphism

$$\mathcal{P} \cong \textsc{MSet}(\mathcal{I})$$

expresses precisely the unique factorization property. Thus, the irreducibles are determined implicitly from the class of all polynomials whose OGF is known. Theorem I.5 implies the identity

$$I(z) = \sum_{k \geq 1} \frac{\mu(k)}{k} \log \frac{1}{1 - pz^k},$$

and, upon extracting coefficients,

$$I_n = \frac{1}{n} \sum_{k \mid n} \mu(k) p^{n/k}.$$

In particular, I_n is asymptotic to p^n/n. This estimate constitutes the density theorem for irreducible polynomials:

> *The fraction of irreducible polynomials amongst all polynomials of degree n over the finite field $\mathbb{F}_p$ is asymptotic to $\frac{1}{n}$.*

This property is analogous to the Prime Number Theorem of number theory (which is technically much harder [**89**]), after which the proportion of prime numbers in the interval $[1, n]$ is asymptotic to $\frac{1}{\log n}$. (The derivation above is in essence due to Gauß. See Knopfmacher's book [**259**] for an abstract discussion of statistical properties of arithmetical semigroups.) END OF EXAMPLE 18. □

▷ **52.** *Square-free polynomials.* Let Q be the class of monic square-free polynomials (*i.e.*, polynomials not divisible by the square of a polynomial). One has by "Vallée's identity" (p. 29) $Q(z) = P(z)/P(z^2)$, hence

$$Q(z) = \frac{1 - pz^2}{1 - pz} \qquad \text{and} \qquad Q_n = p^n - p^{n-1} \quad (n \geq 2).$$

Berlekamp's book [**39**] discusses such facts together with relations to error correcting codes. ◁

▷ **53.** *Balanced trees.* The class $\mathcal{E}$ of balanced 2-3 trees contains all the (rooted planar) trees whose internal nodes have degree 2 or 3 and such that all leaves are at the same distance from the root. Only leaves contribute to size. Such tree trees, which are particular cases of B-trees, are a useful data structure for implementing dynamic dictionaries [**269, 381**]. Balanced trees satisfy an implicit equation based on combinatorial substitution:

$$\mathcal{E} = \mathcal{Z} + \mathcal{E}[(\mathcal{Z} \times \mathcal{Z}) + (\mathcal{Z} \times \mathcal{Z} \times \mathcal{Z})] \qquad \Longrightarrow \qquad E(z) = z + E(z^2 + z^3).$$

The expansion starts as (*EIS* **A014535**) $E(z) = z + z^2 + z^3 + z^4 + 2\,z^5 + 2\,z^6 + 3\,z^7 + 4\,z^8 + 5\,z^9 + 8\,z^{10} + \cdots$. Odlyzko [**328**] has determined the growth of E_n to be roughly φ^n/n, where $\varphi = (1 + \sqrt{5})/2$ is the golden ratio. Cf Section IV. 7.2, p. 267 for a partial analysis. ◁

I. 7. Perspective

This chapter and the next amount to a survey of symbolic combinatorics, organized in a coherent manner summarized in Figure 14. We refer to the process of specifying combinatorial classes using these constructions and then automatically having access to the corresponding generating functions as the *symbolic method*. The symbolic method is the "combinatorics" in analytic combinatorics: it allows us to organize classical results in combinatorics with a unifying overall approach, to derive new results that generalize and extend classical problems, and to address new classes of problems that are arising in computer science, computational biology, statistical physics, and other scientific disciplines.

More important, the symbolic method leaves us with generating functions that we can handle with the "analytic" part of analytic combinatorics. A full treatment of this feature of the approach is premature, but a brief discussion may help place the rest of the book in context.

For a given class of problems, the symbolic method typically leads to a unified treatment that reveals a natural class of functions in which generating functions lie. Even though the symbolic method is completely formal, we can often successfully proceed by using classical techniques from complex and asymptotic analysis. For example, denumerants with a finite set of coin denominations always lead to rational generating functions with poles on the unit circle. Such an observation is useful since then a common strategy for coefficient extraction can be applied (partial fraction expansion, in the case of denumerants with coin denominations). In the same vein, the run statistics constitute a particular case of the general theorem of Chomsky and Schützenberger to the effect that the generating function of a regular language

1. The main constructions of disjoint union (combinatorial sum), product, sequence, set, multiset, and cycle and their translation into generating functions (Theorem I.1).

Construction		*OGF*
Union	$\mathcal{A} = \mathcal{B} + \mathcal{C}$	$A(z) = B(z) + C(z)$
Product	$\mathcal{A} = \mathcal{B} \times \mathcal{C}$	$A(z) = B(z) \cdot C(z)$
Sequence	$\mathcal{A} = \text{SEQ}(\mathcal{B})$	$A(z) = \dfrac{1}{1 - B(z)}$
Set	$\mathcal{A} = \text{SET}(\mathcal{B})$	$A(z) = \exp\left(B(z) - \frac{1}{2}B(z^2) + \cdots\right)$
Multiset	$\mathcal{A} = \text{MSET}(\mathcal{B})$	$A(z) = \exp\left(B(z) + \frac{1}{2}B(z^2) + \cdots\right)$
Cycle	$\mathcal{A} = \text{CYC}(\mathcal{B})$	$A(z) = \log \dfrac{1}{1 - B(z)} + \dfrac{1}{2} \log \dfrac{1}{1 - B(z^2)} + \cdots$

2. The translation for sets, multisets, and cycles constrained by the number of components (Theorem I.3, p. 77).

$$
\begin{aligned}
\text{SEQ}_k(\mathcal{B}) &: \quad B(z)^k \\[6pt]
\text{PSET}_2(\mathcal{B}) &: \quad \tfrac{B(z)^2}{2} - \tfrac{B(z^2)}{2} \\
\text{MSET}_2(\mathcal{B}) &: \quad \tfrac{B(z)^2}{2} + \tfrac{B(z^2)}{2} \\
\text{CYC}_2(\mathcal{B}) &: \quad \tfrac{B(z)^2}{2} + \tfrac{B(z^2)}{2} \\[6pt]
\text{PSET}_3(\mathcal{B}) &: \quad \tfrac{B(z)^3}{6} - \tfrac{B(z)\,B(z^2)}{2} + \tfrac{B(z^3)}{3} \\
\text{MSET}_3(\mathcal{B}) &: \quad \tfrac{B(z)^3}{6} + \tfrac{B(z)\,B(z^2)}{2} + \tfrac{B(z^3)}{3} \\
\text{CYC}_3(\mathcal{B}) &: \quad \tfrac{B(z)^3}{3} + \tfrac{2B(z^3)}{3} \\[6pt]
\text{PSET}_4(\mathcal{B}) &: \quad \tfrac{B(z)^4}{24} - \tfrac{B(z)^2 B(z^2)}{4} + \tfrac{B(z)B(z^3)}{3} + \tfrac{B(z^2)^2}{8} - \tfrac{B(z^4)}{4} \\
\text{MSET}_4(\mathcal{B}) &: \quad \tfrac{B(z)^4}{24} + \tfrac{B(z)^2 B(z^2)}{4} + \tfrac{B(z)B(z^3)}{3} + \tfrac{B(z^2)^2}{8} + \tfrac{B(z^4)}{4} \\
\text{CYC}_4(\mathcal{B}) &: \quad \tfrac{B(z)^4}{4} + \tfrac{B(z^2)^2}{4} + \tfrac{B(z^4)}{2}.
\end{aligned}
$$

3. The additional constructions of pointing and substitution (Section I. 6).

Construction		*OGF*
Pointing	$\mathcal{A} = \Theta\mathcal{B}$	$A(z) = z\frac{d}{dz}B(z)$
Substitution	$\mathcal{A} = \mathcal{B} \circ \mathcal{C}$	$A(z) = B(C(z))$

FIGURE 14. A dictionary of constructions applicable to *unlabelled* structures, together with their translation into ordinary generating functions (OGFs). (The labelled counterpart of this table appears in Figure 16 of Chapter II, p. 137.)

is necessarily a rational function. Theorems of this sort establish a bridge between combinatorial analysis and special functions.

Not all applications of the symbolic method are automatic (though that is certainly a goal underlying the approach). The example of counting set partitions shows that application of the symbolic method may require finding an adequate presentation of the combinatorial structures to be counted. In this way, bijective combinatorics enters the game in a nontrivial fashion.

Our introductory examples of compositions and partitions correspond to classes of combinatorial structures with *explicit* "iterative" definitions, a fact leading in turn to explicit generating function expressions. The tree examples then introduce *recursively defined* structures. In that case, the recursive definition translates into a *functional equation* that only determines the generating function implicitly. In simpler situations (like binary or general trees), the equation can be solved and explicit counting results still follow. In other cases (like non-planar trees) one can usually proceed with complex asymptotic analysis directly from the functional equation and obtain very precise *asymptotic estimates*; see Chapters IV–VII.

Analytic combinatorics is characterized by the focus on constructions that leave us with generating functions that yield to classical techniques in complex analysis and asymptotic analysis. For some combinatorial classes, as we shall see, we have theorems that carry us all the way from purely combinatorial constructions through to asymptotic estimates for counting sequences, under general assumptions. For others, the general theorems are yet to be proved, but the symbolic method lays the groundwork for analysis that leads to the results that we seek.

Modern presentations of combinatorial analysis appear in the books of Comtet [**82**] (a beautiful book largely example-driven), Stanley [**391, 393**] (a rich set with an algebraic orientation), and Wilf [**437**] (generating functions oriented). An elementary but insightful presentation of the basic techniques appears in Graham, Knuth, and Patashnik's classic [**212**], a popular book with a highly original design. An encyclopedic reference is the book of Jackson & Goulden [**208**] whose descriptive approach very much parallels ours.

The sources of the modern approaches to combinatorial analysis are hard to trace since they are usually based on earlier traditions and informally stated mechanisms that were well mastered by practicing combinatorial analysts. (See for instance MacMahon's book [**306**] *Combinatory Analysis* first published in 1917, the introduction of denumerant generating functions by Pólya as exposed in [**350**], or the "domino theory" in [**212**, Sec. 7.1].) One source in recent times is the Chomsky–Schützenberger theory of formal languages and enumerations [**80**]. Rota [**365**] and Stanley [**390, 393**] developed an approach which is largely based on partially ordered sets. Bender and Goldman developed a theory of "prefabs" [**32**] whose purposes are similar to the theory developed here. Joyal [**248**] proposed an especially elegant framework, the "theory of species", that addresses foundational issues in combinatorial theory and constitutes the starting point of the superb exposition by Bergeron, Labelle, and Leroux [**37**]. Parallel (but independent) developments by the "Russian School" are nicely synthetized in the books by Sachkov [**369, 370**].

One of the reasons for the revival of interest in combinatorial enumerations and properties of random structures is the analysis of algorithms (a subject founded in modern times by Knuth [**271**]), where the goal is to predict the performance characteristics of computer programs. The symbolic ideas exposed here have been applied to the analysis of algorithms in surveys [**146, 427**] and are further exposed in our book [**382**]. Flajolet, Salvy, and Zimmermann [**173**] have shown how to use them in order to automate the analysis of some well characterized classes of combinatorial structures. Even more recently, research in statistical physics, computational biology, and other scientific disciplines have been drawn towards the study of the sorts of discrete models that can be specified by the sorts of combinatorial constructions that we have described, and therefore are candidates for study via analytic combinatorics. Research in these fields are the driving force in the study of new kinds of constructions on the combinatorics side that lead to new methods on the analytic side.

II

Labelled Structures and Exponential Generating Functions

Cette approche évacue pratiquement tous les calculs[1].
— DOMINIQUE FOATA &
MARCEL P. SCHÜTZENBERGER [**186**]

Contents

Many objects of classical combinatorics present themselves naturally as labelled structures where atoms of an object (typically nodes in a graph or a tree) are distinguishable from one another by the fact that they bear distinct *labels*. Without loss of generality, we may take the set from which labels are drawn to be the set of positive integers. For instance, a permutation can be viewed as a linear arrangement of distinct labels; its cycle decomposition represents it as an unordered collection of circular directed graphs whose nodes are labelled by integers.

Operations on labelled structures are based on a special product: the *labelled product* that distributes labels between components. This operation is a natural analogue of the cartesian product for plain unlabelled objects. The labelled product in turn leads to labelled analogues of the sequence, set, and cycle constructions.

Labelled constructions translate over *exponential generating functions*. The translation schemes turn out to be analytically even simpler than in the unlabelled case considered in the previous chapter. At the same time, labelled constructions enable us to take into account structures that are in many ways combinatorially richer than their unlabelled counterparts, in particular as regards order properties. They constitute another facet, with powerful descriptive powers, of the symbolic method for combinatorial enumeration.

In this chapter, we examine some of the most important classes of labelled objects, including surjections, set partitions, permutations, labelled graphs and labelled trees, as well as graphs and mappings from a finite set into itself. Certain aspects of words

[1] *"This approach eliminates virtually all calculations."*

can also be treated by this theory, a fact which has numerous consequences not only in combinatorics itself but also in probability and statistics. In particular, labelled constructions of words can be put to use in order to solve elegantly two classical problems, the birthday problem and the coupon collector problem, as well as several of their variants that have numerous applications in other fields, including the analysis of hashing algorithms in computer science.

II. 1. Labelled classes

Throughout this chapter, we consider combinatorial classes in the sense of Chapter I: we deal exclusively with finite objects; a combinatorial class $\mathcal{A}$ is a set of objects, with a notion of size attached, so that the number of objects of each size in $\mathcal{A}$ is finite. To these basic concepts, we now add the idea that the objects are *labelled*, by which we mean that each atom carries with it a distinctive colour, or equivalently an integer label, in such a way that all the labels occurring in an object are distinct. Precisely:

DEFINITION II.1. *A* weakly labelled object *of size n is a graph whose set of vertices is a subset of the integers. Equivalently, we say that the vertices bear labels, with the implied condition that labels are distinct integers from $\mathbb{Z}$. An object of size n is said to be* well-labelled*, or simply* labelled*, if it is weakly labelled and, in addition, its collection of labels is the complete integer interval $[1\,..\,n]$. A* labelled class *is a combinatorial class comprised of well-labelled objects.*

The graphs considered may be directed or undirected. In fact, when the need arises, we shall take "object" to mean any kind of discrete structure enriched by integer labels. Virtually all labelled classes considered in this book can eventually be encoded as graphs of sorts, so that this extended use of the notion of a labelled class is a harmless convenience. (See Section II. 7 for a brief discussion of alternative but logically equivalent frameworks for the notion of a labelled class.)

EXAMPLE 1. *Labelled graphs.* A labelled graph is by definition an undirected graph such that distinct integer labels forming an interval of the form $\{1, 2, \ldots, n\}$ are supported by vertices. A particular labelled graph of size 4 is then

$$g = \begin{array}{ccc} 1 & \text{—} & 3 \\ | & & | \\ 4 & \text{—} & 2 \end{array},$$

which represents a graph whose vertices bear the labels $\{1, 2, 3, 4\}$ and whose set of edges is

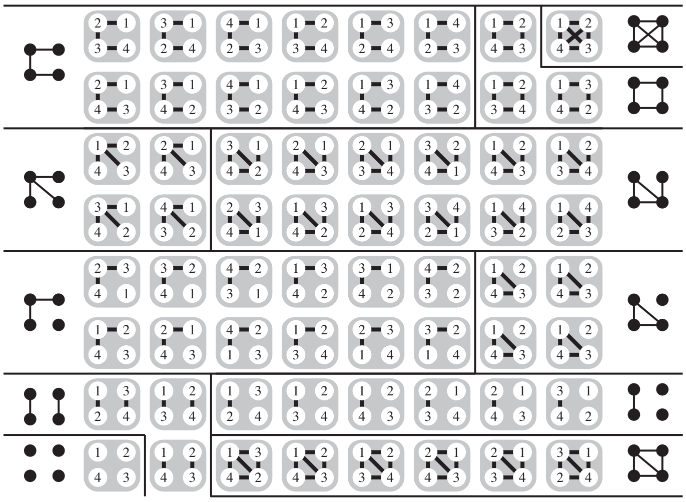
$$\{\,\{1,3\},\ \{2,3\},\ \{2,4\},\ \{1,4\}\,\}\,.$$

Only the graph structure (as defined by its set of edges) counts, so that this is the same abstract graph as in the alternative visual representations

$$g = \begin{array}{ccc} 1 & \text{—} & 4 \\ | & & | \\ 3 & \text{—} & 2 \end{array}, \quad \begin{array}{ccc} 3 & \text{—} & 2 \\ | & & | \\ 1 & \text{—} & 4 \end{array}.$$

However, this graph is different from either of

$$h = \begin{array}{ccc} 4 & \text{—} & 1 \\ | & & | \\ 3 & \text{—} & 2 \end{array}, \qquad j = \begin{array}{ccc} 3 & \text{—} & 1 \\ | & & | \\ 4 & \text{—} & 2 \end{array},$$



There are altogether $G_4 = \mathbf{64} = 2^6$ labelled graphs of size 4, i.e., comprising 4 nodes, in agreement with the general formula (see p. 97 for details): $G_n = 2^{n(n-1)/2}$. The labelled graphs can be grouped into equivalence classes up to arbitrary permutation of the labels, which determines the $\widehat{G}_4 = \mathbf{11}$ unlabelled graphs of size 4. Each unlabelled graph corresponds to a variable number of labelled graphs: for instance, the totally disconnected graph (bottom, left) and the complete graph (top, right) correspond to 1 labelling only, while the line graph admits $\frac{1}{2}\,4! = 12$ possible labellings.

FIGURE 1. Labelled versus unlabelled graphs for size $n = 4$.

since, for instance, 1 and 2 are adjacent in h and j, but not in g. Altogether, there are 3 different labelled graphs (namely, g, h, j), that have the same "shape", corresponding to the unlabelled quadrangle graph

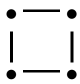
$\mathbf{Q} =$  .

Figure 1 lists all the 64 labelled graphs of size 4 as well as their 11 unlabelled counterparts viewed as equivalence classes of labelled graphs when labels are ignored. END OF EXAMPLE 1. □

In order to count labelled objects, we appeal to exponential generating functions.

DEFINITION II.2. *The* exponential generating function *(EGF) of a sequence* $\{A_n\}$ *is the formal power series*

$$A(z) = \sum_{n\geq 0} A_n \frac{z^n}{n!}. \tag{1}$$

The exponential generating function *(EGF) of a class* $\mathcal{A}$ *is the exponential generating function of the numbers* $A_n = \mathrm{card}(\mathcal{A}_n)$*. Equivalently, the EGF of class* $\mathcal{A}$ *is*

$$A(z) = \sum_{n\geq 0} A_n \frac{z^n}{n!} = \sum_{\alpha\in\mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!}.$$

It is also said that the variable z marks *size in the generating function.*

With the standard notation for coefficients of series, the coefficient A_n in an exponential generating function is then recovered by[2]

$$A_n = n\,! \cdot [z^n]\, A(z),$$

since $[z^n]A(z) = A_n/n!$ by the definition of EGFs and in accordance with the coefficient extractor notation, Eq. (6) of Chapter I.

Note that, like in the previous chapter, we adhere to a systematic naming convention for generating functions of combinatorial structures. A labelled class $\mathcal{A}$, its counting sequence (A_n) (or (a_n)) and its exponential generating function $A(z)$ (or $a(z)$) are all denoted by the same group of letters.

Neutral and atomic classes. Like in the unlabelled universe, it proves useful to introduce a neutral (empty, null) object ϵ that has size 0 and bears no label at all, and consider it as a special labelled object; a *neutral class* $\mathcal{E}$ is then by definition $\mathcal{E} = \{\epsilon\}$. The (labelled) *atomic class* $\mathcal{Z} = \{①\}$ is formed of a unique object of size 1 that, being well-labelled, bears the integer label ①. The EGFs of the neutral class and the atomic class are respectively

$$E(z) = 1, \qquad Z(z) = z.$$

EXAMPLE 2. *Permutations.* The class $\{\mathcal{P}\}$ of all permutations is prototypical of labelled classes. Under the linear representation of permutations, where

$$\sigma = \begin{pmatrix} 1 & 2 & \cdots & n \\ \sigma_1 & \sigma_2 & \cdots & \sigma_n \end{pmatrix}$$

is represented as the sequence $(\sigma_1, \sigma_2, \ldots, \sigma_n)$, the class $\mathcal{P}$ is schematically

$$\mathcal{P} = \left\{ \epsilon\,,\ ①\,,\ \begin{matrix} ①\!-\!② \\ ②\!-\!① \end{matrix}\,,\ \begin{matrix} ①\!-\!②\!-\!③ \\ ②\!-\!③\!-\!① \\ ③\!-\!①\!-\!② \\ ②\!-\!①\!-\!③ \\ ①\!-\!③\!-\!② \\ ③\!-\!②\!-\!① \end{matrix}\,,\ \ldots \right\},$$

so that $P_0 = 1$, $P_1 = 1$, $P_2 = 2$, $P_3 = 6$, etc. There, by definition, all the possible orderings of the distinct labels are taken into account, so that the class $\mathcal{P}$ can be equivalently viewed as the class of all labelled linear digraphs (with an implicit direction, from left to right, say, in the representation). Accordingly, the class $\mathcal{P}$ of permutations has the counting sequence $P_n = n!$ (argument: there are n positions where to place the element 1, then $(n-1)$ possible positions for 2, and so on). Thus the EGF of $\mathcal{P}$ is

$$P(z) = \sum_{n\geq 0} n!\,\frac{z^n}{n!} = \sum_{n\geq 0} z^n = \frac{1}{1-z}.$$

Permutations, as they contain information relative to the order of their elements are essential in many applications related to order statistics. .... END OF EXAMPLE 2. □

[2]Some authors prefer the notation $[\frac{z^n}{n!}]A(z)$ to $n![z^n]A(z)$, which we avoid in this book. Indeed, Knuth [**267**] argues convincingly that the variant notation is not consistent with many desirable properties of a "good" coefficient operator (e.g., bilinearity).

EXAMPLE 3. *Urns.* The class $\mathcal{U}$ of totally disconnected graphs starts as

$$\mathcal{U} = \left\{ \epsilon\,,\ ①\,,\ \boxed{①\ \ ②}\,,\ \boxed{\begin{matrix} ① & ② \\ \multicolumn{2}{c}{③} \end{matrix}}\,,\ \boxed{\begin{matrix} ① & ② \\ ③ & ④ \end{matrix}}\,,\ \boxed{\begin{matrix} ① & & ② \\ & ⑤ & \\ ③ & & ④ \end{matrix}}\,, \dots \right\}.$$

Order between the labelled atoms does *not* count, so that for each n, there is only *one* possible arrangement and $U_n = 1$. The class $\mathcal{U}$ can be regarded as the class of "urns", where an urn of size n contains n distinguishable balls in an unspecified (and irrelevant) order. The corresponding EGF is

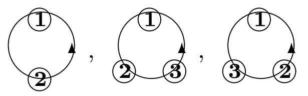
$$U(z) = \sum_{n\geq 0} 1\,\frac{z^n}{n!} = \exp(z) = e^z.$$

(The fact that the EGF of the constant sequence $(1)_{n\geq 0}$ is the exponential function explains the term "exponential generating function".) It also proves convenient, in several applications, to represent elements of an urn in a sorted sequence, which leads to an equivalent representation of urns as *increasing linear graphs*; for instance,

$$①-②-③-④-⑤$$

may be equivalently used to represent the urn of size 5. Though urns look trivial at first glance, they are of particular importance as building blocks of complex labelled structures (e.g., allocations of various sorts), as we shall see shortly. .... END OF EXAMPLE 3. $\square$

EXAMPLE 4. *Circular graphs.* Finally, the class of circular graphs, where cycles are oriented in some conventional manner (say, positively here) is

$\mathcal{C} = \Big\{ ①\,,$  $, \dots \Big\}.$

Cyclic graphs correspond bijectively to *cyclic permutations* . One has $C_n = (n-1)!$ (argument: a directed cycle is determined by the succession of elements that "follow" 1, hence by a permutation of $n-1$ elements). Thus, one has

$$C(z) = \sum_{n\geq 1} (n-1)!\frac{z^n}{n!} = \sum_{n\geq 1} \frac{z^n}{n} = \log\frac{1}{1-z}.$$

As we shall see in the next section, the logarithm is characteristic of circular arrangements of labelled objects. .... END OF EXAMPLE 4. $\square$

$\triangleright$ **1.** *Labelled trees.* Let U_n be now the number of labelled graphs with n vertices that are connected and acyclic; equivalently, U_n is the number of labelled unrooted nonplane trees. Let T_n be the number of labelled rooted nonplane trees. The identity $T_n = nU_n$ is elementary, since all vertices in a labelled tree are distinguishable (by their labels) and a root can be chosen in n possible ways. In Section II. 5, we shall prove that $U_n = n^{n-2}$ and $T_n = n^{n-1}$. $\triangleleft$

II. 2. Admissible labelled constructions

We now describe a toolkit of constructions that make it possible to build complex labelled classes from simpler ones. Combinatorial sum or disjoint union is defined exactly as in Chapter I: it is the union of disjoint copies. To define a product that is adapted to labelled structures, we cannot use the cartesian product, since an ordered pair of two labelled objects is not well-labelled (for instance the label 1 would invariably appear repeated twice). Instead, we define a new operation, the *labelled product*, which translates naturally into exponential generating functions. From there, simple translation rules follow for labelled sequences, sets, and cycles.

Binomial convolutions. As a preparation to the translation of labelled constructions, we first briefly review the effect of products over EGFs. Let $a(z), b(z), c(z)$ be EGFs, with $a(z) = \sum_n a_n z^n/n!$, and so on. The *binomial convolution* formula is:

$$\text{if } a(z) = b(z) \cdot c(z), \text{ then } a_n = \sum_{k=0}^{n} \binom{n}{k} b_k c_{n-k}. \tag{2}$$

This formula results from the usual product of formal power series,

$$\frac{a_n}{n!} = \sum_{k=0}^{n} \frac{b_k}{k!} \cdot \frac{c_{n-k}}{(n-k)!} \quad \text{and} \quad \binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

In the same vein, if $a(z) = a^{(1)}(z)\, a^{(2)}(z) \cdots a^{(r)}(z)$, then

$$a_n = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} a^{(1)}_{n_1} a^{(2)}_{n_2} \cdots a^{(r)}_{n_r}. \tag{3}$$

In Equation (3) there occurs the multinomial coefficient

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!},$$

which counts the number of ways of splitting n elements into r distinguished classes of cardinalities $n_1, \ldots, n_r$. This property lies at the very heart of enumerative applications of binomial convolutions and EGFs.

II. 2.1. Labelled constructions. A labelled object may be relabelled. *We only consider* consistent *relabellings defined by the fact that they preserve the order relations among labels.* Then two dual modes of relabellings prove important:

— *Reduction*: For a weakly labelled structure of size n, this operation reduces its labels to the standard interval $[1\,..\,n]$ while preserving the relative order of labels. For instance, the sequence $\langle 7, 3, 9, 2\rangle$ reduces to $\langle 3, 2, 4, 1\rangle$. We note $\rho(\alpha)$ the canonical reduction of the structure α.

— *Expansion*: This operation is defined relative to a relabelling function $e \in [1\,..\,n] \mapsto \mathbb{Z}$ that is assumed to be strictly increasing. For instance, $\langle 3, 2, 4, 1\rangle$ may expand as $\langle 33, 22, 44, 11\rangle$, $\langle 7, 3, 9, 2\rangle$, and so on. We note $e(\alpha)$ the result of relabelling α by e.

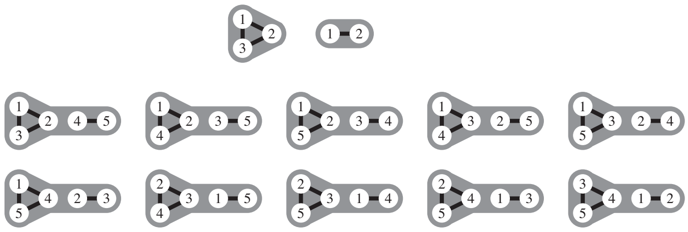


FIGURE 2. The $10 \equiv \binom{5}{2}$ elements in the labelled product of a triangle and a segment.

These notions enable us to devise a product suited to labelled objects.

The *labelled product*, (or simply *product*), of objects and classes was originally formalized under the name of "partitional product" by Foata [**184**]. Given two labelled structures $\beta \in \mathcal{B}$ and $\gamma \in \mathcal{C}$, this product noted as $\beta \star \gamma$ is a set comprised of the collection of well-labelled ordered pairs (β', γ') that reduce to (β, γ):

$$\beta \star \gamma := \{ (\beta', \gamma') \mid (\beta', \gamma') \text{ is well–labelled}, \ \rho(\beta') = \beta, \ \rho(\gamma') = \gamma \}. \tag{4}$$

An equivalent form is via expansion of labels:

$$\beta\star\gamma = \{ (e(\beta), f(\gamma) \mid \mathrm{Im}(e) \cap \mathrm{Im}(f) = \emptyset, \ \mathrm{Im}(e) \cup \mathrm{Im}(f) = [\, 1 \,..\, |\beta| + |\gamma| \,] \}, \tag{5}$$

where e, f are relabelling functions with ranges $\mathrm{Im}(e), \mathrm{Im}(f)$, respectively. Note that elements of a labelled product are, by construction, well-labelled. Figure 2 displays the labelled product of a particular object of size 3 with an other object of size 2.

The labelled product $\beta \star \gamma$ of two elements β, γ of respective sizes n_1, n_2 is a set whose cardinality is, with $n = n_1 + n_2$, expressed as

$$\binom{n_1 + n_2}{n_1, n_2} \equiv \binom{n}{n_1},$$

since this quantity is the number of legal relabellings by expansion of the pair (β, γ). (The example of Figure 2 verifies that the number of relabellings is indeed $\binom{5}{2} = 10$.)

If $\mathcal{B}$ and $\mathcal{C}$ are two labelled classes of combinatorial structures, the labelled product $\mathcal{A} = \mathcal{B} \star \mathcal{C}$ is defined by the usual extension of operations to sets:

$$\mathcal{B} \star \mathcal{C} = \bigcup_{\beta \in \mathcal{B}, \ \gamma \in \mathcal{C}} (\beta \star \gamma). \tag{6}$$

In summary:

DEFINITION II.3. *The* labelled product *of $\mathcal{B}$ and $\mathcal{C}$, denoted $\mathcal{B} \star \mathcal{C}$, is obtained by forming ordered pairs from $\mathcal{B} \times \mathcal{C}$ and performing all possible order-consistent relabellings, ensuring that the resulting pairs are well labelled, as described by (4) or (5), and (6).*

Equipped with this notion, we can build sequences, sets, and cycles, in a way much similar to the unlabelled case. We proceed to do so and, at the same time, establish *admissibility*[3] of the constructions.

Labelled product. When $\mathcal{A} = \mathcal{B} \star \mathcal{C}$, the corresponding counting sequences satisfy the relation,

$$(7) \qquad A_n = \sum_{|\beta|+|\gamma|=n} \binom{|\beta|+|\gamma|}{|\beta|,|\gamma|} B_{|\beta|} C_{|\gamma|} = \sum_{n_1+n_2=n} \binom{n}{n_1,n_2} B_{n_1} C_{n_2}.$$

The product $B_{n_1} C_{n_2}$ keeps track of all the possibilities for the $\mathcal{B}$ and $\mathcal{C}$ components and the binomial coefficient accounts for the number of possible relabellings, in accordance with our earlier discussion. The binomial convolution property (7) then implies admissibility,

$$\mathcal{A} = \mathcal{B} \star \mathcal{C} \qquad \Longrightarrow \qquad A(z) = B(z) \cdot C(z),$$

with the labelled product simply translating into the product operation on EGFs.

$\triangleright$ **2.** *Multiple labelled products.* The (binary) labelled product satisfies the associativity property,

$$\mathcal{B} \star (\mathcal{C} \star \mathcal{D}) \cong (\mathcal{B} \star \mathcal{C}) \star \mathcal{D},$$

which may serve to define $\mathcal{B} \star \mathcal{C} \star \mathcal{D}$. The corresponding EGF is the product $A(z) \cdot B(z) \cdot C(z)$. This product rule generalizes to r factors with coefficients given by a multinomial convolution (3). $\triangleleft$

k-sequences and sequences. The kth (labelled) *power* of $\mathcal{B}$ is defined as $(\mathcal{B} \star \mathcal{B} \cdots \mathcal{B})$, with k factors equal to $\mathcal{B}$. It is denoted $\mathrm{SEQ}_k\{\mathcal{B}\}$ as it corresponds to forming k–sequences and performing all consistent relabellings. The (labelled) *sequence* class of $\mathcal{B}$ is denoted by $\mathrm{SEQ}\{\mathcal{B}\}$ and is defined by

$$\mathrm{SEQ}\{\mathcal{B}\} := \{\epsilon\} + \mathcal{B} + (\mathcal{B} \star \mathcal{B}) + (\mathcal{B} \star \mathcal{B} \star \mathcal{B}) + \cdots = \bigcup_{k \geq 0} \mathrm{SEQ}_k\{\mathcal{B}\}.$$

The product relation for EGFs extends to arbitrary products (Note 2), so that

$$\begin{aligned} \mathcal{A} = \mathrm{SEQ}_k(\mathcal{B}) &\quad\Longrightarrow\quad A(z) = B(z)^k \\ \mathcal{A} = \mathrm{SEQ}(\mathcal{B}) &\quad\Longrightarrow\quad A(z) = \sum_{k=0}^{\infty} B(z)^k = \frac{1}{1-B(z)}, \end{aligned}$$

where the last equation requires $\mathcal{B}_0 = \emptyset$.

[3]We recall that a construction is admissible (Chapter I) if the counting sequence of the result only depends on the counting sequences of the operands. An admissible construction therefore induces a well-defined transformation over exponential generating functions.

k–sets and sets. We denote by $\textsc{Set}_k\{\mathcal{B}\}$ the class of k–sets formed from $\mathcal{B}$. The set class is defined formally, like in the case of the unlabelled multiset: it is the quotient $\textsc{Set}_k\{\mathcal{B}\} := \textsc{Seq}_k\{\mathcal{B}\}/\mathbf{R}$ where the equivalence relation $\mathbf{R}$ identifies two sequences when the components of one are a permutation of the components of the other (p. 25). A "set" is like a sequence, but the order between components is immaterial. The (labelled) *set* construction applied to $\mathcal{B}$, denoted $\textsc{Set}\{\mathcal{B}\}$, is then defined by

$$\textsc{Set}\{\mathcal{B}\} \stackrel{\text{def}}{=} \{\epsilon\} + \mathcal{B} + \textsc{Set}_2\{\mathcal{B}\} + \cdots = \bigcup_{k\geq 0} \textsc{Set}_k\{\mathcal{B}\}.$$

A labelled k–set is associated with exactly $k!$ different sequences. (In the unlabelled case, formulæ are more complex.) Thus in terms of EGFs, one has (assuming $\mathcal{B}_0 = \emptyset$)

$$\begin{array}{lcl} \mathcal{A} = \textsc{Set}_k(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{k!}B(z)^k \\ \mathcal{A} = \textsc{Set}(\mathcal{B}) & \Longrightarrow & A(z) = \displaystyle\sum_{k=0}^{\infty} \frac{1}{k!}B(z)^k = \exp(B(z)). \end{array}$$

Note that the distinction between multisets and powersets that is meaningful for unlabelled structures is here immaterial: by definition components of a labelled set all have distinct labels so that, relative to the labelled universe, we have the correspondence: $\textsc{MSet}, \textsc{PSet} \rightsquigarrow \textsc{Set}$.

k–cycles and cycles. We also introduce the class of k–cycles, $\textsc{Cyc}_k\{\mathcal{B}\}$ and the cycle class. The cycle class is defined formally, like in the unlabelled case, as the quotient $\textsc{Cyc}_k\{\mathcal{B}\} := \textsc{Seq}_k\{\mathcal{B}\}/\mathbf{S}$ where the equivalence relation $\mathbf{S}$ identifies two sequences when the components of one are a cyclic permutation of the components of the other (p. 24). A cycle is like a sequence whose components can be circularly shifted. In terms of EGFs, we have (assuming $\mathcal{B}_0 = \emptyset$)

$$\begin{array}{lcl} \mathcal{A} = \textsc{Cyc}_k(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{k}B(z)^k \\ \mathcal{A} = \textsc{Cyc}(\mathcal{B}) & \Longrightarrow & A(z) = \displaystyle\sum_{k=1}^{\infty} \frac{1}{k}B(z)^k = \log\frac{1}{1-B(z)}, \end{array}$$

since each cycle admits exactly k representations as a sequence.

In summary:

THEOREM II.1. *The constructions of combinatorial sum (disjoint union), labelled product, sequence, cycle and set are all admissible. The associated operators on*

EGFs are:

$$
\begin{array}{llll}
\textit{Sum:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(z) = B(z) + C(z) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \star \mathcal{C} & \Longrightarrow & A(z) = B(z) \cdot C(z) \\
\textit{Sequence:} & \mathcal{A} = \textsc{Seq}(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{1 - B(z)} \\
\textit{— } k \textit{ comp.:} & \mathcal{A} = \textsc{Seq}_k(\mathcal{B}) \equiv (\mathcal{B})^{\star k} & \Longrightarrow & A(z) = B(z)^k \\
\textit{Set:} & \mathcal{A} = \textsc{Set}(\mathcal{B}) & \Longrightarrow & A(z) = \exp(B(z)) \\
\textit{— } k \textit{ comp.:} & \mathcal{A} = \textsc{Set}_k(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{k!} B(z)^k \\
\textit{Cycle:} & \mathcal{A} = \textsc{Cyc}(\mathcal{B}) & \Longrightarrow & A(z) = \log \dfrac{1}{1 - B(z)} \\
\textit{— } k \textit{ comp.:} & \mathcal{A} = \textsc{Cyc}_k(\mathcal{B}) & \Longrightarrow & A(z) = \dfrac{1}{k} B(z)^k
\end{array}
$$

Constructible classes. As in the previous chapter, we say that a class of labelled objects is constructible if it admits a specification in terms of sums (disjoint unions), the labelled constructions of product, sequence, set, cycle, and the initial classes defined by the neutral structure of size 0 and the atomic class $\mathcal{Z} = \{①\}$. Regarding the elementary classes discussed in Section II. 1, it is immediately recognized that

$$\mathcal{P} = \textsc{Seq}\{\mathcal{Z}\}, \quad \mathcal{U} = \textsc{Set}\{\mathcal{Z}\}, \quad \mathcal{C} = \textsc{Cyc}\{\mathcal{Z}\},$$

specify permutations, urns, and circular graphs respectively. These constructions are basic building blocks out of which more complex objects can be constructed. In particular, as we shall explain shortly (Section II. 3 and Section II. 4), set partitions ($\mathcal{S}$), surjections ($\mathcal{R}$), permutations under their cycle decomposition ($\mathcal{P}$), and alignments ($\mathcal{O}$) are constructible classes corresponding to

$$
\begin{array}{lll}
\textit{Surjections:} & \mathcal{R} \simeq \textsc{Seq}\{\textsc{Set}_{\geq 1}\{\mathcal{Z}\}\} & \text{(sequences-of-sets),} \\
\textit{Set partititions:} & \mathcal{S} \simeq \textsc{Set}\{\textsc{Set}_{\geq 1}\{\mathcal{Z}\}\} & \text{(sets-of-sets),} \\
\textit{Alignments:} & \mathcal{O} \simeq \textsc{Seq}\{\textsc{Cyc}_{\geq 1}\{\mathcal{Z}\}\} & \text{(sequences-of-cycles).} \\
\textit{Permutations:} & \mathcal{P} \simeq \textsc{Set}\{\textsc{Cyc}_{\geq 1}\{\mathcal{Z}\}\}, & \text{(sets-of-cycles),}
\end{array}
$$

An immediate consequence of Theorem II.1 is the fact that the EGF of a constructible labelled class can be computed automatically.

THEOREM II.2. *The exponential generating function of a constructible class of labelled objects is a component of a system of generating function equations whose terms are built from* 1 *and* z *using the operators*

$$+\,,\ \times\,,\ Q(f) = \frac{1}{1-f}\,,\ E(f) = e^f\,,\ L(f) = \log\frac{1}{1-f}.$$

If we further allow cardinality restrictions in composite constructions, the operators f^k (for $\textsc{Seq}_k$), $f^k/k!$ (for $\textsc{Set}_k$), and f^k/k (for $\textsc{Cyc}_k$) are to be added to the list.

II. 2.2. Labelled versus unlabelled enumeration. Any labelled class $\mathcal{A}$ has an unlabelled counterpart $\widehat{\mathcal{A}}$: objects in $\widehat{\mathcal{A}}$ are obtained from objects of $\mathcal{A}$ by ignoring the labels. This idea is formalized by identifying two labelled objects if there is an *arbitrary* relabelling (not just an order-consistent one, as has been used so far) that transforms one into the other. For an object of size n, each equivalence class contains a priori between 1 and $n!$ elements. Thus:

PROPOSITION II.1. *The counts of a labelled class $\mathcal{A}$ and its unlabelled counterpart $\widehat{\mathcal{A}}$ are related by*

$$\widehat{A}_n \le A_n \le n!\,\widehat{A}_n \quad \textit{or equivalently} \quad 1 \le \frac{A_n}{\widehat{A}_n} \le n!. \tag{8}$$

EXAMPLE 5. *Labelled and Unlabelled graphs.* This phenomenon has been already encountered in our discussion of graphs (Figure 1). Let generally G_n and $\widehat{G}_n$ be the number of graphs of size n in the labelled and unlabelled case respectively. One finds for $n = 1\,..\,15$

$\widehat{G}_n$ (unlabelled)	G_n (labelled)
1	1
2	2
4	8
11	64
34	1024
156	32768
1044	2097152
12346	268435456
274668	68719476736
12005168	35184372088832
1018997864	36028797018963968
165091172592	73786976294838206464
50502031367952	302231454903657293676544
29054155657235488	2475880078570760549798248448
31426485969804308768	40564819207303340847894502572032

The sequence $\{\widehat{G}_n\}$ constitutes *EIS* **A000088**, which can be obtained by an extension of methods of Chapter I; see [**223**, Ch. 4]. The sequence $\{G_n\}$ is determined directly by the fact that a graph of n vertices can have each of the $\binom{n}{2}$ possible edges either present or not, so that

$$G_n = 2^{\binom{n}{2}} = 2^{n(n-1)/2}.$$

The sequence of labelled counts obviously grows much faster than its unlabelled counterpart. We may then verify the inequality (8) in this particular case. The normalized ratios,

$$\rho_n := G_n/\widehat{G}_n, \quad \sigma_n := G_n/(n!\widehat{G}_n),$$

are observed to be

n	$\rho_n = G_n/\widehat{G}_n$	$\sigma_n = G_n/(n!\widehat{G}_n)$
1	1.000000000	1.0000000000
2	1.000000000	0.5000000000
3	2.000000000	0.3333333333
4	5.818181818	0.2424242424
5	30.11764706	0.2509803922
6	210.0512821	0.2917378918
8	21742.70663	0.5392536367
10	2930768.823	0.8076413203
12	446946830.2	0.9330800361
14	$0.8521603960 \cdot 10^{11}$	0.9774915111
16	$0.2076885783 \cdot 10^{14}$	0.9926428522

From these data, it is natural to conjecture that σ_n tends (fast) to 1 as n tends to infinity. This is indeed a nontrivial fact originally established by Pólya (see Chapter 9 of Harary and Palmer's book [**223**] dedicated to asymptotics of graph enumerations):

$$\widehat{G}_n \sim \frac{1}{n!}\, 2^{\binom{n}{2}} \sim \frac{G_n}{n!}.$$

In other words, "almost all" graphs of size n should admit a number of labellings close to $n!$. (Combinatorially, this corresponds to the fact that in a random unlabelled graph, with high probability, all of the nodes can be distinguished based on the adjacency structure of the graph; in such a case, the graph has no nontrivial automorphism and the number of distinct labellings is $n!$ exactly.) .... END OF EXAMPLE 5. □

The case of urns and totally disconnected graphs resorts to the other extreme situation where

$$\widehat{U}_n = U_n = 1.$$

The examples of graphs and urns illustrate the fact that, beyond the general bounds of Proposition II.1, there is no automatic way to translate between labelled and unlabelled enumerations. At least, if the class $\mathcal{A}$ is constructible, its unlabelled counterpart $\widehat{\mathcal{A}}$ can be obtained by interpreting all the intervening constructions as unlabelled ones in the sense of Chapter I (with $\text{SET} \mapsto \text{MSET}$), both generating functions are computable, and their coefficients can be compared.

▷ **3.** *Permutations and their unlabelled counterparts.* The labelled class of permutations can be specified by $\mathcal{P} = \text{SEQ}(\mathcal{Z})$; the unlabelled counterpart is the set $\widehat{\mathcal{P}}$ of integers in unary notation, and $\widehat{P}_n \equiv 1$, so that $P_n = n! \cdot \widehat{P}_n$ exactly. The specification $\mathcal{P}' = \text{SET}(\text{CYC}(\mathcal{Z}))$ describes sets of cycles and, in the labelled universe, one has $\mathcal{P}' \cong \mathcal{P}$; however the unlabelled counterpart of $\mathcal{P}'$ is the class $\widehat{\mathcal{P}'} \neq \widehat{\mathcal{P}}$ of integer partitions examined in Chapter I. [In the unlabelled universe, there are special combinatorial isomorphisms like: $\text{SEQ}_{\geq 1}(\mathcal{Z}) \cong \text{MSET}_{\geq 1}(\mathcal{Z}) \cong \text{CYC}(\mathcal{Z})$. In the labelled universe, the identity $\text{SET} \circ \text{CYC} \equiv \text{SEQ}$ holds.] ◁

II. 3. Surjections, set partitions, and words

This section and the next are devoted to what could be termed level-two nonrecursive structures defined by the fact that they combine two constructions. In this section, we discuss surjections and set partitions (Section II. 3.1), which constitute labelled analogues of integer compositions and integer partitions in the unlabelled universe. The symbolic method then extends naturally to words over a finite alphabet, where it opens access to an analysis of the frequencies of letters composing words. This in turn has useful consequences for the study of some classical random allocation problems, of which the birthday paradox and the coupon collector problem stand out (Section II. 3.2).

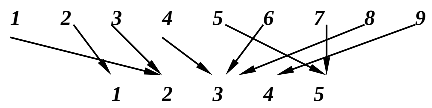
II. 3.1. Surjections and set partitions. We examine classes

$$\mathcal{R} = \textsc{Seq}\{\textsc{Set}_{\geq 1}\{Z\}\} \qquad \text{and} \qquad \mathcal{S} = \textsc{Set}\{\textsc{Set}_{\geq 1}\{Z\}\},$$

corresponding to sequences-of-sets ($\mathcal{R}$) and sets-of-sets ($\mathcal{S}$), or equivalently, sequences of urns and sets of urns, respectively. Such abstract specifications model very classical objects of discrete mathematics, namely surjections ($\mathcal{R}$) and set partitions ($\mathcal{S}$)

Surjections with r images. In elementary mathematics, a surjection from a set A to a set B is a function from A to B that assumes each value *at least once* (an onto mapping). Fix some integer $r \geq 1$ and let $\mathcal{R}_n^{(r)}$ denote the class of all surjections from the set $[1\,..\,n]$ *onto* $[1\,..\,r]$ whose elements are also called r–surjections. Here is a particular object $\phi \in \mathcal{R}_9^{(5)}$:

(9) $\qquad \phi\ :$



(Note that, if $\phi(9)$ were 3, then ϕ would not be a surjection.) We set $\mathcal{R}^{(r)} = \bigcup_n \mathcal{R}_n^{(r)}$ and proceed to compute the corresponding EGF, $R^{(r)}(z)$. First, let us observe that an r–surjection $\phi \in \mathcal{R}_n^{(r)}$ is determined by the *ordered r–tuple* formed with the collection of all preimage sets, $\left(\phi^{-1}(1), \phi^{-1}(2), \ldots, \phi^{-1}(r)\right)$, themselves disjoint nonempty sets of integers that cover the interval $[1\,..\,n]$. In the case of the surjection ϕ of (9), this alternative representation is

$$\phi\ : \qquad (\,\{2\},\ \{1,3\},\ \{4,6,8\},\ \{9\},\ \{5,7\}\,)\,.$$

One has the combinatorial specification and EGF relation:

$$(10) \qquad \mathcal{R}^{(r)} = \textsc{Seq}_r\{\mathcal{V}\},\ \mathcal{V} = \textsc{Set}_{\geq 1}\{\mathcal{Z}\} \qquad \Longrightarrow \qquad R^{(r)}(z) = (e^z - 1)^r.$$

There $\mathcal{V} \equiv \mathcal{U} \setminus \{\epsilon\}$ designates the class of urns ($\mathcal{U}$) that are nonempty, with EGF $V(z) = e^z - 1$, in view of our earlier discussion of urns. In words: "a surjection is a sequence of nonempty sets". See Figure II. 3.1 for an illustration.

Equation (10) does solve the counting problem for surjections. For small r, one finds

$$R^{(2)}(z) = e^{2z} - 2e^z + 1, \qquad R^{(3)}(z) = e^{3z} - 3e^{2z} + 3e^z - 1,$$

whence, by expanding,

$$R_n^{(2)} = 2^n - 2, \qquad R_n^{(3)} = 3^n - 3 \cdot 2^n + 3\,.$$

The general formula follows similarly from expanding the rth power in (10) by the binomial theorem, and then extracting coefficients:

$$(11) \qquad R_n^{(r)} = n!\,[z^n] \sum_{j=0}^{r} \binom{r}{j} (-1)^j e^{(r-j)z} = \sum_{j=0}^{r} \binom{r}{j} (-1)^j (r-j)^n.$$

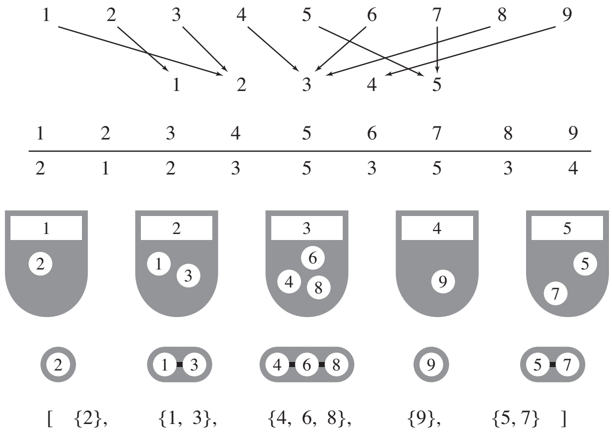


FIGURE 3. The decomposition of surjections as sequences-of-sets: a surjection given by its graph (top), its table (second line), and its sequence of preimages (bottom lines).

$\triangleright$ **4.** *A direct derivation of the surjection EGF.* One can verify the result provided by the symbolic method by returning to first principles. The preimage of value j by a surjection is a nonempty set of some cardinality $n_j \geq 1$, so that

$$R_n^{(r)} = \sum_{(n_1,n_2,\ldots,n_r)} \binom{n}{n_1, n_2, \ldots, n_r}, \tag{12}$$

the sum being taken over $n_j \geq 1,\ n_1 + n_2 + \cdots + n_r = n$. Introduce the numbers $V_n := [\![n \geq 1]\!]$. The formula (12) then assumes the simpler form

$$R_n^{(r)} \equiv \sum_{n_1,n_2,\ldots,n_r} \binom{n}{n_1, n_2, \ldots, n_r} V_{n_1} V_{n_2} \cdots V_{n_r}, \tag{13}$$

where the summation now extends to *all* tuples $(n_1, n_2, \ldots, n_r)$. The EGF of the V_n is $V(z) = \sum V_n z^n/n! = e^z - 1$. Thus the convolution relation (13) leads again to (10). $\triangleleft$

Set partitions into r blocks. Let $\mathcal{S}_n^{(r)}$ denote the number of ways of partitioning the set $[1 \,..\, n]$ into r disjoint and nonempty equivalence classes also known as *blocks*. We set $\mathcal{S}^{(r)} = \bigcup_n \mathcal{S}_n^{(r)}$; the corresponding objects are called *set partitions* (the latter not to be confused with integer partitions examined in Section I. 3). The enumeration problem for set partitions is closely related to that of surjections. Symbolically, a partition is determined as a labelled *set* of classes (blocks), each of which is a nonempty urn. Thus, one has

$$\mathcal{S}^{(r)} = \textsc{Set}_r\{\mathcal{V}\},\ \mathcal{V} = \textsc{Set}_{\geq 1}\{\mathcal{Z}\} \qquad \Longrightarrow \qquad S^{(r)}(z) = \frac{1}{r!}\left(e^z - 1\right)^r. \tag{14}$$

The basic formula connecting the two counting sequences is, in accordance with (10) and (14),

$$S_n^{(r)} = \frac{1}{r!} R_n^{(r)}$$

This can be interpreted directly along the lines of the proof of Theorem II.1: an r–partition is associated with a group of exactly $r!$ distinct r–surjections, two surjections belonging to the same group iff one obtains from the other by permuting the range values, $[1\,..\,r]$.

The numbers $S_n^{(r)} = n![z^n]S^{(r)}(z)$ are known as the Stirling numbers of the second kind, or better, the Stirling "partition" numbers. They were briefly encountered in the previous chapter and discussed in connection with encodings by words (Chapter I, p. 59). Knuth, following Karamata, advocated for the $S_n^{(r)}$ the notation $\left\{ {n \atop r} \right\}$. From (11), an explicit form also exists:

$$S_n^{(r)} \equiv \left\{ {n \atop r} \right\} = \frac{1}{r!} \sum_{j=0}^{r} \binom{r}{j} (-1)^j (r-j)^n. \tag{15}$$

The books by Graham, Knuth, and Patashnik [**212**] and Comtet [**82**] contain a thorough discussion of these numbers; see also APPENDIX A: *Stirling numbers*, p. 652.

All surjections and set partitions. Define now the collection of all surjections and all set partitions by

$$\mathcal{R} = \bigcup_r \mathcal{R}^{(r)}, \qquad \mathcal{S} = \bigcup_r \mathcal{S}^{(r)}.$$

Thus $\mathcal{R}_n$ is the class of all surjections of $[1\,..\,n]$ onto *any* initial segment of the integers, and $\mathcal{S}_n$ is the class of all partitions of the set $[1\,..\,n]$ into *any* number of blocks (Figure 4). Symbolically, one has

$$\begin{array}{lllll} \mathcal{R} = \text{SEQ}(\text{SET}_{\geq 1}\{\mathcal{Z}\}) & \Longrightarrow & R(z) = \dfrac{1}{2-e^z} \\ \mathcal{S} = \text{SET}(\text{SET}_{\geq 1}\{\mathcal{Z}\}) & \Longrightarrow & S(z) = e^{e^z-1}. \end{array} \tag{16}$$

The numbers $R_n = n!\,[z^n]R(z)$ and $S_n = n![z^n]S(z)$ are called *surjection numbers* (also, "preferential arrangements" numbers, *EIS* **A000670**) and *Bell numbers* (*EIS* **A000110**) respectively. These numbers are well determined by expanding the EGFs:

$$\begin{array}{lll} R(z) & = & 1 + z + 3\,\dfrac{z^2}{2!} + 13\,\dfrac{z^3}{3!} + 75\,\dfrac{z^4}{4!} + 541\,\dfrac{z^5}{5!} + 4683\,\dfrac{z^6}{6!} + 47293\,\dfrac{z^7}{7!} + \cdots \\ S(z) & = & 1 + z + 2\,\dfrac{z^2}{2!} + 5\,\dfrac{z^3}{3!} + 15\,\dfrac{z^4}{4!} + 52\,\dfrac{z^5}{5!} + 203\,\dfrac{z^6}{6!} + 877\,\dfrac{z^7}{7!} + \cdots . \end{array}$$

Explicit expressions as finite double sums result from summing Stirling numbers,

$$R_n = \sum_{r \geq 0} r! \left\{ {n \atop r} \right\}, \qquad \text{and} \qquad S_n = \sum_{r \geq 0} \left\{ {n \atop r} \right\},$$

where each Stirling number is itself a sum given by (15). Alternatively, single (though infinite) sums result from the expansions

$$\left\{ \begin{array}{lll} R(z) & = & \dfrac{1}{2}\,\dfrac{1}{1-\frac{1}{2}e^z} \\ & = & \displaystyle\sum_{\ell=0}^{\infty} \frac{1}{2^{\ell+1}} e^{\ell z} \end{array} \right. \qquad \text{and} \qquad \left\{ \begin{array}{lll} S(z) & = & e^{e^z-1} = \dfrac{1}{e} e^{e^z} \\ & = & \dfrac{1}{e} \displaystyle\sum_{\ell=0}^{\infty} \frac{1}{\ell!} e^{\ell z}, \end{array} \right.$$

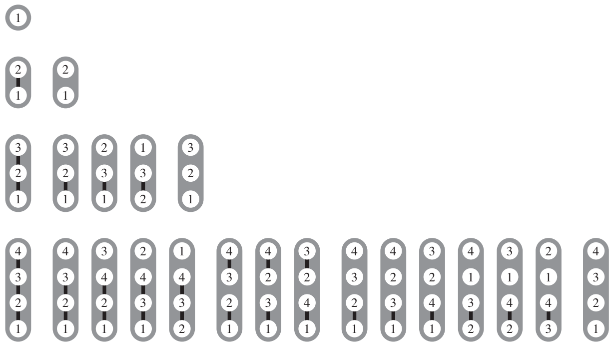


FIGURE 4. A complete listing of all set partitions for sizes $n = 1, 2, 3, 4$. The corresponding sequence $1, 1, 2, 5, 15, \ldots$ is formed of Bell numbers, *EIS* **A000110**.

from which coefficient extraction yields

$$R_n = \frac{1}{2}\sum_{\ell=0}^{\infty}\frac{\ell^n}{2^\ell} \qquad \text{and} \qquad S_n = \frac{1}{e}\sum_{\ell=0}^{\infty}\frac{\ell^n}{\ell!}.$$

The formula for Bell numbers was found by Dobinski in 1877.

The asymptotic analysis of the surjection numbers (R_n) will be performed in Chapter IV as one of the very first illustrations of complex asymptotic methods (the meromorphic case); that of Bell's partition numbers is best done by means of the saddle point method exposed in Chapter IX. The asymptotic forms found are

$$(17) \qquad R_n \sim \frac{n!}{2}\frac{1}{(\log 2)^{n+1}} \qquad \text{and} \qquad S_n \sim n!\,\frac{e^{e^{r(n)}-1}}{r(n)^{n+1}\sqrt{2\pi\exp(r(n))}},$$

where $r(n)$ is the positive root of the equation $re^r = n$. One has $r(n) \sim \log n - \log\log n$, so that

$$\log S_n = n\left(\log n - \log\log n - 1 + o(1)\right).$$

Elementary derivations (i.e., based solely on real analysis) of these asymptotic forms are also possible as discussed briefly in APPENDIX B: *Laplace's method*, p. 667.

The line of reasoning adopted for the enumeration of surjections viewed as sequences-of-sets and partitions viewed as sets-of-sets yields a general result that is applicable to a wide variety of constrained objects.

PROPOSITION II.2. *Let $\mathcal{R}^{(A,B)}$ be the class of surjections where the cardinalities of the preimages lie in $A \subseteq \mathbb{Z}_{\geq 1}$ and the cardinality of the range belongs to B. The corresponding EGF is*

$$R^{(A,B)}(z) = \beta(\alpha(z)) \qquad \textit{where} \qquad \alpha(z) = \sum_{a\in A}\frac{z^a}{a!}, \quad \beta(z) = \sum_{b\in B} z^b.$$

Let $\mathcal{S}^{(A,B)}$ be the class of set partitions with part sizes in $A \subseteq \mathbb{Z}_{\geq 1}$ and with a number of blocks that belongs to B. The corresponding EGF is

$$S^{(A,B)}(z) = \beta(\alpha(z)) \qquad \text{where} \qquad \alpha(z) = \sum_{a \in A} \frac{z^a}{a!}, \quad \beta(z) = \sum_{b \in B} \frac{z^b}{b!}.$$

PROOF. One has

$$\mathcal{R}^{(A,B)} = \text{SEQ}_B\{\text{SET}_A\{\mathcal{Z}\}\} \qquad \text{and} \qquad \mathcal{S}^{(A,B)} = \text{SET}_B\{\text{SET}_A\{\mathcal{Z}\}\},$$

where, as usual, the subscript X specifies a construction with a number of components restricted to the integer set X. □

EXAMPLE 6. *Smallest and largest blocks in set partitions.* Let $e_b(z)$ denote the truncated exponential function,

$$e_b(z) := 1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^b}{b!}.$$

The EGFs $S^{\langle \leq b \rangle}(z) = \exp(e_b(z) - 1)$ and $S^{\langle > b \rangle}(z) = \exp(e^z - e_b(z))$, correspond to partitions with all blocks of size $\leq b$ and all blocks of size $> b$, respectively. END OF EXAMPLE 6. □

▷ **5.** *No singletons.* The EGF of partitions without singleton parts is $e^{e^z - 1 - z}$. The EGF of "double surjections" (each preimage contains at least two elements) is $(2 + z - e^z)^{-1}$. ◁

EXAMPLE 7. *Comtet's square.* An exercise in Comtet's book [**82**, Ex. 13, p. 225] serves beautifully to illustrate the power of the symbolic method. The question is to enumerate set partitions such that a parity constraint is satisfied by the number of blocks and/or the number of elements in each block. Then, the EGFs are tabulated as follows:

Set partitions	Any # of blocks	Odd # of blocks	Even # of blocks
Any block sizes	$e^{e^z - 1}$	$\sinh(e^z - 1)$	$\cosh(e^z - 1)$
Odd block sizes	$e^{\sinh z}$	$\sinh(\sinh z)$	$\cosh(\sinh z)$
Even block sizes	$e^{\cosh z - 1}$	$\sinh(\cosh z - 1)$	$\cosh(\cosh z - 1)$

The proof is a direct application of Proposition II.2, upon noting that

$$e^z, \quad \sinh z, \quad \cosh z$$

are the characteristic EGFs of $\mathbb{Z}_{\geq 0}$, $2\mathbb{Z}_{\geq 0} + 1$, and $2\mathbb{Z}_{\geq 0}$ respectively. The sought EGFs are then obtained by forming the compositions

$$\left\{ \begin{array}{c} \exp \\ \sinh \\ \cosh \end{array} \right\} \circ \left\{ \begin{array}{c} -1 + \exp \\ \sinh \\ -1 + \cosh \end{array} \right\},$$

in accordance with general principles. . . . . END OF EXAMPLE 7. □

II. 3.2. Applications to words and random allocations. Numerous enumerative problems present themselves when analysing statistics on letters in words. They find applications in the study of *random allocations* and the design of *hashing algorithms* of computer science [**382**]. Fix an alphabet

$$\mathcal{X} = \{a_1, a_2, \dots, a_r\}$$

of cardinality r, and let $\mathcal{W}$ be the class of all words over the alphabet $\mathcal{X}$, the size of a word being its length. A word of length n, $w \in \mathcal{W}_n$, is an unconstrained function from $[1 \,..\, n]$ to $[1 \,..\, r]$, the function associating to each position the value of the corresponding letter in the word (canonically numbered from 1 to r). For instance, let $\mathcal{X} = \{a, b, c, d, r\}$ and take the letters of $\mathcal{X}$ canonically numbered as $a_1 = a, \dots, a_5 = r$; for the word $w =$ '*abracadabra*', the table giving the position-to-letter mapping is

$$\left(\begin{array}{ccccccccccc} a & b & r & a & c & a & d & a & b & r & a \\ \hline 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 2 & 5 & 1 & 3 & 1 & 4 & 1 & 2 & 5 & 1 \end{array}\right),$$

which is itself determined by its sequence of preimages:

$$\overbrace{\{1,4,6,8,11\}}^{a=a_1}, \quad \overbrace{\{2,9\}}^{b=a_2}, \quad \overbrace{\{5\}}^{c=a_3}, \quad \overbrace{\{7\}}^{d=a_4}, \quad \overbrace{\{3,10\}}^{r=a_5}.$$

(In this particular case, all preimages are nonempty, but this need not always the case.) The decomposition based on preimages then gives, with $\mathcal{U}$ the class of all urns

(18) $$\mathcal{W} \simeq \mathcal{U}^r \equiv \mathrm{SEQ}_r\{\mathcal{U}\} \qquad \Longrightarrow \qquad W(z) = (e^z)^r = e^{rz},$$

which yields back $W_n = r^n$, as was to be expected. In summary: words over an r-ary alphabet are equivalent to functions into a set of cardinality r and are described by an r-fold labelled product.

For the situation where restrictions are imposed on the number of occurrences of letters, the decomposition (18) generalizes as follows.

PROPOSITION II.3. *Let $\mathcal{W}^{(A)}$ denote the family of words such that the number of occurrences of each letter lies in a set A. Then*

(19) $$W^{(A)}(z) = \alpha(z)^r \qquad \textit{where} \qquad \alpha(z) = \sum_{a \in A} \frac{z^a}{a!}.$$

The proof is a one-liner: $\mathcal{W}^{(A)} \cong \mathrm{SEQ}_r(\mathrm{SET}_A(\mathcal{Z}))$. Though this result is technically a shallow consequence of the symbolic method, it has several important applications in discrete probability; see [**382**, Ch. 8] for a discussion along the lines of the symbolic method.

EXAMPLE 8. *Restricted words.* The EGF of words containing *at most* b times each letter, and that of words containing *more* than b times each letter are

(20) $$\mathcal{W}^{\langle \le b \rangle}(z) = (e_b(z))^r, \qquad \mathcal{W}^{\langle > b \rangle}(z) = (e^z - e_b(z))^r,$$

respectively. (Observe the analogy with Example 6.) Taking $b = 1$ in the first formula gives the number of n–arrangements of r elements (i.e., of ordered combinations of n elements amongst r possibilities),

$$n!\,[z^n](1+z)^r = n!\binom{r}{n} = r(r-1)\cdots(r-n+1), \tag{21}$$

as anticipated; taking $b = 0$, but now in the second formula, gives back the number of r-surjections. For general b, the generating functions of (20) contain valuable information on the least frequent and most frequent letter in random words. . . . . END OF EXAMPLE 8. □

EXAMPLE 9. *Random allocations (balls-in-bins model).* Throw at random n distinguishable balls into m distinguishable bins. A particular realization is described by a word of length n (balls are distinguishable, say, as numbers from 1 to n) over an alphabet of cardinality m (representing the bins chosen). Let Min and Max represent the size of the least filled and most filled bins, respectively. Then[4],

$$\begin{array}{lll} \mathbb{P}\{\text{Max} \le b\} & = & n!\,[z^n] e_b\left(\dfrac{z}{m}\right)^m \\ \mathbb{P}\{\text{Max} > b\} & = & n!\,[z^n]\left(e^{z/m} - e_b\left(\dfrac{z}{m}\right)\right)^m. \end{array} \tag{22}$$

The justification of this formula relies on the easy identity

$$\frac{1}{m^n}\,[z^n]f(z) \equiv [z^n]f(\frac{z}{m}), \tag{23}$$

and on the fact that a probability is determined as the ratio between the number of favourable cases (given by (20)) and the total number of cases (m^n). The formulæ of (22) lend themselves to evaluation using symbolic manipulations systems; for instance, with $m = 100$ and $n = 200$, one finds for $\mathbb{P}(\text{Max} = k)$, where $k = 2, 4, 5, \ldots$, the values:

2	4	5	6	7	8	9	12	15	20
10^{-55}	$1.4 \cdot 10^{-3}$	0.17	0.46	0.26	0.07	0.01	$9.2 \cdot 10^{-5}$	$2.3 \cdot 10^{-7}$	$4.7 \cdot 10^{-10}$

The values $k = 5, 6, 7, 8$ concentrate about 99% of the probability mass.

An especially interesting case is when m and n are asymptotically proportional, that is, $n/m = \alpha$ and α lies in a compact subinterval of $(0, +\infty)$. In that case, with probability tending to 1 as n tends to infinity, one has

$$\text{Min} = 0, \qquad \text{Max} \sim \frac{\log n}{\log\log n}.$$

In other words, there are almost surely empty urns (in fact many of them, see Example 9 in Chapter III) and the most filled urn grows logarithmically in size. Such probabilistic properties are best established by complex analytic methods (especially the saddle point method detailed in Chapter VIII) based on exact generating representations like (20) and (22). They form the core of the reference book [**278**] by Kolchin, Sevastyanov, and Chistyakov. The resulting estimates are in turn invaluable in the analysis of hashing algorithms [**206, 269, 382**] to which the balls-in-bins model has been recognized to apply with great accuracy [**304**]. END OF EXAMPLE 9. □

---

[4]We let $\mathbb{P}(E)$ represent the probability of an event E and $\mathbb{E}(X)$ the expectation of the random variable X; cf APPENDIX C: *Random variables*, p. 685.

$\triangleright$ **6.** *Number of different letters in words.* The probability that a random word of length n over an alphabet of cardinality r contains k different letters is

$$p_{n,k}^{(r)} := \frac{1}{r^n}\binom{r}{k}\left\{ {n \atop k} \right\}k!$$

(Choose k letters amongst r, then split the n positions into k distinguished nonempty classes.) The quantity $p_{n,k}^{(r)}$ is also the probability that a random mapping from $[1\,..\,n]$ to $[1\,..\,r]$ has an image of cardinality k. $\triangleleft$

$\triangleright$ **7.** *Arrangements.* An *arrangement* of size n is an ordered combination of (some) elements of $[1\,..\,n]$. Let $\mathcal{A}$ be the class of all arrangements. Grouping together all the possible elements *not* present in the arrangement into an urn shows that a specification and its companion EGF are

$$\mathcal{A} \simeq \mathcal{U} \star \mathcal{P},\ \mathcal{U} = \textsc{Set}\{\mathcal{Z}\},\ \mathcal{P} = \textsc{Seq}\{\mathcal{Z}\} \qquad \Longrightarrow \qquad A(z) = \frac{e^z}{1-z}.$$

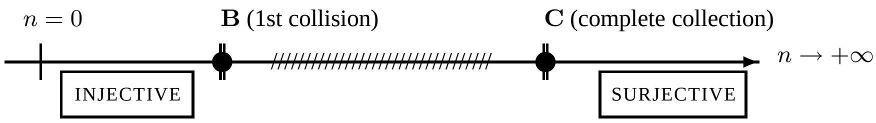
The counting sequence $A_n = \sum_{k=0}^{n} \frac{n!}{k!}$ starts as $1, 2, 5, 16, 65, 326, 1957$ (*EIS* **A000522**); see also Comtet [**82**, p. 75]. $\triangleleft$

$\triangleright$ **8.** *Balls-switching-bins model.* There are m distinguishable balls and two bins (also called "urns") A and B. At any time $t = 1, 2, \ldots$, one of the balls changes bins. The number of moves of length $2n$ that start with urn A full (at $t = 0$) and end with urn A again full (at $t = 2n$) is

$$(2n)! \cdot [z^{2n}]\,(\cosh(z))^m.$$

[Hint: the EGF enumerates mappings where each preimage has an even cardinality.] From there, one can generalize to the case where A contains initially k balls and finally ℓ balls. This is Ehrenfest's simplified model of heat transfer that is analysed thoroughly in [**209**] by combinatorial methods. $\triangleleft$

Birthday paradox and coupon collector problem. The next two examples illustrate applications of EGFs to two classical problems of probability theory, the *birthday paradox* and the *coupon collector problem*. Assume that there is a very long line of persons ready to enter a very large room one by one. Each person is let in and declares her birthday upon entering the room. How many people must enter in order to find two that have the same birthday? The birthday paradox is the counterintuitive fact that on average a birthday collision takes place as early as $n \doteq 24$. Dually, the coupon collector problem asks for the average number of persons that must enter in order to exhaust all the possible days in the year as birthdates. In this case, the answer is the rather large number $n' \doteq 2364$. The term "coupon collection" alludes to the situation where images or coupons of various sorts are inserted in sales items and some premium is given to those who succeed in gathering a complete collection. The birthday problem and the coupon collector problem are relative to a potentially infinite sequence of events; however, the fact that the first birthday collision or the first complete collection occurs at any fixed time n only involves finite events. The following diagram illustrates the events of interest:



In other words, we seek the time at which injectivity *ceases* to hold (the first birthday collision, B) and the time at which surjectivity *begins* to be satisfied (a complete collection, C). In what follows, we consider a year with r days (readers from earth may take $r = 365$) and let $\mathcal{X}$ represent an alphabet with r letters (the days in the year).

EXAMPLE 10. *Birthday paradox.* Let B be the time of the first collision, which is a random variable ranging between 2 and $r+1$ (where the upperbound derives from the pigeonhole principle). A collision has not yet occurred at time n, if the sequence of birthdates $\beta_1, \dots, \beta_n$ has no repetition. In other words, the function β from $[1\,.\,.\,n]$ to $\mathcal{X}$ must be injective; equivalently, $\beta_1, \dots, \beta_n$ is an n-arrangement of r objects. Thus, we have the fundamental relation

$$\begin{aligned} \mathbb{P}\{B > n\} &= \frac{r(r-1)\cdots(r-n+1)}{r^n} \\ &= \frac{n!}{r^n}[z^n](1+z)^r \\ &= n!\,[z^n]\left(1+\frac{z}{r}\right)^r, \end{aligned} \tag{24}$$

where the second line repeats (21) and the third results from the series transformation (23).

The expectation of the random variable B is elementarily

$$\mathbb{E}(B) = \sum_{n=0}^{\infty} \mathbb{P}\{B > n\}, \tag{25}$$

this by virtue of a general formula valid for all discrete random variables (APPENDIX C: *Random variables*, p. 685). From (24), line 1, this gives us a sum expressing the expectation, namely,

$$\mathbb{E}(B) = 1 + \sum_{n=1}^{r} \frac{r(r-1)\cdots(r-n+1)}{r^n}.$$

For instance, with $r = 365$, one finds that the expectation is the rational number,

$$\mathbb{E}(B) = \frac{12681\cdots06674}{51517\cdots40625} \doteq 24.61658,$$

where the denominator comprises as much as 864 digits.

An alternative form of the expectation derives from the generating function involved in (24), line 3. Let $f(z) = \sum_n f_n z^n$ be an entire function with nonnegative coefficients. Then the formula

$$\sum_{n=0}^{\infty} f_n n! = \int_0^{\infty} e^{-t} f(t)\, dt, \tag{26}$$

is valid provided either the sum or the integral on the right converges. The reason is the usual Eulerian representation of factorials,

$$n! = \int_0^{\infty} e^{-t} t^n\, dt.$$

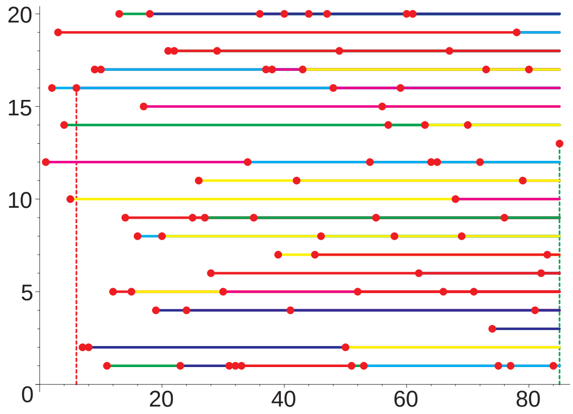


FIGURE 5. A sample realization of the "birthday paradox" and "coupon collection" with an alphabet of $r = 20$ letters. The first collision occurs at time $B = 6$ while the collection becomes complete at time $C = 87$.

Applying this principle to (25) with the probabilities given by (24) (third line), one finds

$$\mathbb{E}(B) = \int_0^\infty e^{-t}\left(1+\frac{t}{r}\right)^r dt. \tag{27}$$

This last form is easily amenable to asymptotic analysis and the Laplace method[5] (see APPENDIX B: *Laplace's method*, p. 667) provides the estimation

$$\mathbb{E}(B) = \sqrt{\frac{\pi r}{2}} + \frac{2}{3} + O(r^{-1/2}), \tag{28}$$

as r tends to infinity. For instance, the asymptotic approximation given by the first two terms of (28) is 24.61*119*, which represents a relative error of only $2 \cdot 10^{-4}$.

The interest of such integral representations based on generating function is that they are *robust*: they adjust naturally to many kinds of combinatorial conditions. For instance, the expected time necessary for the first occurrence of the event "b persons have the same birthday" is found to have expectation given by the integral

$$I(r,b) := \int_0^\infty e^{-t} e_{b-1}\left(\frac{t}{r}\right)^r dt. \tag{29}$$

(The basic birthday paradox corresponds to $b = 2$.) The formula (29) was first derived by Klamkin and Newman in 1967; their paper [**255**] shows in addition that

$$I(r,b) \underset{r\to\infty}{\sim} \sqrt[b]{b!}\,\Gamma\left(1+\frac{1}{b}\right) r^{1-1/b},$$

where the asymptotic form evaluates to 82.87 for $r = 365$ and $b = 3$, while the exact value of the expectation is 88.73891. Thus three-way collisions also tend to occur much sooner than one might think, with about 89 persons on average. Globally, such developments illustrate the versatility of the symbolic approach to many basic probabilistic problems. END OF EXAMPLE 10. □

[5]Knuth [**268**, Sec. 1.2.11.3] uses this calculation as a pilot example for (real) asymptotic analysis; the quantity $\mathbb{E}(B)$ is related to Ramanujan's Q-function (see also Eq. (45) below) by $\mathbb{E}(B) = 1 + Q(r)$.

$\triangleright$ **9.** *The probability distribution of time till a birthday collision.* Elementary approximations show that, for large r, and in the "central" regime $n = t\sqrt{r}$, one has

$$\mathbb{P}(B > t\sqrt{r}) \sim e^{-t^2/2}, \qquad \mathbb{P}(B = t\sqrt{r}) \sim \frac{1}{\sqrt{r}} t e^{-t^2/2}.$$

The continuous probability distribution with density $te^{-t^2/2}$ is called a *Rayleigh distribution*. Saddle point methods (Chapter VIII) may be used to show that for the first occurrence of a b-fold birthday collision: $\mathbb{P}(B > tr^{1-1/b}) \sim e^{-t^b/b!}$. $\triangleleft$

EXAMPLE 11. *Coupon collector problem.* This problem is dual to the birthday paradox. We ask for the first time C when $\beta_1, \ldots, \beta_C$ contains all the elements of $\mathcal{X}$, that is, all the possible birthdates have been "collected". In other words, the event $\{C \le n\}$ means the equality between sets, $\{\beta_1, \ldots, \beta_n\} = \mathcal{X}$. Thus, the probabilities satisfy

$$\begin{aligned} \mathbb{P}\{C \le n\} &= \frac{R_n^{(r)}}{r^n} = \frac{r!\left\{ {n \atop r} \right\}}{r^n} \\ &= \frac{n!}{r^n}[z^n]\,(e^z - 1)^r \\ &= n![z^n]\left(e^{z/r} - 1\right)^r, \end{aligned} \tag{30}$$

by our earlier enumeration of surjections. The complementary probabilities are then

$$\mathbb{P}\{C > n\} = 1 - \mathbb{P}\{C \le n\} = n![z^n]\left(e^z - \left(e^{z/r} - 1\right)^r\right).$$

An application of the Eulerian integral trick of (27) then provides a representation of the expectation of the time needed for a full collection as

$$\mathbb{E}(C) = \int_0^\infty \left(1 - (1 - e^{-t/r})^r\right) dt. \tag{31}$$

A simple calculation (expand by the binomial theorem and integrate termwise) shows that

$$\mathbb{E}(C) = r \sum_{j=1}^{r} \binom{r}{j} \frac{(-1)^{j-1}}{j},$$

which constitutes a first answer to the coupon collector problem in the form of an alternating sum. Alternatively, in (31), perform the change of variables $v = 1 - e^{-t/r}$, then expand and integrate termwise; this process provides the more tractable form

$$\mathbb{E}(C) = r\,\mathrm{H}_r, \tag{32}$$

where H_r is the harmonic number:

$$\mathrm{H}_r = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{r}.$$

Formula (32) is by the way easy to interpret directly[6]: one needs on average $1 = r/r$ trials to get the first day, then $r/(r-1)$ to get a different day, etc.

Regarding (32), one has available the well-known formula (by comparing sums with integrals or by Euler-Maclaurin summation),

$$\mathrm{H}_r = \log r + \gamma + \frac{1}{2r} + O(r^{-2}), \quad \gamma \doteq 0.57721\,56649,$$

[6]Such elementary derivations are very much problem specific: contrary to the symbolic method, they do not usually generalize to more complex situations.

where γ is known as Euler's constant. Thus, the expected time for a full collection satisfies

$$\mathbb{E}(C) = r \log r + \gamma r + \frac{1}{2} + O(r^{-1}). \tag{33}$$

Here the "surprise" lies the nonlinear growth of the expected time for a full collection. For a year on earth, $r = 365$, the exact expected value is $\doteq 2364.64602$ while the approximation provided by the first three terms of (33) yields 2364.64625, representing a relative error of only one in ten millions.

As usual, the symbolic treatment adapts to a variety of situations, for instance, to multiple collections. The expected time till each item (birthday or coupon) is obtained b times (the standard case corresponds to $b = 1$) equals the quantity

$$J(r,b) = \int_0^\infty \left(1 - \left(1 - e_{b-1}(t/r)e^{-t/r}\right)^r\right)\, dt,$$

an expression that vastly generalizes (31). From there, one finds [**325**]

$$J(r,b) = r\left(\log r + (b-1)\log\log r + \gamma - \log(b-1)! + o(1)\right),$$

so that only a few more trials are needed in order to obtain additional collections. END OF EXAMPLE 11. □

▷ **10.** *The little sister.* The coupon collector has a little sister to whom he gives his duplicates. Foata, Lass, and Han [**185**] show that the little sister misses on average H_r coupons when her big brother first obtains a complete collection. ◁

▷ **11.** *The probability distribution of time till a complete collection.* The saddle point method (Chapter VIII) may be used to prove that, in the regime $n = r \log r + tr$:

$$\lim_{t\to\infty} \mathbb{P}(C \le r\log r + tr) = e^{-e^{-t}}.$$

This continuous probability distribution is known a *double exponential distribution*. For the time $C^{(b)}$ till a collection of multiplicity b, one has

$$\lim_{t\to\infty} P(C^{(b)} < r\log r + (b-1)r\log\log r + tr) = \exp(-e^{-t}/(b-1)!),$$

a property known as the Erdős-Rényi law, which finds applications in the study of random graphs [**127**]. ◁

Words as both labelled and unlabelled objects. What distinguishes a labelled structure from an unlabelled one? There is nothing intrinsic there, and everything is in the eye of the beholder—or rather in the type of construction adopted when modelling a specific problem. Take the class of words $\mathcal{W}$ over an alphabet of cardinality r. The two generating functions (an OGF and an EGF respectively),

$$\widehat{W}(z) \equiv \sum_n W_n z^n = \frac{1}{1-rz} \qquad \text{and} \qquad W(z) \equiv \sum_n W_n \frac{z^n}{n!} = e^{rz},$$

leading in both cases to $W_n = r^n$, correspond to two different ways of constructing words: the first one directly as an unlabelled sequence, the other one as a labelled power of letter positions. A similar situation arises for r–partitions, for which we found as OGF and EGF,

$$\widehat{S}^{(r)}(z) = \frac{z^r}{(1-z)(1-2z)\cdots(1-rz)} \qquad \text{and} \qquad S^{(r)}(z) = \frac{(e^z-1)^r}{r!},$$

by viewing these either as unlabelled structures (an encoding via words of a regular language, see Section I.4.3) or directly as labelled structures.

$\triangleright$ **12.** *Balls changing urns: the Ehrenfest[2] urn model.* Consider a system of two urns A and B. There are N distinguishable balls, and, initially, urn A contains them all. At any instant $\frac{1}{2}, \frac{3}{2}, \ldots$, one ball is allowed to change urns. Let $E_n^{[\ell]}$ be the number of possible evolutions that lead to urn A containing ℓ balls at instant n and $E^{[\ell]}(z)$ the corresponding EGF. Then

$$E^{[\ell]}(z) = \binom{N}{\ell}(\cosh z)^{\ell}(\sinh z)^{N-\ell}, \qquad E^{[N]}(z) = 2^{-N}(e^z + e^{-z})^N.$$

In particular the probability that urn A is again full at time $2n$ is $\dfrac{1}{2^N N^{2n}} \displaystyle\sum_{k=0}^{N} \binom{N}{k}(N-2k)^{2n}$.
This famous model was introduced by Paul and Tatiana Ehrenfest [**122**] in 1907, in relation to the apparent contradiction between irreversibility in thermodynamics and recurrence of systems undergoing ergodic transformations. See especially Mark Kac's discussion [**250**]. $\triangleleft$

II. 4. Alignments, permutations, and related structures

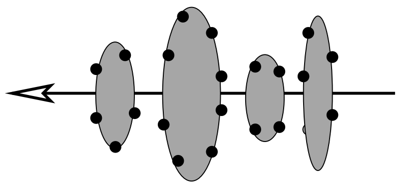
In this section, we start by considering specifications built by piling up two constructions, sequences-of-cycles and sets-of-cycles respectively. They define a new class of objects, alignments, while serving to specify permutations in a novel way as detailed below. (These specifications otherwise parallel surjections and set partitions.) Permutations are in this context examined under their cycle decomposition, the corresponding enumerative results being the most important ones combinatorially (Section II. 4.1). In Section II. 4.2, we recapitulate the meaning of classes that can be defined iteratively by a combination of any two nested labelled constructions.

II. 4.1. Alignments and Permutations. The two specifications under consideration here are

$$\mathcal{O} = \textsc{Seq}\{\textsc{Cyc}\{\mathcal{Z}\}\}, \qquad \text{and} \qquad \mathcal{P} = \textsc{Set}\{\textsc{Cyc}\{\mathcal{Z}\}\}, \tag{34}$$

defining new objects called alignments ($\mathcal{O}$) and an important decomposition of permutations ($\mathcal{P}$).

Alignments. An alignment is a well-labelled sequence of cycles. Let $\mathcal{O}$ be the class of all alignments. Schematically, one can visualize an alignment as a collection of directed cycles arranged in a linear order, somewhat like slices of a sausage fastened on a skewer:

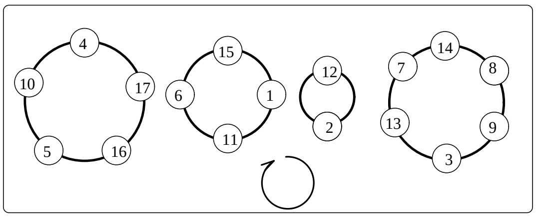


The symbolic method provides,

$$\mathcal{O} = \textsc{Seq}\{\textsc{Cyc}\{\mathcal{Z}\}\} \qquad \Longrightarrow \qquad O(z) = \frac{1}{1 - \log(1-z)^{-1}},$$

and the expansion starts as

$$O(z) = 1 + z + 3\frac{z^2}{2!} + 14\frac{z^3}{3!} + 88\frac{z^4}{4!} + 694\frac{z^5}{5!} + \cdots,$$



A permutation may be viewed as a *set* of cycles that are labelled circular digraphs. The diagram shows the decomposition of the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 \\ 11 & 12 & 13 & 17 & 10 & 15 & 14 & 9 & 3 & 4 & 6 & 2 & 7 & 8 & 1 & 5 & 16 \end{pmatrix}.$$

(Cycles read here clockwise and i is connected to σ_i by an edge in the graph.)

FIGURE 6. The cycle decomposition of permutations.

but the coefficients (*EIS* **A007840**: "ordered factorizations of permutations into cycles") appear to admit of no simple form.

Permutations and cycles. From elementary mathematics, it is known that a permutation admits a unique decomposition into cycles. Let $\sigma = \sigma_1 \ldots \sigma_n$ be a permutation. Start with any element, say 1, and draw a directed edge from 1 to $\sigma(1)$, then continue connecting to $\sigma^2(1), \sigma^3(1)$, and so on; a cycle containing 1 is obtained after at most n steps. If one repeats the construction, taking at each stage an element not yet connected to earlier ones, the cycle decomposition of the permutation σ is obtained. This argument shows that the class of sets-of-cycles (corresponding to $\mathcal{P}$ in (34)) is isomorphic to the class of permutations as defined in Section II. 1:

$$\mathcal{P} = \text{SET}\{\text{CYC}\{\mathcal{Z}\}\} \cong \text{SEQ}\{\mathcal{Z}\}.$$

This combinatorial isomorphism is reflected by the obvious series identity

$$P(z) = \exp\left(\log \frac{1}{1-z}\right) = \frac{1}{1-z}.$$

The property that exp and log are inverse of one another is an analytic reflex of the combinatorial fact that permutations uniquely decompose into cycles!

As regards combinatorial applications, what is especially fruitful is the variety of specializations of the construction of permutations from cycles. We state:

PROPOSITION II.4. *Let $\mathcal{P}^{(A,B)}$ be the class of permutations with cycle lengths in $A \subseteq \mathbb{Z}_{>0}$ and with a number of cycles that belongs to $B \subseteq \mathbb{Z}_{\geq 0}$. The corresponding EGF is*

$$P^{(A,B)}(z) = \beta(\alpha(z)) \qquad \textit{where} \qquad \alpha(z) = \sum_{a\in A} \frac{z^a}{a},\ \beta(z) = \sum_{b \in B} \frac{z^b}{b!}.$$

EXAMPLE 12. *Stirling cycle numbers.* The number of permutations of size n comprised of r cycles is determined by the explicit generating function, to the effect that

$$P_n^{(r)} = \frac{n!}{r!}\,[z^n] \left(\log \frac{1}{1-z}\right)^r. \tag{35}$$

These numbers are fundamental quantities of combinatorial analysis. They are known as the Stirling numbers of the first kind, or better, according to a proposal of Knuth, the *Stirling cycle numbers*. Together with the Stirling partition numbers, the properties of the Stirling cycle numbers are explored in the book by Graham, Knuth, and Patashnik [**212**] where they are denoted by $\left[{n \atop r}\right]$. See APPENDIX A: *Stirling numbers*, p. 652. (Note that the number of alignments formed with r cycles is $r!\left[{n \atop r}\right]$.) As we shall see shortly (p. 131) Stirling numbers also surface in the enumeration of permutations by their number of records.

It is also of interest to determine what happens regarding cycles in a random permutation of size n. Clearly, when the uniform distribution is placed over all elements of $\mathcal{P}_n$, each particular permutation has probability exactly $1/n!$. Since the probability of an event is the quotient of the number of favourable cases over the total number of cases, the quantity

$$p_{n,k} := \frac{1}{n!}\begin{bmatrix} n \\ k \end{bmatrix}$$

is the probability that a random element of P_n has n cycles. This probabilities can be effectively determined for moderate values of n from (35) by means of a computer algebra system. Here are for instance selected values for $n = 100$:

k :	1	2	3	4	5	6	7	8	9	10
$p_{n,k}$:	0.01	0.05	0.12	0.19	0.21	0.17	0.11	0.06	0.03	0.01

For this value $n = 100$, we expect in a vast majority of cases the number of cycles to be in the interval $[1, 10]$. (The residual probability is only about 0.005.) Under this probabilistic model, the mean is found to be about 5.18. Thus: *A random permutation of size 100 has on average a little more than 5 cycles; it rarely has more than 10 cycles.*

Such procedures demonstrate a direct exploitation of symbolic methods. They do not however tell us how the number of cycles could depend on n as n varies. Such questions are to be examined systematically in Chapter III. Here, we shall content ourselves with a brief sketch. First, form the bivariate generating function,

$$P(z,u) := \sum_{r=0}^{\infty} P^{(r)}(z)u^r,$$

and observe that

$$\begin{aligned} P(z,u) &= \sum_{r=0}^{\infty} \frac{u^r}{r!}\left(\log\frac{1}{1-z}\right)^r = \exp\left(u\log\frac{1}{1-z}\right) \\ &= (1-z)^{-u}. \end{aligned}$$

Newton's binomial theorem then provides

$$[z^n](1-z)^{-u} = (-1)^n\binom{-u}{n}.$$

In other words, a simple formula

$$\sum_{k=0}^{n}\begin{bmatrix} n \\ k \end{bmatrix}u^k = u(u+1)(u+2)\cdots(u+n-1) \tag{36}$$

encodes precisely all the Stirling cycle numbers corresponding to a fixed value of n. From there, the expected number of cycles, $\mu_n := \sum_k kp_{n,k}$ is easily found (use logarithmic differentiation of (36)),

$$\mu_n = \mathrm{H}_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

In particular, one has $\mu_{100} \equiv H_{100} \doteq 5.18738$. In general: *The mean number of cycles in a random permutation of size n grows logarithmically with n, $\mu_n \sim \log n$.* END OF EXAMPLE 12. □

EXAMPLE 13. *Involutions and permutations without long cycles.* A permutation σ is an *involution* if $\sigma^2 = Id$ with Id the identity permutation. Clearly, an involution can have only cycles of sizes 1 and 2. The class $\mathcal{I}$ of all involutions thus satisfies

$$\mathcal{I} = \text{SET}\{\text{CYC}_{1,2}\{\mathcal{Z}\}\} \qquad \Longrightarrow \qquad I(z) = \exp\left(z + \frac{z^2}{2}\right). \tag{37}$$

The explicit form of the EGF lends itself to expansion,

$$I_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{(n-2k)! 2^k k!},$$

which solves the counting problem explicitly. A *pairing* is an involution without fixed point. In other words, only cycles of length 2 are allowed, so that

$$\mathcal{J} = \text{SET}(\text{CYC}_2(\mathcal{Z})) \qquad \Longrightarrow \qquad J(z) = e^{z^2/2}, \quad J_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1).$$

(The formula for J_n, hence that of I_n, can be checked by a direct reasoning.)

Generally, the EGF of permutations, all of whose cycles (in particular the largest one) have length at most equal to r satisfies

$$B^{(r)}(z) = \exp\left(\sum_{j=1}^{r} \frac{z^j}{j}\right).$$

The numbers $b_n^{(r)} = [z^n]B^{(r)}(z)$ satisfy the recurrence

$$(n+1)b_{n+1}^{(r)} = (n+1)b_n^{(r)} - b_{n-r}^{(r)},$$

by which they can be computed fast. This gives access to the statistics of the longest cycle in a permutation. .... END OF EXAMPLE 13. □

EXAMPLE 14. *Derangements and permutations without short cycles.* Classically, a derangement is defined as a permutation without fixed points, i.e., $\sigma_i \neq i$ for all i. Given an integer r, an r–derangement is a permutation all of whose cycles (in particular the shortest one) have length larger than r. Let $\mathcal{D}^{(r)}$ be the class of all r–derangements. A specification is

$$\mathcal{D}^{(r)} = \text{SET}\{\text{CYC}_{>r}\{\mathcal{Z}\}\}, \tag{38}$$

the corresponding EGF being then

$$D^{(r)}(z) = \exp\left(\sum_{j>r} \frac{z^j}{j}\right) = \frac{\exp(-\sum_{j=1}^{r} \frac{z^j}{j})}{1-z}. \tag{39}$$

For instance, when $r = 1$, a direct expansion yields

$$\frac{D_n^{(1)}}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!},$$

a truncation of the series expansion of $\exp(-1)$ that converges fast to e^{-1}. Phrased differently, the enumeration of derangements is a famous combinatorial problem with a pleasantly quaint nineteenth century formulation [**82**]: "A number n of people go to opera, leave their hats on hook in the cloakroom and grab them at random when leaving; the probability that nobody gets

All perms $\dfrac{1}{1-z}$ Derangements $\dfrac{e^{-z}}{1-z}$	Involutions $e^{z+z^2/2}$ Pairings $e^{z^2/2}$
Shortest cycle $> r$ $\dfrac{\exp\left(-\frac{z}{1}-\frac{z^2}{2}-\cdots-\frac{z^r}{r}\right)}{1-z}$	Longest cycle $\leq r$ $\exp\left(\frac{z}{1}+\frac{z^2}{2}+\cdots+\frac{z^r}{r}\right)$

FIGURE 7. A summary of major EGFs related to permutations.

back his own hat is asymptotic to $1/e$, which is nearly 37%". (The usual proof uses an inclusion-exclusion argument. Also, it is a sign of changing times that Motwani and Raghavan [**323**, p. 11] describe the problem as one of sailors that return to their ship in state of inebriation and choose random cabins to sleep in.) For the generalized derangement problem, there holds

$$\frac{D_n^{(r)}}{n!} \sim e^{-\mathrm{H}_r}, \tag{40}$$

(for any fixed r), as can be proved easily by complex asymptotic methods (Chapter IV). END OF EXAMPLE 14. □

Like several other structures that we have been considering previously, permutation allow for transparent connections between structural constraints and the forms of generating functions. The major counting results encountered in this section are summarized in Figure 7.

▷ **13.** *Permutations such that $\sigma^f = Id$.* Such permutations are "roots of unity" in the symmetric group. Their EGF is

$$\exp\left(\sum_{d \mid f} \frac{z^d}{d}\right),$$

where the sum extends to all divisors d of f. ◁

▷ **14.** *Parity constraints in permutations.* The EGFs of permutations having only even size cycles ($E(z)$) or odd size cycles ($O(z)$) are

$$E(z) = \exp\left(\frac{1}{2}\log\frac{1}{1-z^2}\right) = \frac{1}{\sqrt{1-z^2}}, \qquad O(z) = \exp\left(\frac{1}{2}\log\frac{1+z}{1-z}\right) = \sqrt{\frac{1+z}{1-z}}.$$

From the EGFs, one finds $E_{2n} = (1 \cdot 3 \cdot 5 \cdots (2n-1))^2$, $O_{2n} = E_{2n}$, $O_{2n+1} = (2n+1)E_{2n}$.

The EGFs of permutations having an even number of cycles ($E^*(z)$) and an odd number of cycles ($O^*(z)$) are

$$E^*(z) = \cosh(\log\frac{1}{1-z}) = \frac{1}{2}\frac{1}{1-z}+\frac{1}{2}-\frac{z}{2}, \ O^*(z) = \sinh(\log\frac{1}{1-z}) = \frac{1}{2}\frac{1}{1-z}-\frac{1}{2}+\frac{z}{2},$$

so that parity of the number of cycles is evenly distributed amongst permutations of size n as soon as $n \geq 2$. The generating functions obtained in this way are analogous to the ones appearing in the discussion of "Comtet's square" in the previous section. ◁

II. 4.2. Second level structures. Consider the three basic constructors of labelled sequence (SEQ), set (SET), and cycle (CYC). We can play the formal game of examining what the various combinations produce as combinatorial objects. Restricting attention to superpositions of two constructors (an external one applied to an internal one) gives nine possibilities summarized by the following table:

ext.\int.	$\text{SEQ}_{\geq 1}$	$\text{SET}_{\geq 1}$	CYC
SEQ	*Labelled compositions* ($\mathcal{L}$) SEQ ∘ SEQ $\dfrac{1-z}{1-2z}$	Surjections ($\mathcal{R}$) SEQ ∘ SET $\dfrac{1}{2-e^z}$	Alignments ($\mathcal{O}$) SEQ ∘ CYC $\dfrac{1}{1-\log(1-z)^{-1}}$
SET	*Fragmented permutations* ($\mathcal{F}$) SET ∘ SEQ $e^{z/(1-z)}$	Set partitions ($\mathcal{S}$) SET ∘ SET e^{e^z-1}	Permutations ($\mathcal{P}$) SET ∘ CYC $\dfrac{1}{1-z}$
CYC	*Supernecklaces* ($\mathcal{S}^I$) CYC ∘ SEQ $\log\dfrac{1-z}{1-2z}$	*Supernecklaces* ($\mathcal{S}^{II}$) CYC ∘ SET $\log(2-e^z)^{-1}$	*Supernecklaces* ($\mathcal{S}^{III}$) CYC ∘ CYC $\log\dfrac{1}{1-\log(1-z)^{-1}}$

The classes of surjections, alignments, set partitions, and permutations appear naturally as SEQ ∘ SET, SEQ ∘ CYC, SET ∘ SET, and SET ∘ CYC (top right corner). The other ones represent essentially nonclassical objects. The case of $\mathcal{L}$ corresponding to SEQ ∘ SEQ describes objects that are (ordered) sequences of linear graphs; this can be interpreted as permutations with separators inserted, e.g, $53|264|1$, or alternatively as integer compositions with a labelling superimposed, so that $L_n = n!\,2^{n-1}$. The class $\mathcal{F} = \text{SET}\{\text{SEQ}_{\geq 1}\{\mathcal{Z}\}\}$ corresponds to unordered collections of permutations; in other words, "fragments" are obtained by breaking a permutation into pieces (pieces must be nonempty for definiteness). The interesting EGF is

$$F(z) = e^{z/(1-z)} = 1 + z + 3\frac{z^2}{2!} + 13\frac{z^3}{3} + 73\frac{z^4}{4!} + \cdots,$$

(*EIS* **A000262**: "sets of lists"). The corresponding asymptotic analysis serves to illustrate an important aspect of the saddle point method in Chapter VIII. What we termed "supernecklaces" in the last row represents cyclic arrangements of composite objects existing in three brands.

All sorts of refinements, of which Figure 7 may give an idea, are clearly possible. We leave to the reader's imagination the task of determining which amongst the level 3 structures may be of combinatorial interest. . .

▷ **15.** *A meta-exercise: Counting specifications of level n.* The algebra of constructions satisfies the combinatorial isomorphism $\text{SET}\{\text{CYC}\{\mathcal{X}\}\} \cong \text{SEQ}\{X\}$ for all $\mathcal{X}$. How many different terms involving n constructions can be built from three symbols CYC, SET, SEQ satisfying a semi-group law ('∘') together with the relation $\text{SET} \circ \text{CYC} = \text{SEQ}$? This determines the number of specifications of level n. [Hint: the OGF is rational as normal forms correspond to words with an excluded pattern.] ◁

II. 5. Labelled trees, mappings, and graphs

In this section, we consider labelled trees as well as other important structures that are naturally associated with them, namely mappings and functional graphs on one side, graphs of small excess on the other side. Like in the unlabelled case considered in Section I. 6, the corresponding combinatorial classes are inherently recursive, the case of trees being typical since a tree is obtained by appending a root to a collection (set, sequence) of subtrees. From there, it is possible to build the graphs associated to mappings from a finite set to itself, as these decompose as sets of connected components that are cycles of trees. Variations of these construction finally open access to the enumeration of graphs having a fixed excess of the number of edges over the number of vertices.

II. 5.1. Trees. The trees to be studied here are invariably labelled, so that nodes bear distinct integer labels. Unless otherwise specified, they are rooted, meaning as usual that one node is distinguished as the root. Labelled trees, like their unlabelled counterparts, exist in two varieties: (i) plane trees where an embedding in the plane is understood (or, equivalently, subtrees dangling from a node are ordered, say, from left to right); (ii) nonplane trees where no such embedding is imposed (such trees are then nothing but connected directed acyclic graphs with a distinguished root). Trees may be further restricted by the additional constraint that the node outdegrees should belong to a fixed set $\Omega \subseteq \mathbb{Z}_{\geq 0}$ where $\Omega \ni 0$.

Plane labelled trees. We first dispose of the plane variety of labelled trees. Let $\mathcal{A}$ be the set of (rooted labelled) plane trees constrained by Ω. This family is specified by

$$\mathcal{A} = \mathcal{Z} \star \text{SEQ}_{\Omega}\{\mathcal{A}\},$$

where $\mathcal{Z}$ represents the atomic class consisting of a single labelled node: $\mathcal{Z} = \{1\}$. The sequence construction appearing here reflects the planar embedding of trees, as

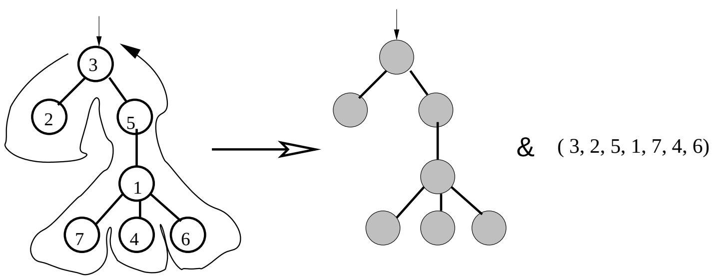


FIGURE 8. A labelled plane tree is determined by an unlabelled tree (the "shape") and a permutation of the labels $1, \ldots, n$.

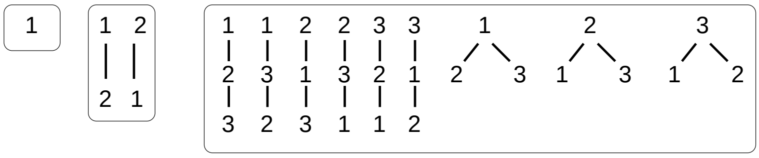


FIGURE 9. There are $T_1 = 1, T_2 = 2, T_3 = 9$, and in general $T_n = n^{n-1}$ Cayley trees of size n.

subtrees stemming from a common root are ordered between themselves. Accordingly, the EGF $A(z)$ satisfies

$$A(z) = z\phi(A(z)) \qquad \text{where} \quad \phi(u) = \sum_{\omega \in \Omega} u^{\omega}.$$

This is exactly the same equation as the one satisfied by the *ordinary* GF of Ω-restricted *unlabelled* plane trees (see Proposition I.5). Thus, $\frac{1}{n!}A_n$ is the number of unlabelled trees. In other words: *in the plane rooted case, the number of labelled trees equals $n!$ times the corresponding number of unlabelled trees.* As illustrated by Figure 8, this is easily understood combinatorially: each labelled tree can be defined by its "shape" that is an unlabelled tree and by the sequence of node labels where nodes are traversed in some fixed order (preorder, say). Finally, one has, by Lagrange inversion,

$$A_n = n![z^n]A(z) = (n-1)![u^{n-1}]\phi(u)^n.$$

This simple analytic–combinatorial relation enables us to transpose all of the enumerative results of Section I.5.1 to plane labelled trees (upon multiplying the evaluations by $n!$, of course). In particular, the total number of "general" plane labelled trees (with no degree restriction imposed, i.e., $\Omega = \mathbb{Z}_{\geq 0}$) is

$$n! \times \frac{1}{n}\binom{2n-2}{n-1} = \frac{(2n-2)!}{(n-1)!} = 2^{n-1}\left(1 \cdot 3 \cdots (2n-3)\right).$$

The corresponding sequence starts as $1, 2, 12, 120, 1680$ and is *EIS* **A001813**.

Nonplane labelled trees. We next turn to labelled nonplane trees (Figure 9) to which the rest of this section will be devoted. The class $\mathcal{T}$ of all such trees is definable by a symbolic equation, which provides an implicit equation satisfies by the EGF:

$$\mathcal{T} = \mathcal{Z} \star \text{SET}\{\mathcal{T}\} \qquad \Longrightarrow \qquad T(z) = ze^{T(z)}. \tag{41}$$

There the set construction translates the fact that subtrees stemming from the root are not ordered between themselves. From the specification (41), the EGF $T(z)$ is defined implicitly by the "functional equation"

$$T(z) = ze^{T(z)}. \tag{42}$$

The first few values are easily found, for instance by the method of indeterminate coefficients,

$$T(z) = z + 2\frac{z^2}{2!} + 9\frac{z^3}{3!} + 64\frac{z^4}{4!} + 625\frac{z^5}{5!} + \cdots.$$

As suggested by the first few coefficients($9 = 3^2, 64 = 4^3, 625 = 5^4$), the general formula is

$$T_n = n^{n-1} \tag{43}$$

which is established (like in the case of plane unlabelled trees, Chapter I) by the Lagrange Inversion Theorem (see APPENDIX A: *Lagrange Inversion*, p. 649).

The enumerative result $T_n = n^{n-1}$ is a famous one, attributed to the prolific British mathematician Arthur Cayley (1821–1895) who had keen interest in combinatorial mathematics and published altogether over 900 papers and notes. Consequently, formula (43) given by Cayley in 1889 is often referred to as "Cayley's formula" and unrestricted nonplane labelled trees are often called "Cayley trees". See [**52**, p. 51] for a historical discussion. The function $T(z)$ is also known as the (Cayley) "tree function"; it is a close relative of the W–function [**83**] defined implicitly by $We^W = z$, which was introduced by the Swiss mathematician Johann Lambert (1728–1777) otherwise famous for first proving the irrationality of the number π.

A similar process gives the number of (nonplane rooted) trees where all (out)degrees of nodes are restricted to lie in a set Ω. This corresponds to the specification:

$$\mathcal{T}^{(\Omega)} = \mathcal{Z} \star \text{SET}_\Omega\{\mathcal{T}^{(\Omega)}\} \quad \Longrightarrow \quad T^{(\Omega)}(z) = z\overline{\phi}(T^{(\Omega)}(z)) \text{ where } \overline{\phi}(u) = \sum_{\omega \in \Omega} \frac{u^\omega}{\omega!}.$$

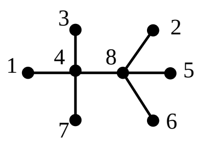
What the last formula involves is the "exponential characteristic" of the degree sequence (as opposed to the ordinary characteristic, in the planar case). It is once more amenable to Lagrange inversion. In summary:

PROPOSITION II.5. *The number of rooted nonplane trees, where all nodes have their outdegree in Ω, is*

$$T_n^{(\Omega)} = (n-1)![u^{n-1}](\overline{\phi}(u))^n \qquad \textit{where} \quad \overline{\phi}(u) = \sum_{\omega \in \Omega} \frac{u^\omega}{\omega!}.$$

In particular, when all node degrees are allowed ($\Omega \equiv \mathbb{Z}_{\geq 0}$), the number of trees is $T_n = n^{n-1}$ and its EGF is the Cayley tree function satisfying $T(z) = ze^{T(z)}$.

$\triangleright$ **16.** *Prüfer's bijective proofs of Cayley's formula.* The simplicity of Cayley's formula calls for a combinatorial explanation. The most famous one is due to Prüfer (in 1918). It establishes as follows a bijective correspondence between unrooted Cayley trees whose number is n^{n-2} for size n and sequences $(a_1, \ldots, a_{n-2})$ with $1 \leq a_j \leq n$ for each j. Given an unrooted tree τ, remove the endnode (and its incident edge) with the smallest label; let a_1 denote the label of the node that was joined to the removed node. Continue with the pruned tree τ' to get a_2 in a similar way. Repeat the construction of the sequence until the tree obtained only consists of a single edge. For instance:



$\longrightarrow \quad (4, 8, 4, 8, 8, 4).$

It can be checked that the correspondence is bijective; see [**52**, p. 53] or [**318**, p. 5]. $\triangleleft$

▷ **17.** *Forests.* The number of unordered k–forests (i.e., k–sets of trees) is

$$F_n^{(k)} = n![z^n]\frac{(T(z))^k}{k!} = \frac{(n-1)!}{(k-1)!}[u^{n-k}](e^u)^n = \binom{n-1}{k-1}n^{n-k},$$

as follows from Bürmann's form of Lagrange inversion. ◁

▷ **18.** *Labelled hierarchies.* The class $\mathcal{L}$ of labelled hierarchies is formed of trees whose internal nodes are unlabelled and are constrained to have outdegree larger than 1, while leaves have labels attached to them. Like for other labelled structure, size is the number of labels (so that internal nodes do not contribute). Hierarchies satisfy the specification

$$\mathcal{L} = \mathcal{Z} + \textsc{Set}_{\geq 2}\{\mathcal{L}\}, \qquad \Longrightarrow \qquad L = z + e^L - 1 - L.$$

This happens to be solvable in terms of the Cayley function: $L(z) = T(\frac{1}{2}e^{z/2-1/2}) + \frac{z}{2} - \frac{1}{2}$. The first few values are $0, 1, 4, 26, 236$ (*EIS* **A000311**): these numbers count phylogenetic trees (used to describe the evolution of a genetically related group of organisms) and correspond to Schröder's "fourth problem"; see [**82**, p. 224] and Section I.5.2 for unlabelled analogues.

The class of binary (labelled) hierarchies defined by the additional fact that internal nodes can have degree 2 only is expressed by

$$\mathcal{M} = \mathcal{Z} + \textsc{Set}_2\{\mathcal{M}\} \qquad \Longrightarrow \qquad M(z) = 1 - \sqrt{1-2z} \quad \text{and} \quad M_n = 1 \cdot 3 \cdots (2n-3),$$

where the counting numbers are now the odd factorials. ◁

II. 5.2. Mappings and functional graphs. Let $\mathcal{F}$ be the class of mappings (or "functions") from $[1\,..\,n]$ to itself. A mapping $f \in [1\,..\,n] \mapsto [1\,..\,n]$ can be represented by a directed graph over the set of vertices $[1\,..\,n]$ with an edge connecting x to $f(x)$, for all $x \in [1\,..\,n]$. The graphs so obtained are called *functional graphs* and they have the characteristic property that the outdegree of each vertex is exactly equal to 1.

Mappings and functional graphs. Given a mapping (or function) f, upon starting from any point x_0, the succession of (directed) edges in the graph traverses the vertices corresponding to iterated values of the mapping,

$$x_0, \quad f(x_0), \quad f(f(x_0)), \ldots .$$

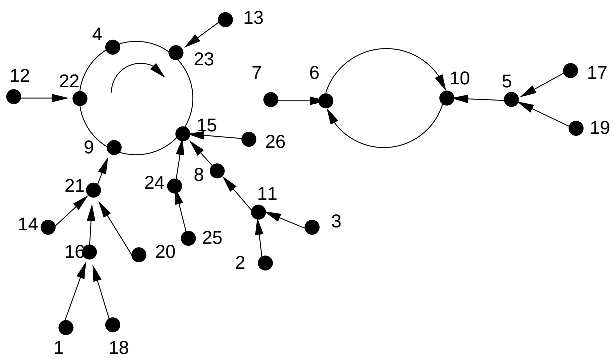


FIGURE 10. A functional graph of size $n = 26$ associated to the mapping φ such that $\varphi(1) = 16$, $\varphi(2) = \varphi(3) = 11$, $\varphi(4) = 23$, and so on.

Since the domain is finite, each such sequence must eventually loop on itself. When the operation is repeated starting each time from an element not previously hit, the vertices group themselves into components. This leads to another characterization of functional graphs (Figure 10): *A functional graph is a set of connected functional graphs. A connected functional graph is a collection of rooted trees arranged in a cycle.*

Thus, with $\mathcal{T}$ being as before the class of all Cayley trees, and with $\mathcal{K}$ the class of all connected functional graphs, we have the specification:

$$(44)\qquad \begin{cases} \mathcal{F} &=& \text{SET}\{\mathcal{K}\} \\ \mathcal{K} &=& \text{CYC}\{\mathcal{T}\} \\ \mathcal{T} &=& \mathcal{Z} \star \text{SET}\{\mathcal{T}\} \end{cases} \qquad \Longrightarrow \qquad \begin{cases} F(z) &=& e^{K(z)} \\ K(z) &=& \log \dfrac{1}{1-T(z)} \\ T(z) &=& ze^{T(z)}. \end{cases}$$

What is especially interesting here is a specification binding three types of related structures. From Equation (44), the EGF $F(z)$ is found to satisfy $F = (1-T)^{-1}$. It can be checked from there, by Lagrange inversion once again, that we have

$$F_n = n^n,$$

as was to be expected (!) from the origin of the problem. More interestingly, Lagrange inversion also provides for the number of connected functional graphs (expand $\log(1-T)^{-1}$ and recover coefficients by Bürmann's form):

$$(45)\quad K_n = n^{n-1}Q(n) \quad \text{where} \quad Q(n) := 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots.$$

The quantity $Q(n)$ that appears in (45) is a famous one that surfaces in many problems of discrete mathematics (including the birthday paradox, Equation (27)). Knuth has proposed to call it "Ramanujan's Q–function" as it already appears in the first letter of Ramanujan to Hardy in 1913. The asymptotic analysis can be done elementarily by developing a continuous approximation of the general term and approximating the resulting Riemann sum by an integral: this is an instance of the Laplace method for sums briefly explained in APPENDIX B: *Laplace's method*, p. 667. (See also [**268**, Sec. 1.2.11.3] and [**382**, Sec. 4.7].) In fact, very precise estimates come out naturally from an analysis of the singularities of the EGF $K(z)$, as we shall see in Chapters VI and VII. The net result is

$$K_n \sim n^n \sqrt{\frac{\pi}{2n}},$$

so that a fraction about $1/\sqrt{n}$ of all the graphs consist of a single component.

Constrained mappings. As is customary with the symbolic method, the constructions (44) also lead to a large number of related counting results. First, the mappings without fixed points, ($(\forall x)\ f(x) \neq x$) and those without $1,2$–cycles, (additionally, $(\forall x)\ f(f(x)) \neq x$), have EGFs

$$\frac{e^{-T(z)}}{1-T(z)}, \qquad \frac{e^{-T(z)-T^2(z)/2}}{1-T(z)}.$$

The first equation is consistent with what a direct count yields, namely $(n-1)^n$, which is asymptotic to $e^{-1}n^n$, so that the fraction of mappings without fixed point is asymptotic to e^{-1}. The second one lends itself easily to complex-asymptotic methods that give

$$n![z^n]\frac{e^{-T-T^2/2}}{1-T} \sim e^{-3/2}n^n,$$

and the proportion is asymptotic to $e^{-3/2}$. These two particular estimates are of the same form as what has been found for permutations (the generalized derangements, Eq. (40)). Such facts that are not quite obvious by elementary probabilistic arguments are in fact neatly explained by the singular theory of combinatorial schemas developed in Part B of this book.

Next, idempotent mappings satisfying $f(f(x)) = f(x)$ for all x correspond to $\mathcal{I} \cong \text{SET}\{\mathcal{Z} \star \text{SET}\{\mathcal{Z}\}\}$, so that

$$I(z) = e^{ze^z} \qquad \text{and} \qquad I_n = \sum_{k=0}^{n} \binom{n}{k} k^{n-k}.$$

(The specification translates the fact that idempotent mappings can have only cycles of length 1 on which are grafted sets of direct antecedents.) The latter sequence is *EIS* **A000248**, which starts as 1,1,3,10,41,196,1057. An asymptotic estimate can be derived either from the Laplace method or, better, from the saddle point method exposed in Chapter VIII.

Several analyses of this type are of relevance to cryptography and the study of random number generators. For instance, the fact that a random mapping over $[1\,..\,n]$ tends to reach a cycle in $O(\sqrt{n})$ steps led Pollard to design a Monte Carlo integer factorization algorithm, see [**269**, p. 371] and [**382**, Sec 8.8]. The algorithm once suitably optimized first led to the factorization of the Fermat number $F_8 = 2^{2^8} + 1$ obtained by Brent in 1980.

$\triangleright$ **19.** *Binary mappings.* The class BF of binary mappings, where each point has either 0 or 2 preimages, is specified by

$$\mathcal{BF} = \text{SET}\{\mathcal{K}\},\ \mathcal{K} = \text{CYC}\{\mathcal{P}\},\ \mathcal{P} = \mathcal{Z} \star \mathcal{B},\ \mathcal{B} = \mathcal{Z} \star \text{SET}_{0,2}\{\mathcal{B}\}$$

(planted trees $\mathcal{P}$ and binary trees $\mathcal{B}$ are needed), so that

$$BF(z) = \frac{1}{\sqrt{1-2z^2}}, \qquad BF_{2n} = \frac{((2n)!)^2}{2^n (n!)^2}.$$

The class $\mathcal{BF}$ is an approximate model of the behaviour of (modular) quadratic functions under iteration. See [**14, 166**] for a general enumerative theory of random mappings including degree-restricted ones. $\triangleleft$

$\triangleright$ **20.** *Partial mappings.* A partial mapping may be undefined at some points, where it can be considered as taking a special value, $\perp$. The iterated preimages of $\perp$ form a forest, while the remaining values organize themselves into a standard mapping. The class $\mathcal{PF}$ of partial mappings is thus specified by $\mathcal{PF} = \text{SET}\{\mathcal{T}\} \star \mathcal{F}$, so that

$$PF(z) = \frac{e^{T(z)}}{1-T(z)} \qquad \text{and} \qquad PF_n = (n+1)^n.$$

All mappings $\frac{1}{1-T}$	Partial $\frac{e^T}{1-T}$	Injective partial $\frac{1}{1-z}e^{z/(1-z)}$	Surjection $\frac{1}{2-e^z}$	Bijection $\frac{1}{1-z}$
Connected ($\mathcal{K}$) $\log\frac{1}{1-T}$	No fixed point $\frac{e^{-T}}{1-T}$	Involution $e^{z+z^2/2}$	Idempotent e^{ze^z}	Binary $\frac{1}{\sqrt{1-2z^2}}$

FIGURE 11. A summary of various counting EGFs relative to mappings.

This construction lends itself to all sorts of variations. For instance, the class PFI of *injective* partial maps is described as sets of chains of linear and circular graphs, $PFI = \text{SET}\{\text{CYC}\{\mathcal{Z}\} + \text{SEQ}_{\geq 1}\{\mathcal{Z}\}\}$, so that

$$PFI(z) = \frac{1}{1-z}e^{z/(1-z)}, \qquad PFI_n = \sum_{i=0}^{n} i!\binom{n}{i}^2$$

(This is a symbolic rewriting of part of the paper [**58**].) ◁

The symbolic method thus gives access to a wide variety of counting results relative to maps satisfying diverse constraints. A summary is offered in Figure 11.

II. 5.3. Labelled graphs. Random graphs form a major chapter of the theory of random discrete structures [**56, 245**]. We examine here enumerative results concerning graphs of low "complexity", that is, graphs which are very nearly trees. (Such graph for instance play an essential rôle in the analysis of early stages of the evolution of a random graph, when edges are successively added, as shown in [**162, 244**].)

Unrooted trees and acyclic graphs. The simplest of all connected graphs are certainly the ones that are acyclic. These are trees, but contrary to the case of Cayley trees, no root is specified. Let $\mathcal{U}$ be the class of all *unrooted* trees. Since a rooted tree (rooted trees are, as we know, counted by $T_n = n^{n-1}$) is an unrooted tree combined with a choice of a distinguished node (there are n possible such choices for trees of size n), one has

$$T_n = nU_n \qquad \text{implying} \qquad U_n = n^{n-2}.$$

At generating function level, this combinatorial equality translates into

$$U(z) = \int_0^z T(w)\,\frac{dw}{w},$$

which integrates to give (take T as the independent variable)

$$U(z) = T(z) - \frac{1}{2}T(z)^2.$$

Since $U(z)$ is the EGF of acyclic connected graphs, the quantity

$$A(z) = e^{U(z)} = e^{T(z)-T(z)^2/2},$$

is the EGF of all acyclic graphs. (Equivalently, these are unordered forests of unrooted trees.) Methods developed in Chapters VI and VII imply the estimate $A_n \sim$

$e^{1/2}\, n^{n-2}$. Surprisingly, perhaps, there are barely more acyclic graphs than unrooted trees—such phenomena are easily explained by singularity analysis.

Unicyclic graphs. The *excess* of a graph is defined as the difference between the number of edges and the number of vertices. For a connected graph, this quantity must be at least -1, the minimal value -1 being precisely attained by unrooted trees. The class $\mathcal{W}_k$ is the class of connected graphs of excess equal to k; in particular $\mathcal{U} = \mathcal{W}_{-1}$. The successive classes $\mathcal{W}_{-1}, \mathcal{W}_0, \mathcal{W}_1, \ldots$, may be viewed as describing connected graphs of increasing complexity.

The class $\mathcal{W}_0$ comprises all connected graphs with the number of edges equal to the number of vertices. Equivalently, a graph in $\mathcal{W}_0$ is a connected graph with exactly one cycle (a sort of "eye"), and for that reason, elements of $\mathcal{W}_0$ are sometimes referred to as "unicyclic components" or "unicycles". In a way, such a graph looks very much like an undirected version of a connected functional graph. Precisely, a graph of $\mathcal{W}_0$ consists of a cycle of length at least 3 (by definition, graphs have neither loops nor multiple edges) that is undirected (the orientation present in the usual cycle construction is killed by identifying cycles isomorphic up to reflection) and on which are grafted trees (these are implicitly rooted by the point at which they are attached to the cycle). With UCYC representing the (new) undirected cycle construction, one thus has

$$\mathcal{W}_0 \cong \mathrm{UC{\small YC}}_{\geq 3}\{\mathcal{T}\}.$$

We claim that this construction is reflected by the EGF equation

$$W_0(z) = \frac{1}{2}\log\frac{1}{1-T(z)} - \frac{1}{2}T(z) - \frac{1}{4}T(z)^2. \tag{46}$$

Indeed one has the isomorphism

$$\mathcal{W}_0 + \mathcal{W}_0 \cong \mathrm{C{\small YC}}_{\geq 3}\{\mathcal{T}\},$$

since we may regard the two disjoint copies on the left as instantiating two possible orientations of the undirected cycle. The result of (46) then follows from the usual translation of the cycle construction. It is originally due to the Hungarian probabilist Rényi in 1959. Asymptotically, one finds (by methods of Chapter VI):

$$n![z^n]W_0 \sim \frac{1}{4}\sqrt{2\pi}n^{n-1/2} - \frac{5}{3}n^{n-1} + \frac{1}{48}\sqrt{2\pi}n^{n-3/2} + \cdots . \tag{47}$$

Finally, the number of graphs made only of trees and unicyclic components is

$$e^{W_{-1}(z)+W_0(z)} = \frac{e^{T/2-3T^2/4}}{\sqrt{1-T}},$$

and asymptotically: $n![z^n]e^{W_{-1}+W_0} \sim \Gamma(3/4)2^{-1/4}e^{-1/2}\pi^{-1/2}n^{n-1/4}$. Such graphs stand just next to acyclic graphs in order of structural complexity. They are the undirected counterparts of functional graphs encountered in the previous section.

$\triangleright$ **21.** *2-Regular graphs.* This is based on Comtet's account [**82**, Sec. 7.3]. A 2–regular graph is an undirected graph in which each vertex has degree exactly 2. Connected 2–regular graphs are thus undirected cycles of length $n \geq 3$, so that the EGF of all 2–regular graphs is

$$R(z) = \frac{e^{-z/2-z^2/4}}{\sqrt{1-z}}.$$

Unrooted trees	$U \equiv W_{-1} = T - T^2/2$	$U_n = n^{n-2}$
Acyclic gr. (forests)	$A = e^{T-T^2/2}$	$A_n \sim e^{1/2} n^{n-2}$
Unicycles	$W_0 = \frac{1}{2}\log\frac{1}{1-T} - \frac{T}{2} - \frac{T^2}{4}$	$W_{0,n} \sim \frac{1}{4}\sqrt{2\pi} n^{n-1/2}$
Trees + unicycles	$B = \dfrac{e^{T/2-3T^2/4}}{\sqrt{1-T}}$	$B_n \sim \Gamma(\dfrac{3}{4})\dfrac{2^{-1/4}}{\sqrt{\pi e}} n^{n-1/4}$
Conn. excess k	$W_k = \dfrac{P_k(T)}{(1-T)^{3k}}$	$W_{k,n} \sim \dfrac{P_k(1)\sqrt{2\pi}}{2^{3k/2}\Gamma(\frac{3}{4}k)} n^{n+(3k-1)/2}$

FIGURE 12. A summary of major enumeration results relative to labelled graphs of small excess.

Given n straight lines in general position, a cloud is defined to be a set of n intersection points no three being collinear. Clouds and 2–regular graphs are equinumerous. [Hint: Use duality.] The asymptotic analysis will serve as a leading example of the singularity analysis process in Chapter VI (Examples VI.1, p. 363 and VI.2, p. 378).

The general enumeration of r–regular graphs becomes somewhat more difficult as soon as $r > 2$. Algebraic aspects are discussed in [**199, 208**] while Bender and Canfield [**30**] have determined the asymptotic formula (for rn even),

$$R_n^{(r)} \sim \sqrt{2} e^{(r^2-1)/4} \frac{r^{r/2}}{e^{r/2} r!} n^{rn/2},$$

for the number of r–regular graphs of size n. ◁

Graphs of fixed excess. The previous discussion suggests considering more generally the enumeration of connected graphs according to excess. E. M. Wright made important contributions in this area [**444, 445, 446**] that are revisited in the famous "giant paper on the giant component" by Janson, Knuth, Łuczak, and Pittel [**244**]. Wright's result are summarized by the following proposition.

PROPOSITION II.6. *The EGF $W_k(z)$ of connected graphs with excess (of edges over vertices) equal to k is, for $k \geq 1$, of the form*

$$W_k(z) = \frac{P_k(T)}{(1-T)^{3k}}, \quad T \equiv T(z), \tag{48}$$

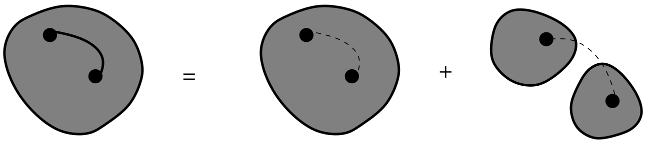
where P_k is a polynomial of degree $3k+2$. For any fixed k, as $n \to \infty$, one has

$$W_{k,n} = n![z^n]W_k(z) = \frac{P_k(1)\sqrt{2\pi}}{2^{3k/2}\Gamma\left(\frac{3}{2}k\right)} n^{n+(3k-1)/2}\left(1 + O(n^{-1/2})\right). \tag{49}$$

The combinatorial part of the proof (see Note 22 below) is an interesting exercise in *graph surgery* and symbolic methods. The analytic part of the statement follows straightforwardly from singularity analysis. The polynomials $P(T)$ and the constants $P_k(1)$ are determined by an explicit nonlinear recurrence; one finds for instance:

$$W_1 = \frac{1}{24}\frac{T^4(6-T)}{(1-T)^3}, \qquad W_2 = \frac{1}{48}\frac{T^4(2+28T-23T^2+9T^3-T^4)}{(1-T)^6}.$$

▷ **22.** *Wright's surgery.* The full proof of Proposition II.6 by symbolic methods requires the notion of pointing in conjunction with multivariate generating function techniques of Chapter III. It is convenient to define $w_k(z,y) := y^k W_k(zy)$, which is a bivariate generating function with y marking the number of edges. Pick up an edge in a connected graph of excess $k+1$, then remove it. This results either in a connected graph of excess k with two pointed vertices (and no edge in between) or in two connected components of respective excess h and $k-h$, each with a pointed vertex. Graphically:



This translates into the differential recurrence on the w_k ($\partial_x := \frac{\partial}{\partial x}$),

$$2\partial_y w_{k+1} = \left(z^2\partial_z^2 w_k - 2y\partial_y w_k\right) + \sum_{h=-1}^{k+1} \left(z\partial_z w_h\right)\cdot\left(z\partial_z w_{k-h}\right),$$

and similarly for $W_k(z) = w_k(z,1)$. From there, it can be verified by induction that each W_k is a rational function of $T \equiv W_{-1}$. (See Wright's original papers [**444, 445, 446**] or [**244**] for details.) $\lhd$

As explained in the giant paper [**244**], such results combined with complex analytic techniques provide with great detail information on the aspect of a random graph $\Gamma(n,m)$ with n nodes and m edges. In the sparse case where m is of the order of n, one finds the following properties to hold "*with high probability*" (w.h.p.)[7], that is, with probability tending to 1 as $n \to \infty$.

- For $m = \mu n$, with $\mu < \frac{1}{2}$, the random graph $\Gamma(m,n)$ has w.h.p. only tree and unicycle components; the largest component is w.h.p. of size $O(\log n)$.
- For $m = \frac{1}{2}n + O(n^{2/3})$, w.h.p. there appear one or several semi-giant components that have size $O(n^{2/3})$.
- For $m = \mu n$, with $\mu > \frac{1}{2}$, there is w.h.p. a unique giant component of size proportional to n.

In each case, refined estimates follow from a detailed analysis of corresponding generating functions, which is a main theme of [**162**] and especially [**244**]. Raw forms of these results were first obtained by Erdős and Rényi who launched the subject in a famous series of papers dating from 1959–60; see the books [**56, 245**] for a probabilistic context and the paper [**31**] for the finest counting estimates available. In contrast, the enumeration of *all* connected graphs (irrespective of the number of edges, that is, without excess being taken into account) is a relatively easy problem treated in the next section. Many other classical aspects of the enumerative theory of graphs are covered in the book *Graphical Enumeration* by Harary and Palmer [**223**].

II. 6. Additional constructions

Like in the unlabelled case, pointing and substitution are available in the world of labelled structures (Section II. 6.1), and implicit definitions enlarge the scope of the symbolic method (Section II. 6.2). The inversion process needed to enumerate implicit structures is even simpler, since in the labelled universe sets and cycles have

[7]Synonymous expressions are "asymptotically almost surely" (a.a.s) and "in probability". The term "almost surely" is sometimes used, though it lends itself to confusion with properties of continuous measures.

more concise translations as operators over EGF. Finally, and this departs significantly from Chapter I, the fact that integer labels are naturally ordered makes it possible to take into account certain order properties of combinatorial structures (Section II. 6.3).

II. 6.1. Pointing and substitution. The pointing operation is entirely similar to its unlabelled counterpart since it consists in distinguishing one atom amongst all the ones that compose an object of size n. The definition of composition for labelled structures is however a bit more subtle as it requires singling out "leaders" in substituends.

Pointing. The *pointing* of a class $\mathcal{B}$ is defined by

$$\mathcal{A} = \Theta\mathcal{B} \qquad \text{iff} \qquad \mathcal{A}_n = [1\,..\,n] \times \mathcal{B}_n.$$

In other words, in order to generate an element of $\mathcal{A}$, select one of the n labels and point at it. Clearly

$$A_n = n \cdot B_n \implies A(z) = z\frac{d}{dz}B(z).$$

Substitution (composition). The *composition* or *substitution* can be defined so that it corresponds *a priori* to composition of generating functions. It is formally defined as

$$\mathcal{B} \circ \mathcal{C} = \sum_{k=0}^{\infty} \mathcal{B}_k \times \text{SET}_k\{\mathcal{C}\},$$

so that its EGF is

$$\sum_{k=0}^{\infty} B_k \frac{(C(z))^k}{k!} = B(C(z)).$$

A combinatorial way of realizing this definition and form an arbitrary object of $\mathcal{B} \circ \mathcal{C}$, is as follows. First select an element of $\beta \in \mathcal{B}$ called the "base" and let $k = |\beta|$ be its size; then pick up a k–set of $\mathcal{C}^k$; the elements of the k–set are naturally ordered by value of their "leader" (the leader of an object being by convention the value of its smallest label); the element with leader of rank r is then substituted to the node labelled by value r of β.

THEOREM II.3. *The combinatorial constructions of pointing and substitution are admissible.*

$$\begin{aligned} \mathcal{A} = \Theta\mathcal{B} \quad &\implies \quad A(z) = z\partial_z B(z), \qquad \partial_z \equiv \frac{d}{dz} \\ \mathcal{A} = \mathcal{B} \circ \mathcal{C} \quad &\implies \quad A(z) = B(C(z)). \end{aligned}$$

For instance, the EGF of (relabelled) pairings of elements drawn from $\mathcal{C}$ is

$$e^{C(z)+C(z)^2/2},$$

since the EGF of involutions is $e^{z+z^2/2}$.

▷ **23.** *Standard constructions based on substitutions.* The sequence class of $\mathcal{A}$ may be defined by composition as $\mathcal{P} \circ \mathcal{A}$ where $\mathcal{P}$ is the set of all permutations. The set class of $\mathcal{A}$ may be defined as $\mathcal{U} \circ \mathcal{A}$ where $\mathcal{U}$ is the class of all urns. Similarly, cycles are obtained by substitution into circular graphs. Thus,

$$\text{SEQ}(\mathcal{A}) \cong \mathcal{P} \circ \mathcal{A}, \qquad \text{SET}(\mathcal{A}) \cong \mathcal{U} \circ \mathcal{A}, \qquad \text{CYC}(\mathcal{A}) \cong \mathcal{C} \circ \mathcal{A}.$$

In this way, permutation, urns and circle graphs appear as archetypal classes in a development of combinatorial analysis based on composition. (Joyal's "theory of species" [**248**] and the book by Bergeron, Labelle, and Leroux [**37**] make a great use of such ideas and show that an extensive theory of combinatorial enumeration can be based on the concept of substitution.) ◁

▷ **24.** *Distinct component sizes.* The EGFs of permutations with cycles of distinct lengths and of set partitions with parts of distinct sizes are

$$\prod_{n=1}^{\infty}(1+\frac{z^n}{n}),\ \prod_{n=1}^{\infty}(1+\frac{z^n}{n!}).$$

The probability that a permutation of $\mathcal{P}_n$ has distinct cycle sizes tends to $e^{-\gamma}$; see [**213**, Sec. 4.1.6] for a Tauberian argument and [**352**] for precise asymptotics. The corresponding analysis for set partitions is treated in the seven author paper [**257**]. ◁

II. 6.2. Implicit structures. Let $\mathcal{X}$ be a labelled class implicitly defined by either of the equations

$$\mathcal{A} = \mathcal{B} + \mathcal{X}, \qquad \mathcal{A} = \mathcal{B} \star \mathcal{X}.$$

Then, solving the corresponding EGF equations leads to

$$X(z) = A(z) - B(z), \qquad X(z) = \frac{A(z)}{B(z)},$$

respectively. For the composite labelled constructions SEQ, SET, CYC, the algebra is equally easy.

THEOREM II.4 (Implicit specifications). *The generating functions associated to the implicit equations in $\mathcal{X}$*

$$\mathcal{A} = \mathrm{SEQ}(\mathcal{X}), \qquad \mathcal{A} = \mathrm{SET}(\mathcal{X}), \qquad \mathcal{A} = \mathrm{CYC}(\mathcal{X}),$$

are respectively

$$X(z) = 1 - \frac{1}{A(z)}, \qquad X(z) = \log A(z), \qquad X(z) = 1 - e^{-A(z)}.$$

EXAMPLE 15. *Connected graphs.* In the context of graphical enumerations, the labelled set construction takes the form of an enumerative formula relating a class of graphs $\mathcal{G}$ and the subclass of its connected graphs $\mathcal{K} \subset \mathcal{G}$:

$$\mathcal{G} = \mathrm{SET}(\mathcal{K}) \implies G(z) = e^{K(z)}.$$

This basic formula is known in graph theory [**223**] as the *exponential formula*.

Consider the class $\mathcal{G}$ of all (undirected) labelled graphs, the size of a graph being the number of its nodes. Since a graph is determined by the choice of its set of edges, there are $\binom{n}{2}$ potential edges each of which may be taken in or out, so that $G_n = 2^{\binom{n}{2}}$. Let $\mathcal{K} \subset \mathcal{G}$ be the subclass of all connected graphs. The exponential formula determines $K(z)$ implicitly,

$$\begin{aligned} K(z) &= \log\left(1 + \sum_{n\ge 1} 2^{\binom{n}{2}} \frac{z^n}{n!}\right) \\ &= z + \frac{z^2}{2!} + 4\frac{z^3}{3!} + 38\frac{z^4}{4!} + 728\frac{z^5}{5!}, \end{aligned}$$

where the sequence is *EIS* **A001187**. The series is divergent, that is, it has radius of convergence 0. It can nonetheless be manipulated as a formal series (APPENDIX A: *Formal power* *series*, p. 648). Expanding by means of $\log(1+u) = u - u^2/2 + \cdots$, yields a complicated convolution expression for K_n:

$$K_n = 2^{\binom{n}{2}} - \frac{1}{2}\sum \binom{n}{n_1, n_2} 2^{\binom{n_1}{2}+\binom{n_2}{2}} + \frac{1}{3}\sum \binom{n}{n_1, n_2, n_3} 2^{\binom{n_1}{2}+\binom{n_2}{2}+\binom{n_3}{2}} - \cdots .$$

(The kth term is a sum over $n_1 + \cdots + n_k = n$, with $0 < n_j < n$.) Given the very fast increase of G_n with n, for instance

$$2^{\binom{n+1}{2}} = 2^n \, 2^{\binom{n}{2}},$$

a detailed analysis of the various terms of the expression of K_n shows predominance of the first sum, and, in that sum itself, predominance of the extreme terms corresponding to $n_1 = n - 1$ or $n_2 = n - 1$, so that

$$K_n = 2^{\binom{n}{2}} \left(1 - 2n2^{-n} + o(2^{-n})\right). \tag{50}$$

Thus, almost all labelled graphs of size n are connected. In addition, the error term decreases very fast: for instance, for $n = 18$, an exact computation based on the generating function formula reveals that a proportion only 0.0001373291074 of all the graphs are not connected—this is *extremely* close to the value 0.0001373291016 predicted by the main terms in the asymptotic formula (50). Notice that here good use could be made of a purely divergent generating function for asymptotic enumeration purposes. .... END OF EXAMPLE 15. □

▷ **25.** *Bipartite graphs.* A plane bipartite graph is a pair (G, ω) where G is a labelled graph, $\omega = (\omega_W, \omega_E)$ is a bipartition of the nodes (into *West* and *East* categories), and the edges are such that they only connect nodes from ω_W to nodes of ω_E. A direct count shows that the EGF of plane bipartite graphs is

$$\Gamma(z) = \sum_n \gamma_n \frac{z^n}{n!} \text{ with } \gamma_n = \sum_k \binom{n}{k} 2^{k(n-k)}.$$

The EGF of plane bipartite graphs that are connected is $\log \Gamma(z)$.

A bipartite graph is a labelled graph whose nodes can be partitioned into two groups so that edges only connect nodes of different groups. The EGF of bipartite graphs is

$$\exp\left(\frac{1}{2}\log\Gamma(z)\right) = \sqrt{\Gamma(z)}.$$

[Hint. The EGF of a connected bipartite graph is $\frac{1}{2}\log\Gamma(z)$ as a factor of $\frac{1}{2}$ kills the East–West orientation present in a connected plane bipartite graph. See Wilf's book [**437**, p. 78] for details.] ◁

Note. The class of all graphs is not fully constructible in the sense that it does not admit a complete construction starting from single atoms and involving only sums, products, sets and cycles. In a sense, it is too "large" to be constructible. This assertion is established rigorously by complex analysis since EGFs of constructible classes must have a nonzero radius of convergence, as proved in Chapter IV.

II. 6.3. Order constraints. A construction well suited to taking into account many order properties of combinatorial structures is the modified labelled product,

$$\mathcal{A} = (\mathcal{B}^{\square} \star \mathcal{C}).$$

This denotes the subset of the product $\mathcal{B} \star \mathcal{C}$ formed with elements such that the smallest label is constrained to lie in the $\mathcal{B}$ component. (To make this definition consistent, it must be assumed that $B_0 = 0$.) We call this binary operation on structures the *boxed* product.

THEOREM II.5. *The boxed product is admissible.*

$$(51) \qquad \mathcal{A} = (\mathcal{B}^{\square} \star \mathcal{C}) \quad \Longrightarrow \quad A(z) = \int_0^z (\partial_t B(t)) \cdot C(t)\, dt, \qquad \partial_t \equiv \frac{d}{dt}.$$

PROOF. The definition of boxed products implies the coefficient relation

$$A_n = \sum_{k=1}^{n} \binom{n-1}{k-1} B_k C_{n-k}.$$

The binomial coefficient that appears in the standard labelled product is now modified since only $n-1$ labels need to be distributed between the two components, $k-1$ going to the $\mathcal{B}$ component (that is constrained to contain the label 1 already) and $n-k$ to the $\mathcal{C}$ component. From the equivalent form

$$A_n = \frac{1}{n} \sum_{k=0}^{n} \binom{n}{k} (kB_k)\, C_{n-k},$$

the result follows by taking EGFs. □

A useful special case is the min–rooting operation,

$$\mathcal{A} = \{1\}^{\square} \star \mathcal{C},$$

for which a variant definition goes as follows. Take in all possible ways elements $\gamma \in \mathcal{C}$, prepend an atom with a label smaller than the labels of γ, for instance 0, and relabel in the canonical way over $[1\,..\,(n+1)]$ by shifting all label values by 1. Clearly $A_{n+1} = C_n$ which yields

$$A(z) = \int_0^z C(t)\, dt,$$

a result also consistent with the general formula of boxed products.

For some applications, it is easier to impose constraints on the *maximal* label rather than the minimum. The max-boxed product written

$$\mathcal{A} = (\mathcal{B}^{\blacksquare} \star \mathcal{C}),$$

is then defined by the fact the maximum is constrained to lie in the $\mathcal{B}$-component of the labelled product. Naturally, the translation by an integral in (51) remains valid for this trivially modified boxed product.

▷ **26.** *Combinatorics of integration.* In the perspective of this book, integration by parts has an immediate interpretation. Indeed, the equality,

$$\int_0^z A'(t) \cdot B(t)\, dt = A(z) \cdot B(z) - \int_0^z A(t) \cdot B'(t)\, dt,$$

reads off as: *"The smallest label in an ordered pair, if it appears on the left, cannot appear on the right."* ◁

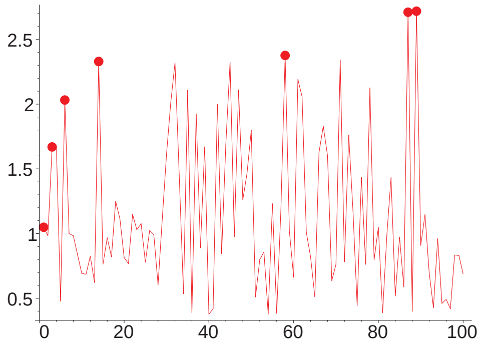


FIGURE 13. A numerical sequence of size 100 with records marked by circles: there are 7 records that occur at times $1, 3, 5, 11, 60, 86, 88$.

EXAMPLE 16. *Records in permutations.* Given a sequence of numerical data, $x = (x_1, \ldots, x_n)$ assumed all distinct, a *record* in that sequence is defined to be an element x_j such that $x_k < x_j$ for all $k < j$. (A record is an element "better" than its predecessors!) Figure 13 displays a numerical sequence of length $n = 100$ that has 7 records. Confronted to such data, a statistician will typically want to determine whether the data obey purely random fluctuations or there could be some indications of a "trend" or of a "bias" [**90**, Ch. 10]. (Think of the data as reflecting share prices or athletic records, say.) In particular, if the x_j are independently drawn from a continuous distribution, then the number of records obeys the same laws as in a random permutation of $[1 \mathinner{.\,.} n]$. This statistical preamble then invites the question: *How many permutations of n have k records?*

First, we start with a special brand of permutations, the ones that have their *maximum* at the beginning. Such permutations are defined as ('$\blacksquare$' indicates the boxed product based on the maximum label)

$$\mathcal{Q} = (\mathcal{Z}^{\blacksquare} \star \mathcal{P}),$$

where $\mathcal{P}$ is the class of all permutations. Observe that this gives the EGF

$$Q(z) = \int_0^z \left(\frac{d}{dt}t\right) \cdot \frac{1}{1-t}\,dt = \log\frac{1}{1-z},$$

implying the obvious result $Q_n = (n-1)!$ for all $n \geq 1$. These are exactly the permutations with *one* record. Next, consider the class

$$\mathcal{P}^{(k)} = \mathrm{SET}_k(\mathcal{Q}).$$

The elements of $\mathcal{P}^{(k)}$ are unordered sets of cardinality k with elements of type $\mathcal{Q}$. Define the (max) leader of any component of $\mathcal{P}^{(k)}$ as the value of its maximal element. Then, if we place the components in sequence, ordered by increasing values of their leaders, then read off the whole sequence, we obtain a permutation with k records exactly. The correspondence[8] is clearly revertible. Here is an illustration, with leaders underlined:

$$\begin{aligned}\{(\underline{\mathbf{7}},2,6,1),\ (\underline{\mathbf{4}},3),\ (\underline{\mathbf{9}},8,5)\} &\cong [(\underline{\mathbf{4}},3),\ (\underline{\mathbf{7}},2,6,1),\ (\underline{\mathbf{9}},8,5))] \\ &\cong \underline{\mathbf{4}},3,\underline{\mathbf{7}},2,6,1,\underline{\mathbf{9}},8,5.\end{aligned}$$

[8]This correspondence can also be viewed as a transformation on permutations that maps the number of records to the number of cycles—it is known as Foata's fundamental correspondence [**294**, Sec. 10.2].

Thus, the number of permutations with k records is determined by

$$P^{(k)}(z) = \frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k, \qquad P_n^{(k)} = \begin{bmatrix} n \\ k \end{bmatrix},$$

where we recognize Stirling cycle numbers from Example 12. In other words:

The number of permutations of size n having k records is counted by the Stirling "cycle" number $\left[{n \atop k}\right]$.

Returning to our statistical problem, the treatment of Example 12 p. 113 (to be revisited in Chapter III) shows that the expected number of records in a random permutation of size n equals H_n, the harmonic number. One has $\mathrm{H}_{100} \doteq 5.18$, so that for 100 data items, a little more than 5 records are expected on average. The probability of observing 7 records or more is still about 23%, an altogether not especially rare event. In contrast, observing twice as many records, that is, 14, would be a fairly strong indication of a bias since, on random data, the event has probability very close to 10^{-4}. Altogether, the present discussion is consistent with the hypothesis for the data of Figure 13 to have been generated independently at random (and indeed they were). . . . . END OF EXAMPLE 16. □

It is possible to base a fair part of the theory of labelled constructions on sums and products in conjunction with the boxed product. In effect, consider the three relations

$$\begin{array}{llll}
\mathcal{F} = \mathrm{SEQ}\{\mathcal{G}\} & \Longrightarrow & f(z) = \dfrac{1}{1-g(z)}, & f = 1 + gf \\
\mathcal{F} = \mathrm{SET}\{\mathcal{G}\} & \Longrightarrow & f(z) = e^{g(z)}, & f = 1 + \displaystyle\int g' f \\
\mathcal{F} = \mathrm{CYC}\{\mathcal{G}\} & \Longrightarrow & f(z) = \log\dfrac{1}{1-g(z)}, & f = \displaystyle\int g' \frac{1}{1-g}
\end{array}$$

The last column is easily checked to provide an alternative form of the standard operator corresponding to sequences, sets, and cycles. Each case is then itself deduced directly from Theorem II.5 and the labelled product rule:

Sequences: they obey the recursive definition

$$\mathcal{F} = \mathrm{SEQ}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong \{\epsilon\} + (\mathcal{G} \star \mathcal{F}).$$

Sets: we have

$$\mathcal{F} = \mathrm{SET}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong \{\epsilon\} + (\mathcal{G}^{\blacksquare} \star \mathcal{F}),$$

which means that, in a set, one can always single out the component with the largest label, the rest of the components forming a set. In other words, when this construction is repeated, the elements of a set can be canonically arranged according to increasing values of their largest labels, the "leaders". (We recognize here a generalization of the construction used for records in permutations.)

Cycles: The element of a cycle that contains the largest label can be taken canonically as the cycle "starter", which is then followed by an arbitrary sequence of elements upon traversing the cycle in circular order. Thus

$$\mathcal{F} = \mathrm{CYC}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong (\mathcal{G}^{\blacksquare} \star \mathrm{SEQ}\{\mathcal{G}\}).$$

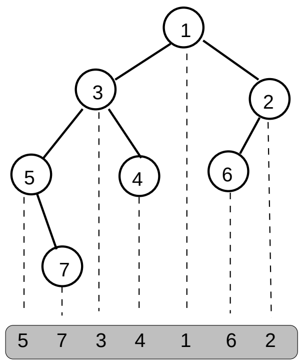


FIGURE 14. A permutation of size 7 and its increasing binary tree lifting.

Greene [**215**] has developed a complete framework of labelled grammars based on standard and boxed labelled products. In its basic form, its expressive power is essentially equivalent to ours, because of the above relations. More complicated order constraints, dealing simultaneously with a collection of larger and smaller elements, can be furthermore taken into account within this framework.

▷ **27.** *Higher order constraints, after Greene.* Let the symbols $\square$, $\boxdot$, $\blacksquare$ represent smallest, second smallest, and largest labels respectively. One has the correspondences (with $\partial_z = \frac{d}{dz}$)

$$\begin{array}{ll}
\mathcal{A} = \left(\mathcal{B}^{\square} \star \mathcal{C}^{\blacksquare}\right) & \qquad \partial_z^2 A(z) = (\partial_z B(z)) \cdot (\partial_z C(z)) \\
\mathcal{A} = \left(\mathcal{B}^{\square\,\blacksquare} \star \mathcal{C}\right) & \qquad \partial_z^2 A(z) = \left(\partial_z^2 B(z)\right) \cdot C(z) \\
\mathcal{A} = \left(\mathcal{B}^{\square} \star \mathcal{C}^{\boxdot} \star \mathcal{D}^{\blacksquare}\right) & \qquad \partial_z^3 A(z) = (\partial_z B(z)) \cdot (\partial_z C(z)) \cdot (\partial_z D(z)) \,,
\end{array}$$

and so on. These can be transformed into (iterated) integral representations. [See [**215**] for more.] ◁

The next two examples demonstrate the usefulness of min-rooting used in conjunction with recursion. In this way, trees satisfying some order conditions can be constructed and enumerated easily. This is in turn gives access to new characteristics of permutations.

EXAMPLE 17. *Increasing binary trees and alternating permutations.* To each permutation, one can associate bijectively a binary tree of a special type called an *increasing binary tree* and sometimes a heap–ordered tree or a tournament tree. This is a plane rooted binary tree in which internal nodes bear labels in the usual way, but with the additional constraint that node labels increase along any branch stemming from the root. Such trees are closely related to classical data structures of computer science, like heaps and binomial queues [**84, 381**].

The correspondence (Figure 14) is as follows: Given a permutation of a set written as a word, $\sigma = \sigma_1\sigma_2 \ldots \sigma_n$, factor it in the form $\sigma = \sigma_L \cdot \min(\sigma) \cdot \sigma_R$, with $\min(\sigma)$ the smallest label value in the permutation, and σ_L, σ_R the factors left and right of $\min(\sigma)$. Then the binary tree $\beta(\sigma)$ is defined recursively in the format $\langle$root, left,right$\rangle$ by

$$\beta(\sigma) = \langle \min(\sigma), \beta(\sigma_L), \beta(\sigma_R) \rangle, \qquad \beta(\epsilon) = \epsilon.$$

The empty tree (consisting of a unique external node of size 0) goes with the empty permutation ϵ. Conversely, reading the labels of the tree in symmetric (infix) order gives back the original permutation. (The correspondence is described for instance in Stanley's book [**391**, p. 23–25] who says that "it has been primarily developed by the French", pointing at [**186**].)

Thus, the family $\mathcal{I}$ of binary increasing trees satisfies the recursive definition

$$\mathcal{I} = \{\epsilon\} + \left(\mathcal{Z}^{\square} \star \mathcal{I} \star \mathcal{I}\right),$$

which implies the nonlinear integral equation for the EGF

$$I(z) = 1 + \int_0^z I(t)^2 \, dt.$$

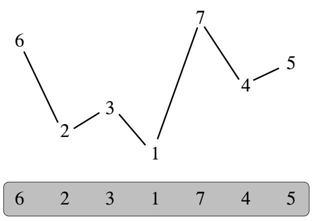
This equation reduces to $I'(z) = I(z)^2$ and, under the initial condition $I(0) = 1$, it admits the solution $I(z) = (1-z)^{-1}$. Thus $I_n = n!$, which is consistent with the fact that there are as many increasing trees as there are permutations.

The construction of increasing trees associated with permutation is instrumental in deriving EGFs relative to various local order patterns in permutations. We illustrate its use here by counting the number of *up-and-down* (or *zig-zag*) permutations, also known as *alternating* permutations. The result, already mentioned in our *Invitation* chapter, was first derived by Désiré André in 1881 by means of a direct recurrence argument.

A permutation $\sigma = \sigma_1 \sigma_2 \ldots \sigma_n$ is an alternating permutation if

$$\sigma_1 > \sigma_2 < \sigma_3 > \sigma_4 < \cdots, \tag{52}$$

so that pairs of consecutive elements form a succession of ups and downs; for instance,



Consider first the case of an alternating permutation of *odd* size. It can be checked that the corresponding increasing trees have no one–way branching nodes, so that they consist solely of binary nodes and leaves. Thus, the corresponding specification is

$$\mathcal{J} = \mathcal{Z} + \left(\mathcal{Z}^{\square} \star \mathcal{J} \star \mathcal{J}\right),$$

so that

$$J(z) = z + \int_0^z J(t)^2 \, dt \qquad \text{and} \qquad \frac{d}{dz} J(z) = 1 + J(z)^2.$$

The equation admits separation of variables, which implies (with $J(0) = 0$)

$$J(z) = \tan(z) = z + 2\frac{z^3}{3!} + 16\frac{z^5}{5!} + 272\frac{z^7}{7!} + \cdots.$$

The coefficients J_{2n+1} are known as the *tangent numbers* or the *Euler numbers* of odd index (*EIS* **A000182**).

Alternating permutations of *even* size defined by the constraint (52) and denoted by $\overline{\mathcal{J}}$ can be determined from

$$\overline{\mathcal{J}} = \{\epsilon\} + \left(\mathcal{Z}^\square \star \mathcal{J} \star \overline{\mathcal{J}}\right),$$

since now all internal nodes of the tree representation are binary, except for the rightmost one that only branches on the left. Thus, $\overline{J}'(z) = \tan(z)\overline{J}(z)$, and the EGF is

$$\overline{J}(z) = \frac{1}{\cos(z)} = 1 + 1\frac{z^2}{2!} + 5\frac{z^4}{4!} + 61\frac{z^6}{6!} + 1385\frac{z^8}{8!} + \cdots,$$

where the coefficients $\overline{J}_{2n}$ are the *secant numbers* also known as Euler numbers of even index (*EIS* **A000364**). .... END OF EXAMPLE 17. □

Use will be made later in this book (Chapter III, p. 22) of this important tree representation of permutations as it opens access to parameters like the number of descents, runs, and (once more!) records in permutations. Analyses of increasing trees also inform us of crucial performance issues regarding binary search trees, quicksort, and heap–like priority queue structures [**307, 382, 427, 429**].

▷ **28.** *Combinatorics of trigonometrics.* Interpret $\tan\frac{z}{1-z}$, $\tan\tan z$, $\tan(e^z-1)$ as EGFs of combinatorial classes. ◁

EXAMPLE 18. *Increasing Cayley trees and regressive mappings.* An increasing Cayley tree is a Cayley tree (i.e., it is nonplane and rooted) whose labels along any branch stemming from the root form an increasing sequence. In particular, the minimum must occur at the root, and no plane embedding is implied. Let $\mathcal{K}$ be the class of such trees. The recursive specification is now

$$\mathcal{K} = \left(\mathcal{Z}^\square \star \textsc{Set}\{\mathcal{K}\}\right).$$

The generating function thus satisfies the functional relations

$$K(z) = \int_0^z e^{K(t)}\,dt, \qquad K'(z) = e^{K(z)},$$

with $K(0) = 0$. Integration of $K'e^{-K} = 1$ shows that

$$K(z) = \log\frac{1}{1-z} \qquad \text{and} \qquad K_n = (n-1)!.$$

Thus the number of increasing Cayley trees is $(n-1)!$, which is also the number of permutations of size $n-1$. These trees have been studied by Meir and Moon [**312**] under the name of “recursive trees”, a terminology that we do not however retain here.

The simplicity of the formula $K_n = (n-1)!$ certainly calls for a combinatorial interpretation. In fact, an increasing Cayley tree is fully determined by its child parent relationship (Figure 15). Otherwise said, to each increasing Cayley tree τ, we associate a partial map $\phi = \phi_\tau$ such that $\phi(i) = j$ iff the label of the parent of i is j. Since the root of tree is an orphan, the value of $\phi(1)$ is undefined, $\phi(1) = \perp$; since the tree is increasing, one has $\phi(i) < i$ for all $i \geq 2$. A function satisfying these last two conditions is called a *regressive mapping*. The correspondence between trees and regressive mappings is then easily seen to be a bijective one.

Thus regressive mappings on the domain $[1\,..\,n]$ and increasing Cayley trees are equinumerous, so that we may as well use $\mathcal{K}$ to denote the class of regressive mappings. Now, a regressive mapping of size n is evidently determined by a single choice for $\phi(2)$ (since $\phi(2) = 1$), two possible choices for $\phi(3)$ (either of $1, 2$), and so on. Hence the formula

$$K_n = 1 \cdot 2 \cdot 3 \cdots (n-1)$$

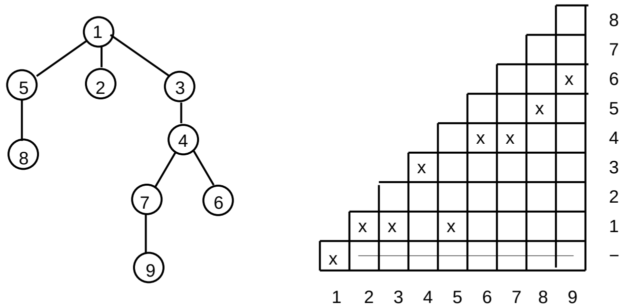


FIGURE 15. An increasing Cayley tree (left) and its associated regressive mapping (right).

receives a natural interpretation. . . . . END OF EXAMPLE 18. □

▷ **29.** *Regressive mappings and permutations.* Regressive mappings can be related directly to permutations. The construction that associates a regressive mapping to a permutation is called the "inversion table" construction; see [**269, 382**]. Given a permutation $\sigma = \sigma_1, \dots, \sigma_n$, associate to it a function $\psi = \psi_\sigma$ from $[1\,..\,n]$ to $[0\,..\,n-1]$ by the rule

$$\psi(j) = \operatorname{card}\left\{k < j \;\middle|\; \sigma_k > \sigma_j\right\}.$$

The function ψ is a trivial variant of a regressive mapping. ◁

▷ **30.** *Rotations and increasing trees.* An increasing Cayley tree can be canonically drawn by ordering descendants of each node from left to right according to their label values. The rotation correspondence (p. 69) then gives rise to a binary increasing tree. Hence, increasing Cayley trees and increasing binary trees are also directly related. Summarizing this note and the previous one, we have a quadruple combinatorial connection,

Increasing Cayley tree ≅ *Regressive mappings* ≅ *Permutations* ≅ *Increasing binary trees,*

that opens the way to yet more permutation enumerations. ◁

II. 7. Perspective

Together with the previous chapter and Figure I.14, this chapter and Figure 16 provide the basis for the symbolic method that is at the core of analytic combinatorics. The translations of the basic constructions for labelled classes to EFGs could hardly be simpler, but, as we have seen, they are sufficiently powerful to embrace numerous classical results in combinatorics, ranging from the birthday and coupon collector problems to graph enumeration.

The examples that we have considered for second-level structures, trees, mappings, and graphs lead to EGFs that are simple to express and natural to generalize. (Often, the simple form is misleading—direct derivations of many of these EGFs that do not appeal to the symbolic method can be rather intricate.) Indeed, the symbolic method provides a framework that allows us to understand the nature of many of these combinatorial classes. From there, numerous seemingly scattered counting problems can be organized into broad structural categories and solved in an almost mechanical manner.

1. The main constructions of union, and product, sequence, set, and cycle for labelled structures together with their translation into exponential generating functions.

Construction		EGF
Union	$\mathcal{A} = \mathcal{B} + \mathcal{C}$	$A(z) = B(z) + C(z)$
Product	$\mathcal{A} = \mathcal{B} \star \mathcal{C}$	$A(z) = B(z) \cdot C(z)$
Sequence	$\mathcal{A} = \mathrm{SEQ}\{\mathcal{B}\}$	$A(z) = \dfrac{1}{1 - B(z)}$
Set	$\mathcal{A} = \mathrm{SET}\{\mathcal{B}\}$	$A(z) = \exp(B(z))$
Cycle	$\mathcal{A} = \mathrm{CYC}\{\mathcal{B}\}$	$A(z) = \log \dfrac{1}{1 - B(z)}$

2. The translation for sets, multisets, and cycles of fixed cardinality.

Construction		EGF
Sequence	$\mathcal{A} = \mathrm{SEQ}_k\{\mathcal{B}\}$	$A(z) = B(z)^k$
Set	$\mathcal{A} = \mathrm{SET}_k\{\mathcal{B}\}$	$A(z) = \dfrac{1}{k!} B(z)^k$
Cycle	$\mathcal{A} = \mathrm{CYC}_k\{\mathcal{B}\}$	$A(z) = \dfrac{1}{k} B(z)^k$

3. The additional constructions of pointing and substitution.

Construction		EGF
Pointing	$\mathcal{A} = \Theta\mathcal{B}$	$A(z) = z\frac{d}{dz}B(z)$
Substitution	$\mathcal{A} = \mathcal{B} \circ \mathcal{C}$	$A(z) = B(C(z))$

4. The "boxed" product.

$$\boxed{\mathcal{A} = (\mathcal{B}^{\square} \star \mathcal{C}) \implies A(z) = \int_0^z \left(\frac{d}{dt} B(t)\right) \cdot C(t)\, dt.}$$

FIGURE 16. A "dictionary" of *labelled* constructions together with their translation into *exponential* generating functions (EGFs). The first constructions are counterparts of the unlabelled constructions of the previous chapter (the multiset construction is not meaningful here). The translation for composite constructions of bounded cardinality appears to be simple. Finally, the boxed product is specific to labelled structures. (Compare with the unlabelled counterpart, Figure 14 of Chapter I, p. 14.)

Again, the symbolic method is only half of the story (the "combinatorics" in analytic combinatorics), leading to EGFs for the counting sequences of numerous interesting combinatorial classes. While some of these EGFs lead immediately to explicit counting results, others require the classical techniques in complex analysis and asymptotic analysis that are covered in Part B (the "analytic" part of analytic combinatorics) to deliver asymptotic estimates. Together with these techniques, the basic constructions, translations, and applications that we have discussed in this chapter reinforce the overall message that the symbolic method is a systematic approach that is successful for addressing classical and new problems in combinatorics, generalizations, and applications.

We have been focussing on *enumeration problems*—counting the number of objects of a given size in a combinatorial class. In the next chapter, we consider how to extend the symbolic method to help analyse other properties of combinatorial classes.

The labelled set construction and the exponential formula were recognized early by researchers working in the area of graphical enumerations [**223**]. Foata [**184**] proposed a detailed formalization in 1974 of labelled constructions, especially sequences and sets, under the names of partitional complex; a brief account is also given by Stanley in his survey [**390**]. This is parallel to the concept of "prefab" due to Bender and Goldman [**32**]. The books by Comtet [**82**], Wilf [**437**], Stanley [**391**], or Goulden and Jackson [**208**] have many examples of the use of labelled constructions in combinatorial analysis.

Greene [**215**] has introduced a general framework of "labelled grammars" largely based on the boxed product with implications for the random generation of combinatorial structures in his 1983 dissertation. Joyal's theory of species dating from 1981 (see [**248**] for the original article and the book by Bergeron, Labelle, and Leroux [**37**] for a rich exposition), is based on category theory; it presents the advantage of uniting in a common theory the unlabelled and the labelled worlds.

Flajolet, Salvy, and Zimmermann have developed a specification language closely related to the system exposed here. They show in [**173**] how to compile automatically specifications into generating functions; this is complemented by a calculus that produces fast random generation algorithms [**183**].

III

Combinatorial Parameters and Multivariate Generating Functions

Generating functions find averages, etc.
— HERBERT WILF [**437**]

Je n'ai jamais été assez loin pour bien sentir l'application de l'algèbre à la géométrie. Je n'aimais point cette manière d'opérer sans voir ce qu'on fait, et il me sembloit que résoudre un problème de géométrie par les équations, c'étoit jouer un air en tournant une manivelle[1].
— JEAN-JACQUES ROUSSEAU, *Les Confessions*, Livre VI

Contents

Many scientific endeavours demand precise quantitative information on probabilistic properties of *parameters* of combinatorial objects. For instance, when designing, analysing, and optimizing a sorting algorithm, it is of interest to determine what the typical disorder of data obeying a given model of randomness is, and do so in the mean, or even in distribution, either exactly or asymptotically. Similar situations arise in a broad variety of fields, including probability theory and statistics, computer science, information theory, statistical physics, and computational biology. The exact problem is then a refined counting problem with two parameters, namely, size and additional characteristic: this is the subject addressed in this chapter and treated by a natural extension of the generating function framework. (The asymptotic problem can be viewed as one of characterizing in the limit a family of probability laws indexed by the values of the possible sizes: this is a topic to be discussed in Chapter IX.) As demonstrated here, the symbolic methods initially developed for counting combinatorial objects adapt gracefully to the analysis of various sorts of parameters of constructible classes, unlabelled and labelled alike.

[1]*"I never went far enough to get a good feel for the application of algebra to geometry. I was not pleased with this method of operating according to the rules without seeing what one does; solving geometrical problems by means of equations seemed like playing a tune by turning a crank."*

Multivariate generating functions (MGFs)—ordinary or exponential—can keep track of a collection of parameters defined over combinatorial objects. From the knowledge of such generating functions, there result either explicit probability distributions or, at least, mean and variance evaluations. For *inherited* parameters, all the combinatorial classes discussed so far are amenable to such a treatment and technically, the translation schemes that relate combinatorial constructions and multivariate generating functions present no major difficulty—they appear to be natural (notational, even) refinements of the paradigm developed in Chapters I and II for the univariate case. Typical applications from classical combinatorics are the number of summands in a composition, the number of blocks in a set partition, the number of cycles in a permutation, the root degree or path length of a tree, the number of fixed point in a permutation, the number of singleton blocks in a set partition, the number of leaves in trees of various sorts, and so on.

Beyond its technical aspects anchored in symbolic methods, this chapter also serves as a first encounter with the general area of random combinatorial structures. The general question is: *What does a random object of large size look like?* Multivariate generating functions first provide an easy access to *moments* of combinatorial parameters—typically the mean and variance. In addition, when combined with basic probabilistic inequalities, moment estimates often lead to precise characterizations of properties of large random structures that hold with high probability. For instance, a large integer partition conforms with high probability to a deterministic profile, a large random permutation almost surely has at least one long cycle and a few short ones, and so on. Such a highly constrained behaviour of large objects may in turn serve to design dedicated algorithms and optimize data structures; or it may serve to build statistical tests—when does one depart from randomness and detect a "signal" in large sets of observed data?. Randomness aspects form a recurrent theme of the book: they will be developed much further in Chapter IX, where complex-asymptotic methods of Part B are grafted on the exact modelling by multivariate generating functions presented in this chapter.

This chapter is organized as follows. First a few pragmatic developments related to bivariate generating functions, the multivariate paradigm specialized to two variables, are presented in Section III. 1. Section III. 2 then presents the notion of bivariate enumeration and its relation to discrete probabilistic models, including the determination of moments, as the language of elementary probability theory does provide an intuitively appealing way to conceive of bivariate counting data. The symbolic method *per se* declined in its general multivariate version is centrally developed in Sections III. 3 and III. 4: with suitable multi-index notations, the extension of the symbolic method to the multivariate case is almost immediate. Recursive parameters that often arise in particular from tree statistics form the subject of Section III. 5, while complete generating functions and associated combinatorial models are discussed in Section III. 6. Additional constructions like pointing, substitution, and order constraints lead to interesting developments, in particular, an original treatment of the inclusion-exclusion principle in Section III. 7. The chapter concludes with Section III. 8, which presents a brief abstract discussion of extremal parameters like height in trees or smallest and largest components in composite structures— such parameters are best treated via families of univariate generating functions.

III. 1. An introduction to bivariate generating functions (BGFs)

We have seen in Chapters I and II that a number sequence (f_n) can be encoded by means of a generating function in one variable, either ordinary or exponential:

$$(f_n) \qquad \rightsquigarrow \qquad f(z) = \begin{cases} \displaystyle\sum_n f_n z^n & \text{ordinary GF} \\ \displaystyle\sum_n f_n \frac{z^n}{n!} & \text{exponential GF.} \end{cases}$$

This encoding is powerful, since many combinatorial constructions admit of a translation as operations over such generating functions. In this way, one gains access to many useful counting formulæ.

Similarly, consider a sequence of numbers $(f_{n,k})$ depending on two integer valued indices, n and k. Usually, in this book, $(f_{n,k})$ will be an array of number (often a triangular array), where $f_{n,k}$ is the number of objects φ in some class $\mathcal{F}$, such that $|\varphi| = n$ and some parameter $\chi(\varphi)$ is equal to k. We can encode this sequence by means of a *bivariate generating function (BGF)*, which involves two variables, z attached to n and u attached to k.

DEFINITION III.1. *The* bivariate generating functions (BGFs)*, either of the ordinary or exponential type, of an array* $(f_{n,k})$ *are the formal power series* $f(z,u)$ *in two variables defined by*

$$(f_{n,k}) \qquad \rightsquigarrow \qquad f(z,u) = \begin{cases} \displaystyle\sum_{n,k} f_{n,k} z^n u^k & \textit{ordinary BGF} \\ \displaystyle\sum_{n,k} f_{n,k} \frac{z^n}{n!} u^k & \textit{exponential BGF.} \end{cases}$$

(The case of a "double exponential" GF corresponding to $\frac{z^n}{n!}\frac{u^k}{k!}$ is not used in the book.)

As we shall see shortly, many parameters of constructible classes become accessible through such BGFs. According to the point of view adopted momentarily here, one starts with an array of numbers and forms a BGF by a double summation process. We present here two examples related to binomial coefficients and Stirling cycle numbers illustrating how such BGFs can be be determined, then manipulated. In what follows it is convenient to refer to the *horizontal* and *vertical* generating functions that are each a one-parameter family of GFs in a single variable defined by

$$\begin{aligned} \text{horizontal GF:} \quad & f_n(u) := \sum_k f_{n,k} u^k; \\ \text{vertical GF:} \quad & f^{\langle k \rangle}(z) := \sum_n f_{n,k} z^n \quad \text{(ordinary case)} \\ & f^{\langle k \rangle}(z) := \sum_n f_{n,k} \frac{z^n}{n!} \quad \text{(exponential case).} \end{aligned}$$

$$
\begin{array}{|llllll}
\hline
f_{00} & & & \longrightarrow & f_0(u) \\
f_{10} & f_{11} & & \longrightarrow & f_1(u) \\
f_{20} & f_{21} & f_{22} & \longrightarrow & f_1(u) \\
\vdots & \vdots & \vdots & & \\
\downarrow & \downarrow & \downarrow & & \\
f^{\langle 0 \rangle}(z) & f^{\langle 1 \rangle}(z) & f^{\langle 2 \rangle}(z) & &
\end{array}
$$

FIGURE 1. An array of numbers and its associated horizontal and vertical GFs.

The terminology is transparently explained if the elements $(f_{n,k})$ are arranged as an infinite matrix, with $f_{n,k}$ placed in row n and column k, since the horizontal and vertical GFs appear as the GFs of the rows and columns respectively (Figure 1). Naturally, one has

$$f(z,u) = \sum_k u^k f^{\langle k \rangle}(z) = \begin{cases} \displaystyle\sum_n f_n(u) z^n & \text{ordinary BGF} \\ \displaystyle\sum_n f_n(u) \frac{z^n}{n!} & \text{exponential BGF.} \end{cases}$$

EXAMPLE 1. *The BGF of binomial coefficients.* The binomial coefficient $\binom{n}{k}$, counts the binary words of length n having k occurrences of a designated letter; see Figure 2. In order to compose the bivariate GF, start from the simplest case of Newton's binomial theorem and form directly the horizontal GFs corresponding to a fixed n:

$$W_n(u) := \sum_{k=0}^{n} \binom{n}{k} u^k = (1+u)^n, \tag{1}$$

Then a summation over all values of n gives the ordinary BGF

$$W(z,u) = \sum_{k,n \geq 0} \binom{n}{k} u^k z^n = \sum_{n \geq 0} (1+u)^n z^n = \frac{1}{1 - z(1+u)}. \tag{2}$$

Such calculations are typical of BGF manipulations. What we have done amounts to starting from a sequence of numbers, determining the horizontal GFs $W_n(u)$ in (1), then the bivariate GF $W(z,u)$ in (2), according to the scheme:

$$W_{n,k} \quad \rightsquigarrow \quad W_n(u) \quad \rightsquigarrow \quad W(z,u).$$

Observe that (2) reduces to the OGF $(1-2z)^{-1}$ of binary words, as it should, upon setting $u = 1$.

In addition, one can deduce from (2) the vertical GFs of the binomial coefficients corresponding to a fixed value of k,

$$W^{\langle k \rangle}(z) = \sum_{n \geq 0} \binom{n}{k} z^n = \frac{z^k}{(1-z)^{k+1}},$$

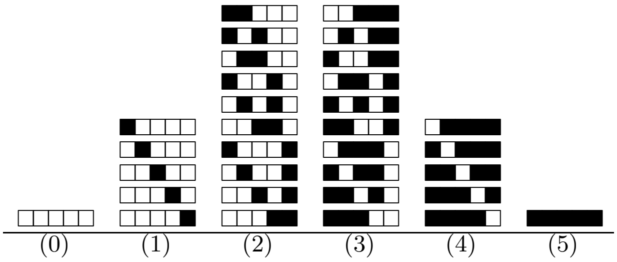


FIGURE 2. The set $\mathcal{W}_5$ of the 32 binary words over the alphabet $\{\square, \blacksquare\}$ enumerated according to the number of occurrences of the letter '$\blacksquare$' gives rise to the bivariate counting sequence $\{W_{5,j}\} = 1, 5, 10, 10, 5, 1$.

from an expansion of the BGF with respect to u,

$$W(z,u) = \frac{1}{1-z}\frac{1}{1-u\frac{z}{1-z}} = \sum_{k\geq 0} u^k \frac{z^k}{(1-z)^{k+1}}, \tag{3}$$

and the result naturally matches what a direct calculation would give. END OF EXAMPLE 1. $\square$

$\triangleright$ **1.** *The exponential BGF of binomial coefficients.* It is

$$\widetilde{W}(z,u) = \sum_{k,n} \binom{n}{k} u^k \frac{z^n}{n!} = \sum (1+u)^n \frac{z^n}{n!} = e^{z(1+u)}. \tag{4}$$

The vertical GFs are $e^z z^k/k!$. The horizontal GFs are $(1+u)^n$, like in the ordinary case. $\triangleleft$

EXAMPLE 2. *The BGF of Stirling cycle numbers.* As seen in Chapter II Example 12, the number of permutations of size n having k cycles is the Stirling cycle number $\left[{n \atop k}\right]$ with a vertical EGF being

$$P^{\langle k\rangle}(z) := \sum_n \begin{bmatrix} n \\ k \end{bmatrix} \frac{z^n}{n!} = \frac{L(z)^k}{k!}, \qquad L(z) := \log\frac{1}{1-z}.$$

From there, the exponential BGF is formed as follows (this revisits some of the calculations on p. 113):

$$\begin{aligned} P(z,u) &:= \sum_k P^{\langle k\rangle}(z) u^k = \sum_k \frac{u^k}{k!} L(z)^k = e^{uL(z)} \\ &= (1-z)^{-u}. \end{aligned} \tag{5}$$

The simplification is quite remarkable but altogether quite typical, as we shall see shortly, in the context of a labelled set construction. The starting point is thus a collection of vertical EGFs and the scheme is now

$$P_n^{\langle k\rangle} \quad \leadsto \quad P^{\langle k\rangle}(z) \quad \leadsto \quad P(z,u).$$

Observe that (5) reduces to the EGF of permutations at $u=1$.

Numbers	Horizontal GFs
$\binom{n}{k}$	$(1+u)^n$
Vertical OGFs $\frac{z^k}{(1-z)^{k+1}}$	Ordinary BGF $\frac{1}{1-z(1+u)}$

Numbers	Horizontal GFs
$\left[\begin{matrix} n \\ k \end{matrix}\right]$	$u(u+1)\cdots(u+n-1)$
Vertical EGFs $\frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k$	Exponential BGF $(1-z)^{-u}$

FIGURE 3. The various GFs associated to binomial coefficients (left) and Stirling cycle numbers (right).

In addition, an expansion of the BGF according to the variable z provides a useful information, namely, the horizontal GFs by virtue of Newton's binomial theorem:

$$\begin{aligned} P(z,u) &= \sum_{n\geq 0}\binom{n+u-1}{n}z^n = \sum_{n\geq 0} P_n(u)\frac{z^n}{n!} \\ \text{where}\quad P_n(u) &= u(u+1)\cdots(u+n-1). \end{aligned} \tag{6}$$

This last polynomial is called the *Stirling cycle polynomial* of index n and it describes completely the distribution of the number of cycles in all permutations of size n. In addition, note that the relation

$$P_n(u) = P_{n-1}(u)(u+(n-1)),$$

is equivalent to a recurrence

$$\left[\begin{matrix} n \\ k \end{matrix}\right] = (n-1)\left[\begin{matrix} n-1 \\ k \end{matrix}\right] + \left[\begin{matrix} n-1 \\ k-1 \end{matrix}\right],$$

by which Stirling numbers are often defined and easily evaluated numerically; see also APPENDIX A: *Stirling numbers*, p. 652. (The recurrence is susceptible to a direct combinatorial interpretation—add n either to an existing cycle or as a "new" singleton.) END OF EXAMPLE 2. □

Concise expressions for BGFs like (2), (3), (5), or (17) summarized in Figure 3 are precious for deriving moments, variance, and even finer characteristics of distributions, as we see next. The determination of such BGFs can be covered by a simple extension of the symbolic method along the lines of what was done in Chapters I and II, as detailed in Sections III. 3 and III. 4.

III. 2. Bivariate generating functions and probability distributions

Our purpose in this book is to analyse characteristics of combinatorial structures of very diverse types. We shall be principally interested in enumeration according to size *and* an auxiliary parameter, the corresponding problems being naturally treated by means of BGFs. In order to avoid redundant definitions, it proves convenient to introduce the sequence of *fundamental factors* $(\omega_n)_{n\geq 0}$, defined by

$$\omega_n = 1 \quad \text{for ordinary GFs}, \qquad \omega_n = n! \quad \text{for exponential GFs}. \tag{7}$$

Then, the OGF and EGF of a sequence (f_n) are jointly represented as

$$f(z) = \sum f_n \frac{z^n}{\omega_n} \qquad \text{and} \qquad f_n = \omega_n \, [z^n] f(z).$$

DEFINITION III.2. *Given a combinatorial class $\mathcal{A}$, a* (scalar) parameter *is a function from $\mathcal{A}$ to $\mathbb{Z}_{\geq 0}$ that associates to any object $\alpha \in \mathcal{A}$ an integer value $\chi(\alpha)$. The sequence*

$$A_{n,k} = \text{card}\left(\{\alpha \in \mathcal{A} \;\middle|\; |\alpha| = n, \; \chi(\alpha) = k\}\right),$$

is called the counting sequence *of the pair $\mathcal{A}, \chi$. The* bivariate generating function (BGF) *of $\mathcal{A}, \chi$ is defined as*

$$A(z, u) := \sum_{n,k \geq 0} A_{n,k} \frac{z^n}{\omega_n} u^k,$$

and is of ordinary type *if $\omega_n \equiv 1$ and of* exponential type *if $\omega_n \equiv n!$. One says that the variable z* marks size *and the variable u* marks the parameter χ.

Naturally $A(z, 1)$ reduces to the usual counting generating function $A(z)$ associated to $\mathcal{A}$, and the cardinality of $\mathcal{A}_n$ is expressible as

$$A_n = \omega_n [z^n] A(z, 1).$$

III. 2.1. Distributions and moments. As indicated in the introduction to this chapter, the eventual goal of multivariate enumeration is the quantification of properties present with high regularity in large random structures. Within this section, we discuss the relationship between probabilistic models needed to interpret bivariate counting sequences and bivariate generating functions. The elementary notions needed are recalled in APPENDIX A: *Combinatorial probability*, p. 644.

Consider a combinatorial class $\mathcal{A}$. The *uniform probability distribution* over $\mathcal{A}_n$ assigns to any $\alpha \in \mathcal{A}_n$ a probability equal to $1/A_n$. We shall use the symbol $\mathbb{P}$ to denote probability and occasionally subscript it with an indication of the probabilistic model used, whenever this model needs to be stressed: we shall then write $\mathbb{P}_{\mathcal{A}_n}$ (or simply $\mathbb{P}_n$ if $\mathcal{A}$ is understood) to indicate probability relative to the uniform distribution over $\mathcal{A}_n$.

Probability generating functions. Consider a parameter χ. It determines over each $\mathcal{A}_n$ a discrete *random variable* defined over the discrete probability space $\mathcal{A}_n$:

$$\mathbb{P}_{\mathcal{A}_n}\{\chi = k\} = \frac{A_{n,k}}{A_n} = \frac{A_{n,k}}{\sum_k A_{n,k}}. \tag{8}$$

Given a discrete random variable X, we recall that its its *probability generating function (PGF)* is the quantity

$$p(u) = \sum_k \mathbb{P}(X = k) u^k, \tag{9}$$

a generating function whose coefficients are probabilities. From (8) and (9), one has immediately:

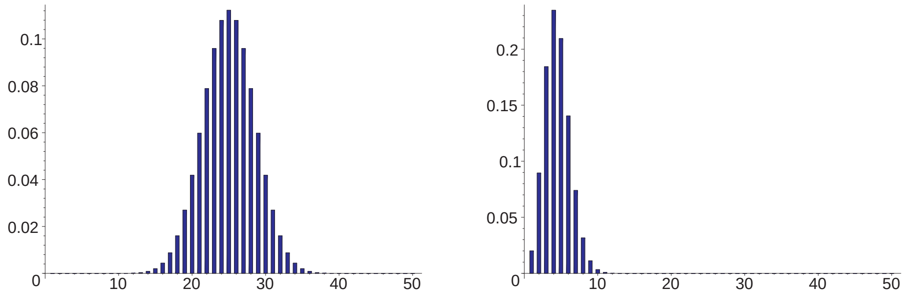


FIGURE 4. Histograms of two combinatorial distributions. Left: the number of occurrences of a designated letter in a random binary word of length 50 (binomial distribution). Right: the number of cycles in a random permutation of size 50 (Stirling cycle distribution).

PROPOSITION III.1 (PGFs from BGFs). *Let $A(z,u)$ be the bivariate generating function of a parameter χ defined over a combinatorial class $\mathcal{A}$. The probability generating function of χ over $\mathcal{A}_n$ is given by*

$$\sum_k \mathbb{P}_{\mathcal{A}_n}(\chi = k)u^k = \frac{[z^n]A(z,u)}{[z^n]A(z,1)},$$

and is thus a normalized version of a horizontal generating function.

The translation into the language of probability enables us to make use of whichever intuition might be available in any particular case, while allowing for a natural interpretation of data (Figure 4). Indeed, instead of noting that the quantity 381922055502195 represents the number of permutations of size 20 that have 10 cycles, it is perhaps more informative to state the probability of the event, which is 0.00015, i.e., about 1.5 per ten thousand. Discrete distributions are conveniently represented by *histograms* or "bar charts", where the height of the bar at abscissa k indicates the value of $\mathbb{P}\{X = k\}$. Figure 4 displays in this way two classical combinatorial distributions. Given the uniform probabilistic model that we have been adopting, such histograms are eventually nothing but a condensed form of the "stacks" corresponding to exhaustive listings, like the one displayed in Figure 2.

Moments. Important information is conveyed by *moments*. Given a discrete random variable X, the *expectation* of $f(X)$ is by definition the linear functional

$$\mathbb{E}(f(X)) := \sum_k \mathbb{P}\{X = k\} \cdot f(k).$$

The (power) *moments* are

$$\mathbb{E}(X^r) := \sum_k \mathbb{P}\{X = k\} \cdot k^r.$$

Then the expectation (or average, mean) of X, its variance, and its standard deviation are expressed as

$$\mathbb{E}(X), \qquad \mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2, \qquad \sigma(X) = \sqrt{\mathbb{V}(X)}.$$

The expectation corresponds to what is typically seen when forming the arithmetic mean value of a large number of observations: this property is the *weak law of large numbers* [**133**, Ch X]. The standard deviation then measures the dispersion of values observed from the expectation and it does so in a mean-quadratic sense.

The *factorial moment* defined for order r as

$$E\left(X(X-1)\cdots(X-r+1)\right)$$

is also of interest for computational purposes, since it is obtained plainly by differentiation of PGFs (APPENDIX A: *Combinatorial probability*, p. 644). Power moments are then easily recovered as linear combinations of factorial moments, see Note 7 of Appendix A. In summary:

PROPOSITION III.2 (Moments from BGFs). *The factorial moment of order r of a parameter χ is determined from the BGF $A(z,u)$ by r-fold differentiation followed by specialization at 1:*

$$\mathbb{E}_{\mathcal{A}_n}\left(\chi(\chi-1)\cdots(\chi-r+1)\right) = \frac{[z^n]\partial_u^r A(z,u)_{u=1}}{[z^n]A(z,1)}.$$

In particular, the first two moments satisfy

$$\mathbb{E}_{\mathcal{A}_n}(\chi) = \frac{[z^n]\partial_u A(z,1)}{[z^n]A(z,1)}, \qquad \mathbb{E}_{\mathcal{A}_n}(\chi^2) = \frac{[z^n]\partial_u^2 A(z,1)}{[z^n]A(z,1)} + \frac{[z^n]\partial_u A(z,1)}{[z^n]A(z,1)},$$

the variance and standard deviation being the determined by

$$\mathbb{V}(\chi) = \sigma(\chi)^2 = \mathbb{E}(\chi^2) - \mathbb{E}(\chi)^2.$$

PROOF. The PGF $p_n(u)$ of χ over $\mathcal{A}_n$ is given by Proposition III.1. On the other hand, factorial moments are on general grounds obtained from a PGF by differentiation and specialization at $u = 1$ (APPENDIX A: *Combinatorial probability*, p. 644). The result follows. □

In other words, the quantities

$$\Omega_n^{(k)} := \omega_n \cdot \left([z^n]\ \partial_u^k A(z,u)\Big|_{u=1}\right)$$

give, after a simple normalization (by $[z^n]A(z,1)$), the factorial moments.

$$\mathbb{E}\left(\chi(\chi-1)\cdots(\chi-k+1)\right) = \frac{1}{A_n}\Omega_n^{(k)}.$$

Most notably, $\Omega_n^{(1)}$ is the *cumulated value* of χ over all objects of $\mathcal{A}_n$:

$$\Omega_n^{(1)} \equiv \omega_n \cdot [z^n]\ \partial_u A(z,u)\big|_{u=1} = \sum_{\alpha\in\mathcal{A}_n} \chi(\alpha) \equiv A_n \cdot \mathbb{E}_{\mathcal{A}_n}(\chi).$$

Accordingly, the GF (ordinary or exponential) of the $\Omega_n^{(1)}$ is sometimes named the *cumulative* generating function. It can be viewed as an unnormalized generating function of the sequence of expected values. These considerations explain Wilf's suggestive motto quoted on p. 139:

"Generating functions find averages, etc."

The "*etc*" is to be interpreted as a token for higher moments.

▷ **2.** *A combinatorial form of cumulative GFs.* One has

$$\Omega^{(1)}(z) \equiv \sum_n \mathbb{E}_{\mathcal{A}_n}(\chi) A_n \frac{z^n}{\omega_n} = \sum_{\alpha \in \mathcal{A}} \chi(\alpha) \frac{z^{|\alpha|}}{\omega_{|\alpha|}},$$

where $\omega_n = 1$ (ordinary case) or $\omega_n = n!$ (exponential case). ◁

EXAMPLE 3. *Moments of the binomial distribution.* The binomial distribution of index n can be defined as the distribution of the number of a's in a random word of length n over the binary alphabet $\{a, b\}$. The determination of moments results easily from the ordinary BGF,

$$W(z, u) = \frac{1}{1 - z - zu}.$$

By differentiation, one finds

$$\left.\frac{\partial^r}{\partial u^r} W(z, u)\right| = \frac{r! z^r}{(1 - 2z)^r}.$$

Coefficient extraction then gives the form of the factorial moments of orders $1, 2, 3, \ldots, r$ as

$$\frac{n}{2}, \quad \frac{n(n-1)}{4}, \quad \frac{n(n-1)(n-2)}{8}, \ldots, \quad \frac{r!}{2^r}\binom{n}{r}.$$

In particular, the mean and the variance are $\frac{1}{2}n$ and $\frac{1}{4}n$. The standard deviation is thus $\frac{1}{2}\sqrt{n}$ which is of an order much smaller than the mean: this indicates that the distribution is somehow concentrated around its mean value, as suggested by Figure 4; see the next subsection for quantitative estimates. . . . . END OF EXAMPLE 3. □

▷ **3.** *De Moivre's approximation of the Gaussian coefficients.* The fact that the mean and the standard deviation of the binomial distribution are respectively $\frac{1}{2}n$ and $\frac{1}{2}\sqrt{n}$ suggests to examine what goes on at a distance of x standard deviations from the mean. Consider for simplicity the case of $n = 2\nu$ even. From the ratio

$$r(\nu, \ell) := \frac{\binom{2\nu}{\nu+\ell}}{\binom{2\nu}{\nu}} = \frac{(1 - \frac{1}{\nu})(1 - \frac{2}{\nu}) \cdots (1 - \frac{k-1}{\nu})}{(1 + \frac{1}{\nu})(1 + \frac{2}{\nu}) \cdots (1 + \frac{k}{\nu})}$$

an estimate of the logarithm shows that for any fixed $x \in \mathbb{R}$,

$$\lim_{n \to \infty,\ \ell = \nu + x\sqrt{\nu/2}} \frac{\binom{2\nu}{\nu+\ell}}{\binom{2\nu}{\nu}} = e^{-x^2/2}.$$

(Alternatively, Stirling's formula can be employed.) This Gaussian approximation for the binomial distribution was first discovered in 1733 by Abraham de Moivre (1667–1754), a close friend of Newton. Much more general methods for establishing such approximations form the subject of Chapter IX. ◁

EXAMPLE 4. *Moments of the Stirling cycle distribution.* Let us return to the example of cycles in permutations which is of interest in connection with certain sorting algorithms like bubble sort or insertion sort, maximum finding, and *in situ* rearrangement [**263**].

We are dealing with labelled objects, hence exponential generating functions. As seen earlier on p. 143, the BGF of permutations counted according to cycles is

$$P(z, u) = (1 - z)^{-u}.$$

We have $P_n = n!$, while $\omega_n = n!$ since the BGF is exponential. (The number of permutations of size n being $n!$, the combinatorial normalization happens to coincide with the factor of $1/n!$ present in all exponential generating functions.)

By differentiation of the BGF with respect to u, then setting $u = 1$, we next get the expected number of cycles in a random permutation of size n as a Taylor coefficient

$$\mathbb{E}_n(\chi) = [z^n]\frac{1}{1-z}\log\frac{1}{1-z} = 1 + \frac{1}{2} + \cdots + \frac{1}{n}, \tag{10}$$

which is the harmonic number H_n. Thus, on average, a random permutation of size n has about $\log n + \gamma$ cycles, a well known fact of discrete probability theory, derived on p. 113 by means of horizontal generating functions.

For the variance, a further differentiation of the bivariate EGF gives

$$\sum_{n\geq 0}\mathbb{E}_n(\chi(\chi-1))z^n = \frac{1}{1-z}\left(\log\frac{1}{1-z}\right)^2. \tag{11}$$

From this expression and Note 4 (or directly from the Stirling polynomials), a calculation shows that

$$\sigma_n^2 = \left(\sum_{k=1}^{n}\frac{1}{k}\right) - \left(\sum_{k=1}^{n}\frac{1}{k^2}\right) = \log n + \gamma - \frac{\pi^2}{6} + O\left(\frac{1}{n}\right). \tag{12}$$

Thus, asymptotically,

$$\sigma_n \sim \sqrt{\log n}.$$

The standard deviation is of an order smaller than the mean, and therefore deviations from the mean have an asymptotically negligible probability of occurrence (see below the discussion of moment inequalities). Furthermore, the distribution was proved to be asymptotically Gaussian by V. Gončarov, around 1942, see [**204**] and Chapter IX. . . . . END OF EXAMPLE 4. $\square$

$\triangleright$ **4.** *Stirling cycle numbers and harmonic numbers.* By the "exp-log trick" of Chapter I, the PGF of the Stirling cycle distribution satisfies

$$\frac{1}{n!}u(u+1)\cdots(u+n-1) = \exp\left(v\,\mathrm{H}_n - \frac{v^2}{2}\,\mathrm{H}_n^{(2)} + \frac{v^3}{3}\,\mathrm{H}_n^{(3)} + \cdots\right), \qquad u = 1+v$$

where $\mathrm{H}_n^{(r)}$ is the generalized harmonic number $\sum_{j=1}^{n} j^{-r}$. Consequently, any moment of the distribution is a polynomial in generalized harmonic numbers, cf (10) and (12). Also, the kth moment satisfies $\mathbb{E}_{\mathcal{P}_n}(\chi^k) \sim (\log n)^k$. (The same technique expresses the Stirling cycle number $\left[{n \atop k}\right]$ as a polynomial in generalized harmonic numbers $\mathrm{H}_{n-1}^{(r)}$.)

Alternatively, start from the expansion of $(1-z)^{-\alpha}$ and differentiate repeatedly with respect to α; for instance, one has

$$(1-z)^{-\alpha}\log\frac{1}{1-z} = \sum_{n\geq 0}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1} + \cdots + \frac{1}{n-1+\alpha}\right)\binom{n+\alpha-1}{n}z^n,$$

which provides (10) upon setting $\alpha = 1$, while the next differentiation gives access to (12). $\triangleleft$

The situation encountered with cycles in permutations is typical of iterative (non–recursive) structures. In many other cases, especially when dealing with recursive structures, the bivariate GF may satisfy complicated functional equations in two variables (see the example of path length in trees, Section III. 5 below) that do not make them available under an explicit form. Thus, exact expressions for the distributions are not always available, but asymptotic laws can be determined in a large number of cases (Chapter IX). In all cases, the BGFs are the central tool in obtaining mean and variance estimates, since their derivatives instantiated at $u = 1$ become univariate GFs that usually satisfy much simpler relations than the BGFs themselves.

III. 2.2. Moment inequalities and concentration of distributions. Qualitatively speaking, families of distributions can be classified into two categories: (i) distributions that at are spread, i.e., the standard deviation is of order at least as large as the mean (e.g.the uniform distributions over $[0\,..\,n]$, which have totally flat histograms, are spread); (ii) distributions such that the standard deviation is of an order smaller than the mean. Figure 4 illustrates the phenomena at stake and suggests that both the Stirling cycle distributions and the binomial distributions belong to the second category and are somehow concentrated around their mean value. Such informal observations are indeed supported by the Markov-Chebyshev inequalities, which take advantage of information provided by the first two moments. (A proof is found in APPENDIX A: *Combinatorial probability*, p. 644.)

> ***Markov-Chebyshev inequalities.*** *Let X be a nonnegative random variable and Y an arbitrary real variable. One has for an arbitrary $t > 0$:*
>
> $$\mathbb{P}\{X \geq t\mathbb{E}(X)\} \leq \frac{1}{t} \qquad \textit{(Markov inequality)}$$
>
> $$\mathbb{P}\{|Y - \mathbb{E}(Y)| \geq t\sigma(Y)\} \leq \frac{1}{t^2} \qquad \textit{(Chebyshev inequality)}.$$

This result informs us that the probability of being much larger than the mean must decay (Markov) and that an upperbound on the decay is measured in units given by the standard deviation (Chebyshev).

The next proposition formalizes a concentration property of distributions. It applies to a *family* of distributions indexed by the integers.

PROPOSITION III.3 (Concentration of distribution). *Consider a family of random variables X_n, typically, a scalar parameter χ on the subclass $\mathcal{A}_n$. Assume that the means $\mu_n = \mathbb{E}(X_n)$ and the standard deviations $\sigma_n = \sigma(X_n)$ satisfy the condition*

$$\lim_{n\to+\infty} \frac{\sigma_n}{\mu_n} = 0.$$

Then the distribution of X_n is concentrated *in the sense that, for any $\epsilon > 0$, there holds*

$$\lim_{n\to+\infty} \mathbb{P}\left\{1 - \epsilon \leq \frac{X_n}{\mu_n} \leq 1 + \epsilon\right\} = 1. \tag{13}$$

PROOF. It is a direct consequence of Chebyshev's inequality. □

The concentration property (13) expresses the fact that values of X_n tend to become closer and closer (in relative terms) to the mean μ_n as n increases. Another figurative way to describe concentration, much used in random combinatorics, is by saying that "X_n/μ_n *tends to 1 in probability*". When this property is satisfied, the expected value is in a strong sense a typical value. This fact is an extension of the *weak law of large numbers* of probability theory. In that field, the concentration property (13) is

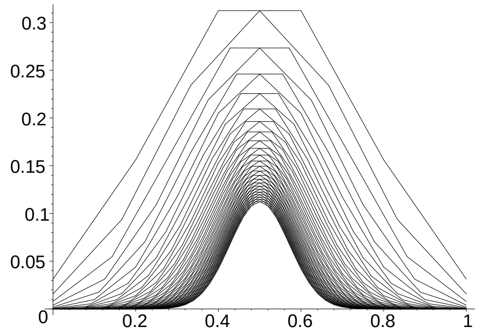


FIGURE 5. Plots of the binomial distributions for $n = 5, \ldots, 50$. The horizontal axis is normalized (by a factor of $1/n$) and rescaled to 1, so that the curves display $\left\{\mathbb{P}(\frac{X_n}{n} = x)\right\}$, for $x = 0, \frac{1}{n}, \frac{2}{n}, \ldots$.

also known as *convergence in probability* and is then written more concisely:

$$\frac{X_n}{\mu_n} \xrightarrow{P} 1.$$

Concentration properties of the binomial and Stirling cycle distributions. The binomial distribution *is* concentrated, since the mean of the distribution is $n/2$ and the standard deviation is $\sqrt{n/4}$, a much smaller quantity. Figure 5 illustrates concentration by displaying the graphs (as polygonal lines) associated to the binomial distributions for $n = 5, \ldots, 50$. Concentration is also quite perceptible on simulations as n gets large: the table below describes the results of batches of ten (sorted) simulations from the binomial distribution $\left\{\frac{1}{2^n}\binom{n}{k}\right\}_{k=0}^{n}$:

$n = 100$	$39, 42, 43, 49, 50, 52, 54, 55, 55, 57$
$n = 1000$	$487, 492, 494, 494, 506, 508, 512, 516, 527, 545$
$n = 10,000$	$4972, 4988, 5000, 5004, 5012, 5017, 5023, 5025, 5034, 5065$
$n = 100,000$	$49798, 49873, 49968, 49980, 49999, 50017, 50029, 50080, 50101, 50284$;

the maximal deviations from the mean observed on such samples are 22% ($n = 10^2$), 9% ($n = 10^3$), 1.3% ($n = 10^4$), and 0.6% ($n = 10^5$).

Similarly, the mean and variance computations of (10) and (12) imply that the number of cycles in a random permutation of large size is concentrated.

Finer estimates on distributions form the subject of our Chapter IX dedicated to limit laws. The reader may get a feeling of some of the phenomena at stake when re-examining Figure 5: the visible emergence of a continuous curve (the bell shaped curve) corresponds to a common asymptotic shape for the whole family of distributions—the Gaussian law.

III. 3. Inherited parameters and ordinary MGFs

We have seen so far basic manipulations of BGFs (Section III. 1) as well as their use in order to determine moments of combinatorial distributions (Section III. 2). In this section and its labelled counterpart, Section III. 4, we address the question of determining directly BGFs from combinatorial specifications. The answer is provided by a simple extension of the symbolic method, which is formulated in terms of *multivariate generating functions* (MGFs). Such generating functions have the capability of taking into account a finite collection (equivalently, a vector) of combinatorial parameters. On the one hand, the theory specializes immediately to BGFs, which correspond to the particular case of a single (scalar) parameter. On the other hand, it provides "complete" (multivariate) generating functions discussed in Section III. 6.

III. 3.1. Multivariate generating functions (MGFs). The theory is best developed in full generality for the joint analysis of a fixed finite collection of parameters.

DEFINITION III.3. *Consider a combinatorial class $\mathcal{A}$. A* (multidimensional) parameter $\chi = (\chi_1, \ldots, \chi_d)$ *on the class is a function from $\mathcal{A}$ to the set $\mathbb{Z}^d_{\geq 0}$ of d-tuples of natural numbers. The* counting sequence *of $\mathcal{A}$ with respect to size and the parameter χ is then defined by*

$$A_{n,k_1,\ldots,k_d} = \operatorname{card}\left\{\alpha \;\middle|\; |\alpha| = n,\ \chi_1(\alpha) = k_1, \ldots, \chi_d(\alpha) = k_d\right\}.$$

We sometimes refer to such a parameter as a "multiparameter" when $d > 1$, and a "simple" or "scalar" parameter otherwise. For instance, one may take the class $\mathcal{P}$ of all permutations σ, and for χ_j ($j = 1, 2, 3$) the number of cycles of length j in σ. Alternatively, we may consider the class $\mathcal{W}$ of all words w over an alphabet with four letters, $\{\alpha_1, \ldots, \alpha_4\}$ and take for χ_j ($j = 1, \ldots, 4$) the number of occurrences of the letter α_j in w, and so on.

The *multi-index convention* employed in various branches of mathematics greatly simplifies notations: let $\mathbf{u} = (u_1, \ldots, u_d)$ be a vector of d formal variables and $\mathbf{k} = (k_1, \ldots, k_d)$ be a vector of integers of the same dimension; then, the multi-power $\mathbf{u}^{\mathbf{k}}$ is defined as the monomial

$$\mathbf{u}^{\mathbf{k}} := u_1^{k_1} u_2^{k_2} \cdots u_d^{k_d}. \tag{14}$$

With this notation, we have:

DEFINITION III.4. *Let $A_{n,\mathbf{k}}$ be a multi-index sequence of numbers, where $\mathbf{k} \in \mathbb{N}^d$. The* multivariate generating function (MGF) *of the sequence of either ordinary or exponential type is defined as the formal power series*

$$\begin{aligned} A(z, \mathbf{u}) &= \sum_{n,k} A_{n,\mathbf{k}} \mathbf{u}^{\mathbf{k}} z^n \quad \textit{(ordinary MGF)} \\ A(z, \mathbf{u}) &= \sum_{n,k} A_{n,\mathbf{k}} \mathbf{u}^{\mathbf{k}} \frac{z^n}{n!} \quad \textit{(exponential MGF).} \end{aligned} \tag{15}$$

Given a class $\mathcal{A}$ and a parameter χ, the multivariate generating function (MGF) of the pair $\langle \mathcal{A}, \chi \rangle$ is the MGF of the corresponding counting sequence. In particular, one has the combinatorial forms

$$\begin{aligned} A(z, \mathbf{u}) &= \sum_{\alpha \in \mathcal{A}} \mathbf{u}^{\chi(\alpha)} z^{|\alpha|} \quad \textit{(ordinary MGF; unlabelled case)} \\ A(z, \mathbf{u}) &= \sum_{\alpha \in \mathcal{A}} \mathbf{u}^{\chi(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!} \quad \textit{(exponential MGF; labelled case).} \end{aligned} \tag{16}$$

One also says that $A(z, \mathbf{u})$ is the MGF of the combinatorial class with the formal variable u_j marking *the parameter χ_j and z* marking *size.*

From the very definition, $A(z, \mathbf{1})$ (with $\mathbf{1}$ a vector of all 1's) coincides with the counting generating function of $\mathcal{A}$, either ordinary or exponential as the case may be. One can then view an MGF as a deformation of a univariate GF by way of a (vector) parameter $\mathbf{u}$, with the property for the multivariate GF to reduce to the univariate counting GF at $\mathbf{u} = \mathbf{1}$. If all but one of the u_j are set to 1, then a BGF results. Thus, the symbolic calculus that we are going to develop opens full access to BGFs and hence moments. In fact, it has the capacity of determining the *joint probability distribution* of a finite collection of parameters.

$\triangleright$ **5.** *Specializations of MGFs.* The exponential MGF of permutations with u_1, u_2 marking the number of 1-cycles and 2-cycles respectively turns out to be

$$P(z, u_1, u_2) = \frac{\exp\left((u_1 - 1)z + (u_2 - 1)\frac{z^2}{2}\right)}{1 - z}. \tag{17}$$

(This is to be proved later in this chapter, p. 176.) The formula is checked to be consistent with three already known specializations derived in Chapter II: (i) setting $u_1 = u_2 = 1$ gives back the counting of *all* permutations, $P(z, 1, 1) = (1 - z)^{-1}$, as it should; (ii) setting $u_1 = 0$ and $u_2 = 1$ gives back the EGF of derangements, namely $e^{-z}/(1 - z)$; (iii) setting $u_1 = u_2 = 0$ gives back the EGF of permutations with cycles all of length greater than 2, $P(z, 0, 0) = e^{-z-z^2/2}/(1 - z)$, a generalized derangement GF. In addition, the specialized BGF

$$P(z, u, 1) = \frac{e^{(u-1)z}}{1 - z},$$

enumerates permutations according to singleton cycles. This last BGF interpolates between the EGF of derangements $(u = 0)$ and the EGF of all permutations $(u = 1)$. $\triangleleft$

III. 3.2. Inheritance and MGFs. Parameters that are *inherited* from substructures can be taken into account by a direct extension of the symbolic method. With a suitable use of the multi-index conventions, it is even the case that the translation rules previously established in Chapters I and II can be copied verbatim. This approach opens the way to a large quantity of multivariate enumeration results that then follow automatically by the symbolic method.

Let us consider a pair $\langle \mathcal{A}, \chi \rangle$, where A is a combinatorial class endowed with its usual size function $|\cdot|$ and $\chi = (\chi_1, \ldots, \chi_d)$ is a d-dimensional (multi)parameter. Write χ_0 for size and z_0 for the variable marking size (previously denoted by z). The key point for theoretical developments is to define an extended multiparameter $\overline{\chi} = (\chi_0, \chi_1, \ldots, \chi_d)$, that is, we treat size and parameters on an equal basis. Then the ordinary MGF in (15) assumes an extremely simple and symmetrical form:

$$\begin{aligned} A(\mathbf{z}) &= \sum_{\mathbf{k}} A_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} \\ &= \sum_{\alpha \in \mathcal{A}} \mathbf{z}^{\overline{\chi}(\alpha)}. \end{aligned} \tag{18}$$

There, the indeterminates are the vector $\mathbf{z} = (z_0, z_1, \ldots, z_d)$, the indices are $\mathbf{k} = (k_0, k_1, \ldots, k_d)$ (where k_0 indexes size, previously denoted by n), and the usual multi-index convention introduced in (14) is in force,

$$\mathbf{z}^{\mathbf{k}} := z_0^{k_0} z_1^{k_1} \cdots {z_d}^{k_d}, \tag{19}$$

but it is now applied to $(d+1)$-dimensional vectors.

Next, we define inherited parameters.

DEFINITION III.5. *Let $\langle \mathcal{A}, \chi \rangle$, $\langle \mathcal{B}, \xi \rangle$, $\langle \mathcal{C}, \zeta \rangle$ be three combinatorial classes endowed with parameters of the same dimension d. The parameter χ is said to be* inherited *in the following cases:*

- *Disjoint union: when $\mathcal{A} = \mathcal{B} + \mathcal{C}$, the parameter χ is inherited from ξ, ζ iff its value is determined by cases from ξ, ζ:*

$$\chi(\omega) = \begin{cases} \xi(\omega) & \text{if } \omega \in \mathcal{B} \\ \zeta(\omega) & \text{if } \omega \in \mathcal{C}. \end{cases}$$

- *Cartesian product: when $\mathcal{A} = \mathcal{B} \times \mathcal{C}$, the parameter χ is inherited from ξ, ζ iff its value is obtained additively from the values of ξ, ζ:*

$$\chi(\langle \beta, \gamma \rangle) = \xi(\beta) + \zeta(\gamma).$$

- *Composite constructions: when $\mathcal{A} = \mathfrak{K}\{B\}$, where $\mathfrak{K}$ is a metasymbol representing any of* SEQ, MSET PSET, CYC, *the parameter χ is inherited from ξ iff its value is obtained additively from the values of ξ on components; for instance, for sequences:*

$$\chi([\beta_1, \ldots, \beta_r]) = \xi(\beta_1) + \cdots + \xi(\beta_r).$$

With a natural extension of the notation used for constructions, one shall write

$$\langle \mathcal{A}, \chi \rangle = \langle \mathcal{B}, \xi \rangle + \langle \mathcal{C}, \zeta \rangle, \quad \langle \mathcal{A}, \chi \rangle = \langle \mathcal{B}, \xi \rangle \times \langle \mathcal{C}, \zeta \rangle, \quad \langle \mathcal{A}, \chi \rangle = \mathfrak{K}\left\{\langle \mathcal{B}, \xi \rangle\right\}.$$

This definition of inheritance is seen to be a natural extension of the axioms that size itself has to satisfy (Chapter I): size of a disjoint union is defined by cases, while size of a pair, and similarly of a composite construction, is obtained by addition.

THEOREM III.1 (Inherited parameters and ordinary MGFs). *Let $\mathcal{A}$ be a combinatorial class constructed from $\mathcal{B}, \mathcal{C}$, and let χ be a parameter inherited from ξ defined on $\mathcal{B}$ and (as the case may be) from ζ on $\mathcal{C}$. Then the translation rules of admissible constructions stated in Theorem I.1 apply provided the multi-index convention* (18) *is used. The associated operators on ordinary MGFs are then:*

$$\begin{array}{llcl}
\textit{Union:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) + C(\mathbf{z}) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \times \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) \cdot C(\mathbf{z}) \\
\textit{Sequence:} & \mathcal{A} = \text{S}\textsc{eq}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \dfrac{1}{1 - B(\mathbf{z})} \\
\textit{Cycle:} & \mathcal{A} = \text{C}\textsc{yc}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \displaystyle\sum_{\ell=1}^{\infty} \frac{\varphi(\ell)}{\ell} \log \frac{1}{1 - B(\mathbf{z}^\ell)}. \\
\textit{Multiset:} & \mathcal{A} = \text{MS}\textsc{et}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \exp\left(\displaystyle\sum_{\ell=1}^{\infty} \frac{1}{\ell} B(\mathbf{z}^\ell)\right) \\
\textit{Powerset:} & \mathcal{A} = \text{PS}\textsc{et}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \exp\left(\displaystyle\sum_{\ell=1}^{\infty} \frac{(-1)^{\ell-1}}{\ell} B(\mathbf{z}^\ell)\right)
\end{array}$$

PROOF. The verification for sums and products is immediate, given the combinatorial forms of OGFs. For disjoint unions, one has

$$A(\mathbf{z}) = \sum_{\alpha \in \mathcal{A}} \mathbf{z}^{\overline{\chi}(\alpha)} = \sum_{\beta \in \mathcal{B}} \mathbf{z}^{\overline{\xi}(\beta)} + \sum_{\gamma \in \mathcal{C}} \mathbf{z}^{\overline{\zeta}(\gamma)},$$

as results from the fact that inheritance is defined by cases on unions. For cartesian products, one has

$$A(\mathbf{z}) = \sum_{\alpha \in \mathcal{A}} \mathbf{z}^{\overline{\chi}(\alpha)} = \sum_{\beta \in \mathcal{B}} \mathbf{z}^{\overline{\xi}(\beta)} \times \sum_{\gamma \in \mathcal{C}} \mathbf{z}^{\overline{\zeta}(\gamma)},$$

as results from the fact that inheritance is defined additively on products.

The translation of composite constructions in the case of sequences, powersets, and multisets are then built up from the union and product schemes, in exactly the same manner as in the proof of Theorem I.1. Cycles are dealt with by the methods of APPENDIX A: *Cycle construction*, p. 646. □

This theorem is a straightforward extension of the symbolic method, but it is important because it can be applied in a wide range of combinatorial applications. The reader is especially encouraged to study carefully the treatment of integer compositions below, as it illustrates in its bare bones version the power of the symbolic method for taking into account combinatorial parameters.

The multi-index notation is a crucial ingredient for developing the general theory of multivariate enumerations. However, in most cases, we work with only a small number of parameters, typically one or two. In such cases, we often use vectors of variables like (z, u) or (z, u, v), the corresponding monomials being then written as $z^n u^k$ or $z^n u^k v^\ell$. This has the advantage of avoiding unnecessary subscripts.

Integer compositions and marks. The class $\mathcal{C}$ of all integer compositions (Chapter I) is specified by

$$\mathcal{C} = \text{S}\textsc{eq}(\mathcal{I}), \qquad \mathcal{I} = \text{S}\textsc{et}_{\geq 1}(\mathcal{Z}),$$

where $\mathcal{I}$ is the set of all positive numbers. The corresponding OGFS are

$$I(z) = \frac{1}{1 - I(z)}, \qquad I(z) = \frac{z}{1 - z},$$

so that $C_n = 2^{n-1}$ $(n \geq 1)$. Say we want to enumerate compositions according to the number χ of summands. One way to proceed, in accordance with the formal definition of inheritance, is as follows. Let ξ be the parameter that takes the constant value 1 on all elements of $\mathcal{I}$. The parameter χ on compositions is inherited from the (almost trivial) parameter $\xi \equiv 1$ defined on summands. The ordinary MGF of $\langle \mathcal{I}, \xi \rangle$ is obviously

$$I(z,u) = zu + z^2u + z^3u + \cdots = \frac{zu}{1-z}.$$

Let $C(z,u)$ be the BGF of $\langle \mathcal{C}, \chi \rangle$. By Theorem III.1, the schemes translating admissible constructions in the univariate case carry over to the multivariate case, so that

$$(20) \qquad C(z,u) = \frac{1}{1 - I(z,u)} = \frac{1}{1 - u\frac{z}{1-z}} = \frac{1-z}{1-z(u+1)}.$$

Et voila!

Here is an alternative way of arriving at (20), which is important and is of much use in the sequel. One may regard the enumeration of compositions with respect to the number of summands as the enumeration of compositions with respect to both size (i.e., number of atoms) and number of *marks*, where each summand carries a mark, say 'μ', which is an object of size 0. The number of marks is clearly *inherited* from summands to compositions. Then, one has an enriched specification, and its translation into MGFs,

$$(21) \qquad \mathcal{C} = \text{SEQ}\left(\mu\, \text{SEQ}_{\geq 1}(\mathcal{Z})\right) \qquad \Longrightarrow \qquad C(z,u) = \frac{1}{1-uI(z)}.$$

as granted by Theorem III.1 and based on the correspondence: $\mathcal{Z} \mapsto z$, $\mu \mapsto u$. This notion of mark when used in conjunction with Theorem III.1 provides access to many joint parameters, as shown in Example 5 below.

EXAMPLE 5. *Summands in integer compositions.* Consider the double parameter $\chi = (\chi_1, \chi_2)$ where χ_1 is the number of parts equal to 1 and χ_2 the number of parts equal to 2. One can write down an extended specification, with μ_1 a combinatorial mark for summands equal to 1 and μ_2 for summands equal to 2,

$$(22) \qquad \begin{aligned} &\mathcal{C} = \text{SEQ}\left(\mu_1 \mathcal{Z} + \mu_2 \mathcal{Z}^2 + \text{SEQ}_{\geq 3}(\mathcal{Z})\right) \\ &\Longrightarrow \qquad C(z,u_1,u_2) = \frac{1}{1-(u_1z + u_2z^2 + z^3(1-z)^{-1})}, \end{aligned}$$

where u_j $(j = 1, 2)$ records the number of marks of type μ_j.

Similarly, let μ mark each summand and μ_1 mark summands equal to 1. Then, one has,

$$(23) \quad \mathcal{C} = \text{SEQ}\left(\mu\mu_1\mathcal{Z} + \mu\, \text{SEQ}_{\geq 2}(\mathcal{Z})\right) \implies C(z,u_1,u) = \frac{1}{1-(uu_1z + uz^2(1-z)^{-1})},$$

where u keeps track of the total number of summands and u_1 records the number of summands equal to 1.

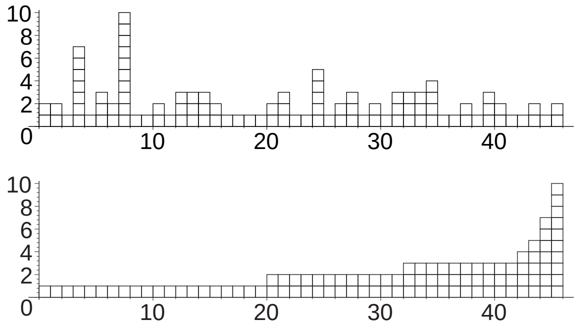


FIGURE 6. A random composition of $n = 100$ represented as a ragged landscape (top); its associated profile $1^{20}2^{12}3^{10}4^{1}5^{1}7^{1}10^{1}$, defined as the partition obtained by sorting the summands (bottom).

MGFs obtained in this way via the multivariate extension of the symbolic method can then provide explicit counts, after suitable series expansions. For instance, the number of compositions of n with k parts is, by (20),

$$[z^n u^k]\frac{1-z}{1-(1+u)z} = \binom{n}{k} - \binom{n-1}{k} = \binom{n-1}{k-1},$$

a result otherwise obtained in Chapter I by direct combinatorial reasoning (the balls-and-bars model). The number of compositions of n containing k parts equal to 1 is obtained from the special case $u_2 = 1$ in (22),

$$[z^n u^k]\frac{1}{1-uz-\frac{z^2}{(1-z)}} = [z^{n-k}]\frac{(1-z)^{k+1}}{(1-z-z^2)^{k+1}},$$

where the last OGF closely resembles a power of the OGF of Fibonacci numbers.

Following the discussion of Section III. 2, such MGFs also carry complete information on moments. In particular, the cumulated value of the number of parts in all compositions of n has OGF

$$\partial_u C(z,u)\big|_{u=1} = \frac{z(1-z)}{(1-2z)^2},$$

as seen from Section III. 2.1, since cumulated values are obtained via differentiation of a BGF. Therefore, the expected number of parts in a random composition of n is exactly ($n \geq 1$)

$$\frac{1}{2^{n-1}}\,[z^n]\frac{z(1-z)}{(1-2z)^2} = \frac{1}{2}(n+1).$$

A further differentiation will give access to the variance. The standard deviation is found to be $\frac{1}{2}\sqrt{n-1}$, which is of an order (much) smaller than the mean. Thus, the distribution of the number of summands in a random composition satisfies the concentration property as $n \to \infty$.

In the same vein, the number of parts equal to a fixed number r in compositions is determined by

$$\mathcal{C} = \text{SEQ}\left(\mu\mathcal{Z}^r + \text{SEQ}_{\neq r}(\mathcal{Z})\right) \quad\Longrightarrow\quad C(z,u) = \left(1 - \left(\frac{z}{1-z} + (u-1)z^r\right)\right)^{-1}.$$

It is then easy to pull out the expected number of r-summands in a random composition of size n. The differentiated form

$$\partial_u C(z,u)\big|_{u=1} = \frac{z^r(1-z)^2}{(1-2z)^2}$$

gives by partial fraction expansion

$$\partial_u \widehat{C}(z,u)\Big|_{u=1} = \frac{2^{-r-2}}{(1-2z)^2} + \frac{2^{-r-1} - r2^{-r-2}}{1-2z} + q(z),$$

for a polynomial $q(z)$ that we do not need to make explicit. Extracting the nth coefficient of the cumulative GF $C'_u(z,1)$ and dividing by 2^{n-1} yields the mean number of r-parts in a random composition. Another differentiation gives access to the second moment. One finds:

PROPOSITION III.4 (Summands in integer compositions). *The total number of summands in a random composition of size n has mean $\frac{1}{2}(n+1)$ and a distribution that is concentrated around the mean. The number of r summands in a composition of size n has mean*

$$\frac{n}{2^{r+1}} + O(1);$$

and a standard deviation of order $\sqrt{n}$, which also ensures concentration of distribution.

Clearly, suitable MGFs can keep track of any finite collection of summand types in compositions, and the method is extremely general. Much use of this way of envisioning multivariate enumeration will be made throughout this book. .... END OF EXAMPLE 5. □

From the point of view of random structures, the example of summands shows that random compositions of large size tend to conform to a global "profile". With high probability, a composition of size n should have about $n/4$ parts equal to 1, $n/8$ parts equal to 2, and so on. Naturally, there are statistically unavoidable fluctuations, and for any finite n, the regularity of this law cannot be perfect: it tends to fade away especially as regards to largest summands that are $\log_2(n) + O(1)$ with high probability. (In this region mean and standard deviation both become of the same order and are $O(1)$, so that concentration no longer holds.) However, such observations *do* tell us a great deal about what a typical random composition must (probably) look like—it should conform to a "logarithmic profile",

$$1^{n/4}\, 2^{n/8}\, 3^{n/16}\, 4^{n/32} \, \cdots \, .$$

Here are for instance the profiles of two compositions of size $n = 1024$ drawn uniformly at random:

$$1^{250}\, 2^{138}\, 3^{70}\, 4^{29}\, 5^{15}\, 6^{10}\, 7^4\, 8^0, 9^1, \qquad 1^{253}\, 2^{136}\, 3^{68}\, 4^{31}\, 5^{13}\, 6^8\, 7^3\, 8^1\, 9^1\, 10^2$$

to be compared to the "ideal" profile

$$1^{256}\, 2^{128}\, 3^{64}\, 4^{32}\, 5^{16}\, 6^8\, 7^4\, 8^2\, 9^1.$$

It is a striking fact that samples of a very few elements or even just *one* element (this would be ridiculous by the usual standards of statistics) are often sufficient to illustrate asymptotic properties of large random structures. The reason is once more to be attributed to concentration of distributions whose effect is manifest here. Profiles of a similar nature present themselves amongst objects defined by the sequence construction, as we shall see throughout this book. (Establishing such general laws is often not difficult but it requires the full power of complex-analytic methods developed in Chapters IV–VIII.)

$\triangleright$ **6.** *Largest summands in compositions.* For any $\epsilon > 0$, with probability tending to 1 as $n \to \infty$, the largest summand in a random integer composition of size n is almost surely of size in the interval $[(1-\epsilon)\log_2 n, (1+\epsilon)\log_2 n]$. (Hint: use the first and second moment methods. More precise estimates are given in Chapter V.) $\triangleleft$

In the sequel, it proves convenient to adopt a simplifying notation, much in the spirit of our basic convention, where the atom $\mathcal{Z}$ is systematically reflected by the name z of the variable in GFs.

> ***Simplified notation for marks.*** *The same symbol (usually $u, v, u_1, u_2 \ldots$) is freely employed to designate a combinatorial mark (of size 0) and the corresponding marking variable in MGFs.*

For instance, we allow ourselves to write directly, for compositions,

$$\mathcal{C} = \mathrm{SEQ}(u\,\mathrm{SEQ}_{\geq 1}\,\mathcal{Z})), \qquad \mathcal{C} = \mathrm{SEQ}(uu_1\mathcal{Z} + u\,\mathrm{SEQ}_{\geq 2}\,\mathcal{Z})),$$

where u marks all summands and u_1 marks summands equal to 1, giving rise to (21) and (23). Note that the symbolic scheme of Theorem III.1 invariably applies to enumeration according to the number of zero-size marks inserted into specifications.

III. 3.3. Number of components in abstract unlabelled schemas. Consider a construction $\mathcal{A} = \mathfrak{K}(\mathcal{B})$, where the metasymbol $\mathfrak{K}$ designates any standard *unlabelled* constructor amongst $\mathrm{SEQ}, \mathrm{MSET}, \mathrm{PSET}, \mathrm{CYC}$. What is sought is the BGF $A(z,u)$ of class $\mathcal{A}$, with u marking each component. The specification is then of the form

$$\mathcal{A} = \mathfrak{K}(u\mathcal{B}), \qquad \mathfrak{K} = \mathrm{SEQ}, \mathrm{MSET}, \mathrm{PSET}, \mathrm{CYC}\,.$$

Theorem III.1 applies and yields immediately the BGF $A(z,u)$. In addition, differentiating with respect to u then setting $u = 1$ provides the GF of cumulated values (hence, in a non-normalized form, the OGF of the sequence of mean values of the number of components):

$$\Omega(z) = \left.\frac{\partial}{\partial u} A(z,u)\right|_{u=1}.$$

In summary:

PROPOSITION III.5 (Components in unlabelled schemas). *Given a construction, $\mathcal{A} = \mathfrak{K}(\mathcal{B})$, the BGF $A(z,u)$ and the cumulated GF $\Omega(z)$ associated to the number of components are given by the following table:*

$\mathfrak{K}$	*BGF* $(A(z,u))$	*Cumulative OGF* $(\Omega(z))$
SEQ :	$\dfrac{1}{1-uB(z)}$	$A(z)^2 \cdot B(z) = \dfrac{B(z)}{(1-B(z))^2}$
PSET :	$\begin{cases} \exp\left(\sum_{k=1}^{\infty}(-1)^{k-1}\frac{u^k}{k}B(z^k)\right) \\ \prod_{n=1}^{\infty}(1+uz^n)^{B_n} \end{cases}$	$A(z)\cdot\sum_{k=1}^{\infty}(-1)^{k-1}B(z^k)$
MSET :	$\begin{cases} \exp\left(\sum_{k=1}^{\infty}\frac{u^k}{k}B(z^k)\right) \\ \prod_{n=1}^{\infty}(1-uz^n)^{-B_n} \end{cases}$	$A(z)\cdot\sum_{k=1}^{\infty}B(z^k)$
CYC :	$\sum_{k=1}^{\infty}\frac{\varphi(k)}{k}\log\frac{1}{1-u^kB(z^k)}$	$\sum_{k=1}^{\infty}\varphi(k)\frac{B(z^k)}{1-B(z^k)}.$

Mean values are then recovered with the usual formula,

$$\mathbb{E}_{\mathcal{A}_n}(\text{\# components}) = \frac{[z^n]\Omega(z)}{[z^n]A(z)}.$$

A similar process applies to the number of components of a fixed size r in an $\mathcal{A}$-object.

$\triangleright$ **7.** *r-Components in abstract unlabelled schemas.* Consider unlabelled structures. The BGF of the number of r-components in $\mathcal{A} = \mathfrak{K}\{\mathcal{B}\}$ is given by

$$A(z,u) = (1 - B(z) - (u-1)B_r z^r)^{-1}, \qquad A(z,u) = A(z)\cdot\left(\frac{1-z^r}{1-uz^r}\right)^{B_r},$$

in the case of sequences ($\mathfrak{K} = \text{SEQ}$) and multisets ($\mathfrak{K} = \text{MSET}$), respectively. Similar formulæ hold for the other basic constructions and the cumulative GFs. $\triangleleft$

$\triangleright$ **8.** *Number of distinct components in a multiset.* The specification and the BGF are

$$\prod_{n\geq 1}\left(\mathbf{1} + u\,\text{SEQ}_{\geq 1}(\mathcal{Z})\right) \Longrightarrow \qquad \Longrightarrow \prod_{n\geq 1}\left(1 + \frac{uz^n}{1-z^n}\right)^{B_n},$$

as follows from first principles. $\triangleleft$

As an illustration, we discuss the profile of random partitions (Figure 7).

EXAMPLE 6. *The profile of partitions.* Let $\mathcal{P} = \text{MSET}(\mathcal{I})$ be the class of all integer partitions, where $\mathcal{I} = \text{SEQ}_{\geq 1}(\mathcal{Z})$ represents integers in unary notation. The BGF of $\mathcal{P}$ with u marking the number χ of parts (or summands) is obtained from the specification

$$\mathcal{P} = \text{MSET}(u\mathcal{I}) \quad\Longrightarrow\quad P(z,u) = \exp\left(\sum_{k=1}^{\infty}\frac{u^k}{k}\frac{z^k}{1-z^k}\right).$$

Equivalently, from first principles,

$$\mathcal{P} \cong \prod_{n=1}^{\infty}\text{SEQ}(u\mathcal{I}_n) \quad\Longrightarrow\quad \prod_{n=1}^{\infty}\frac{1}{1-uz^n}.$$

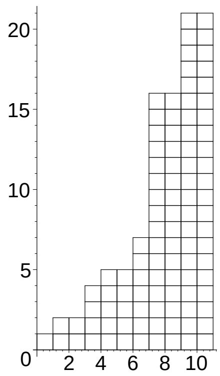


FIGURE 7. A random partition of size $n = 100$ has an aspect rather different from the profile of a random composition of the same size (Figure 6).

The OGF of cumulated values then results from the second form of the BGF by logarithmic differentiation:

$$\Omega(z) = P(z) \cdot \sum_{k=1}^{\infty} \frac{z^k}{1 - z^k}. \tag{24}$$

Now, the factor on the right in (24) can be expanded as

$$\sum_{k=1}^{\infty} \frac{z^k}{1 - z^k} = \sum_{n=1}^{\infty} d(n) z^n,$$

with $d(n)$ the number of divisors of n. Thus, the mean value of χ is

$$\mathbb{E}_n(\chi) = \frac{1}{P_n} \sum_{j=1}^{n} d(j) P_{n-j}. \tag{25}$$

The same technique applies to the number of parts equal to r. The form of the BGF is

$$\widetilde{\mathcal{P}} \cong \text{SEQ}(u\mathcal{I}_r) \times \prod_{n \neq r} \text{SEQ}(\mathcal{I}_n) \quad \Longrightarrow \quad \widetilde{P}(z, u) = \frac{1 - z^r}{1 - uz^r} \cdot P(z),$$

which implies that the mean value of the number $\widetilde{\chi}$ of r-parts satisfies

$$\mathbb{E}_n(\widetilde{\chi}) = \frac{1}{P_n} [z^n] \left(P(z) \cdot \frac{z^r}{1 - z^r} \right) = \frac{1}{P_n} \left(P_{n-r} + P_{n-2r} + P_{n-3r} + \cdots \right).$$

From these formulæ and a decent symbolic manipulation package, the means are calculated easily till values of n well in the range of several thousand. .... END OF EXAMPLE 6. □

The comparison between Figures 6 and 7 together with the supporting analysis shows that different combinatorial models may well lead to rather different types of probabilistic behaviours. Figure 8 displays the exact value of the mean number of parts in random partitions of size $n = 1, \ldots, 500$, (as calculated from (25)) accompanied with the observed values of one random sample for each value of n in the range. The

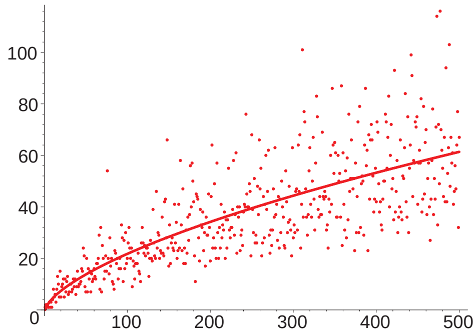


FIGURE 8. The number of parts in random partitions of size $1, \ldots, 500$: exact values of the mean and simulations (circles, one for each value of n).

mean number of parts is known to be asymptotic to

$$\frac{\sqrt{n}\log n}{\pi\sqrt{2/3}},$$

and the distribution, though it admits a comparatively large standard deviation ($O(\sqrt{n})$), is still concentrated in the technical sense; see [**128**].

In recent years, Vershik and his collaborators [**100, 425**] have shown that most integer partitions tend to conform to a definite profile given (after normalization by $\sqrt{n}$) by the continuous plane curve $y = \Psi(x)$ defined implicitly by

$$y = \Psi(x) \qquad \text{iff} \qquad e^{-\alpha x} + e^{-\alpha y} = 1, \qquad \alpha = \frac{\pi}{\sqrt{6}}. \tag{26}$$

This is illustrated in Figure 9 by two randomly drawn elements of $\mathcal{P}_{1000}$ represented together with the "most likely" limit shape. The theoretical result explains the huge

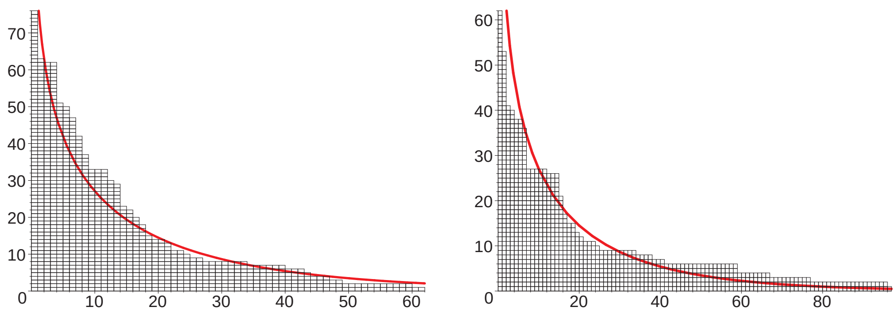


FIGURE 9. Two partitions of $\mathcal{P}_{1000}$ drawn at random, compared to the limiting shape $\Psi(x)$ defined by (26).

differences that are manifest on simulations between integer compositions and integer partitions.

The last example demonstrates the application of BGFs to estimates regarding the root degree of a tree drawn uniformly at random amongst the class $\mathcal{G}_n$ of general Catalan trees of size n. Tree parameters such as number of leaves and path length that are more global in nature and need a recursive definition will be discussed in Section III. 5 below.

EXAMPLE 7. *Root degree in general Catalan trees.* Consider the parameter χ equal to the degree of the root in a tree, and take the class $\mathcal{G}$ of all plane unlabelled trees, i.e., general Catalan trees. The specification is obtained by first defining trees ($\mathcal{G}$), then defining trees with a mark for subtrees($\mathcal{G}^\circ$) dangling from the root:

$$\begin{cases} \mathcal{G} = \mathcal{Z} \times \textsc{Seq}(\mathcal{G}) \\ \mathcal{G}^\circ = \mathcal{Z} \times \textsc{Seq}(u\mathcal{G}) \end{cases} \qquad \Longrightarrow \qquad \begin{cases} G(z) = \dfrac{z}{1-G(z)} \\ G(z,u) = \dfrac{z}{1-uG(z)}. \end{cases}$$

This set of equations reveals that the probability that the root degree equals r is

$$\mathbb{P}_n\{\chi = r\} = \frac{1}{G_n}[z^{n-1}]G(z)^r = \frac{r}{n-1}\binom{2n-3-r}{n-2} \sim \frac{r}{2^{r+1}},$$

this by Lagrange inversion and elementary asymptotics. Also, the cumulative GF is found to be

$$\Omega(z) = \frac{zG(z)}{(1-G(z))^2}.$$

The relation satisfied by G entails a further simplification,

$$\Omega(z) = \frac{1}{z}G(z)^3 = \left(\frac{1}{z}-1\right)G(z) - 1,$$

so that the mean root degree admits a closed form,

$$\mathbb{E}_n(\chi) = \frac{1}{G_n}\left(G_{n+1} - G_n\right) = 3\frac{n-1}{n+1},$$

a quantity clearly asymptotic to 3.

A random plane tree is thus usually composed of a small number of root subtrees, at least one of which should accordingly be fairly large. .... END OF EXAMPLE 7. □

III. 4. Inherited parameters and exponential MGFs

The theory of inheritance developed in the last section applies almost *verbatim* to labelled objects. The only difference is that the variable marking size must carry a factorial coefficient dictated by the needs of relabellings. Once more, with a suitable use of multi-index conventions, the translation mechanisms developed in the univariate case (Chapter II) remain in vigour, this in a way that parallels the unlabelled case.

Let us consider a pair $\langle \mathcal{A}, \chi \rangle$, where A is a labelled combinatorial class endowed with its size function $|\cdot|$ and $\chi = (\chi_1, \ldots, \chi_d)$ is a d-dimensional parameter. Like before, the parameter χ is extended into $\overline{\chi}$ by inserting size as zeroth coordinate and a vector $\mathbf{z} = (z_0, \ldots, z_d)$ of $d+1$ indeterminates is introduced, with z_0 marking size and z_j marking χ_j. Once the multi-index convention of (19) defining $\mathbf{z}^{\mathbf{k}}$ has been brought into the game, the exponential MGF of $\langle \mathcal{A}, \chi \rangle$ (see Definition III.4) can be rephrased as

$$
(27) \qquad \begin{aligned} A(\mathbf{z}) &= \sum_{\mathbf{k}} A_{\mathbf{k}} \frac{\mathbf{z}^{\mathbf{k}}}{k_0!} \\ &= \sum_{\alpha \in \mathcal{A}} \frac{\mathbf{z}^{\overline{\chi}(\alpha)}}{|\alpha|!}. \end{aligned}
$$

In a sense, this MGF is exponential in z (alias z_0) but ordinary in the other variables; only the factorial $k_0!$ is needed to take into account relabelling induced by labelled products.

We *a priori* restrict attention to parameters that do not depend on the absolute values of labels (but may well depend on the relative order of labels): a parameter is said to be *compatible* if, for any α, it assumes the same value on any labelled object α and all the order-consistent relabellings of α. A parameter is said to be *inherited* if it is compatible and it is defined by cases on disjoint unions and determined additively on labelled products—this is Definition III.5 with labelled products replacing cartesian products. In particular, for a compatible parameter, *inheritance signifies additivity on components of labelled sequences, sets, and cycles*. We can then cut-and-paste (with minor adjustments) the statement of Theorem III.1:

THEOREM III.2 (Inherited parameters and exponential MGFs). *Let $\mathcal{A}$ be a labelled combinatorial class constructed from $\mathcal{B}, \mathcal{C}$, and let χ be a parameter inherited from ξ defined on $\mathcal{B}$ and (as the case may be) from ζ on $\mathcal{C}$. Then the translation rules of admissible constructions stated in Theorem II.1 apply. is used. The associated operators on exponential MGFs are:*

$$
\begin{array}{llll}
\textit{Union:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) + C(\mathbf{z}) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \star \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) \cdot C(\mathbf{z}) \\
\textit{Sequence:} & \mathcal{A} = \textsc{Seq}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \dfrac{1}{1 - B(\mathbf{z})} \\
\textit{Cycle:} & \mathcal{A} = \textsc{Cyc}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \log \dfrac{1}{1 - B(\mathbf{z})}. \\
\textit{Set:} & \mathcal{A} = \textsc{Set}(\mathcal{B}) & \Longrightarrow & A(\mathbf{z}) = \exp\big(B(\mathbf{z})\big).
\end{array}
$$

PROOF. Disjoint unions are treated like in the unlabelled multivariate case. Labelled products result from

$$
A(\mathbf{z}) = \sum_{\alpha \in \mathcal{A}} \frac{\mathbf{z}^{\overline{\chi}(\alpha)}}{|\alpha|!} = \sum_{\beta \in \mathcal{B}, \gamma \in \mathcal{C}} \binom{|\beta| + |\gamma|}{|\beta|, |\gamma|} \frac{\mathbf{z}^{\overline{\xi}(\beta)}\, \mathbf{z}^{\overline{\zeta}(\gamma)}}{(|\beta| + |\gamma|)!},
$$

and the usual translation of binomial convolutions that reflect labellings by means of products of exponential generating functions (like in the univariate case detailed in Chapter II). The translation for composite constructions is then immediate. $\square$

This theorem can be exploited to determine moments, in a way that entirely parallels its unlabelled counterpart.

EXAMPLE 8. *The profile of permutations.* Let $\mathcal{P}$ be the class of all permutations and χ the number of components. Using the concept of marking, the specification and the exponential

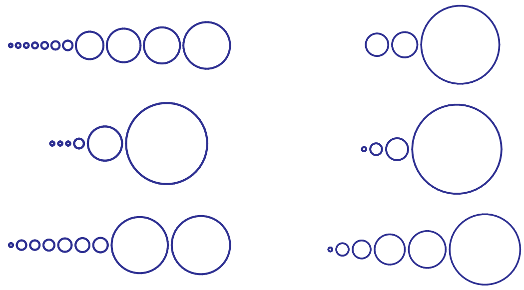


FIGURE 10. The profile of permutations: a rendering of the cycle structure of six random permutations of size 500, where circle areas are drawn in proportion to cycle lengths. Permutations tend to have a few small cycles (of size $O(1)$), a few large ones (of size $\Theta(n)$), and altogether have $\mathrm{H}_n \sim \log n$ cycles on average.

BGF are

$$\mathcal{P} = \mathrm{SET}\,(u\,\mathrm{CYC}(\mathcal{Z})) \qquad \Longrightarrow \qquad P(z,u) = \exp\left(u \log \frac{1}{1-z}\right) = (1-z)^{-u},$$

as was already obtained by an *ad hoc* calculation in (5). We also know (page 149) that the mean number of cycles is the harmonic number H_n and that the distribution is concentrated since the standard deviation is much smaller than the mean.

Regarding the number $\overline{\chi}$ of cycles of length r, the specification and the exponential BGF are now

$$(28) \qquad \begin{aligned} &\mathcal{P} = \mathrm{SET}\,(\mathrm{CYC}_{\neq r}(\mathcal{Z}) + u\,\mathrm{CYC}_{=r}(\mathcal{Z})) \\ &\Longrightarrow \quad P(z,u) = \exp\left(\log\frac{1}{1-z} + (u-1)\frac{z^r}{r}\right) = \frac{e^{(u-1)z^r/r}}{1-z}. \end{aligned}$$

The EGF of cumulated values is then

$$(29) \qquad \widetilde{\Omega}(z) = \frac{z^r}{r}\,\frac{1}{1-z}.$$

The result is a remarkably simple one: *In a random permutation of size n, the mean number of r-cycles is equal to $\frac{1}{r}$ for any $r \leq n$.*

Thus, the profile of a random permutation, where profile is defined as the ordered sequence of cycle lengths departs significantly from what has been encountered for integer compositions and partitions. Formula (29) also sheds a new light on the harmonic number formula for the mean number of cycles—each term $\frac{1}{r}$ in the harmonic number expresses the mean number of r cycles.

Since formulæ are so simple, one can get more information. By (28) one has, as seen above,

$$\mathbb{P}\{\overline{\chi} = k\} = \frac{1}{k!\,r^k}[z^{n-kr}]\frac{e^{-z^r/r}}{1-z},$$

where the last factor counts permutations without cycles of length r. From this (and the asymptotics of generalized derangement numbers in Chapter IV), one proves easily that the asymptotic

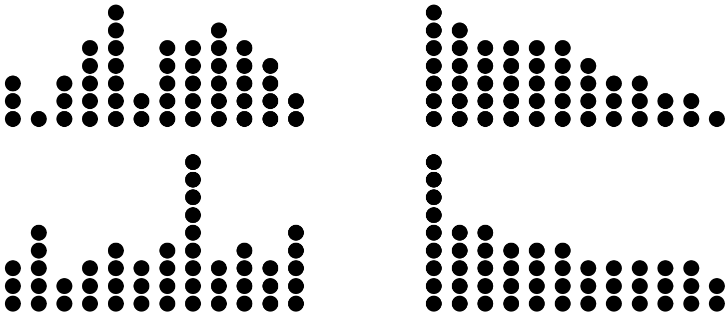


FIGURE 11. Two random allocations with $m = 12$, $n = 48$. The rightmost diagrams display the bins sorted by decreasing order of occupancy.

law of the number of r-cycles is Poisson[2] of rate $\frac{1}{r}$; in particular it is not concentrated. (This interesting property to be established in later chapters constitutes the starting point of an important study by Shep and Lloyd [**383**].)

Also, the mean number of cycles whose size is between $n/2$ and n is $\mathrm{H}_n - \mathrm{H}_{\lfloor n/2 \rfloor}$ a quantity that is approximately $\log 2 \doteq 0.69314$. In other words, we expect a random permutation of size n to have one or a few large cycles. (See the paper [**383**] for the original discussion of largest and smallest cycles). .... END OF EXAMPLE 8. □

EXAMPLE 9. *Allocations, balls-in-bins models, and the Poisson law.* Random allocations and the balls-in-bins model have been introduced in Chapter II in connection with the birthday paradox and the coupon collector problem. Under this model, there are n balls thrown into m bins in all possible ways, the total number of allocations being thus m^n. By the labelled construction of words, the bivariate EGF with z marking the number of balls and u marking the number $\chi^{(s)}$ of bins that contain s balls (s a fixed parameter) is given by

$$\mathcal{A} = \mathrm{SEQ}_m\left(\mathrm{SET}_{\neq s}(\mathcal{Z}) + u\,\mathrm{SET}_{=s}(\mathcal{Z})\right) \quad \Longrightarrow \quad A^{(s)}(z,u) = \left(e^z + (u-1)\frac{z^s}{s!}\right)^m.$$

In particular, the distribution of the number of empty bins ($\chi^{(0)}$) is expressible in terms of Stirling partition numbers:

$$\mathbb{P}_{m,n}(\chi^{(0)} = k) \equiv \frac{n!}{m^n}[u^k z^n]A^{(0)}(z,u) = \frac{(m-k)!}{m^n}\binom{m}{k}\left\{ {n \atop m-k} \right\}.$$

By differentiation of the BGF, there results an exact expression for the mean (any $s \geq 0$):

$$\frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)}) = \frac{1}{s!}\left(1 - \frac{1}{m}\right)^{n-s}\frac{n(n-1)\cdots(n-s+1)}{m^s}. \tag{30}$$

Let m and n tend to infinity in such a way that $\frac{n}{m} = \lambda$ is a fixed constant. This regime is extremely important in many applications, some of which are listed below. The average proportion

[2] The Poisson distribution of rate $\lambda > 0$ is supported by the nonnegative integers and determined by

$$\mathbb{P}\{k\} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

of bins containing s elements is $\frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)})$, and from (30), one obtains by straightforward calculations the asymptotic limit estimate,

$$\lim_{n/m=\lambda,\ n\to\infty} \frac{1}{m}\mathbb{E}_{m,n}(\chi^{(s)}) = e^{-\lambda}\frac{\lambda^s}{s!}. \tag{31}$$

In other words, a Poisson formula describes the average proportion of bins of a given size in a large random allocation. (Equivalently, the occupancy of a random bin in a random allocation satisfies a Poisson law in the limit.)

The variance of each $\chi^{(s)}$ (with fixed s) is estimated similarly via a second derivative and one finds:

$$\mathbb{V}_{m,n}(\chi^{(s)}) \sim me^{-2\lambda}\frac{\lambda^s}{s!}E(\lambda), \quad E(\lambda) := \left(e^{\lambda} - \frac{s\lambda^{s-1}}{(s-1)!} - (1-2s)\frac{\lambda^s}{s!} - \frac{\lambda^{s+1}}{s!}\right).$$

As a consequence, one has the convergence in probability,

$$\frac{1}{m}\chi^{(s)} \xrightarrow{P} e^{-\lambda}\frac{\lambda^s}{s!},$$

valid for any *fixed* $s \geq 0$. . . . . END OF EXAMPLE 9. □

▷ **9.** *Hashing and random allocations.* Random allocations of balls into bins are central in the understanding of a class of important algorithms of computer science known as *hashing* [**175, 269, 381, 382, 427**]: given a universe $\mathcal{U}$ of data, set up a function (called a hashing function) $h : \mathcal{U} \longrightarrow [1\,..\,m]$ and arrange for an array of m bins; an element $x \in \mathcal{U}$ is placed in bin number $h(x)$. If the hash function scrambles the data in a way that is suitably (pseudo)uniform, then the process of hashing a file of n records (keys, data items) into m bins is adequately modelled by a random allocation scheme. If $\lambda = \frac{n}{m}$, representing the "load", is kept reasonably bounded (say, $\lambda \leq 10$), the previous analysis implies that hashing allows for an almost direct access to data. ◁

Number of components in abstract labelled schemas. Like in the unlabelled universe, a general formula gives the distribution of the number of components for the basic constructions.

PROPOSITION III.6. *Consider labelled structures and the parameter χ equal to the number of components in a construction $\mathcal{A} = \mathfrak{K}\{\mathcal{B}\}$, where $\mathfrak{K}$ is one of* SEQ, SET CYC. *The exponential BGF $A(z,u)$ and the exponential GF $\Omega(z)$ of cumulated values are given by the following table:*

(32)

$\mathfrak{K}$	*exp. MGF* $(A(z,u))$	*Cumul. EGF* $(\Omega(z))$
SEQ :	$\dfrac{1}{1-uB(z)}$	$A(z)^2 \cdot B(z) = \dfrac{B(z)}{(1-B(z))^2}$
SET :	$\exp(uB(z))$	$A(z)\cdot B(z) = B(z)e^{B(z)}$
CYC :	$\log\dfrac{1}{1-uB(z)}$	$\dfrac{B(z)}{1-B(z)}$.

Mean values are then easily recovered, and one finds

$$\mathbb{E}_n(\chi) = \frac{\Omega_n}{A_n} = \frac{[z^n]\Omega(z)}{[z^n]A(z)},$$

by the same formula as in the unlabelled case.

$\triangleright$ **10.** *r-Components in abstract labelled schemas.* The BGF $A(z,u)$ and the cumulative EGF $\Omega(z)$ are given by the following table,

$$\begin{array}{lll} \text{SEQ}: & \dfrac{1}{1-\left(B(z)+(u-1)\frac{B_r z^r}{r!}\right)} & \dfrac{1}{(1-B(z))^2}\cdot\dfrac{B_r z^r}{r!} \\ \text{SET}: & \exp\left(B(z)+(u-1)\dfrac{B_r z^r}{r!}\right) & e^{B(z)}\cdot\dfrac{B_r z^r}{r!} \\ \text{CYC}: & \log\dfrac{1}{1-\left(B(z)+(u-1)\frac{B_r z^r}{r!}\right)} & \dfrac{1}{(1-B(z))}\cdot\dfrac{B_r z^r}{r!}, \end{array}$$

in the labelled case. $\triangleleft$

EXAMPLE 10. *Set partitions.* Set partitions $\mathcal{S}$ are sets of blocks, themselves nonempty sets of elements. The enumeration of set partitions according to the number of blocks is then given by

$$\mathcal{S}=\text{SET}(u\,\text{SET}_{\geq 1}(\mathcal{Z})) \quad\Longrightarrow\quad S(z,u)=e^{u(e^z-1)}.$$

Since set partitions are otherwise known to be enumerated by the Stirling partition numbers, one has the BGF and the vertical EGFs as a corollary,

$$\sum_{n,k}\left\{{n\atop k}\right\}u^k\frac{z^n}{n!}=e^{u(e^z-1)},\qquad \sum_{n}\left\{{n\atop k}\right\}\frac{z^n}{n!}=\frac{1}{k!}(e^z-1)^k,$$

which is consistent with earlier calculations of Chapter II.

The EGF of cumulated values, $\Omega(z)$ is then

$$\Omega(z)=(e^z-1)e^{e^z-1},$$

which is almost a derivative of $S(z)$:

$$\Omega(z)=\frac{d}{dz}S(z)-S(z).$$

Thus, the mean number of blocks in a random partition of size n is

$$\frac{\Omega_n}{S_n}=\frac{S_{n+1}}{S_n}-1,$$

a quantity directly expressible in terms of Bell numbers. A delicate computation based on the asymptotic expansion of the Bell numbers reveals that the expected value and the standard deviation are asymptotic to (Chapter VIII)

$$\frac{n}{\log n},\qquad \frac{\sqrt{n}}{\log n},$$

respectively. Similarly the exponential BGF of the number of blocks of size k is

$$\mathcal{S}=\text{SET}(u\,\text{SET}_{=k}(\mathcal{Z})+\text{SET}_{\neq 0,k}(\mathcal{Z})) \qquad\Longrightarrow\qquad S(z,u)=e^{e^z-1+(u-1)z^k/k!},$$

out of which mean and variance can be derived. .... END OF EXAMPLE 10. $\square$

EXAMPLE 11. *Root degree in Cayley trees.* Consider the class $\mathcal{T}$ of Cayley trees (nonplane labelled trees) and the parameter "root-degree". The basic specifications are

$$\left\{\begin{array}{lll}\mathcal{T} & = & \mathcal{Z}\star\text{SET}(\mathcal{T}) \\ \mathcal{T}^\circ & = & \mathcal{Z}\star\text{SET}(u\mathcal{T})\end{array}\right. \qquad\Longrightarrow\qquad \left\{\begin{array}{lll}T(z) & = & ze^{T(z)} \\ T(z,u) & = & ze^{uT(z)}.\end{array}\right.$$

The set construction reflects the non-planar character of Cayley trees and the specification $\mathcal{T}^\circ$ is enriched by a mark associated to subtrees dangling from the root. Lagrange inversion provides the fraction of trees with root degree k,

$$\frac{1}{(k-1)!}\frac{n!}{(n-1-k)!}\frac{(n-1)^{n-2-k}}{n^{n-1}} \sim \frac{e^{-1}}{(k-1)!}, \qquad k \geq 1.$$

Similarly, the cumulative GF is found to be $\Omega(z) = T(z)^2$, so that the mean root degree satisfies

$$\mathbb{E}_{\mathcal{T}_n}(\text{root degree}) = 2(1-\frac{1}{n}) \sim 2.$$

Thus the law of root degree is asymptotically a Poisson law of rate 1 (shifted by 1). Probabilistic phenomena qualitatively similar to those encountered in plane trees are observed here as the mean root degree is asymptotic to a constant. However a Poisson law eventually reflecting the nonplanarity condition replaces the modified geometric law (known as a negative binomial law) present in plane trees. . . . . END OF EXAMPLE 11. □

▷ **11.** *Numbers of components in alignments.* Alignments ($\mathcal{O}$) are sequences of cycles (Chapter II). The expected number of components in a random alignment of $\mathcal{O}_n$ is

$$\frac{[z^n]\log(1-z)^{-1}(1-\log(1-z)^{-1})^{-2}}{[z^n](1-\log(1-z)^{-1})^{-1}}.$$

Methods of Chapter V imply that the number of components in a random alignment has expectation $\sim n/(e-1)$ and standard deviation $\Theta(\sqrt{n})$. ◁

▷ **12.** *Image cardinality of a random surjection.* The expected cardinality of the image of a random surjection in $\mathcal{R}_n$ (see Chapter II) is

$$\frac{[z^n]e^z(2-e^z)^{-2}}{[z^n](2-e^z)^{-1}}.$$

The number of values whose preimages have cardinality k is obtained by replacing the single exponential factor e^z by $z^k/k!$. Methods of Chapters IV and V imply that the image cardinality of a random surjection has expectation $n/(2\log 2)$ and standard deviation $\Theta(\sqrt{n})$. ◁

▷ **13.** *Distinct component sizes in set partitions.* Take the number of *distinct* block sizes and cycle sizes in set partitions and permutations. The bivariate EGFs are

$$\prod_{n=1}^{\infty}\left(1-u+ue^{z^n/n!}\right), \qquad \prod_{n=1}^{\infty}\left(1-u+ue^{z^n/n}\right),$$

as follows from first principles. ◁

Postscript: Towards a theory of schemas. Let us look back and recapitulate some of the information gathered in pages 156—169 regarding the number of components in composite structures. The classes considered in the table below are compositions of two constructions, either in the unlabelled or the labelled universe. Each entry contains the BGF for the number of components (e.g., cycles in permutations, parts in integer partitions, and so on), and the asymptotic orders of the mean and standard deviation of the number of components for objects of size n.

Some obvious facts stand out from the data and call for explanation. First the outer construction appears to play the essential rôle: outer *sequence* constructs (cf integer compositions, surjections and alignments) tend to dictate a number of components that is $\Theta(n)$ on average, while outer *set* constructs (cf integer compositions,

Unlabelled structures	
Integer partitions, MSET $\circ$ SEQ $\exp\left(u\frac{z}{1-z}+\frac{u^2}{2}\frac{z^2}{1-z^2}+\cdots\right)$ $\sim \frac{\sqrt{n}\log n}{\pi\sqrt{2/3}}, \quad \Theta(\sqrt{n})$	Integer compositions, SEQ $\circ$ SEQ $\left(1-u\frac{z}{1-z}\right)^{-1}$ $\sim \frac{n}{2}, \quad \Theta(\sqrt{n})$
Labelled structures	
Set partitions, SET $\circ$ SET $\exp\left(u\left(e^z-1\right)\right)$ $\sim \frac{n}{\log n} \quad \sim \frac{\sqrt{n}}{\log n}$	Surjections, SEQ $\circ$ SET $\left(1-u\left(e^z-1\right)\right)^{-1}$ $\sim \frac{n}{2\log 2}, \quad \Theta(\sqrt{n})$
Permutations, SET $\circ$ CYC $\exp\left(u\log(1-z)^{-1}\right)$ $\sim \log n, \quad \sim \sqrt{\log n}$	Alignments, SEQ $\circ$ CYC $\left(1-u\log(1-z)^{-1}\right)^{-1}$ $\sim \frac{n}{e-1}, \quad \Theta(\sqrt{n})$

FIGURE 12. Major properties of the number of components in six level-two structures. For each class, from top to bottom: (i) specification type; (ii) BGF; (iii) mean and variance of the number of components.

set partitions, and permutations) are associated with a greater variety of asymptotic regimes. Eventually, such facts can be organized into broad analytic schemas, as will be seen in Chapters IV–IX.

$\triangleright$ **14.** *Singularity and probability.* The differences in behaviour are to be assigned to the rather different types of singularity involved: on the one hand sets corresponding algebraically to an $\exp(\cdot)$ operator induce an exponential blow up of singularities; on the other hand sequences expressed algebraically by quasi-inverses $(1-\cdot)^{-1}$ are likely to induce polar singularities. Recursive structures like trees lead to yet other types of phenomena with a number of components, i.e., the root degree, that is bounded in probability. $\triangleleft$

III. 5. Recursive parameters

In this section, we adapt the general methodology of previous sections in order to treat parameters that are defined by recursive rules over structures that are themselves recursively specified. Typical applications concern trees and tree-like structures.

Regarding the number of leaves, or more generally, the number of nodes of some fixed degree, in a tree, the method of placing marks applies like in the non-recursive case. It suffices to distinguish elements of interest and mark them by an auxiliary variable. For instance, in order to mark composite objects made of r components, where r is an integer and $\mathfrak{K}$ designates any of SEQ, SET (or MSET, PSET), CYC, one should split a construction $\mathfrak{K}(\mathcal{C})$ according to the identity

$$\mathfrak{K}(\mathcal{C}) = \mathfrak{K}_{=r}(\mathcal{C}) + \mathfrak{K}_{\neq r}(\mathcal{C}),$$

then introduce a mark (u) in front of the first term of the sum. This technique gives rise to specifications decorated by marks to which Theorems III.1 and III.2 apply. For a recursively defined structure, the outcome is a functional equation defining the BGF recursively. This technique is illustrated by Examples 12 ands 13 below in the case of Catalan trees and the parameter number of leaves.

EXAMPLE 12. *Leaves in general Catalan trees.* How many leaves does a random tree of some variety have? Can different varieties of trees be somehow distinguished by the proportion of their leaves? Beyond the botany of combinatorics, such considerations are for instance relevant to the analysis of algorithms since tree leaves, having no descendants, can be stored more economically; see [**268**, Sec. 2.3] for an algorithmic motivation for such questions.

Consider once more the class $\mathcal{G}$ of plane unlabelled trees, $\mathcal{G} = \mathcal{Z} \times \text{SEQ}(\mathcal{G})$, enumerated by the Catalan numbers: $G_n = \frac{1}{n}\binom{2n-2}{n-1}$. The class $\mathcal{G}^\circ$ where each leaf is marked is

$$\mathcal{G}^\circ = \mathcal{Z}u + \mathcal{Z} \times \text{SEQ}_{\geq 1}(\mathcal{G}^\circ) \qquad \Longrightarrow \qquad G(z,u) = zu + \frac{zG(z,u)}{1-G(z,u)}.$$

The induced quadratic equation can be solved explicitly

$$G(z,u) = \frac{1}{2}\left(1 + (u-1)z - \sqrt{1 - 2(u+1)z + (u-1)^2z^2}\right).$$

It is however simpler to expand using the Lagrange inversion theorem which provides

$$\begin{aligned} G_{n,k} &= [u^k]\left([z^n]G(z,u)\right) = [u^k]\left(\frac{1}{n}[y^{n-1}](u + \frac{y}{1-y})^n\right) \\ &= \frac{1}{n}\binom{n}{k}[y^{n-1}]\frac{y^{n-k}}{(1-y)^{n-k}} = \frac{1}{n}\binom{n}{k}\binom{n-2}{k-1}. \end{aligned}$$

These numbers are known as Narayana numbers, see *EIS* **A001263**, and they surface repeatedly in connexion with ballot problems. The mean number of leaves derives from the cumulative GF, which is

$$\Omega(z) = \left.\partial_u G(z,u)\right|_{u=1} = \frac{1}{2}\,z + \frac{1}{2}\,\frac{z}{\sqrt{1-4z}},$$

so that the mean is $n/2$ exactly for $n \geq 2$. The distribution is concentrated since the standard deviation is easily calculated to be $O(\sqrt{n})$. .... END OF EXAMPLE 12. □

EXAMPLE 13. *Leaves and node types in binary trees.* The class $\mathcal{B}$ of binary plane trees, also enumerated by Catalan numbers ($B_n = \frac{1}{n+1}\binom{2n}{n}$) can be specified as

$$\mathcal{B} = \mathcal{Z} + (\mathcal{B} \times \mathcal{Z}) + (\mathcal{Z} \times B) + (\mathcal{B} \times \mathcal{Z} \times \mathcal{B}), \tag{33}$$

which stresses the distinction between four types of nodes: leaves, left branching, right branching, and binary. Let u_0, u_1, u_2 be variables that mark nodes of degree 0,1,2, respectively. Then the root decomposition (33) provides for the MGF $B = B(z, u_0, u_1, u_2)$ the functional equation

$$B = zu_0 + 2zu_1B + zu_2B^2,$$

which, by Lagrange inversion, gives

$$B_{n,k_0,k_1,k_2} = \frac{2^{k_1}}{n}\binom{n}{k_0, k_1, k_2},$$

subject to the natural conditions: $k_0 + k_1 + k_2 = n$ and and $k_0 = k_2 + 1$. Specializations and moments can be easily calculated from such an approach [**356**]. In particular, the mean number of nodes of each type is asymptotically:

$$\text{leaves: } \sim \frac{n}{4}, \qquad \text{1-nodes}: \sim \frac{n}{2}, \qquad \text{2-nodes}: \sim \frac{n}{4}.$$

There is an equal asymptotic proportion of leaves, double nodes, left branching, and right branching nodes. Also, the standard deviation is in each case $O(\sqrt{n})$, so that each of the corresponding distributions is concentrated. .... END OF EXAMPLE 13. □

▷ **15.** *Leaves and node-degree profile in Cayley trees.* For Cayley trees, the bivariate EGF with u marking the number of leaves is the solution to

$$T(z,u) = uz + z(e^{T(z,u)} - 1).$$

The distribution is expressed in terms of Stirling partition numbers. The mean number of leaves in a random Cayley tree is asymptotic to ne^{-1}.

More generally, the mean number of nodes of outdegree k in a random Cayley tree of size n is asymptotic to

$$n \cdot e^{-1} \frac{1}{k!}.$$

Degrees of nodes are thus approximately given by a Poisson law of rate 1. ◁

▷ **16.** *Node-degree profile in simple varieties of trees.* For a family of trees generated by $T(z) = z\phi(T(z))$ with ϕ a power series, the BGF of the number of nodes of degree k satisfies

$$T(z,u) = z\left(\phi(T(z,u)) + \phi_k(u-1)T(z,u)^k\right),$$

where $\phi_k = [u^k]\phi(u)$. The cumulative GF is

$$\Omega(z) = z\frac{\phi_k T(z)^k}{1 - z\phi'(T(z))} = \phi_k z^2 T(z)^{k-1} T'(z),$$

from which expectations can be determined. ◁

▷ **17.** *Marking in functional graphs.* Consider the class $\mathcal{F}$ of finite mappings discussed in Chapter II:

$$\mathcal{F} = \textsc{Set}(\mathcal{K}), \qquad \mathcal{K} = \textsc{Cyc}(\mathcal{T}), \qquad \mathcal{T} = \mathcal{Z} \star \textsc{Set}(\mathcal{T}).$$

The translation into EGFs is

$$F(z) = e^{K(z)}, \qquad K(z) = \log\frac{1}{1 - T(z)}, \qquad T(z) = ze^{T(z)}.$$

Here are bivariate EGFs for (i) the number of components, (ii) the number of maximal trees, (iii) the number of leaves:

$$(i)\ e^{uK(z)}, \qquad (ii)\ \frac{1}{1 - uT(z)},$$
$$(iii)\ \frac{1}{1 - T(z,u)} \quad \text{with} \quad T(z,u) = (u-1)z + ze^{T(z,u)}.$$

The trivariate EGF $F(u_1, u_2, z)$ of functional graphs with u_1 marking components and u_2 marking trees is

$$F(z, u_1, u_2) = \exp(u_1 \log(1 - u_2 T(z))^{-1}) = \frac{1}{(1 - u_2 T(z))^{u_1}}.$$

An explicit expression for the coefficients involves the Stirling cycle numbers. ◁

We shall stop here these examples that could be multiplied *ad libitum* since such calculations greatly simplify when interpreted in the light of asymptotic analysis. The phenomena observed asymptotically are, for good reasons, especially close to what the classical theory of branching processes provides (see the book by Harris [**226**]).

Linear transformations on parameters and path length in trees. We have so far been dealing with a parameter defined directly by recursion. Next, we turn to other parameters such as path length. As a preamble, one needs a simple linear transformation on combinatorial parameters. Let $\mathcal{A}$ be a class equipped with two scalar parameters, χ and ξ, related by

$$\chi(\alpha) = |\alpha| + \xi(\alpha).$$

Then, the combinatorial form of BGFs yields

$$\sum_{\alpha\in\mathcal{A}} z^{|\alpha|}u^{\chi(\alpha)} = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|}u^{|\alpha|+\xi(\alpha)} = \sum_{\alpha\in\mathcal{A}} (zu)^{|\alpha|}u^{\xi(\alpha)},$$

that is,

$$A_\chi(z,u) = A_\xi(zu,u). \tag{34}$$

This is clearly a general mechanism:

> **Linear transformations and MGFs:** *A linear transformation on parameters induces a monomial substitution on the corresponding marking variables in MGFs.*

We now put this mechanism to use in the recursive analysis of path length in trees.

EXAMPLE 14. *Path length in trees.* The path length of a tree is defined as the sum of distances of all nodes to the root of the tree, where distances are measured by the number of edges on the minimal connecting path of a node to the root. Path length is an important characteristic of trees. For instance, when a tree is used as a data structure with nodes containing additional information, path length represents the total cost of accessing all data items when a search is started from the root. For this reason, path length surfaces, under various models, in the analysis of algorithms like algorithms and data structures for searching and sorting (e.g., treesort, quicksort, radix-sort); see [**268, 382**].

The definition of path length as

$$\lambda(\tau) := \sum_{\nu\in\tau} \text{dist}(\nu, \text{root}(\tau)),$$

transforms into an inductive definition:

$$\lambda(\tau) = \sum_{\upsilon \text{ root subtree of } \tau} (\lambda(\upsilon) + |\upsilon|)\,. \tag{35}$$

To establish this identity, distribute nodes in their corresponding subtrees; correct distances to the subtree roots by 1, and regroup terms.

From this point on, we specialize the discussion to general Catalan trees (see Note 18 for other cases): $\mathcal{G} = \mathcal{Z} \times \textsc{Seq}(calG)$. Introduce momentarily the parameter $\mu(\tau) = |\tau| + \lambda(\tau)$. Then, one has from the inductive definition (35) and the general transformation rule (34):

$$G_\lambda(z,u) = \frac{z}{1-G_\mu(z,u)} \qquad \text{and} \qquad G_\mu(z,u) = G_\lambda(zu,u). \tag{36}$$

In other words, $G(z, u) \equiv G_\lambda(z, u)$ satisfies a nonlinear functional equation of the difference type:

$$G(z, u) = \frac{z}{1 - G(uz, u)}.$$

(This functional equation will be encountered again in connection with area under Dyck paths: see Chapter V, p. 309.) The generating function $\Omega(z)$ of cumulated values of λ then obtains by differentiation with respect to u upon setting $u = 1$. We find in this way that $\Omega(z) := \partial_u G(z, u)\big|_{u=1}$ satisfies

$$\Omega(z) = \frac{z}{(1 - G(z))^2} \left(zG'(z) + \Omega(z) \right),$$

which is a linear equation that solves to

$$\Omega(z) = z^2 \frac{G'(z)}{(1 - G(z))^2 - z} = \frac{z}{2(1 - 4z)} - \frac{z}{2\sqrt{1 - 4z}}.$$

Consequently, one has ($n \geq 1$)

$$\Omega_n = 2^{2n-3} - \frac{1}{2}\binom{2n-2}{n-1},$$

where the sequence starting 1, 5, 22, 93, 386 for $n \geq 2$ constitutes *EIS* **A000346**. By an elementary asymptotic analysis, we get:

> *The mean path length of a random Catalan tree of size n is asymptotic to $\frac{1}{2}\sqrt{\pi n^3}$; in short: a branch from the root to a random node in a random Catalan tree of size n has expected length of the order of $\sqrt{n}$.*

Random Catalan trees thus tend to be somewhat imbalanced—by comparison, a fully balanced binary tree has all paths of length at most $\log_2 n + O(1)$. .... END OF EXAMPLE 14. □

The imbalance in random Catalan trees is a general phenomenon—it holds for binary Catalan and more generally for all simple varieties of trees. Note 18 below and Chapter VII (p. 435) imply that path length is invariably of order $n\sqrt{n}$ on average in such cases. Height is of typical order $\sqrt{n}$ as shown by Rényi and Szekeres [**361**], de Bruijn, Knuth and Rice [**95**], Kolchin [**276**], as well as Flajolet, and Odlyzko [**165**]. Figure 13 borrowed from [**382**] illustrates this on a simulation. (The contour of the histogram of nodes by levels, once normalized, has been proved to converge to the process known as Brownian excursion.)

▷ **18.** *Path length in simple varieties of trees.* The BGF of path length in a variety of trees generated by $T(z) = z\phi(T(z))$ satisfies

$$T(z, u) = z\phi(T(zu, u)).$$

In particular, the cumulative GF is

$$\Omega(z) \equiv \partial_u \left(T(z, u) \right)_{u=1} = \frac{\phi'(T(z))}{\phi(T(z))} \left(zT'(z) \right)^2,$$

from which coefficients can be extracted. ◁

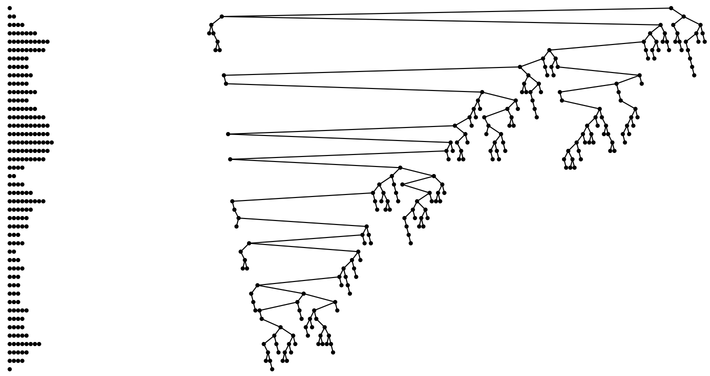


FIGURE 13. A random pruned binary tree of size 256 and its associated level profile: the histogram on the left displays the number of nodes at each level in the tree.

III. 6. Complete generating functions and discrete models

By a *complete* generating function, we mean, loosely speaking, a generating function in a (possibly large, and even infinite in the limit) number of variables that mark a homogeneous collection of characteristics of a combinatorial class[3]. For instance one may be interested in the joint distribution of *all* the different letters composing words, the number of cycles of *all* lengths in permutations, and so on. A complete MGF naturally entails very detailed knowledge on the enumerative properties of structures to which it is relative. Complete generating functions, given their expressive power, also make weighted models accessible to calculation, a situation that covers in particular Bernoulli trials (p. 179) and branching processes from classical probability theory (p. 185).

Complete GFs for words. As a basic example, consider the class of all words $\mathcal{W} = \mathrm{S}\textsc{eq}\{\mathcal{A}\}$ over some finite alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$. Let $\chi = (\chi_1, \ldots, \chi_r)$, where $\chi_j(w)$ is the number of occurrences of the letter a_j in word w. The MGF of $\mathcal{A}$ with respect to χ is

$$\mathcal{A} = u_1 a_1 + u_2 a_2 + \cdots u_r a_r \qquad \Longrightarrow \qquad A(z, \mathbf{u}) = zu_1 + zu_2 + \cdots + zu_r,$$

and χ on $\mathcal{W}$ is clearly inherited from χ on $\mathcal{A}$. Thus, by the sequence rule, one has

$$(37) \qquad \mathcal{W} = \mathrm{S}\textsc{eq}(\mathcal{A}) \qquad \Longrightarrow \qquad W(z, \mathbf{u}) = \frac{1}{1 - z(u_1 + u_2 + \cdots + u_r)},$$

[3]Complete GFs are *not* new objects. They are simply a an avatar of multivariate GFs. Thus the term is only meant to be suggestive of a particular usage of MGFs, and essentially no new theory is needed in order to cope with them.

which describes all words according to their compositions into letters. In particular, the number of words with n_j occurrences of letter a_j and $n = \sum n_j$ is in this framework obtained as

$$[u_1^{n_1} u_2^{n_2} \cdots u_r^{n_r}]\,(u_1 + u_2 + \cdots + u^r)^n = \binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r}.$$

We are back to the usual multinomial coefficients.

$\triangleright$ **19.** *After Bhaskara Acharya* (*circa 1150*AD). Consider all the numbers formed in decimal with digit 1 used once, with digit 2 used twice,..., with digit 9 used nine times. Such numbers all have 45 digits. Compute their sum S and discover, much to your amazement that S equals

$$458755596000061532190847692863\underline{999999999999999}5412444039999384678091523071360000 0.$$

This number has a long run of nines (and further nines are hidden!). Is there a simple explanation? This exercise is inspired by the Indian mathematician Bhaskara Acharya who discovered multinomial coefficients near 1150AD; see [**268**, p. 23] for a brief historical note. $\triangleleft$

Complete GFs for permutations and set partitions. Consider permutations and the various lengths of their cycles. The MGF where u_k marks cycles of length k for $k = 1, 2, \ldots$ can be written as an MGF in *infinitely many* variables:

$$P(z, \mathbf{u}) = \exp\left(u_1 \frac{z}{1} + u_2 \frac{z^2}{2} + u_3 \frac{z^3}{3} + \cdots\right). \tag{38}$$

The MGF expression U has the neat feature that, upon specializing all but a finite number of u_j to 1, we derive all the particular cases of interest with respect to any finite collection of cycles lengths. Observe also that one can calculate in the usual way any coefficient $[z^n]P$ as it only involves the variables $u_1, \ldots, u_n$.

$\triangleright$ **20.** *The theory of formal power series in infinitely many variables.* (This note is for formalists.) Mathematically, an object like P in (38) is perfectly well defined. Let $U = \{u_1, u_2, \ldots\}$ be an infinite collection of indeterminates. First, the ring of polynomials $R = \mathbb{C}[U]$ is well defined and a given element of R involves only finitely many indeterminates. Then, from R, one can define the ring of formal power series in z, namely $R[\![z]\!]$. (Note that, if $f \in R[\![z]\!]$, then each $[z^n]f$ involves only finitely many of the variables u_j.) The basic operations and the notion of convergence, as described in APPENDIX A: *Formal power series*, p. 648, apply in a standard way.

For instance, in the case of (38), the complete GF $P(z, \mathbf{u})$ is obtainable as the formal limit

$$P(z, \mathbf{u}) = \lim_{k \to \infty} \exp\left(u_1 \frac{z}{1} + \cdots + u_k \frac{z^k}{k} + \frac{z^{k+1}}{k+1} + \cdots\right)$$

in $R[\![z]\!]$ equipped with the formal topology. (In contrast, the quantity evocative of a generating function of words over an infinite alphabet

$$W \stackrel{!}{=} \left(1 - z \sum_{j=1}^{\infty} u_j\right)^{-1}$$

cannot receive a sound definition as a element of the formal domain $R[\![z]\!]$.) $\triangleleft$

Henceforth, we shall keep in mind that verifications of formal correctness regarding power series in infinitely many indeterminates are always possible by returning to basic definitions.

Complete generating functions are often surprisingly simple to expand. For instance, the equivalent form of (38)

$$P(z, \mathbf{u}) = e^{u_1 z/1} \cdot e^{u_2 z^2/2} \cdot e^{u_3 z^3/3} \cdots$$

implies immediately that the number of permutations with k_1 cycles of size 1, k_2 of size 2, and so on, is

$$\frac{n!}{k_1!\, k_2! \cdots k_n!\; 1^{k_1}\, 2^{k_2} \cdots n^{k_n}}, \tag{39}$$

provided $\sum j k_j = n$. This is a result originally due to Cauchy. Similarly, the EGF of set partitions with u_j marking the number of blocks of size j is

$$S(z, \mathbf{u}) = \exp\left(u_1 \frac{z}{1!} + u_2 \frac{z^2}{2!} + u_3 \frac{z^3}{3!} + \cdots\right).$$

A formula analogous to (39) follows: the number of partitions with k_1 blocks of size 1, k_2 of size 2, and so on, is

$$\frac{n!}{k_1!\, k_2! \cdots k_n!\; 1!^{k_1}\, 2!^{k_2} \cdots n!^{k_n}}.$$

Several examples of such complete generating functions are presented in Comtet's book; see [**82**], pages 225 and 233.

▷ **21.** *Complete GFs for compositions and surjections.* The complete GFs of integer compositions and surjections with u_j marking the number of components of size j are

$$\frac{1}{1 - \sum_{j=1}^{\infty} u_j z^j}, \qquad \frac{1}{1 - \sum_{j=1}^{\infty} u_j \frac{z^j}{j!}}.$$

The associated counts with $n = \sum_j j k_j$ are given by

$$\binom{k_1 + k_2 + \cdots}{k_1, k_2, \ldots}, \qquad \frac{n!}{1!^{k_1} 2!^{k_2} \cdots} \binom{k_1 + k_2 + \cdots}{k_1, k_2, \ldots}.$$

These factored forms derive directly from the multinomial expansion. The symbolic form of the multinomial expansion of powers of a generating function is sometimes expressed in terms of Bell polynomials, themselves nothing but a rephrasing of the multinomial expansion; see Comtet's book [**82**, Sec. 3.3] for a fair treatment of such polynomials. ◁

▷ **22.** *Faà di Bruno's formula.* The formulæ for the successive derivatives of a functional composition $h(z) = f(g(z))$

$$\partial_z h(z) = f'(g(z))g'(z), \quad \partial_z^2 h(z) = f''(g(z))g'(z)^2 + f'(z)g''(z), \ldots,$$

are clearly equivalent to the expansion of a formal power series composition. Indeed, assume without loss of generality that $z = 0$ and $g(0) = 0$; set $f_n := \partial_z^n f(0)$, and similarly for g, h. Then:

$$h(z) \equiv \sum_n h_n \frac{z^n}{n!} = \sum_k \frac{f_k}{k!} \left(g_1 z + \frac{g_2}{2!} z^2 + \cdots\right)^k.$$

Thus in one direct application of the multinomial expansion, one finds

$$\frac{h_n}{n!} = \sum_k \frac{f_k}{k!} \sum_{\mathcal{C}} \binom{k}{\ell_1, \ell_2, \ldots, \ell_k} \left(\frac{g_1}{1!}\right)^{\ell_1} \left(\frac{g_2}{2!}\right)^{\ell_2} \cdots \left(\frac{g_k}{k!}\right)^{\ell_k}$$

where the summation condition $\mathcal{C}$ is: $1\ell_1 + 2\ell_2 + \cdots + k\ell_k = n$, $\ell_1 + \ell_2 + \cdots + \ell_k = k$. This shallow identity is known as Faà di Bruno's formula [**82**, p. 137]. (Faà di Bruno (1825–1888) was canonized by the Catholic Church in 1988, presumably for reasons not related to his formula.) ◁

▷ **23.** *Relations between symmetric functions.* Symmetric functions may be manipulated by mechanisms that are often reminiscent of the set and multiset construction. They appear in many areas of combinatorial enumeration. Let $X = \{x_i\}_{i=1}^r$ be a collection of formal variables. Define the symmetric functions

$$\prod_i (1 + x_i z) = \sum_n a_n z^n, \quad \prod_i \frac{1}{1 - x_i z} = \sum_n b_n z^n, \quad \sum_i \frac{x_i z}{1 - x_i z} = \sum_n c_n z^n.$$

The a_n, b_n, c_n, called resp. elementary, monomial, and power symmetric functions are expressible as

$$a_n = \sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} x_{i_2} \cdots x_{i_r}, \quad b_n = \sum_{i_1 \le i_2 \le \cdots \le i_r} x_{i_1} x_{i_2} \cdots x_{i_r}, \quad c_n = \sum_{i=1}^r x_i^r.$$

The following relations hold for the OGFs $A(z), B(z), C(z)$ of a_n, b_n, c_n::

$$\begin{array}{rclcrcl} B(z) & = & \dfrac{1}{A(-z)}, & & A(z) & = & \dfrac{1}{B(-z)}, \\ C(z) & = & z\dfrac{d}{dz}\log B(z), & & B(z) & = & \exp \displaystyle\int_0^z C(t)\,\frac{dt}{t}. \end{array}$$

Consequently, each of a_n, b_n, c_n is polynomially expressible in terms of any of the other quantities. (The connection coefficients, like in Note 22, involve multinomials.) ◁

▷ **24.** *Regular graphs.* A graph is r–regular iff each node has degree exactly equal to r. The number of r–regular graphs of size n is

$$[x_1^r x_2^r \cdots x_n^r] \prod_{1 \le i < j \le n} (1 + x_i x_j).$$

[Gessel [**199**] has shown how to extract explicit expressions from such huge symmetric functions.] ◁

III. 6.1. Word models. . The enumeration of words constitutes a rich chapter of combinatorial analysis, and complete GFs serve to generalize many results to the case of nonuniform letter probabilities, like the coupon collector problem and the birthday paradox considered in Chapter II. Applications are to be found in classical probability theory and statistics [**90**] (the so-called Bernoulli trial models), as well as in computer science [**401**] and mathematical models of biology [**432**].

EXAMPLE 15. *Words and records.* Fix an alphabet $\mathcal{A} = \{a_1, \dots, a_r\}$ and let $\mathcal{W} = \text{SEQ}\{\mathcal{A}\}$ be the class of all words over $\mathcal{A}$, where $\mathcal{A}$ is naturally ordered by $a_1 < a_2 < \cdots < a_r$. Given a word $w = w_1 \cdots w_n$, a (strict) record is an element w_j that is larger than all preceding elements: $w_j > w_i$ for all $i < j$. (Refer to Figure 13 of Chapter II for a graphical rendering of records in the case of permutations.)

Consider first the subset of $\mathcal{W}$ comprising all words that have the letters $a_{i_1}, \dots, a_{i_k}$ as successive records, where $i_1 < \cdots < i_k$. The symbolic description of this set is in the form of a product of k terms

$$\text{SEQ}\Big(a_{i_1}(a_1 + \cdots + a_{i_1})\Big) \quad \cdots \quad \text{SEQ}\Big(a_{i_k}(a_1 + \cdots + a_{i_k})\Big). \tag{40}$$

Consider now MGFs of words where z marks length, v marks the number of records, and each u_j marks the number of occurrences of letter a_j. The MGF associated to the subset described in (40) is then

$$\left(zvu_{i_1}(1-z(u_1+\cdots+u_{i_1}))^{-1}\right) \quad \cdots \quad \left(zvu_{i_k}(1-z(u_1+\cdots+u_{i_k}))^{-1}\right).$$

Summing over all values of k and of $i_1 < \cdots < i_k$ gives

$$W(z,v,\mathbf{u}) = \prod_{s=1}^{r} \left(1 + zvu_s\,(1-z(u_1+\cdots+u_s))^{-1}\right), \tag{41}$$

the rationale being that, for arbitrary quantities y_s, one has by distributivity:

$$\sum_{k=0}^{r} \sum_{1\le i_1<\cdots<i_k\le r} y_{i_1}y_{i_2}\cdots y_{i_k} = \prod_{s=1}^{r}(1+y_s).$$

We shall encounter more applications of (41) below. For the time being let us simply examine the mean number of records in a word of length n over the alphabet $\mathcal{A}$, when all such words are taken equally likely. One should set $u_j \mapsto 1$ (the composition into specific letters is forgotten), so that W assumes the simpler form

$$W(z,v) = \prod_{j=1}^{r}\left(1+\frac{vz}{1-jz}\right).$$

Logarithmic differentiation then gives access to the generating function of cumulated values,

$$\Omega(z) \equiv \left.\frac{\partial}{\partial v}W(z,v)\right|_{v=1} = \frac{z}{1-rz}\sum_{j=1}^{r}\frac{1}{1-(j-1)z}.$$

Thus, by partial fraction expansion, the mean number of records in $\mathcal{W}_n$ (whose cardinality is r^n) has the exact value

$$\mathbb{E}_{\mathcal{W}_n}(\text{\# records}) = \mathrm{H}_r - \sum_{j=1}^{r-1}\frac{(j/r)^n}{r-j}. \tag{42}$$

There appears the harmonic number H_r, like in the permutation case, but now with a negative correction term which, for fixed r, vanishes exponentially fast with n (this betrays the fact that some letters from the alphabet might be missing). . . . . END OF EXAMPLE 15. $\square$

EXAMPLE 16. *Weighted word models and Bernoulli trials.* Let $\mathcal{A} = \{a_1,\dots,a_r\}$ be an alphabet of cardinality r, and let $\Lambda = \{\lambda_1,\dots,\lambda_r\}$ be a system of numbers called ***weights***, where weight λ_j is viewed as attached to letter a_j. Weights may be extended from letters to words multiplicatively by defining the weight $\pi(w)$ of word w as

$$\begin{aligned} \pi(w) &= \lambda_{i_1}\lambda_{i_2}\cdots\lambda_{i_n} \qquad \text{if} \quad w = a_{i_1}a_{i_2}\cdots a_{i_n} \\ &= \prod_{j=1}^{r}\lambda_j^{\chi_j(w)}, \end{aligned}$$

where $\chi_j(w)$ is the number of occurrences of letter a_j in w. Finally, the weight of a set is by definition the *sum* of the weights of its elements.

Combinatorially, weights of sets are immediately obtained once the corresponding generating function is known. Indeed, let $\mathcal{S} \subseteq \mathcal{W} = \text{SEQ}\{\mathcal{A}\}$ have complete GF

$$S(z, u_1, \ldots, u_r) = \sum_{w \in S} z^{|w|} u_1^{\chi_1(w)} \cdots u_r^{\chi_r(w)},$$

where $\chi_j(w)$ is the number of occurrences of letter a_j in w. Then one has

$$S(z, \lambda_1, \ldots, \lambda_r) = \sum_{w \in S} z^{|w|} \pi(w),$$

so that extracting the coefficient of z^n gives the total weight of $\mathcal{S}_n = \mathcal{S} \cap \mathcal{W}_n$ under the weight system Λ. In other words, *the GF of a weighted set is obtained by substitution of the numerical values of the weights inside the associated complete MGF.*

In probability theory, Bernoulli trials refer to sequences of independent draws from a fixed distribution with finitely many possible values. One may think of the succession of flippings of a coin or castings of a die. If any trial has r possible outcomes, then the various possibilities can be described by letters of the r-ary alphabet $\mathcal{A}$. If the probability of the jth outcome is taken to be λ_j, then the Λ-weighted models on words becomes the usual probabilistic model of independent trials. (In this situation, the λ_j's are often written as p_j's.) Observe that, in the probabilistic situation, one must have $\lambda_1 + \cdots + \lambda_r = 1$ with each λ_j satisfying $0 \leq \lambda_j \leq 1$. The equiprobable case, where each outcome has probability $1/r$ can be obtained by setting $\lambda_j = 1/r$ and it then becomes equivalent to the usual enumerative model. In terms of GFs, the coefficient $[z^n]S(z, \lambda_1, \ldots, \lambda_r)$ then represents the probability that a random word of $\mathcal{W}_n$ belongs to $\mathcal{S}$. Multivariate generating functions and cumulative generating functions then obey properties similar to their usual (ordinary, exponential) counterparts.

As an illustration, assume one has a biased coin with probability p for heads (H) and $q = 1 - p$ for tails (T). Consider the event: "*in n tosses of the coin, there never appear ℓ contiguous heads*". The alphabet is $\mathcal{A} = \{H, T\}$. The construction describing the events of interest is, as seen in Chapter I,

$$\mathcal{S} = \text{SEQ}_{<\ell}\{H\}\, \text{SEQ}\{T\, \text{SEQ}_{<\ell}\{H\}\}.$$

Its GF with u marking heads and v marking tails is then

$$W(z, u, v) = \frac{1 - z^\ell u^\ell}{1 - zu} \left(1 - zv \frac{1 - z^\ell u^\ell}{1 - zu}\right)^{-1}.$$

Thus, the probability of the absence of ℓ-runs amongst a sequence of n random coin tosses is obtained after the substitution $u \to p$, $v \to q$ in the MGF,

$$[z^n] \frac{1 - p^\ell z^\ell}{1 - z + qp^\ell z^{\ell+1}},$$

leading to an expression which is amenable to numerical or asymptotic analysis. Feller's book [**134**, p. 322–326] offers for instance a classical discussion of the problem. END OF EXAMPLE 16. $\square$

EXAMPLE 17. *Records in Bernoulli trials.* To conclude the discussion of probabilistic models on words, we come back to the analysis of records. Assume now that the alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$ has in all generality the probability p_j associated with the letter a_j. The mean number of records is analysed by a process entirely parallel to the derivation of (42): one finds by logarithmic differentiation of (41)

$$(43) \quad \mathbb{E}_{\mathcal{W}_n}(\#\,\text{records}) = [z^n]\Omega(z) \quad \text{where} \quad \Omega(z) = \frac{z}{1 - z} \sum_{j=1}^{r} \frac{p_j}{1 - z(p_1 + \cdots + p_{j-1})}.$$

The cumulative GF $\Omega(z)$ in (43) has simple poles at the points $1, 1/P_{r-1}, 1/P_{r-2}$, and so on, where $P_s = p_1 + \cdots + p_s$. For asymptotic purposes, only the dominant poles at $z = 1$ counts (see Chapter IV for a systematic discussion), near which

$$\Omega(z) \underset{z\to 1}{\sim} \frac{1}{1-z} \sum_{j=1}^{r} \frac{p_j}{1-P_{j-1}}.$$

Consequently, one has an elegant asymptotic formula generalizing the case of permutations that has a harmonic mean (10):

The mean number of records in a random word of length n with nonuniform letter probabilities p_j satisfies asymptotically $(n \to +\infty)$

$$\mathbb{E}_{\mathcal{W}_n}(\text{\# records}) \sim \sum_{j=1}^{r} \frac{p_j}{p_j + p_{j+1} + \cdots + p_r}.$$

This relation and similar ones were obtained by Burge [**69**]; analogous ideas may serve to analyse the sorting algorithm *Quicksort* under equal keys [**380**] as well as the hybrid data structures of Bentley and Sedgewick; see [**36, 81**]. . . . . END OF EXAMPLE 17. □

Coupon collector problem and birthday paradox. Similar considerations apply to weighted EGFs of words, as considered in Chapter II. For instance, the probability of having attained a complete coupon collection at time n in case a company issues coupon j with probability p_j, for $1 \le j \le r$, is (coupon collector problem, Chapter II)

$$\mathbb{P}(C \le n) = n![z^n] \prod_{j=1}^{r} \left(e^{p_j z} - 1\right).$$

The probability that all coupons are different at time n is (birthday paradox, Chapter II)

$$\mathbb{P}(B > n) = n![z^n] \prod_{j=1}^{r} \left(1 + p_j z\right),$$

which corresponds to the birthday problem in the case of nonuniform mating periods. Integral representations comparable to the ones of Chapter II are also available:

$$\mathbb{E}(C) = \int_0^\infty \left(1 - \prod_{j=1}^{r} (1 - e^{-p_i t})\right) dt, \qquad \mathbb{E}(B) = \int_0^\infty \prod_{j=1}^{r} \left(1 + p_j t\right) e^{-t}\, dt.$$

See the study by Flajolet, Gardy, and Thimonier [**151**] for several variations on this theme.

▷ **25.** *Birthday paradox with leap years.* Assume that the 29th of February exists precisely once every fourth year. Estimate the effect on the expectation of the first birthday collision. ◁

EXAMPLE 18. *Rises in Bernoulli trials: Simon Newcomb's problem.* Simon Newcomb (1835–1909), otherwise famous for his astronomical work, was reportedly fond of playing the following patience game: one draws from a deck of 52 playing cards, stacking them in piles in such a way that one new pile is started each time a card appears whose number is smaller than its predecessor. What is the probability of obtaining t piles? A solution to this famous problem is found in MacMahon's book [**306**] and a concise account by Andrews appears in [**10**, §4.4].

Simon Newcomb's problem can be rephrased in terms of rises. Given a word $w = w_1 \cdots w_n$ over the alphabet $\mathcal{A}$ ordered by $a_1 < a_2 < \cdots$, a *weak rise* is a position $j < n$ such that $w_j \le w_{j+1}$. (The numbers of piles in Newcomb's problem is the number of cards minus 1 minus the number of rises.) Let $W(z, v, \mathbf{u})$ be the MGF of all words where z marks length, v marks the number of weak rises, and u_j marks the number of occurrences of letter j. Set $z_j = zu_j$ and let $W_j(z, v, \mathbf{u})$ be the MGF relative to those nonempty words that start with letter a_j, so that

$$W = 1 + (z_1 W_1 + \cdots + z_r W_r).$$

The W_j satisfy the set of equations ($j = 1, \ldots, r$),

$$W_j = z_j + z_j \left(W_1 + \cdots + W_{j-1}\right) + v z_j \left(W_j + \cdots + W_r\right), \tag{44}$$

as seen by considering the first letter of each word. The linear system (44) is easily solved upon setting $W_j = z_j X_j$. Indeed, by differencing, one finds that

$$X_{j+1} - X_j = z_j X_j (1 - v), \qquad X_{j+1} = X_j (1 + z_j(1 - v)), \tag{45}$$

In this way, each X_j can be determined in terms of X_1. Then transporting the resulting expressions into the relation (44) instantiated at $j = 1$, and solving for X_1 leads to an expression for X_1, hence for all the X_j and finally for W itself:

$$W = \frac{v - 1}{v - P^{-1}}, \qquad P := \prod_{j=1}^{r} (1 + (1 - v) z_j). \tag{46}$$

Goulden and Jackson provide a similar looking expressions in [**208**] (pp. 72 and 236).

The result of (46) gives access to moments (e.g., mean and variance) of the number of rises in a Bernoulli sequence as well as to counting results, once coefficients of the MGF are extracted. (See also [**199, 208**] for some of the possible tools from the theory of symmetric functions.) The OGF (46) can alternatively be derived by an inclusion-exclusion argument: refer to the particular case of rises in permutations and Eulerian numbers which is discussed below. .... END OF EXAMPLE 18. □

▷ **26.** *The final solution to Simon Newcomb's problem.* Consider a deck of cards with a suits and r distinct card values. Set $N = ra$. (The original problem has $r = 13$, $a = 4$, $N = 52$.) One has from (46): $W = (v-1)P/(1-vP)^{-1}$. The expansion of $(1-y)^{-1}$ and the collection of coefficients yields

$$[z_1^a \cdots z_r^a] W = (1 - v) \sum_{k \ge 1} v^{k-1} [z_1^a \cdots z_r^a] P^k = (1 - v)^{N+1} \sum_{k \ge 1} \binom{k}{a}^r v^{k-1},$$

so that $[z_1^a \cdots z_r^a v^t] W = \displaystyle\sum_{k=0}^{t+1} (-1)^{t+1-k} \binom{N+1}{t+1-k} \binom{k}{a}^r.$ ◁

III. 6.2. Tree models. We examine here two important GFs associated with tree models; these provide valuable information concerning the *degree profile* and the *level profile* of trees, while being tightly coupled with an important class of stochastic processes, namely the *branching processes*.

The major classes of trees that we have encountered so far are the unlabelled plane trees and the labelled nonplane trees, prototypes being the general Catalan trees

(Chapter I) and the Cayley trees (Chapter II). In both cases, the counting generating functions satisfy a relation of the form

$$Y(z) = z\phi(Y(z)), \tag{47}$$

where the GF is either ordinary (plane unlabelled trees) or exponential (nonplane labelled trees). Corresponding respectively to the two cases, the function ϕ is determined by

$$\phi(w) = \sum_{\omega\in\Omega} u^{\omega}, \qquad \phi(w) = \sum_{\omega\in\Omega} \frac{u^{\omega}}{\omega!}, \tag{48}$$

where $\Omega \subseteq \mathbb{N}$ is the set of allowed node degrees. Meir and Moon in an important paper [**312**] have described some common properties of tree families that are determined by the Axiom (47). (For instance mean path length is invariably of order $n\sqrt{n}$, see Chapter VII, and height is $O(\sqrt{n})$.) Following these authors, we call *simple variety of trees* any class whose counting GF is defined by an equation of type (47). For each of the two cases of (48), we write

$$\phi(w) = \sum_{j=0}^{\infty} \phi_j w^j. \tag{49}$$

Degree profile of trees. First we examine the *degree profile* of trees. Such a profile is determined by the collection of parameters χ_j, where $\chi_j(\tau)$ is the number of nodes of outdegree j in τ. The variable u_j will be used to mark χ_j, that is, nodes of outdegree j. The discussion already conducted regarding recursive parameters shows that the GF $Y(z,\mathbf{u})$ satisfies the equation

$$Y(z,\mathbf{u}) = z\Phi(Y(z,\mathbf{u})) \qquad \text{where} \quad \Phi(w) = u_0\phi_0 + u_1\phi_1 w + u_2\phi_2 w^2 + \cdots .$$

Formal Lagrange inversion can then be applied to $Y(z,\mathbf{u})$, to the effect that its coefficients are given by the coefficients of the powers of Φ.

PROPOSITION III.7 (Degree profile of trees). *The number of trees of size n and degree profile $(n_0, n_1, n_2, \ldots)$ in a simple variety of trees defined by the "generator"* (49) *is*

$$Y_{n;n_0,n_1,n_2,\ldots} = \omega_n \cdot \frac{1}{n}\binom{n}{n_0, n_1, n_2, \ldots} \phi_0^{n_0}\phi_1^{n_1}\phi_2^{n_2}\cdots . \tag{50}$$

There, $\omega_n = 1$ in the unlabelled case, whereas $\omega_n = n!$ in the labelled case. The values of the n_j are assumed to satisfy the two consistency conditions: $\sum_j n_j = n$ and $\sum_j jn_j = n-1$.

PROOF. The consistency conditions translate the fact that the total number of nodes should be n while the total number of edges should equal $n-1$ (each node of degree j is the originator of j edges). The result follows from Lagrange inversion

$$Y_{n;n_0,n_1,n_2,\ldots} = \omega_n \cdot [u_0^{n_0}u_1^{n_1}u_2^{n_2}\cdots]\left(\frac{1}{n}[w^{n-1}]\Phi(w)^n\right),$$

to which a standard multinomial expansion applies, yielding (50).

For instance, for general Catalan trees ($\phi_j = 1$) and for Cayley trees ($\phi_j = 1/j!$) these formulæ become

$$\frac{1}{n}\binom{n}{n_0, n_1, n_2, \ldots} \qquad \text{and} \qquad \frac{(n-1)!}{0!^{n_0}1!^{n_1}2!^{n_2}\cdots}\binom{n}{n_0, n_1, n_2, \ldots}.$$

□

The proof above also reveals the logical equivalence between the general tree counting result of Proposition III.7 and the most general case of Lagrange inversion. (This results from the fact that Φ can be specialized to any particular series.) Put otherwise, any direct proof of (50) provides a combinatorial proof of the Lagrange inversion theorem. Such direct derivations have been proposed by Raney [**359**] and are based on simple but cunning surgery performed on lattice path representations of trees (the "conjugation principle" of which a particular case is the "cycle lemma" of Dvoretzky–Motzkin [**120**]).

Level profile of trees. The next example demonstrates the utility of complete generating functions for investigating the level profile of trees.

EXAMPLE 19. *Trees and level profile.* Given a rooted tree τ, its *level profile* is defined as the vector $(n_0, n_1, n_2, \ldots)$ where n_j is the number of nodes present at level j (i.e., at distance j from the root) in tree τ. Continuing within the framework of a simple variety of trees, we now define the quantity $Y_{n;n_0,n_1,n_2}$ to be the number of trees with size n and level profile given by the n_j. The corresponding complete GF $Y(z, \mathbf{u})$ with z marking size and u_j marking nodes at level j is expressible in terms of the fundamental "generator" ϕ:

$$Y(z, \mathbf{u}) = zu_0\phi\left(zu_1\phi\left(zu_2\phi\left(zu_3\phi(\cdots)\right)\right)\right). \tag{51}$$

We may call this a "continued ϕ-form". For instance general Catalan trees have generator $\phi(w) = (1-w)^{-1}$, so that in this case the complete GF is the continued fraction:

$$Y(z, \mathbf{u}) = \cfrac{u_0 z}{1 - \cfrac{u_1 z}{1 - \cfrac{u_2 z}{1 - \cfrac{u_3 z}{\ddots}}}}. \tag{52}$$

(See Section V. 3 for complementary aspects.) In contrast, Cayley trees are generated by $\phi(w) = e^w$, so that

$$Y(z, \mathbf{u}) = zu_0 e^{zu_1 e^{zu_2 e^{zu_3 e^{\cdot^{\cdot^{\cdot}}}}}},$$

which is a "continued exponential", that is, a tower of exponentials. Expanding such generating functions with respect to $u_0, u_1, \ldots$, in order gives straightforwardly:

PROPOSITION III.8 (Level profile of trees). *The number of trees of size n and level profile $(n_0, n_1, n_2, \ldots)$ in a simple variety of trees defined by the "generator" $\phi(w)$ of* (49) *is*

$$Y_{n;n_0,n_1,n_2,\ldots} = \omega_{n-1}\cdot \phi_{n_1}^{(n_0)}\phi_{n_2}^{(n_1)}\phi_{n_3}^{(n_2)}\cdots \qquad \textit{where} \quad \phi_\nu^{(\mu)} := [w^\nu]\phi(w)^\mu.$$

There, the consistency conditions are $n_0 = 1$ and $\sum_j n_j = n$. In particular, the counts for general Catalan trees and for Cayley trees are respectively

$$\binom{n_0+n_1-1}{n_1}\binom{n_1+n_2-1}{n_2}\binom{n_2+n_3-1}{n_3}\cdots, \qquad \frac{(n-1)!}{n_0!n_1!n_2!\cdots}n_0^{n_1}n_1^{n_2}n_2^{n_3}\cdots.$$

(Note that one must always have $n_0 = 1$ for a single tree; the general formula with $n_0 \neq 1$ and ω_{n-1} replaced by ω_{n-n_0} gives the level profile of forests.) The first of these enumerative results is due to Flajolet [**139**] and it places itself within a general combinatorial theory of continued fractions (Chapter V); the second one is due to Rényi and Szekeres [**361**] who developed such a formula in the context of a deep study of the distribution of height in random Cayley trees. END OF EXAMPLE 19. □

▷ **27.** *Continued forms for path length.* The BGF of path length are obtained from the level profile MGF by means of the substitution $u_j \mapsto q^j$. For general Catalan trees and Cayley trees, this gives

$$G(z,q) = \cfrac{z}{1-\cfrac{zq}{1-\cfrac{zq^2}{\ddots}}}, \qquad T(z,q) = ze^{zqe^{zq^2e^{\cdot^{\cdot^{\cdot}}}}}, \tag{53}$$

where q marks path length. The MGFs are ordinary and exponential respectively. (Combined with differentiation, such MGFs represent an attractive option for mean value analysis.) ◁

Trees and processes. The next example is an especially important application of complete GFs, as these GFs provide a bridge between combinatorial models and a major class of stochastic processes, the *branching processes* of probability theory.

EXAMPLE 20. *Weighted tree models and branching processes.* Consider the family $\mathcal{G}$ of all general plane trees. Let $\Lambda = (\lambda_0, \lambda_1, \ldots)$ be a system of numeric weights. The weight of a node of outdegree j is taken to be λ_j and the weight of a tree is the product of the individual weights of its nodes:

$$\pi(\tau) = \prod_{j=0}^{\infty} \lambda_j^{\chi_j(\tau)}, \tag{54}$$

with $\chi_j(\tau)$ the number of nodes of degree j in τ. One can view the weighted model of trees as a model in which a tree receives a probability proportional to $\pi(w)$. Precisely, the probability of selecting a particular tree τ under this model is, for a fixed size n

$$\mathbb{P}_{\mathcal{G}_n,\Lambda}(\tau) = \frac{\pi(\tau)}{\sum_{|\tau|=n}\pi(\tau)}. \tag{55}$$

This defines a probability measure over the set $\mathcal{G}_n$ and one can consider events and random variables under this weighted model.

The weighted model defined by (54) and (55) covers any simple variety of trees: just replace each λ_j by the quantity ϕ_j given by the "generator' (49) of the model. For instance, plane unlabelled unary-binary trees are obtained by $\Lambda = (1, 1, 1, 0, 0, \ldots)$, while Cayley trees correspond to $\lambda_j = 1/j!$. Two *equivalence-preserving transformations* are then especially important in this context:

(i) Let Λ^* be defined by $\lambda_j^* = c\lambda_j$ for some nonzero constant c. Then the weight corresponding to Λ^* satisfies $\pi^*(\tau) = c^{|\tau|}\pi(w)$. Consequently, the models associated to Λ and Λ^* are equivalent as regards (55).

(ii) Let Λ° be defined by $\lambda_j^\circ = \theta^j\lambda_j$ for some nonzero constant θ. Then the weight corresponding to Λ° satisfies $\pi^\circ(\tau) = c^{|\tau|-1}\pi(w)$, since $\sum_j j\chi_j(\tau) = |\tau| - 1$ for any tree τ. Thus the models Λ° and Λ are again equivalent.

Each transformation has a simple effect on the generator ϕ, namely:

$$\phi(w) \mapsto \phi^*(w) = c\phi(w) \qquad \text{and} \qquad \phi(w) \mapsto \phi^\circ(w) = \phi(\theta w). \tag{56}$$

Once equipped with such equivalence transformations, it becomes possible to describe probabilistically the process that generates trees according to a weighted model. Assume that $\lambda_j \geq 0$ and that the λ_j are summable. Then the normalized quantities

$$p_j = \frac{\lambda_j}{\sum_j \lambda_j}$$

form a probability distribution over $\mathbb{N}$. By the first equivalence-preserving transformation the model induced by the weights p_j is the same as the original model induced by the λ_j. (By the second equivalence transformation, one can furthermore assume that the generator ϕ is the probability generating function of the p_j.)

Such a model defined by nonnegative weights $\{p_j\}$ summing to 1 is nothing but the classical model of *branching processes* (also known as Galton-Watson processes) ; see [**17**]. In effect, a realization T of the branching process is classically defined by the two rules: (i) produce a root node of degree j with probability p_j; (ii) if $j \geq 1$, attach to the root node a collection $T_1, \ldots, T_j$ of independent realizations of the process. This may be viewed as the development of a "family" stemming from a common ancestor where any individual has probability p_j of giving birth to j descendants. Clearly, the probability of obtaining a particular finite tree τ has probability $\pi(\tau)$, where π is given by (54) and the weights are $\lambda_j = p_j$. The generator

$$\phi(w) = \sum_{j=0}^{\infty} p_j w^j$$

is then nothing but the probability generating function of (one-generation) offspring, with the quantity $\mu = \phi'(1)$ being its mean size.

For the record, we recall that branching processes can be classified into three categories depending on the values of μ:

Subcriticality: when $\mu < 1$, the random tree produced is finite with probability 1 and its expected size is also finite.
Criticality: when $\mu = 1$, the random tree produced is finite with probability 1 but its expected size is infinite.
Supercriticality: when $\mu > 1$, the random tree produced is finite with probability strictly less than 1.

From the discussion of equivalence transformations (56), there furthermore results that, regarding trees of a *fixed size* n, there is complete equivalence between all branching processes with generators of the form

$$\phi_\theta(w) = \frac{\phi(\theta w)}{\phi(\theta)}.$$

Such families of related functions are known as "exponential families" in probability theory. In this way, one may always regard at will the random tree produced by a weighted model of some fixed size n as originating from a branching process of subcritical, critical, or supercritical type conditioned upon the size of the total progeny.

Finally, take a set $\mathcal{S} \subseteq \mathcal{G}$ for which the complete generating function of $\mathcal{S}$ with respect to the degree profile is available,

$$S(z, u_0, u_1, \ldots) = \sum_{\tau \in \mathcal{S}} z^{|\tau|} \left(u_0^{\chi_0(\tau)} u_1^{\chi_1(\tau)} \cdots \right).$$

Then, for a system of weights Λ, one has

$$S(z, \lambda_0, \lambda_1, \ldots) = \sum_{\tau \in \mathcal{S}} \pi(\tau) z^{|\tau|}.$$

Thus, the probability that a weighted tree of size n belongs to $\mathcal{S}$ becomes accessible by extracting the coefficient of z^n. This applies *a fortiori* to branching processes as well. In summary, *the analysis of parameters of trees of size n under either weighted models or branching process models derives from substituting weights or probability values inside the corresponding combinatorial generating functions.* .... END OF EXAMPLE 20. □

The reduction of combinatorial tree models to branching processes has been pursued early, most notably by the "Russian School": see especially the books by Kolchin [**276, 277**] and references therein. (For asymptotic purposes, the equivalence between combinatorial models and critical branching processes often turns out to be most fruitful.) Conversely, symbolic-combinatorial methods may be viewed as a systematic way of obtaining equations relative to characteristics of branching processes. We do not elaborate further along these lines as this would take us outside of the scope of the present book.

▷ **28.** *Catalan trees, Cayley trees, and branching processes.* Catalan trees of size n are defined by the weighted model in which $\lambda_j \equiv 1$, but also equivalently by $\widehat{\lambda}_j = c\theta^j$, for any $c > 0$ and $\theta \leq 1$. In particular they coincide with the random tree produced by the critical branching process whose offspring probabilities are geometric: $p_j = 1/2^{j+1}$.

Cayley trees are *a priori* defined by $\lambda_j = 1/j!$. They can be generated by the critical branching process with Poisson probabilities, $p_j = e^{-1}/j!$, and more generally with an arbitrary Poisson distribution $p_j = e^{-\lambda}\lambda^j/j!$. ◁

III. 7. Additional constructions

We discuss here additional constructions already examined in earlier chapters, namely pointing and substitution (Section III. 7.1) as well as order constraints (Section III. 7.2) on the one hand, implicit structures (Section III. 7.3) on the other hand. Given the that basic translation mechanisms can be directly adapted to the multivariate realm, such extensions involve basically no new concept and the methods of Chapters I and II can be recycled. In Section III. 7.4, we revisit the classical principle of inclusion-exclusion under a generating function perspective. In this light, the principle appears as a typically multivariate device well-suited to enumerating objects according the number of occurrences of sub-configurations.

III. 7.1. Pointing and substitution. Let $\langle \mathcal{F}, \chi \rangle$ be a class–parameter pair, where χ is multivariate of dimension $r \geq 1$ and let $F(\mathbf{z})$ be the MGF associated to it in the notations of (18) and (27). In particular $z_0 = z$ marks size, and z_k marks the component j of the multiparameter k. If z marks size, then, like in the univariate case, θ_z translates the fact of distinguishing one atom. Generally, pick up a variable $x \equiv z_j$ for some j with $0 \leq j \leq r$. Then since

$$x\partial_x(s^a t^b x^f) = f \cdot (s^a t^b x^f),$$

the interpretation of the operator $\theta_x \equiv x\partial_x$ is immediate; it means "pick up in all possible ways in objects of $\mathcal{F}$ a configuration marked by x and point to it". For instance, if $F(z, u)$ is the BGF of trees where z marks size and u marks leaves, then $\theta_u F(z, u) = u\partial_u F(z, u)$ enumerates trees with one distinguished leaf.

Similarly, the substitution $x \mapsto S(\mathbf{z})$ in a GF F, where $S(\mathbf{z})$ is the MGF of a class $\mathcal{S}$, means attaching an object of type $\mathcal{S}$ to configurations marked by the variable x in $\mathcal{F}$. We refrain from giving detailed definitions (that would be somewhat clumsy and uninformative) as the process is better understood by practice than by long formal developments. Justification in each particular case is easily obtained by returning to the combinatorial representation of generating functions as images of combinatorial classes.

EXAMPLE 21. *Constrained integer compositions and "slicing".* This example illustrates variations around the substitution scheme. Consider compositions of integers where successive summands have sizes that are constrained to belong to a fixed set $\mathcal{R} \subseteq \mathbb{N}^2$. For instance, the relations

$$\mathcal{R}_1 = \{(x, y) \mid 1 \leq x \leq y\}, \quad \mathcal{R}_2 = \{(x, y) \mid 1 \leq y \leq 2x\},$$

correspond to weakly increasing summands in the case of $\mathcal{R}_1$ and to summands that can at most double at each stage in the case of $\mathcal{R}_2$. In the "ragged landscape" representation of compositions, this means considering diagrams of unit cells aligned in columns along the horizontal axis, with successive columns obeying the constraint imposed by $\mathcal{R}$.

Let $F(z, u)$ be the BGF of such $\mathcal{R}$–restricted compositions, where z marks total sum and u marks the value of the last summand, that is, the height of the last column. The function $F(z, u)$ satisfies a functional equation of the form

$$F(z, u) = f(zu) + \left(\mathcal{L}\left[F(z, u)\right]\right)_{u \mapsto zu}, \tag{57}$$

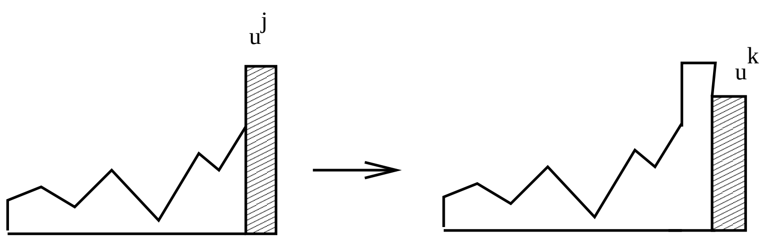


FIGURE 14. The technique of "adding a slice" for enumerating constrained compositions.

where $f(z)$ is the generating function of the one-column objects and $\mathcal{L}$ is a linear operator over formal series in u given by

$$\mathcal{L}[u^j] := \sum_{(j,k)\in\mathcal{R}} u^k. \tag{58}$$

In effect, Equation (57) describes inductively objects as comprising either one column ($f(zu)$) or else being formed by adding a new column to an existing one. In the latter case, the last column added has a size k that must be such that $(j,k)\in\mathcal{R}$, if it was added after a column of size j, and it will contribute $u^k z^k$ to the BGF $F(z,u)$; this is precisely what (58) expresses. In particular, $F(z,1)$ gives back the enumeration of $\mathcal{F}$–objects irrespective of the size of the last column.

For a rule $\mathcal{R}$ that is "simple enough", the basic equation (57) will often involve a substitution. Let us first rederive in this way the enumeration of partitions. We take $\mathcal{R}=\mathcal{R}_1$ and assume that the first column can have any positive size. Compositions into increasing summands are clearly the same as partitions. Since

$$\mathcal{L}[u^j] = u^j + u^{j+1} + u^{j+2} + \cdots = \frac{u^j}{1-u},$$

the function $F(z,u)$ satisfies a functional equation involving a substitution,

$$F(z,u) = \frac{zu}{1-zu} + \frac{1}{1-zu}F(z,zu). \tag{59}$$

This relation iterates: *any linear functional equation of the substitution type*

$$\phi(u) = \alpha(u) + \beta(u)\phi(\sigma(u))$$

is solved formally by

$$\phi(u) = \alpha(u) + \beta(u)\alpha(\sigma(u)) + \beta(u)\beta(\sigma(u))\alpha(\sigma^{\langle 2\rangle}(u)) + \cdots, \tag{60}$$

where $\sigma^{\langle j\rangle}(u)$ designates the jth iterate of u.

Returning to compositions into increasing summands, that is, partitions, the turnkey solution (60) gives, upon iterating on the second argument with the first argument treated as a parameter:

$$F(z,u) = \frac{zu}{1-zu} + \frac{z^2u}{(1-zu)(1-z^2u)} + \frac{z^3u}{(1-zu)(1-z^2u)(1-z^3u)} + \cdots. \tag{61}$$

Equivalence with the alternative form

$$F(z,u) = \frac{zu}{1-z} + \frac{z^2u^2}{(1-z)(1-z^2)} + \frac{z^3u^3}{(1-z)(1-z^2)(1-z^3)} + \cdots \tag{62}$$

is then easily verified from (59) upon expanding $F(z,u)$ as a series in u and applying the method of indeterminate coefficients to the form $(1-zu)F(z,u) = zu + F(z,zu)$. The presentation (62) is furthermore consistent with the treatment of partitions given in Chapter I since the quantity $[u^k]F(z,u)$ clearly represents the OGF of (nonempty) partitions whose largest summand is k. (In passing, the equality between (61) and (62) is a shallow but curious identity that is quite typical of the area of q-analogues.)

This same method has been applied in [**169**] to compositions satisfying condition $\mathcal{R}_2$ above. In this case, successive summands are allowed to double at most at each stage. The associated linear operator is

$$\mathcal{L}[u^j] = u + \cdots + u^{2j} = u\frac{1-u^{2j}}{1-u}.$$

For simplicity, it is assumed that the first column has size 1. Thus, F satisfies a functional equation of the substitution type:

$$F(z,u) = zu + \frac{zu}{1-zu}\left(F(z,1) - F(z,z^2u^2)\right).$$

This can be solved by means of the general iteration mechanism (60), treating momentarily $F(z,1)$ as a known quantity: with $a(u) := zu + F(z,1)/(1-zu)$, one has

$$F(z,u) = a(u) - \frac{zu}{1-zu}a(z^2u^2) + \frac{zu}{1-zu}\frac{z^2u^2}{1-z^2u^2}a(z^6u^4) - \cdots.$$

Then, the substitution $u = 1$ in the solution becomes permissible. Upon solving for $F(z,1)$, one eventually gets the somewhat curious GF for compositions satisfying $\mathcal{R}_2$:

$$\begin{aligned} F(z,1) \quad &= \quad \frac{\sum_{j\geq 1}(-1)^{j-1}z^{2^{j+1}-j-2}/Q_{j-1}(z)}{\sum_{j\geq 0}(-1)^{j}z^{2^{j+1}-j-2}/Q_{j}(z)} \\ &\qquad \text{where} \quad Q_j(z) = (1-z)(1-z^3)(1-z^7)\cdots(1-z^{2^j-1}). \end{aligned}$$

The sequence of coefficients starts as $1,1,2,3,5,9,16,28,50$ and is *EIS* **A002572**: it represents for instance the number of possible level profiles of binary trees, or equivalently the number of partitions of 1 into summands of the form $1, \frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ (this is related to the number of solutions to Kraft's inequality). See [**169**] for details including very precise asymptotic estimates and Tangora's paper [**406**] for relations to algebraic topology. END OF EXAMPLE 21. □

The reason for presenting the slicing method in some detail is that it is very general. It has been in particular employed to derive a number of original enumerations of polyominoes by area, a topic of interest in some branches of statistical mechanics: for instance, the book by Janse van Rensburg [**423**] discusses many applications of such lattice models to polymers and vesicles. See Bousquet-Mélou's review paper [**62**] for a methodological perspective. Some of the origins of the method point to Pólya in the 1930's, see [**348**], and independently to Temperley [**408**, pp. 65–67].

▷ **29.** *Pointing-erasing and the combinatorics of Taylor's formula.* The derivative operator ∂_x corresponds combinatorially to a "pointing-erasing" operation: select in all possible ways an atom marked by x and make it transparent to x-marking (e.g., by replacing it by a neutral object). The operator

$$\mu^k[f](x) := \frac{1}{k!}\partial_x^k f(x),$$

then corresponds to picking up in all possible way a subset of k configurations marked by x and unmarking them. The identity (Taylor's formula)

$$f(x+y) = \sum_{k\geq 0}\left(\frac{1}{k!}\partial_x^k f(x)\right)y^k$$

can then receive a simple combinatorial interpretation: Given a population of individuals ($\mathcal{F}$ enumerated by f), form the bicoloured population of individuals enumerated by $f(x+y)$, where each atom of each object can be repainted either in x-colour or y-colour; this is equivalent to deciding a priori for each individual to repaint k of its atoms from x to y, this for all possible values of $k \geq 0$. Taylor's formula follows. ◁

▷ **30.** *Carlitz compositions I.* Let $\mathcal{K}$ be the class of compositions such that all pairs of adjacent summands are formed of distinct values. These can be generated by the operator $\mathcal{L}[u^j] =$

valley:	$\sigma_{i-1} > \sigma_i < \sigma_{i+1}$	leaf node (u_0)
double rise:	$\sigma_{i-1} < \sigma_i < \sigma_{i+1}$	unary right-branching (u_1)
double fall:	$\sigma_{i-1} > \sigma_i > \sigma_{i+1}$	unary left-branching (u_1')
peak:	$\sigma_{i-1} < \sigma_i > \sigma_{i+1}$	binary node (u_2)

FIGURE 15. Local order patters in a permutation and the four types of nodes in the corresponding increasing binary tree.

$\frac{uz}{1-uz} - u^j z^j$, so that $L[f(u)] = \frac{uz}{1-uz} f(1) - f(uz)$; The BGF $K(z,u)$, with u marking the value of the last summand, then satisfies a functional equation,

$$K(z,u) = \frac{uz}{1-uz} + \frac{uz}{1-uz} K(z,1) - K(z,zu),$$

giving eventually $K(z) \equiv K(z,1)$ under the form

$$\begin{aligned} K(z) &= \left(1 + \sum_{j\geq 1} \frac{(-z)^j}{1-z^j}\right)^{-1} \\ &= \ = 1 + z + z^2 + 3z^3 + 4z^4 + 7z^5 + 14z^6 + 23z^7 + 39z^8 + \cdots . \end{aligned} \tag{63}$$

The sequence of coefficients constitutes *EIS* **A003242**. Such compositions have been introduced by Carlitz in 1976; the derivation above is from a paper by Knopfmacher and Prodinger [**258**] who provide early references and asymptotic properties. (We resume this thread in Note 32 below and in Chapter IV, p. 250.) $\triangleleft$

III. 7.2. Order constraints. We refer in this subsection to the discussion of order constraints in labelled products that has been given in Chapter II. We recall that the modified labelled product

$$\mathcal{A} = (\mathcal{B}^\square \star \mathcal{C})$$

only includes the elements of $(\mathcal{B} \star \mathcal{C})$ such that the minimal label lies in the $\mathcal{A}$ component. Once more the univariate rules generalize verbatim for parameters that are inherited and the corresponding exponential MGFs are related by

$$A(z, \mathbf{u}) = \int_0^z (\partial_t B(t, \mathbf{u})) \cdot C(t, \mathbf{u})\, dt.$$

To illustrate this multivariate extension, we shall consider a quadrivariate statistic on permutations.

EXAMPLE 22. *Local order patterns in permutations.* An element σ_i of a permutation written $\sigma = \sigma_1, \ldots, \sigma_n$ when compared to its immediate neighbours can be categorized into one of four types[4] summarized in the first two columns of Figure 15. The correspondence with binary increasing trees described in Example 17 of Chapter II then shows the following: peaks and valleys correspond to binary nodes and leaves, respectively, while double rises and double falls are associated with right-branching and left-branching unary nodes. Let u_0, u_1, u_1', u_2 be

[4]Here, for $|\sigma| = n$, we regard σ as *bordered* by $(-\infty, -\infty)$, i.e., we set $\sigma_0 = \sigma_{n+1} = -\infty$ and let the index i in Figure 15 as varying in $[1 \,..\, n]$. Alternative bordering conventions prove occasionally useful.

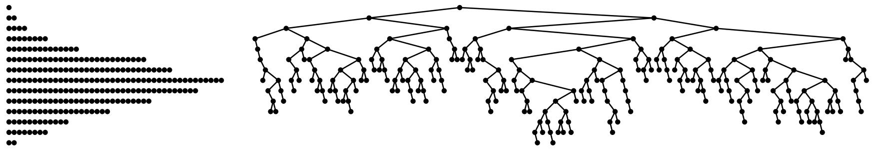


FIGURE 16. The level profile of a random increasing binary tree of size 256. (Compare with Figure 13 for binary trees under the uniform Catalan statistic.)

markers for the number of nodes of each type, as summarized in Figure 15. Then the exponential MGF of increasing trees under this statistic satisfies

$$\frac{\partial}{\partial z} I(z, \mathbf{u}) = u_0 + (u_1 + u_1')I(z, \mathbf{u}) + u_2 I(z, \mathbf{u})^2.$$

This is solved by separation of variables as

$$(64) \qquad I(z, \mathbf{u}) = \frac{\delta}{u_2} \frac{v_1 + \delta \tan(z\delta)}{\delta - v_1 \tan(z\delta)} - \frac{v_1}{u_2},$$

where the following abbreviations are used:

$$v_1 = \frac{1}{2}(u_1 + u_1'), \qquad \delta = \sqrt{u_0 u_2 - v_1^2}.$$

One has

$$I = u_0 z + u_0(u_1 + u_1')\frac{z^2}{2!} + u_0((u_1 + u_1')^2 + 2u_0u_2)\frac{z^3}{3!},$$

which agrees with the small cases. This calculation is consistent with what has been found in Chapter II regarding the EGF of all nonempty permutations and of alternating permutations,

$$\frac{z}{1-z}, \quad \tan(z),$$

that derive from the substitutions $\{u_0 = u_1 = u_1' = u_2 = 1\}$ and $\{u_0 = u_2 = 1, u_1 = u_1' = 0\}$, respectively. The substitution $\{u_0 = u_1 = u, u_1' = u_2 = 1\}$ gives a simple variant (without the empty permutation) of the BGF of Eulerian numbers (73) derived below by other means (p. 198).

By specialization of the quadrivariate GF, there results that, in a tree of size n the mean number of nodes of nullary, unary, or binary type is asymptotic to $n/3$, with a variance that is $O(n)$, thereby ensuring concentration of distribution. .... END OF EXAMPLE 22. $\square$

A similar analysis yields path length. It is found that a random increasing binary tree of size n has mean path length

$$2n \log n + O(n).$$

Contrary to what the uniform combinatorial model give, such tree tend to be rather well balanced, and a typical branch is only about 38.6% longer than in a perfect binary tree (since $2/\log 2 \doteq 1.386$). This fact applies to binary search trees (Note 31) and it justifies that the performance of such trees is quite good when they are applied to random data [**269, 307, 382**] or subjected to randomization [**367, 323**].

$\triangleright$ **31.** *Binary search trees (BSTs).* Given a permutation τ, one defines inductively a tree $\mathrm{BST}(\tau)$ by

$$\mathrm{BST}(\epsilon) = \emptyset; \qquad \mathrm{BST}(\tau) = \langle \tau_1, \mathrm{BST}(\tau|_{<\tau_1}, \mathrm{BST}(\tau|_{>\tau_1}).$$

(There, $\tau|_P$ represents the subword of τ consisting of those elements that satisfy predicate P.) Let $\mathrm{IBT}(\sigma)$ be the increasing binary tree canonically associated to σ. Then one has the fundamental *Equivalence Principle*,

$$\mathrm{IBT}(\sigma) \overset{\text{shape}}{\equiv} \mathrm{BST}(\sigma^{-1}),$$

where $A \overset{\text{shape}}{\equiv} B$ means that A and B have identical tree shapes. $\triangleleft$

III. 7.3. Implicit structures. Here again, we note that equations involving sums and products, either labelled or not, are easily solved just like in the univariate case. The same applies for the sequence construction and for the set construction, especially in the labelled case—refer to the corresponding sections of Chapters I and II. Again, the process is best understood by examples.

Suppose for instance one wants to enumerate connected labelled graphs by the number of nodes (marked by z) and the number of edges (marked by u). The class $\mathcal{K}$ of connected graphs and the class $\mathcal{G}$ of all graphs are related by the set construction,

$$\mathcal{G} = \mathrm{SET}\{\mathcal{K}\},$$

meaning that every graph decomposes uniquely into connected components. The corresponding exponential BGFs then satisfy

$$G(z,u) = e^{K(z,u)} \qquad \text{implying} \qquad K(z,u) = \log G(z,u),$$

since the number of edges in a graph is inherited (additively) from the corresponding numbers in connected components. Now, the number of graphs of size n having k edges is $\binom{n(n-1)/2}{k}$, so that

$$K(z,u) = \log\left(1 + \sum_{n=1}^{\infty} (1+u)^{n(n-1)/2} \frac{z^n}{n!}\right). \tag{65}$$

This formula, which appears as a refinement of the univariate formula of Chapter II, then simply reads: *connected graphs are obtained as components (the* $\log$ *operator) of general graphs, where a general graph is determined by the presence or absence of an edge (corresponding to* $(1+u)$*) between any pair of nodes (the exponent* $n(n-1)/2$*).*

Pulling out information out of the formula (65) is however not obvious due to the alternation of signs in the expansion of $\log(1+w)$ and due to the strongly divergent character of the involved series. As an aside, we note here that the quantity

$$\widehat{K}(z,u) = K\left(\frac{z}{u}, u\right)$$

enumerates connected graphs according to size (marked by z) and excess (marked by u) of the number of edges over the number of nodes. This means that the results of Section 5.3 of Chapter II obtained by Wright's decomposition can be rephrased as the expansion (within $\mathbb{C}(u)[\![z]\!]$):

$$
\begin{aligned}
\log\left(1+\sum_{n=1}^{\infty}(1+u)^{n(n-1)/2}\frac{z^n u^{-n}}{n!}\right) &= \frac{1}{u}W_{-1}(z)+W_0(z)+\cdots \\
&= \frac{1}{u}\left(T-\frac{1}{2}T^2\right)+\left(\frac{1}{2}\log\frac{1}{1-T}-\frac{1}{2}T-\frac{1}{4}T^2\right)+\cdots,
\end{aligned}
\tag{66}
$$

with $T\equiv T(z)$. See Temperley's early works [**407, 408**] as well as the "giant paper on the giant component" [**244**] and the paper [**172**] for direct derivations that eventually constitute analytic alternatives to Wright's combinatorial approach.

EXAMPLE 23. *Smirnov words.* Following the treatment of Goulden and Jackson [**208**], we define a Smirnov word to be any word that has no consecutive equal letters. Let $\mathcal{W}=\mathrm{SEQ}\{\mathcal{A}\}$ be the set of words over the alphabet $\mathcal{A}=\{a_1,\dots,a_r\}$ of cardinality r, and $\mathcal{X}$ be the set of Smirnov words. Let also u_j mark the number of occurrences of the jth letter in a word. One has

$$W(z,\mathbf{u})=\frac{1}{1-(v_1+\cdots+v_r)}\quad\text{with}\quad v_j=zu_j.$$

Start from a Smirnov word and substitute to any letter a_j that appears in it an arbitrary nonempty sequence of letters a_j. When this operation is done at all places of a Smirnov word, it gives rise to an unconstrained word. Conversely, any word is associated to a unique Smirnov word by collapsing into single letters maximal groups of contiguous equal letters. In other terms, words derive from Smirnov words by a simultaneous substitution:

$$\mathcal{W}=\mathcal{S}\big[a_1\mapsto\mathrm{SEQ}_{\geq 1}\{a_1\},\ \dots\ ,a_r\mapsto\mathrm{SEQ}_{\geq 1}\{a_r\}\big].$$

There results the relation

$$W(v_1,\dots,v_r)=S\left(\frac{v_1}{1-v_1},\ \dots\ ,\frac{v_r}{1-v_r}\right).\tag{67}$$

This relation determines the MGF $S(v_1,\dots,v_r)$ *implicitly*. Indeed, since the inverse function of $v/(1-v)$ is $v/(1+v)$, one finds

$$S(v_1,\dots,v_r)=W\left(\frac{v_1}{1+v_1},\ \dots\ ,\frac{v_r}{1+v_r}\right).\tag{68}$$

For instance, if we set $v_j=z$, that is, we "forget" the composition of the words into letters, we get the OGF of Smirnov word counted according to length as

$$\frac{1}{1-r\frac{z}{1+z}}=\frac{1+z}{1-(r-1)z}=1+\sum_{n\geq 1}r(r-1)^{n-1}z^n.$$

This is consistent with elementary combinatorics since a Smirnov word of length n is determined by the choice of its first letter (r possibilities) followed by a sequence of $n-1$ choices constrained to avoid one letter amongst r (and corresponding to $r-1$ possibilities for each position). The interest of (68) is to apply equally well to the Bernoulli model where letters may receive unequal probabilities and where a direct combinatorial argument does not appear to be easy: it suffices to perform the substitution $v_j\mapsto p_j z$ in this case.

From these developments, one can next build the GF of words that never contain more than m consecutive equal letters. It suffices to effect in (68) the substitution $v_j\mapsto v_j+\cdots+v_j^m$.

In particular for the univariate problem (or, equivalently, the case where letters are equiprobable), one finds the OGF

$$\frac{1}{1 - r\dfrac{z\frac{1-z^m}{1-z}}{1+z\frac{1-z^m}{1-z}}} = \frac{1-z^{m+1}}{1-rz+(r-1)z^{m+1}}.$$

This extends to an arbitrary alphabet the analysis of single runs and double runs in binary words that was performed in Section 4 of Chapter I. Naturally, this approach applies equally well to nonuniform letter probabilities and to a collection of run-length upperbounds dependent on each particular letter. For instance, this topic is pursued in several works of Karlin and coauthors (see, e.g., [**319**]), themselves motivated by biological applications. END OF EXAMPLE 23. □

▷ **32.** *Carlitz compositions II.* Here is an alternative derivation of the OGF of Carlitz compositions (Note 30, p. 190). Carlitz compositions with largest summand $\leq r$ are obtained from the OGF of Smirnov words by the substitution $v_j \mapsto z^j$:

$$K^{[r]}(z) = \left(1 - \sum_{j=1}^{r} \frac{z^j}{1+z^j}\right)^{-1}, \tag{69}$$

The OGF of all Carlitz compositions the results from letting $r \to \infty$:

$$K(z) = \left(1 - \sum_{j=1}^{\infty} \frac{z^j}{1+z^j}\right)^{-1}. \tag{70}$$

The asymptotic form of the coefficients is derived in Chapter IV, p. 250. ◁

III. 7.4. Inclusion-Exclusion. Inclusion-exclusion is a familiar type of reasoning rooted in elementary mathematics. Its principle, in order to count *exactly*, consists in grossly *overcounting*, then performing a simple correction of the overcounting, then correcting the correction, and so on. Characteristically, enumerative results provided by inclusion exclusion involve an alternating sum. We revisit this process here in the perspective of multivariate generating functions, where it essentially reduces to a combined use of substitution and implicit definitions. Our approach follows Goulden and Jackson's encyclopedic treatise [**208**].

Let $\mathcal{E}$ be a set endowed with a real or complex valued measure $|\cdot|$ in such a way that, for $A, B \subset \mathcal{E}$, there holds

$$|A \cup B| = |A| + |B| \qquad \text{whenever} \qquad A \cap B = \emptyset.$$

Thus, $|\cdot|$ is an additive measure, typically taken as set cardinality (i.e., $|e| = 1$ for $e \in E$) or a discrete probability measure on $\mathcal{E}$ (i.e., $|e| = p_e$ for $e \in E$). The general formula

$$|A \cup B| = |A| + |B| - |AB| \qquad \text{where} \quad AB := A \cap B,$$

follows immediately from basic set-theoretic principles:

$$\sum_{c \in A \cup B} |c| = \sum_{a \in A} |a| + \sum_{b \in B} |b| - \sum_{i \in A \cap B} |i|.$$

What is called the *inclusion-exclusion principle* or *sieve formula* is the following multivariate generalization, for an arbitrary family $A_1, \ldots, A_r \subset \mathcal{E}$:

(71)
$$\begin{aligned} |A_1 \cup \cdots \cup A_r| &\equiv \left|\mathcal{E} \setminus (\overline{A}_1 \overline{A}_2 \cdots \overline{A}_r)\right| \qquad \text{where} \quad \overline{A} := \mathcal{E} \setminus A \\ &= \sum_{1 \le i \le r} |A_i| - \sum_{1 \le i_1 < i_2 \le r} |A_{i_1} A_{i_2}| + \cdots + (-1)^{r-1} |A_1 A_2 \cdots A_r|. \end{aligned}$$

(The easy proof by induction results from elementary properties of the boolean algebra formed by the subsets of $\mathcal{E}$; see, e.g., [**82**, Ch. IV].) An alternative formulation results from setting $B_j = \overline{A}_j$, $\overline{B}_j = A_j$:

$$\begin{aligned} |B_1 B_2 \cdots B_r| = |\mathcal{E}| - \sum_{1 \le i \le r} |\overline{B}_i| \\ + \sum_{1 \le i_1 < i_2 \le r} |\overline{B}_{i_1} \overline{B}_{i_2}| - \cdots + (-1)^r |\overline{B}_1 \overline{B}_2 \cdots \overline{B}_r|. \end{aligned} \tag{72}$$

In terms of measure, this equality quantifies the set of objects satisfying *exactly* a collection of *simultaneous* conditions (all the B_j) in terms of those that violate *at least some* of the conditions (the $\overline{B}_j$).

Derangements. Here is a textbook example of an inclusion–exclusion argument, namely, the enumeration of *derangements*. Recall that a derangement is a permutation σ such that $\sigma_i \neq i$, for all i. Fix $\mathcal{E}$ as the set of all permutations of $[1, n]$, take the measure $|\cdot|$ to be set cardinality, and let B_i be the subset of permutations in $\mathcal{E}$ associated to the property $\sigma_i \neq i$. (There are consequently $r = n$ conditions.) Thus, B_i means having no fixed point at i, while $\overline{B}_i$ means having a fixed point at the *distinguished* value i. Then, the left hand side of (72) is the number of permutations that are derangements, that is, D_n. As regards the right hand side, the kth sum comprises itself $\binom{n}{k}$ terms counting possibilities attached to the choices of indices $i_1 < \cdots < i_k$; each such choice is associated to a factor $\overline{B}_{i_1} \cdots \overline{B}_{i_k}$ that describes all permutations with fixed points at the distinguished points $i_1, \ldots, i_k$ (i.e., $\sigma(i_1) = i_1, \ldots, \sigma_{i_k} = i_k$). Clearly, $|\overline{B}_{i_1} \cdots \overline{B}_{i_k}| = (n-k)!$. Therefore one has

$$D_n = n! - \binom{n}{1}(n-1)! + \binom{n}{2}(n-2)! - \cdots + (-1)^n \binom{n}{n} 0!,$$

which rewrites into the more familiar form

$$\frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}.$$

This gives an elementary derivation of the derangement numbers already encountered in Chapter II and obtained there by means of the labelled set and cycle constructions.

The derivation above is perfectly fine but carrying it out on complex examples may represent somewhat of a challenge. In contrast, as we now explain, there exists a parallel approach based on multivariate generating functions, which is technically easy to deal with and has great versatility.

Let us now reexamine derangements in a generating function perspective. Consider the set $\mathcal{P}$ of all permutations and build a superset $\mathcal{Q}$ as follows. The set $\mathcal{Q}$

is comprised of permutations in which an arbitrary number of fixed points—some, maybe none, not necessarily all—have been *distinguished*. (This corresponds to arbitrary products of the $\overline{B}_j$ in the argument above.). For instance $\mathcal{Q}$ contains elements like

$$\underline{1}, 3, 2, \quad 1, 3, 2, \quad \underline{1}, 2, 3, \quad 1, \underline{2}, \underline{3}, \quad \underline{1}, 2, \underline{3}, \quad \underline{1}, \underline{2}, \underline{3},$$

where distinguished fixed points are underlined. Clearly, if one removes the distinguished elements of a $\gamma \in \mathcal{Q}$, what is left constitutes an arbitrary permutation of the remaining elements. One has

$$\mathcal{Q} \cong \mathcal{U} \star \mathcal{P},$$

where $\mathcal{U}$ denotes the class of urns that are sets of atoms. In particular, the EGF of $\mathcal{Q}$ is $Q(z) = e^z/(1-z)$. What we've just done is enumerating the quantities that appear in (72), but with the signs "wrong", i.e., all pluses.

Introduce now the variable v to mark the distinguished fixed points in objects of $\mathcal{Q}$. The exponential BGF is then by general principles of this chapter:

$$Q(z, v) = e^{vz} \frac{1}{1-z}.$$

Let now $P(z, u)$ be the BGF of permutations where u marks the number of fixed points. (Let us ignore momentarily the fact that $P(z, u)$ is otherwise known.) Permutations with *some* fixed points distinguished are generated by the substitution $u \mapsto 1 + v$ inside $P(z, u)$. In other words one has the fundamental inclusion-exclusion relation

$$Q(z, v) = P(z, 1 + v).$$

This is then easily solved as

$$P(z, u) = Q(z, u - 1),$$

so that knowledge of (the easy) Q gives (the harder) P. For the case at hand, this yields

$$P(z, u) = \frac{e^{(u-1)z}}{1-z}, \qquad P(z, 0) = D(z) = \frac{e^{-z}}{1-z},$$

and, in particular, the EGF of derangements has been retrieved. Note that the sought $P(z, 0)$ comes out as $Q(z, -1)$, so that signs corresponding to the sieve formula (72) have now been put "right", i.e., alternating.

The process employed for derangements is clearly very general. It is a generating function analogue of the inclusion-exclusion principle: counting objects that satisfy a number of *simultaneous* constraints is reduced to counting objects that violate *some* of the constraints at distinguished "places"—the latter is usually a simpler problem The generating function analogue of inclusion exclusion is then simply the substitution $v \mapsto u - 1$, if a bivariate GF is sought, or $v \mapsto -1$ in the univariate case.

Rises in permutations and patterns in words. The book by Goulden and Jackson [**208**, pp. 45–48] describes a useful formalization of the inclusion process operating on MGFs. Conceptually, it combines substitution and implicit definitions. Once again, the *modus operandi* is best grasped through examples, two of which are detailed below.

EXAMPLE 24. *Rises and ascending runs in permutations.* A *rise* (also called an *ascent*) in a permutation $\sigma = \sigma_1 \cdots \sigma_n$ is a pair of consecutive elements $\sigma_i \sigma_{i+1}$ satisfying $\sigma_i < \sigma_{i+1}$ (with $1 \leq i < n$). The problem is to determine the number $A_{n,k}$ of permutations of size having exactly k rises, together with the BGF $A(z,u)$. By symmetry, we are also enumerating descents (defined by $\sigma_i > \sigma_{i+1}$) as well as ascending runs that are each terminated by a descent.

Guided by the inclusion-exclusion principle, we tackle the easier problem of enumerating permutations with *distinguished* rises, of which the set is denoted by $\mathcal{B}$. For instance, $\mathcal{B}$ contains elements like

$$2\ 1\ \boxed{3 \nearrow 4 \nearrow 8 \nearrow 9 \nearrow 11}\ 15\ 12\ \boxed{5 \nearrow 10}\ 13\ 7\ 14,$$

where those rises that are distinguished are represented by arrows. (Note that some rises may *not* be distinguished.) Maximal sequences of adjacent distinguished rises (boxed in the representation) will be called *clusters*. Then, $\mathcal{B}$ can be specified by the sequence construction applied to atoms ($\mathcal{Z}$) and clusters ($\mathcal{C}$) as

$$\mathcal{B} = \text{SEQ}(\mathcal{Z} + \mathcal{C}), \qquad \text{where} \quad \mathcal{C} = (\mathcal{Z} \nearrow \mathcal{Z}) + (\mathcal{Z} \nearrow \mathcal{Z} \nearrow \mathcal{Z}) + \cdots = \text{SET}_{\geq 2}(\mathcal{Z}).$$

since a cluster is an ordered sequence, or equivalently a set, furthermore having at least two elements. This gives the EGF of $\mathcal{B}$ as

$$B(z) = \frac{1}{1 - (z + (e^z - 1 - z))} = \frac{1}{2 - e^z},$$

which happens to coincide with the EGF of surjections.

For inclusion-exclusion purposes, we need the BGF of $\mathcal{B}$ with v marking the number of distinguished rises. A cluster of size k contains $k-1$ rises, so that

$$B(z,v) = \frac{1}{1 - (z + (e^{zv} - 1 - zv)/v)} = \frac{v}{v + 1 - e^{zv}}.$$

Now, the usual argument applies: the BGF $A(z,u)$ satisfies $B(z,v) = A(z, 1+v)$, so that $A(z,u) = B(z, u-1)$, which yields the particularly simple form

$$A(z,u) = \frac{u-1}{u - e^{z(u-1)}}. \tag{73}$$

In particular, this GF expands as

$$A(z,u) = 1 + z + (u+1)\frac{z^2}{2!} + (u^2 + 4u + 1)\frac{z^3}{3!} + (u^3 + 11u^2 + 11u + 1)\frac{z^4}{4!} + \cdots.$$

The coefficients $A_{n,k}$ are known as the *Eulerian numbers*. In combinatorial analysis, these numbers are almost as classic as the Stirling numbers. A detailed discussion of their properties is to be found in classical treatises like [**82**] or [**212**]. (From Eq. (73), permutations without rises are enumerated by $B(z,-1) = e^z$, an altogether obvious result.)

Moments derive easily from an expansion of (73) at $u = 1$, which gives

$$A(z,u) = \frac{1}{1-z} + \frac{1}{2}\frac{z^2}{(1-z)^2}(u-1) + \frac{1}{12}\frac{z^3(2+z)}{(1-z)^3}(u-1)^2 + \cdots.$$

In particular: *the mean of the number of rises in a random permutation of size n is $\frac{1}{2}(n-1)$ and the variance is $\sim \frac{1}{12}n$, ensuring concentration of distribution.*

The same method applies to the enumeration of *ascending runs*: for a fixed parameter ℓ, an ascending run of length ℓ is a sequence of consecutive elements $\sigma_i\sigma_{i+1}\cdots\sigma_{i+\ell}$ such that $\sigma_i < \sigma_{i+1} < \cdots < \sigma_{i+\ell}$. (Thus, a rise is an ascending run of length 1.) We define a cluster as a sequence of distinguished runs which overlap in the sense that they share some of the elements of the permutation. The exponential BGF of permutations with distinguished ascending runs is then

$$B(z,v) = \frac{1}{1 - z - \widehat{I}(z,v)}, \qquad \text{where} \quad \widehat{I}(z,v) = \sum_{n,k} I_{n,k} v^k \frac{z^n}{n!},$$

and $I_{n,k}$ is the number of ways of covering the segment $[1,n]$ with k distinct intervals of length ℓ that are contained in $[1,n]$ and have integral end points. The numbers $I_{n,k}$ themselves result from elementary combinatorics (see also the case of patterns in words below) and one has for the OGF corresponding to $\widehat{I}$:

$$I(z,v) \equiv \sum_{n,k} I_{n,k} v^k z^n = \frac{z^{\ell+1} v}{1 - v(z + z^2 + \cdots + z^\ell)}.$$

(Proof: The first segment in the covering must be placed on the left, the other ones appear in succession, each shifted right by 1 to ℓ positions from the previous one.) The last two equations finally determine the exponential BGF of permutations with size marked by z and ascending runs of length $\ell + 1$ marked by u,

$$A(z,u) = B(z, u-1), \tag{74}$$

given the inclusion-exclusion principle.

The resulting formulæ are checked to generalize the case of rises ($\ell = 1$). They can be made explicit by first expanding the OGF $I(z,v)$ into partial fractions, then applying the transformation $(1-\omega z)^{-1} \mapsto e^{\omega z}$ in order to translate $I(z,v)$ into $\widehat{I}(z,v)$. The net result is

$$A(z,u) = \frac{1}{1 - z - \widehat{I}(z,u-1)}, \qquad \text{where} \quad \widehat{I}(z,v) = (1-z)(v+1) + \sum_{j=1}^{\ell} c_j(v) e^{\omega_j(v) z}$$

involves a sum of exponentials. In this last equation, the $\omega_j(v)$ are the roots of the characteristic equation $\omega^\ell = v(1 + \cdots + \omega^{\ell-1})$ and the $c_j(v)$ are the corresponding coefficients in the partial fraction decomposition of $I(z,v)$. These expressions were first published by Elizalde and Noy [**124**] who obtained them by means of tree decompositions.

The BGF (74) can be exploited in order to determine quantitative information on long runs in permutations. First, an expansion at $u = 1$ (also, a direct reasoning: see the discussion of hidden words in Chapter I) shows that the mean number of ascending runs of length $\ell - 1$ is $(n - \ell + 1)/\ell!$ exactly, as soon as $n \geq \ell$. This entails that, if $n = o(\ell!)$, the probability of finding an ascending run of length $\ell - 1$ tends to 0 as $n \to \infty$. What is used in passing in this argument is the general fact that for a discrete variable X with values in $0, 1, 2, \ldots$, one has (with Iverson's notation)

$$\mathbb{P}(X \geq 1) = \mathbb{E}([\![X \geq 1]\!]) = \mathbb{E}(\min(X,1)) \leq \mathbb{E}(X).$$

An inequality in the converse direction can be obtained from the second moment method. In effect, the variance of the number of ascending runs of length $\ell - 1$ is found to be of the exact form $\alpha_\ell n + \beta_\ell$ where α_ℓ is essentially $1/\ell!$ and β_ℓ is of comparable order (details omitted). Then, by Chebyshev's inequalities, concentration of distribution holds as long as ℓ is such that $(\ell+1)! = o(n)$. In this case, with high probability (i.e., with probability tending to 1 as n tends to ∞), there are many ascending runs of length $\ell - 1$. In particular:

Let L_n be the length of the longest ascending run in a random permutation of n elements. Let $\ell_0(n)$ be the smallest integer such that $\ell! \geq n$. Then the distribution of L_n is concentrated in the sense that $L_n/\ell_0(n)$ converges in probability to 1: for any $\epsilon > 0$, one has

$$\lim_{n\to\infty} \mathbb{P}\left(1-\epsilon < \frac{L_n}{\ell_0(n)} < 1+\epsilon\right) = 1.$$

What has been found here is a fairly sharp threshold phenomenon. END OF EXAMPLE 24. □

▷ **33.** *Permutations without ℓ–ascending runs.* The EGF of permutations without 1–, 2– and 3–ascending runs are respectively

$$\left(\sum_{i\geq 0} \frac{x^{2i}}{(2i)!} - \frac{x^{2i+1}}{(2i+1)!}\right)^{-1}, \quad \left(\sum_{i\geq 0} \frac{x^{3i}}{(3i)!} - \frac{x^{3i+1}}{(3i+1)!}\right)^{-1}, \quad \left(\sum_{i\geq 0} \frac{x^{4i}}{(4i)!} - \frac{x^{4i+1}}{(4i+1)!}\right)^{-1},$$

and so on. (See Elizalde and Noy's article [**124**] for similar computations, as well as interesting results involving other types of order patterns in permutations.) ◁

Many variations on the theme of rises and ascending runs are clearly possible. Local order patterns in permutations have been intensely researched, notably by Carlitz in the 1970's. Goulden and Jackson [**208**, Sec. 4.3] offer a general theory of patterns in sequences and permutations. Special permutations patterns associated with binary increasing trees are also studied by Flajolet, Gourdon, and Martínez [**154**] (by combinatorial methods) and Devroye [**104**] (by probabilistic arguments). On another register, the longest ascending run has been found above to be of order $(\log n)/\log\log n$ in probability. The superficially resembling problem of analysing the length of the *longest increasing sequence* in random permutations (elements must be in ascending order but need not be adjacent) has attracted a lot of attention, but is considerably harder. This quantity is $\sim 2\sqrt{n}$ on average and in probability, as shown by a penetrating analysis of the shape of random Young tableaus due to Logan, Shepp, Vershik, and Kerov [**293, 426**] Solving a problem open for over 20 years, Baik, Deift, and Johansson [**19**] have eventually determined its limiting distribution. The undemanding survey by Aldous and Diaconis [**6**] discusses some of the background of this problem, while Chapter VIII shows how to derive bounds that are of the right order of magnitude but rather crude, using saddle-point methods.

EXAMPLE 25. *Patterns in words.* Take the set of all words $W = \text{SEQ}\{\mathcal{A}\}$ over a finite alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$. A pattern $\mathfrak{p} = p_1 p_2 \cdots p_k$, which is particular word of length k has been fixed. What is sought is the BGF $W(z,u)$ of $\mathcal{W}$, where u marks the number of occurrences of pattern $\mathfrak{p}$ inside a word of $\mathcal{W}$. Results of Chapter I already give access to $W(z,0)$, which is the OGF of words not containing the pattern.

In accordance with the inclusion-exclusion principle, one should introduce the class $\mathcal{X}$ of words augmented by distinguishing an arbitrary number of occurrences of $\mathfrak{p}$. Define a *cluster* as a maximal collection of distinguished occurrences that have an overlap. For instance, if $\mathfrak{p} = aaaaa$, a particular word may be give rise to the particular cluster:

```
a b a a a a a a a a a a a a b a a a a a a a a a b b
---------------------------------------------------
      a a a a a
          a a a a a
```

a a a a a

Then objects of $\mathcal{X}$ decompose as sequences of either arbitrary letters from $\mathcal{A}$ or clusters:

$$\mathcal{X} = \mathrm{S}\small{\mathrm{EQ}}\,(\mathcal{Z} + \mathcal{C})\,,$$

with $\mathcal{C}$ the class of all clusters.

Clusters are themselves obtained by repeatedly sliding the pattern, but with the constraint that it should constantly overlap partly with itself. Let $c(z)$ be the autocorrelation polynomial of $\mathfrak{p}$ as defined in Chapter I, and set $\widehat{c}(z) = c(z) - 1$. A moment's reflection should convince the reader that $z^k \widehat{c}(z)^{s-1}$ when expanded describes all the possibilities for forming clusters of s overlapping occurrences. On the example above, one has $\widehat{c}(z) = 1 + z + z^2 + z^3 + z^4$, and a particular cluster of 3 overlapping occurrences corresponds to one of the terms in $z^k \widehat{c}(z)^2$ as follows:

$$\begin{array}{l|l}
\overbrace{\mathtt{a\ a\ a\ a\ a}}^{z^5} & z^5 \\
\phantom{\mathtt{a\ a\ a\ }}\mathtt{a\ a\ a}\overbrace{\mathtt{a\ a}}^{z^2} & \times\,(z + \underline{\mathbf{z^2}} + z^3 + z^4) \\
\phantom{\mathtt{a\ a\ a\ a\ a\ a\ a\ }}\mathtt{a}\overbrace{\mathtt{a\ a\ a\ a}}^{z^4} & \times\,(z + z^2 + z^3 + \underline{\mathbf{z^4}}).
\end{array}$$

The OGF of clusters is consequently $C(z) = z^k/(1 - \widehat{c}(z))$ since this quantity describes all the ways to write the pattern (z^k) and then slide it so that it should overlap with itself (this is given by $(1 - \widehat{c}(z))^{-1}$). A slightly different way on obtaining this expressions of $C(z)$ is described in Note 36 below.

By a similar reasoning, the BGF of clusters is $vz^k/(1 - v\widehat{c}(z))$, and the BGF of $\mathcal{X}$ with the supplementary variable v marking the number of distinguished occurrences is

$$X(z, v) = \frac{1}{1 - rz - vz^k/(1 - v\widehat{c}(z))}.$$

Finally, the usual inclusion-exclusion argument (change v to $u - 1$) yields $W(z, u) = X(z, u - 1)$. As a result:

> *For a pattern $\mathfrak{p}$ with correlation polynomial $c(z)$ and length k, the BGF of words over an alphabet of cardinality r, where u marks the number of occurrences of $\mathfrak{p}$, is*

$$W(z, u) = \frac{(u - 1)c(z) - u}{(1 - rz)((u - 1)c(z) - u) + (u - 1)z^k}.$$

The specialization $u = 0$ gives back the formula already found in Chapter I. The same principles clearly apply to weighted models corresponding to unequal letter probabilities, provided a suitably weighted version of the correlation polynomial is introduced (Note 36 below). END OF EXAMPLE 25. $\square$

There are a very large number of formulæ related to patterns in strings. For instance, BGFs are known for occurrences of one or several patterns under either Bernoulli or Markov models; see Note 36 below. We refer to Szpankowski's book [**401**], where such questions are treated systematically and in great detail. Bourdon and Vallée [**61**] have even succeeded in extending this approach to *dynamical sources* of information, thereby extending a large number of previously known results. Their approach even makes it possible to analyse the occurrence of patterns in continued fraction representations of real numbers.

$\triangleright$ **34.** *Moments of number of occurrences.* The derivatives of $X(z,v)$ at $v=0$ give access to the factorial moments of the number of occurrences of a pattern. In this way or directly, one determines

$$W(z,u) = \frac{1}{1-rz} + \frac{z^k}{(1-rz)^2}(u-1) + 2\frac{z^k((1-rz)(c(z)-1)+z^k)}{(1-rz)^3}\frac{(u-1)^2}{2!} + \cdots .$$

The mean number of occurrences is r^{-n} times the coefficient of z^n in the coefficient of $(u-1)$ and is $(n-k+1)r^{-k}$, as anticipated. The coefficient of of $(u-1)^2/2!$ is of the form

$$\frac{2r^{-2k}}{(1-rz)^3} + \frac{2r^{-k}(1+2kr^{-k}-c(1/r))}{(1-rz)^2} + \frac{P(z)}{1-rz},$$

with P a polynomial. There results that the variance of the number of occurrences is of the form

$$\alpha n + \beta, \qquad \alpha = r^{-k}(2c(1/r) - 1 + r^{-k}(1-2k)).$$

Consequently, the distribution is concentrated around its mean. (See also the discussion of "Borges' Theorem" in Chapter I, p. 58.) $\triangleleft$

$\triangleright$ **35.** *Words with fixed repetitions.* Let $W^{\langle s\rangle}(z) = [u^s]W(z,u)$ be the OGF of words containing a pattern exactly s times. One has, for $s>0$ and $s=0$ respectively,

$$W^{\langle s\rangle}(z) = \frac{z^k N(z)^{s-1}}{D(z)^{s+1}}, \qquad W^{\langle 0\rangle}(z) = \frac{c(z)}{D(z)},$$

with $N(z)$ and $D(z)$ given by

$$N(z) = (1-rz)(c(z)-1)+z^k, \qquad D(z) = (1-rz)c(z)+z^k.$$

The expression of $W^{\langle 0\rangle}$ is in agreement with Chapter I, Equation (48). $\triangleleft$

$\triangleright$ **36.** *Patterns in Bernoulli sequences.* Let $\mathcal{A}$ be an alphabet where letter α has probability π_α and consider the Bernoulli model where letters in words are chosen independently. Fix a pattern $\mathfrak{p} = p_1 \cdots p_k$ and define the finite language of *protrusions* as

$$\Gamma = \bigcup_{i\,:\,c_i\neq 0} \{p_{i+1}p_{i+2}\cdots p_k\},$$

where the union is over all correlation positions of the pattern. Define now the correlation polynomial $\gamma(z)$ (relative to $\mathfrak{p}$ and the π_α) as the generating polynomial of the finite language of protrusions weighted by π_α. For instance, $\mathfrak{p} = \mathsf{ababa}$ gives rise to $\Gamma = \{\epsilon, \mathsf{ba}, \mathsf{baba}\}$ and

$$\gamma(z) = 1 + \pi_a\pi_b z^2 + \pi_a^2\pi_b^2 z^4.$$

Then, the BGF of words with z marking length and U marking the number of occurrences of $\mathfrak{p}$ is

$$W(z,u) = \frac{(u-1)\gamma(z) - u}{(1-z)((u-1)\gamma(z)-u) + \pi[\mathfrak{p}]z^k},$$

where $\pi[\mathfrak{p}]$ is the product of the probabilities of letters of $\mathfrak{p}$. $\triangleleft$

$\triangleright$ **37.** *Patterns in binary trees.* Consider the class $\mathcal{B}$ of pruned binary trees. An occurrence of pattern $\mathfrak{t}$ in a tree τ is defined by a node whose "dangling subtree" is isomorphic to $\mathfrak{t}$. Let p be the size of $\mathfrak{t}$. The BGF $B(z,u)$ of class $\mathcal{B}$ where u marks the number of occurrences of $\mathfrak{t}$ is sought.

The OGF of $\mathcal{B}$ is $B(z) = (1-\sqrt{1-4z})/(2z)$. The quantity $vB(zv)$ is the BGF of $\mathcal{B}$ with v marking external nodes. By virtue of the pointing operation, the quantity

$$U_k := \left(\frac{1}{k!}\partial_v^k\,(vB(zv))\right)_{v=1},$$

describes trees with k distinct external nodes distinguished (pointed). The quantity

$$V := \sum U_k u^k (z^p)^k \quad \text{satisfies} \quad V = (vB(zv))_{v=1+uz^p},$$

by virtue of Taylor's formula. It is also the BGF of trees with distinguished occurrences of $\mathfrak{t}$. Setting $v \mapsto u-1$ in V then gives back $B(z,u)$ as

$$B(z,u) = \frac{1}{2z}\left(1 - \sqrt{1-4z-4(u-1)z^{p+1}}\right).$$

In particular

$$B(z,0) = \frac{1}{2z}\left(1-\sqrt{1-4z+4z^{p+1}}\right)$$

gives the OGF of trees *not* containing pattern $\mathfrak{t}$. The method generalizes to any simple variety of trees and it can be used to prove that the factored representation (as a directed acyclic graph) of a random tree of size n has expected size $O(n/\sqrt{\log n})$; see [**176**]. ◁

III. 8. Extremal parameters

Apart from additively inherited parameters already examined at length in this chapter, another important category is that of parameters defined by a maximum rule. Two major cases are the largest component in a combinatorial structure (for instance, the largest cycle of a permutation) and the maximum degree of nesting of constructions in a recursive structure (typically, the height of a tree). In this case, bivariate generating functions are of little help. The standard technique consists in introducing a collection of univariate generating functions defined by imposing a bound on the parameter of interest. Such GFs can then be constructed by the symbolic method in its univariate version.

III. 8.1. Largest components. Consider a construction $\mathcal{B} = \Phi\{\mathcal{A}\}$, where Φ may involve an arbitrary combination of basic constructions, and assume here for simplicity that the construction for $\mathcal{B}$ is a non–recursive one. This corresponds to a relation between generating functions

$$B(z) = \Psi[A(z)],$$

where Ψ is the functional that is the "image" of the combinatorial construction Φ. Elements of $\mathcal{A}$ thus appear as components in an object $\beta \in \mathcal{B}$. Let $\mathcal{B}^{\langle b\rangle}$ denote the subclass of $\mathcal{B}$ formed with objects whose $\mathcal{A}$–components all have a size at most b. The GF of $\mathcal{B}^{\langle b\rangle}$ is obtained by the same process as that of $\mathcal{B}$ itself, save that $A(z)$ should be replaced by the GF of elements of size at most b. Thus,

$$B^{\langle b\rangle}(z) = \Psi[\mathbf{T}_b A(z)],$$

where the *truncation operator* is defined on series by

$$\mathbf{T}_b f(z) = \sum_{n=0}^{b} f_n z^n \qquad (f(z) = \sum_{n=0}^{\infty} f_n z^n).$$

Several cases of this situation have already been encountered in earlier chapters. For instance, the cycle decomposition of permutations translated by

$$P(z) = \exp\left(\log\frac{1}{1-z}\right)$$

gives more generally the EGF of permutations with longest cycle $\le b$,

$$P^{\langle b\rangle}(z) = \exp\left(\frac{z}{1} + \frac{z^2}{2} + \cdots + \frac{z^b}{b}\right),$$

which involves the truncated logarithm. Similarly, the EGF of words over an m–ary alphabet

$$W(z) = (e^z)^m$$

leads to the EGF of words such that each letter occurs at most b times:

$$W^{\langle b\rangle}(z) = \left(1 + \frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^b}{b!}\right)^m,$$

which now involves the truncated exponential. One finds similarly the EGF of set partitions with largest block of size at most b,

$$S^{\langle b\rangle}(z) = \exp\left(\frac{z}{1!} + \frac{z^2}{2!} + \cdots + \frac{z^b}{b!}\right).$$

A slightly less direct example is that of the longest run in a sequence of binary draws. The collection $\mathcal{W}$ of binary strings over the alphabet $\{a, b\}$ admits the decomposition

$$\mathcal{W} = \text{SEQ}(a) \cdot \text{SEQ}(b\ \text{SEQ}(a)),$$

corresponding to a "scansion " dictated by the occurrences of the letter b. The corresponding OGF then appears under the form

$$W(z) = Y(z) \cdot \frac{1}{1 - zY(z)} \qquad \text{where } Y(z) = \frac{1}{1-z}$$

corresponds to $\mathcal{Y} = \text{SEQ}(a)$. Thus, the OGF of strings with at most $k-1$ consecutive occurrences of the letter a obtains upon replacing $Y(z)$ by its truncation:

$$W^{\langle k\rangle}(z) = Y^{\langle k\rangle}(z)\frac{1}{1 - zY^{\langle k\rangle}(z)} \text{ where } Y^{\langle k\rangle}(z) = 1 + z + z^2 + \cdots + z^{k-1},$$

so that

$$W^{\langle k\rangle}(z) = \frac{1 - z^k}{1 - 2z + z^{k+1}}.$$

Such generating functions are thus easy to derive. The asymptotic analysis of their coefficients is however often hard when compared to additive parameters, owing to the need to rely on complex analytic properties of the truncation operator. The bases of a general asymptotic theory have been laid by Gourdon [**210**].

$\triangleright$ **38.** *Smallest components.* The EGF of permutations with smallest cycle of size $> b$ is

$$\frac{\exp(-\frac{z}{1} - \frac{z^2}{2} - \cdots - \frac{z^b}{b})}{1-z}.$$

A symbolic theory of *smallest* components in combinatorial structures is easily developed as regards GFs. Elements of the corresponding asymptotic theory are provided by Panario and Richmond in [**338**]. $\triangleleft$

III. 8.2. Height. The degree of nesting of a recursive construction is a generalization of the notion of height in the simpler case of trees. Consider for instance a recursively defined class

$$\mathcal{B} = \Phi\{\mathcal{B}\},$$

where Φ is a construction. Let $\mathcal{B}^{[h]}$ denote the subclass of $\mathcal{B}$ composed solely of elements whose construction involves at most h applications of Φ. We have by definition

$$\mathcal{B}^{[h+1]} = \Phi\{\mathcal{B}^{[h]}\}.$$

Thus, with Ψ the image functional of construction Φ, the corresponding GFs are defined by a *recurrence*,

$$B^{[h+1]} = \Psi[B^{[h]}].$$

It is usually convenient to start the recurrence with the initial condition $B^{[-1]}(z) = 0$. (This discussion is related to semantics of recursion, p. 31)

Consider for instance general plane trees defined by

$$\mathcal{G} = \mathcal{N} \times \textsc{Seq}(\mathcal{G}) \qquad \text{so that} \qquad G(z) = \frac{z}{1 - G(z)}.$$

Define the height of a tree as the number of nodes on its longest branch. Then the set of trees of height $\le h$ satisfies the recurrence

$$\mathcal{G}^{[0]} = \mathcal{N},\ \mathcal{G}^{[h+1]} = \mathcal{N} \times \textsc{Seq}(\mathcal{G}^{[h]}).$$

Accordingly, the OGF of trees of bounded height satisfies

$$G^{[-1]}(z) = 0,\ G^{[0]}(z) = z,\ G^{[h+1]}(z) = \frac{z}{1 - G^{[h]}(z)}.$$

The recurrence unwinds and one finds

$$G^{[h]}(z) = \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{\ddots \over 1 - z}}}, \tag{75}$$

where the number of stages in the fraction equals b. This is the finite form (technically known as a "convergent") of a *continued fraction* expansion. From implied linear recurrences and an analysis based on Mellin transforms, de Bruijn, Knuth, and Rice [**95**] have determined the average height of a general plane tree to be $\sim \sqrt{\pi n}$. We provide a proof of this fact in Chapter V dedicated to applications of rational and meromorphic asymptotics.

For plane binary trees defined by

$$\mathcal{B} = \mathcal{Z} + \mathcal{B} \times \mathcal{B} \qquad \text{so that} \qquad B(z) = z + (B(z))^2,$$

(size is the number of external nodes), the recurrence is

$$B^{[0]}(z) = z,\ B^{[h+1]}(z) = z + (B^{[h]}(z))^2.$$

In this case, the $B^{[h]}$ are the approximants to a "continuous quadratic form", namely

$$B^{[h]}(z) = z + (z + (z + (\cdots)^2)^2)^2.$$

These are polynomials of degree 2^h for which no closed form expression is known, nor even likely to exist[5]. However, using complex asymptotic methods and singularity analysis, Flajolet and Odlyzko [**165**] have shown that the average height of a binary plane tree is $\sim 2\sqrt{\pi n}$.

For Cayley trees, finally, the defining equation is

$$\mathcal{T} = \{1\} \star \text{SET}(\mathcal{T}) \qquad \text{so that} \qquad T(z) = ze^{T(z)}.$$

The EGF of trees of bounded height satisfy the recurrence

$$T^{[0]}(z) = z,\ T^{[h+1]}(z) = ze^{T^{[h]}(z)}.$$

We are now confronted with a "continuous exponential",

$$T^{[h]}(z) = ze^{ze^{ze^{\cdot^{\cdot^{\cdot ze^{z}}}}}}.$$

The average height was found by Rényi and Szekeres who appealed again to complex asymptotics and found it to be $\sim \sqrt{2\pi n}$.

These examples show that height statistics are closely related to iteration theory. Except in a few cases like general plane trees, normally no algebra is available and one has to resort to complex analytic methods as exposed in forthcoming chapters.

III. 8.3. Averages and moments. For extremal parameters, the GF of mean values obey a general pattern. Let $\mathcal{F}$ be some combinatorial class with GF $f(z)$. Consider for instance an extremal parameter χ such that $f^{[h]}(z)$ is the GF of objects with χ-parameter *at most* h. The GF of objects for which $\chi = k$ *exactly* is equal to

$$f^{[h]}(z) - f^{[h-1]}(z).$$

Thus differencing gives access to the probability distribution of height over $\mathcal{F}$. The generating function of cumulated values (providing mean values after normalization) is then

$$\begin{aligned} \Xi(z) &= \sum_{h=0}^{\infty} h\left[f^{[h]}(z) - f^{[h-1]}(z)\right] \\ &= \sum_{h=0}^{\infty} \left[f(z) - f^{[h]}(z)\right], \end{aligned}$$

as is readily checked by rearranging the second sum, or equivalently using summation by parts.

For maximum component size, the formulæ involve truncated Taylor series. For height, analysis involves in all generality the differences between the fixed point of a functional Φ (the GF $f(z)$) and the approximations to the fixed point ($f^{[h]}(z)$) provided by iteration. This is a common scheme in extremal statistics.

[5]These polynomials are exactly the much studied Mandelbrot polynomials whose behaviour in the complex plane gives rise to extraordinary graphics.

▷ **39.** *Hierarchical partitions.* Let $\varepsilon(z) = e^z - 1$. The generating function

$$\varepsilon(\varepsilon(\cdots(\varepsilon(z)))) \qquad (h \text{ times}).$$

can be interpreted as the EGF of certain hierarchical partitions. (Such structures show up in statistical classification theory [**417, 418**].) ◁

▷ **40.** *Balanced trees.* Balanced structures lead to counting GFs close to the ones obtained for height statistics. The OGF of balanced 2-3 trees of height h counted by the number of leaves satisfies the recurrence

$$Z^{[h+1]}(z) = Z^{[h]}(z^2 + z^3) = (Z^{[h]}(z))^2 + (Z^{[h]}(z))^3,$$

which can be expressed in terms of the iterates of $\sigma(z) = z^2 + z^3$. It is also possible to express the OGF of cumulated values of the number of internal nodes in such trees. ◁

▷ **41.** *Extremal statistics in random mappings.* One can express the EGFs relative to the largest cycle, longest branch, and diameter of functional graphs. Similarly for the largest tree, largest component. [Hint: see [**166**] for details.] ◁

▷ **42.** *Deep nodes in trees.* The BGF giving the number of nodes at maximal depth in a general plane tree or a Cayley tree can be expressed in terms of a continued fraction or a continuous exponential. ◁

III. 9. Perspective

The message of this chapter is that we can use the symbolic method not just to count combinatorial objects but also to quantify their properties. The relative ease with which we are able to do so is testimony to the power of the method as major organizing principle of analytic combinatorics.

The global framework of the symbolic method leads us to a natural structural categorization of parameters of combinatorial objects. First, the concept of *inherited parameters* permits a direct extension of the already seen formal translation mechanisms from combinatorial structures to GFs, for both labelled and unlabelled objects—this leads to MGFs useful for solving a broad variety of classical combinatorial problems. Second, the adaptation of the theory to *recursive parameters* provides information about trees and similar structures, this even in the absence of an explicit representation of the associated MGFs. Third, *extremal parameters* which are defined by a maximum rule (rather than an additive rule) can be studied by analysing families of univariate GFs. Yet another illustration of the power of the symbolic method is found in the notion of *complete GFs*, which in particular enable us to study Bernoulli trials and branching processes.

As we shall see starting with Chapter IV, these approaches become especially powerful since they serve as the basis for the *asymptotic analysis of properties of structures*. Not only does the symbolic method provide precise information about particular parameters, but also it paves the way for the discovery of general theorems that tell us what to expect about a broad variety of combinatorial types.

Multivariate generating functions are a common tool from classical combinatorial analysis. Comtet's book [**82**] is once more an excellent source of examples. A systematization of multivariate generating functions for inherited parameters is given in the book by Goulden and Jackson [**208**].

In contrast generating functions for cumulated values of parameters (related to averages) seemed to have received relatively little attention until the advent of digital computers and the analysis of algorithms. Many important techniques are implicit in Knuth's treatises, especially [**268, 269**]. Wilf discusses related issues in his book [**437**] and the paper [**435**]. Early systems specialized to tree algorithms have been proposed by Flajolet and Steyaert in the 1980s [**140, 180, 181, 398**]; see also Berstel and Reutenauer's work [**42**]. Some of the ideas developed there initially drew their inspiration from the well established treatment of formal power series in noncommutative indeterminates, see the books by Eilenberg [**123**] and Salomaa–Soittola [**372**] as well as the proceedings edited by Berstel [**43**]. Several computations in this area can nowadays even be automated with the help of computer algebra systems, as shown by Flajolet, Salvy, and Zimmermann [**173, 373, 451**].

Part B

COMPLEX ASYMPTOTICS

IV

Complex Analysis, Rational and Meromorphic Asymptotics

The shortest path between two truths in the real domain passes through the complex domain.

— JACQUES HADAMARD[1]

Contents

Generating functions are a central concept of combinatorial theory. In Part A, we have treated them as formal objects, that is, as formal power series. Indeed, the major theme of Chapters I–III has been to demonstrate how the algebraic structure of generating functions directly reflects the structure of combinatorial classes. From now on, we examine generating functions in the light of *analysis*. This point of view involves assigning *values* to the variables that appear in generating functions.

Comparatively little benefit results from assigning only real values to the variable z that figures in a univariate generating function. In contrast, assigning *complex* values turns out to have serendipitous consequences. When we do so, a generating function becomes a geometric transformation of the complex plane. This transformation is very regular near the origin—one says that it is *analytic* (or *holomorphic*). In other words, near 0, it only effects a smooth distortion of the complex plane. Farther away from the origin, some cracks start appearing in the picture. These cracks—the dignified name is *singularities*—correspond to the disappearance of smoothness. It turns out that a function's singularities provide a wealth of information regarding the function's coefficients, and especially their asymptotic rate of growth. Adopting a geometric point of view for generating functions has a large pay-off.

By focussing on singularities, analytic combinatorics treads in the steps of many respectable older areas of mathematics. For instance, Euler recognized that the fact for the Riemann zeta function $\zeta(s)$ to become infinite at 1 implies the existence of infinitely many prime numbers, while Riemann, Hadamard, and de la Vallée-Poussin uncovered deeper connections between quantitative properties of prime numbers and singularities of $1/\zeta(s)$.

[1]Quoted in The Mathematical Intelligencer, v. 13, no. 1, Winter 1991.

The purpose of this chapter is largely to serve as an accessible introduction or a refresher of basic notions regarding analytic functions. We start by recalling the elementary theory of functions and their singularities in a style tuned to the needs of analytic combinatorics. Cauchy's integral formula expresses coefficients of analytic functions as contour integrals. Suitable uses of Cauchy's integral formula then make it possible to estimate such coefficients by suitably selecting an appropriate contour of integration. For the common case of functions that have singularities at a finite distance, the exponential growth formula relates the *location* of the singularities closest to the origin—these are also known as dominant singularities—to the *exponential order of growth* of coefficients. The *nature* of these singularities then dictates the fine structure of the asymptotics of the function's coefficients, especially the *subexponential factors* involved.

As regards generating functions, combinatorial enumeration problems can be broadly categorized according to a hierarchy of increasing structural complexity. At the most basic level, we encounter scattered classes, which are simple enough, so that the associated generating function and coefficients can be made explicit. (Examples of Part A include binary and general plane trees, Cayley trees, derangements, mappings, and set partitions). In that case, elementary real-analysis techniques usually suffice to estimate asymptotically counting sequences. At the next, intermediate, level, the generating function is still explicit, but its form is such that no simple expression is available for coefficients. This is where the theory developed in this and the next chapters comes into play. It usually suffices to have an expression for a generating function, but *not* necessarily its coefficients, so as to be able to deduce precise asymptotic estimates of its coefficients. (Surjections, generalized derangements, unary-binary trees are easily subjected to this method. A striking example, that of trains, is detailed in Section IV. 4.) Properties of analytic functions then make this analysis depend only on *local properties* of the generating function at a few points, its dominant singularities. The third, highest, level, within the perspective of analytic combinatorics, comprises generating functions that can no longer be made explicit, but are only determined by a *functional equation*. This covers structures defined recursively or implicitly by means of the basic constructors of Part A. The analytic approach even applies to a large number of such cases. (Examples include simple families of trees, balanced trees, and the enumeration of certain molecules treated at the end of this chapter. Another characteristic example is that of nonplane unlabelled trees treated in Chapter VII.)

As we are going to see in this chapter and the next four ones, the analytic methodology applies to almost all the combinatorial classes studied in Part A, which are provided by the symbolic method. In the present chapter we carry out this programme for *rational functions* and *meromorphic functions*, where the latter are defined by the fact their singularities are simply *poles*.

IV. 1. Generating functions as analytic objects

Generating functions, considered in Part A as purely *formal* objects subject to algebraic operations, are now going to be interpreted as *analytic* objects. In so doing

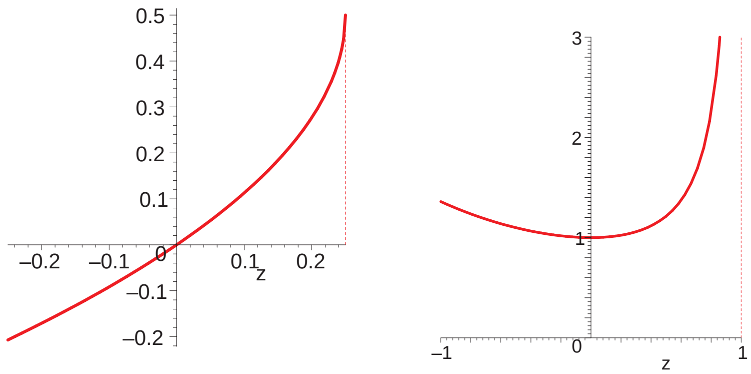


FIGURE 1. Left: the graph of the Catalan OGF, $f(z)$, for $z \in (-\frac{1}{4}, +\frac{1}{4})$; right: the graph of the derangement EGF, $g(z)$, for $z \in (-1, +1)$.

one gains an easy access to the asymptotic form of their coefficients. This informal section offers a glimpse of themes that form the basis of Chapters IV–VII.

In order to introduce the subject softly, let us start with two simple generating functions, one, $f(z)$, being the OGF of the Catalan numbers (cf $G(z)$, p. 33), the other, $g(z)$, being the EGF of derangements (cf $D^{(1)}(z)$, p. 114):

$$f(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right), \qquad g(z) = \frac{\exp(-z)}{1-z}. \tag{1}$$

At this stage, the forms above are merely compact descriptions of formal power series built from the elementary series

$$\begin{aligned} (1-y)^{-1} &= 1 + y + y^2 + \cdots, \qquad & (1-y)^{1/2} &= 1 - \frac{1}{2}y - \frac{1}{8}y^2 - \cdots, \\ \exp(y) &= 1 + \frac{1}{1!}y + \frac{1}{2!}y^2 + \cdots, \end{aligned}$$

by standard composition rules. Accordingly, the coefficients of both GFs are known in explicit form

$$f_n := [z^n]f(z) = \frac{1}{n}\binom{2n-2}{n-1}, \quad g_n := [z^n]g(z) = \left(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\right).$$

Stirling's formula and comparison with the alternating series giving $\exp(-1)$ provide respectively

$$f_n \underset{n\to\infty}{\sim} \frac{4^n}{\sqrt{\pi n^3}}, \qquad g_n = \underset{n\to\infty}{\sim} e^{-1} \doteq 0.36787. \tag{2}$$

Our purpose now is to provide intuition on how such approximations could be derived without a recourse to explicit forms. We thus examine, heuristically for the moment, the direct relationship between the asymptotic forms (2) and the structure of the corresponding generating functions in (1).

Granted the growth estimates available for f_n and g_n, it is legitimate to substitute in the power series expansions of the GFs $f(z)$ and $g(z)$ any real or complex value of a small enough modulus, the upper bounds on modulus being $\rho_f = \frac{1}{4}$ (for f) and

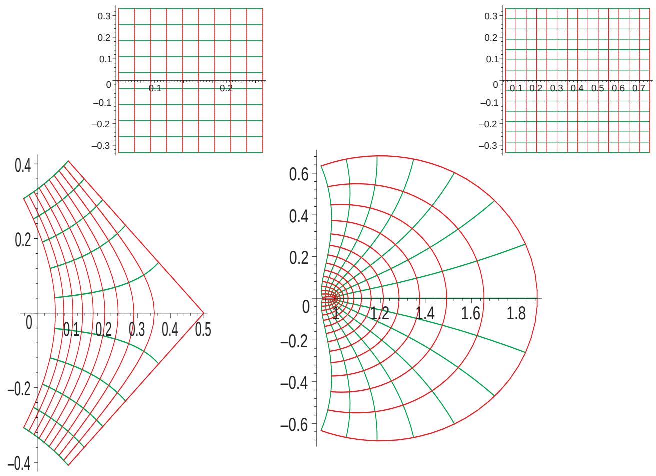


FIGURE 2. The images of regular grids by $f(z)$ (left) and $g(z)$ (right).

$\rho_g = 1$ (for g). Figure 1 represents the graph of the resulting functions when such *real* values are assigned to z. The graphs are smooth, representing functions that are differentiable any number of times for z interior to the interval $(-\rho, +\rho)$. However, at the right boundary point, smoothness stops: $g(z)$ become infinite at $z = 1$, and so it even ceases to be finitely defined; $f(z)$ does tend to the limit $\frac{1}{2}$ as $z \to (\frac{1}{4})^-$, but its derivative becomes infinite there. Such special points at which smoothness stops are called *singularities*, a term that will acquire a precise meaning in the next sections.

Observe also that, in spite of the series expressions being divergent outside the specified intervals, the functions $f(z)$ and $g(z)$ can be *continued* in certain regions: it suffices to make use of the global expressions of Equation (1), with $\exp$ and $\sqrt{\ }$ being assigned their usual real-analytic interpretation. For instance:

$$f(-1) = \frac{1}{2}\left(1 - \sqrt{5}\right), \qquad g(-2) = \frac{e^2}{3}.$$

Such continuation properties, most notably to the *complex* realm, will prove essential in developing efficient methods for coefficient asymptotics.

One may proceed similarly with complex numbers, starting with numbers whose modulus is less than the radius of convergence of the series defining the GF. Figure 2 displays the images of regular grids by f and g, as given by (1). This illustrates the fact that a regular grid transforms into an orthogonal network of curves and more precisely that f and g preserve angles—this property corresponds to complex differentiability and is equivalent to analyticity to be introduced shortly. The singularity of f is clearly perceptible on the right of its diagram, since, at $z = \frac{1}{4}$ (corresponding to $f(z) = \frac{1}{2}$), the function f folds lines and divides angles by a factor of 2.

Let us now turn to coefficient asymptotics. As is expressed by (2), the coefficients f_n and g_n each belong to a general asymptotic type for coefficients of a function F, namely,

$$[z^n]F(z) = A^n\theta(n),$$

corresponding to an exponential growth factor A^n modulated by a tame factor $\theta(n)$, which is subexponential. Here, one has $A = 4$ for f_n and $A = 1$ for g_n; also, $\theta(n) \sim \frac{1}{4}(\sqrt{\pi n^3})^{-1}$ for f_n and $\theta(n) \sim e^{-1}$ for g_n. Clearly, A should be related to the radius of convergence of the series. We shall see that invariably, for combinatorial generating functions, the exponential rate of growth is given by $A = 1/\rho$, where ρ is the first singularity encountered along the positive real axis (Theorem IV.6). In addition, under general complex-analytic conditions, it will be established that $\theta(n) = O(1)$ is systematically associated to a simple pole of the generating function (Theorem IV.10, p. 245), while $\theta(n) = O(n^{-3/2})$ systematically arises from a singularity that is of the square-root type (Chapters VI and VII). In summary, as this chapter and the next ones will copiously illustrate, the coefficient formula

$$[z^n]F(z) = A^n\theta(n), \tag{3}$$

with its exponentially dominating term and its subexponential factor, is central. We have:

> **First Principle of Coefficient Asymptotics.** *The* location *of a function's singularities dictates the* exponential growth *(A^n) of its coefficients.*
> **Second Principle of Coefficient Asymptotics.** *The* nature *of the function's singularities determines the associate* subexponential factor *($\theta(n)$).*

Observe that the rescaling rule,

$$[z^n]F(z) = \rho^{-n}[z^n]F(\rho z),$$

enables one to normalize functions so that they are singular at 1. Then various theorems, starting with Theorems IV.9 and IV.10, provide sufficient conditions under which the following central implication is valid,

$$h(z) \sim \sigma(z) \quad \Longrightarrow \quad [z^n]h(z) \sim [z^n]\sigma(z). \tag{4}$$

There $h(z)$, whose coefficients are to be estimated, is a function singular at 1 and $\sigma(z)$ is a local approximation near the singularity; usually σ is a much simpler function, typically like $(1-z)^\alpha \log^\beta(1-z)$ whose coefficients are comparatively easy to estimate (Chapter VI). The relation (4) expresses *a mapping between asymptotic scales of functions near singularities and asymptotics scales of coefficients*. Under suitable conditions, it then suffices to estimate a function locally at a few distinguished points (singularities), in order to estimate its coefficients asymptotically.

$\triangleright$ **1.** *Euler, the discrete, and the continuous.* Eulers's proof of the existence of infinitely many prime numbers illustrates in a striking manner the way analysis of generating functions can inform us on the discrete realm. Define, for real $s > 1$ the function

$$\zeta(s) := \sum_{n=1}^{\infty} \frac{1}{n^s},$$

known as the Riemann zeta function. The decomposition (p ranges over the prime numbers $2, 3, 5, \ldots$)

$$\begin{aligned}\zeta(s) &= \left(1+\frac{1}{2^s}+\frac{1}{2^{2s}}+\cdots\right)\left(1+\frac{1}{3^s}+\frac{1}{3^{2s}}+\cdots\right)\left(1+\frac{1}{5^s}+\frac{1}{5^{2s}}+\cdots\right)\cdots \\ &= \prod_p \left(1-\frac{1}{p^s}\right)^{-1}\end{aligned} \tag{5}$$

expresses precisely the fact that each integer has a unique decomposition as a product of primes. Analytically, the identity (5) is easily checked to be valid for all $s > 1$. Now suppose that there were only finitely many primes. Let s tend to 1^+ in (5). Then, the left hand side becomes infinite, while the right hand side tends to the finite limit $\prod_p (1-1/p)^{-1}$: a contradiction has been reached. $\lhd$

$\rhd$ **2.** *Elementary transfers.* Elementary series manipulation yield the following general result: *Let $h(z)$ be a power series with radius of convergence > 1 and assume that $h(1) \neq 0$; then one has*

$$[z^n]\frac{h(z)}{1-z} \sim h(1), \qquad [z^n]h(z)\sqrt{1-z} \sim -\frac{h(1)}{2\sqrt{\pi n^3}}, \qquad [z^n]h(z)\log\frac{1}{1-z} \sim \frac{h(1)}{n}.$$

See Bender's survey [**28**] for many similar statements. $\lhd$

$\rhd$ **3.** *Asymptotics of generalized derangements.* The EGF of permutations without cycles of length 1 and 2 satisfies (p. 114)

$$j(z) = \frac{e^{-z-z^2/2}}{1-z} \qquad \text{with} \qquad j(z) \underset{z\to 1}{\sim} \frac{e^{-3/2}}{1-z}.$$

Analogy with derangements suggests that $[z^n]j(z) \underset{n\to\infty}{\sim} e^{-3/2}$. [For a proof, use Note 2 or refer to Example 8.] Here is a table of exact values of $[z^n]j(z)$ (with relative error of the approximation by $e^{-3/2}$ in parentheses):

	$n=5$	$n=10$	$n=20$	$n=50$
j_n :	0.2	0.22317	0.2231301600	0.2231301601484298289332804707640122
error :	(10^{-1})	$(2\cdot 10^{-4})$	$(3\cdot 10^{-10})$	(10^{-33})

The quality of the asymptotic approximation is extremely good, such a property being invariably attached to polar singularities. $\lhd$

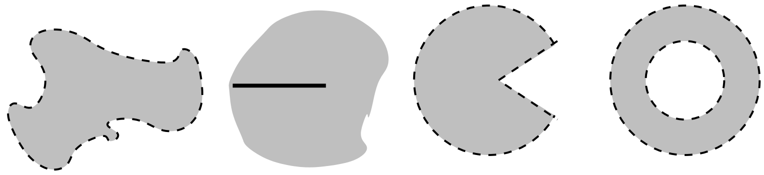
IV. 2. Analytic functions and meromorphic functions

Analytic functions are a primary mathematical concept of asymptotic theory. They can be characterized in two essentially equivalent ways (see IV. 2.1): by means of convergent series expansions (à la Cauchy and Weierstraß) and by differentiability properties (à la Riemann). The first aspect is directly related to the use of generating functions for enumeration; the second one allows for a powerful abstract discussion of closure properties that usually requires little computation.

Integral calculus with analytic functions (see IV. 2.2) assumes a shape radically different from what it is in the real domain: integrals become quintessentially independent of details of the integration contour—certainly the prime example of this fact is Cauchy's famous residue theorem. Conceptually, this independence makes it possible to relate properties of a function at a point (e.g., the coefficients of its expansion at 0) to its properties at another far-away point (e.g., its residue at a pole).

The presentation in this section and the next one constitutes an informal review[2] of basic properties of analytic functions tuned to the needs of asymptotic analysis of counting sequences. The entry in APPENDIX B: *Equivalent definitions of analyticity*, p. 659 provides further information, in particular a proof of the Basic Equivalence Theorem, Theorem IV.1 below. For a detailed treatment, we refer the reader to one of the many excellent treatises on the subject, like the books by Dieudonné [**106**], Henrici [**229**], Hille [**232**], Knopp [**261**], Titchmarsh [**411**], or Whittaker and Watson [**433**].

IV. 2.1. Basics. We shall consider functions defined in certain *regions* of the complex domain $\mathbb{C}$. By a region is meant an *open* subset Ω of the complex plane that is *connected*. Here are some examples:



Classical treatises teach us how to extend to the complex domain the standard functions of real analysis: polynomials are immediately extended as soon as complex addition and multiplication have been defined, while the exponential is definable by means of Euler's formula. One has for instance

$$z^2 = (x^2 - y^2) + 2ixy, \qquad e^z = e^x \cos y + ie^x \sin y,$$

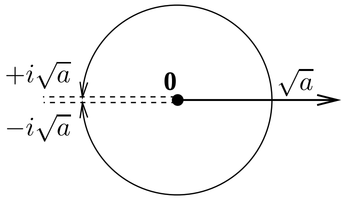
if $z = x + iy$, that is, $x = \Re(z)$ and $y = \Im(z)$ are the real and imaginary parts of z. Both functions are consequently defined over the whole complex plane $\mathbb{C}$.

The square-root and the logarithm are conveniently described in polar coordinates by

$$\sqrt{z} = \sqrt{\rho}e^{i\theta/2}, \qquad \log z = \log \rho + i\theta, \tag{6}$$

if $z = \rho e^{i\theta}$. One can take the domain of validity of (6) to be the complex plane slit along the axis from 0 to $-\infty$, that is, restrict θ to the open interval $(-\pi, +\pi)$, in which case the definitions above specify what is known as the *principal determination*. There is no way for instance to extend by continuity the definition of $\sqrt{z}$ in any domain containing 0 in its interior since, for $a > 0$ and $z \to -a$, one has $\sqrt{z} \to i\sqrt{a}$ as $z \to -a$ from above, while $\sqrt{z} \to -i\sqrt{a}$ as $z \to -a$ from below. This situation is depicted here:

[2]The reader previously unfamilar with the theory of analytic functions should essentially be able to adopt Theorems IV.1 and IV.2 as "*axioms*" and start from there using basic definitions and a fair knowledge of elementary calculus. Figure 18 at the end of this chapter (p. 274) recapitulates the main results of relevance to *Analytic Combinatorics*.



The values of $\sqrt{z}$ as z varies along $|z| = a$.

The point $z = 0$ where several determinations "meet" is accordingly known as a *branch point*.

Analytic functions. First comes the main notion of an analytic function that arises from convergent series expansions and is closely related to the notion of generating function encountered in previous chapters.

DEFINITION IV.1. *A function $f(z)$ defined over a region Ω is* analytic *at a point $z_0 \in \Omega$ if, for z in some open disc centred at z_0 and contained in Ω, it is representable by a convergent power series expansion*

$$f(z) = \sum_{n\geq 0} c_n (z - z_0)^n. \tag{7}$$

A function is analytic in a region Ω iff it is analytic at every point of Ω.

As derived from an elementary property of power series, given a function f that is analytic at a point z_0, there exists a disc (of possibly infinite radius) with the property that the series representing $f(z)$ is convergent for z inside the disc and divergent for z outside the disc. The disc is called the *disc of convergence* and its radius is the *radius of convergence* of $f(z)$ at $z = z_0$, which will be denoted by $\mathrm{R_{conv}}(f; z_0)$. Quite elementarily, the radius of convergence of a power series conveys information regarding the rate at which its coefficients grow; see Subsection IV. 3.2 below for developments. It is also easy to prove by simple series rearrangement (see APPENDIX B: *Equivalent definitions of analyticity*, p. 659) that if a function is analytic at z_0, it is then analytic at all points interior to its disc of convergence.

Consider for instance the function $f(z) = 1/(1 - z)$ defined over $\mathbb{C} \setminus \{1\}$ in the usual way via complex division. It is analytic at 0 by virtue of the geometric series sum,

$$\frac{1}{1-z} = \sum_{n\geq 0} 1 \cdot z^n,$$

which converges in the disc $|z| < 1$. At a point $z_0 \neq 1$, we may write

$$\begin{aligned} \frac{1}{1-z} &= \frac{1}{1 - z_0 - (z - z_0)} = \frac{1}{1-z_0} \frac{1}{1 - \frac{z-z_0}{1-z_0}} \\ &= \sum_{n\geq 0} \left(\frac{1}{1-z_0}\right)^{n+1} (z - z_0)^n. \end{aligned} \tag{8}$$

The last equation shows that $f(z)$ is analytic in the disc centred at z_0 with radius $|1 - z_0|$, that is, the interior of the circle centred at z_0 and passing through the point 1. In particular $\mathrm{R_{conv}}(f, z_0) = |1 - z_0|$ and $f(z)$ is globally analytic in the punctured plane $\mathbb{C} \setminus \{1\}$.

The last example illustrates the definition of analyticity. However, the series rearrangement approach that it uses might be difficult to carry out for more complicated functions. In other words, a more manageable approach to analyticity is called for. The differentiability properties developed next provide such an approach.

Differentiable (holomorphic) functions. The next important notion is a geometric one based on differentiability.

DEFINITION IV.2. *A function $f(z)$ defined over a region Ω is called* complex-differentiable *(also* holomorphic*) at z_0 if the limit, for complex δ,*

$$\lim_{\delta\to 0}\frac{f(z_0+\delta)-f(z_0)}{\delta}$$

exists. (In particular, the limit is independent of the way δ tends to 0 in $\mathbb{C}$.) This limit is denoted as usual by $f'(z_0)$ or $\frac{d}{dz}f(z)\big|_{z_0}$. A function is complex-differentiable in Ω iff it is complex-differentiable at every $z_0\in\Omega$.

Clearly, if $f(z)$ is complex-differentiable at z_0 and $f'(z_0)\neq 0$, it acts locally as a linear transformation:

$$f(z)-f(z_0)\sim f'(z_0)(z-z_0)\qquad (z\to z_0).$$

Then $f(z)$ behaves in small regions almost like a similarity transformation (composed of a translation, a rotation, and a scaling). In particular, it preserves angles[3] and infinitesimal squares get transformed into infinitesimal squares; see Figure 3 for a rendering.

For instance the function $\sqrt{z}$, defined by (6) in the complex plane slit along the ray $(-\infty,0)$, is complex-differentiable at any z of the slit plane since

$$\lim_{\delta\to 0}\frac{\sqrt{z+\delta}-\sqrt{z}}{\delta}=\lim_{\delta\to 0}\sqrt{z}\,\frac{\sqrt{1+\delta/z}-1}{\delta}=\frac{1}{2\sqrt{z}},\tag{9}$$

which extends the customary proof of real analysis. Similarly, $\sqrt{1-z}$ is analytic in the complex plane slit along the ray $(1,+\infty)$. More generally, the usual proofs from real analysis carry over almost verbatim to the complex realm, to the effect that

$$(f+g)'=f'+g',\quad (fg)'=f'g+fg',\quad \left(\frac{1}{f}\right)'=-\frac{f'}{f^2},\quad (f\circ g)'=(f'\circ g)g'.$$

The notion of complex differentiability is thus much more manageable than the notion of analyticity.

It follows from a well known theorem of Riemann (see for instance [**229**, vol. 1, p 143] and APPENDIX B: *Equivalent definitions of analyticity*, p. 659) that analyticity and complex differentiability are equivalent notions.

THEOREM IV.1 (Basic Equivalence Theorem). *A function is analytic in a region Ω if and only if it is complex-differentiable in Ω.*

The following are known facts (see again Appendix B): if a function is analytic (equivalently complex-differentiable) in Ω, it admits (complex) derivatives of any order there. This property markedly differs from real analysis: complex differentiable

[3]A mapping of the plane that locally preserves angles is also called a *conformal* map.

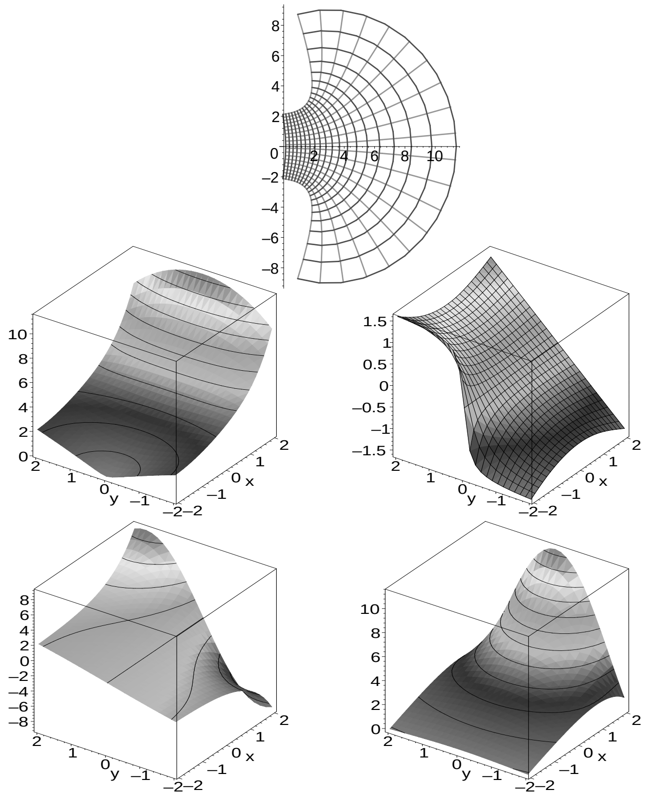


FIGURE 3. Multiple views of an analytic function. The image of the domain $\Omega = \{z \mid |\Re(z)| < 2, |\Im(z)| < 2\}$ by $f(z) = \exp(z) + z + 2$: [top] transformation of a square grid in Ω by f; [middle] the modulus and argument of $f(z)$; [bottom] the real and imaginary parts of $f(z)$.

(equivalently, analytic) functions are all smooth. Also derivatives of a function are obtained through term-by-term differentiation of the series representation of the function.

Meromorphic functions. We finally introduce *meromorphic*[4] functions that are mild extensions of the concept of analyticity (or holomorphy) and are essential to the theory.

The quotient of two analytic functions $f(z)/g(z)$ ceases to be analytic at a point a where $g(a) = 0$. However, a simple structure for quotients of analytic functions prevails.

DEFINITION IV.3. *A function $h(z)$ is* meromorphic *at z_0 iff, for z in a neighbourhood of z_0 with $z \neq z_0$, it it can be represented as $f(z)/g(z)$, with $f(z)$ and $g(z)$ being*

[4]"Holomorphic" and "meromorphic" are words coming from Greek, meaning respectively "of complete form" and "of partial form".

analytic at z_0*. In that case, it admits near* z_0 *an expansion of the form*

$$h(z) = \sum_{n \geq -M} h_n (z - z_0)^n. \tag{10}$$

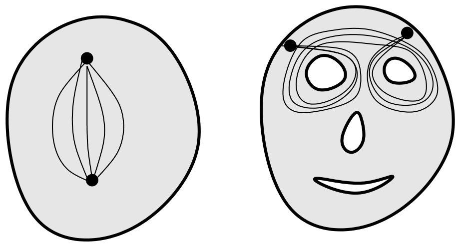
If $h_{-M} \neq 0$ *and* $M \geq 1$*, then* $h(z)$ *is said to have a* pole *of order* M *at* $z = a$*. The coefficient* h_{-1} *is called the* residue *of* $h(z)$ *at* $z = a$ *and is written as*

$$\operatorname{Res}[h(z); z = a].$$

A function is meromorphic in a region iff it is meromorphic at any point of the region.

IV. 2.2. Integrals and residues. A path in a region Ω is described by its parameterization, which is a continuous function γ mapping $[0, 1]$ into Ω. Two paths γ, γ' in Ω having the same end points are said to be *homotopic* (in Ω) if one can be continuously deformed into the other while staying within Ω as in the following examples:

homotopic paths:



A closed path[5] is defined by the fact that its end points coincide: $\gamma(0) = \gamma(1)$, and a path is *simple* if the mapping γ is one-to-one. A closed path is said to be a *loop* of Ω if it can be continuously deformed *within* Ω to a single point; in this case one also says that the path is homotopic to 0. In what follows we implicitly restrict attention to paths that are assumed to be rectifiable. Unless otherwise stated, all integration paths will be assumed to be oriented positively.

Integrals along curves in the complex plane are defined in the usual way as curvilinear integrals of complex-valued functions. Explicitly: let $f(x + iy)$ be a function and γ be a path; then,

$$\begin{aligned} \int_\gamma f(z)\, dz &:= \int_0^1 f(\gamma(t))\gamma'(t)\, dt \\ &= \int_0^1 [AC - BD]\ dt + i \int_0^1 [AD + BC]\ dt, \end{aligned}$$

where $f = A + iB$ and $\gamma' = C + iD$. However integral calculus in the complex plane is of a radically different nature from what it is on the real line—in a way it is much simpler and much more powerful. One has:

THEOREM IV.2 (Null Integral Property). *Let* f *be analytic in* Ω *and let* λ *be a simple loop of* Ω*. Then* $\int_\lambda f = 0$*.*

[5]By default, paths used in this book are assumed to be positively oriented piecewise continuously differentiable (hence rectifiable); in addition, closed paths are assumed to be positively oriented.

Equivalently, integrals are largely independent of details of contours: for f analytic in Ω, one has

$$\int_\gamma f = \int_{\gamma'} f, \tag{11}$$

provided γ and γ' are homotopic (not necessarily closed) paths in Ω. A proof of Theorem IV.2 is sketched in APPENDIX B: *Equivalent definitions of analyticity*, p. 659.

Residues. The important *Residue Theorem* due to Cauchy relates *global* properties of a meromorphic function (its integral along closed curves) to purely *local* characteristics at designated points (the residues at poles).

THEOREM IV.3 (Cauchy's residue theorem). *Let $h(z)$ be meromorphic in the region Ω and let λ be a simple loop in Ω along which the function is analytic. Then*

$$\frac{1}{2i\pi}\int_\lambda h(z)\,dz = \sum_s \operatorname{Res}[h(z); z = s],$$

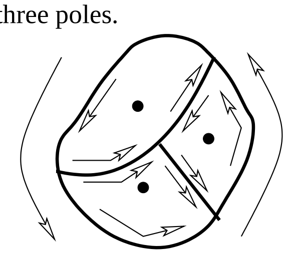
where the sum is extended to all poles s of $h(z)$ enclosed by λ.

PROOF. (Sketch) To see it in the representative case where $h(z)$ has only a pole at $z = 0$, observe by appealing to primitive functions that

$$\int_\lambda h(z)\,dz = \sum_{\substack{n \geq -M \\ n \neq -1}} h_n \left[\frac{z^{n+1}}{n+1}\right]_\lambda + h_{-1}\int_\lambda \frac{dz}{z},$$

where the bracket notation $\big[u(z)\big]_\lambda$ designates the variation of the function $u(z)$ along the contour λ. This expression reduces to its last term, itself equal to $2i\pi h_{-1}$, as is checked by using integration along a circle (set $z = re^{i\theta}$). The computation extends by translation to the case of a unique pole at $z = a$.

In the case of multiple poles, we observe that the simple loop can only enclose finitely many poles (by compactness). The proof then follows from a simple decomposition of the interior domain of λ into cells each containing only one pole. Here is an illustration in the case of three poles.



(Contributions from internal edges cancel.) $\square$

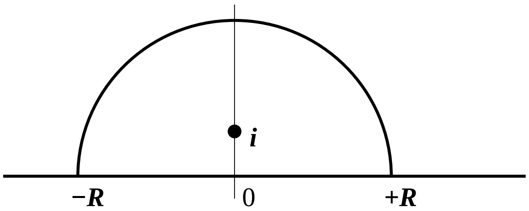
Global (integral) to local (residues) connections. Here is a textbook example of a reduction from global to local properties of analytic functions. Define the integrals

$$I_m := \int_{-\infty}^{\infty} \frac{dx}{1 + x^{2m}},$$

and consider specifically I_1. Elementary calculus teaches us that $I_1 = \pi$ since the antiderivative of the integrand is an arc tangent:

$$I_1 = \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = [\arctan x]_{-\infty}^{+\infty} = \pi.$$

Here is an alternative, and in many ways more fruitful, derivation. In the light of the residue theorem, we consider the integral over the whole line as the limit of integrals over large intervals of the form $[-R, +R]$, then complete the contour of integration by means of a large semi-circle in the upper half-plane, as shown below:



Let γ be the contour comprised of the interval and the semi-circle. Inside γ, the integrand has a pole at $x = i$, where

$$\frac{1}{1+x^2} \equiv \frac{1}{(x+i)(x-i)} = -\frac{i}{2}\frac{1}{x-i} + \cdots ,$$

so that its residue there is $-i/2$. By the residue theorem, the integral taken over γ is equal to $2i\pi$ times the residue of the integrand at i. As $R \to \infty$, the integral along the semi-circle vanishes (it is less than $\pi R/(1+R^2)$ in modulus), while the integral along the real segment gives I_1 in the limit. There results the relation giving I_1:

$$I_1 = 2i\pi \operatorname{Res}\left(\frac{1}{1+x^2}; x = i\right) = (2i\pi)\left(-\frac{i}{2}\right) = \pi.$$

The evaluation of the integral in the framework of complex analysis rests solely upon the local expansion of the integrand at special points (here, the point i). This is a remarkable feature of the theory, one that confers it much simplicity, when compared to real analysis.

$\triangleright$ **4.** *The general integral* I_m. Let $\alpha = \exp(\frac{i\pi}{2m})$ so that $\alpha^{2m} = -1$. Contour integration of the type used for I_1 yields

$$I_m = 2i\pi \sum_{j=1}^{m} \operatorname{Res}\left(\frac{1}{1+x^{2m}}; x = \alpha^{2j-1}\right),$$

while, for any $\beta = \alpha^{2j-1}$ with $1 \le j \le m$, one has

$$\frac{1}{1+x^{2m}} \underset{x\to\beta}{\sim} \frac{1}{2m\beta^{2m-1}}\frac{1}{x-\beta} \equiv -\frac{\beta}{2m}\frac{1}{x-\beta}.$$

As a consequence,

$$I_{2m} = -\frac{i\pi}{m}\left(\alpha + \alpha^3 + \cdots + \alpha^{2m-1}\right) = \frac{\pi}{m \sin \frac{\pi}{2m}}.$$

In particular, $I_2 = \pi/\sqrt{2}$, $I_3 = 2\pi/3$, $I_4 = \frac{\pi}{4}\sqrt{2}\sqrt{2+\sqrt{2}}$, and $\frac{1}{\pi}I_5$, $\frac{1}{\pi}I_6$ are expressible by radicals, but $\frac{1}{\pi}I_7$, $\frac{1}{\pi}I_9$ are not. The special cases $\frac{1}{\pi}I_{17}$, $\frac{1}{\pi}I_{257}$ are expressible by radicals. $\triangleleft$

$\triangleright$ **5.** *Integrals of rational fractions.* Generally, all integrals of rational functions taken over the whole real line are computable by residues. In particular,

$$J_m = \int_{-\infty}^{+\infty} \frac{dx}{(1+x^2)^m}, \qquad K_m = \int_{-\infty}^{+\infty} \frac{dx}{(1^2+x^2)(2^2+x^2)\cdots(m^2+x^2)}$$

can be explicitly evaluated. $\triangleleft$

Cauchy's coefficient formula. Many function-theoretic consequences derive from the residue theorem. For instance, if f is analytic in Ω, $z_0 \in \Omega$ and λ is a simple loop of Ω encircling z_0, one has

$$f(z_0) = \frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{\zeta - z_0}. \tag{12}$$

This follows directly since

$$\operatorname{Res}[f(\zeta)/(\zeta - z_0); \zeta = z_0] = f(z_0).$$

Then, by differentiation with respect to z_0 under the integral sign, one gets similarly

$$\frac{1}{k!} f^{(k)}(z_0) = \frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{(\zeta - z_0)^{k+1}}. \tag{13}$$

The values of a function and its derivatives at a point can thus be obtained as values of integrals of the function away from that point. The world of analytic functions is a very gentle one in which to live: contrary to real analysis, a function is differentiable *any number of times* as soon as it is differentiable *once*. Also, Taylor's formula invariably holds: as soon as $f(z)$ is analytic at z_0, one has

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{1}{2!} f''(z_0)(z - z_0)^2 + \cdots, \tag{14}$$

with the representation being convergent in a small disc centred at z_0. [Proof: a verification from (12) and (13), or a series rearrangement as in (B.7), p. 660.]

A very important application of the residue theorem concerns coefficients of analytic functions.

THEOREM IV.4 (Cauchy's Coefficient Formula). *Let $f(z)$ be analytic in a region containing 0 and let λ be a simple loop around 0 that is positively oriented. Then the coefficient $[z^n]f(z)$ admits the integral representation*

$$f_n \equiv [z^n]f(z) = \frac{1}{2i\pi} \int_\lambda f(z) \frac{dz}{z^{n+1}}.$$

PROOF. This formula follows directly from the equalities

$$\frac{1}{2i\pi} \int_\lambda f(z) \frac{dz}{z^{n+1}} = \operatorname{Res}\left[f(z) z^{-n-1}; z = 0\right] = [z^n]f(z),$$

of which the first follows from the residue theorem, and the second from the identification of the residue at 0 as a coefficient. $\square$

Analytically, the coefficient formula allows one to deduce information about the coefficients from the values of the function itself, using adequately chosen contours of integration. It thus opens the possibility of estimating the coefficients $[z^n]f(z)$ in the expansion of $f(z)$ near 0 by using information on $f(z)$ *away* from 0. The rest of this chapter will precisely illustrate this process in the case of rational and meromorphic functions. Observe also that the residue theorem provides the simplest known proof of the Lagrange inversion theorem (see APPENDIX A: *Lagrange Inversion*, p. 649) whose rôle is central to tree enumerations, as we saw in Chapters I and II. The notes below explore some independent consequences of the residue theorem and the coefficient formula.

▷ **6.** *Liouville's Theorem*. If a function $f(z)$ is analytic in the whole of $\mathbb{C}$ and is of modulus bounded by an absolute constant, $|f(z)| \le B$, then it must be a constant. [By trivial bounds, upon integrating on a large circle, it is found that the Taylor coefficients at the origin of index ≥ 1 are all equal to 0.] Similarly, if $f(z)$ is of at most polynomial growth, $|f(z)| \le B\,(|z|+1)^r$ over the whole of $\mathbb{C}$, then it must be a polynomial. ◁

▷ **7.** *Lindelöf integrals.* Let $a(s)$ be analytic in $\Re(s) > \frac{1}{4}$ where it is assumed to satisfy $a(s) = O(\exp((\pi-\delta)|s|))$ for some δ with $0 < \delta < \pi$. Then, one has for $|\arg(z)| < \delta$,

$$\sum_{k=1}^{\infty} a(k)(-z)^k = -\frac{1}{2i\pi}\int_{1/2-i\infty}^{1/2+i\infty} a(s) z^s \,\frac{\pi}{\sin \pi s}\, ds,$$

in the sense that the integral exists and provides the analytic continuation of the sum in $|\arg(z)| < \delta$. [Close the integration contour by a large semi-circle on the right and evaluate by residues.] Such integrals, sometimes called Lindelöf integrals, provide representations for many functions whose Taylor coefficients are given by an explicit rule [**187, 292**]. ◁

▷ **8.** *Continuation of polylogarithms.* As a consequence of Lindelöf's representation, the generalized *polylogarithm* functions,

$$\mathrm{Li}_{\alpha,k}(z) = \sum_{n\ge 1} n^{-\alpha}(\log n)^k z^n \qquad (\alpha \in \mathbb{R}, \quad k \in \mathbb{Z}_{\ge 0}),$$

are analytic in the complex plane $\mathbb{C}$ slit along $(1+, \infty)$. (More properties are presented in Section VI. 8; see also [**147, 187**].) For instance, one obtains in this way

$$\text{``}\sum_{n=1}^{\infty}(-1)^n \log n\text{''} = -\frac{1}{4}\int_{-\infty}^{+\infty} \frac{\log(\frac{1}{4}+t^2)}{\cosh(\pi t)}\,dt = 0.22579\cdots = \log\sqrt{\frac{\pi}{2}},$$

when the divergent series on the left is interpreted as $\mathrm{Li}_{0,1}(-1) = \lim_{z\to -1^+} \mathrm{Li}_{0,1}(z)$. ◁

▷ **9.** *Magic duality.* Let ϕ be a function initially defined over the nonnegative integers but admitting a meromorphic extension over the whole of $\mathbb{C}$. Under growth conditions in the style of Note 7, the function

$$F(z) := \sum_{n\ge 1} \phi(n)(-z)^n,$$

which is analytic at the origin, is such that, near positive infinity,

$$F(z) \underset{z\to+\infty}{\sim} E(z) - \sum_{n\ge 1} \phi(-n)(-z)^{-n},$$

for some elementary function $E(z)$. [Starting from the representation of Note 7, close the contour of integration by a large semicircle to the left.] In such cases, the function is said to satisfy the principle of *magic duality*—its expansion at 0 and ∞ are given by one and the same rule. Functions

$$\frac{1}{1+z}, \quad \log(1+z), \quad \exp(-x), \quad \mathrm{Li}_2(-z), \quad \mathrm{Li}_3(-z),$$

as well as hypergeometric functions (and many other!) satisfy a form of magic duality. Ramanujan [**40**] made a great use of this principle, which applies to a wide class of functions including hypergeometric ones; see Hardy's insightful discussion [**224**, Ch XI]. ◁

▷ **10.** *Euler–Maclaurin and Abel–Plana summations.* Under simple conditions on the analytic function f, one has Plana's (also known as Abel's) complex variables version of the Euler–Maclaurin summation formula:

$$\sum_{n=0}^{\infty} f(n) = \frac{1}{2}f(0) + \int_0^{\infty} f(x)\,dx + \int_0^{\infty} \frac{f(iy)-f(-iy)}{e^{2i\pi y}-1}\,dy.$$

(See [**230**, p. 274] for a proof and validity conditions.) ◁

▷ **11.** *Nörlund-Rice integrals.* Let $a(z)$ be analytic for $\Re(z) > k_0 - \frac{1}{2}$ and of at most polynomial growth in this right half plane. Then, with γ a simple loop around the interval $[k_0, n]$, one has

$$\sum_{k=k_0}^{n} \binom{n}{k} (-1)^{n-k} a(k) = \frac{1}{2i\pi} \int_\gamma a(s)\, \frac{n!\,ds}{s(s-1)(s-2)\cdots(s-n)}.$$

If $a(z)$ is meromorphic in a larger region, then the integral can be estimated by residues. For instance, with

$$S_n = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k}, \qquad T_n = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^2+1},$$

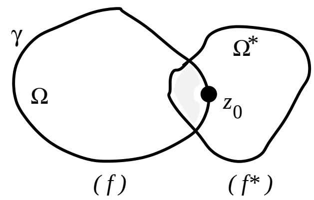
it is found that $S_n = -H_n$ (a harmonic number), while T_n oscillates boundedly as $n \to +\infty$. [This technique is a classical one in the calculus of finite differences, going back to Nörlund [**327**]. In computer science it is known as the method of Rice's integrals [**174**] and is used in the analysis of many algorithms and data structures including digital trees and radix sort [**269, 401**].] ◁

IV. 3. Singularities and exponential growth of coefficients

For a given function, a singularity can be informally defined as a point where the function ceases to be analytic. (Poles are the very simplest type of singularity.) Singularities are, as we have stressed repeatedly, essential to coefficient asymptotics. This section presents the bases of a discussion within the framework of analytic function theory.

IV. 3.1. Singularities. Let $f(z)$ be an analytic function defined over the interior region determined by a simple closed curve γ, and let z_0 be a point of the bounding curve γ. If there exists an analytic function $f^*(z)$ defined over some open set Ω^* containing z_0 and such that $f^*(z) = f(z)$ in $\Omega^* \cap \Omega$, one says that f is *analytically continuable* at z_0 and that $f^\star$ is an *immediate analytic continuation* of f.

Analytic continuation:



$f^*(z) = f(z)$ on $\Omega^* \cap \Omega$.

Consider for instance the quasi-inverse function, $f(z) = 1/(1-z)$. Its power series representation $f(z) = \sum_{n\ge 0} z^n$ initially converges in $|z| < 1$. However, the calculation of (8) shows that it is representable locally by a convergent series near any point $z_0 \neq 1$. In particular, it is continuable at any point of the unit disc except 1. (Alternatively, one may appeal to complex-differentiability to verify directly that $f(z)$, which is given by a "global" expression, is holomorphic, hence analytic, in the punctured plane $\mathbb{C} \setminus \{1\}$.)

In sharp contrast to real analysis where a function admits of many smooth extensions, analytic continuation is essentially *unique*: if f^* (in $\Omega^\star$) and f^{**} (in $\Omega^{\star\star}$) continue f at z_0, then one must have $f^*(z) = f^{**}(z)$ in the intersection $\Omega^\star \cap \Omega^{\star\star}$, which in particular includes a small disc around z_0. Thus, the notion of immediate analytic continuation at a boundary point is intrinsic. The process can be iterated and we say that g is an *analytic continuation*[6] of f along a path γ, even if the domains of definition of f and g do not overlap, provided a finite chain of intermediate functions connects f and g. This notion is once more intrinsic—this is known as the principle of *unicity of analytic continuation* (Rudin [**368**, Ch. 16] provides a thorough discussion). An analytic function is then much like a hologram: as soon as it is specified in any tiny region, it is rigidly determined in any wider region where it can be continued.

DEFINITION IV.4. *Given a function f defined in the region interior to the simple closed curve γ, a point z_0 on the boundary (γ) of the region is a* singular point *or a* singularity[7] *if f is* not *analytically continuable at z_0.*

Granted the intrinsic character of analytic continuation, we can usually dispense with a detailed description of the original domain Ω and the curve γ. In simple terms, a function is singular at z_0 if it cannot be continued as an analytic function beyond z_0. A point at which a function is analytic is also called by contrast a *regular point*.

The two functions $f(z) = 1/(1-z)$ and $g(z) = \sqrt{1-z}$ may be taken as initially defined over the open unit disk by their power series representation. Then, as we already know, they can be analytically continued to larger regions, the punctured plane $\Omega = \mathbb{C} \setminus \{1\}$ for f [e.g., by the calculation of (8)] and the complex plane slit along $(1, +\infty)$ for g [e.g., by virtue of differentiability as in (9)]. But both are singular at 1: for f, this results from the fact that (say) $f(z) \to \infty$ as $z \to 1$; for g this is due to the branching character of the square-root. Figure 4 displays a few types of singularities that are traceable by the way they deform a regular grid near a boundary point.

It is easy to check from the definitions that a converging power series is analytic inside its disc of convergence. In other words, it can have no singularity inside this disc. However, it *must* have at least one singularity on the boundary of the disc, as asserted by the theorem below. In addition, a classical theorem, called Pringsheim's theorem, provides a refinement of this property in the case of functions with nonnegative coefficients, which includes all combinatorial generating functions.

THEOREM IV.5 (Boundary singularities). *A function analytic $f(z)$ at the origin whose expansion at the origin has a finite radius of convergence R necessarily has a singularity on the boundary of its disc of convergence, $|z| = R$.*

PROOF. Consider the expansion

$$f(z) = \sum_{n \geq 0} f_n z^n, \tag{15}$$

[6] The collection of all function elements continuing a given function gives rise to the notion of *Riemann surface*, for which many good books exist, e.g., [**129, 389**]. We shall normally avoid appealing to this theory.

[7] For a detailed discussion, see [**106**, p. 229], [**261**, vol. 1, p. 82], or [**411**].

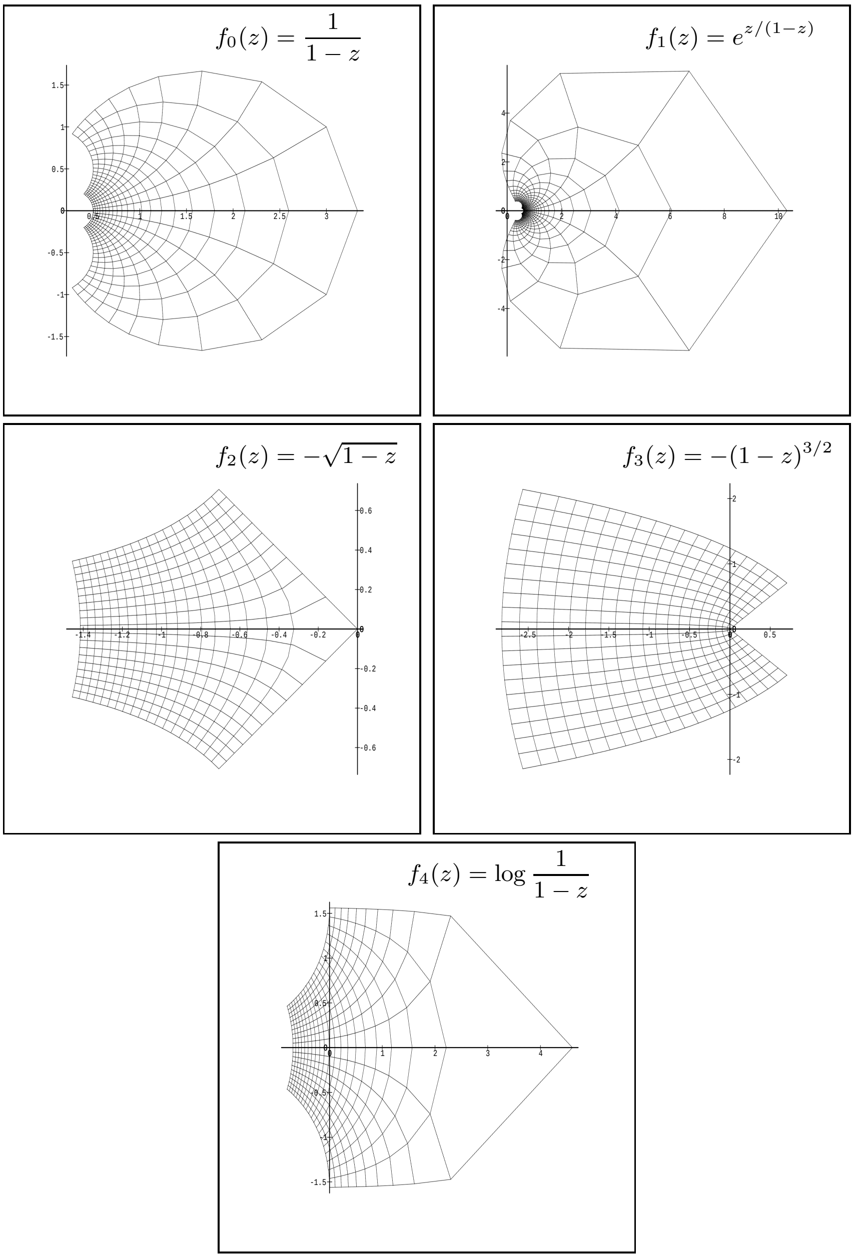


FIGURE 4. The images of a grid on the unit square (with corners $\pm 1 \pm i$) by various functions singular at $z = 1$ reflect the nature of the singularities involved. Singularities are apparent near the right of each diagram where small grid squares get folded or unfolded in various ways. (In the case of functions f_0, f_1, f_4 that become infinite at $z = 1$, the grid has been slightly truncated to the right.)

assumed to have radius of convergence exactly R. We already know that there can be no singularity of f within the disc $|z| < R$. To prove that there is a singularity on $|z| = R$, suppose a contrario that $f(z)$ is analytic in the disc $|z| < \rho$ for some ρ satisfying $\rho > R$. By Cauchy's coefficient formula (Theorem IV.4), upon integrating along the circle of radius $r = (R + \rho)/2$, and by trivial bounds, it is seen that the coefficient $[z^n]f(z)$ is $O(r^{-n})$. But then, the series expansion of f would have to converge in the disc of radius $r > R$, a contradiction. $\square$

Pringsheim's Theorem stated and proved now is a refinement of Theorem IV.5 that applies to *all* series having nonnegative coefficients, in particular, generating functions. It is central to asymptotic enumeration as the remainder of this section will amply demonstrate.

THEOREM IV.6 (Pringsheim's Theorem). *If $f(z)$ is representable at the origin by a series expansion that has nonnegative coefficients and radius of convergence R, then the point $z = R$ is a singularity of $f(z)$.*

$\triangleright$ **12.** *Proof of Pringsheim's Theorem.* (See also [**411**, Sec. 7.21].) In a nutshell, the idea of the proof is that if f has positive coefficients and is analytic at R, then its expansion slightly to the left of R has positive coefficients. Then the power series of f would converge in a disc larger than the postulated sic of convergence—a clear contradiction.

Suppose a contrario that $f(z)$ is analytic at R, implying that it is analytic in a disc of radius r centred at R. We choose a number h such that $0 < h < \frac{1}{3}r$ and consider the expansion of $f(z)$ around $z_0 = R - h$:

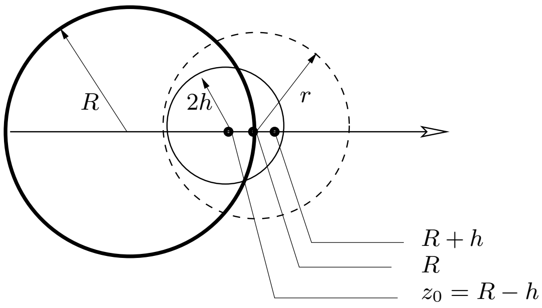
$$f(z) = \sum_{m \geq 0} g_m (z - z_0)^m. \tag{16}$$

By Taylor's formula and the representability of $f(z)$ together with its derivatives at z_0 by means of (15), we have

$$g_m = \sum_{n \geq 0} \binom{n}{m} f_n z_0^{n-m},$$

and in particular, $g_m \geq 0$.

Given the way h was chosen, the series (16) converges at $z = R + h$ (so that $z - z_0 = 2h$) as illustrated by the following diagram:



Consequently, one has

$$f(R + h) = \sum_{m \geq 0} \left(\sum_{n \geq 0} \binom{n}{m} f_n z_0^{m-n} \right) (2h)^m.$$

This is a converging double sum of positive terms, so that the sum can be reorganized in any way we like. In particular, one has convergence of all the series involved in

$$\begin{aligned} f(R+h) &= \sum_{m,n\geq 0} \binom{n}{m} f_n (R-h)^{m-n}(2h)^m \\ &= \sum_{n\geq 0} f_n \left[(R-h)+(2h)\right]^n \\ &= \sum_{n\geq 0} f_n (R+h)^n. \end{aligned}$$

This establishes the fact that $f_n = o((R+h)^n)$, thereby reaching a contradiction with the assumption that the serie representation of f has radius of convergence exactly R. Pringsheim's theorem is proved. $\triangleleft$

Singularities of a function analytic at 0 which lie on the boundary of the disc of convergence are called *dominant singularities*. Pringsheim's theorem appreciably simplifies the search for dominant singularities of combinatorial generating functions since these have nonnegative coefficients—it is then sufficient to investigate analyticity along the positive real line and detect the first place at which it ceases to hold.

For instance, the derangement EGF and the surjection EGF,

$$D(z) = \frac{e^{-z}}{1-z}, \qquad S(z) = (2-e^z)^{-1}$$

are analytic except for a simple pole at $z = 1$ in the case of $D(z)$, and except for points $z_k = \log 2 + 2ik\pi$ that are simple poles in the case of $S(z)$. Thus the dominant singularities for derangements and surjections are at 1 and $\log 2$ respectively.

It is known that $\sqrt{Z}$ cannot be unambiguously defined as an analytic function in a neighbourhood of $Z = 0$. As a consequence, the function

$$C(z) = (1-\sqrt{1-4z})/2,$$

which is the generating function of the Catalan numbers, is an analytic function in regions that must exclude $1/4$; for instance, one may opt to take the complex plane slit along the ray $(1/4, +\infty)$. Similarly, the function

$$L(z) = \log \frac{1}{1-z}$$

which is the EGF of cyclic permutations is analytic in the complex plane slit along $(1, +\infty)$.

A function having no singularity at a finite distance is called *entire*; its Taylor series then converges everywhere in the complex plane. The EGFs,

$$e^{z+z^2/2} \qquad \text{and} \qquad e^{e^z-1},$$

associated respectively with involutions and set partitions, are entire.

IV. 3.2. The Exponential Growth Formula. We say that a number sequence $\{a_n\}$ is of *exponential order* K^n which we abbreviate as (the symbol $\bowtie$ is a "bowtie")

$$a_n \bowtie K^n \qquad \text{iff} \qquad \limsup |a_n|^{1/n} = K.$$

The relation $X \bowtie Y$ reads as "X *is of exponential order* Y". It expresses both an upper bound and a lower bound, and one has, for any $\epsilon > 0$:

(i) $|a_n| >_{i.o} (K-\epsilon)^n$, that is to say, $|a_n|$ exceeds $(K-\epsilon)^n$ infinitely often (for infinitely many values of n);

(ii) $|a_n| <_{a.e.} (K+\epsilon)^n$, that is to say, $|a_n|$ is dominated by $(K+\epsilon)^n$ almost everywhere (except for possibly finitely many values of n).

This relation can be rephrased as $a_n = K^n \vartheta(n)$, where ϑ is a *subexponential factor* satisfying

$$\limsup |\theta(n)|^{1/n} = 1;$$

such a factor is thus bounded from above almost everywhere by any increasing exponential (of the form $(1+\epsilon)^n$) and bounded from below infinitely often by any decaying exponential (of the form $(1-\epsilon)^n$). Typical subexponential factors are

$$1,\ n^3,\ (\log n)^2,\ \sqrt{n},\ \frac{1}{\sqrt[3]{\log n}},\ n^{-3/2},\ \log\log n.$$

(Functions like $e^{\sqrt{n}}$ and $\exp(\log^2 n)$ are to be treated as subexponential factors for the purpose of this discussion.) The $\limsup$ definition also allows in principle for factors that are infinitely often very small or 0, like $n^2 \sin n\frac{\pi}{2}$, $\log n \cos \sqrt{n}\frac{\pi}{2}$, and so on. In this and the next chapters, we shall develop systematic methods that enable one to extract such subexponential factors from generating functions.

It is an elementary observation that the radius of convergence of the series representation of $f(z)$ at 0 is related to the exponential growth rate of the coefficients $f_n = [z^n]f(z)$. To wit, if $\mathrm{R}_{\mathrm{conv}}(f;0) = R$, then we claim that

$$(17)\qquad f_n \bowtie \left(\frac{1}{R}\right)^n, \qquad \text{i.e.,}\quad f_n = R^{-n}\theta(n) \quad \text{with } \limsup |\theta(n)|^{1/n} = 1.$$

$\triangleright$ **13.** *Radius of convergence and exponential growth.* This only requires the basic definition of a power series. (i) By definition of the radius of convergence, we have for any small $\epsilon > 0$, $f_n(R-\epsilon)^n \to 0$. In particular, $|f_n|(R-\epsilon)^n < 1$ for all sufficiently large n, so that $|f_n|^{1/n} < (R-\epsilon)^{-1}$ "almost everywhere". (ii) In the other direction, for any $\epsilon > 0$, $|f_n|(R+\epsilon)^n$ cannot be a bounded sequence, since otherwise, $\sum_n |f_n|(R+\epsilon/2)^n$ would be a convergent series. Thus, $|f_n|^{1/n} > (R+\epsilon)^{-1}$ "infinitely often". $\triangleleft$

A global approach to the determination of growth rates is desirable. This is made possible by Theorem IV.5.

THEOREM IV.7 (Exponential Growth Formula). *If $f(z)$ is analytic at 0 and R is the modulus of a singularity nearest to the origin in the sense that*[8]

$$R := \sup\left\{ r \geq 0 \;\middle|\; f \textit{ is analytic in } |z| < r \right\},$$

then the coefficient $f_n = [z^n]f(z)$ satisfies

$$f_n \bowtie \left(\frac{1}{R}\right)^n.$$

[8]One should think of the process defining R as follows: take discs of increasing radii r and stop as soon as a singularity is encountered on the boundary. (The dual process that would start from a large disc and restrict its radius is in general ill-defined—think of $\sqrt{1-z}$.)

For functions with nonnegative coefficients, including all combinatorial generating functions, one can also adopt

$$R := \sup \left\{ r \geq 0 \ \middle|\ f \text{ is analytic at all points of } 0 \leq z < r \right\}.$$

PROOF. Let R be as stated. We cannot have $R < \mathrm{R}_{\mathrm{conv}}(f;0)$ since a function is analytic everywhere in the interior of the disc of convergence of its series representation. We cannot have $R > \mathrm{R}_{\mathrm{conv}}(f;0)$ by the Boundary Singularity Theorem. Thus $R = \mathrm{R}_{\mathrm{conv}}(f;0)$. The statement then follows from (17). The adaptation to nonnegative coefficients results from Pringsheim's theorem. □

The exponential growth formula thus directly relates the exponential order of growth of coefficients of a function to the *location* of its singularities nearest to the origin. This is precisely expressed by the *First Principle of Coefficient Asymptotics* (p. 215), which, given its importance, we repeat here:

> **First Principle of Coefficient Asymptotics.** *The* location *of a function's singularities dictates the* exponential growth *(A^n) of its coefficient.*

Several direct applications to combinatorial enumeration are given below.

EXAMPLE 1. *Exponential growth and combinatorial enumeration.* Here are a few immediate applications of of exponential bounds.

Surjections. The function

$$R(z) = (2 - e^z)^{-1}$$

is the EGF of surjections. The denominator is an entire function, so that singularities may only arise from its zeros, to be found at the points

$$\chi_k = \log 2 + 2ik\pi, \qquad k \in \mathbb{Z}.$$

The dominant singularity of R is then at $\rho = \chi_0 = \log 2$. Thus, with $r_n = [z^n]R(z)$,

$$r_n \bowtie \left(\frac{1}{\log 2}\right)^n.$$

Similarly, if "double" surjections are considered (each value in the range of the surjection is taken at least twice), the corresponding EGF is

$$R^*(z) = \frac{1}{2 + z - e^z},$$

with the counts starting as 1,0,1,1,7,21,141 (*EIS* **A032032**). The dominant singularity is at ρ^* defined as the positive root of equation $e^{\rho^*} - \rho^* = 2$, and the coefficient r_n^* satisfies: $r_n^* \bowtie (\frac{1}{\rho^*})^n$ Numerically, this gives

$$r_n \bowtie 1.44269^n \qquad \text{and} \qquad r_n^* \bowtie 0.87245^n,$$

with the actual figures for the corresponding logarithms being

n	$\frac{1}{n}\log r_n$	$\frac{1}{n}\log r_n^*$
10	0.33385	-0.22508
20	0.35018	-0.18144
50	0.35998	-0.154449
100	0.36325	-0.145447
∞	0.36651	-0.13644
	$(\log 1/\rho)$	$(\log(1/\rho^*)$

These estimates constitute a weak form of a more precise result to be established later in this chapter: If random surjections of size n are taken equally likely, the probability of a surjection being a double surjection is exponentially small.

Derangements. There, for $d_{1,n} = [z^n]e^{-z}(1-z)^{-1}$ and $d_{2,n} = [z^n]e^{-z-z^2/2}(1-z)^{-1}$ we have, from the poles at $z = 1$,

$$d_{1,n} \bowtie 1^n \quad \text{and} \quad d_{2,n} \bowtie 1^n.$$

The upper bound is combinatorially trivial. The lower bound expresses that the probability for a random permutation to be a derangement is *not* exponentially small. For $d_{1,n}$, we have already proved by an elementary argument the stronger result $d_{1,n} \to e^{-1}$; in the case of $d_{2,n}$, we shall establish later the precise asymptotic equivalent $d_{2,n} \to e^{-3/2}$, in accordance with what was announced in the introduction.

Unary-Binary trees. The expression

$$U(z) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z} = z + z^2 + 2\,z^3 + 4\,z^4 + 9\,z^5 + \cdots,$$

represents the OGF of (plane unlabelled) unary-binary trees. From the equivalent form,

$$U(z) = \frac{1-z-\sqrt{(1-3z)(1+z)}}{2z},$$

it follows that $U(z)$ is analytic in the complex plane slit along $(\frac{1}{3}, +\infty)$ and $(-\infty, -1)$ and is singular at $z = -1$ and $z = 1/3$ where it has branch points. The closest singularity to the origin being at $\frac{1}{3}$, one has

$$U_n \bowtie 3^n.$$

In this case, the stronger upper bound $U_n \le 3^n$ results directly from the possibility of encoding such trees by words over a ternary alphabet using Łukasiewicz codes (Chapter I). A complete asymptotic expansion will be obtained in Chapter VI. . . . . END OF EXAMPLE 1. □

▷ **14.** *Coding theory bounds.* Let $\mathcal{C}$ be a combinatorial class. We say that it *can be encoded with $f(n)$ bits* if, for all sufficiently large values of n, elements of $\mathcal{C}_n$ can be encoded as words of $f(n)$ bits. Assume that $\mathcal{C}$ has OGF $C(z)$ with radius of singularity R satisfying $0 < R < 1$. Then, for any ϵ, $\mathcal{C}$ can be encoded with $(1+\epsilon)\kappa n$ bits where $\kappa = -\log_2 R$, but $\mathcal{C}$ cannot be encoded with $(1-\epsilon)\kappa n$ bits.

Similarly, if $\mathcal{C}$ has EGF $\widehat{C}(z)$ with radius of convergence R satisfying $0 < R < \infty$, $\mathcal{C}$ can be encoded with $n\log(n/e) + (1+\epsilon)\kappa n$ bits where $\kappa = -\log_2 R$, but $\mathcal{C}$ cannot be encoded with $n\log(n/e) + (1-\epsilon)\kappa n$ bits. *Singularities convey information on optimal codes!* ◁

Saddle-point bounds. The exponential growth formula (Theorem IV.7) can be supplemented by effective upper bounds which are very easy to derive and often turn out to be surprisingly accurate. We state:

PROPOSITION IV.1 (Saddle-Point bounds). *Let $f(z)$ be analytic in the disc $|z| < R$ with $0 < R \le \infty$. Define $M(f;r)$ for $r \in (0,R)$ by $M(f;r) := \sup_{|z|=r} |f(z)|$. Then, one has, for* any r *in* $(0,R)$*, the family of* saddle point upper bounds

$$(18) \qquad [z^n]f(z) \le \frac{M(f;r)}{r^n} \qquad \textit{implying} \qquad [z^n]f(z) \le \inf_{r\in(0,R)} \frac{M(f;r)}{r^n}.$$

If in addition $f(z)$ has nonnegative coefficients at 0, then

$$(19) \qquad [z^n]f(z) \le \frac{f(r)}{r^n} \qquad \textit{implying} \qquad [z^n]f(z) \le \inf_{r\in(0,R)} \frac{f(r)}{r^n}.$$

PROOF. In the general case of (18), the first inequality results from trivial bounds applied to the Cauchy coefficient formula, when integration is performed along a circle:

$$[z^n]f(z) = \frac{1}{2i\pi}\int_{|z|=r} f(z)\,\frac{dz}{z^{n+1}}.$$

It is consequently valid for any r smaller than the radius of convergence of f at 0. The second inequality in (18) plainly represents the best possible bound of this type.

In the positive case of (19), the bounds can be viewed as a direct specialization of (18). (Alternatively, they can be obtained elementarily since, in the case of positive coefficients,

$$f_n \le \frac{f_0}{r^n} + \cdots + \frac{f_{n-1}}{r^{n-1}} + f_n + \frac{f_{n+1}}{r^{n+1}} + \cdots\,,$$

whenever the f_k are nonnegative.) □

Note that the value s that provides the best bound in (19) can be determined by cancelling a derivative,

$$s\frac{f'(s)}{f(s)} = n. \tag{20}$$

Thanks to the universal character of the first bound, *any* approximate solution of this last equation will in fact provide a valid upper bound.

For reasons well explained by the saddle point method (Chapter VIII), these bounds usually capture the actual asymptotic behaviour up to a polynomial factor only. A typical instance is the weak form of Stirling's formula,

$$\frac{1}{n!} \equiv [z^n]e^z \le \frac{e^n}{n^n},$$

which only overestimates the true asymptotic value by a factor of $\sqrt{2\pi n}$.

▷ **15.** *A suboptimal but easy saddle-point bound.* Let $f(z)$ be analytic in $|z| < 1$ with nonnegative coefficients. Assume that $f(x) \le (1-x)^{-\beta}$ for some $\beta \ge 0$ and all $x \in (0,1)$. Then

$$[z^n]f(z) = O(n^\beta).$$

(Better bounds of the form $O(n^{\beta-1})$ are usually obtained by the method of singularity analysis exposed in Chapter VI.) ◁

EXAMPLE 2. *Combinatorial examples of saddle point bounds.* Here are applications to fragmented permutations, set partitions (Bell numbers), involutions, and integer partitions.

Fragmented permutations. Consider first fragmented permutations defined by $\mathcal{F} = \mathrm{SET}(\mathrm{SEQ}_{\ge 1}(\mathcal{Z}))$ in the labelled universe (Chapter II, p. 116). The EGF is $e^{z/(1-z)}$, and we claim that

$$\frac{1}{n!}F_n \equiv [z^n]e^{z/(1-z)} \le e^{2\sqrt{n}-\frac{1}{2}+O(n^{-1/2})}. \tag{21}$$

Indeed, the minimizing radius of the saddle point bound (19) is s such that

$$0 = \frac{d}{ds}\left(\frac{s}{1-s} - n\log s\right) = \frac{1}{(1-s)^2} - \frac{n}{s}.$$

The equation is solved by $s = (2n+1-\sqrt{4n+1})/(2n)$. One can either use this exact value and compute an asymptotic approximation of $f(s)/s^n$, or adopt right away the approximate

n	$\widetilde{I}_n$	I_n
100	$0.106579 \cdot 10^{85}$	$0.240533 \cdot 10^{83}$
200	$0.231809 \cdot 10^{195}$	$0.367247 \cdot 10^{193}$
300	$0.383502 \cdot 10^{316}$	$0.494575 \cdot 10^{314}$
400	$0.869362 \cdot 10^{444}$	$0.968454 \cdot 10^{442}$
500	$0.425391 \cdot 10^{578}$	$0.423108 \cdot 10^{576}$

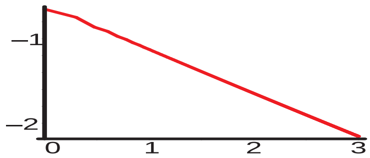


FIGURE 5. A comparison of the exact number of involutions I_n to its approximation $\widetilde{I}_n = n!e^{\sqrt{n}+n/2}n^{-n/2}$: [left] a table; [right] a plot of $\log_{10}(I_n/\widetilde{I}_n)$ against $\log_{10} n$ suggesting that the ratio satisfies $I_n/\widetilde{I}_n \sim K \cdot n^{-1/2}$, the slope of the line being $\approx \frac{1}{2}$.

value $s_1 = 1 - 1/\sqrt{n}$, which leads to simpler calculations. The estimate (21) results. It is off from the actual asymptotic value only by a factor of order $n^{-3/4}$ (cf Example VIII.5, p. 509).

Bell numbers and set partitions. Another immediate applications is an upper bound on Bell numbers enumerating set partitions, $\mathcal{S} = \text{SET}(\text{SET}_{\geq 1}(\mathcal{Z}))$, with EGF e^{e^z-1}. According to (20), the best saddle point bound is obtained for s such that $se^s = n$. Thus,

$$\frac{1}{n!}S_n \leq e^{e^s-1-n\log s}, \qquad s\ :\ se^s = n, \tag{22}$$

where, additionally, $s = \log n - \log\log n + o(\log\log n)$. See Chapter VIII, p. 507 for the complete saddle point analysis.

Involutions. Involutions are specified by $\mathcal{I} = \text{SET}(\text{CYC}_{1,2}(\mathcal{Z}))$ and have EGF $I(z) = \exp(z + \frac{1}{2}z^2)$. One determines, by choosing $s = \sqrt{n}$ as an approximate solution to (20):

$$\frac{1}{n!}I_n \leq \frac{e^{\sqrt{n}+n/2}}{n^{n/2}}. \tag{23}$$

(See Figure 5 for numerical data and Example VIII.3, p. 505 for a full analysis.) Similar bounds hold for permutations with all cycle lengths $\leq k$ and permutations σ such that $\sigma^k = Id$.

Integer partitions. The function

$$P(z) = \prod_{k=1}^{\infty} \frac{1}{1-z^k} = \exp\left(\sum_{\ell=1}^{\infty} \frac{1}{\ell}\frac{z^\ell}{1-z^\ell}\right) \tag{24}$$

is the OGF of integer partitions, an unlabelled analogue of set partitions. Its radius of convergence is *a priori* bounded from above by 1, since the set $\mathcal{P}$ is infinite and the second form of $P(z)$ shows that it is exactly equal to 1. Therefore $P_n \bowtie 1^n$. A finer upper bound results from the estimate

$$\Lambda(t) := \log P(e^{-t}) \sim \frac{\pi^2}{6t} + \log\sqrt{\frac{t}{2\pi}} - \frac{1}{24}t + O(t^2), \tag{25}$$

which obtains from Euler–Maclaurin summation or, better, from a Mellin analysis following APPENDIX B: *Mellin transform*, p. 674. Indeed, the Mellin transform of Λ is, by the harmonic sum rule,

$$\Lambda^\star(s) = \zeta(s)\zeta(s+1)\Gamma(s), \qquad s \in \langle 1, +\infty\rangle,$$

and the successive leftmost poles at $s = 1$ (simple pole), $s = 0$ (double pole), and $s = -1$ (simple pole) translate into the asymptotic expansion (25). When $z \to 1^-$, we have

$$P(z) \sim \frac{e^{-\pi^2/12}}{\sqrt{2\pi}}\sqrt{1-z}\exp\left(\frac{\pi^2}{6(1-z)}\right), \tag{26}$$

from which we derive (choose $s = D\sqrt{n}$ as an approximate solution to (20))

$$P_n \leq Cn^{-1/4} e^{\pi\sqrt{2n/3}},$$

for some $C > 0$. This last bound is once more only off by a polynomial factor, as we shall prove when studying the saddle point method (Proposition VIII.3, p. 513). END OF EXAMPLE 2. □

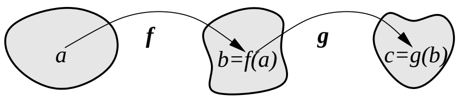
▷ **16.** *A natural boundary.* One has $P(re^{i\theta}) \to \infty$ as $r \to 1^-$, for any angle θ that is a rational multiple of 2π. Points $e^{2i\pi p/q}$ being dense on the unit circle, the function $P(z)$ admits the unit circle as a *natural boundary*, i.e., it cannot be analytically continued beyond this circle. ◁

IV. 4. Closure properties and computable bounds

Analytic functions are robust: they satisfy a rich set of closure properties. This fact makes possible the determination of exponential growth constants for coefficients of a wide range of classes of functions. Theorem IV.8 below expresses computability of growth rate for all specifications associated with iterative specifications. It is the first result of this sort that relates symbolic methods of Part A with analytic methods developed here.

Closure properties of analytic functions. The functions analytic at a point $z = a$ are closed under sum and product, and hence form a ring. If $f(z)$ and $g(z)$ are analytic at $z = a$, then so is their quotient $f(z)/g(z)$ provided $g(a) \neq 0$. Meromorphic functions are furthermore closed under quotient and hence form a field. Such properties are proved most easily using complex-differentiability and extending the usual relations from real analysis, for instance, $(f + g)' = f' + g'$, $(fg)' = fg' = f'g$.

Analytic functions are also closed under composition: if $f(z)$ is analytic at $z = a$ and $g(w)$ is analytic at $b = f(a)$, then $g \circ f(z)$ is analytic at $z = a$. Graphically:



The proof based on complex-differentiability closely mimicks the real case. Inverse functions exist conditionally: if $f'(a) \neq 0$, then $f(z)$ is locally linear near a, hence invertible, so that there exists a g satisfying $f \circ g = g \circ f = Id$, where Id is the identity function, $Id(z) \equiv z$. The inverse function is itself locally linear, hence complex differentiable, hence analytic. In short, the inverse of an analytic function f at a place where the derivative does not vanish is an analytic function.

▷ **17.** *The analytic inversion lemma.* Let f be analytic on $\Omega \ni z_0$ and satisfy $f'(z_0) \neq 0$. Then there esists a small region $\Omega_1 \subseteq \Omega$ containing z_0 and a $C > 0$ such that $|f(z) - f(z')| > C|z - z'|$, for all $z, z' \in \Omega_1$. Consequently, f maps bijectively Ω_1 on $f(\Omega_1)$. ◁

One way to establish closure properties, as suggested above, is to deduce analyticity criteria from complex differentiability by way of the Basic Equivalence Theorem (Theorem IV.1). An alternative approach, closer to the original notion of analyticity, can be based on a two-step process: (i) closure properties are shown to hold true for formal power series; (ii) the resulting formal power series are proved to be locally convergent by means of suitable majorizations on their coefficients. This is the basis of the classical method of *majorant series* originating with Cauchy.

$\triangleright$ **18.** *The majorant series technique.* Given two power series, define $f(z) \preceq g(z)$ if $|[z^n]f(z)| \leq [z^n]g(z)$ for all $n \geq 0$. The following two conditions are equivalent: (i) $f(z)$ is analytic in the disc $|z| < \rho$; (ii) for any $r > \rho^{-1}$ there exists a c such that

$$f(z) \preceq \frac{c}{1 - rz}.$$

If f, g are majorized by $c/(1-rz), d/(1-rz)$ respectively, then $f+g$ and $f \cdot g$ are majorized,

$$f(z) + g(z) \preceq \frac{c+d}{1-rz}, \qquad f(z) \cdot g(z) \preceq \frac{e}{1-sz},$$

for any $s > r$ and for some e dependent on s. Similarly, the composition $f \circ g$ is majorized:

$$f \circ g(z) \preceq \frac{c}{1 - r(1+d)z}.$$

Constructions for $1/f$ and for the functional inverse of f can be similarly developed. See Cartan's book [**73**] and van der Hoeven's study [**419**] for a systematic treatment. $\triangleleft$

For functions defined by analytic expressions, singularities can be determined inductively in an intuitively transparent manner. If $\mathrm{Sing}(f)$ and $\mathrm{Zero}(f)$ are respectively the set of singularities and zeros of function f, then, due to closure properties of analytic functions, the following informally stated guidelines apply.

$$\begin{cases} \mathrm{Sing}(f \pm g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \\ \mathrm{Sing}(f \times g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \\ \mathrm{Sing}(f/g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \cup \mathrm{Zero}(g) \\ \mathrm{Sing}(f \circ g) & \subseteq & \mathrm{Sing}(g) \cup g^{(-1)}(\mathrm{Sing}(f)) \\ \mathrm{Sing}(\sqrt{f}) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Zero}(f) \\ \mathrm{Sing}(\log(f)) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Zero}(f) \\ \mathrm{Sing}(f^{(-1)}) & \subseteq & f(\mathrm{Sing}(f)) \cup f(\mathrm{Zero}(f')). \end{cases}$$

A mathematically rigorous treatment would require considering multivalued functions and Riemann surfaces, so that we do not state detailed validity conditions and, at this stage, keep for these formulæ the status of useful heuristics. In fact, because of Pringsheim's theorem, the search of dominant singularities of combinatorial generating function can normally avoid considering the complete multivalued structure of functions, since only some initial segment of the positive real half–line needs to be considered. This in turn implies a powerful and easy way of determining the exponential order of coefficients of a wide variety of generating functions, as we explain next.

Computability of exponential growth constants. As defined in Chapters I and II, a combinatorial class is *constructible* or *specifiable* if it can be specified by a finite set of equations involving only the basic constructors. A specification is *iterative* or *non-recursive* if in addition the dependency graph of the specification is acyclic, that is, no recursion is involved and a single functional term (written with sums, products, as well as sequence, set, and cycle constructions) describes the specification.

Our interest here is in effective computability issues. We recall that a real number α is computable iff there exists a program Π_α which on input m outputs a rational number α_m guaranteed to be within $\pm 10^{-m}$ of α. We state:

THEOREM IV.8 (Computability of growth). *Let $\mathcal{C}$ be a* constructible *unlabelled class that admits of an* iterative *specification in terms of* $(\textsc{Seq}, \textsc{PSet}, \textsc{MSet}, \textsc{Cyc}; +, \times)$

starting with $(1, \mathcal{Z})$. *Then the radius of convergence* ρ_C *of the OGF* $C(z)$ *of* $\mathcal{C}$ *is either* $+\infty$ *or a (strictly) positive computable real number.*

Let $\mathcal{D}$ *be a* constructible *labelled class that admits of an* iterative *specification in terms of* $(\text{SEQ}, \text{SET}, \text{CYC}; +, \star)$ *starting with* $(1, \mathcal{Z})$. *Then the radius of convergence* ρ_D *of the EGF* $D(z)$ *of* $\mathcal{D}$ *is either* $+\infty$ *or a (strictly) positive computable real number.*

Accordingly, if finite, the constants ρ_C, ρ_D *in the exponential growth estimates,*

$$[z^n]C(z) \equiv C_n \bowtie \left(\frac{1}{\rho_C}\right)^n, \qquad [z^n]D(z) \equiv \frac{1}{n!}D_n \bowtie \left(\frac{1}{\rho_D}\right)^n,$$

are computable numbers.

PROOF. In both cases, the proof proceeds by induction on the structural specification of the class. For each class $\mathcal{F}$, with generating function $F(z)$, we associate a *signature*, which is an ordered pair $\langle \rho_F, \tau_F \rangle$, where ρ_F is the radius of convergence of F and τ_F is the value of F at ρ_F, precisely,

$$\tau_F := \lim_{x \to \rho_F^-} F(x).$$

(The value τ_F is well defined as an element of $\mathbb{R} \cup \{+\infty\}$ since F, being a counting generating function, is necessarily increasing on $(0, \rho_F)$.)

Unlabelled case. An unlabelled class $\mathcal{G}$ is either finite, in which case its OGF $G(z)$ is a polynomial, or infinite, in which case it diverges at $z = 1$, so that $\rho_G \leq 1$. It is clearly decidable, given the specification, whether a class is finite or not: a necessary and sufficient condition is that one of the unary constructors $(\text{SEQ}, \text{MSET}, \text{CYC})$ intervenes in the specification. We prove (by induction) the assertion of the theorem together with the stronger property that $\tau_F = \infty$ as soon as the class is infinite.

First, the signatures of the neutral class 1 and the atomic class $\mathcal{Z}$, with OGF 1 and z, are $\langle +\infty, 1 \rangle$ and $\langle +\infty, +\infty \rangle$. Any nonconstant polynomial which is the OGF of a finite set has the signature $\langle +\infty, +\infty \rangle$. The assertion is thus easily verified in these cases.

Next, let $\mathcal{F} = \text{SEQ}(\mathcal{G})$. The OGF $G(z)$ must be nonconstant and in fact satisfy $G(0) = 0$ in order for the sequence construction to be properly defined. Thus, by the induction hypothesis, one has $0 < \rho_G \leq +\infty$ and $\tau_G = +\infty$. Now, the function G being increasing and continuous along the positive axis, there must exist a value β such that $0 < \beta < \rho_G$ with $G(\beta) = 1$. For $z \in (0, \beta)$, the quasi-inverse $F(z) = (1 - G(z))^{-1}$ is well defined and analytic; as z approaches β from the left, $F(z)$ increases unboundedly. Thus, the smallest singularity of F along the positive axis is at β, and by Pringsheim's theorem, one has $\rho_F = \beta$. The argument simultaneously shows that $\tau_F = +\infty$. There only remains to check that β is computable. The coefficients of G form a computable sequence of integers, so that $G(x)$, which can be well approximated via truncated Taylor series, is an effectively computable number[9]

[9]The present argument only establishes non-constructively the *existence* of a program, based on the fact that truncated Taylor series converge geometrically fast at an interior point of their disc of convergence. Making explict this program and the involved parameters from the specification itself however represents a much harder problem (that of "uniformity" with respect to specifications) that is not addressed here.

if x is itself a positive computable number less than ρ_G. Then binary search provides an effective procedure for determining β.

Next, we consider the multiset construction, $\mathcal{F} = \text{MSET}(\mathcal{G})$, whose translation into OGFs necessitates the Pólya exponential:

$$F(z) = \text{Exp}(G(z)) \quad \text{where } \text{Exp}(h(z)) := \exp\left(h(z) + \frac{1}{2}h(z^2) + \frac{1}{3}h(z^3) + \cdots\right).$$

Once more, the induction hypothesis is assumed for G. If G is a polynomial, then F is a rational function with poles at roots of unity only. Thus, $\rho_F = 1$ and $\tau_F = \infty$ in that particular case. In the general case of $\mathcal{F} = \text{MSET}(\mathcal{G})$ with $\mathcal{G}$ infinite, we start by fixing arbitrarily a number r such that $0 < r < \rho_G \leq 1$ and examine $F(z)$ for $z \in (0, r)$. The expression for F rewrites as

$$\text{Exp}(G(z)) = e^{G(z)} \cdot \exp\left(\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots\right).$$

The first factor is analytic for z on $(0, \rho_G)$ since, the exponential function being entire, e^G has the singularities of G. As to the second factor, one has $G(0) = 0$ (in order for the set construction to be well-defined), while $G(x)$ is convex for $x \in [0, r]$ (since its second derivative is positive). Thus, there exists a positive constant K such that $G(x) \leq Kx$ when $x \in [0, r]$. Then, the series $\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots$ has its terms dominated by those of the convergent series

$$\frac{K}{2}r^2 + \frac{K}{3}r^3 + \cdots = K \log(1 - r)^{-1} - Kr.$$

By a well known theorem of analytic function theory, a uniformly convergent sum of analytic functions is itself analytic; consequently, $\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots$ is analytic at all z of $(0, r)$. Analyticity is then preserved by the exponential, so that $F(z)$, being analytic at $z \in (0, r)$ for any $r < \rho_G$ has a radius of convergence that satisfies $\rho_F \geq \rho_G$. On the other hand, since $F(z)$ dominates termwise $G(z)$, one has $\rho_F \leq \rho_G$. Thus finally one has $\rho_F = \rho_G$. Also, $\tau_G = +\infty$ implies $\tau_F = +\infty$.

A parallel discussion covers the case of the powerset construction (PSET) whose associated functional $\overline{\text{Exp}}$ is a minor modification of the Pólya exponential Exp. The cycle construction can be treated by similar arguments based on consideration of "Pólya's logarithm" as $\mathcal{F} = \text{CYC}(\mathcal{G})$ corresponds to

$$F(z) = \text{Log}\,\frac{1}{1 - G(z)}, \qquad \text{where} \quad \text{Log}\, h(z) = \log h(z) + \frac{1}{2}\log h(z^2) + \cdots.$$

In order to conclude with the unlabelled case, there only remains to discuss the binary constructors $+$, $\times$, which give rise to $F = G + H$, $F = G \cdot H$. It is easily verified that $\rho_F = \min(\rho_G, \rho_H)$. Computability is granted since the minimum of two computable numbers is computable. That $\tau_F = +\infty$ in each case is immediate.

Labelled case. The labelled case is covered by the same type of argument as above, the discussion being even simpler, since the ordinary exponential and logarithm replace the Pólya operators Exp and Log. It is still a fact that all the EGFs of infinite nonrecursive classes are infinite at their dominant positive singularity, though the radii of convergence can now be of any magnitude (compared to 1). $\square$

▷ **19.** *Relativized constructions.* This is an exercise in induction. Theorem IV.8 is stated for specifications involving the basic constructors. Show that the conclusion still holds if the corresponding relativized constructions ($\mathfrak{K}_{=r}, \mathfrak{K}_{<r}, \mathfrak{K}_{>r}$ with $\mathfrak{K}$ being any of the basic constructors) are also allowed. ◁

▷ **20.** *Syntactically decidable properties.* For unlabelled classes $\mathcal{F}$, the property $\rho_F = 1$ is decidable. For labelled and unlabelled classes, the property $\rho_F = +\infty$ is decidable. ◁

▷ **21.** *Pólya–Carlson and a curious property of OGFs.* Here is a statement first conjectured by Pólya, then proved by Carlson in 1921 (see [**105**, p. 323]): *If a function is represented by a power series with integer coefficients that converges inside the unit disc, then either it is a rational function or it admits the unit circle as a natural boundary.* This theorem applies in particular to the OGF of any combinatorial class. ◁

▷ **22.** *Trees are recursive structures only!* General and binary trees cannot receive an iterative specification since their OGFs assume a finite value at their Pringsheim singularity. [The same is true of most simple families of treee; cf Proposition IV.5 p. 265]. ◁

▷ **23.** *Nonconstructibility of permutations and graphs.* The class $\mathcal{P}$ of all permutations cannot be specified as a constructible unlabelled class since the OGF $P(z) = \sum_n n! z^n$ has radius of convergence 0. (It is of course constructible as a labelled class.) Graphs, whether labelled or unlabelled, are too numerous to form a constructible class. ◁

Theorem IV.8 establishes a link between analytic combinatorics, computability theory, and symbolic manipulation systems. It is based on an article of Flajolet, Salvy, and Zimmermann [**173**] devoted to such computability issues in exact and asymptotic enumeration. (Recursive specifications are not discussed now since they tend to give rise to branch points, themselves amenable to singularity analysis techniques to be developed in Chapters VI and VII.) The inductive process, implied by the proof of Theorem IV.8, that decorates a specification with the radius of convergence of each of its subexpressions provides a practical basis for determining the exponential growth rate of counts associated to a nonrecursive specification. The example of trains detailed below is typical.

EXAMPLE 3. *Combinatorial trains.* This somewhat artificial example from [**144**] (see Figure 6) serves to illustrate the scope of Theorem IV.8 and demonstrate its inner mechanisms at work. Define the class of all *labelled trains* by the following specification,

$$(27) \qquad \begin{cases} \mathcal{T}r &=& \mathcal{W}a \star \mathrm{SEQ}(\mathcal{W}a \star \mathrm{SET}(\mathcal{P}a)), \\ \mathcal{W}a &=& \mathrm{SEQ}_{\geq 1}(\mathcal{P}\ell), \\ \mathcal{P}\ell &=& \mathcal{Z} \star \mathcal{Z} \star (\mathbf{1} + \mathrm{CYC}(\mathcal{Z})), \\ \mathcal{P}a &=& \mathrm{CYC}(\mathcal{Z}) \star \mathrm{CYC}(\mathcal{Z}). \end{cases}$$

In figurative terms, a train ($\mathcal{T}$) is composed of a first wagon ($\mathcal{W}a$) to which is appended a sequence of passenger wagons, each of the latter capable of containing a set of passengers ($\mathcal{P}a$). A wagon is itself composed of "planks" ($\mathcal{P}\ell$) determined by their end points ($\mathcal{Z} \star \mathcal{Z}$) and to which a circular wheel ($\mathrm{CYC}(\mathcal{Z})$) may be attached. A passenger is composed of a head and

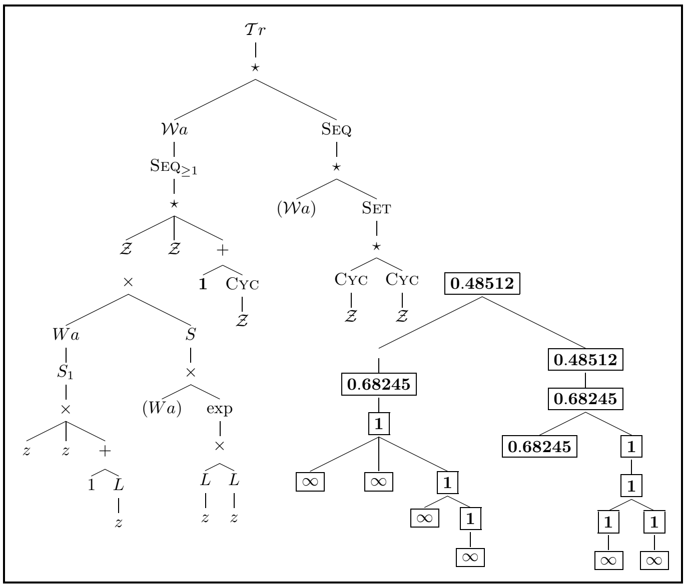
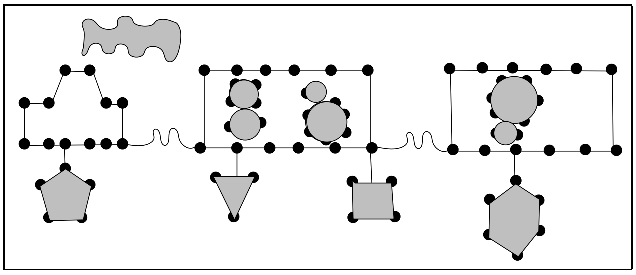


FIGURE 6. The inductive determination of the radius of convergence of the EGF of trains: (top) a hierarchical view of the specification of $\mathcal{T}r$; (bottom left) the corresponding expression tree of the EGF $Tr(z)$; (bottom right) the value of the radii for each subexpression of $Tr(z)$ (with $L(y) = \log(1-y)^{-1}$, $S(y) = (1-y)^{-1}$, $S_1(y) = yS(y)$).

a belly that are each circular arrangements of atoms. Here is a depiction of a random train:



The translation into a set of EGF equations is immediate and a symbolic manipulation system readily provides the form of the EGF of trains as

$$Tr(z) = \frac{z^2\left(1+\log((1-z)^{-1})\right)}{\left(1-z^2\left(1+\log((1-z)^{-1})\right)\right)}\left(1-\frac{z^2\left(1+\log((1-z)^{-1})\right)e^{\left(\log((1-z)^{-1})\right)^2}}{1-z^2\left(1+\log((1-z)^{-1})\right)}\right)^{-1},$$

together with the expansion

$$Tr(z) = 2\,\frac{z^2}{2!} + 6\,\frac{z^3}{3!} + 60\,\frac{z^4}{4!} + 520\,\frac{z^5}{5!} + 6660\,\frac{z^6}{6!} + 93408\,\frac{z^7}{7!} + \cdots .$$

The specification (27) has a hierarchical structure, as suggested by the top representation of Figure 6, and this structure is itself directly reflected by the form of the expression tree of the GF $Tr(z)$. Then each node in the expression tree of $Tr(z)$ can be tagged with the corresponding value of the radius of convergence. This is done according to the principles of Theorem IV.8; see the bottom-right part of Figure 6. For instance, the quantity 0.68245 associated to $Wa(z)$ is given by the sequence rule and is determined as smallest positive solution to the equation

$$z^2\left(1 - \log(1-z)^{-1}\right) = 1.$$

The tagging process works upwards till the root of the tree is reached; here the radius of convergence of Tr is determined to be $\rho \doteq 0.48512\cdots$, a quantity that happens to coincide with the ratio $[z^{49}]Tr(z)/[z^{50}]Tr(z)$ to more than 15 decimal places. ... END OF EXAMPLE 3. □

IV. 5. Rational and meromorphic functions

The last section has fully justified the *First Principle* of coefficient asymptotics leading to the exponential growth formula $f_n \bowtie A^n$ for the coefficients of an analytic function $f(z)$. Indeed, as we saw, one has $A = 1/\rho$, where ρ equals both the radius of convergence of the series representing f and the distance of the origin to the dominant, i.e., closest, singularities. We are going to start examining here the *Second Principle*, already quoted on p. 215 and relative to the form,

$$f_n = A^n\theta(n),$$

with $\theta(n)$ the subexponential factor:

> **Second Principle of Coefficient Asymptotics.** *The* nature *of the function's singularities determines the associate* subexponential factor *($\theta(n)$).*

In this section, we develop a complete theory in the case of rational functions (that is, quotients of polynomials) and, more generally, meromorphic functions. The net result is that, for such functions, the subexponential factors are essentially polynomials:

Polar singularities $\quad\rightsquigarrow\quad$ Subexponential factors $\theta(n)$ are of *polynomial growth*.

A distinguishing feature is the extremely good quality of the asymptotic approximations obtained; for naturally occuring combinatorial problems, 15 digits of accuracy is not uncommon in coefficients of index as low as 50 (see Figure 7 below for a striking example).

IV. 5.1. Rational functions. A function $f(z)$ is a *rational function* iff it is of the form $f(z) = N(z)/D(z)$, with $N(z)$ and $D(z)$ being polynomials, which we may without loss of generality assume to be relatively prime. For rational functions that are analytic at the origin (e.g., generating functions), we have $D(0) \neq 0$.

Sequences $\{f_n\}_{n\geq 0}$ that are coefficients of rational functions satisfy linear recurrence relations with constant coefficients. This fact is easy to establish: compute $[z^n]f(z)\cdot D(z)$; then, with $D(z) = d_0 + d_1 z + \cdots + d_m z^m$, one has, for all $n > \deg(N(z))$,

$$\sum_{j=0}^{m} d_j f_{n-j} = 0.$$

The main theorem we prove here provides an *exact* finite expression for coefficients of $f(z)$ in terms of the poles of $f(z)$. Individual terms in these expressions are sometimes called *exponential polynomials*.

THEOREM IV.9 (Expansion of rational functions). *If $f(z)$ is a rational function that is analytic at zero and has poles at points $\alpha_1, \alpha_2, \ldots, \alpha_m$, then its coefficients are a sum of* exponential polynomials*: there exist m polynomials $\{\Pi_j(x)\}_{j=1}^m$ such that, for n larger than some fixed n_0,*

$$f_n \equiv [z^n]f(z) = \sum_{j=1}^{m} \Pi_j(n)\alpha_j^{-n}. \tag{28}$$

Furthermore the degree of Π_j is equal to the order of the pole of f at α_j minus one.

PROOF. Since $f(z)$ is rational it admits a partial fraction expansion. To wit:

$$f(z) = Q(z) + \sum_{(\alpha,r)} \frac{c_{\alpha,r}}{(z-\alpha)^r},$$

where $Q(z)$ is a polynomial of degree $n_0 := \deg(N) - \deg(D)$ if $f = N/D$. There α ranges over the poles of $f(z)$ and r is bounded from above by the multiplicity of α as a pole of f. Coefficient extraction in this expression results from Newton's expansion,

$$[z^n]\frac{1}{(z-\alpha)^r} = \frac{(-1)^r}{\alpha^r}[z^n]\frac{1}{(1-\frac{z}{\alpha})^r} = \frac{(-1)^r}{\alpha^r}\binom{n+r-1}{r-1}\alpha^{-n}.$$

The binomial coefficient is a polynomial of degree $r-1$ in n, and collecting terms associated with a given α yields the statement of the theorem. □

Notice that the expansion (28) is also an asymptotic expansion in disguise: when grouping terms according to the α's of increasing modulus, each group appears to be *exponentially smaller* than the previous one. In particular, if there is a unique dominant pole, $|\alpha_1| < |\alpha_2| \le |\alpha_3| \le \cdots$, then

$$f_n \sim \alpha_1^{-n}\Pi_1(n),$$

and the error term is exponentially small as it is $O(\alpha_2^{-n}n^r)$ for some r. A classical instance is the OGF of Fibonacci numbers,

$$f(z) = \frac{z}{1-z-z^2},$$

with poles at $\dfrac{-1+\sqrt{5}}{2} \doteq 0.61803$ and $\dfrac{-1-\sqrt{5}}{2} \doteq -1.61803$, so that

$$F_n = \frac{1}{\sqrt{5}}\varphi^n - \frac{1}{\sqrt{5}}\bar{\varphi}^n = \frac{\varphi^n}{\sqrt{5}} + O(\frac{1}{\varphi^n}),$$

with $\varphi = (1+\sqrt{5})/2$ the golden ratio, and $\bar{\varphi}$ its conjugate.

▷ **24.** *A simple exercise.* Let $f(z)$ be as in Theorem IV.9, assuming additionally a unique dominant pole α_1 of multiplicity r. Then, by inspection of the proof of Theorem IV.9:

$$f_n = \frac{C}{(r-1)!}\alpha_1^{-n-1}n^{r-1}\left(1 + O\left(\frac{1}{n}\right)\right) \qquad \text{with} \quad C = \lim_{z\to\alpha_1}(z-\alpha_1)^r f(z).$$

This is certainly the most direct illustration of the Second Principle: under the assumptions, a one-term asymptotic expansion of the functon at its dominant singularity suffices to determine the asymptotic form of the coefficients. ◁

EXAMPLE 4. *Qualitative analysis of a rational function.* This is an artificial example designed to demonstrate that all the details of the full decomposition are usually not required. The rational function

$$f(z)=\frac{1}{(1-z^3)^2(1-z^2)^3(1-\frac{z^2}{2})}$$

has a pole of order 5 at $z=1$, poles of order 2 at $z=\omega,\omega^2$ ($\omega=e^{2i\pi/3}$ a cubic root of unity), a pole of order 3 at $z=-1$, and simple poles at $z=\pm\sqrt{2}$. Therefore,

$$f_n = P_1(n)+P_2(n)\omega^{-n}+P_3(n)\omega^{-2n}+P_4(n)(-1)^n+$$
$$+P_5(n)2^{-n/2}+P_6(n)(-1)^n 2^{-n/2}$$

where the degrees of $P_1,\ldots,P_6$ are respectively $4,1,1,2,0,0$. For an asymptotic equivalent of f_n, only the poles at roots of unity need to be considered since they corresponds to the fastest exponential growth; in addition, only $z=1$ needs to be considered for first order asymptotics; finally, at $z=1$, only the term of fastest growth needs to be taken into account. In this way, we find: the correspondence

$$f(z)\sim\frac{1}{3^2\cdot 2^3\cdot(\frac{1}{2})}\frac{1}{(1-z)^5}\implies f_n\sim\frac{1}{3^2\cdot 2^3\cdot(\frac{1}{2})}\binom{n+4}{4}\sim\frac{n^4}{864}.$$

The way the analysis can be developed *without computing details* of partial fraction expansion is typical. .... END OF EXAMPLE 4. □

Theorem IV.9 applies to any specification leading to a GF that is a rational function[10]. Combined with the qualitative approach to rational coefficient asymptotics, it gives access to a large number of effective asymptotic estimates for combinatorial counting sequences.

EXAMPLE 5. *Asymptotics of denumerants.* Denumerants are integer partitions with summands restricted to be from a *fixed* finite set (Chapter I, p. 41). We let $\mathcal{P}^{\mathcal{T}}$ be the class relative to set $\mathcal{T}\subset\mathbb{Z}_{>0}$, with the known OGF,

$$P^{\mathcal{T}}(z)=\prod_{\omega\in\mathcal{T}}\frac{1}{1-z^\omega}.$$

A particular case is the one of integer partitions whose summands are in $\{1,2,\ldots,r\}$,

$$P^{\{1,\ldots,r\}}(z)=\prod_{m=1}^{r}\frac{1}{1-z^m}.$$

The GF has all its poles that are roots of unity. At $z=1$, the order of the pole is r, and one has

$$P^{\{1,\ldots,r\}}(z)\sim\frac{1}{r!}\frac{1}{(1-z)^r},$$

as $z\to 1$. Other poles have smaller multiplicity: for instance the multiplicity of $z=-1$ is equal to the number of factors $(1-z^{2j})^{-1}$ in $P^{\{1,\ldots,r\}}$, that is $\lfloor r/2\rfloor$; in general a primitive qth root of unity is found to have multiplicity $\lfloor r/q\rfloor$. There results that $z=1$ contributes a term of

[10]In Part A, we have been occasionally led to discuss coefficients of rational functions, thereby anticipating the statement of the theorem: see for instance the discussion of parts in compositions (p. 158) and of records in sequences (p. 179).

the form n^{r-1} to the coefficient of order n, while each of the other poles contributes a term of order at most $n^{\lfloor r/2 \rfloor}$. We thus find

$$P_n^{\{1,\dots,r\}} \sim c_r n^{r-1} \qquad \text{with} \quad c_r = \frac{1}{r!(r-1)!}.$$

The same argument provides the asymptotic form of $P_n^{\mathcal{T}}$, since, to first order asymptotics, only the pole at $z = 1$ counts. One then has:

PROPOSITION IV.2. *Let $\mathcal{T}$ be a finite set of integers without a common divisor ($\gcd(\mathcal{T}) = 1$). The number of partitions with summands restricted to $\mathcal{T}$ satisfies*

$$P_n^{\mathcal{T}} \sim \frac{1}{\tau} \frac{n^{r-1}}{(r-1)!}, \qquad \textit{with} \quad \tau := \prod_{n \in \mathcal{T}} n, \quad r := \operatorname{card}(\mathcal{T}).$$

For instance, in a strange country that would have pennies (1 cent), nickels (5 cents), dimes (10 cents), and quarters (25 cents), the number of ways to make change for a total of n cents is

$$[z^n] \frac{1}{(1-z)(1-z^5)(1-z^{10})(1-z^{25})} \sim \frac{1}{1 \cdot 5 \cdot 10 \cdot 25} \frac{n^3}{3!} \equiv \frac{n^3}{7500},$$

asymptotically. . . . . END OF EXAMPLE 5. □

IV. 5.2. Meromorphic Functions. An expansion similar to that of Theorem IV.9 holds true for coefficients of a larger class—meromorphic functions.

THEOREM IV.10 (Expansion of meromorphic functions). *Let $f(z)$ be a function meromorphic for $|z| \le R$ with poles at points $\alpha_1, \alpha_2, \dots, \alpha_m$, and analytic at all points of $|z| = R$ and at $z = 0$. Then there exist m polynomials $\{\Pi_j(x)\}_{j=1}^m$ such that:*

$$f_n \equiv [z^n]f(z) = \sum_{j=1}^{m} \Pi_j(n) \alpha_j^{-n} + \mathcal{O}(R^{-n}). \tag{29}$$

Furthermore the degree of Π_j is equal to the order of the pole of f at α_j minus one.

PROOF. We offer two different proofs, one based on subtracted singularities, the other one based on contour integration.

(i) *Subtracted singularities.* Around any pole α, $f(z)$ can be expanded locally:

$$f(z) = \sum_{k \ge -M} c_{\alpha,k}(z-\alpha)^k \tag{30}$$

$$= S_\alpha(z) + H_\alpha(z) \tag{31}$$

where the "singular part" $S_\alpha(z)$ is obtained by collecting all the terms with index in $[-M \mathinner{.\,.} -1]$ (that is, forming $S_\alpha(z) = N_\alpha(z)/(z-\alpha)^M$ with $N_\alpha(z)$ a polynomial of degree less than M) and $H_\alpha(z)$ is analytic at α. Thus setting $S(z) := \sum_j S_{\alpha_j}(z)$, we observe that $f(z) - S(z)$ is analytic for $|z| \le R$. In other words, by collecting the singular parts of the expansions and subtracting them, we have "removed" the singularities of $f(z)$, whence the name of *method of subtracted singularities* sometimes given to the method [**229**, vol. 2, p. 448].

Taking coefficients, we get:

$$[z^n]f(z) = [z^n]S(z) + [z^n](f(z) - S(z)).$$

The coefficient of $[z^n]$ in the rational function $S(z)$ is obtained from Theorem 1. It suffices to prove that the coefficient of z^n in $f(z)-S(z)$, a function analytic for $|z| \le R$, is $\mathcal{O}(R^{-n})$. This fact follows from trivial bounds applied to Cauchy's integral formula with the contour of integration being $\lambda = \{z \,/\, |z| = R\}$, as in the proof of Theorem IV.7:

$$\left|[z^n](f(z)-S(z))\right| = \frac{1}{2\pi}\left|\int_{|z|=R}(f(z)-S(z))\,\frac{dz}{z^{n+1}}\right| \le \frac{1}{2\pi}\frac{\mathcal{O}(1)}{R^{n+1}}2\pi R.$$

(ii) *Contour integration.* There is another line of proof for Theorem IV.10 which we briefly sketch as it provides an insight which is useful for applications to other types of singularities treated in Chapter VI. It consists in using directly Cauchy's coefficient formula and "pushing" the contour of integration past singularities. In other words, one computes directly the integral

$$I_n = \frac{1}{2i\pi}\int_{|z|=R} f(z)\,\frac{dz}{z^{n+1}}$$

by residues. There is a pole at $z=0$ with residue f_n and poles at the α_j with residues corresponding to the terms in the expansion stated in Theorem IV.10; for instance, if $f(z) \sim c/(z-a)$ as $z \to a$, then

$$\operatorname{Res}(f(z)z^{-n-1}; z=a) = \operatorname{Res}(\frac{c}{(z-a)}z^{-n-1}; z=a) = \frac{c}{a^{n+1}}.$$

Finally, by the same trivial bounds as before, I_n is $\mathcal{O}(R^{-n})$. ☐

▷ **25.** *Effective error bounds.* The error term $O(R^n)$ in (29), call it ε_n, satisfies

$$|\varepsilon_n| \le \sup_{|z|=R} |f(z)|.$$

This results immediately from the second proof. This bound may be useful, even in the case of rational functions. ◁

EXAMPLE 6. *Surjections.* These are defined as sequences of sets ($\mathcal{R} = \mathrm{SEQ}(\mathrm{SET}_{\ge 1}(\mathcal{Z}))$) with EGF $R(z) = (2-e^z)^{-1}$ (see p. 98). We have already determined the poles, the one of smallest modulus being at $\log 2 \doteq 0.69314$. At this dominant pole, one finds $R(z) \sim -\frac{1}{2}(z-\log 2)^{-1}$. This implies an approximation for the number of surjections:

$$R_n \equiv n![z^n]R(z) \sim \xi(n), \qquad \text{with} \quad \xi(n) := \frac{n!}{2}\cdot\Big(\frac{1}{\log 2}\Big)^{n+1}.$$

Here is, for $n = 2, 4, \ldots, 32$, a table of the values of the surjection numbers (left) compared with the asymptotic approximation rounded[11] to the nearest integer, $\lceil \xi(n) \rfloor$: It is piquant to see that $\lceil \xi(n) \rfloor$ provides the exact value of R_n for all values of $n = 1, \ldots, 15$, and it starts losing one digit for $n = 17$, after which point a few "wrong" digits gradually appear, but in very limited number; see Figure 7. (A similar situation holds for tangent numbers discussed in our *Invitation*, p. 4.) The explanation of such a faithful asymptotic representation owes to the fact that the error terms provided by meromorphic asymptotics are exponentially small. In

[11]The notation $\lceil x \rfloor$ represents x rounded to the nearest integer: $\lceil x \rfloor := \lfloor x + \frac{1}{2} \rfloor$.

3	3
75	75
4683	4683
545835	545835
102247563	102247563
28091567595	28091567595
10641342970443	10641342970443
5315654681981355	5315654681981355
3385534663256845323	338553466325684532 ***6***
2677687796244384203115	2677687796244384203 ***088***
2574844419803190384544203	2574844419803190384544 ***450***
2958279121074145472650648875	2958279121074145472650064 ***6597***
4002225759844168492486127539083	4002225759844168492486127 5 ***55859***
6297562064950066033518373935334635	6297562064950066033518373935 ***416161***
11403568794011880483742464196184901963	114035687940118804837424641961 7 ***4527074***
2354515408573489664918449063714485547639 5	2354515408573489664918449063714 ***5314147690***

FIGURE 7. The surjection numbers pyramid: for $n = 2, 4, \dots, 32$, the exact values of the numbers R_n (left) compared to the approximation $\lceil \xi(n) \rfloor$ with discrepant digits in boldface (right).

effect, there is no other pole in $|z| \le 6$, the next ones being at $\log 2 \pm 2i\pi$ with modulus of about 6.32. Thus, for $r_n = [z^n]R(z)$, there holds

$$\frac{R_n}{n!} \sim \frac{1}{2} \cdot \Big(\frac{1}{\log 2}\Big)^{n+1} + \mathcal{O}(6^{-n}). \tag{32}$$

For the double surjection problem, $R^*(z) = (2 + z - e^z)$, we get similarly

$$[z^n]R^*(z) \sim \frac{1}{e^{\rho^*} - 1} (\rho^*)^{-n-1},$$

with $\rho^* = 1.14619$ the smallest positive root of $e^{\rho^*} - \rho^* = 2$. ..... END OF EXAMPLE 6. $\square$

It is worth reflecting on this example as it is representative of a production chain based on the two successive implications reflecting the spirit of Part A and Part B of the book:

$$\left\{\begin{array}{ccc} \mathcal{S} = \mathrm{SEQ}(\mathrm{SET}_{\ge 1}(\mathcal{Z})) & \Longrightarrow & S(z) = \dfrac{1}{2 - e^z} \\ S(z) \underset{z \to \log 2}{\sim} -\dfrac{1}{2} \dfrac{1}{(z - \log 2)} & \rightsquigarrow & \frac{1}{n!} S_n \sim \frac{1}{2} (\log 2)^{-n-1}. \end{array}\right.$$

There the first implication (written '$\Longrightarrow$' as usual) is provided automatically by the symbolic method. The second one (written here '$\rightsquigarrow$') is a formal translation from the expansion of the GF at its dominant singularity to the asymptotic form of coefficients, validity being granted by complex-analytic conditions.

EXAMPLE 7. *Alignments.* These are sequences of cycles ($\mathcal{O} = \mathrm{SEQ}(\mathrm{CYC}(\mathcal{Z}))$, p. 111) with EGF

$$O(z) = \frac{1}{1 - \log(1 - z)^{-1}}.$$

There is a singularity when $\log(1 - z)^{-1} = 1$, which is at $\rho = 1 - e^{-1}$ and arises before $z = 1$ where the logarithm becomes singular. Then, the computation of the asymptotic form of

$[z^n]O(z)$ only requires a local expansion near ρ,

$$O(z) \sim \frac{-e^{-1}}{z-1+e^{-1}} \qquad \Longrightarrow \qquad [z^n]O(z) \sim \frac{e^{-1}}{(1-e^{-1})^{n+1}},$$

and the coefficient estimates result from Theorem IV.10. . . . . END OF EXAMPLE 7. □

▷ **26.** *Some "supernecklaces".* One estimates

$$[z^n] \log\left(\frac{1}{1-\log\frac{1}{1-z}}\right) \sim \frac{1}{n}(1-e^{-1})^{-n},$$

where the EGF enumerates labelled cycles of cycles (supernecklaces, p. 116). [Hint: Take derivatives.] ◁

EXAMPLE 8. *Generalized derangements.* The probability that the shortest cycle in a random permutation of size n has length larger than k is

$$[z^n]D^{(k)}(z), \qquad \text{where} \quad D^{(k)}(z) = \frac{e^{-\frac{z}{1}-\frac{z^2}{2}-\cdots-\frac{z^k}{k}}}{1-z},$$

as results from the specification $\mathcal{D}^{(k)} = \text{SET}(\text{CYC}_{>k}(\mathcal{Z}))$. For any *fixed* k, one has (easily) $D^{(k)}(z) \sim e^{-H_k}/(1-z)$ as $z \to 1$, with 1 being a simple pole. Accordingly the coefficients $[z^n]D^{(k)}(z)$ tend to e^{-H_k} as $n \to \infty$. Thus, due to meromorphy, we have the characteristic implication

$$D^{(k)}(z) \sim \frac{e^{-H_k}}{1-z} \qquad \Longrightarrow \qquad [z^n]D^{(k)}(z) \sim e^{-H_k}.$$

Since there is no other singularity at a finite distance, the error in the approximation is (at least) exponentially small,

$$(33) \qquad [z^n]\frac{e^{-\frac{z}{1}-\frac{z^2}{2}-\cdots-\frac{z^k}{k}}}{1-z} = e^{-H_k} + O(R^{-n}),$$

for *any* $R > 1$. The cases $k = 1, 2$ in particular justify the estimates mentioned at the beginning of this chapter, on p. 216. . . . . END OF EXAMPLE 8. □

This example is also worth reflecting upon. In prohibiting cycles of length $< k$, we modify the EGF of all permutations, $(1-z)^{-1}$ by a factor $e^{-z/1-\cdots-z^k/k}$. The resulting EGF is meromorphic at 1; thus only the value of the modifying factor at $z = 1$ matters, so that this value, namely e^{H_k}, provides the asymptotic proportion of k-derangements. We shall encounter more and more shortcuts of this sort as we progress into the book.

▷ **27.** *Shortest cycles of permutations are not too long.* Let S_n be the random variable denoting the length of the shortest cycle in a random permutation of size n. Using the circle $|z| = 2$ to estimate the error in the approximation e^{-H_k} above, one finds that, for $k \leq \log n$,

$$\left|\mathbb{P}(S_n > k) - e^{-H_k}\right| \leq \frac{1}{2^n}e^{2^{k+1}},$$

which is exponentially small in this range of k-values. Thus, the approximation e^{-H_k} remains good when k is allowed to tend sufficiently slowly to ∞ with n. One can also explore the possibility of better bounds and larger regions of validity of the main approximation. (See Panario and Richmond's study [**338**] for a general theory of smallest components in sets.) ◁

$\triangleright$ **28.** *Expected length of the shortest cycle.* The classical approximation of the harmonic numbers, $H_k \approx \log k + \gamma$ suggests $e^{-\gamma}/k$ as a possible approximation to (33) for *both* large n and large k in suitable regions. In agreement with this heuristic argument, the expected length of the shortest cycle in a random permutation of size n is effectively asymptotic to

$$\sum_{k=1}^{n} \frac{e^{-\gamma}}{k} \sim e^{-\gamma} \log n,$$

a property first discovered by Shepp and Lloyd [**383**]. $\triangleleft$

The next example illustrates the analysis of a collection of rational generating functions (Smirnov words) paralleling nicely the enumeration of a special type of integer composition (Carlitz compositions) that resorts to meromorphic asymptotics.

EXAMPLE 9. *Smirnov words and Carlitz compositions.* Bernoulli trials have been discussed in Chapter III, in relation to weighted word models. Take the class $\mathcal{W}$ of all words over an r-ary alphabet, where letter j is assigned probability p_j and letters of words are drawn independently. With this weighting, the GF of all words is $W(z) = 1/(1 - \sum p_j z) = (1-z)^{-1}$. Consider the problem of determining the probability that a random word of length n is of Smirnov type, that is, all blocks of length 2 are formed with distinct letters. In order to avoid degeneracies, we impose $r \geq 3$ (since for $r = 2$, the only Smirnov words are `ababa`... and `babab`...).

By our discussion of Section III. 7 (p. 194), the GF of Smirnov words (again with the probabilistic weighting) is

$$S(z) = \frac{1}{1 - \sum \frac{p_j z}{1+p_j z}}.$$

By monotonicity of the denominator, this rational function has a unique dominant singularity at ρ such that

$$\sum_{j=1}^{r} \frac{p_j \rho}{1 + p_j \rho} = 1, \tag{34}$$

and $z = \rho$ is a simple pole. Consequently, ρ is a well-characterized algebraic number defined implicitly by an equation of degree r. There results that the probability for a word to be Smirnov is (not too surprisingly) exponentially small, with the precise formula being

$$[z^n]S(z) \sim C \cdot \rho^{-n}, \qquad C = \left(\sum_{j=1}^{r} \frac{p_j \rho}{(1+p_j\rho)^2}\right)^{-1}.$$

A similar analysis, using bivariate generating functions, shows that in a random word of length n conditioned to be Smirnov, the letter j appears with asymptotic frequency

$$q_j = \frac{1}{Q} \frac{p_j}{(1+p_j\rho)^2}, \qquad Q := \sum_{j=1}^{r} \frac{p_j}{(1+p_j\rho)^2}, \tag{35}$$

in the sense that the mean number of occurrences of letter j is asymptotic to $q_j n$. All these results are seen to be consistent with the equiprobable letter case $p_j = 1/r$, for which $\rho = r/(r-1)$.

Carlitz compositions illustrate a limit situation, in which the alphabet is infinite, while letters have different sizes. Recall that a Carlitz composition of the integer n is a composition of n such that no two adjacent summands have equal value. By Note III.30, p. 190, such compositions can be obtained by substitution from Smirnov words, to the effect that

$$K(z) = \left(1 - \sum_{j=1}^{\infty} \frac{z^j}{1+z^j}\right)^{-1}. \tag{36}$$

The asymptotic form of the coefficients then results from an analysis of dominant poles. The OGF has a simple pole at ρ, which is the smallest positive root of the equation

$$\sum_{j=1}^{\infty} \frac{\rho^j}{1+\rho^j} = 1. \tag{37}$$

(Note the analogy with (34) due to commonality of the combinatorial argument.) Thus:

$$C_n \sim C \cdot \beta^n, \qquad C \doteq 0.45636\,34740, \quad \beta \doteq 1.75024\,12917.$$

There, $\beta = 1/\rho$ with ρ as in (37). In a way analogous to Smirnov words, the asymptotic frequency of summand k appears to be proportional to $k\rho^k/(1+\rho^k)^2$; see [**258, 301**] for further properties. .... END OF EXAMPLE 9. □

IV. 6. Localization of singularities

There are situations where a function possesses several dominant singularities, that is, several singularities are present on the boundary of the disc of convergence. We examine here the induced effect on coefficients and discuss ways to localize such dominant singularities.

IV. 6.1. Multiple singularities. In the case when there exists more than one dominant singularity, several geometric terms of the form β^n sharing the same modulus (and each carrying its own subexponential factor) must be combined. In simpler situations, such terms globally induce a pure periodic behaviour for coefficients that is easy to describe. In the general case, irregular fluctuations of a somewhat arithmetic nature may prevail.

Pure periodicities. When several dominant singularities of $f(z)$ have the same modulus and are regularly spaced on the boundary of the disc of convergence, they may induce complete cancellations of the main exponential terms in the asymptotic expansion of the coefficient f_n. In that case, different regimes will be present in the coefficients f_n based on congruence properties of n. For instance, the functions

$$\frac{1}{1+z^2} = 1 - z^2 + z^4 - z^6 + z^8 - \cdots, \quad \frac{1}{1-z^3} = 1 + z^3 + z^6 + z^9 + \cdots,$$

exhibit patterns of periods 4 and 3 respectively, this corresponding to poles that are roots of unity or order 4 ($\pm i$), and 3 ($\omega : \omega^3 = 1$). Accordingly, the function

$$\phi(z) = \frac{1}{1+z^2} + \frac{1}{1-z^3} = \frac{2 - z^2 + z^3 + z^4 + z^8 + z^9 - z^{10}}{1 - z^{12}}$$

has coefficients that obey a pattern of period 12 (for example, the coefficients ϕ_n such that $n \equiv 1, 5, 6, 7, 11$ modulo 12 are zero). Accordingly, the coefficients of

$$[z^n]\psi(z) \qquad \text{where} \qquad \psi(z) = \phi(z) + \frac{1}{1 - z/2},$$

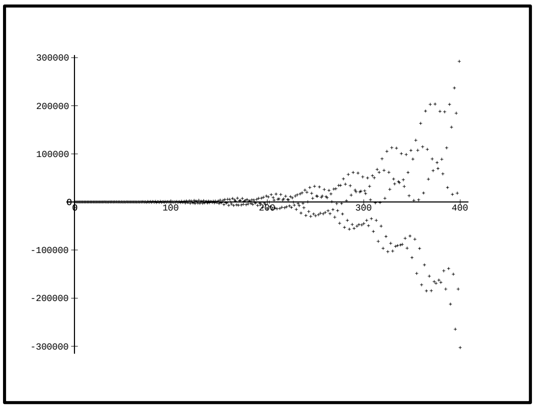


FIGURE 8. The coefficients $[z^n]f(z)$, where $f(z) = \left(1+1.02z^4\right)^{-3}\left(1-1.05z^5\right)^{-1}$ illustrate a periodic superposition of smooth behaviours that depend on the residue class of n modulo 20.

manifest a different exponential growth when n is congruent to $1, 5, 6, 7, 11 \bmod 12$. See Figure 8 for such a superposition of pure periodicities. In many combinatorial applications, generating functions involving periodicities can be decomposed at sight, and the corresponding asymptotic subproblems generated are then solved separately.

$\triangleright$ **29.** *Decidability of polynomial properties.* Given a polynomial $p(z) \in \mathbb{Q}[z]$, the following properties are decidable: (i) whether one of the zeros of p is a root of unity; (ii) whether one of the zeros of p has an argument that is commensurate with π. [One can use resultants. An algorithmic discussion of this and related issues is given in [**211**].] $\triangleleft$

Nonperiodic fluctuations. As a representative example, consider the polynomial $D(z) = 1 - \frac{6}{5}z + z^2$, whose roots are

$$\alpha = \frac{3}{5} + i\frac{4}{5}, \qquad \bar{\alpha} = \frac{3}{5} - i\frac{4}{5},$$

both of modulus 1 (the numbers $3, 4, 5$ form a Pythagorean triple), with argument $\pm\theta_0$ where $\theta_0 = \arctan(\frac{4}{3}) \doteq 0.92729$. The expansion of the function $f(z) = 1/D(z)$ starts as

$$\frac{1}{1-\frac{6}{5}z+z^2} = 1 + \frac{6}{5}z + \frac{11}{25}z^2 - \frac{84}{125}z^3 - \frac{779}{625}z^4 - \frac{2574}{3125}z^5 + \cdots,$$

the sign sequence being

$$+++---++++---+++----+++----+++---,$$

which indicates a somewhat irregular oscillating behaviour, where blocks of 3 or 4 pluses follow blocks of 3 or 4 minuses.

The exact form of the coefficients of f results from a partial fraction expansion:

$$f(z) = \frac{a}{1-z/\alpha} + \frac{b}{1-z/\bar{\alpha}} \qquad \text{with} \quad a = \frac{1}{2} + \frac{3}{8}i, \quad b = \frac{1}{2} - \frac{3}{8}i,$$

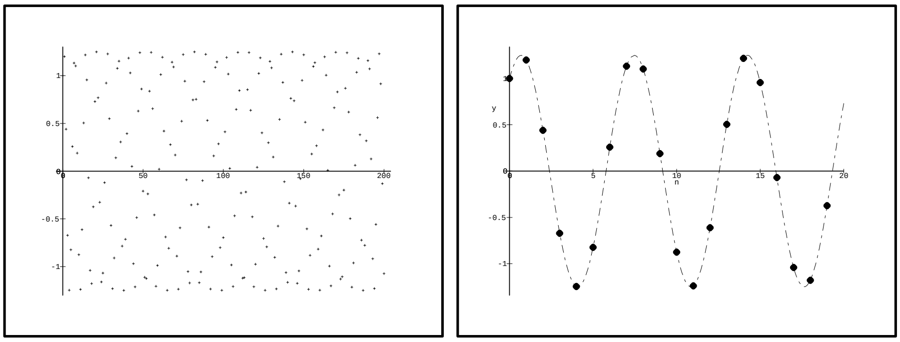


FIGURE 9. The coefficients of $f = 1/(1 - \frac{6}{5}z + z^2)$ exhibit an apparently chaotic behaviour (left) which in fact corresponds to a discrete sampling of a sine function (right), reflecting the presence of two conjugate complex poles.

where $\alpha = e^{i\theta_0}$, $\overline{\alpha} = e^{-i\theta_0}$ Accordingly,

$$f_n = ae^{-in\theta_0} + be^{in\theta_0} = \frac{\sin((n+1)\theta_0)}{\sin(\theta_0)}. \tag{38}$$

This explains the sign changes observed. Since the angle θ_0 is not commensurate with π, the coefficients fluctuate but, unlike in our earlier examples, no exact periodicity is present in the sign patterns. See Figure 9 for a rendering and Figure 13 of Chapter V (p. 321) for a meromorphic case linked to compositions into prime summands.

Complicated problems of an arithmetical nature may occur if several such singularities with non–commensurate arguments combine, and some open problem even remain in the analysis of linear recurring sequences. (For instance no decision procedure is known to determine whether such a sequence ever vanishes.) Fortunately, such problems occur infrequently in combinatorial applications, where dominant poles of rational functions (as well as many other functions) tend to have a simple geometry as we explain next.

▷ **30.** *Irregular fluctuations and Pythagorean triples.* The quantity $\frac{1}{\pi}\theta_0$ is an irrational number, so that the sign fluctuations of (38) are "irregular" (i.e., non purely periodic). [Proof: *a contrario*. Indeed, otherwise, $\alpha = (3+4i)/5$ would be a root of unity. But then the minimal polynomial of α would be a cyclotomic polynomial with nonintegral coefficients, a contradiction; see [**286**, VIII.3] for the latter property.] ◁

▷ **31.** *Skolem-Mahler-Lech Theorem.* Let f_n be the sequence of coefficients of a rational function, $f(z) = A(z)/B(z)$, where $A, B \in \mathbb{Q}[z]$. The set of all n such that $f_n = 0$ is the union of a finite (possibly empty) set and a finite number (possibly zero) of infinite arithmetic progressions. (The proof is based on p-adic analysis, but the argument is intrinsically nonconstructive; see [**324**] for an attractive introduction to the subject and references.) ◁

Periodicity conditions for positive generating functions. By the previous discussion, it is of interest to locate dominant singularities of combinatorial generating

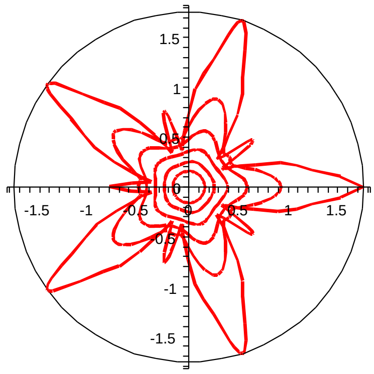


FIGURE 10. Illustration of the "Daffodil Lemma": the images of circles $z = Re^{i\theta}$ ($R = 0.4\,..\,0.8$) rendered by a polar plot of $|f(z)|$ in the case of $f(z) = z^7 e^{z^{25}} + z^2/(1-z^{10}))$, which has span 5.

functions, and, in particular, determine whether their arguments (the "dominant directions") are commensurate to 2π. In the latter case, different asymptotic regimes of the coefficients manifest themselves, depending on congruence properties of n.

First a few definitions. For a sequence (f_n) with GF $f(z)$, the *support* of f, denoted $\mathrm{Supp}(f)$, is the set of all n such that $f_n \neq 0$. The sequence (also its GF) is said to admit *span*, or *period*, d if for some r, there holds

$$\mathrm{Supp}(f) \subseteq r + d\mathbb{Z}_{\geq 0} \equiv \{r,\ r+d,\ r+2d, \ldots\}.$$

In that case, if f is analytic at 0, then there exists a function g analytic at 0 such that $f(z) = z^r g(z^d)$. The largest span, p, is often plainly referred to as *the* period, all other spans being divisors of p. With $E := \mathrm{Supp}(f)$, this maximal span is attainable as $p = \gcd(E - E)$ (pairwise differences) as well as $p = \gcd(E - \{r\})$ where $r := \min(E)$. For instance $\sin(z)$ has period 2, $\cos(z) + \cosh(z)$ has period 4, $z^3 e^{z^5}$ has period 5, and so on.

In the context of periodicities, a basic property is expressed by what we have chosen to name figuratively the "Daffodil Lemma". By virtue of this lemma, the span of a function f with nonnegative coefficients is related to the behaviour of $|f(z)|$ as z varies along circles centred at the origin (Figure 10).

LEMMA IV.1 ("Daffodil Lemma"). *Let $f(z)$ be analytic in $|z| < \rho$ and have nonnegative coefficients at 0. Assume that f does not reduce to a monomial and that for* some *nonzero nonpositive z satisfying $|z| < \rho$, one has*

$$|f(z)| = f(|z|).$$

Then, the following hold: (i) the argument of z must be commensurate to 2π, i.e., $z = Re^{i\theta}$ with $\theta/(2\pi) = \frac{r}{p} \in \mathbb{Q}$ (an irreducible fraction) and $0 < r < p$; (ii) f admits p as a span.

PROOF. This classical lemma is a simple consequence of the strong triangle inequality. Indeed, with $z = Re^{i\theta}$, the equality $|f(z)| = f(|z|)$ implies that the complex numbers

$f_n R^n e^{in\theta}$ for $n \in \mathrm{Supp}(f)$ all lie on the same ray (a half-line emanating from 0). This is impossible if θ is irrational, as soon as the expansion of f contains at least two monomials. $\square$

Berstel [**41**] first realized that rational generating functions arising from regular languages can only have dominant singularities of the form $\rho\omega^j$, where ω is a certain root of unity. This property in fact extends to many nonrecursive specifications, as shown by Flajolet, Salvy, and Zimmermann in [**173**].

PROPOSITION IV.3 (Commensurability of dominant directions). *Let $\mathcal{S}$ be a constructible labelled class that is nonrecursive, in the sense of Theorem IV.8. Assume that the EGF $S(z)$ has a finite radius of convergence ρ. Then there exists a computable integer $d \geq 1$ such that the set of dominant singularities of $S(z)$ is contained in the set $\{\rho\omega^j\}$, where $\omega^d = 1$.*

PROOF. (Sketch; see [**41, 173**]) By definition, a nonrecursive class $\mathcal{S}$ is obtained from **1** and $\mathcal{Z}$ by means of a finite number of union, product, sequence, set, and cycle constructions. We have seen earlier, in Section IV. 4, an inductive algorithm that determines radii of convergence. It is then easy to enrich that algorithm and determine simultaneously (by induction on the specification) the period of its GF and the set of dominant directions.

The period is determined by simple rules. For instance, if $\mathcal{S} = \mathcal{T} \star \mathcal{U}$ ($S = T \cdot U$) and T, U are infinite series with respective periods p, q, one has the implication

$$\mathrm{Supp}(T) \subseteq a + p\mathbb{Z}, \quad \mathrm{Supp}(U) \subseteq b + q\mathbb{Z} \qquad \Longrightarrow \qquad \mathrm{Supp}(S) \subseteq a + b + \xi\mathbb{Z},$$

with $\xi = \gcd(p, q)$. Similarly, for $\mathcal{S} = \mathrm{SEQ}(\mathcal{T})$,

$$\mathrm{Supp}(T) \subseteq a + p\mathbb{Z} \qquad \Longrightarrow \qquad \mathrm{Supp}(S) \subseteq \delta\mathbb{Z},$$

where now $\delta = \gcd(a, p)$.

Regarding dominant singularities, the case of a sequence construction is typical. It corresponds to $g(z) = (1 - f(z))^{-1}$. Assume that $f(z) = z^a h(z^p)$, with p the maximal period, and let $\rho > 0$ be such that $f(\rho) = 1$. The equations determining any dominant singularity ζ are $f(\zeta) = 1$, $\zeta = |\rho|$. In particular, the equations imply $|f(\zeta)| = f(|\zeta|)$, so that, by the Daffodil Lemma, the argument of ζ must be of the form $2\pi r/s$. An easy refinement of the argument shows that, for $\delta = \gcd(a, p)$, all the dominant directions coincide with the multiples of $2\pi/\delta$. The discussion of cycles is entirely similar since $\log(1 - f)^{-1}$ has the same dominant singularities as $(1 - f)^{-1}$. Finally, for exponentials, it suffices to observe that e^f does not modify the singularity pattern of f, since $\exp(z)$ is an entire function. $\square$

$\triangleright$ **32.** *Daffodil lemma and unlabelled classes.* Proposition IV.3 applies to any unlabelled class $\mathcal{S}$ that admits a nonrecursive specification, provided its radius of convergence ρ satisfies $\rho < 1$. (When $\rho = 1$, there is a possibility of having the unit circle as a natural boundary—a property that is otherwise decidable.) $\triangleleft$

Exact formulæ. The error terms appearing in the asymptotic expansion of coefficients of meromorphic functions are already exponentially small. By peeling off the singularities of a meromorphic function layer by layer, in order of increasing modulus, one is led to extremely precise—or even exact—expansions for the coefficients. Such exact representations are found for Bernoulli numbers B_n, surjection numbers R_n, as well as Secant numbers E_{2n} and Tangent numbers E_{2n+1}, defined by

$$\begin{array}{lcll}
\displaystyle\sum_{n=0}^{\infty} B_n \frac{z^n}{n!} & = & \displaystyle\frac{z}{e^z-1} & \text{(Bernoulli numbers)} \\
\displaystyle\sum_{n=0}^{\infty} R_n \frac{z^n}{n!} & = & \displaystyle\frac{1}{2-e^z} & \text{(Surjection numbers)} \\
\displaystyle\sum_{n=0}^{\infty} E_{2n} \frac{z^{2n}}{(2n)!} & = & \displaystyle\frac{1}{\cos(z)} & \text{(Secant numbers)} \\
\displaystyle\sum_{n=0}^{\infty} E_{2n+1} \frac{z^{2n+1}}{(2n+1)!} & = & \tan(z) & \text{(Tangent numbers).}
\end{array}$$

Bernoulli numbers. These numbers traditionally written B_n can be defined by their EGF $B(z) = z/(e^z - 1)$. The function $B(z)$ has poles at the points $\chi_k = 2ik\pi$, with $k \in \mathbb{Z} \setminus \{0\}$, and the residue at χ_k is equal to χ_k,

$$\frac{z}{e^z-1} \sim \frac{\chi_k}{z-\chi_k} \qquad (z \to \chi_k).$$

The expansion theorem for meromorphic functions is applicable here: start with the Cauchy integral formula, and proceed as in the proof of Theorem IV.10, using as external contours a large circle of radius R that passes half way between poles. As R tends to infinity, the integrand tends to 0 (as soon as $n \geq 2$) because the Cauchy kernel z^{-n-1} decreases as an inverse power of R while the EGF remains $O(R)$. In the limit, corresponding to an infinitely large contour, the coefficient integral becomes equal to the sum of all residues of the meromorphic function over the whole of the complex plane.

From this argument, we get the representation $B_n = -n! \sum_{k \in \mathbb{Z}\setminus\{0\}} \chi_k^{-n}$. This verifies that $B_n = 0$ if n is odd and $n \geq 3$. If n is even, then grouping terms two by two, we get the exact representation (which also serves as an asymptotic expansion):

$$\frac{B_{2n}}{(2n)!} = (-1)^{n-1} 2^{1-2n} \pi^{-2n} \sum_{k=0}^{\infty} \frac{1}{k^{2n}}. \tag{39}$$

Reverting the equality, we have also established that

$$\zeta(2n) = (-1)^{n-1} 2^{2n-1} \pi^{2n} \frac{B_{2n}}{(2n)!} \qquad \text{with} \quad \zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}, \quad B_n = n![z^n]\frac{z}{e^z-1},$$

a well-known identity that provides values of the Riemann zeta function $\zeta(s)$ at even integers as rational multiples of powers of π.

Surjection numbers. In the same vein, the surjection numbers have EGF $R(z) = (2 - e^z)^{-1}$ with simple poles at

$$\chi_k = \log 2 + 2ik\pi \qquad \text{where} \qquad R(z) \sim \frac{1}{2}\frac{1}{\chi_k - z}.$$

Since $R(z)$ stays bounded on circles passing half way in between poles, we find the exact formula, $R_n = \frac{1}{2}n!\sum_{k\in\mathbb{Z}} \chi_k^{-n-1}$. An equivalent real formulation is

$$\text{(40)}\quad \frac{R_n}{n!} = \frac{1}{2}\left(\frac{1}{\log 2}\right)^{n+1} + \sum_{k=1}^{\infty} \frac{\cos((n+1)\theta_k)}{(\log^2 2 + 4k^2\pi^2)^{(n+1)/2}}, \quad \theta_k := \arctan(\frac{2k\pi}{\log 2}),$$

which exhibits infinitely many harmonics of fast decaying amplitude.

▷ **33.** *Alternating permutations, tangent and secant numbers.* The relation (39) also provides a representation of the *tangent numbers* since $E_{2n-1} = (-1)^{n-1}B_{2n}4^n(4^n-1)/(2n)$. The secant numbers E_{2n} satisfy

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)^{2n+1}} = \frac{(\pi/2)^{2n+1}}{2\,(2n)!}E_{2n},$$

which can be read either as providing an asymptotic expansion of E_{2n} or as an evaluation of the sums on the left (the values of a Dirichlet L-function) in terms of π. The asymptotic number of alternating permutations (Chapter II) is consequently known to great accuracy. ◁

▷ **34.** *Solutions to the equation* $\tan(x) = x$. Let x_n be the nth positive root of the equation $\tan(x) = x$. For any integer $r \geq 1$, the sum $S(r) := \sum_n x_n^{-2r}$ is a computable rational number. [From folklore and *The American Mathematical Monthly*.] ◁

IV. 6.2. Localization of zeros and poles. We gather here a few results that often prove useful in determining the location of zeros of analytic functions, and hence of poles of meromorphic functions. A detailed treatment of this topic may be found in Henrici's book [**229**].

Let $f(z)$ be an analytic function in a region Ω and let γ be a simple closed curve interior to Ω, and on which f is assumed to have no zeros. We claim that the quantity

$$\text{(41)}\qquad N(f;\gamma) = \frac{1}{2i\pi}\int_\gamma \frac{f'(z)}{f(z)}\,dz$$

exactly equals the number of zeros of f inside γ counted with multiplicity. [Proof: the function f'/f has its poles exactly at the zeros of f, and the residue at each pole α equals the multiplicity of α as a root of f; the assertion then results from the residue theorem.]

Since a primitive function of f'/f is $\log f$, the integral also represents the variation of $\log f$ along γ, which is written $[\log f]_\gamma$. This variation itself reduces to $2i\pi$ times the variation of the argument of f along γ, since $\log(re^{i\theta}) = \log r + i\theta$ and the modulus r has variation equal to 0 along a closed contour ($[\log r]_\gamma = 0$). The quantity $[\theta]_\gamma$ is, by its definition, 2π multiplied by the number of times the transformed contour $f(\gamma)$ winds about the origin. This observation is known as the *Argument Principle*:

> **Argument Principle.** *The number of zeros of $f(z)$ (counted with multiplicities) inside γ equals the winding number of the transformed contour $f(\gamma)$ around the origin.*

By the same argument, if f is meromorphic in $\Omega \ni \gamma$, then $N(f;\gamma)$ equals the difference between the number of zeros and the number of poles of f inside γ, multiplicities being taken into account. Figure 11 exemplifies the use of the argument principle in localizing zeros of a polynomial.

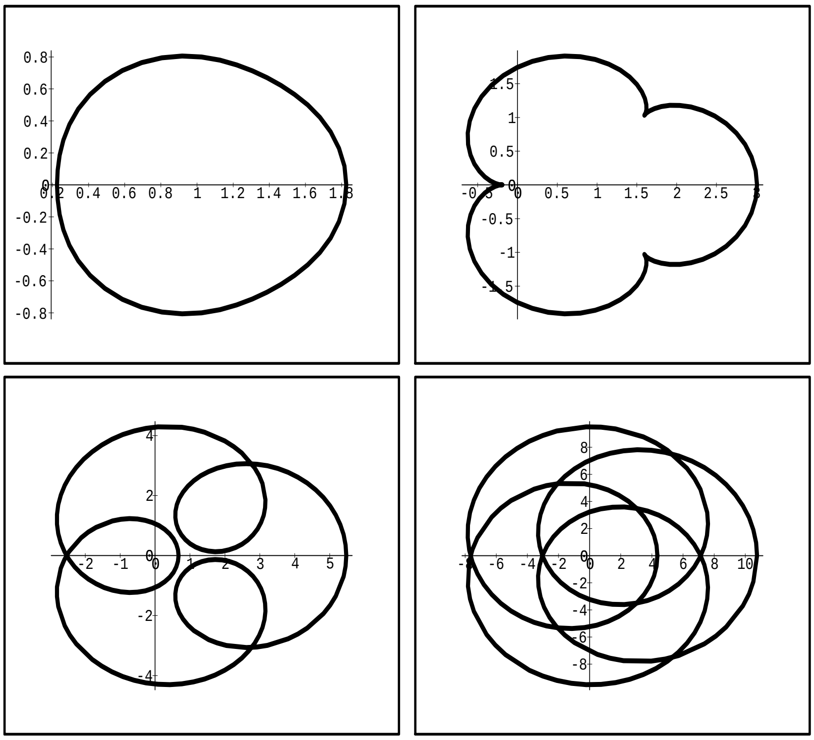


FIGURE 11. The transforms of $\gamma_j = \{|z| = \frac{4j}{10}\}$ by $P_4(z) = 1 - 2z + z^4$, for $j = 1, 2, 3, 4$, demonstrate that $P_4(z)$ has no zero inside $|z| < 0.4$, one zero inside $|z| < 0.8$, two zeros inside $|z| < 1.2$ and four zeros inside $|z| < 1.6$. The actual zeros are at $\rho_4 = 0.54368$, 1 and $1.11514 \pm 0.77184i$.

By similar devices, we get Rouché's theorem:

> **Rouché's theorem.** *Let the functions $f(z)$ and $g(z)$ be analytic in a region containing in its interior the closed simple curve γ. Assume that f and g satisfy* $|g(z)| < |f(z)|$ on *the curve γ. Then $f(z)$ and $f(z) + g(z)$ have the same number of zeros* inside *the interior domain delimited by γ.*

An intuitive way to visualize Rouché's Theorem is as follows: since $|g| < |f|$, then $f(\gamma)$ and $(f + g)(\gamma)$ must have the same winding number.

$\triangleright$ **35.** *Proof of Rouché's theorem.* Under the hypothesis of Rouché's theorem, for $0 \le t \le 1$ $h(z) = f(z) + tg(z)$ is such that $N(h; \gamma)$ is both an integer and an analytic, hence continuous, function of t in the given range. The conclusion of the theorem follows. $\triangleleft$

$\triangleright$ **36.** *The Fundamental Theorem of Algebra.* Every complex polynomial $p(z)$ of degree n has exactly n roots. A proof follows by Rouché's theorem from the fact that, for large enough $|z| = R$, the polynomial assumed to be monic is a "perturbation" of its leading term, z^n. $\triangleleft$

▷ **37.** *Symmetric function of the zeros.* Let $S_k(f;\gamma)$ be the sum of the kth powers of the roots of equation $f(z) = 0$ inside γ. One has

$$S_k(f;\gamma) = \frac{1}{2i\pi}\frac{f'(z)}{f(z)}z^k\,dz,$$

by a variant of the proof of the Argument Principle. ◁

These principles form the basis of numerical algorithms for locating zeros of analytic functions, in particular the ones closest to the origin, which are of most interest to us. One can start from an initially large domain and recursively subdivide it until roots have been isolated with enough precision—the number of roots in a subdomain being at each stage determined by numerical integration; see Figure 11 and refer for instance to [**99**] for a discussion. Such algorithms even acquire the status of full proofs if one operates with guaranteed precision routines (using, e.g., careful implementations of interval arithmetics).

IV. 6.3. Patterns in words: a case-study. Analysing the coefficients of a single generating function that is rational is a simple task, often even bordering on the trivial, granted the exponential-polynomial formula for coefficients (Theorem IV.9). However, in analytic combinatorics, we are often confronted with problems that involve an *infinite family* of functions. In that case, Rouché's Theorem and the Argument Principle provide decisive tools for localizing poles, while Theorems IV.3 (Residue Theorem) and IV.10 (Expansion of meromorphic functions) serve to determine effective error terms. An illustration of this situation is the analysis of patterns in words for which GFs have been derived in Chapters I (p. 50) and III (p. 200).

All patterns are not born equal. Surprisingly, in a random sequence of coin tossings, the pattern HTT is likely to occur much sooner (after 8 tosses on average) than the pattern HHH (needing 14 tosses on average); see the preliminary discussion in Example I.12 (p. 56). Questions of this sort are of obvious interest in the statistical analysis of genetic sequences [**295, 432**]. Say you discover that a sequence of length 100,000 on the four letters A, G, C, T contains the pattern TACTAC twice. Can this be assigned to chance or is this is likely to be a meaningful signal of some yet unknown structure? The difficulty here lies in quantifying precisely where the asymptotic regime starts, since, by Borges's Theorem (Note I.32, p. 58), sufficiently long texts will almost certainly contain any fixed pattern. The analysis of rational generating functions supplemented by Rouché's theorem provides definite answers to such questions, under Bernoulli models at least.

We consider here the class $\mathcal{W}$ of words over an alphabet $\mathcal{A}$ of cardinality $m \geq 2$. A pattern $\mathfrak{p}$ of some length k is given. As seen in Chapters I and III, its autocorrelation polynomial is central to enumeration. This polynomial is defined as $c(z) = \sum_{j=0}^{k-1} c_j z^j$, where c_j is 1 if $\mathfrak{p}$ coincides with its jth shifted version and 0 otherwise. We consider here the enumeration of words containing the pattern $\mathfrak{p}$ at least once, and dually of words excluding the pattern $\mathfrak{p}$. In other words, we look at problems such as: What is the probability that a random text of length n does (or does not) contain your name as a block of consecutive letters?

Length (k)	*Types*	$c(z)$	ρ
$k=3$	`aab, abb, bba, baa`	1	0.61803
	`aba, bab`	$1+z^2$	0.56984
	`aaa, bbb`	$1+z+z^2$	0.54368
$k=4$	`aaab, aabb, abbb, bbba, bbaa, baaa`	1	0.54368
	`aaba, abba, abaa, bbab, baab, babb`	$1+z^3$	0.53568
	`abab, baba`	$1+z^2$	0.53101
	`aaaa, bbbb`	$1+z+z^2+z^3$	0.51879

FIGURE 12. Patterns of length $3,4$: autocorrelation polynomial and dominant poles of $S(z)$.

The OGF of the class of words excluding $\mathfrak{p}$ is, we recall,

$$S(z) = \frac{c(z)}{z^k + (1-mz)c(z)}. \tag{42}$$

(Proposition I.4, p. 57), and we shall start with the case $m=2$ of a binary alphabet. The function $S(z)$ is simply a rational function, but the location and nature of its poles is yet unknown. We only know *a priori* that it should have a pole in the positive interval somewhere between $\frac{1}{2}$ and 1 (by Pringsheim's Theorem and since its coefficients are in the interval $[1, 2^n]$, for n large enough). Figure 12 gives a small list, for patterns of length $k=3,4$, of the pole ρ of $S(z)$ that is nearest to the origin. Inspection of the figure suggests ρ to be close to $\frac{1}{2}$ as soon as the pattern is long enough. We are going to prove this fact, based on Rouché's Theorem applied to the denominator of (42).

As regards termwise domination of coefficients, the autocorrelation polynomial lies between 1 (for less correlated patterns like `aaa...b`) and $1+z+\cdots+z^{k-1}$ (for the special case `aaa...a`). We set aside the special case of $\mathfrak{p}$ having only equal letters, i.e., a "maximal" autocorrelation polynomial—this case is discussed at length in the next chapter. Thus, in this scenario, the autocorrelation polynomial starts as $1+z^\ell+\cdots$ for some $\ell \geq 2$. Fix the number $A=0.6$. On $|z|=A$, we have

$$|c(z)| \geq \left|1-(A^2+A^3+\cdots)\right| = \left|1-\frac{A^2}{1-A}\right| = \frac{1}{10}. \tag{43}$$

In addition, the quantity $(1-2z)$ ranges over the circle of diameter $[-0.2, 1.2]$ as z varies along $|z|=A$, so that $|1-2z| \geq 0.2$. All in all, we have found that, for $|z|=A$,

$$|(1-2z)c(z)| \geq 0.02.$$

On the other hand, for $k>7$, we have $|z^k| < 0.017$ on the circle $|z|=A$. Then, amongst the two terms composing the denominator of (42), the first is strictly dominated by the second along $|z|=A$. By virtue of Rouché's Theorem, the number of roots of the denominator inside $|z| \leq A$ is then same as the number of roots of $(1-2z)c(z)$. The latter number is 1 (due to the root $\frac{1}{2}$) since $c(z)$ cannot be 0 by the argument of (43). Figure 13 exemplifies the extremely well-behaved characters of the complex zeros.

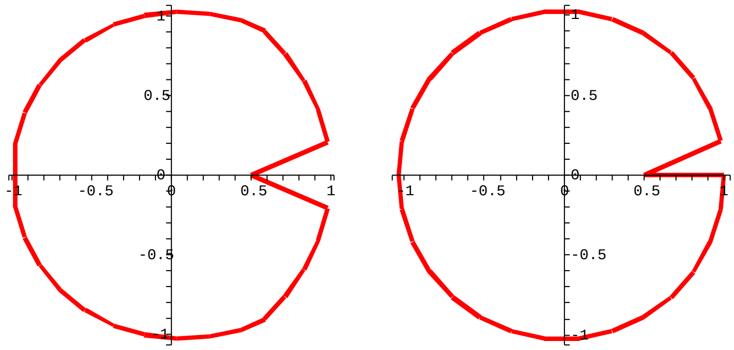


FIGURE 13. Complex zeros of $z^{31}+(1-2z)c(z)$ represented as joined by a polygonal line: (left) correlated pattern $a(ba)^{15}$; (right) uncorrelated pattern $a(ab)^{15}$.

In summary, we have found that for all patterns with at least two different letters ($\ell \geq 2$) and length $k \geq 8$, the denominator has a unique root in $|z| \leq A = 0.6$. The property for lengths k satisfying $4 \leq k \leq 7$ is then easily verified directly. The case $\ell = 1$ where we are dealing with long runs of identical letters can be subjected to an entirely similar argument (see also Example V.2, p. 287, for details). Therefore, unicity of a simple pole ρ of $S(z)$ in the interval $(0.5, 0.6)$ is granted.

It is then a simple matter to determine the local expansion of $s(z)$ near $z = \rho$,

$$S(z) \underset{z\to\rho}{\sim} \frac{\widetilde{\Lambda}}{\rho - z}, \qquad \widetilde{\Lambda} := \frac{c(\rho)}{2c(\rho) - (1-2\rho)c'(\rho) - k\rho^{k-1}},$$

from which a precise estimate for coefficients derives by Theorems IV.9 and IV.10.

The computation finally extends almost verbatim to nonbinary alphabets, with ρ being now close to $\frac{1}{m}$. It suffices to use the disc of radius $A = 1.2/m$. The Rouché part of the argument grants us unicity of the dominant pole in the interval $(1/m, A)$ for $k \geq 5$ when $m = 3$, and for $k \geq 4$ and any $m \geq 4$. (The remaining cases are easily checked individually.)

PROPOSITION IV.4. *Consider an m-ary alphabet. Let $\mathfrak{p}$ be a fixed pattern of length $k \geq 4$, with autocorrelation polynomial $c(z)$. Then the probability that a random word of length n does not contain $\mathfrak{p}$ as a pattern (a block of consecutive letters) satisfies*

$$\mathbb{P}_{\mathcal{W}_n}(\mathfrak{p} \text{ does not occur}) = \Lambda_{\mathfrak{p}}(m\rho)^{-n-1} + O\left((\frac{5}{6})^n\right), \tag{44}$$

where $\rho \equiv \rho_{\mathfrak{p}}$ is the unique root in $(\frac{1}{m}, \frac{6}{5m})$ of the equation $z^k + (1-mz)c(z) = 0$ and $\Lambda_{\mathfrak{p}} := mc(\rho)/(mc(\rho) - c'(\rho)(1-m\rho) - k\rho^{k-1})$.

Despite their austere appearance, these formulæ have indeed an a fairly concrete content. First, the equation satisfied by ρ can be put under the form $mz = 1 + z^k/c(z)$, and, since ρ is close to $\frac{1}{m}$, we may expect the approximation (remember the use of

"≈" as meaning "numerically approximately equal")

$$m\rho \approx 1 + \frac{1}{\gamma m^k},$$

where $\gamma := c(m^{-1})$ satisfies $1 \le \gamma < m/(m-1)$. By similar principles, the probabilities in (44) should be approximately

$$\mathbb{P}_{\mathcal{W}_n}(\mathfrak{p} \text{ does not occur}) \approx \left(1 + \frac{1}{\gamma m^k}\right)^{-n} \approx e^{-n/(\gamma m^k)}.$$

For a binary alphabet, this tells us that the occurrence of a pattern of length k starts becoming likely when n is of the order of 2^k, that is, when k is of the order of $\log_2 n$. The more precise moment when this happens must depend (via γ) on the autocorrelation of the pattern, with strongly correlated patterns having a tendency to occur a little late. (This vastly generalizes our empirical observations of Chapter I.) However, the mean number of occurrences of a pattern in a text of length n does not depend on the shape of the pattern. The apparent paradox is easily resolved: correlated patterns tend to occur late, while being prone to appear in clusters. For instance, the "late" pattern aaa, when it occurs, still has probability $\frac{1}{2}$ to occur at the next position as well and cash in another occurrence; in contrast no such possibility is available to the "early" uncorrelated pattern aab, whose occurrences must be somewhat spread out.

Such analyses are important as they can be used to develop a precise understanding of the behaviour of data compression algorithms (the Lempel–Ziv scheme); see Julien Fayolle's contribution [**132**] for details.

▷ **38.** *Multiple pattern occurrences.* A similar analysis applies to the generating function $S^{\langle s\rangle}(z)$ of words containing a fixed number s of occurrences of a pattern $\mathfrak{p}$. The OGF is obtained by expanding (with respect to u) the BGF $W(z,u)$ obtained in Chapter III by means of an inclusion-exclusion argument. For $s \ge 1$, one finds

$$S^{\langle s\rangle}(z) = z^k \frac{N(z)^{s-1}}{D(z)^{s+1}}, \qquad D(z) = z^k + (1-mz)c(z), \quad N(z) = z^k + (1-mz)(c(z)-1)),$$

which now has a pole of multiplicity $s+1$ at $z=\rho$. ◁

▷ **39.** *Patterns in Bernoulli sequences—asymptotics.* Similar results hold when letters are assigned nonuniform probabilities, $p_j = \mathbb{P}(a_j)$, for $a_j \in \mathcal{A}$. The weighted autocorrelation polynomial is then defined by protrusions, as in Note III.36 (p. 202). Multiple pattern occurrences can be also analysed. ◁

IV. 7. Singularities and functional equations

In the various combinatorial examples discussed so far in this chapter, we have been dealing with functions that are given by explicit expressions. Such situations essentially cover nonrecursive structures as well as the very simplest recursive ones, like Catalan or Motzkin trees, whose generating functions are expressible in terms of radicals. In fact, as will shall see extensively in this book, complex analytic methods are instrumental in analysing coefficients of functions *implicitly* specified by functional equations. In other words: *the nature of a functional equation can often provide information regarding the singularities of its solution*. Chapter V will illustrate this philosophy in the case of rational functions defined by systems of positive equations;

a very large number of examples will then be given in Chapters VI and VII, where singularities much more general than poles are treated.

In this section, we discuss three representative functional equations,

$$f(z) = ze^{f(z)}, \qquad f(z) = z + f(z^2 + z^3), \qquad f(z) = \frac{1}{1 - zf(z^2)}.$$

that illustrate the use of fundamental inversion or iteration properties to locate dominant singularities and derive exponential growth estimates for coefficients.

IV. 7.1. Inverse functions. We start with a generic problem: given a function ψ analytic at a point y_0 with $z_0 = \psi(y_0)$ what can be said about its inverse, namely the solution(s) to the equation $\psi(y) = z$ when z is near z_0 and y near y_0?

Let us examine what happens when $\psi'(y_0) \neq 0$, first without paying attention to analytic rigour. One has locally ('$\approx$' means as usual 'approximately equal')

$$\psi(y) \approx \psi(y_0) + \psi'(y_0)(y - y_0), \tag{45}$$

so that the equation $\psi(y) = z$ should admit, for z near z_0, a solution satisfying

$$y \approx y_0 + \frac{1}{\psi'(y_0)}(z - z_0). \tag{46}$$

If this is granted, the solution being locally linear, it is differentiable, hence analytic. The Analytic Inversion Lemma[12] provides a firm foundation for this calculation.

LEMMA IV.2 (Analytic Inversion). *Let $\psi(z)$ be analytic at y_0, with $\psi(y_0) = z_0$. Assume that $\psi'(y_0) \neq 0$. Then, for z in some small neighbourhood Ω_0 of z_0, there exists an analytic function $y(z)$ that solves the equation $\psi(y) = z$ and is such that $y(z_0) = y_0$.*

PROOF. (Sketch) The proof involves ideas analogous to those used to establish Rouché's Theorem and the Argument Principle (see especially the argument justifying Equation (41), p. 256) As a preliminary step, define the integrals ($j \in \mathbb{Z}_{\geq 0}$)

$$\sigma_j(z) := \frac{1}{2i\pi} \int_\gamma \frac{\psi'(y)}{\psi(y) - z}\, y^j\, dy, \tag{47}$$

where γ is a small enough circle centred at y_0 in the y-plane.

First consider σ_0. This function satisfies $\sigma_0(z_0) = 1$ [by the Residue Theorem] and is a continuous function of z whose value can only be an integer, this value being the number of roots of the equation $\psi(y) = z$. Thus, for z close enough to z_0, one must have $\sigma_0(z) \equiv 1$. In other words, the equation $\psi(y) = z$ has exactly one solution, the function ψ is locally invertible and a solution $y = y(z)$ that satisfies $y(z_0) = y_0$ is well-defined.

Next examine σ_1. By the Residue Theorem once more, the integral defining $\sigma_1(z)$ is the sum of the roots of the equation $\psi(y) = z$ that lie inside γ, that is, in our case, the value of $y(z)$ itself. (This is also a particular case of Note 37.) Thus, one has $\sigma_1(z) \equiv y(z)$. Since the integral defining $\sigma_1(z)$ depends analytically on z for z close enough to z_0, analyticity of $y(z)$ results. $\square$

[12] A more general statement and several proof techniques are also discussed in APPENDIX B: *Implicit Function Theorem*, p. 665.

$\triangleright$ **40.** *Details.* Let ψ be analytic in an open disc D centred at y_0. Then, there exists a small circle γ centred at y_0 and contained in D such that $\psi(y) \neq y_0$ on γ. [Zeros of analytic functions are isolated, a fact that results from the definition of an analytic expansion]. The integrals $\sigma_j(z)$ are thus well defined for z restricted to be close enough to z_0, which ensures that there exists a $\delta > 0$ such that $|\psi(y) - z| > \delta$ for all $y \in \gamma$. One can then expand the integrand as a power series in $(z - z_0)$, integrate the expansion termwise, and form in this way the analytic expansions of σ_0, σ_1 at z_0. [This line of proof follows [**232**, I, §9.4].] $\triangleleft$

$\triangleright$ **41.** *Inversion and majorant series.* The process corresponding to (45) and (46) can be transformed into a sound proof: first derive a formal power series solution, then verify that the formal solution is locally convergent using the method of majorant series (p. 237). $\triangleleft$

The Analytic Inversion Lemma states the following: *An analytic function locally admits an analytic inverse near any point where its first derivative is nonzero.* However, as we see next, a function cannot be analytically inverted in a neighbourhood of a point where its first derivative vanishes.

Consider now a function $\psi(y)$ such that $\psi'(y_0) = 0$ but $\psi''(y_0) \neq 0$, then, by the Taylor expansion of ψ, one expects

$$\psi(y) \approx \psi(y_0) + \frac{1}{2}(y - y_0)^2 \psi''(y_0). \tag{48}$$

Solving formally for y now indicates a *locally quadratic* dependency

$$(y - y_0)^2 \approx \frac{2}{\psi''(y_0)}(z - z_0),$$

and the inversion problem admits *two* solutions satisfying

$$y \approx y_0 \pm \sqrt{\frac{2}{\psi''(y_0)}}\sqrt{z - z_0}. \tag{49}$$

What this informal argument suggests is that the solutions have a singularity at z_0, and, in order for them to be suitably specified, one must somehow restrict their domain of definition: the case of $\sqrt{z}$ (the root(s) of $y^2 - z = 0$) discussed on p. 217 is typical.

Given some point z_0 and a neighbourhood Ω, the *slit neighbourhood* along direction θ, is the set

$$\Omega^{\setminus\theta} := \left\{ z \in \Omega \;\middle|\; \arg(z - z_0) \not\equiv \theta \bmod 2\pi \right\}.$$

We state:

LEMMA IV.3 (Singular Inversion). *Let $\psi(y)$ be analytic at y_0, with $\psi(y_0) = z_0$. Assume that $\psi'(y_0) = 0$ and $\psi''(y_0) \neq 0$. There exists a small neighbourhood Ω_0 such that the following holds: for any direction θ, there exist two functions, $y_1(z)$ and $y_2(z)$ defined on $\Omega_0^{\setminus\theta}$ that satisfy $\psi(y(z)) = z$; each is analytic in $\Omega_0^{\setminus\theta}$, has a singularity at the point z_0, and satisfies $\lim_{z\to z_0} y(z) = y_0$.*

PROOF. (Sketch) Define the functions $\sigma_j(z)$ as in the proof of the previous lemma, Equation (47). One now has $\sigma_0(z) = 2$, that is, the equation $\psi(y) = z$ possesses *two* roots near y_0, when z is near z_0. In other words ψ effects a double covering of a small neighbourhood Ω of y_0 onto the image neighbourhood $\Omega_0 = \psi(\Omega) \ni z_0$. By possibly restricting Ω, we may furthermore assume that $\psi'(y)$ only vanishes at y_0 in Ω (zeros of analytic functions are isolated) and that Ω is simply connected.

Fix any direction θ and consider the slit neighbourhood $\Omega_0^{\setminus\theta}$. Fix a point ζ in this slit domain; it has two preimages, $\eta_1, \eta_2 \in \Omega$. Pick up the one named η_1. Since $\psi'(\eta_1)$ is nonzero, the Analytic Inversion lemma applies: there is a local analytic inverse $y_1(z)$ of ψ. This $y_1(z)$ can then be uniquely continued[13] to the whole of $\Omega_0^{\setminus\theta}$, and similarly for $y_2(z)$. We have thus obtained two *distinct* analytic inverses.

Assume *a contrario* that $y_1(z)$ can be analytically continued at z_0. It would then admit a local expansion

$$y_1(z) = \sum_{n\geq 0} c_n (z - z_0)^n,$$

while satisfying $\psi(y_1(z)) = z$. But then, composing the expansions of ψ and y would entail

$$\psi(y_1(z)) = z_0 + O\left((z - z_0)^2\right) \qquad (z \to z_0),$$

which cannot coincide with the identity function (z). A contradiction has been reached. The point z_0 is thus a singular point for y_1 (as well as for y_2). □

▷ **42.** *Singular inversion and majorant series.* In a way that parallels Note 41, the process summarized by Equations (48) and (49) can be justified by the method of majorant series, which leads to an alternative proof of the Singular Inversion Lemma. ◁

▷ **43.** *Higher order branch points.* If all derivatives of ψ till order $r-1$ inclusive vanish at y_0, there are r inverses, $y_1(z), \ldots, y_r(z)$, defined over a slit neighbourhood of z_0. ◁

Tree enumeration. We can now consider the problem of obtaining information on the coefficients of a function $y(z)$ defined by an implicit equation

$$y(z) = z\phi(y(z)), \tag{50}$$

when $\phi(u)$ is analytic at $u = 0$. In order for the problem to be well-posed (algebraically, in terms of formal power series, as well as analytically, near the origin), we assume that $\phi(0) \neq 0$. Equation (50) may then be rephrased as

$$\psi(y(z)) = z \qquad \text{where} \qquad \psi(u) = \frac{u}{\phi(u)}, \tag{51}$$

so that it is in fact an instance of the inversion problem for analytic functions.

Equation (50) occurs in the counting of various types of trees, as seen in Subsections I. 5.1 (p. 61), II. 5.1 (p. 117), and III. 6.2 (p. 182). A typical case is $\phi(u) = e^u$, which corresponds to labelled nonplane trees, known as Cayley trees. The function $\phi(u) = (1+u)^2$ is associated to unlabelled plane binary trees and $\phi(u) = 1 + u + u^2$ to unary–binary trees (Motzkin trees). A full analysis was developed by Meir and Moon [**312**], themselves elaborating on earlier ideas of Pólya [**347, 349**] and Otter [**335**]. In all these cases, the exponential growth rate of the number of trees can be automatically determined.

[13] The fact of slitting Ω_0 makes the resulting domain simply connected, so that analytic continuation becomes uniquely defined. In contrast, the punctured domain $\Omega_0 \setminus \{z_0\}$ is *not* simply connected, so that the argument cannot be applied to it. As a matter of fact, $y_1(z)$ gets continued to $y_2(z)$, when the ray of angle θ is crossed: the point z_0 where two determinations meet is a *branch point*.

PROPOSITION IV.5. *Let ϕ be a function analytic at 0, having nonnegative Taylor coefficients, and such that $\phi(0) \neq 0$. Let $R \leq +\infty$ be the radius of convergence of the series representing ϕ at 0. Under the condition,*

$$\lim_{x \to R^-} \frac{x\phi'(x)}{\phi(x)} > 1, \tag{52}$$

there exists a unique solution $\tau \in (0, R)$ of the characteristic equation,

$$\frac{\tau\phi'(\tau)}{\phi(\tau)} = 1. \tag{53}$$

Then, the formal solution $y(z)$ of the equation $y(z) = z\phi(y(z))$ is analytic at 0 and its coefficients satisfy the exponential growth formula:

$$[z^n]\, y(z) \bowtie \left(\frac{1}{\rho}\right)^n \qquad \textit{where} \quad \rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}.$$

Note that condition (52) is automatically realized as soon as $\phi(R^-) = +\infty$, which covers our earlier examples as well as all the cases where ϕ is an entire function (e.g., a polynomial). Figure 14 displays graphs of functions on the real line associated to a typical inversion problem, that of Cayley trees, where $\phi(u) = e^u$.

PROOF. By Note 44 below, the function $x\phi'(x)/\phi(x)$ is an increasing function of x for $x \in (0, R)$. Condition (52) thus guarantees the existence and unicity of a solution of the characteristic equation. (Alternatively, rewrite the characteristic equation as $\phi_0 = \phi_2\tau^2 + 2\phi_3\tau^3 + \cdots$, where the right side is clearly an increasing function.)

Next, we observe that the equation $y = z\phi(y)$ admits a unique formal power series solution, which furthermore has nonnegative coefficients. (This solution can for instance be built by the method of indeterminate coefficients.) The Analytic Inversion Lemma (Lemma IV.2) then implies that this formal solution represents a function, $y(z)$, that is analytic at 0, where it satisfies $y(0) = 0$.

Now comes the hunt for singularities and, by Pringsheim's Theorem, one may restrict attention to the positive real axis. Let $r \leq +\infty$ be the radius of convergence of $y(z)$ at 0 and set $y(r) := \lim_{x \to r^-} y(x)$, which is well defined (though possibly infinite), given positivity of coefficients. Our goal is to prove that $y(r) = \tau$.

— Assume *a contrario* that $y(r) < \tau$. One would then have $\psi'(y(r)) \neq 0$. By the Analytic Inversion Lemma, $y(z)$ would be analytic at r, a contradiction.
— Assume *a contrario* that $y(r) > \tau$. There would then exist $r^* \in (0, r)$ such that $\psi'(y(r^*)) = 0$. But then y would be singular at r^*, by the Singular Inversion Lemma, also a contradiction.

Thus, one has $y(r) = \tau$, which is finite. Finally, since y and ψ are inverse functions, one must have

$$r = \psi(\tau) = \tau/\phi(\tau) = \rho,$$

by continuity as $x \to r^-$, which completes the proof. □

Proposition IV.5 thus yields an *algorithm* that produces the exponential growth rate associated to tree functions. This rate is itself invariably a computable number as

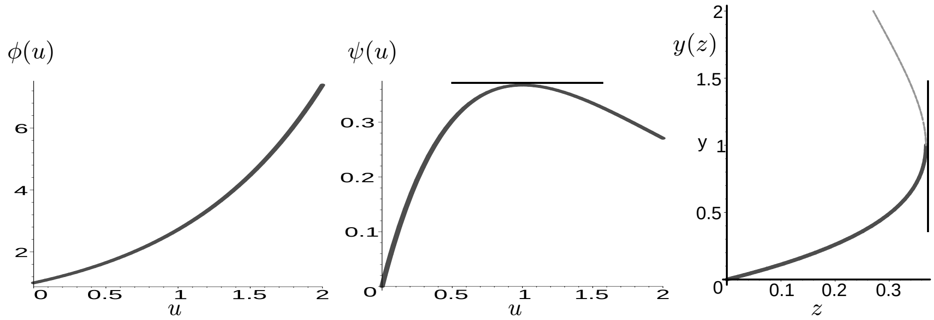


FIGURE 14. Singularities of inverse functions: $\phi(u) = e^u$ (left); $\psi(u) = u/\phi(u)$ (middle); $y = \mathrm{Inv}(\psi)$ (right).

Type	$\phi(u)$	(R)	τ	ρ	$y_n \bowtie \rho^{-n}$
binary tree	$(1+u)^2$	(∞)	1	$\frac{1}{4}$	$y_n \bowtie 4^n$
Motzkin tree	$1+u+u^2$	(∞)	1	$\frac{1}{3}$	$y_n \bowtie 3^n$
gen. Catalan tree	$\dfrac{1}{1-u}$	(1)	$\frac{1}{2}$	$\frac{1}{4}$	$y_n \bowtie 4^n$
Cayley tree	e^u	(∞)	1	e^{-1}	$y_n \bowtie e^n$

FIGURE 15. Exponential growth for classical tree families.

soon as ϕ is computable (i.e., its sequence of coefficients is computable). This computability result complements Theorem IV.8 which is relative to nonrecursive structures only.

As an example of application of Proposition IV.5, general Catalan trees correspond to $\phi(y) = (1-y)^{-1}$, whose radius of convergence is $R = 1$. The characteristic equation is $\tau/(1-\tau) = 1$, which implies $\tau = \frac{1}{2}$ and $\rho = \frac{1}{4}$. We obtain (not a suprise!) $y_n \bowtie 4^n$, a weak asymptotic formula for the Catalan numbers. Similarly, for Cayley trees, $\phi(u) = e^u$ and $R = +\infty$. The characteristic equation reduces to $(\tau-1)e^\tau = 0$, so that $\tau = 1$ and $\rho = e^{-1}$, giving a weak form of Stirling's formula: $[z^n]y(z) = \frac{n^{n-1}}{n!} \bowtie e^n$. Figure 15 summarizes the application of the method to a few already encountered tree families.

As our previous discussion suggests, the dominant singularity of tree generating functions is, under mild conditions, of the square-root type. Such a singular behaviour can then be analysed by the methods of Chapter VI: the coefficients admit an asymptotic form

$$[z^n]\, y(z) \sim C \cdot \rho^{-n} n^{-3/2},$$

with a subexponential factor of the form $n^{-3/2}$; see Section VI. 7, p. 385.

$\triangleright$ **44.** *Convexity of GFs and the Variance Lemma.* Let $\phi(z)$ be a nonconstant analytic function with nonnegative coefficients and a nonzero radius of convergence R, such that $\phi(0) \neq 0$. For $x \in (0, R)$ a parameter, define the *Boltzmann random variable* Ξ (of parameter x) by the property

$$\mathbb{P}(\Xi = n) = \frac{\phi_n x^n}{\phi(x)}, \qquad \text{with} \quad \mathbb{E}(s^\Xi) = \frac{\phi(sx)}{\phi(x)} \tag{54}$$

the probability generating function of Ξ. By differentiation, the first two moments of Ξ are

$$\mathbb{E}(\Xi) = \frac{x\phi'(x)}{\phi(x)}, \qquad \mathbb{E}(\Xi^2) = \frac{x^2\phi''(x)}{\phi(x)} + \frac{x\phi'(x)}{\phi(x)}.$$

There results, for any nonconstant GF ϕ, the general convexity inequality valid for $0 < x < R$:

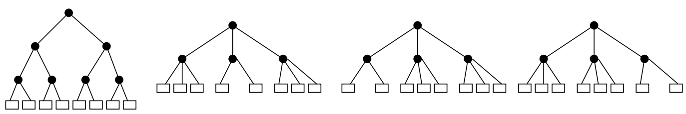
$$\frac{d}{dx}\left(\frac{x\phi'(x)}{\phi(x)}\right) > 0,$$

due to the fact that the variance of a nondegenerate random variable is always positive. Equivalently, the function $\log(\phi(e^t))$ is convex for $t \in (-\infty, \log R)$. (In statistical physics, a Boltzmann model (of parameter x) corresponds to a class Φ (with OGF ϕ) from which elements are drawn according to the size distribution (54).) $\lhd$

$\rhd$ **45.** *A variant form of the inversion problem.* Consider the equation $y = z + \phi(y)$, where ϕ is assumed to have nonegative coefficients and be entire, with $\phi(u) = O(u^2)$ at $u = 0$. This corresponds to a simple variety of trees in which trees are counted by the number of their leaves only. For instance, we have already encountered labelled hierarchies (phylogenetic trees in Section II. 5, p. 120) corresponding to $\phi(u) = e^u - 1 - u$, which gives rise to one of "Schröder's problems". Let τ be the root of $\phi'(\tau) = 1$ and set $\rho = \tau - \phi(\tau)$. Then $[z^n]y(z) \bowtie \rho^{-n}$. For the EGF L of labelled hierarchies ($L = z + e^L - 1 - L$), this gives $L_n/n! \bowtie (2\log 2 - 1)^{-n}$. (Observe that Lagrange inversion also provides $[z^n]y(z) = \frac{1}{n}[w^{n-1}](1 - y^{-1}\phi(y))^{-n}$.) $\lhd$

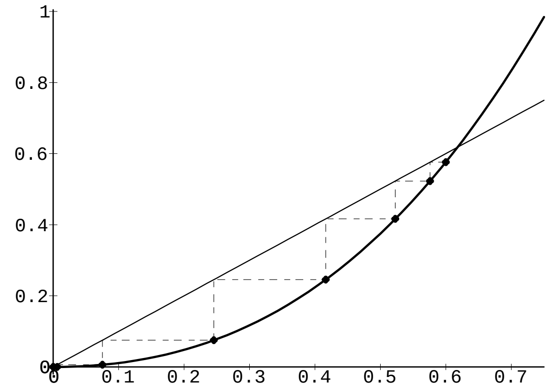
IV. 7.2. Iteration. The study of iteration of analytic functions was launched by Fatou and Julia in the first half of the twentieth century. Our reader is certainly aware of the beautiful images associated with the name of Mandelbrot whose works have triggered renewed interest in these questions now classified as resorting to the field of "complex dynamics" [**24, 103, 317, 340**]. In particular, the sets that appear in this context are often of a fractal nature. Mathematical objects of this sort are occasionally encountered in analytic combinatorics. We present here the first steps of a classic analysis of balanced trees published by Odlyzko [**328**] in 1982.

Consider the class $\mathcal{E}$ of *balanced 2–3 trees* defined as trees whose node degrees are restricted to the set $\{0, 2, 3\}$, with the additional property that all leaves are at the same distance from the root (Note 53, p. 83). We adopt as notion of size the number of leaves (also called external nodes), the list of all 4 trees of size 8 being:



Given an existing tree, a new tree is obtained by substituting in all possible ways to each external node ($\square$) either a pair $(\square, \square)$ or a triple $(\square, \square, \square)$, and symbolically, one has

$$\mathcal{E}[\square] = \square + \mathcal{E}\Big[\square \to (\square\square + \square\square\square)\Big].$$



$$\begin{aligned} x_0 &= 0.6 \\ x_1 &= 0.576 \\ x_2 &= 0.522878976 \\ x_3 &\doteq 0.416358802 \\ x_4 &\doteq 0.245532388 \\ x_5 &\doteq 0.075088357 \\ x_6 &\doteq 0.006061629 \\ x_7 &\doteq 0.000036966 \\ x_8 &\doteq 0.000000001 \\ x_9 &\doteq 1.867434390 \times 10^{-18} \\ x_{10} &\doteq 3.487311201 \times 10^{-36} \end{aligned}$$

FIGURE 16. The iterates of a point $x_0 \in (0, \frac{1}{\varphi})$, here $x_0 = 0.6$, by $\sigma(z) = z^2 + z^3$ converge fast to 0.

In accordance with the specification, the OGF of $\mathcal{E}$ satisfies the functional equation

$$E(z) = z + E(z^2 + z^3), \tag{55}$$

corresponding to the seemingly innocuous recurrence

$$E_n = \sum_{k=0}^{n} \binom{k}{n-2k} E_k \qquad \text{with} \quad E_0 = 0,\ E_1 = 1.$$

Let $\sigma(z) = z^2 + z^3$. Equation (55) can be expanded by iteration in the ring of formal power series,

$$E(z) = z + \sigma(z) + \sigma^{[2]}(z) + \sigma^{[3]}(z) + \cdots , \tag{56}$$

where $\sigma^{[j]}(z)$ denotes the jth iterate of the polynomial σ: $\sigma^{[0]}(z) = z$, $\sigma^{[h+1]}(z) = \sigma^{[h]}(\sigma(z)) = \sigma(\sigma^{[h]}(z))$. Thus, $E(z)$ is nothing but the sum of all iterates of σ. The problem is to determine the radius of convergence of $E(z)$, and by Pringsheim's theorem, the quest for dominant singularities can be limited to the positive real line.

For $z > 0$, the polynomial $\sigma(z)$ has a unique fixed point, $\rho = \sigma(\rho)$, at

$$\rho = \frac{1}{\varphi} \qquad \text{where} \qquad \varphi = \frac{1+\sqrt{5}}{2}$$

is the golden ratio. Also, for any positive x satisfying $x < \rho$, the iterates $\sigma^{[j]}(x)$ do converge to 0; see Figure 16. Furthermore, since $\sigma(z) \sim z^2$ near 0, these iterates converge to 0 doubly exponentially fast (Note 46). By the triangle inequality, $|\sigma(z)| \leq \sigma(|z|)$, the sum in (56) is a normally converging sum of analytic functions, and is thus itself analytic. Consequently $E(z)$ is analytic in the whole of the open disk $|z| < \rho$.

It remains to prove that the radius of convergence of $E(z)$ is exactly equal to ρ. To that purpose it suffices to observe that $E(z)$, as given by (56), satisfies

$$E(x) \to +\infty \qquad \text{as} \qquad x \to \rho^-.$$

Let N be an arbitrarily large but fixed integer. It is possible to select a positive x_N sufficiently close to ρ with $x_N < \rho$, such that the Nth iterate $\sigma^{[N]}(x_N)$ is larger than

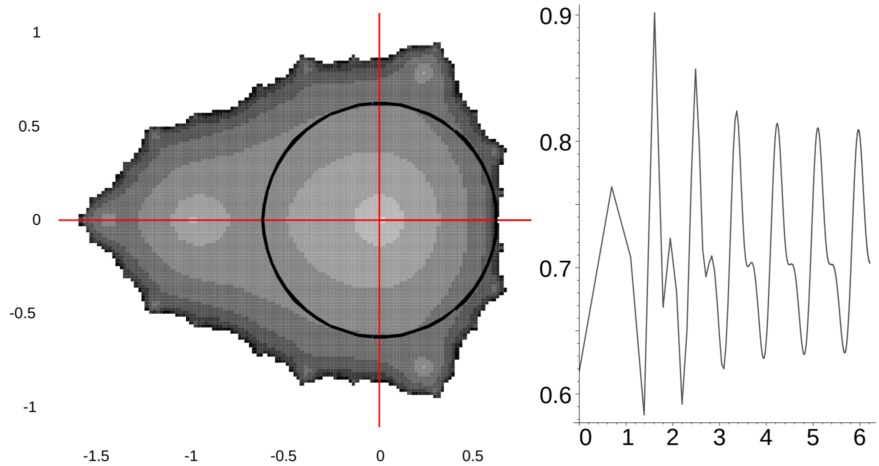


FIGURE 17. Left: the circle of convergence of $E(z)$ and its fractal domain of analyticity (in gray with darker areas representing slower convergence of iterates of σ). Right: the ratio $E_n/(\varphi^n n^{-1})$ plotted against $\log n$ for $n = 1\,..\,500$ confirms that $E_n \bowtie \varphi^n$ and illustrates the periodic fluctuations expressed by Equation (58).

$\frac{1}{2}$ (the function $\sigma^{[N]}(x)$ admits ρ as a fixed point and it is continuous and increasing at ρ). Given the sum expression (56), this entails the lower bound $E(x_N) > \frac{N}{2}$ for such an $x_N < \rho$. Thus $E(x)$ is unbounded as $x \to \rho^-$ and ρ is a singularity.

The dominant positive real singularity of $E(z)$ is thus $\rho = \varphi^{-1}$, and the Exponential Growth Formula gives:

PROPOSITION IV.6. *The number of balanced 2–3 trees satisfies:*

$$[z^n]\,E(z) \bowtie \left(\frac{1+\sqrt{5}}{2}\right)^n. \tag{57}$$

It is notable that this estimate could be established so simply by a purely qualitative examination of the basic functional equation and of a fixed point of the associated iteration scheme.

The complete asymptotic analysis of the E_n requires the full power of singularity analysis methods to be developed in Chapter VI. Equation (58) below states the end result, which involves fluctuations that are clearly visible on Figure 17. There is overconvergence of the representation (56), that is, convergence in certain domains beyond the disc of convergence of $E(z)$. Figure 17 displays the domain of analyticity of $E(z)$ and reveals its fractal nature.

$\triangleright$ **46.** *Quadratic convergence.* First, for $x \in [0, \frac{1}{2}]$, one has $\sigma(x) \le \frac{3}{2}x^2$, so that $\sigma^{[j]}(x) \le (3/2)^{2^j-1}\,x^{2^j}$. Second, for $x \in [0, A]$, where A is any number $< \rho$, there is a number k_A such that $\sigma^{[k_A]}(x) < \frac{1}{2}$, so that $\sigma^{[k]}(x) \le (3/2)\,(3/4)^{2^{k-k_A}}$. Thus, for any $A < \rho$, the series of iterates of σ is quadratically convergent when $z \in [0, A]$. $\triangleleft$

$\triangleright$ **47.** *The asymptotic number of 2–3 trees.* This analysis is from [**328, 330**]. The number of 2–3 trees satisfies asymptotically

$$E_n = \frac{\varphi^n}{n}\Omega(\log n) + O\left(\frac{\varphi^n}{n^2}\right), \tag{58}$$

where Ω is a periodic function with mean value $(\phi \log(4-\varphi))^{-1} \doteq 0.71208$ and period $\log(4-\phi) \doteq 0.86792$. Thus oscillations are inherent in E_n. A plot of the ratio $E_n/(\phi^n/n)$ is offered in Figure 17. $\triangleleft$

IV. 7.3. Complete asymptotics of a functional equation. George Pólya (1887–1985) is mostly remembered by combinatorialists for being at the origin of Pólya theory, a branch of combinatorics that deals with the enumeration of objects invariant under symmetry groups. However, in his classic article [**347, 349**] which founded this theory, Pólya discovered at the same time a number of startling applications of complex analysis to asymptotic enumeration[14] . We detail one of these now.

The combinatorial problem of interest here is the determination of the number M_n of chemical isomeres of alcohols $C_nH_{2n+1}OH$ without asymmetric carbon atoms. The OGF $M(z) = \sum_n M_n z^n$ that starts as (*EIS* **A000621**)

$$M(z) = 1 + z + z^2 + 2z^3 + 3z^4 + 5z^5 + 8z^6 + 14z^7 + 23z^8 + 39z^9 + \cdots, \tag{59}$$

is accessible through a functional equation:

$$M(z) = \frac{1}{1 - zM(z^2)}. \tag{60}$$

Iteration of the functional equation leads to a continued fraction representation,

$$M(z) = \cfrac{1}{1 - \cfrac{z}{1 - \cfrac{z^2}{1 - \cfrac{z^4}{\ddots}}}},$$

from which Pólya found:

PROPOSITION IV.7. *Let $M(z)$ be the solution analytic around 0 of the functional equation*

$$M(z) = \frac{1}{1 - zM(z^2)}.$$

Then, there exist constants K, β, and $B > 1$, such that

$$M_n = K \cdot \beta^n \left(1 + O(B^{-n})\right), \qquad \beta \doteq 1.68136\,75244, \quad K \doteq 0.36071\,40971.$$

PROOF. We offer two proofs. The first one is based on direct consideration of the functional equation and is of a fair degree of applicability. The second one, following Pólya, makes explicit a special linear structure present in the problem. As suggested by the main estimate, the dominant singularity of $M(z)$ is a simple pole.

[14] In many ways, Pólya can be regarded as the grand father of the field of analytic combinatorics.

First proof. By positivity of the functional equation, $M(z)$ dominates coefficientwise any GF $(1 - zM^{<m}(z^2))^{-1}$, where $M^{<m}(z) := \sum_{0\le j<m} M_n z^n$ is the mth truncation of $M(z)$. In particular, one has the domination relation (use $M^{<2}(z) = 1 + z$)

$$M(z) \succeq \frac{1}{1-z-z^3}.$$

Since the rational fraction has its dominant pole at $z \doteq 0.68232$, this implies that the radius ρ of convergence of $M(z)$ satisfies $\rho < 0.69$. In the other direction, since $M(z^2) < M(z)$ for $z \in (0, \rho)$, then, one has the numerical inequality

$$M(z) \le \frac{1}{1 - zM(z)}, \qquad 0 \le z < \rho.$$

This can be used to show (Note 48) that the Catalan generating function $C(z) = (1 - \sqrt{1-4z})/(2z)$ is a majorant of $M(z)$ on the interval $(0, \frac{1}{4})$, which implies that $M(z)$ is well defined and analytic for $z \in (0, \frac{1}{4})$. In other words, one has $\frac{1}{4} \le \rho < 0.69$. Altogether, the radius of convergence of M lies strictly between 0 and 1.

$\triangleright$ **48.** *Alcohols, trees, and bootstrapping.* Since $M(z)$ starts as $1 + z + z^2 + \cdots$ while $C(z)$ starts as $1 + z + 2z^2 + \cdots$, there is a small interval $(0, \epsilon)$ such that $M(z) \le C(z)$. By the functional equation of $M(z)$, one has $M(z) \le C(z)$ for z in the larger interval $(0, \sqrt{\epsilon})$. Bootstrapping then shows that $M(z) \le C(z)$ for $z \in (0, \frac{1}{4})$. $\triangleleft$

Next, as $z \to \rho^-$, one must have $zM(z^2) \to 1$. (Indeed, if this was not the case, we would have $zM(z^2) < A < 1$ for some A. But then, since $\rho^2 < \rho$, the quantity $(1 - zM(z^2))^{-1}$ would be analytic at $z = \rho$, a clear contradiction.) Thus, ρ is determined implicitly by the equation

$$\rho M(\rho^2) = 1, \qquad 0 < \rho < 1.$$

One can estimate ρ numerically (Note 49), and the statement follows with $\beta = 1/\rho$. (Pólya determined ρ to five decimals by hand!)

The previous discussion also implies that ρ is a pole of $M(z)$, which must be simple (since $\partial_z(zM(z^2)\big|_{z=\rho} > 0)$. Thus

$$M(z) \underset{z\to\rho}{\sim} K\frac{1}{1 - z/\rho}, \qquad K := \frac{1}{\rho M(\rho^2) + 2\rho^3 M'(\rho^2)}. \tag{61}$$

The argument shows at the same time that $M(z)$ is meromorphic in $|z| < \sqrt{\rho} \doteq 0.77$. That ρ is the only pole of $M(z)$ on $|z| = \rho$ results from the fact that $zM(z^2) = z + z^3 + \cdots$ can be subjected to the type of argument encountered in the context of the Daffodil Lemma (see the discussion of quasi-inverses in the proof of Proposition IV.3, p. 254). The translation of the singular expansion (61) then yields the statement.

$\triangleright$ **49.** *The growth constant of molecules.* The quantity ρ can be obtained as the limit of the ρ_m satisfying $\sum_{n=0}^{m} M_n \rho_m^{2n+1} = 1$, together with $\rho \in [\frac{1}{4}, 0.069]$. In each case, only a few of the M_n (provided by the functional equation) are needed. One obtains: $\rho_{10} \doteq 0.595$, $\rho_{20} \doteq 0.594756$, $\rho_{30} \doteq 0.59475397$, $\rho_{40} \doteq 0.594753964$. This algorithms constitutes a geometrically convergent scheme with limit $\rho \doteq 0.59475\,39639$. $\triangleleft$

Second proof. First, a sequence of formal approximants follows from (60) starting with

$$1, \quad \frac{1}{1-z}, \quad \frac{1}{1-\dfrac{z}{1-z^2}} = \frac{1-z^2}{1-z-z^2}, \quad \frac{1}{1-\dfrac{z}{1-\dfrac{z^2}{1-z^4}}} = \frac{1-z^2-z^4}{1-z-z^2-z^4+z^5},$$

which permits us to compute any number of terms of the series $M(z)$. Closer examination of (60) suggests to set

$$M(z) = \frac{\psi(z^2)}{\psi(z)},$$

where $\psi(z) = 1 - z - z^2 - z^4 + z^5 - z^8 + z^9 + z^{10} - z^{16} + \cdots$. Back substitution into (60) yields

$$\frac{\psi(z^2)}{\psi(z)} = \frac{1}{1 - z\dfrac{\psi(z^4)}{\psi(z^2)}} \text{ or } \frac{\psi(z^2)}{\psi(z)} = \frac{\psi(z^2)}{\psi(z^2) - z\psi(z^4)},$$

which shows $\psi(z)$ to be a solution of the functional equation

$$\psi(z) = \psi(z^2) - z\psi(z^4), \qquad \psi(0) = 1.$$

The coefficients of ψ satisfy the recurrence

$$\psi_{4n} = \psi_{2n}, \qquad \psi_{4n+1} = -\psi_n, \qquad \psi_{4n+2} = \psi_{2n+1}, \qquad \psi_{4n+3} = 0,$$

which implies that their values are all contained in the set $\{0, -1, +1\}$.

Thus, $M(z)$ appears to be the quotient of two function, $\psi(z^2)/\psi(z)$, each analytic in the unit disc, and $M(z)$ is meromorphic in the unit disc. A numerical evaluation then shows that $\psi(z)$ has its smallest positive real zero at $\rho \doteq 0.59475$, which is a simple root. The quantity ρ is thus a pole of $M(z)$ (since, numerically, $\psi(\rho^2) \neq 0$). Thus

$$M(z) \sim \frac{\psi(\rho^2)}{(z-\rho)\psi'(\rho)} \implies M_n \sim -\frac{\psi(\rho^2)}{\rho\psi'(\rho)} \left(\frac{1}{\rho}\right)^n.$$

Numerical computations then yield Pólya's estimate. Et voilà! □

The example of Pólya's alcohols is exemplary, both from a historical point of view and from a methodological perspective. As the first proof of Proposition IV.7 demonstrates, quite a lot of information can be pulled out of a functional equation without solving it. (A similar situation will be encountered in relation to coin fountains, Example V.7, p. 309.) Here, we have made great use of the fact that if $f(z)$ is analytic in $|z| < r$ and some *a priori* bounds imply the strict inequalities $0 < r < 1$, then one can regard functions like $f(z^2)$, $f(z^3)$, and so on, as "known" since they are analytic in the disc of convergence of f and even beyond, a situation also evocative of our earlier discussion of Pólya operators in Subsection IV. 4. Globally, the lesson is that functional equations, even complicated ones, can be used to bootstrap the local singular behaviour of solutions, and one can often do so even in the absence of any explicit generating function solution. The transition from singularities to coefficient asymptotics is then a simple jump.

▷ **50.** *An arithmetic exercise.* The coefficients $\psi_n = [z^n]\psi(z)$ can be characterized simply in terms of the binary representation of n. Find the asymptotic proportion of the ψ_n for $n \in [1\,..\,2^N]$ that assume each of the values 0, $+1$, and -1. ◁

IV. 8. Perspective

In this chapter, we have started examining generating functions under a new light. Instead of being merely *formal algebraic* objects—power series—that encode *exactly* counting sequences, generating functions can be regarded as *analytic* objects—transformations of the complex plane—whose singularities provide a wealth of information concerning *asymptotic* properties of structures.

Singularities provide a royal road to coefficient asymptotics. We could treat here, with a relatively simple apparatus, singularities that are poles. In this perspective, the two main statements of this chapter are the theorems relative to the expansion of rational and meromorphic functions, (Theorems IV.9 and IV.10). These are classical results of analysis. Issai Schur (1875–1941) is to be counted amongst the very first mathematicians who recognized their rôle in combinatorial enumerations (denumerants, Example 5, p. 244). The complex-analytic thread was developed much further by George Pólya in his famous paper of 1937 (see [**347, 349**]), which Read in [**349**, p. 96] describes as a "landmark in the history of combinatorial analysis". There, Pólya laid the groundwork of combinatorial chemistry, the enumeration of objects under group actions, as well as the complex-asymptotic theory of graphs and trees.

The present chapter serves as the foundation stone of a rich theory to be developed in future chapters. In particular the method of singularity analysis exposed in Chapter VI considerably extends the range of applicability of the Second Principle to functions having singularities appreciably more complicated that poles (e.g., the ones involving fractional powers, logarithms, iterated logarithms, and so on).

As we hope to convince our reader, a consequence of the theory developed in Part B is that most combinatorial classes amenable to symbolic descriptions can be thoroughly analysed, as regards their asymptotic properties, by means of a selected collection of basic theorems of complex analysis. The case of structures like balanced trees and molecules, where only a functional equation of sorts is available, is exemplary.

This chapter has been designed to serve as a refresher of basic complex analysis, with special emphasis on methods relevant for analytic combinatorics. See Figure 18 for a concise summary of results. References most useful for the discussion given here include the books of Titchmarsh [**411**] (oriented towards classical analysis), Whittaker and Watson [**433**] (stressing special functions), Dieudonné [**106**], Hille [**232**], and Knopp [**261**]. Henrici [**229**] presents complex analysis under the perspective of constructive and numerical methods, a highly valuable point of view for this book.

De Bruijn's classic booklet [**93**] is a wonderfully concrete introduction to effective asymptotic theory, and it contains many examples from discrete mathematics thoroughly worked out using a complex-analytic approach. The use of such analytic methods in combinatorics was pioneered in modern times by Bender and Odlyzko, whose first publications in this area go back to the 1970's. The state of affairs in 1995 regarding analytic methods in combinatorial enumeration is superbly summarized in Odlyzko's scholarly chapter [**330**]. Wilf devotes his

Basics. The theory of analytic functions benefits from the equivalence between two notions, analyticity and differentiability. It is the basis of a powerful integral calculus, much different from its real variable counterpart. The following two results can serve as "axioms" of the theory.

THEOREM IV.1 [Basic Equivalence Theorem] (p. 219): Two fundamental notions are equivalent, namely, analyticity (defined by convergent power series) and holomorphy (defined by differentiability). Combinatorial generating functions, *a priori* determined by their expansions at 0 thus satisfy the rich set of properties associated with these two equivalent notions.

THEOREM IV.2 [Null Integral Property] (p. 221): The integral of an analytic function along a simple loop (closed path that can be contracted to a single point) is 0. Consequently, integrals are largely independent of particular details of the integration contour.

Residues. For meromorphic functions (functions with poles), residues are essential. Coefficients of a function can be evaluated by means of integrals. The following two theorems provide connections between local properties of a function (e.g., coefficients at one point) and global properties of the function elsewhere (e.g., an integral along a distant curve).

THEOREM IV.3 [Cauchy's residue theorem] (p. 222): In the realm of meromorphic functions, integrals of a function can be evaluated based on local properties of the function at a few specific points, its poles.

THEOREM IV.4 [Cauchy's Coefficient Formula] (p. 224): This is an almost immediate consequence of Cauchy's residue theorem: The coefficients of an analytic function admit of a representation by a contour integral. Coefficients can then be evaluated or estimated using properties of the function at points away from the origin.

Singularities and growth. Singularities (places where analyticity stops), provide essential information on the growth rate of a function's coefficients. The "First Principle" relates the exponential growth rate of coefficients to the location of singularities.

THEOREM IV.5 [Boundary singularities] (p. 227): A function (given by its series expansion at 0) always has a singularity on the boundary of its disc of convergence.

THEOREM IV.6 [Pringsheim's Theorem] (p. 229): This theorem refines the previous one for functions with non-negative coefficients. It implies that, in the case of combinatorial generating functions, the search for a dominant singularity can be restricted to the positive real axis.

THEOREM IV.7 [Exponential Growth Formula] (p. 231): The exponential growth rate of coefficients of is dictated by the *location* of the singularities nearest to the origin—the *dominant* singularities.

THEOREM IV.8 [Computability of growth] (p. 237): For any combinatorial class that is nonrecursive (iterative), the exponential growth rate of coefficients is invariably a computable number. This statement can be regarded as the first general theorem of analytic combinatorics.

Coefficient asymptotics. The "Second Principle" relates subexponential factors of coefficients to the nature of singularities. For rational and meromorphic functions, everything is simple.

THEOREM IV.9 [Expansion of rational functions] (p. 243): Coefficients of rational functions are explicitly expressible in terms of the poles, given their location (values) and nature (multiplicity).

THEOREM IV.10 [Expansion of meromorphic functions] (p. 245): Coefficients of meromorphic functions admit of a precise asymptotic form with exponentially small error terms, given the location and nature of the dominant poles.

FIGURE 18. A summary of the main results of Chapter IV.

Chapter 5 of *Generatingfunctionoloy* [**437**] to this question. The books by Hofri [**233**] and Szpankowski [**401**] contain useful accounts in the perspective of analysis of algorithms. See also our book [**382**] for a light introduction and the chapter by Vitter and Flajolet [**427**] for more on this specific topic.

V

Applications of Rational and Meromorphic Asymptotics

Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision.

— ANDREW ODLYZKO [**330**]

Contents

The primary goal of this chapter is to provide combinatorial illustrations of the power of complex analytic methods, and specifically of the rational-meromorphic framework exposed in the previous chapter. At the same time, we shift gears and envisage counting problems at a *new level of generality*. Precisely, we organize combinatorial problems into wide *families* of combinatorial types amenable to a common treatment and associated with a common collection of asymptotic properties. Without attempting a formal definition, we call *schema* any such family determined by combinatorial and analytic conditions that covers an infinity of combinatorial classes.

The first schema comprises *regular specifications* and languages, which *a priori* lead to rational generating functions and thus systematically resort to Theorem IV.9 (p. 243), to the effect that coefficients are described as exponential-polynomials. In the case of regular specifications, much additional structure is present, especially positivity. As a consequence, fluctuations can be systematically circumvented. Applications include the analysis of longest runs, corresponding to maximal sequences of good (or bad) luck in games of chance, pure birth processes, and the occurrence of hidden patterns (subsequences) in random texts.

We then consider an important subset of regular specifications, the ones that are built on *nested sequences* and combinatorially correspond to a variety of lattice paths. Such nested sequences naturally lead to nested quasi-inverses, which are none other than continued fractions, A wealth of combinatorial, algebraic, and analytic properties then surround such constructions. A prime illustration is the very explicit analysis of height in Dyck paths and general Catalan trees; other interesting applications relate to coin fountain and interconnection networks.

Next, we discuss a general schema of analytic combinatorics known as the *supercritical sequence* schema, which provides a neat illustration of the power of meromorphic asymptotics while being of a very wide applicability. For instance, one can predict very precisely (and easily) the number of ways in which an integer can be decomposed additively as a sum of primes (or twin primes), this even though many details of the distribution of primes are still surrounded in mystery.

Finally, the last two sections examine *positive linear systems of generating functions*, starting with the simplest case of graphs and automata and concluding with the general framework of transfer matrices. Although the resulting generating functions are once more bound to be rational, there is benefit in examining them as defined implicitly (rather than solving explicitly) and work out singularities directly. The spectrum of matrices (the set of eigenvalues) then plays a central rôle. The crucial technical tool is the Perron-Frobenus theory of nonnegative matrices, whose importance has been long recognized in the theory of finite Markov chains. A general discussion of singularities can then be conducted, leading to valuable consequences on a variety of models—paths in graphs, finite automata, and transfer matrices. The last example discussed in this chapter treats locally constrained permutations, where rational functions combined with inclusion-exclusion provide an entry to the world of value-constrained permutations.

In the various combinatorial examples encountered in this chapter, the generating functions are generally meromorphic in some domain extending beyond their disc of convergence at 0. As a consequence, the asymptotic estimates of coefficients involve main terms that are explicit exponential polynomials and error terms that are exponentially smaller. This is a situation which is well summarized by Odlyzko's aphorism: "*Analytic methods [...] often yield estimates of unparalleled precision*".

V. 1. A roadmap to rational and meromorphic asymptotics

The key character in this chapter is the combinatorial sequence construction SEQ. Since its translation into generating functions involves a *quasi-inverse*, $(1-f)^{-1}$, the construction should in many cases be expected to induce polar singularities. Also, linear systems of equations, of which the simplest case is $X = 1 + AX$, are solvable by means of inverses: the solution is $X = (1-A)^{-1}$ in the scalar case, and it is otherwise expressible as a quotient of determinants, by Cramer's rule, in the vectorial case. Consequently, linear systems of equations are also conducive to polar singularities.

This chapter accordingly develops along two main lines. First, we study non-recursive families of combinatorial problems that are, in a suitable sense, driven by a sequence construction. Second, we examine families of recursive problems that are naturally described by linear systems of equations. Clearly, the general theorems giving the asymptotic forms of coefficients of rational and meromorphic functions apply. As we see here, the additional positivity structure arising from combinatorics often entails notable simplifications in the asymptotic form of counting sequences.

Regular specification and languages. This topic is treated in Section V. 2. Regular specifications are non-recursive specifications that only involve the constructions $(+, \times, \text{SEQ})$. In the unlabelled case, they can always be interpreted as describing a regular language in the sense of Chapter I. The main result here is the following: given a regular specification $\mathcal{R}$, it is possible to determine constructively a number D, so that an asymptotic estimate of the form

$$R_n = P(n)\beta^n + O(B^n), \qquad 0 \le B < \beta, \quad P \text{ a polynomial}, \tag{1}$$

holds, once the index n is restricted to a fixed congruence class modulo D. (Naturally, the quantities P, β, B may depend on the particular congruence class considered.) In other words, a "pure" exponential polynomial form holds for each of the D sections of the counting sequence $(R_n)_{n\ge 0}$. In particular, irregular fluctuations, which might otherwise arise from the existence of several dominant poles sharing the same modulus but having incommensurable arguments (see the discussion in Subsection IV. 6.1, p. IV. 6.1 dedicated to multiple singularities), are simply not present in regular specifications and languages. Similar estimates hold for profiles of regular specifications, where profile of an object is understood as the number of times any fixed construction is employed.

Nested sequences, lattice paths, and continued fractions. What is considered here could be termed the $\text{SEQ} \circ \cdots \circ \text{SEQ}$ schema, corresponding to nested sequences. The resulting GFs are chains of quasi-inverses, that is, continued fractions. Though the general theory of regular specifications applies, the additional structure resulting from nested sequences implies in essence uniqueness and simplicity of the dominant pole, resulting directly in an estimate of the form

$$S_n = c\beta^n + O(B^n), \qquad 0 \le B < \beta, \quad c \in \mathbb{R}_{>0}, \tag{2}$$

for objects enumerated by nested sequences. This schema covers lattice paths of bounded height, their weighted versions, as well as several other bijectively equivalent classes, like interconnection networks. In each case, profiles can be fully characterized, the estimates being of a simple form.

The supercritical sequence. This is a schema of the general form $\mathcal{F} = \text{SEQ}(\mathcal{G})$ with a simple analytic condition, "supercriticality", attached to the the generating function $G(z)$ of $\mathcal{G}$. Under this condition, the sequence (F_n) happens to be predictable and an asymptotic estimate,

$$F_n = cS^n + O(T^n), \qquad 0 \le T < S, \quad c \in \mathbb{R}_{>0}, \tag{3}$$

applies with S such that $G(1/S) = 1$. Integer compositions, surjections, and alignments presented in Chapters I and II can then be treated in a unified manner. The supercritical sequence schema even covers many situations where $\mathcal{G}$ is *not* necessarily constructible—this includes compositions into summands that are prime numbers or twin primes. Parameters, like the number of components and more generally profiles, are under these circumstances governed by laws that hold with a high probability.

Paths in graphs and automata. The framework of paths in directed graphs is of considerable generality. In particular, it covers the case of finite automata introduced in Chapter I. Although, in the abstract, the descriptive power of this framework is formally equivalent to the one of regular specifications (APPENDIX A: *Regular languages*, p. 650), there is great advantage in considering directly problems whose natural formulation is recursive and phrased in terms of graphs or automata. (The reduction of automata to regular expressions is nontrivial so that it does not tend to preserve the original combinatorial structure.) The algebraic theory is that of matrices of the form $(I - zT)^{-1}$, where T is a matrix with nonnegative entries. The analytic theory behind the scene is now that of positive matrices and the companion Perron-Frobenius theory. Uniqueness and simplicity of dominant poles of generating functions can be guaranteed under easily testable structural conditions—principally, the condition of irreducibility that corresponds to a strong connectedness of the system. Then a pure exponential polynomial form of the simplest type holds,

$$C_n \sim c \cdot \lambda_1^n + O(\Lambda^n), \qquad 0 \le \Lambda < \lambda_1, \quad c \in \mathbb{R}_{>0}, \tag{4}$$

where λ_1 is the (unique) dominant eigenvalue of the transition matrix T. Applications include walks over various types of graphs (the interval graph, the devil's staircase) and words excluding one or several patterns (walks on the De Bruijn graph).

Transfer matrices. This framework, whose origins lie in statistical physics, is an extension of automata and paths in graphs. What is retained is the notion of a finite state system, but transitions can now take place at different speeds. Algebraically, one is dealing with matrices of the form $(I - T(z))^{-1}$, where T is a matrix whose entries are polynomials (in z) with nonnegative coefficients. Perron-Frobenius theory can be adapted to cover such cases, that, to a probabilist, involve a mixture of Markov chain and renewal theory. The consequence is once more an estimate of the type (4) for this category of models. A striking application of transfer matrices is a study, with an experimental mathematics flavour, of self-avoiding walks and polygons in the plane, which predicts with a high degree of certainty what the number of polygons is and which distribution of area is to be expected. A combination of the transfer matrix approach with a suitable use of inclusion-exclusion finally provides (Subsection V. 6.4) a solution to the classic *ménage problem* of combinatorial theory as well as to many related questions regarding value-constrained permutations.

Sections V. 2 to V. 6 are organized following a common pattern: first, we discuss "combinatorial aspects", then "analytic aspects", and finally "applications". Each of Sections V. 2 to V. 5 is furthermore centred around two analytic-combinatorial theorems, one describing *asymptotic enumeration*, the other quantifying the *asymptotic profiles* of combinatorial structures. The last section (Section V. 6) departs slightly from this general pattern: transfer matrices are reducible rather simply the framework of paths in graphs and automata, presented in the immediately preceding section, so that, in order to avoid redundancy, the corresponding theorems are not explicitly stated.

V. 2. Regular specification and languages

The purpose of this section is the general study of the $(+, \times, \textsc{Seq})$ schema, which covers all regular specifications. As we show here, pure exponential-polynomial forms with a single dominating exponential can always be extracted. Theorems V.1 and V.2 thus provide a universal framework for the asymptotic analysis of regular classes. Additional structural conditions to be introduced in later sections (nested sequences, irreducibility of the dependency graph and of transfer matrices) will then be seen to induce further simplifications in asymptotic formuæ.

V. 2.1. Combinatorial aspects. For convenience and without loss of analytic generality, we consider here unlabelled structures. According to Chapter I, a combinatorial specification is *regular* if it is nonrecursive ("iterative") and it involves only the constructions of Atom, Union, Product, and Sequence. A language $\mathcal{L}$ is S-regular if it is combinatorially isomorphic to a class $\mathcal{M}$ described by a regular specification. Alternatively, a language is S-regular if all the operations involved in its description (unions, catenation products and star operations) are unambiguous. See Definition I.10 (p. 48) and the companion Proposition I.2 (p. 48).

The dictionary translating constructions into OGFs is

$$\mathcal{F}+\mathcal{G} \mapsto F+G, \qquad \mathcal{F}\times\mathcal{G} \mapsto F\times G, \qquad \mathrm{SEQ}(\mathcal{F}) \mapsto (1-F)^{-1}, \tag{5}$$

and for languages, under the essential condition of *non-ambiguity*,

$$\mathcal{L}\cup\mathcal{M} \mapsto L+M, \qquad \mathcal{L}\cdot\mathcal{M} \mapsto L\times M, \qquad \mathcal{L}^{\star} \mapsto (1-L)^{-1}. \tag{6}$$

The rules (5) and (6) then give rise to generating functions that are invariably *rational functions*. Consequently, given a regular class $\mathcal{C}$, the exponential-polynomial form of coefficients expressed by Theorem IV.9 systematically applies, and one has

$$C_n \equiv [z^n]C(z) = \sum_{j=1}^{m} \Pi_j(n)\alpha_j^{-n}, \tag{7}$$

for a family of algebraic numbers α_j (the poles of $C(z)$) and a family of polynomials Π_j.

As we know from the discussion of periodicities in Section IV. 6.1 (p. 250, the collective behaviour of the sum in (7) depends on whether or not a single α dominates all others in modulus. In the case where several dominant singularities coexist, fluctuations of sorts (either periodic or irregular) may manifest themselves. In contrast, if a single α dominates, then the exponential-polynomial formula acquires a transparent asymptotic meaning. Accordingly, we set:

DEFINITION V.1. *An exponential-polynomial form $\sum_{j=1}^{m}\Pi_j(n)\alpha_j^{-n}$ is said to be* pure *if $|\alpha_1| < |\alpha_j|$, for all $j \geq 2$. In that case, a single exponential dominates all the other ones.*

As we see next for regular languages and specifications, the corresponding counting coefficients can always be described by a *finite collection* of pure exponential polynomial forms. The fundamental reason is that we are dealing with a special subset of rational functions, one that enjoys strong positivity properties.

$\triangleright$ **1.** *Positive rational functions.* Define a the class Rat^{+} of *positive rational functions* as the smallest class containing polynomials with positive coefficients ($\mathbb{R}_{\geq 0}[z]$) and closed under sum, product, and quasi-inverse, where $Q(f) = (1-f)^{-1}$ is applied to elements f such that $f(0) = 0$. The OGF of any regular class with positive weights attached to neutral structures and atoms is in Rat^{+}. Conversely, any function in Rat^{+} is the OGF of a positively weighted regular class. The notion of a Rat^{+} function is for instance useful in the analysis of weighted word models and Bernoulli trials, as discussed in Section III. 6.1, p. 178. $\triangleleft$

V. 2.2. Analytic aspects. First we need the notion of sections of a sequence.

DEFINITION V.2. *Let (f_n) be a sequence of numbers. Its* section *of parameters D, r, where $D \in \mathbb{Z}_{>0}$ and $r \in \mathbb{Z}_{\geq 0}$ is the subsequence (f_{nD+r}). The numbers D and r are referred to as the modulus and the base respectively.*

The main theorem describing the asymptotic behaviour of regular classes is a consequence of Proposition IV.3 (p. 254) and is originally due to Berstel. (See Soittola's article [**387**] as well as the books by Eilenberg [**123**, Ch VII] and Berstel–Reutenauer [**42**] for context and proofs of some of the assertions below.)

THEOREM V.1 (Asymptotics of regular classes). *Let $\mathcal{S}$ be a class described by a regular specification. Then there exists an integer D such that each section of modulus D of S_n that is not eventually 0 admits a pure exponential polynomial form: for n larger than some n_0, and any such section of base r, one has*

$$S_n = \Pi(n)\beta^n + \sum_{j=1}^{m} P_j(n)\beta_j^n \qquad n \equiv r \bmod D,$$

where $\beta > |\beta_j|$, and Π, P_j are polynomials that depend on the base r, with $\Pi(x) \not\equiv 0$.

PROOF. Let α_1 be the dominant pole of $S(z)$ that is positive. Proposition IV.3 asserts that any dominant pole, α is such that $\alpha/|\alpha|$ is a root of unity. Let D_0 be such that the dominant singularities are all contained in the set $\{\alpha_1\omega^{j-1}\}_{j=0}^{D_0-1}$, where $\omega = \exp(2i\pi/D_0)$. By collecting all contributions arising from dominant poles in the general expansion (7) and by restricting n to a fixed congruence class modulo D_0, namely $n = r + D_0\nu$ with $0 \leq r < D_0$, one gets

$$S_{r+D_0\nu} = \Pi^{[r]}(n)\alpha_1^{-D_0\nu} + O(A^{-n}). \tag{8}$$

There $\Pi^{[r]}$ is a polynomial depending on r and the remainder term represents an exponential polynomial with growth at most $O(A^{-n})$ for some $A > \alpha_1$.

The sections with modulus D_0 that are not eventually 0 can be categorized into two classes.

— Let $\mathcal{R}_{\neq 0}$ be the set of those values of r such that $\Pi^{[r]}$ is not identically 0. The set $\mathcal{R}_{\neq 0}$ is nonempty (else the radius of convergence of $S(z)$ would be larger than α_1.) For any base $r \in \mathcal{R}_{\neq 0}$, the assertion of the theorem is then established with $\beta = 1/\alpha_1$.

— Let $\mathcal{R}_0$ be the set of those values of r such that $\Pi^{[r]}(x) \equiv 0$, with $\Pi^{[r]}$ as given by (8). Then one needs to examine the next layer of poles of $S(z)$, as detailed below.

Consider a number r such that $r \in \mathcal{R}_0$, so that the polynomial $\Pi^{[r]}$ is identically 0. First, we isolate in the expansion of $S(z)$ those indices that are congruent to r modulo D_0. This is achieved by means of a Hadamard product:

$$g(z) = S(z) \odot \left(\frac{z^r}{1 - z^{D_0}}\right).$$

A classical theorem [**45, 123**] from the theory of positive rational functions in the sense of Note 1 asserts that such functions are closed under Hadamard product. (A dedicated construction is also possible.) Then the resulting function $G(z)$ is of the form

$$g(z) = z^r \gamma(z^{D_0}),$$

with the rational function $\gamma(z)$ being analytic at 0. Note that we have $[z^\nu]\gamma(z) = S_{\nu D_0 + r}$, so that γ is exactly the generating function of the section of base r of $S(z)$. One verifies next that $\gamma(z)$, which is obtained by the substitution $z \mapsto z^{1/D_0}$ in $S(z)z^{-r}$, is itself a positive rational function. Then, by a fresh application of Berstel's Theorem (Proposition IV.3, p. 254) this function, if not a polynomial, has a radius of convergence ρ with all its dominant poles σ being such that σ/ρ is a D_1 root of unity for some $D_1 \geq 1$. The argument originally applied to $S(z)$ can thus be repeated, with $\gamma(z)$ replacing $S(z)$. In particular, one finds at least one section (of modulus D_1) of the coefficients of $\gamma(z)$ that admits a pure exponential-polynomial form. The other sections of modulus D_1 can themselves be further refined, and so on

In other words, successive refinements of the sectioning process provide at each stage at least one pure exponential-polynomial form, possibly leaving a few congruence classes open for further refinements. Define the *layer index* of a rational function f as the integer $\kappa(f)$, such that

$$\kappa(f) = \operatorname{card}\left\{|\zeta| \;\middle|\; f(\zeta) = \infty\right\}.$$

(This index is thus the number of different moduli of poles of f.) It is seen that each successive refinement step decreases by at least 1 the layer index of the rational function involved, thereby ensuring termination of the whole refinement process. Finally, the collection of the iterated sectionings obtained can be reduced to a single sectioning according to a common modulus D, which is the least common multiple of the collection of all the finite products $D_0 D_1 \cdots$ that are generated by the algorithm. $\square$

For instance the coefficients (Figure 1) of the function

$$F(z) = \frac{1}{(1-z)(1-z^2-z^4)} + \frac{z}{1-3z^3}, \tag{9}$$

associated to the regular language $a^\star(bb+cccc)^\star + d(ddd+eee+fff)^\star$, exhibit an apparently irregular behaviour, with the expansion of $F(z)$ starting as

$$1 + 2z + 2z^2 + 2z^3 + 7z^4 + 4z^5 + 7z^6 + 16z^7 + 12z^8 + 12z^9 + 47z^{10} + 20z^{11} + \cdots.$$

However the sections modulo 6 each admit a pure exponential-polynomial form and consequently become easy to describe.

$\triangleright$ **2.** *Extension to* Rat^+ *functions.* The conclusions of Theorem V.1 hold for any function in Rat^+ in the sense of Note 1. $\triangleleft$

$\triangleright$ **3.** *Soittola's Theorem.* This is a converse to Theorem V.1 proved in [**387**]. Assume that coefficients of an *arbitrary* rational function $f(z)$ are nonnegative and that there exists a sectioning such that each section admits a pure exponential-polynomial form. Then $f(z)$ is in Rat^+ in the sense of Note 1; in particular, f is the OGF of a (weighted) regular class. $\triangleleft$

Theorem V.1 is useful for interpreting the enumeration of regular classes and languages. It serves a similar purpose with regards to structural parameters of regular classes. Consider a regular specification $\mathcal{C}$ augmented with a mark u that is, as usual, a neutral object of size 0 (see Chapter III). We let $C(z,u)$ be the corresponding BGF

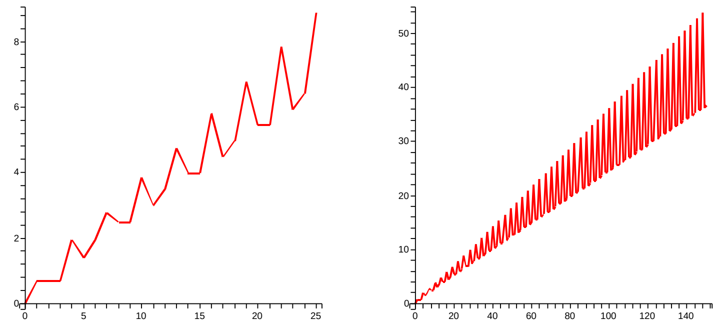


FIGURE 1. Plots of $\log F_n$ with $F_n = [z^n]F(z)$ and $F(z)$ as in (9) display fluctuations that disappear as soon as sections of modulus 6 are considered.

of $\mathcal{C}$, so that $C_{n,k} = [z^n u^k]C(z,u)$ is the number of $\mathcal{C}$-objects of size n that bear k marks. A suitable placement of marks makes it possible to record the number of times any given construction enters an object. For instance, in the augmented specification of binary words,

$$\mathcal{C} = \text{SEQ}(b)\,\text{SEQ}(a(\text{SEQ}_{<r}(b) + u\,\text{SEQ}_{\geq r}(b))),$$

all maximal runs of b having length at least r are marked by a u. There results the following BGF for the corresponding parameter "number of b-runs of length $\geq r$",

$$C(z,u) = \frac{1}{1-z}\frac{1}{1-z\left(\frac{1-z^r}{1-z} + \frac{uz^r}{1-z}\right)},$$

from which mean and variance can be determined. In a sense, marks make it possible to analyse profile, with respect to constructions entering the specification, of a random object.

THEOREM V.2 (Profile of regular classes). *Consider a regular specification $\mathcal{C}$ augmented with a mark and let χ be the parameter corresponding to the number of occurrences of that mark. There exists a sectioning index d such that for any fixed section of (C_n) of modulus d, the following hold: Any moment of integral order $s \geq 1$ of χ satisfies an asymptotic formula*

$$\mathbb{E}_{\mathcal{C}_n}[\chi^s] = Q(n)\gamma^n + O(G^n), \tag{10}$$

where $0 < \gamma \leq 1$, $Q(n)$ is a rational fraction, and $G < \gamma$.

In this statement, it is tacitly assumed that only sections that are not eventually 0 are considered.

PROOF. The case of expectations suffices to indicate the lines of a general proof. One possible approach[1] is to build a derived specification $\mathcal{D}$ such that

$$\mathbb{E}_{\mathcal{C}_n}[\chi] = \frac{D_n}{C_n},$$

which is also a regular specification. To this purpose, define a transformation on specifications defined inductively by the rules

$$\partial(A+B) = \partial A + \partial B,\ \partial(A\times B) = \partial A\times B + A\times\partial B,\ \partial\,\mathrm{SEQ}(A) = \mathrm{SEQ}(A)\times\partial A\times\partial A,$$

together with the initial conditions $\partial u = \mathbf{1}$ and $\partial\mathcal{Z} = \emptyset$. This is a form of combinatorial differentiation: an object $\gamma\in\mathcal{C}$ corresponds to $\chi(\gamma)$ objects in $\mathcal{D}$, namely, one for each choice of an occurrence of the mark.

As a consequence, D_n is the cumulated value of χ over $\mathcal{C}_n$, so that $D_n/C_n = \mathbb{E}_{\mathcal{C}_n}[\chi]$. On the other hand, $\mathcal{D}$ is a regular specification to which Theorem V.1 applies. The result follows upon considering (if necessary) a sectioning that refines the sectionings of both $\mathcal{C}$ and $\mathcal{D}$.

□

▷ **4.** *An example.* Consider the regular language $\mathcal{C} = a^\star(b+c)^\star d(b+c)^\star$. Let χ be the length of the initial run of a's. Then one finds

$$C(z) = \frac{z}{(1-z)(1-2z)^2}, \qquad D(z) = \frac{z^2}{(1-z)^2(1-2z)^2}.$$

Thus the mean of χ satisfies

$$\mathbb{E}_{\mathcal{C}_n}[\chi] = \frac{D_n}{C_n} = \frac{(n-3)2^n+(n+3)}{(n-1)2^n+1} = \frac{n-3}{n-1} + O\left(\left(\frac{3}{2}\right)^n\right).$$

Generally, in the statement of Theorem V.2, let $Q(n) = A(n)/B(n)$ with A, B polynomials and $a = \deg(A), b = \deg(B)$. The following combinations prove to be possible: $\gamma = 1$ and (a, b) any pair such that $0 \le a \le b+1$; $\gamma < 1$ and (a, b) any pair of elements ≥ 0. ◁

▷ **5.** *Shuffle products.* Let $\mathcal{L}, \mathcal{M}$ be two languages over two disjoint alphabets. Then, the shuffle product $\mathcal{S}$ of $\mathcal{L}$ and $\mathcal{M}$ is such that $\widehat{S}(z) = \widehat{L}(z)\cdot\widehat{M}(z)$, where $\widehat{S}, \widehat{L}, \widehat{M}$ are the exponential generating functions of $\mathcal{S}, \mathcal{L}, \mathcal{M}$. Accordingly, if the OGF $L(z)$ and $M(z)$ are rational then the OGF $S(z)$ is also rational. [This technique may be used to analyse generalized birthday paradox and coupon collector problems; see [**151**].] ◁

V. 2.3. Applications. This subsection details several examples that illustrate the explicit determination of exponential-polynomial forms in regular specifications. Various types of estimates conforming to Theorems V.1 and V.2 are obtained.

— We start by recapitulating a collection of combinatorial problems (a "pot-pourri", Example 1) already encountered in Chapters I–III, where rational function asymptotics has been used *en passant* .

— Next, we show how to develop a complete analysis of runs of consecutive equal letters in random sequences (Example 2): this is in theory a special case of the analysis of patterns in random texts (Section IV. 6.3, p. 258), but

[1] Equivalently, one may operate at generating function level and observe that the derivative of a Rat^+ function is Rat^+; cf Notes 1 and 2.

Class	*Asymptotics*	
Integer compositions	2^{n-1}	
— k summands	$\sim \frac{n^{k-1}}{(k-1)!}$	(§I. 3.1, p. 42)
— summands $\leq r$	$\sim c\beta_r^n$	(§I. 3.1, p. 40)
Integer partitions		
— k summands	$\sim \frac{n^{k-1}}{k!\,(k-1)!}$	(§I. 3.1, p. 42)
— summands $\leq r$	$\sim \frac{n^{r-1}}{r!\,(r-1)!}$	(§I. 3.1, p. 41)
Set partitions, k classes	$\sim \frac{k^n}{k!}$	(§I. 4.3, p. 59)
Words excluding a pattern $\mathfrak{p}$	$\sim c\beta_{\mathfrak{p}}^n$	(§IV. 6.3, p. 258)

FIGURE 2. A potpourri of regular classes and their asymptotics.

the particular nature of the patterns makes it possible to derive much more explicit results, including limit distributions for longest runs.

— We then examine walks of the pure birth type (Example 3) that turn out to have applications to the analysis of a probabilistic algorithm (Approximate Counting, Example 4).

— Finally, we present a mean and variance analysis of the occurrence of hidden patterns in random texts (subsequences, Example 5), which is sufficient to entail the concentration of distribution property.

EXAMPLE 1. *A potpourri of regular specifications.* We gather here a few combinatorial problems to be found scattered across Chapters I–IV that are reducible to regular specifications; see also Figure 2 for a summary.

Compositions of integers (Section I. 3, p. 37) are specified by $\mathcal{C} = \text{SEQ}(\text{SEQ}_{\geq 1}(\mathcal{Z}))$, whence the OGF $(1-z)/(1-2z)$ and the closed form $C_n = 2^{n-1}$, an especially trivial exponential-polynomial form. Polar singularities are also present for compositions into k summands that are described by $\text{SEQ}_k(\text{SEQ}_{\geq 1}(\mathcal{Z}))$ and for compositions whose summands are restricted to the interval $[1\,..\,r]$ (i.e., $\text{SEQ}(\text{SEQ}_{1\,..\,r}(\mathcal{Z}))$, with corresponding generating functions

$$\frac{z^k}{(1-z)^k}, \qquad \frac{1-z}{1-2z+z^{r+1}}.$$

In the first case, there is an explicit form for the coefficients, $\binom{n-1}{k-1}$, which constitutes a particular exponential-polynomial form (with the basis of the exponential being 1). The second case requires a dedicated analysis of the dominant polar singularity. (Example 2 below treats the closely related problem of determining longest runs in random binary words.)

Integer partitions involve the multiset construction. However, when summands are restricted to the interval $[1\,..\,r]$, the specification and the OGF are given by

$$\text{MSET}(\text{SEQ}_{1\,..\,r}(\mathcal{Z})) \simeq \text{SEQ}(\mathcal{Z}) \times \text{SEQ}(\mathcal{Z}^2) \times \cdots \text{SEQ}(\mathcal{Z}^r) \quad \Longrightarrow \quad \prod_{j=1}^{r} \frac{1}{1-z^j}.$$

This case first introduced in Section I. 3 (p. 37) has also served as a leading example in our discussion of denumerants in Example IV.5 (p. 244), where the analysis of the pole at 1 furnishes the dominant asymptotic behaviour, $n^{r-1}/(r!(r-1)!)$, for such special partitions. The enumeration of partitions by number of parts then follows, by duality, from the staircase representation.

Set partitions are typically labelled objects. However, when suitably constrained, they can be encoded by regular expressions; see Section I. 4.3 (p. 59) for partitions into k classes, where the OGF found is

$$S^{(k)}(z) = \frac{z^k}{(1-z)(1-2z)\cdots(1-kz)} \qquad \text{implying} \qquad S_n^{(k)} \sim \frac{k^n}{k!},$$

and the asymptotic estimate results from the dominant pole at $1/k$.

Words lead to many problems that are prototypical of the regular specification framework. In Section I. 4 (p. 47), we saw that one could give a regular expression describing the set of words containing the pattern abb, from which the exact and asymptotic forms of counting coefficients derive. For a general pattern $\mathfrak{p}$, the generating functions of words constrained to include (or dually exclude) $\mathfrak{p}$ are rational. The corresponding asymptotic analysis has been given in Section IV. 6.3 (p. 258).

Words can also be analysed under the Bernoulli model, where letter i is selected with probability p_i; cf Section III. 6.1 for a general discussion including the analysis of records in random words (p. 179). . . . . END OF EXAMPLE 1. $\square$

$\triangleright$ **6.** *Partially commutative monoids.* Let $\mathcal{W} = \mathcal{A}^\star$ be the set of all words over a finite alphabet $\mathcal{A}$. Consider a collection C of commutation rules between pairs of elements of A. For instance, if $\mathcal{A} = \{a, b, c\}$, then $\mathrm{C} = \{ab = ba, ac = ca\}$ means that a commutes with both b and c, but bc is not a commuting pair: $bc \neq cb$. Let $\mathcal{M} = \mathcal{W}/[\mathrm{C}]$ be the set of equivalent classes of words (monomials) under the rules induced by C. The set $\mathcal{M}$ is said to be a *partially commutative monoid* or a trace monoid [**74**].

If $A = \{a, b\}$, then the two possibilities for C are $\mathrm{C} = \emptyset$ and $\mathrm{C} := \{ab = ba\}$. Normal forms for $\mathcal{M}$ are given by the regular expressions $(a+b)^\star$ and $a^\star b^\star$ corresponding to the OGFs

$$\frac{1}{1-a-b}, \qquad \frac{1}{1-a-b+ab}.$$

If $\mathcal{A} = \{a, b, c\}$, the possibilities for C, the corresponding normal forms, and the OGFs M are as follows. If $\mathrm{C} = \emptyset$, then $\mathcal{M} \simeq (a+b+c)^\star$ with OGF $(1-a-b-c)^{-1}$; the other cases are

$$\begin{array}{ccc} ab = ba & ab = ba,\ ac = ca & ab = ba,\ ac = ca,\ bc = cb \\ (a^\star b^\star c)^\star a^\star b^\star & a^\star (b+c)^\star & a^\star b^\star c^\star \\ \dfrac{1}{1-a-b-c+ab} & \dfrac{1}{1-a-b-c+ab+ac} & \dfrac{1}{1-a-b-c+ab+ac+bc-abc} \end{array}.$$

Cartier and Foata [**74**] have discovered the general form (based on extended Möbius inversion),

$$M = \left(\sum_F (-1)^{|F|} F \right)^{-1},$$

where the sum is over all monomials F composed of distinct letters that all commute pairwise. Goldwurm and Santini [**203**] have shown that $[z^n]M(z) \sim K \cdot \alpha^n$ for $K, \alpha > 0$. $\triangleleft$

EXAMPLE 2. *Longest runs in words* Longest runs in words introduced in Section I. 4.1 (p. 47) provide an illustration of the technique of localizing dominant singularities in rational functions and of the corresponding coefficient extraction process. The probabilistic problem is a famous one, discussed by Feller in [**133**], as it represents a basic question in the analysis of runs of good (or bad) luck in a succession of independent events. Our presentation closely follows an insightful note of Knuth [**265**] whose motivation was the analysis of carry propagation in certain binary adders.

Start from the class $\mathcal{W}$ of all binary words over the alphabet $\{a,b\}$. Our interest lies in the length L of the longest consecutive block of a's in a word. For the the property $L<k$, the specification and the corresponding OGF are

$$\mathcal{W}^{\langle k\rangle} = \text{SEQ}_{<k}(a)\,\text{SEQ}(b\,\text{SEQ}_{<k}(a)) \qquad \Longrightarrow \qquad W^{\langle k\rangle}(z) = \frac{1-z^k}{1-z}\cdot\frac{1}{1-z\frac{1-z^k}{1-k}},$$

that is,

$$W^{\langle k\rangle}(z) = \frac{1-z^k}{1-2z+z^{k+1}}. \tag{11}$$

This represents a collection of OGFs indexed by k, which contain all the information relative to the distribution of longest runs in random words. We propose to prove:

PROPOSITION V.1. *The longest run parameter L taken over the set of binary words of length n (endowed with the uniform distribution) satisfies the uniform estimate*[2]

$$\mathbb{P}_n\left(L<\lfloor \lg n\rfloor + h\right) = e^{-\alpha(n)2^{-h-1}} + O\left(\frac{\log n}{\sqrt{n}}\right), \qquad \alpha(n) := 2^{\{\lg n\}}. \tag{12}$$

In particular, the mean satisfies

$$\mathbb{E}_n(L) = \lg n + \frac{\gamma}{\log 2} - \frac{3}{2} + P(\lg x) + O\left(\frac{\log^2 n}{\sqrt{n}}\right),$$

where P is a continuous periodic function whose Fourier expansion is given by (20). *The variance satisfies $\mathbb{V}_n(L)=O(1)$ and the distribution is concentrated around its mean.*

The probability distributions appearing in (12) are know as *double exponential distributions* (Figure 3). The formula (12) does not represent a single limit distribution in the usual sense of Chapter IX, but rather a whole *family of distributions* indexed by the fractional part of $\lg n$, thus dictated by the way n places itself with respect to powers of 2.

PROOF. The proof consists of the following steps: locate the dominant pole; estimate the corresponding contribution; separate the dominant pole from the other poles in order to derive constructive error terms; finally approximate the main quantities of interest.

(i) *Location of the dominant pole.* The OGF $W^{\langle k\rangle}$ has, by the first form of (11) a dominant pole ρ_k which is a root of the equation $1=s(\rho_k)$, where $s(z)=z(1-z^k)/(1-z)$. We consider $k\geq 2$. Since $s(z)$ is an increasing polynomial and $s(0)=0$, $s(\frac{1}{2})<1$, $s(1)=1$, the root ρ_k must lie in the open interval $(\frac{1}{2},1)$. In fact, as one easily verifies, the condition $k\geq 2$ guarantees that $s(0.6)>1$, hence the refined estimate

$$\frac{1}{2}<\rho_k<\frac{3}{5} \qquad (k\geq 2). \tag{13}$$

It now becomes possible to derive very precise estimates by bootstrapping. (This technique is a form of iteration for approaching a fixed point—its use in the context of asymptotic expansions is detailed in De Bruijn's book [**93**].) Writing the defining equation for ρ_k as a fixed point equation,

$$z=\frac{1}{2}(1+z^{k+1}),$$

and making use of the rough estimates (13) yields next

$$\frac{1}{2}\left(1+(\frac{1}{2})^{k+1}\right) < \rho_k < \frac{1}{2}\left(1+(\frac{3}{5})^{k+1}\right). \tag{14}$$

[2]The symbol $\lg x$ denotes the binary logarithm, $\lg x = \log_2 x$.

Thus, ρ_k is exponentially close to $\frac{1}{2}$, and further iteration from (14) shows

$$\rho_k = \frac{1}{2} + \frac{1}{2^{k+2}} + O\left(\frac{k}{2^{2k}}\right), \tag{15}$$

which constitutes a very precise estimate.

(ii) *Contribution from the dominant pole.* A straightforward calculation provides the value of the residue,

$$R_{n,k} := -\operatorname{Res}\left[W^{\langle k\rangle}(z)z^{-n-1}; z = \rho_k\right] = \frac{1-\rho_k^k}{2-(k+1)\rho^k}\rho_k^{-n-1}, \tag{16}$$

which is expected to provide the main approximation to the coefficients of $W^{\langle k\rangle}$ as $n \to \infty$. The quantity in (16) is of the rough form $2^n e^{-n/2^{k+1}}$: we shall return to such approximations shortly.

(iii) *Separation of the subdominant poles.* Consider the circle $|z| = \frac{3}{4}$ and take the second form of the denominator of $W^{\langle k\rangle}$, namely,

$$1 - 2z + z^{k+1}.$$

In view of Rouché's theorem, we may regard this polynomial as the sum $f(z) + g(z)$, where $f(z) = 1 - 2z$ and $g(z) = z^{k+1}$. The term $f(z)$ has on the circle a modulus that varies between $\frac{1}{2}$ and $\frac{5}{2}$; the term $g(z)$ is at most $\frac{27}{64}$ for any $k \geq 2$. Thus, on the circle $|z| = \frac{3}{4}$, one has $|g(z)| < |f(z)|$, so that $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside the circle. Since $f(z)$ admits $z = \frac{1}{2}$ as only zero there, the denominator must also have a unique root in $|z| \leq \frac{3}{4}$, and that root must coincide with ρ_k.

Similar arguments also give bounds on the error term when the number of words w satisfying $L(w) < k$ is estimated by the residue (16) at the dominant pole. On the circle $|z| = \frac{3}{4}$, the denominator of $W^{\langle k\rangle}$ stays bounded away from 0 (its modulus is at least $\frac{5}{64}$ when $k \geq 2$, by previous considerations). Thus, the modulus of the remainder integral is $O((4/3)^n)$, and in fact bounded from above by $35(4/3)^n$. In summary, letting $q_{n,k}$ represent the probability that the longest run in a random word of length n is less than k, one obtains the the main estimate

$$q_{n,k} := \mathbb{P}_n(L < k) = \frac{1-\rho_k^k}{1-(k+1)\rho_k^k/2}\left(\frac{1}{2\rho_k}\right)^{n+1} + O\left((\frac{2}{3})^n\right), \tag{17}$$

which holds *uniformly* with respect to k. Here is table of the numerical values of the quantities appearing in the approximation of $q_{n,k}$ when written under the form $c_k \cdot (2\rho_k)^{-n}$:

k	$c_k \cdot (2\rho_k)^{-n}$
2	$1.17082 \cdot 0.80901^n$
3	$1.13745 \cdot 0.91964^n$
4	$1.09166 \cdot 0.96378^n$
5	$1.05753 \cdot 0.98297^n$
10	$1.00394 \cdot 0.99950^n$

(iv) *Final approximations.* There only remains to transform the main estimate (17) into the limit form asserted in the statement. First, the "tail inequalities" ($\lg x \equiv \log_2 x$)

$$\mathbb{P}_n\left(L < \frac{3}{4}\lg n\right) = O\left(e^{-\frac{1}{2}\sqrt[4]{n}}\right), \qquad \mathbb{P}_n\left(L \geq 2\lg n + y\right) = O\left(\frac{e^{-2y}}{n}\right), \tag{18}$$

describe the tail of the probability distribution of L_n. They derive from simple bounding techniques applied to the main approximation (17) using (15). Thus, for asymptotic purposes, only a small region around $\lg n$ needs to be considered.

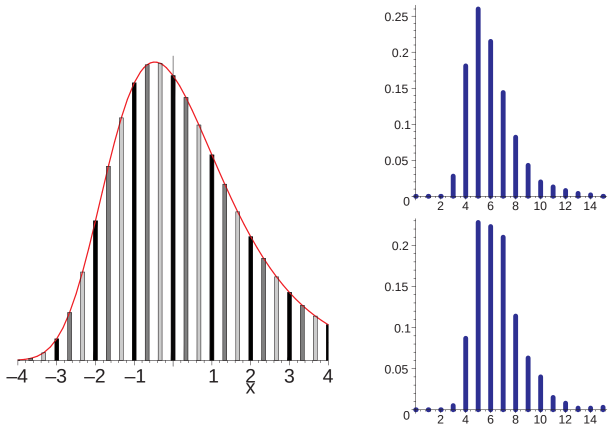


FIGURE 3. The double exponential laws: Left, histograms for n at 2^p (black), $2^{p+1/3}$ (dark gray), and $2^{p+2/3}$ (light gray), where $x = k - \lg n$. Right, empirical histograms for 1000 simulations with $n = 100$ (top) and $n = 140$ (bottom).

Regarding the central regime, for $k = \lg n + x$ and x in $[-\frac{1}{4}\lg n, \lg n]$, the approximation (15) of ρ_k and related quantities applies, and one finds

$$(2\rho_k)^{-n} = \exp\left(-\frac{n}{2^{k+1}} + O(kn2^{-2k})\right) = e^{-n/2^{k+1}}\left(1 + O(\frac{\log n}{\sqrt{n}})\right).$$

(This results from standard expansions like $(1-a)^n = e^{-na}\exp(O(na^2))$.) At the same time, the coefficient of this quantity in (17) is

$$1 + O(k\rho_k^k) = 1 + O\left(\frac{\log n}{\sqrt{n}}\right).$$

Thus a double exponential approximation holds (Figure 3): for $k = \lg n + x$ with x in $[-\frac{1}{4}\lg n, \lg n]$, one has (uniformly)

$$q_{n,k} = e^{-n/2^{k+1}}\left(1 + O\left(\frac{\log n}{\sqrt{n}}\right)\right). \tag{19}$$

In particular, upon setting $k = \lfloor \lg n \rfloor + h$ and making use of the tail inequalities (18), the first part of the statement, namely Equation (12), follows. (The floor function takes into account the fact that k must be an integer.)

The mean and variance estimates derive from the fact that the distribution quickly decays at values away from $\lg n$ (by (18)) while it satisfies Equation (19) in the central region. The mean satisfies

$$\mathbb{E}_n(L) := \sum_{h\geq 1}[1 - \mathbb{P}_n(L < h)] = \Phi(\frac{n}{2}) - 1 + O\left(\frac{\log^2 n}{n}\right), \quad \Phi(x) := \sum_{h\geq 0}\left[1 - e^{-x/2^h}\right].$$

Consider the three cases $h < h_0$, $h \in [h_0, h_1]$, and $h > h_1$ with $h_0 = \lg x - \log\log x$ and $h_1 = \lg x + \log\log x$, where the general term is (respectively) close to 1, between 0 and 1, and close to 0. By summing, one finds elementarily $\Phi(x) = \lg x + O(\log\log x)$ as $x \to \infty$. (An elementary way of catching the next $O(1)$ term is discussed for instance in [**382**, p. 403].)

The method of choice for precise asymptotics is to treat $\Phi(x)$ as a harmonic sum and apply Mellin transform techniques (APPENDIX B: *Mellin Transform*, p. 674). The Mellin transform of $\Phi(x)$ is

$$\Phi^\star(s) := \int_0^\infty \Phi(x) x^{s-1}\, dx = \frac{\Gamma(s)}{1-2^s} \qquad \Re(s) \in (-1, 0).$$

The double pole of $\Phi^\star$ at 0 and the simple poles at $s = \frac{2ik\pi}{\log 2}$ are reflected by the asymptotic expansion:

(20)

$$\Phi(x) = \lg x + \frac{\gamma}{\log 2} + \frac{1}{2} + P(\lg x) + O(x^{-1}), \quad P(w) := -\frac{1}{\log 2} \sum_{k \in \mathbb{Z}\setminus\{0\}} \Gamma\left(\frac{2ik\pi}{\log 2}\right) e^{-2ik\pi w}.$$

The oscillating function $P(w)$ has amplitude of the order of 10^{-6}. (See [**153, 216, 265, 401**] for more on this topic.) The variance is similarly analysed. This concludes the proof of Proposition V.1. $\square$

What is striking is the existence of a family of distributions indexed by the fractional part of $\lg n$. This fact is naturally closely related to the presence of oscillating functions in moments of the random variable L. .... END OF EXAMPLE 2. $\square$

$\triangleright$ **7.** *Longest runs in Bernoulli sequences.* Consider an alphabet $\mathcal{A} = \{a_j\}$ with letter a_j independently chosen with probability $\{p_j\}$. The OGF of words where each letter is repeated at most k times derives from the construction of Smirnov words and is

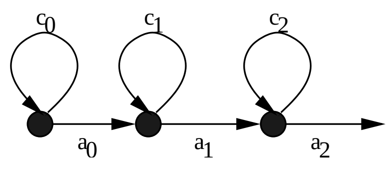
$$W^{[k]}(z) = \left(1 - \sum_i p_i z \frac{1-(p_i z)^k}{1-(p_i z)^{k+1}}\right)^{-1}.$$

Let $p_{\max}$ be the largest of the p_j. Then the expected length of the longest run of any letter is $\log n / \log p_{\max} + O(1)$, and precise quantitative information can be derived from the OGFs by methods akin to Example IV.9 (Smirnov words and Carlitz compositions, p. 249). $\triangleleft$

Walks of the pure birth type. The next two examples develop the analysis of walks in a special type of graphs. These examples serve two purposes: they illustrate further cases of modelling by means of regular specifications, and, at the same time, provide a bridge to the analysis of lattice paths in the next section.

EXAMPLE 3. *Walks of the pure-birth type.* Consider a walk on the nonnegative integers that starts at 0 and is only allowed either to stay at the same place or move by an increment of $+1$. Our goal is to enumerate the walks that start from 0 and reach point $m-1$ in n steps. A step from j to $j+1$ will be encoded by a letter a_j; a step from j to j will be encoded by c_j, in accordance with the following state diagram:

(21)



The language encoding all legal walks from state 0 to state m can be described by a regular expression,

$$\mathcal{H}_{0,m} = \text{SEQ}(c_0)a_0\, \text{SEQ}(c_1)a_1 \cdots \text{SEQ}(c_{m-1})a_{m-1}\, \text{SEQ}(c_m).$$

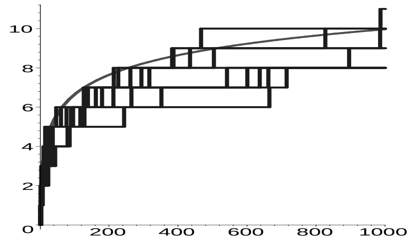


FIGURE 4. A simulation of 10 trajectories of the pure-birth process till $n = 1024$, with geometric probabilities corresponding to $q = 1/2$, compared to the curve $\log_2 x$.

Symbolicly using letters as variables, the corresponding ordinary multivariate generating function is then

$$H_{0,m}(\vec{a}, \vec{c}) = \frac{a_0 a_1 \cdots a_{m-1}}{(1-c_0)(1-c_1)\cdots(1-c_m)}.$$

Assume now that the steps are assigned weights, with α_j corresponding to a_j and γ_j to c_j. Weights of letters are extended multiplicatively to words in the usual way (cf Section III. 6.1, p. 178). In addition, upon taking $\gamma_j = 1 - \alpha_j$, one obtains a probabilistic weighting: the walker starts from position 0, and, if at j, at each clock tick, she either stays at the same place with probability $1 - \alpha_j$ or moves to the right with probability α_j. The OGF of such weighted walks then becomes

$$(22) \qquad H_{0,m}(z) = \frac{\alpha_0 \alpha_1 \cdots \alpha_{m-1} z^m}{(1-(1-\alpha_0)z)(1-(1-\alpha_1)z)\cdots(1-(1-\alpha_m)z)},$$

and $[z^n]H_{0,m}$ is the probability for the walker to be found at position m at (discrete) time n. This walk process can be alternatively interpreted as a (discrete-time) *pure-birth process*[3] in the usual sense of probability theory: There is a population of individuals and, at each discrete epoch, a new birth may take place, the probability of a birth being α_j when the population is of size j.

The form (22) readily lends itself to a partial fraction decomposition. Assume for simplicity that the α_j are all distinct. The poles of $H_{0,m}$ are at the points $(1 - \alpha_j)^{-1}$ and one finds as $z \to (1 - \alpha_j)^{-1}$:

$$H_{0,m}(z) \sim \frac{r_{j,m}}{1 - z(1-\alpha_j)} \quad \text{where} \quad r_{j,m} := \frac{\alpha_0 \alpha_1 \cdots \alpha_{m-1}}{\prod_{k\in[0,m],\ k\neq j}(\alpha_k - \alpha_j)}.$$

Thus, the probability of being in state m at time n is given by a sum:

$$(23) \qquad [z^n]H_{0,m}(z) = \sum_{j=0}^{m} r_{j,m}(1-\alpha_j)^n.$$

An especially interesting case of the pure-birth walk is when the quantities α_k are geometric: $\alpha_k = q^k$ for some q with $0 < q < 1$. In that case, the probability of being in state m after n transitions becomes (cf (23))

$$(24) \qquad \sum_{j=0}^{m} \frac{(-1)^j q^{\binom{j+1}{2}}}{(q)_j (q)_{m-j}} (1 - q^{m-j})^n, \quad (q)_j := (1-q)(1-q^2)\cdots(1-q^j).$$

This corresponds to a stochastic progression in a medium with exponentially increasing hardness or, equivalently, to the growth of a population whose size adversely affects fertility in an

[3]The theory of pure-birth processes is discussed under a calculational and non measure-theoretic angle in the book by Bharucha-Reid [**48**]. See also the *Course* by Karlin and Taylor [**252**] for a concrete presentation.

exponential manner. On intuitive grounds, we expect an evolution of the process to stay reasonably close to the curve $y = \log_{1/q} x$; see Figure 4 for a simulation confirming this fact, which can be justified by means of formula (24). This particular analysis is borrowed from [**143**], where it was initially developed in connection with the "approximate counting" algorithm to be studied next. .... END OF EXAMPLE 3. □

EXAMPLE 4. *Approximate Counting.* Assume you need to keep a counter that is able to record the number of certain events (say impulses) and should have the capability of keeping counts till a certain maximal value N. A standard information-theoretic argument (with ℓ bits, one can only keep track of 2^ℓ possibilities) implies that one needs $\lceil \log_2 N + 1 \rceil$ bits to perform the task—a standard binary counter will indeed do the job. However, in 1977, Robert Morris has proposed a way to maintain counters that only requires of the order of $\log \log N$ bits. What's the catch?

Morris' elegant idea consists in relaxing the constraint of exactness in the counting process and, by playing with probabilities, tolerate a small error on the counts obtained. Precisely, his solution maintains a random quantity Q which is initialized by $Q = 0$. Upon receiving an impulse, one updates Q according to the following simple procedure (with $q \in (0, 1)$ a design parameter):

procedure Update(Q);
    with probability q^{Q+1} do Q := $Q + 1$ (else keep Q unchanged).

When asked the number of impulses (number of times the update procedure was called) at any moment, simply use the following procedure to return an estimate:

procedure Answer(Q);
    output $\dfrac{q^{-Q} - 1}{q^{-1} - 1}$.

Let Q_n be the value of the random quantity Q after n executions of the update procedure and X_n the corresponding estimate output by the algorithm. It is easy to verify (by recurrence or by generating functions; see Note 8 below for higher moments) that

$$\mathbb{E}(q^{-Q_n}) = n(q^{-1} - 1) + 1, \qquad \text{so that} \quad \mathbb{E}(X_n) = n. \tag{25}$$

Thus the answer provided at any instant is an *unbiased estimator* (in a mean value sense) of the actual count n. On the other hand, the analysis of the geometric pure-birth process in the previous example applies. In particular, the exponential approximation $(1 - \alpha)^n \approx e^{-n\alpha}$ in conjunction with the basic formula (24) shows that for large n and m sufficiently near to $\log_{1/q} n$, one has (asymptotically) the *geometric-birth distribution*

$$\mathbb{P}\Big(Q_n = \log_{1/q} n + x\Big) = \sum_{j=0}^{\infty} \frac{(-1)^j q^{\binom{j+1}{2}}}{(q)_j (q)_\infty} \exp(-q^{x-j}) + o(1). \tag{26}$$

(We refer to [**143**] for details.) Such calculations imply that Q_n is with high probability (w.h.p.) close to $\log_{1/q} n$. Thus, if $n \leq N$, the value of Q_n will be w.h.p. bounded from above by $(1 + \epsilon) \log_{1/q} N$, with ϵ a small constant. But this means that the integer Q, which can itself be represented in binary, will only require

$$\log_2 \log n + O(1) \tag{27}$$

bits for storage, for fixed q.

A closer examination of the formulæ reveals that the accuracy of the estimate improves considerably when q becomes close to 1. The *standard error* is defined as $\frac{1}{n}\sqrt{\mathbb{V}(X_n)}$ and it measures (in a mean quadratic sense) the relative error to likely to be made. The variance of Q_n is, like the mean, determined by recurrence or generating functions, and one finds

$$\mathbb{V}(q^{-Q_n}) = \binom{n}{2}\frac{(1-q)^3}{q}, \qquad \frac{1}{n}\sqrt{\mathbb{V}(X_n)} \sim \sqrt{\frac{1-q}{q}} \tag{28}$$

(see also Note 8 below). This means that accuracy increases as q approaches 1 and, by suitably dimensioning q, one can make it asymptotically as small as desired. In summary, (25), (28), and (27) express the following property: *Approximate counting makes it possible to count till N using only about $\log\log N$ bits of storage, while achieving a standard error that is asymptotically a constant and can be set to any prescribed small value.* Morris' trick is now fully understood.

For instance, with $q = 2^{-1/16}$, it proves possible to count up to $2^{16} = 65536$ using only 8 bits (instead of 16), with an error likely not to exceed 20%. Naturally, there's not too much reason to appeal to the algorithm when a *single* counter needs to be managed. (Everybody can afford a few bits!) Approximate Counting turns out to be useful when a very large number of counts need to be kept *simultaneously*. It constitutes one of the early examples of a probabilistic algorithm in the extraction of information from large volumes of data, an area also known as *data mining*; see [**148**] for a review of connections with analytic combinatorics and references.

Functions akin to those of (26) also surface in other areas of probability theory. Guillemin, Robert, and Zwart [**219**] have detected them in processes that combine an additive increase and a multiplicative decrease (AIMD processes), in a context motivated by the adaptive transmission of "windows" of varying sizes in large communication networks (the TCP protocol of the internet). Biane, Bertoin, and Yor [**46**] encountered a function identical to (26) in their study of exponential functionals of Poisson processes. .... END OF EXAMPLE 4. □

▷ **8.** *Moments of q^{-Q_n}.* It is a perhaps surprising fact that any integral moment of q^{-Q_n} is a polynomial in n and q, like in (25), (28). To see it, define

$$\Phi(w) \equiv \Phi(w,\xi,q) := \sum_{m\ge 0} q^{m(m+1)/2}\frac{\xi^m w^m}{(1+\xi q)(1+\xi q^2)\cdots(1+\xi q^{m+1})}.$$

By (22), one has

$$\sum_{m\ge 0} H_{0,m} w^m = \frac{1}{1-z}\Phi\left(w; \frac{z}{1-z}, q\right).$$

On the other hand, Φ satisfies $\Phi(w) = 1 - q\xi(1-w)\Phi(qw)$, hence the q–identity,

$$\Phi(w) = \sum_{j\ge 0}(-q\xi)^j\left[(1-w)(1-qw)\cdots(1-q^{j-1}w)\right],$$

which resorts to q-calculus[4]. Thus $\Phi(q^{-r};\xi,q)$ is a polynomial for any $r\in\mathbb{Z}_{\ge 0}$, as the expansion terminates. See Prodinger's study [**355**] for connections with basic hypergeometric functions and Heine's transformation. ◁

[4]By q–calculus is roughly meant the collection of special function identities relating power series of the form $\sum a_n(q)z^n$, where $a_n(q)$ is a rational fraction whose degree is quadratic in n. See [**11**, Ch. 10] for basics and [**196**] for more advanced (q–hypergeometric) material.

Hidden patterns: Regular expression modelling and moments. We return here to the analysis of the number of occurrences of a pattern $\mathfrak{p}$ as a *subsequence* in a random text. The mean number of occurrences can be obtained by enumerating contexts of occurrences: in a sense we are then enumerating the language of all words by means of a dedicated regular expression where the ambiguity coefficient (the multiplicity) of a word is precisely equal to the number of occurrences of the pattern. This technique, which gives an easy access to expectations, also works for higher moments. It supplements the fact that there is no easy way to get a BGF in such cases.

EXAMPLE 5. *Occurrences of "hidden" patterns in texts.* Fix an alphabet $\mathcal{A} = \{a_1, \dots, a_r\}$ of cardinality r and assume a probability distribution on $\mathcal{A}$ to be given, with p_j the probability of letter a_j. We consider the Bernoulli model on $\mathcal{W} = \text{SEQ}(\mathcal{A})$, where the probability of a word is the product of the probabilities of its letters (cf Section III. 6.1, 178). A word $\mathfrak{p} = y_1 \cdots y_k$ called the pattern is fixed. The problem is to gather information on the random variable X representing the number of occurrences of $\mathfrak{p}$ in the set $\mathcal{W}_n$, where occurrences as a *"hidden pattern"*, i.e., as a *subsequence*, are counted (Example I.11, p. 51).

The generating function associated to $\mathcal{W}$ endowed with its probabilistic weighting is

$$W(z) = \frac{1}{1 - \sum p_j z} = \frac{1}{1-z}.$$

The regular specification

$$\mathcal{O} = \text{SEQ}(\mathcal{A}) y_1 \,\text{SEQ}(\mathcal{A}) \cdots \text{SEQ}(\mathcal{A}) y_{k-1} \,\text{SEQ}(\mathcal{A}) y_k \,\text{SEQ}(\mathcal{A}) \tag{29}$$

describes all *contexts of occurrences* of $\mathfrak{p}$ as a subsequence in all words. Graphically, this may be rendered as follows for a pattern of length 3, $\mathfrak{p} = y_1 y_2 y_3$:

$$\text{———}\boxed{y_1}\text{———}\boxed{y_2}\text{———}\boxed{y_3}\text{———} \tag{30}$$

There the boxes indicate distinguished positions where letters of the pattern appear and the horizontal lines represent arbitrary separating words ($\text{SEQ}(\mathcal{A})$). The corresponding OGF

$$O(z) = \frac{\pi(\mathfrak{p}) z^k}{(1-z)^{k+1}}, \qquad \pi(\mathfrak{p}) := p_{y_1} \cdots p_{y_{k-1}} p_{y_k} \tag{31}$$

counts elements of $\mathcal{W}$ with *multiplicity*[5], where the multiplicity coefficient $\lambda(w)$ of a word $w \in \mathcal{W}$ is precisely equal to the number of occurrences of $\mathfrak{p}$ as a subsequence in w:

$$O(z) \equiv \sum_{w \in \mathcal{A}^\star} \lambda(w) \pi(w) z^{|w|}.$$

There results that the mean number of hidden occurrences of $\mathfrak{p}$ in a random word of length n is

$$[z^n] O(z) = \pi(\mathfrak{p}) \binom{n}{k}, \tag{32}$$

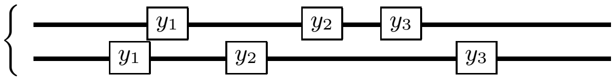
which is consistent with what a direct probabilistic reasoning would give.

We next proceed to determine the variance of X over $\mathcal{W}_n$. In order to do so, we need contexts in which *pairs* of occurrences appear. Let $\mathcal{Q}$ denote the set of all words in $\mathcal{W}$ with *two* occurrences (i.e., an ordered pair of occurrences) of $\mathfrak{p}$ as a subsequence being distinguished.

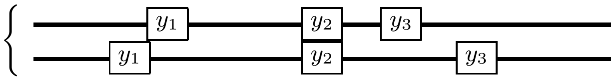
---

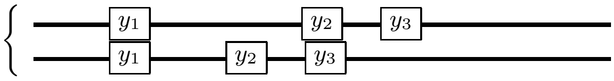
[5] In language-theoretic terms, we are making use of the regular expression $\mathcal{O} = \mathcal{A}^\star y_1 \mathcal{A}^\star \cdots y_{k-1} \mathcal{A}^\star y_k \mathcal{A}^\star$, that describes a subset of $\mathcal{A}^\star$ in an ambiguous manner and take into account the *ambiguity coefficients*.

Then clearly $[z^n]Q(z)$ represents $\mathbb{E}_{\mathcal{W}_n}[X^2]$. There are several cases to be considered. Graphically, a pair of occurrences may share no common position, like in what follows:

(33) 

But they may also have one or several overlapping positions, like in

(34) 

(35) 

(This last situation necessitates $y_2 = y_3$, typical patterns being abb and aaa.)

In the first case corresponding to (33), where there are no overlapping positions, the configurations of interest have OGF

$$Q^{[0]}(z) = \binom{2k}{k} \frac{\pi(\mathfrak{p})^2 z^{2k}}{(1-z)^{2k+1}}. \tag{36}$$

There, the binomial coefficient $\binom{2k}{k}$ counts the total number of ways of freely interleaving two copies of $\mathfrak{p}$; the quantity $\pi(\mathfrak{p})^2 z^{2k}$ takes into account the $2k$ distinct positions where the letters of the two copies appear; the factor $(1-z)^{-2k-1}$ corresponds to all the possible $2k+1$ fillings of the gaps between letters.

In the second case, let us start by considering pairs where exactly one position is overlapping, like in (34). Say this position corresponds to the rth and sth letters of $\mathfrak{p}$ (r and s may be unequal). Obviously, we need $y_r = y_s$ for this to be possible. The OGF of the configurations is now

$$\binom{r+s-2}{r-1}\binom{2k-r-s}{k-r} \frac{\pi(\mathfrak{p})^2 (p_{y_r})^{-1} z^{2k-1}}{(1-z)^{2k}}.$$

There, the first binomial coefficient $\binom{r+s-2}{r-1}$ counts the total number of ways of interleaving $y_1 \cdots y_{r-1}$ and $y_1 \cdots y_{s-1}$; the second binomial $\binom{2k-r-s}{k-r}$ is similarly associated to the interleavings of $y_{r+1} \cdots y_k$ and $y_{s+1} \cdots y_k$; the numerator takes into account the fact that $2k-1$ positions are now occupied by predetermined letters; finally the factor $(1-z)^{-2k}$ corresponds to all the $2k$ fillings of the gaps between letters. Summing over all possibilities for r, s gives the OGF of pairs with one overlapping position as

$$Q^{[1]}(z) = \left(\sum_{1 \le r,s \le k} \binom{r+s-2}{r-1}\binom{2k-r-s}{k-r} \frac{[\![y_r = y_s]\!]}{p_{y_r}} \right) \frac{\pi(\mathfrak{p})^2 z^{2k-1}}{(1-z)^{2k}}. \tag{37}$$

Similar arguments show that the OGF of pairs of occurrences with at least *two* shared positions (see, e.g., 35)) is of the form, with P a polynomial,

$$Q^{[\ge 2]}(z) = \frac{P(z)}{(1-z)^{2k-1}}, \tag{38}$$

for the essential reason that, in the finitely many remaining situations, there are at most $(2k-1)$ possible gaps.

We can now examine (36), (37), (38) in the light of singularities. The coefficient $[z^n]Q^{[0]}(z)$ is seen to cancel to first asymptotic order with the square of the mean as given in (32). The contribution of the coefficient $[z^n]Q^{[\ge 2]}(z)$ appears to be negligible as it is $O(n^{2k-2})$. The coefficient $[z^n]Q^{[1]}(z)$, which is $O(n^{2k-1})$, is seen to contribute to the asymptotic growth of the variance. In summary, after a trite calculation, we obtain:

PROPOSITION V.2. *The number X of occurrences of a hidden pattern $\mathfrak{p}$ in a random text of size n obeying a Bernoulli model satisfies*

$$\mathbb{E}_{\mathcal{W}_n}[X] = \pi(\mathfrak{p})\binom{n}{k} \sim \frac{\pi(\mathfrak{p})}{k!}n^k, \qquad \mathbb{V}_{\mathcal{W}_n}[X] = \frac{\pi(\mathfrak{p})^2\kappa(\mathfrak{p})^2}{(2k-1)!}n^{2k-1}\left(1+O(\frac{1}{n})\right),$$

where the "correlation coefficient" $\kappa(\mathfrak{p})^2$ is given by

$$\kappa(\mathfrak{p})^2 = \sum_{1\le r,s\le k}\binom{r+s-2}{r-1}\binom{2k-r-s}{k-r}\left(\frac{[\![y_r=y_s]\!]}{p_{y_r}}-1\right).$$

In particular, the distribution of X is concentrated around its mean.

This example is based on an article by Flajolet, Szpankowski, and Vallée [**182**]. There the authors show further that the asymptotic behaviour of moments of higher order can be worked out. By the moment convergence theorem, this calculation entails that *the distribution of X over $\mathcal{W}_n$ is asymptotically normal*. The method also extends to a much more general notion of "hidden" pattern, e.g., distances between letters of $\mathfrak{p}$ can be constrained in various ways so as to determine a valid occurrence in the text [**182**]. It also extends to the very general framework of dynamical sources [**61**], which include Markov models as a special case. The two references [**61, 182**] thus provide a set of analyses that interpolate between the two extreme notions of pattern occurrence—as a block of consecutive symbols or as a subsequence ("hidden pattern"). Such studies demonstrate that hidden patterns are with high probability bound to occur an extremely large number of times in a long enough text—this might cast some doubts on numerological interpretations encountered in various cultures: see in particular the critical discussion of the "Bible Codes" by McKay *et al.* in [**310**]. .... END OF EXAMPLE 5. □

▷ **9.** *Hidden patterns and shuffle relations.* To each pairs u, v of words over $\mathcal{A}$ associate the weighted-shuffle polynomial in the indeterminates $\mathcal{A}$ denoted by $\left(\!\binom{u}{v}\!\right)_t$ and defined by the properties

$$\begin{cases} \left(\!\binom{xu}{yv}\!\right)_t = x\left(\!\binom{u}{yv}\!\right)_t + y\left(\!\binom{xu}{v}\!\right)_t + t[\![x=y]\!]x\left(\!\binom{u}{v}\!\right)_t \\ \left(\!\binom{\mathbf{1}}{u}\!\right)_t = \left(\!\binom{u}{\mathbf{1}}\!\right)_t = u \end{cases}$$

where t is a parameter, x, y are elements of $\mathcal{A}$, and $\mathbf{1}$ is the empty word. Then the OGF of $Q(z)$ above is

$$Q(z) = \sigma\left[\left(\!\binom{\mathfrak{p}}{\mathfrak{p}}\!\right)_{(1-z)}\right]\frac{1}{(1-z)^{2k+1}},$$

where σ is the substitution $a_j \mapsto p_j z$. ◁

V. 3. Nested sequences, lattice paths, and continued fractions.

This section treats *nested sequence* constructions corresponding to a schema involving a cascade of sequences of the rough form SEQ ∘ SEQ ∘ ⋯ ∘ SEQ. Such a schema covers Dyck and Motzkin path, a particular type of Łukasiewicz paths already encountered in Section I. 5.3 (p. 69). Equipped with probabilistic weights, these paths appear as trajectories of birth-and-death processes (the case of pure-birth processes has already be dealt with in Example 3 above). They also have great descriptive power since, once endowed with integer weights, they can encode a large variety of combinatorial classes, including trees, permutations, set partitions, and surjections.

Since a combinatorial sequence translates into a quasi-inverse, $Q(f) = (1-f)^{-1}$, a class described by nested sequences has its generating function expressed by a cascade of fractions, that is, a *continued fraction*[6]. Analytically, these GFs have at most two dominant poles (the Dyck case) or a single pole (the Motzkin case) on their disc of convergence, so that the implementation of the process underlying Theorem V.1 is easy: we encounter a pure polynomial form of the simplest type that describes all counting sequences of interest. The profile of a nested sequence can also be easily characterized.

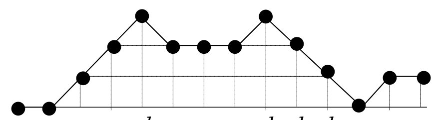
This section starts with a statement of the "Continued Fraction Theorem" taken from an old study of Flajolet [**139**], which provides the general set up for the rest of the section. It then proceeds with the general analytic treatment of nested sequences. A number of examples from various areas of discrete mathematics are then detailed. Some of these make use of structures that are described as infinitely nested sequences, that is, infinite continued fractions, to which the finite theory often extends—the analysis of coin fountains below is typical.

V. 3.1. Combinatorial aspects. We discuss here a special type of lattice paths connecting points of the discrete Cartesian plane $\mathbb{Z} \times \mathbb{Z}$.

DEFINITION V.3 (Lattice path). *A* Motzkin path $\upsilon = (U_0, U_1, \ldots, U_n)$ *is a sequence of points in the discrete quarter plane* $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ *such that* $U_j = (j, y_j)$*, and the jump condition* $|y_{j+1} - y_j| \leq 1$ *is satisfied. An edge* $\langle U_j, U_{j+1} \rangle$ *is called an* ascent *if* $y_{j+1} - y_j = +1$*, a* descent *if* $y_{j+1} - y_j = -1$*, and a* level step *if* $y_{j+1} - y_j = 0$*. A path that has no level steps is called a* Dyck path.

The quantity n *is the* length *of the path,* $\operatorname{ini}(\upsilon) := y_0$ *is the* initial altitude, $\operatorname{fin}(\upsilon) := y_n$ *is the* final altitude. *A path is called an* excursion *if both its initial and final altitudes are zero. The extremal quantities* $\sup\{\upsilon\} := \max_j y_j$ *and* $\inf\{\upsilon\} := \min_j y_j$ *are called the* height *and* depth *of the path.*

A path can always be encoded by a word with a, b, c representing ascents, descents, and level steps, respectively. What we call the *standard encoding* is such a word in which each step a, b, c is (redundantly) subscripted by the value of the y-coordinate of its initial point. For instance,

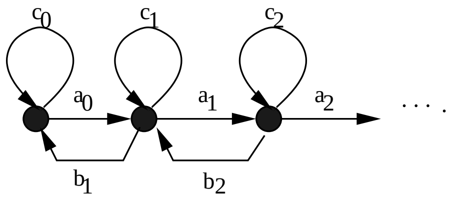


$$w = c_0\, a_0\, a_1\, a_2\, b_3\, c_2\, c_2\, a_2\, b_3\, b_2\, b_1\, a_0\, c_1$$

encodes a path that connects the initial point $(0, 0)$ to the point $(13, 1)$. Such a path can also be regarded as the evolution in discrete time of a walk over the integer line

[6]Characteristically, the German term for "continued fraction", is "*Kettenbruch*", literally "*chain-fraction*".

with jumps restricted to $\{-1, 0, +1\}$, or equivalently as a path in the graph:



Lattice paths can also be interpreted as trajectories of birth-and-death processes, where a population can evolve at any discrete time by a birth or a death. (Compare with the pure-birth case in (21) above.)

As a preparation for later developments, let us examine the description of the class written $\mathcal{H}_{0,0}^{[<1]}$ of Motzkin excursions of height < 1. We have

$$\mathcal{H}_{0,0}^{[<1]} \cong \mathrm{SEQ}(c_0) \qquad \Longrightarrow H_{0,0}^{[<1]}(z) = \frac{1}{1-c_0}.$$

The class of excursions of height < 2 is obtained from there by a substitution

$$c_0 \ \mapsto \ c_0 + a_0\,\mathrm{SEQ}(c_1)b_1,$$

to the effect that

$$\begin{aligned} &\mathcal{H}^{[<2]} \cong \mathrm{SEQ}\left(c_0 + a_0\,\mathrm{SEQ}(c_1)b_1\right) \\ &\quad \Longrightarrow \qquad H^{[<2]}(z) = \cfrac{1}{1-c_0-\cfrac{a_0b_1}{1-c_1}} = \frac{1-c_1}{1-c_0-c_1+c_0c_1-a_0b_1}. \end{aligned}$$

Iteration of this simple mechanism lies at the heart of the calculations performed below. Clearly, generating functions written in this way are nothing but a concise description of usual counting generating functions: for instance if individual weights[7] $\alpha_j, \beta_j, \gamma_j$ are assigned to the letters a_j, b_j, c_j respectively, then the OGF of multiplicatively weighted paths with z marking length is obtained by setting

$$a_j = \alpha_j z, \qquad b_j = \beta_j z, \qquad c_j = \gamma_j. \tag{39}$$

The general class of paths of interest in this subsection is defined by arbitrary combinations of *flooring* (by m) *ceiling* (by h), as well as fixing initial (k) and final (l) altitudes. Accordingly, we define the following subclasses of the class $\mathcal{H}$ of all Motzkin paths:

$$\mathcal{H}_{k,l}^{[m\le\bullet<h]} \quad := \quad \{w \in \mathcal{H} : \mathrm{ini}(w) = k,\ \mathrm{fin}(w) = l,\ m \le \inf\{w\},\ \sup\{w\} < h\}.$$

We shall also need the specializations,

$$\mathcal{H}_{k,l}^{[<h]} = \mathcal{H}_{k,l}^{[0\le\bullet<h]}, \quad \mathcal{H}_{k,l}^{[\ge m]} = \mathcal{H}_{k,l}^{[m\le\bullet<\infty]}, \quad \mathcal{H}_{k,l} = \mathcal{H}_{k,l}^{[0\le\bullet<\infty]}.$$

(Thus, the supercript indicates the condition that is to be satisfied by *all abscissae* of vertices of the path.) Three simple combinatorial decompositions of paths (Figure 5) then suffice to derive all the basic formulæ.

[7]Throughout this chapter, all weights are assumed to be *nonnegative*.

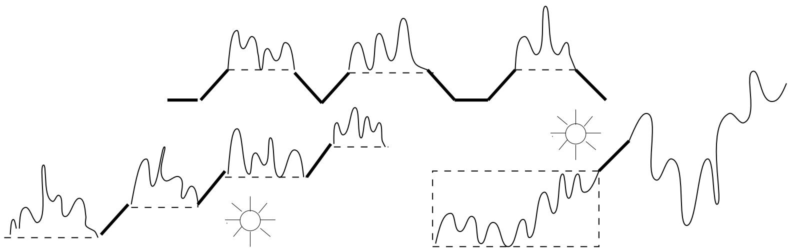


FIGURE 5. The three major decompositions of lattice paths: the arch decomposition (top), the last passages decomposition (bottom left), and the first passage decomposition (bottom right).

— *Arch decomposition*: An excursion from and to level 0 consists of a sequence of “arches”, each made of either a c_0 or a $a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1$, so that

$$\mathcal{H}_{0,0} = \text{SEQ}\left(c_0 \cup a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1\right), \tag{40}$$

which relativizes to height $< h$.

— *Last passages decomposition.* Recording the times at which each level $0, \ldots, k$ is last traversed gives

$$\mathcal{H}_{0,k} = \mathcal{H}_{0,0}^{[\geq 0]}a_0\mathcal{H}_{1,1}^{[\geq 1]}a_1 \cdots a_{k-1}\mathcal{H}_{k,k}^{[\geq k]} \tag{41}$$

— *First passage decomposition.* The quantities $H_{k,l}$ with $k \leq l$ are implicitly determined by the first passage through k in a path connecting level 0 to l, so that

$$\mathcal{H}_{0,l} \quad = \quad \mathcal{H}_{0,k-1}^{[<k]}a_{k-1}\mathcal{H}_{k,l} \quad (k \leq l), \tag{42}$$

(A dual decomposition holds when $k \geq l$.)

The basic results express the generating functions in terms of a fundamental continued fraction and its associated convergent polynomials. They involve the “numerator” and “denominator” polynomials, denoted by P_h and Q_h that are defined as solutions to the second order (or “three-term”) recurrence equation

$$Y_{h+1} = (1 - c_h)Y_h - a_{h-1}b_hY_{h-1}, \ h \geq 0, \tag{43}$$

together with the initial conditions $(P_{-1}, Q_{-1}) = (-1, 0)$, $(P_0, Q_0) = (0, 1)$, and with the convention $a_{-1}b_0 = 1$. In other words, setting $C_j = 1 - c_j$ and $A_j = a_{j-1}b_j$, we have:

(44)

$$\begin{array}{llll} P_0 = 0, & P_1 = 1, & P_2 = C_1, & P_3 = C_1C_2 - A_2 \\ Q_0 = 1, & Q_1 = C_0, & Q_2 = C_0C_1 - A_1, & Q_3 = C_0C_1C_2 - C_2A_1 - C_0A_2. \end{array}$$

These polynomials are also known as continuant polynomials [**270, 430**].

$\triangleright$ **10.** *Combinatorics of continuant polynomials.* The polynomial Q_h is obtained by the following process: start with the product $\Pi := C_0C_1 \cdots C_{h-1}$; then cross out in all possible ways pairs of adjacent elements $C_{j-1}C_j$, replacing each such crossed pair by $-A_j$. For instance, Q_4 is obtained as

$$C_0C_1C_2C_3 + \overbrace{\cancel{C_0C_1}}^{-A_1}C_2C_3 + C_0\overbrace{\cancel{C_1C_2}}^{-A_2}C_3 + C_0C_1\overbrace{\cancel{C_2C_3}}^{-A_3} + \overbrace{\cancel{C_0C_1}}^{-A_1}\overbrace{\cancel{C_2C_3}}^{-A_3}.$$

The polynomials P_h are obtained similarly after a shift of indices. (These observations are due to Euler; see [**212**, §6.7].) $\lhd$

PROPOSITION V.3 (Continued Fraction Theorem [**139**]). (i) *The generating function $H_{0,0}$ of all excursions is represented by the fundamental continued fraction:*

$$H_{0,0} = \cfrac{1}{1-c_0-\cfrac{a_0b_1}{1-c_1-\cfrac{a_1b_2}{1-c_2-\cfrac{a_2b_3}{\ddots}}}}. \tag{45}$$

(ii) *The generating function of ceiled excursion $H_{0,0}^{[<h]}$ is given by a convergent of the fundamental continued fraction* (45)*, with P_h, Q_h as in Equation* (43)*:*

$$H_{0,0}^{[<h]} = \cfrac{1}{1-c_0-\cfrac{a_0b_1}{1-c_1-\cfrac{a_1b_2}{\cfrac{\ddots}{1-c_{h-1}}}}} = \frac{P_h}{Q_h}. \tag{46}$$

(iii) *The generating function of floored excursions is given by a truncation of the fundamental fraction:*

$$H_{h,h}^{[\geq h]} = \cfrac{1}{1-c_h-\cfrac{a_hb_{h+1}}{1-c_{h+1}-\cfrac{a_{h+1}b_{h+2}}{\ddots}}} \tag{47}$$

$$= \frac{1}{a_{h-1}b_h}\,\frac{Q_hH_{0,0}-P_h}{Q_{h-1}H_{0,0}-P_{h-1}}, \tag{48}$$

PROOF. Repeated use of the arch decomposition (40) provides a form of $H_{0,0}^{[<h]}$ with nested quasi-inverses $(1-f)^{-1}$ that is the finite fraction representation (46), for instance,

$$\mathcal{H}_{00}^{[<1]} \cong \text{SEQ}\{c_0\}, \quad \mathcal{H}_{00}^{[<2]} \cong \text{SEQ}\{c_0 + a_0\,\text{SEQ}\{c_1\}b_1\},$$
$$\mathcal{H}_{00}^{[<3]} \cong \text{SEQ}\{c_0 + a_0\,\text{SEQ}\{c_1 + a_0\,\text{SEQ}\{c_2\}b_2\}b_1\}.$$

The continued fraction representation for basic paths without height constraints (namely $H_{0,0}$) is then obtained by letting $h \to \infty$ in (46). Finally, the continued fraction form (47) for ceiled excursions is nothing but the fundamental form (45), when the indices are shifted. The three continued fraction expansions (45), (46), (47) are hence established.

Finding explicit expressions for the fractions $H_{0,0}^{[<h]}$ and $H_{h,h}^{[\geq h]}$ next requires determining the polynomials that appear in the convergents of the basic fraction (45).

By definition, the convergent polynomials P_h and Q_h are the numerator and denominator of the fraction $H_{0,0}^{[<h]}$. For the computation of $H_{0,0}^{[<h]}$ and P_h, Q_h, one classically introduces the linear fractional transformations

$$g_j(y) = \frac{1}{1 - c_j - a_j b_{j+1} y},$$

so that

$$H_{0,0}^{[<h]} = g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(0) \text{ and } H_{0,0} = g_0 \circ g_1 \circ g_2 \circ \cdots \ ,. \tag{49}$$

Now, linear fractional transformations are representable by 2×2-matrices

$$\frac{ay+b}{cy+d} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{50}$$

in such a way that composition corresponds to matrix product. By induction on the compositions that build up $H_{0,0}^{[<h]}$, there follows the equality

$$g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(y) = \frac{P_h - P_{h-1} a_{h-1} b_h y}{Q_h - Q_{h-1} a_{h-1} b_h y}, \tag{51}$$

where P_h and Q_h are seen to satisfy the recurrence (43). Setting $y = 0$ in (51) proves (46).

Finally, $H_{h,h}^{[\geq h]}$ is determined implicitly as the root y of the equation $g_0 \circ \cdots \circ g_{h-1}(y) = H_{0,0}$, an equation that, when solved using (51), yields the form (48). $\square$

A large number of generating functions can be derived by similar techniques. We refer to the article [**139**], where this theory was first systematically developed and to the exposition given in [**208**, Chapter 5]. Our presentation also draws upon [**157**] where the theory was put to use in order to develop a formal algebraic theory of general birth-and-death processes in continuous time.

▷ **11.** *Transitions and crossings.* The lattice paths $\mathcal{H}_{0,l}$ corresponding to the transitions from altitude 0 to l and $\mathcal{H}_{k,0}$ (from k to 0) have OGFs

$$H_{0,l} = \frac{1}{\mathfrak{B}_l}\left(Q_l H_{0,0} - P_l\right), \qquad H_{k,0} = \frac{1}{\mathfrak{A}_k}(Q_k H_{0,0} - P_k).$$

The crossings $\mathcal{H}_{0,h-1}^{[<h]}$ and $\mathcal{H}_{h-1,0}^{[<h]}$ have OGFs,

$$H_{0,h-1}^{[<h]} = \frac{\mathfrak{A}_{h-1}}{Q_h}, \qquad H_{h-1,0}^{[<h]} = \frac{\mathfrak{B}_{h-1}}{Q_h}.$$

(Abbreviations used here are: $\mathfrak{A}_m = a_0 \cdots a_{m-1}$, $\mathfrak{B}_m = b_1 \cdots b_m$.) These extensions provide combinatorial interpretations for fractions of the form $1/Q$. They result from the basic decompositions combined with Proposition V.3; see [**139, 157**] for details. ◁

▷ **12.** *Denominator polynomials and orthogonality.* Let $H_n = [z^n]H_{0,0}(z)$ represent the number of all excursions of length n equipped with *nonnegative* weights. Define a linear functional $\mathcal{L}$ on the space $\mathbb{C}(z)$ of polynomials by $\mathcal{L}[z^n] = H_n$. Introduce the reciprocal polynomials: $\overline{Q}_h(z) = z^h Q(1/z)$. The fact deducible from Note 11 that $Q_l H_{0,0} - P_l = O(z^l)$ corresponds to the property $\mathcal{L}[z^j \overline{Q}_l) = 0$ for all $0 \leq j < l$. In other words, the polynomials $\overline{Q}_l$ are orthogonal with respect to the special scalar product $\langle f, g \rangle := \mathcal{L}[fg]$. (Historically, the theory of orthogonal polynomials evolved from the theory of continued fractions before living a life of its own; see [**79, 240, 400**] for its many facets.) ◁

▷ **13.** *Discrete time birth-and-death processes.* Assume that, at discrete times $n = 0, 1, 2, \ldots$, a population of size j can grow by one element [a birth] with probability α_j, decrease by one element [a death] with probability β_j, and stay the same with probability $\gamma_j = 1 - \alpha_j - \beta_j$. Let ω_n be the probability that an initially empty population is again empty at time n. Then the GF of the sequence (ω_n) is

$$\sum_{n\geq 0} \omega_n z^n = \cfrac{1}{1 - \gamma_0 z - \cfrac{\alpha_0\beta_1 z^2}{1 - \gamma_1 z - \cfrac{\alpha_1\beta_2 z^2}{\cdots}}}.$$

This result was found by I. J. Good in 1958: see [**207**]. ◁

▷ **14.** *Continuous time birth-and-death processes.* Consider a continuous time birth-and-death process, where a transition from state j to $j + 1$ takes place according to an exponential distribution of rate λ_j and a transition from j to $j - 1$ has rate μ_j. Let $\varpi(t)$ be the probability to be in state 0 at time t starting from state 0 at time 0. One has

$$\int_0^\infty e^{-st}\varpi(t)\,dt = \cfrac{1}{s + \lambda_0 - \cfrac{\lambda_0\mu_1}{s + \lambda_1 + \mu_1 - \cfrac{\lambda_1\mu_2}{\cdots}}} = \cfrac{1}{s + \cfrac{\lambda_0}{1 + \cfrac{\mu_1}{s + \cfrac{\lambda_1}{\cdots}}}}.$$

Thus, continued fractions and orthogonal polynomials may be used to analyse birth-and-death processes. (This fact was originally discovered by Karlin and McGregor [**251**], with later additions due to Jones and Magnus [**247**]. See [**157**] for a systematic discussion in relation to combinatorial theory.) ◁

V. 3.2. Analytic aspects. We now consider the general asymptotic properties of lattice paths of height bounded from above by a fixed integer $h \geq 1$. Letters denoting elementary steps are weighted, as previously indicated, with

$$a_j = \alpha_j z, \qquad b_j = \beta_j z, \qquad c_j = \gamma_j z,$$

the weights being invariably nonnegative. We shall limit the discussion to excursions, which are often the most interesting objects from the combinatorial point of view.

As a preamble, in the Dyck case, where all γ_j are 0 (level steps are disallowed), the GF $H^{[<h]}$ is a function of z^2 only, since it takes an even number of steps to return to altitude 0 when starting from altitude 0. In such a case, we shall systematically assume that, when considering $[z^n]H^{[<h]}$, the index $n = 2\nu$ is *even*. In order to avoid trivialities, we also assume that none of the coefficients attached to ascents and descents are 0.

THEOREM V.3 (Asymptotics of nested sequences). *Consider the class $\mathcal{H}_{0,0}^{[<h]}$ of weighted Motzkin excursions of height $< h$. Their number satisfies a pure exponential-polynomial formula,*

$$H_{0,0,n}^{[<h]} = cB^n + O(C^n),$$

where $B > 0$ and $0 \leq C < B$. In the Dyck case, it is assumed furthermore that $n \equiv 0 \pmod 2$.

PROOF. The proof[8] proceeds by induction according to the depth of nesting of the sequence constructions. Write

$$f_j(z) := H^{[<h]}_{h-j-1,h-j-1}(z),$$

and let ρ_j denote the dominant singularity of f_j that is positive (existence is guaranteed by Pringsheim's Theorem).

For ease of discussion, we first examine the case where all γ_j are nonzero. The function $f_0(z)$ is

$$f_0(z) = \frac{1}{1-\gamma_{h-1}z},$$

and one has $\rho_0 = 1/\gamma_{h-1}$. The function f_1 is given by

$$f_1(z) = \frac{1}{1-\gamma_{h-2}z - \alpha_{h-2}\beta_{h-1}z^2 f_0(z)}.$$

The quantity $\gamma_{h-2}z + \alpha_{h-2}\beta_{h-1}z^2 f_0(z)$ in its denominator increases continuously from 0 to $+\infty$ as z increases from 0 to ρ_0; consequently, it crosses the value 1 at some point which must be ρ_1. In particular, one must have $\rho_1 < \rho_0$. Our assumption that all the γ_j are nonzero implies the absence of periodicities, so that ρ_1 is the unique dominant singularity. The argument can be repeated, implying that the sequence of radii is decreasing $\rho_0 > \rho_1 > \rho_2 > \cdots$, the corresponding poles are all simple, and they are uniquely dominating. The statement is thus established in the case that all the γ_j are nonzero.

Dually, in the Dyck case where all the γ_j are zero, one can reason in a similar manner, operating with the collection of "condensed" series $f_j(\sqrt{z})$, which are seen to have a unique dominant singularity. This implies that $f_j(z)$ itself has exactly two dominant singularities, namely ρ_h and $-\rho_h$, both being simple poles.

In the mixed case, the f_j are initially of the Dyck type, till a certain $\gamma_{h-1-j_0} \neq 0$ is encountered. In that case the function f_{j_0} is aperiodic (its span in the sense of Chapter IV is 1). The reasoning then continues like in the Motzkin case, with all the subsequent f_j (for $j \geq j_0$) including $f_{h-1}(z) \equiv H^{[<h]}_{0,0}(z)$ having a unique dominant singularity. $\square$

Similar devices yield a characterization of the profile of a random path, that is, the number of times a given step appears in a random excursion.

THEOREM V.4 (Profile of nested sequences). *Let X_n be the random variable representing the number of times a given step (of type a_j, b_j, or c_j) with nonzero weight appears in a random excursion of length n and height $< h$. The moments of X_n satisfy*

$$\mathbb{E}(X_n) = c_1 n + d_1 + O(D^n), \qquad \mathbb{V}(X_n) = c_2 n + d_2 + O(D^n),$$

for constants c_1, c_2, d_1, d_2, D, with $c_1, c_2 > 0$ and $0 \leq D < 1$. In particular the distribution of X_n is concentrated.

[8] The present discussion is also related to the analysis of the supercritical sequence schema in the next section.

PROOF. Introduce an auxiliary variable u with u marking the number of designated steps, and form the corresponding BGF $H(z,u)$. We only discuss the expectation. The function H is a linear fractional transformation in u of the form

$$H(z,u) = A(z) + \frac{1}{C(z) + uD(z)}.$$

(The coefficients A, B, C are a priori in $\mathbb{C}(z)$; they are in fact computable from Proposition V.3.) Then, one has

$$\frac{\partial}{\partial u} H(z,u)\bigg|_{u=1} = -\frac{D(z)}{(C(z)+D(z))^2}.$$

This function resembles $H(z,1)^2$. An application of the chain rule permits us to verify that indeed

$$\frac{\partial}{\partial u} H(z,u)\bigg|_{u=1} = E(z) H(z,1)^2,$$

where $E(z)$ is analytic in disc larger than the disc of analyticity of $H(z,1)$. The analysis of the dominant double pole then yields the result. (The determination of the second moment follows along similar lines, though the computations become more intricate.) $\square$

$\triangleright$ **15.** *All poles are real.* Assume again $\alpha_j \beta_{j+1} > 0$ and $\gamma_j \geq 0$. By Note 12, the denominator polynomials Q_h are reciprocals of a family of polynomials $\overline{Q}_h$ that are formally orthogonal with respect to a scalar product. Thus the zeros of any of the $\overline{Q}_h$ are all real, and so are the zeros of Q_h. Consequently: *The poles of the OGF of ceiled excursions* $H_{0,0}^{[<h]}$ *are all real*. (See for instance [**400**, §3.3] for the basic argument.) $\triangleleft$

V. 3.3. Applications. Lattice paths corresponding to nested sequences have a quite a wide range of descriptive power, especially when weights are allowed. We illustrate this fact by three types of examples.

— Example 6 provides a complete analysis of height in Dyck paths and general plane rooted trees, as regards moments as well as distribution. This is the simplest case of a continued fraction with constant coefficients attached to the OGF of Catalan numbers and Fibonacci-Chebyshev polynomials.
— Example 7 discusses coin fountains. There, we are dealing with an infinite continued fraction to which the techniques of the previous subsection can be extended. The developments also takes us close to the realm of q-calculus and to the analysis of alcohols seen in Chapter IV.
— Example 8 constitutes a typical application of the possibility of encoding combinatorial structures—here we examine interconnection networks—by means of lattice path weighted by integers. The enumeration involves Hermite polynomials. Other examples related to set partitions and permutations are described in the accompanying notes.

EXAMPLE 6. *Height of Dyck paths and plane rooted trees.* In order to count lattice paths of the Dyck (D) or Motzkin (M) type, it suffices to effect one of the substitutions,

$$\sigma_M : \ a_j \mapsto z, \ b_j \mapsto z, c_j \mapsto z; \qquad \sigma_D : \ a_j \mapsto z, \ b_j \mapsto z, c_j \mapsto 0.$$

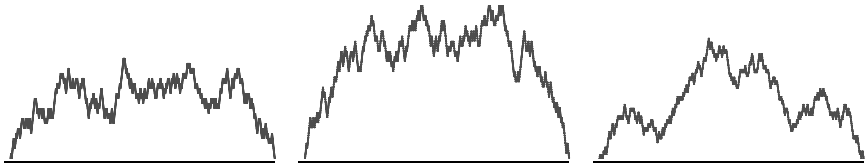


FIGURE 6. Three random Dyck paths of length $2n = 500$ have heights resp. 20, 31, 24: the distribution is spread, see Proposition V.4.

We henceforth restrict attention to the case of Dyck paths. See Figure 6 for three simulations suggesting that the distribution of height is somewhat spread. Given the parenthesis system representation (Note I.45, p.73), the height of a Dyck path automatically translates into as height of the corresponding plane rooted tree.

The continued fraction expressing $H_{0,0}$ results immediately from Proposition V.3 and is in this case periodic (here, in the sense that its stages are all alike), so that it represents a quadratic function,

$$H_{0,0}(z) = \cfrac{1}{1 - \cfrac{z^2}{1 - \cfrac{z^2}{1 - \ddots}}} = \frac{1}{2z^2}\left(1 - \sqrt{1-4z^2}\right),$$

since $H_{0,0}$ satisfies $y = (1-z^2 y)^{-1}$. The families of polynomials P_h, Q_h are in this case determined by a recurrence with constant coefficients. Define classically the Fibonacci polynomials by the recurrence

$$F_{h+2}(z) = F_{h+1}(z) - zF_h(z), \quad F_0(z) = 0, \quad F_1(z) = 1. \tag{52}$$

One finds $Q_h = F_{h+1}(z^2)$ and $P_h = F_h(z^2)$. (The Fibonacci polynomials are reciprocals of Chebyshev polynomials; see Note 16.) By Proposition V.3, the GF of paths of height $< h$ is then

$$H_{00}^{[<h]}(z) = \frac{F_h(z^2)}{F_{h+1}(z^2)}.$$

(We get more and, for instance, the number of ways of crossing a strip of width $h-1$ is $H_{0,h-1}^{[<h]}(z) = z^{h-1}/F_{h+1}(z^2)$.) Note that the polynomials have an explicit form,

$$F_h(z) = \sum_{k=0}^{\lfloor (h-1)/2 \rfloor} \binom{h-1-k}{k} (-z)^k,$$

as follows from the generating function expression: $\sum_h F_h(z) y^h = y/(1-y+zy^2)$.

The equivalence between Dyck paths and (general) plane tree traversals discussed in Chapter I implies that trees of height at most h and size $n+1$ are equinumerous with Dyck paths of length $2n$ and height at most h. Set for convenience

$$G^{[h]}(z) = zH_{00}^{[<h+1]}(z^{1/2}) = z\frac{F_{h+1}(z)}{F_{h+2}(z)},$$

which is precisely the OGF of general plane trees having height $\leq h$. (This is otherwise in agreement with the continued fraction forms obtained directly in Chapter III: cf (52), p. 184 and (75), p. 205.) It is possible to go much further as first shown by De Bruijn, Knuth, and Rice in a landmark paper [**95**], which also constitutes the historic application of Mellin transforms in analytic combinatorics. (We refer to this paper for historical context and references.)

First, solving the linear recurrence (52) with z treated as a parameter yields the alternative closed form expression

$$F_h(z) = \frac{G^h - \overline{G}^h}{G - \overline{G}}, \qquad G = \frac{1-\sqrt{1-4z}}{2}, \quad \overline{G} = \frac{1+\sqrt{1-4z}}{2}. \tag{53}$$

There, $G(z)$ is the OGF of all trees, and an equivalent form of $G^{[h]}$ is provided by

$$G - G^{[h-2]} = \sqrt{1-4z}\frac{u^h}{1-u^h}, \quad \text{where} \quad u = \frac{1-\sqrt{1-4z}}{1+\sqrt{1-4z}} = \frac{G^2}{z}, \tag{54}$$

as is easily verified. Thus $G^{[h]}$ can be expressed in terms of $G(z)$ and z:

$$G - G^{[h-2]} = \sqrt{1-4z}\sum_{j\geq 1} z^{-jh} G(z)^{2jh}.$$

The Lagrange-Bürmann inversion theorem then gives after a simple calculation

$$G_{n+1} - G_{n+1}^{[h-2]} = \sum_{j\geq 1} \Delta^2 \binom{2n}{n-jh}, \tag{55}$$

where

$$\Delta^2 \binom{2n}{n-m} := \binom{2n}{n+1-m} - 2\binom{2n}{n-m} + \binom{2n}{n-1-m}.$$

Consequently, the number of trees of height $\geq h-1$ admits a closed form: it is a "sampled" sum, by steps of h, of the $2n$th line of Pascal's triangle (upon taking second order differences).

The relation (55) leads easily to the asymptotic distribution of height in random trees of size n. Stirling's formula yields the Gaussian approximation of binomial numbers: for $k = o(n^{3/4})$ and with $w = k/\sqrt{n}$, one finds

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \sim e^{-w^2}\left(1 - \frac{w^4-3w^2}{6n} + \frac{5w^8 - 54w^6 + 135w^4 - 60w^2}{360n^2} + \cdots\right). \tag{56}$$

The use of the Gaussian approximation (56) inside the exact formula (55) then implies: *The probability that a tree of size* $n+1$ *has height at least* $h-1$ *satisfies uniformly for* $h \in [\alpha\sqrt{n}, \beta\sqrt{n}]$ *(for any* α, β *such that* $0 < \alpha < \beta < \infty$*) the estimate*

$$\frac{G_{n+1} - G_{n+1}^{[h-2]}}{G_{n+1}} = \Theta\left(\frac{h}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right), \qquad \Theta(x) := \sum_{j\geq 1} e^{-j^2x^2}(4j^2x^2 - 2). \tag{57}$$

The function $\Theta(x)$ is a "theta function" which classically arises in the theory of elliptic functions [**433**]. Since binomial coefficients decay fast away from the center, simple bounds also show that the probability of height to be at least $n^{1/2+\epsilon}$ decays like $\exp(-n^{2\epsilon})$, hence is exponentially small. Note also that the probability distribution of height H itself admits of an exact expression obtained by differencing (55), which is reflected asymptotically by differentiation of the estimate of (57):

$$\mathbb{P}_{\mathcal{G}_{n+1}}\left[H = \lfloor x\sqrt{n}\rfloor\right] = -\frac{1}{\sqrt{n}}\Theta'(x) + O\left(\frac{1}{n}\right), \qquad \Theta'(x) := \sum_{j\geq 1} e^{-j^2x^2}(12j^2x - 8j^4x^3). \tag{58}$$

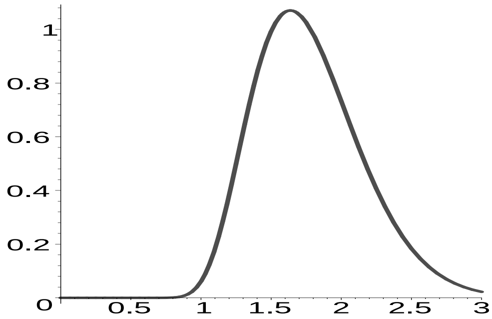


FIGURE 7. The limit density of the distribution of height $-\Theta'(x)$.

The forms (57) and (58) also give access to moments of the distribution of height. We find

$$\mathbb{E}_{\mathcal{G}_{n+1}}\left[H^r\right] \sim \frac{1}{\sqrt{n}} S_r\left(\frac{1}{\sqrt{n}}\right), \qquad \text{where} \quad S_r(y) := -\sum_{h\geq 1} h^r \Theta'(hy).$$

The quantity $y^{r+1}S_r(y)$ is a Riemann sum relative to the function $-x^r\Theta'(x)$, and the step $y = n^{-1/2}$ decreases to 0 as $n \to \infty$. Approximating the sum by the integral, one gets:

$$\mathbb{E}_{\mathcal{G}_{n+1}}\left[H^r\right] \sim n^{r/2}\mu_r \qquad \text{where} \quad \mu_r := -\int_0^\infty x^r \Theta'(x)\, dx.$$

The integral giving μ_r is a Mellin transform in disguise (set $s = r+1$) to which the treatment of harmonic sums applies. We then get upon replacing $n+1$ to n:

PROPOSITION V.4. *The expected height of a random plane rooted tree comprising $n+1$ nodes is*

$$\sqrt{\pi n} - \frac{3}{2} + o(1). \tag{59}$$

More generally, the moment of order r of height is asymptotic to

$$\mu_r n^{r/2} \qquad \textit{where} \quad \mu_r = r(r-1)\Gamma(r/2)\zeta(r). \tag{60}$$

The random variable $H/\sqrt{n}$ obeys asymptotically a Theta distribution, in the sense of both the "central" estimate (57) and the "local" estimate (58). The same asymptotic estimates hold for height of Dyck paths having length $2n$.

The improved estimate of the mean (59) is from [**95**]. The general form of moments in (59) is in fact valid for any real r (not just integers). An alternative formula for the Theta function appears in the Note below. Figure 7 plots the limit density $-\Theta'(x)$. . END OF EXAMPLE 6. □

▷ **16.** *Height and Fibonacci–Chebyshev polynomials.* The reciprocal polynomials $\overline{F}_h(z) = F_{h-1}(z) = z^{h-1}F_h(1/z^2)$ are related to the classical Chebyshev polynomials by $\overline{F}_h(2z) = U_h(z)$, where $U_h(\cos(\theta)) = \sin((h+1)\theta)/\sin(\theta)$. (This is readily verified from the recurrence (52) and elementary trigonometry.) Thus, the roots of $F_h(z)$ are $(4\cos^2 j\pi/(h+1))^{-1}$ and the partial fraction expansion of $G^{[h]}(z)$ can be worked out explicitly [**95**]. There results, for $n \geq 1$,

$$G_{n+1}^{[h-2]} = \frac{4^{n+1}}{h} \sum_{1\leq j<h/2} \sin^2\frac{j\pi}{h} \cos^{2n}\frac{j\pi}{h}, \tag{61}$$

which provides in particular an asymptotic form for any fixed h. (This formula can also be found directly from the sampled sum (55) by multisection of series.) Asymptotic analysis of this last expression when $h = x\sqrt{n}$ yields the alternative expression

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}_{n+1}}\left[H \le x\sqrt{n}\right] = 4\pi^{5/2}x^{-3}\sum_{j\ge 0} j^2 e^{-j^2\pi^2/x^2} \quad (\equiv 1 - \Theta(x)),$$

which, when compared with (57), reflects an important transformation formula of elliptic functions [**433**]. See the study by Biane, Pitman, and Yor [**50**] for fascinating connections with Brownian motion and the functional equation of the Riemann zeta function. $\triangleleft$

$\triangleright$ **17.** *Motzkin paths.* The OGF of Motzkin paths of height $< h$ is $\frac{1}{1-z} \cdot {}^{\mathrm{D}}H_{0,0}^{[<h]}\left(\frac{z}{1-z}\right)$, where ${}^{\mathrm{D}}H_{0,0}^{[<h]}$ refers to Dyck paths. Therefore, such paths of length n can be enumerated exactly by formulæ derived from (55–61). In particular, the mean height is $\sim \sqrt{\pi n/3}$. $\triangleleft$

$\triangleright$ **18.** *Height in simple varieties of trees.* Consider a simple variety of trees corresponding to the GF equation $Y(z) = z\phi(Y(z))$ (see Chapter III) and values of n such that there exists a tree of size n. Assume that there exists a positive τ strictly within the disc of convergence of ϕ such that $\tau\phi'(\tau) - \phi(\tau) = 0$. Then, the rth moment of height ($\overline{H}$) is asymptotically $\xi^{r/2} r(r-1)\Gamma(r/2)\zeta(r)n^{r/2}$. The normalized quantity $H = H/\xi$ obeys asymptotically a Theta distribution in the sense of both the central estimate (57) and the local estimate (58). [This is from [**165**] and [**150**] respectively.] For instance, $\xi = 2$ for plane binary trees and $\xi = \sqrt{2}$ for Cayley trees. $\triangleleft$

EXAMPLE 7. *Area under Dyck path and coin fountains.* Consider Dyck paths and the parameter equal to area below the path. *Area* under a lattice path is taken here as the sum of the indices (i.e., the starting altitudes) of all the variables that enter the standard encoding of the path. Thus, the BGF $D(z, q)$ of Dyck path with z marking half-length and q marking area is obtained by the substitution

$$a_j \mapsto q^j z, \quad b_j \mapsto q^j, \quad c_j \mapsto 0$$

inside the fundamental continued fraction (45). (We rederive here Equation (53) of Chapter III, p. 185.) It proves convenient to operate with the continued fraction

$$F(z,q) = \cfrac{1}{1 - \cfrac{zq}{1 - \cfrac{zq^2}{\ddots}}}, \tag{62}$$

so that $D(z,q) = F(q^{-1}z, q^2)$. Since F and D satisfy difference equations, for instance,

$$F(z,q) = \frac{1}{1 - zqF(qz,q)}, \tag{63}$$

moments of area can be determined by differentiating and setting $q = 1$ (see Chapter III for such a direct approach).

A general trick from q–calculus is effective for deriving an alternative form of F. Attempt to express the continued fraction F of (62) as a quotient $F(z,q) = A(z)/B(z)$. Then, the relation (63) implies

$$\frac{A(z)}{B(z)} = \frac{1}{1 - qz\frac{A(qz)}{B(qz)}}, \quad \text{hence } A(z) = B(qz),\ B(z) = B(qz) - qzB(q^2 z),$$

where q is treated as a parameter. The difference equation satisfied by $B(z)$ is then readily solved by indeterminate coefficients. (This classical technique was introduced in the theory of integer partitions by Euler.) With $B(z) = \sum b_n z^n$, the coefficients satisfy the recurrence

$$b_0 = 1, \quad b_n = q^n b_n - q^{2n-1} b_{n-1}.$$

This is a first order recurrence on b_n that unwinds to give

$$b_n = (-1)^n \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}.$$

In other words, introducing the "q-exponential function",

$$E(z,q) = \sum_{n=0}^{\infty} \frac{(-z)^n q^{n^2}}{(q)_n}, \quad \text{where} \quad (q)_n = (1-q)(1-q^2)\cdots(1-q^n), \tag{64}$$

one finds

$$F(z,q) = \frac{E(qz,q)}{E(z,q)}. \tag{65}$$

Given the importance of the functions under discussion in various branches of mathematics, we cannot resist a quick digression. The name of the q-exponential comes form the obvious property that $E(z(1-q), q)$ reduces to e^{-z} as $q \to 1^-$. The explicit form (64) constitutes in fact the "easy half" of the proof of the celebrated Rogers-Ramanujan identities, namely,

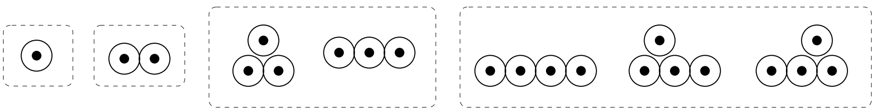
$$\begin{aligned} E(-1,q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+1})^{-1}(1-q^{5n+4})^{-1} \\ E(-q,q) = \sum_{n=0}^{\infty} \frac{q^{n(n+1)}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+2})^{-1}(1-q^{5n+3})^{-1}, \end{aligned} \tag{66}$$

that relate the q-exponential to modular forms. See Andrews' book [**10**, Ch. 7] for context.

Here is finally a cute application of these ideas to the asymptotic enumeration of some special polyominoes. Odlyzko and Wilf define in [**330, 333**] an (n,m) coin fountain as an arrangement of n coins in rows in such a way that there are m coins in the bottom row, and that each coin in a higher row touches exactly two coins in the next lower row. Let $C_{n,m}$ be the number of (n,m) fountains and $C(z,q)$ be the corresponding BGF with q marking n and z marking m. Set $C(q) = C(1,q)$. The question is to determine the total number of coin fountains of area n, $[q^n]C(q)$. The series starts as (this is *EIS* **A005169**)

$$C(q) = 1 + q + q^2 + 2q^3 + 3q^4 + 5q^5 + 9q^6 + 15q^7 + 26q^8 + \cdots,$$

as results from inspection of the first few cases.



There is a clear bijection with Dyck paths that takes area into account: a coin fountain of size n with m coins on its base is equivalent to a Dyck path of length $2m$ and area $2n-m$ (with our ealier definition of area of Dyck paths). From this bijection, one has $C(z,q) = F(z,q)$

Objects	*Weights* $(\alpha_j, \beta_j\gamma_j)$	*Counting*	*Orth. pol.*
Simple paths	$1, 1, 0$	Catalan #	Chebyshev
Permutations	$j+1, j, 2j+1$	Factorial #	Laguerre
Alternating perm.	$j+1, j, 0$	Secant #	Meixner
Involutions	$1, j, 0$	Odd factorial #	Hermite
Set partition	$1, j, j+1$	Bell #	Poisson-Charlier
Nonoverlap. set part.	$1, 1, j+1$	Bessel #	Lommel

FIGURE 8. Some special families of combinatorial objects together with corresponding weights, counting sequences, and orthogonal polynomials. (See also Notes 20— 22.)

(with F as defined earlier) and, in particular, $C(q) = F(1, q)$. Consequently,

$$C(q) = \cfrac{1}{1 - \cfrac{q}{1 - \cfrac{q^2}{1 - \cfrac{q^3}{\ddots}}}},$$

which is (62) with $z = 1$. The identity (65) implies next:

$$C(q) = \frac{E(q,q)}{E(1,q)}.$$

The rest of the discussion is analogous to Section IV. 7.3 (p. 270) relative to alcohols. The function $C(q)$ is *a priori* meromorphic in $|q| < 1$. An exponential lower bound of the form 1.6^n holds for $[q^n]C(q)$, since $(1-q)/(1-q-q^2)$ is dominated by $C(q)$ for $q > 0$. At the same time, the number $[q^n]C(q)$ is majorized by the number of compositions, which is 2^{n-1}. Thus, the radius of convergence of $C(q)$ has to lie somewhere between 0.5 and $0.61803\ldots$. It is then easy to check by numerical analysis the existence of a simple zero of the denominator, $E(1, q)$, near $\rho \doteq 0.57614$. Routine computations based on Rouché's theorem then make it possible to verify formally that ρ is the only pole in $|q| \leq 3/5$ and that this pole is simple (the process is detailed in [**330**]). Thus, singularity analysis of meromorphic functions applies:

PROPOSITION V.5. *The number of coin fountains made of n coins satisfies asymptotically*

$$[q^n]C(q) = cA^n + O((5/3)^n), \quad c \doteq 0.31236, \quad A = \rho^{-1} \doteq 1.73566.$$

This example illustrates the power of modelling by continued fractions as well as the smooth articulation with meromorphic function asymptotics. .... END OF EXAMPLE 7. □

Lattice path encodings of classical structures. The systematic theory of lattice path enumerations and continued fractions was developed initially because of the need to count weighted lattice paths, notably in the context of the analysis of dynamic data structures in computer science [**149**]. In this framework, a system of multiplicative weights $\alpha_j, \beta_j, \gamma_j$ is associated with the steps a_j, b_j, c_j, each weight being an integer that represents a number of "possibilities" for the corresponding step type. A system of weighted lattice paths has counting generating functions given by the usual specialization of the corresponding multivariate expressions we have just developed, namely,

$$a_j \mapsto \alpha_j z, \quad b_j \mapsto \beta_j z, \quad c_j \mapsto \gamma_j z, \tag{67}$$

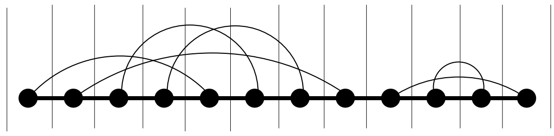


FIGURE 9. An interconnection network on $2n = 12$ points.

where z marks the length of paths. One can then sometimes solve an enumeration problem expressible in this way by reverse-engineering the known collection of continued fractions as found in a reference book like Wall's treatise [**430**]. Next, for general reasons, the polynomials P, Q are always elementary variants of a family of orthogonal polynomials that is determined by the weights (see Note 12 and [**79, 400**]). When the multiplicities have enough structural regularity, the weighted lattice paths are likely to correspond to classical combinatorial objects and to classical families of orthogonal polynomials; see [**139, 149, 202, 208**] and Figure 8 for an outline. We illustrate this by a simple example due to Lagarias, Odlyzko, and Zagier [**284**], which is relative to involutions without fixed points.

EXAMPLE 8. *Interconnection networks and involutions.* The problem treated here has been introduced by Lagarias, Odlyzko, and Zagier in [**284**]: *There are $2n$ points on a line, with n point-to-point connections between pairs of points. What is the probable behaviour of the* width *of such an interconnection network?* Imagine the points to be $1, \ldots, 2n$, the connections as circular arcs between points, and let a vertical line sweep from left to right; width is defined as the maximum number of edges encountered by such a line. One may freely imagine a tunnel of fixed capacity (this corresponds to the width) inside which wires can be placed to connect points pairwise. See Figure 9.

Let $\mathcal{J}_{2n}$ be the class of all interconnection networks on $2n$ points, which is precisely the collection of ways of grouping $2n$ elements into n pairs, or, equivalently, the class of all involutions without fixed points, i.e., permutations with cycles of length 2 only. The number J_{2n} equals the "odd factorial",

$$J_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1),$$

whose EGF is $e^{z^2/2}$ (see Chapter II, p. 114). The problem calls for determining the quantity $J_{2n}^{[h]}$ that is the number of networks corresponding to a width $\le h$.

The relation to lattice paths is as follows. First, when sweeping a vertical line across a network, define an active arc at an abscissa as one that straddles that abscissa. Then build the sequence of active arcs counts at half-integer positions $\frac{1}{2}, \frac{3}{2}, \ldots, 2n - \frac{1}{2}, 2n + \frac{1}{2}$. This constitutes a sequence of integers where each member is ± 1 the previous one, that is, a lattice path without level steps. In other words, there is an ascent in the lattice path for each element that is smaller in its cycle and a descent otherwise. One may view ascents as associated to situations where a node "opens" a new cycle, while descents correspond to "closing" a cycle.

Involutions are much more numerous than lattice paths, so that the correspondence from involutions to lattice paths has to be many-to-one. However, one can easily enrich lattice paths, so that the enriched objects are in one-to-one correspondence with involutions. Consider again a scanning position at a half-integer where the vertical line crosses ℓ (active) arcs. If the next node is of the closing type, there are ℓ possibilities to choose from. If the next node is of the opening type, then there is only one possibility, namely, to start a new cycle. A complete

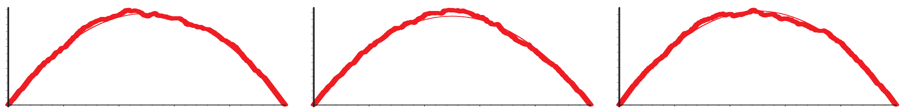


FIGURE 10. Three simulations of random networks with $2n = 1000$ illustrate the tendency of the profile to conform to a parabola with height close to $n/2 = 250$.

encoding of a network is obtained by recording additionally the sequence of the n possible choices corresponding to descents in the lattice path (some canonical order is fixed, for instance, oldest first). If we write these choices as superscripts, this means that the set of all enriched encodings of networks is obtained from the set of standard lattice path encodings by effecting the substitutions

$$b_j \mapsto \sum_{k=1}^{j} b_j^{(k)}.$$

The OGF of all involutions is obtained from the generic continued fraction of Proposition V.3 by the substitution

$$a_j \mapsto z, \quad b_j \mapsto j \cdot z,$$

where z records the number of steps in the enriched lattice path, or equivalently, the number of nodes in the network. In other words, we have obtained combinatorially a *formal* continued fraction representation,

$$\sum_{n=0}^{\infty}(1 \cdot 3 \cdots (2n-1))z^{2n} = \cfrac{1}{1 - \cfrac{1 \cdot z^2}{1 - \cfrac{2 \cdot z^2}{1 - \cfrac{3 \cdot z^2}{\ddots}}}},$$

which was originally discovered by Gauß [**430**]. Proposition V.3 also gives immediately the OGF of involutions of width at most h as a quotient of polynomials. Define

$$J^{[h]}(z) := \sum_{n \geq 0} J_{2n}^{[h]} z^{2n}.$$

One has

$$J^{[h]}(z) = \cfrac{1}{1 - \cfrac{1 \cdot z^2}{1 - \cfrac{2 \cdot z^2}{\cfrac{\ddots}{1 - h \cdot z^2}}}} = \frac{P_{h+1}(z)}{Q_{h+1}(z)}$$

where P_h and Q_h satisfy the recurrence

$$Y_{h+1} = Y_h - hz^2 Y_{h-1}.$$

The polynomials are readily determined by their generating functions that satisfies a first-order linear differential equation reflecting the recurrence. In this way, the denominator polynomials are identified to be reciprocals of the Hermite polynomials,

$$H_h(z) = (2z)^h Q_h\left(\frac{1}{z\sqrt{2}}\right),$$

themselves defined classically [**2**, Ch. 22] as orthogonal with respect to the measure $e^{-x^2}\,dx$ on $(-\infty,\infty)$ and expressible via

$$H_m(x) = \sum_{m=0}^{\lfloor m/2\rfloor} \frac{(-1)^j m!}{j!(m-2j)!}(2x)^{m-2j}, \qquad \sum_{m\ge 0} H_m(x)\frac{t^m}{m!} = e^{2xt-t^2}.$$

In particular, one finds

$$J^{[0]} = 1, \quad J^{[1]} = \frac{1}{1-z^2}, \quad J^{[2]} = \frac{1-2z^2}{1-3z^2}, \quad J^{[3]} = \frac{1-5z^2}{1-6z^2+3z^4}, \quad \&c.$$

The interesting analysis of the dominant poles of the rational GF's, for any fixed h, is discussed in the paper [**284**]. Furthermore, simulations strongly suggest that the width of a random interconnection network on $2n$ nodes is tightly concentrated around $n/2$; see Figure 10. Louchard [**298**] succeeded in proving this fact and a good deal more: With high probability, the altitude (the altitude is defined here as the number of active arcs as time evolves) of a random network conforms asymptotically to a deterministic parabola $2nx(1-x)$ (with $x \in [0,1]$) to which are superimposed random fluctuations of a smaller amplitude, $O(\sqrt{n})$, well-characterized by a Gaussian process. In particular, *the width of a random network of $2n$ nodes converges in probability to $\frac{n}{2}$.* . . . . END OF EXAMPLE 8. □

▷ **19.** *Bell numbers and continued fractions.* With $S_n = n![z^n]e^{e^z-1}$ a Bell number:

$$\sum_{n\ge 0} S_n z^n = \cfrac{1}{1-1z-\cfrac{1z^2}{1-2z-\cfrac{2z^2}{\cdots}}}.$$

[Hint: Define an encoding like for networks, with level steps representing intermediate elements of blocks [**139**].] Refinements include Stirling partition numbers and involution numbers. ◁

▷ **20.** *Factorial numbers and continued fractions.* One has

$$\sum_{n\ge 0} n! z^n = \cfrac{1}{1-1z-\cfrac{1^2z^2}{1-3z-\cfrac{2^2z^2}{\cdots}}}.$$

Refinements include tangent and secant numbers, as well as Stirling cycle numbers and Eulerian numbers. (This continued fraction is due to Euler; see [**139**] for a proof based on a bijection of Françon and Viennot [**188**] and Biane's paper [**49**] for alternative combinatorics.) ◁

▷ **21.** *Surjection numbers and continued fractions.* Let $R_n = n![z^n](2-e^z)^{-1}$. Then

$$\sum_{n=0}^{\infty} R_n z^n = \cfrac{1}{1-1z-\cfrac{2\cdot 1^2z^2}{1-4z-\cfrac{2\cdot 2^2z^2}{1-7z-\cdots}}}.$$

This continued fraction is due to Flajolet [**141**]. ◁

$\triangleright$ **22.** *The Ehrenfest*[2] *urn model.* See Note II.12, p. 111 for context. The OGF of the number of evolutions that lead to urn A full satisfies

$$\sum_{n\geq 0} E_n^{[N]} z^n = \cfrac{1}{1 - \cfrac{1Nz^2}{1 - \cfrac{2(N-1)z^2}{\cdots}}} = \frac{1}{2^N}\sum_{k=0}^{n} \frac{\binom{N}{k}}{1-(N-2k)z}.$$

This results from the EGF of Note II.12, the Continued Fraction Theorem, and basic properties of the Laplace transform. (This continued fraction expansion is originally due to Stieltjes and Rogers. See [**209**] for additional formulæ.) $\triangleleft$

V. 4. The supercritical sequence and its applications

This schema is combinatorially the simplest of all the ones treated in this chapter, since it plainly deals with the sequence construction. An auxiliary analytic condition, named "supercriticality" ensures that meromorphic asymptotics applies and entails strong statistical regularities. This paradigm of supercritical sequences unifies the asymptotic properties of a number of seemingly different combinatorial types, including integer compositions, surjections, and alignments.

V. 4.1. Combinatorial aspects. We consider a sequence construction, $\mathcal{F} = \text{SEQ}(\mathcal{G})$, which may be taken in either the unlabelled or the labelled universe. In either case, we have for the corresponding generating functions the relation

$$F(z) = \frac{1}{1-G(z)},$$

with as usual $G(0) = 0$. It will prove convenient to set

$$f_n = [z^n]F(z), \qquad g_n = [z^n]G(z),$$

so that the number of $\mathcal{F}_n$ structures is f_n in the unlabelled case and $n!f_n$ otherwise.

From Chapter III, the BGF of $\mathcal{F}$-structures with u marking the number of $\mathcal{G}$-components is

$$F(z,u) = \frac{1}{1-uG(z)}. \tag{68}$$

We also have access to the BGF of $\mathcal{F}$ with u marking the number of $\mathcal{G}_k$-components:

$$F^{\langle k\rangle}(z,u) = \frac{1}{1-(G(z)+(u-1)g_k z^k)}. \tag{69}$$

V. 4.2. Analytic aspects. We restrict attention to the case where the radius of convergence ρ of $G(z)$ is nonzero, in which case, the radius of convergence of $F(z)$ is also nonzero by virtue of closure properties of analytic functions. Here is the basic notion of this section.

DEFINITION V.4. *Let F, G be generating functions with nonnegative coefficients that are analytic at 0, with $G(0) = 0$. The analytic relation $F(z) = (1-G(z))^{-1}$ is said to be* supercritical *if $G(\rho) > 1$, where $\rho = \rho_G$ is the radius of convergence of G. A combinatorial schema $\mathcal{F} = \text{SEQ}(\mathcal{G})$ is said to be supercritical if the relation $F(z) = (1-G(z))^{-1}$ between the corresponding generating functions is supercritical.*

Note that $G(\rho)$ is well defined in $\mathbb{R} \cup \{+\infty\}$ as the limit $\lim_{x\to\rho^-} G(x)$ since $G(x)$ increases along the positive real axis, for $x \in (0,\rho)$. (The value $G(\rho)$ corresponds to what has been denoted earlier by τ_G when discussing "signatures" in Section IV. 4, p. 236.) From now on we assume that $G(z)$ is *aperiodic* in the sense that there does not exist an integer $d \geq 2$ such that $G(z) = h(z^d)$ for some h analytic at 0. Put otherwise, the span of $G(z)$ as defined on p. 253 is equal to 1. (This condition entails no loss of analytic generality.)

THEOREM V.5 (Asymptotics of supercritical sequence). *Let the schema $\mathcal{F} = \text{SEQ}(\mathcal{G})$ be supercritical and assume that $G(z)$ is aperiodic. Then, one has*

$$[z^n]F(z) = \frac{1}{\sigma G'(\sigma)} \cdot \sigma^{-n} \left(1 + O(A^n)\right),$$

where σ is the root in $(0,\rho_G)$ of $G(\sigma) = 1$ and A is a number less than 1. The number X of $\mathcal{G}$–components in a random $\mathcal{C}$–structure of size n has mean and variance satisfying

$$\begin{aligned}
\mathbb{E}_n(X) &= \frac{1}{\sigma G'(\sigma)} \cdot (n+1) - 1 + \frac{G''(\sigma)}{G'(\sigma)^2} + O(A^n) \\
\mathbb{V}_n(X) &= \frac{\sigma G''(\sigma) + G'(\sigma) - \sigma G'(\sigma)^2}{\sigma^2 G'(\sigma)^3} \cdot n + O(1).
\end{aligned}$$

In particular, the distribution of X on $\mathcal{F}_n$ is concentrated.

PROOF. See also [**179, 388**]. The basic observation is that G increases continuously from $G(0) = 0$ to $G(\rho_G) = \tau_G$ (with $\tau_G > 1$ by assumption) when x increases from 0 to ρ_G. Therefore, the positive number σ, which satisfies $G(\sigma) = 1$ is well defined. Then, F is analytic at all points of the interval $(0,\sigma)$. The function G being analytic at σ, satisfies, in a neighbourhood of σ

$$G(z) = 1 + G'(\sigma)(z-\sigma) + \frac{1}{2!}G''(\sigma)(z-\sigma)^2 + \cdots .$$

so that $F(z)$ has a pole at $z = \sigma$; also, this pole is simple since $G'(\sigma) > 0$, by positivity of the coefficients of G. Pringsheim's theorem then implies that the radius of convergence of F must coincide with σ.

There remains to show that $F(z)$ is meromorphic in a disc of some radius $R > \sigma$ with the point σ as the only singularity inside the disc. This results from the assumption that G is aperiodic. In effect, by the Daffodil Lemma (Lemma IV.3, p. 254), one has $|G(\sigma e^{i\theta})| < 1$ for all $\theta \not\equiv 0 \pmod{2\pi}$. Thus, by compactness, there exists a disc of radius $R > \sigma$ in which F is analytic except for a unique pole at σ. Take r such that $\sigma < r < R$ and apply the main theorem of meromorphic function asymptotics to deduce the stated formula with $A = \sigma/r$.

Consider next the number of $\mathcal{G}$-components in a random $\mathcal{F}$ structure of size n. Bivariate generating functions give access to the expectation of this random variable:

$$\begin{aligned}
\mathbb{E}_n(X) &= \frac{1}{f_n}\,[z^n]\,\left.\frac{\partial}{\partial u}\frac{1}{1-uG(z)}\right|_{u=1} \\
&= \frac{1}{f_n}\,[z^n]\,\frac{G(z)}{(1-G(z))^2}.
\end{aligned}$$

The problem is now reduced to extracting coefficients in a univariate generating function with a double pole at $z = \sigma$, and it suffices to expand the GF locally at σ. The variance calculation is similar though it involves a triple pole. $\square$

When a sequence construction is supercritical, the number of components is in the mean of order n while its standard deviation is $O(\sqrt{n})$. Thus, the distribution is concentrated (see Section III. 2.2, p. 150). In fact, there results from a general theorem of Bender [**27**] that the distribution of the number of components is asymptotically Gaussian; see Chapter IX for details.

Profiles of supercritical sequences. We have seen in Chapter III that integer compositions and integer partitions, when sampled at random, tend to assume rather different aspects. Given a sequence construction, $\mathcal{F} = \text{SEQ}(\mathcal{G})$, the profile of an element $\alpha \in \mathcal{F}$ is the vector $(X^{\langle 1\rangle}, X^{\langle 2\rangle}, \ldots)$ where $X^{\langle 1\rangle}(\alpha)$ is the number of $\mathcal{G}$–components in α that have size j. In the case of (unrestricted) integer compositions, it could be proved elementarily that, on average and for size n, the number of 1 summands is $\sim n/2$, the number of 2 summands is $\sim n/4$, and so on. Now that meromorphic asymptotic is available, such a property can be placed in a much wider perspective.

THEOREM V.6 (Profiles of supercritical sequences). *Consider a supercritical sequence construction, $\mathcal{F} = \text{SEQ}(\mathcal{G})$, with the aperiodicity condition. The number of $\mathcal{G}$–components of any fixed size k in a random $\mathcal{F}$–object of size n satisfies*

$$\mathbb{E}_n(X^{\langle k\rangle}) = \frac{g_k\sigma^k}{\sigma G'(\sigma)} n + O(1), \quad \mathbb{V}_n(X^{\langle k\rangle}) = O(n). \tag{70}$$

There, σ is the root in $(0, \sigma_G)$ of $G(\sigma) = 1$, and $g_k = [z^k]G(z)$.

PROOF. The bivariate GF with u marking the number of $\mathcal{G}$–components of size k is

$$F(z, u) = \frac{1}{1 - (G(z) + (u-1)g_k z^k)},$$

as results from the theory developed in Chapter III. The mean value is then given by a quotient,

$$\mathbb{E}_n(X^{\langle k\rangle}) = \frac{1}{f_n} [z^n] \left. \frac{\partial}{\partial u} F(z, u) \right|_{u=1} = \frac{1}{f_n} [z^n] \frac{g_k z^k}{(1 - G(z))^2}.$$

The GF of cumulated values has a double pole at $z = \sigma$, and the estimate of the mean value follows. The variance is estimated similarly, after two successive differentiations and the analysis of a triple polar singularity. $\square$

The total number of components X satisfies $X = \sum X^{\langle k\rangle}$, and, by Theorem V.5, its mean is asymptotic to $n/(\sigma G'(\sigma))$. Thus, Equation (70) indicates that, at least in some average-value sense, the "proportion" of components of size k amongst all components is given by $g_k\sigma^k$.

$\triangleright$ **23.** *Proportion of k–components and convergence in probability.* For any fixed k, the random variable $X_n^{\langle k\rangle}/X_n$ converges in probability to the value $g_k\sigma^k$,

$$\frac{X_n^{\langle k\rangle}}{X_n} \xrightarrow{P} g_k\sigma^k, \quad \text{i.e.,} \quad \lim_{n\to\infty} \mathbb{P}\left\{ g_k\sigma^k(1-\epsilon) \le \frac{X_n^{\langle k\rangle}}{X_n} \le g_k\sigma^k(1+\epsilon) \right\} = 1,$$

for any $\epsilon > 0$. The proof is an easy consequence of the Chebyshev inequalities (the distributions of X_n and $X_n^{\langle k\rangle}$ are both concentrated). $\lhd$

V. 4.3. Applications. We examine here two types of applications of the supercritical sequence schema.

— Example 9 makes explicit the asymptotic enumeration and the analysis of profiles of compositions, surjections and alignments. What stands out is the way the mean profile of a structure reflects the underlying inner construction $\mathfrak{K}$ in schemas of the form $\text{SEQ}(\mathfrak{K}(\mathcal{Z}))$.

– Example 10 discusses compositions into restricted summands, including the striking case of compositions into primes.

EXAMPLE 9. *Compositions, surjections, and alignments*. The three classes of interest here are integer compositions ($\mathcal{C}$), surjections ($\mathcal{R}$) and alignments ($\mathcal{O}$), which are specified as

$$\mathcal{C} = \text{SEQ}(\text{SEQ}_{\geq 1}(\mathcal{Z})), \qquad \mathcal{R} = \text{SEQ}(\text{SET}_{\geq 1}(\mathcal{Z})), \qquad \mathcal{C} = \text{SEQ}(\text{CYC}(\mathcal{Z}))$$

and belong to either the labelled universe ($\mathcal{C}$) or to the labelled universe ($\mathcal{R}$ and $\mathcal{O}$). The generating functions (of type OGF, EGF, and EGF, respectively) are

$$C(z) = \frac{1}{1 - \frac{z}{1-z}}, \qquad R(z) = \frac{1}{1-(e^z-1)}, \qquad O(z) = \frac{1}{1-\log(1-z)^{-1}}.$$

A direct application of Theorem V.5 gives us the already known results

$$C_n = 2^{n-1}, \qquad \frac{1}{n!}R_n \sim \frac{1}{2}(\log 2)^{-n-1}, \qquad \frac{1}{n!}O_n = e^{-1}(1-e^{-1})^{-n-1},$$

corresponding to σ equal to $\frac{1}{2}$, $\log 2$, and $1-e^{-1}$, respectively.

Similarly, the expected number of summands in a random composition of the integer n is $\sim \frac{n+1}{2}$. The expected cardinality of the range of a random surjection whose domain has cardinality n is asymptotic to βn with $\beta = 1/(2\log 2)$; The expected number of components in a random alignment of size n is asymptotic to $n/(e-1)$.

Theorem V.6 also applies and gives the mean number of components of size k in each case. The following table summarizes the conclusions:

Structures	*Specif.*	*Law* ($g_k\sigma^k$)	*Type*	σ
Compositions	$\text{SEQ}(\text{SEQ}_{\geq 1}(\mathcal{Z}))$	$\frac{1}{2^k}$	Geometric	$\frac{1}{2}$
Surjections	$\text{SEQ}(\mathfrak{P}_{\geq 1}(\mathcal{Z}))$	$\frac{1}{k!}(\log 2)^k$	Poisson	$\log 2$
Alignments	$\text{SEQ}(\mathfrak{C}(\mathcal{Z}))$	$\frac{1}{k}(1-e^{-1})^k$	Logarithmic	$1-e^{-1}$

Note that the stated laws necessitate $k \geq 1$. The geometric and Poisson law are classical; the *logarithmic distribution* (also called "logarithmic-series distribution") of parameter λ is by definition the law of a discrete random variable Y such that

$$\mathbb{P}(Y=k) = \frac{1}{\log(1-\lambda)^{-1}}\frac{\lambda^k}{k}, \qquad k \geq 1.$$

The way the internal construction $\mathfrak{K}$ in the schema $\text{SEQ}(\mathfrak{K}(\mathcal{Z}))$ determines the law of component sizes,

$$\text{Sequence} \mapsto \text{Geometric}; \quad \text{Set} \mapsto \text{Poisson}; \quad \text{Cycle} \mapsto \text{Logarithmic},$$

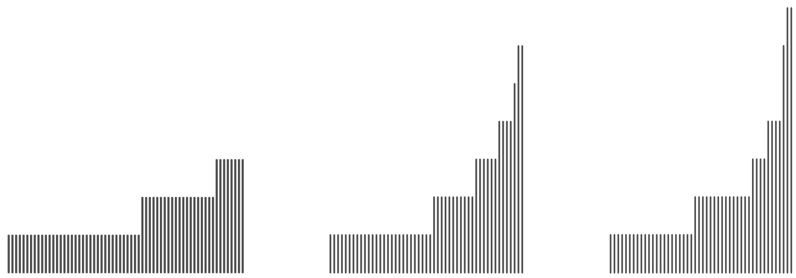


FIGURE 11. Profile of structures drawn at random represented by the sizes of their components in sorted order: (from left to right) a random surjection, alignment, and composition of size $n = 100$.

stands out. Figure 11 exemplifies the phenomenon by displaying components sorted by size and represented by vertical segments of corresponding lengths for three randomly drawn objects of size $n = 100$. . . . . END OF EXAMPLE 9. □

EXAMPLE 10. *Compositions with restricted summands, compositions into primes.* Unrestricted integer compositions are well understood as regards enumeration: their number is exactly $C_n = 2^{n-1}$, their OGF is $C(z) = (1-z)/(1-2z)$, and compositions with k summands are enumerated by binomial coefficients. Such simple exact formulæ disappear when restricted compositions are considered, but, as we now show, asymptotics is much more robust to changes in specifications.

Let S be a subset of the integers $\mathbb{Z}_{\geq 1}$ such that $\gcd(S) = 1$, i.e., not all members of S are multiples of a common divisor $d \geq 2$. In order to avoid trivialities, we also assume that $S \neq \{1\}$. The class $\mathcal{C}^S$ of compositions with summands constrained to the set S then satisfies:

$$\begin{array}{rl} \textit{Specification}: & \mathcal{C}^S = \mathrm{SEQ}(\mathrm{SEQ}_S(\mathcal{Z})); \\ \textit{OGF}: & C^S(z) = \dfrac{1}{1 - S(z)}, \quad S(z) = \displaystyle\sum_{s \in S} z^s. \end{array}$$

By assumption, $S(z)$ is aperiodic, so that Theorem V.5 applies directly. There is a well-defined number σ such that

$$S(\sigma) = 1, \qquad 0 < \sigma < 1,$$

and the number of S–restricted compositions satisfies

$$C_n^S := [z^n]C^S(z) = \frac{1}{\sigma S'(\sigma)} \cdot \sigma^{-n} \left(1 + O(A^n)\right). \tag{71}$$

Amongst the already discussed cases, $S = \{1, 2\}$ gives rise to Fibonacci numbers and, more generally, $S = \{1, \dots, r\}$ corresponds to partitions with summands at most r. In this case, the OGF,

$$C^{\{1,\dots,r\}}(z) = \frac{1}{1 - z\frac{1-z^r}{1-z}} = \frac{1-z}{1 - 2z + z^{r+1}}$$

is a simple variant of the OGF associated to longest runs in strings. The treatment of the latter can be copied almost verbatim to the effect that the largest component in a random composition of n is found to be $\lg n + O(1)$, both on average and with high probability.

10	16	1 ***5***
20	732	73 ***4***
30	36039	360 ***57***
40	1772207	17722 ***61***
50	87109263	871092 ***48***
60	4281550047	42815 ***49331***
70	210444532770	21044453 ***0095***
80	10343662267187	1034366226 ***5182***
90	508406414757253	5084064147 ***81706***
100	24988932929490838	24988932929 ***612479***

FIGURE 12. The pyramid relative to compositions into prime summands for $n = 10\,..\,100$: (left: exact values; right: asymptotic formula rounded).

Here is a surprising application of the general theory. Consider the case where S is taken to be the set of prime numbers, $\text{Prime} = \{2, 3, 5, 7, 11, \ldots\}$, thereby defining the class of *compositions into prime summands*. The sequence starts as

$$1, 0, 1, 1, 1, 3, 2, 6, 6, 10, 16, 20, 35, 46, 72, 105,$$

corresponding to $G(z) = z^2 + z^3 + z^5 + \cdots$, and is *EIS* **A023360** in Sloane's encyclopedia. The formula (71) provides the asymptotic form of the number of such compositions. It is also worth noting that the constants appearing in (71) are easily determined to great accuracy, as we now explain.

By (71) and the preceding equation, the dominant singularity of the OGF of compositions into prime is the positive root $\sigma < 1$ of the characteristic equation

$$S(z) \equiv \sum_{p \ \text{Prime}} z^p = 1.$$

Fix a threshold value m_0 (for instance $m_0 = 10$ or 100) and introduce the two series

$$S^-(z) := \sum_{s\in S,\ s<m_0} z^s, \quad S^+(z) := \left(\sum_{s\in S,\ s<m_0} z^s\right) + \frac{z^{m_0}}{1-z}.$$

Clearly, for $x \in (0, 1)$, one has $S^-(x) < S(x) < S^+(x)$. Define then two constants σ^-, σ^+ by the conditions

$$S^-(\sigma^-) = 1, \quad S^+(\sigma^+) = 1, \qquad 0 < \sigma^-, \sigma^+ < 1.$$

These constants are algebraic numbers that are accessible to computation. At the same time, they satisfy $\sigma^+ < \sigma < \sigma^-$. As the order of truncation, m_0, increases, the values of σ^+, σ^- provide better and better approximations to σ, together with an interval in which σ provably lies. For instance, $m_0 = 10$ is enough to determine that $0.66 < \sigma < 0.69$, and the choice $m_0 = 100$ gives σ to 15 guaranteed digits of accuracy, namely, $\sigma \doteq 0.67740\,17761\,30660$. Then, the asymptotic formula (71) instantiates as

$$C_n^{\text{Prime}} \sim g(n), \qquad g(n) := 0.30365\,52633 \cdot 1.47622\,87836^n. \tag{72}$$

The constant $\sigma^{-1} \doteq 1.47622$ is akin to the family of Backhouse constants described in [**137**].

Once more, the asymptotic approximation is very good as exemplified by the pyramid of Figure 12. The difference between C_n^{Prime} and its approximation $g(n)$ from Eq. (72) is plotted on the left of Figure 13. The seemingly haphazard oscillations that manifest themselves are well explained by the principles discussed in Section IV. 6.1 (p. 250). It appears that the next poles

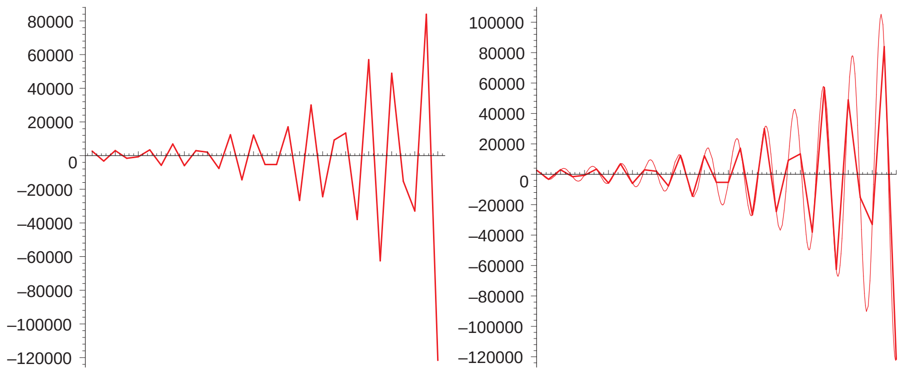


FIGURE 13. Errors in the approximation of the number of compositions into primes for $n = 70 \,..\, 100$: left, the values of $C_n^{\mathrm{Prime}} - g(n)$; right, the correction $g_2(n)$ arising from the next two poles, which are complex conjugate and the continuous extrapolation of this approximation.

of the OGF are complex conjugate and lie near $-0.76 \pm 0.44i$, having modulus about 0.88. The corresponding residues then jointly contribute a quantity of the form

$$g_2(n) = c \cdot A^n \sin(\omega n + \omega_0), \qquad A \doteq 1.13290,$$

for some constants c, ω, ω_0. Comparing the left and right parts of Figure 13, we see that this next layer of poles explains quite well the residual error $C_n^{\mathrm{Prime}} - g(n)$.

Here is a final example that demonstrates in a striking way the scope of the method. Define the set Prime_2 of "twinned primes" as the set of primes that belong to a twin prime pair, that is, $p \in \mathrm{Prime}_2$ if one of $p-2, p+2$ is prime. The set Prime_2 starts as $3, 5, 7, 11, 13, 17, 19, 29, 31, \ldots$ (numbers like 23 or 37 are thus excluded). The asymptotic formula for the number of compositions of the integer n into summands that are twinned primes, is

$$C_n^{\mathrm{Prime}_2} \sim 0.18937 \cdot 1.29799^n,$$

where the constants are found by methods analogous to the case of all primes. It is quite remarkable that the constants involved are still computable real numbers (and of low complexity, even), this despite the fact that it is not known whether the set of twinned primes is finite or infinite. Incidentally, a sequence that starts like $C_n^{\mathrm{Prime}_2}$,

$$1, 0, 0, 1, 0, 1, 1, 1, 2, 1, 3, 4, 3, 7, 7, 8, 14, 15, 21, 28, 33, 47, 58, \ldots$$

and coincides till index 22 included (!), but not beyond, was encountered by P. A. MacMahon[9], as the authors discovered, much to their astonishment, from scanning Sloane's Encyclopedia, where it appears as *EIS* **A002124**. .... END OF EXAMPLE 10. □

[9]See "Properties of prime numbers deduced from the calculus of symmetric functions", *Proc. London Math. Soc.*, 23 (1923), 290-316). MacMahon's sequence corresponds to compositions into arbitrary *odd* primes, and 23 is the first such prime that is not twinned.

$\triangleright$ **24.** *Random generation of supercritical sequences.* Let $\mathcal{F} = \text{SEQ}(\mathcal{G})$ be a supercritical sequence scheme. Consider a sequence of i.i.d. (independently identically distributed) random variables $Y_1, Y_2, \ldots$ each of them obeying the discrete law

$$\mathbb{P}(Y = k) = g_k \sigma^k, \qquad k \geq 1.$$

A sequence is said to be hitting n if $Y_1 + \cdots + Y_r = n$ for some $r \geq 1$. The vector $(Y_1, \ldots, Y_r)$ for a sequence conditioned to hit n has the same distribution as the sequence of the lengths of components in a random $\mathcal{F}$–object of size n.

For probabilists, this explains the shape of the formulæ in Theorem V.5, which resemble renewal relations [**133**, Sec. XIII.10]. It also implies that, given a uniform random generator for $\mathcal{G}$–objects, one can generate a random $\mathcal{F}$–object of size n in $O(n)$ steps on average [**115**]. This applies to surjections, alignments, and compositions in particular. $\triangleleft$

$\triangleright$ **25.** *Largest components in supercritical sequences.* Let $\mathcal{F} = \text{SEQ}(\mathcal{G})$ be a supercritical sequence. Assume that $g_k = [z^k]G(z)$ satisfies the asymptotic "smoothness" condition

$$g_k \underset{k \to \infty}{\sim} c\rho^{-k} k^{\beta}, \qquad c, \rho \in \mathbb{R}_{>0}, \ \beta \in \mathbb{R}.$$

Then the size L of the largest $\mathcal{G}$ component in a random $\mathcal{F}$ object satisfies, for size n,

$$\mathbb{E}_{\mathcal{F}_n}(X) = \frac{1}{\log(\rho/\sigma)} \left(\log n + \beta \log\log n\right) + o(\log\log n).$$

This covers integer compositions ($\rho = 1$, $\beta = 0$) and alignments ($\rho = 1$, $\beta = -1$). [The analysis generalizes the case of longest runs in Example 2 and is based on similar principles. The GF of $\mathcal{F}$ objects with $L \leq m$ is $F^{\langle m \rangle}(z) = \left(1 - \sum_{k \leq m} g_k z^k\right)^{-1}$, according to Section III.7. For m large enough, this has a dominant singularity which is a simple pole at σ_m such that $\sigma_m - \sigma \sim c_1(\sigma/\rho)^m m^{\beta}$. There follows a double-exponential approximation

$$\mathbb{P}_{\mathcal{F}_n}(L \leq m) \approx \exp\left(-c_2 n m^{\beta} (\sigma/\rho)^m\right)$$

in the "central" region. See Gourdon's study [**210**] for details.] $\triangleleft$

V. 5. Paths in graphs and automata

In this section, we first develop the framework of *paths in graphs*: given a graph G, a source node, and a destination node, the problem is to enumerate all paths from the source to the destination. Nonnegative weights may be attached to edges. Applications include the analysis of walks in various types of graphs as well as languages described by finite automata. The set of *eigenvalues* (the spectrum) of the corresponding adjacency matrix intervenes in all quantitative estimates. Under a fundamental structural condition, known as *irreducibility* and corresponding to strong-connectedness of the graph, only the largest positive eigenvalue matters asymptotically: at generating function level, this is reflected by simplicity of the dominant real pole and leads to pure exponential forms for coefficients (possibly tempered by explicit congruence conditions in the periodic case).

V. 5.1. Combinatorial aspects. A *directed graph* or *digraph* is determined by the pair (V, E) of its vertex set V and its edge set $E \subseteq V \times V$. Here, self loops corresponding to edges of the form (v, v) are allowed. Given an edge, $e = (a, b)$, we denote its origin by $\text{orig}(e) := a$ and its destination by $\text{destin}(e) := b$. Let G be a digraph with vertex set $\{1, \ldots, m\}$. Each edge (a, b) will be weighted by a quantity

$g_{a,b}$ which we take to be a formal indeterminate. (We also allow ourselves later to substitute positive values for weights.) We introduce the matrix $\mathbf{G}$ such that

$$\mathbf{G}_{a,b} = g_{a,b} \text{ if the edge } (a,b) \in G, \qquad \mathbf{G}_{a,b} = 0 \text{ otherwise,} \tag{73}$$

which is called the *formal adjacency matrix* of G.

A *path* is a sequence of edges, $\varpi = (e_1, \ldots, e_n)$, such that, for all j with $1 \leq j < n$, one has $\operatorname{destin}(e_j) = \operatorname{orig}(e_{j+1})$. The parameter n is called the length of the path and we define: $\operatorname{orig}(\varpi) := \operatorname{orig}(e_1)$, $\operatorname{destin}(\varpi) := \operatorname{destin}(e_n)$. A *circuit* is a path whose origin and destination are the same vertex. Note that, with our definition, a circuit has its origin that is distinguished. We do *not* identify here two circuits such that one is obtained by circular permutation from the other and also refer to circuits in that sense as *rooted circuits*.

From the standard definition of matrix products, the powers $\mathbf{G}^r$ have elements that are path polynomials. More precisely, one has the simple but essential relation,

$$(\mathbf{G})^r_{i,j} = \sum_{w \in \mathcal{P}_n^{\langle i,j \rangle}} w, \tag{74}$$

where $\mathcal{P}_n^{\langle i,j \rangle}$ is the set of paths in G that connect i to j and have length r, and a path w is assimilated to the monomial in indeterminates $\{g_{i,j}\}$ that represents multiplicatively the succession of its edges; for instance:

$$(\mathbf{G})^3_{i,j} = \sum_{m_1 = i, m_2, m_3, m_4 = j} g_{m_1,m_2} g_{m_2,m_3} g_{m_3,m_4},$$

In other words, powers of the matrix associated to a graph "generate" all paths in a graph. (This fact probably constitutes the most basic result of algebraic graph theory [**51**, p. 9].) One may then treat simultaneously all lengths of paths (and all powers of matrices) by introducing the variable z to record length.

PROPOSITION V.6. (i) *Let G be a digraph and let $\mathbf{G}$ be the formal adjacency matrix of G as given by* (73)*. The OGF $F^{\langle i,j \rangle}(z)$ of the set of all paths from i to j in a digraph G with z marking length and $g_{a,b}$ marking the occurrence of edge (a,b) is the entry i,j of the matrix $(I - z\mathbf{G})^{-1}$, namely*

$$F^{\langle i,j \rangle}(z) = (I - z\mathbf{G})^{-1}\Big|_{i,j} = (-1)^{i+j} \frac{\Delta^{\langle i,j \rangle}(z)}{\Delta(z)}, \tag{75}$$

where $\Delta(z) = \det(I - z\mathbf{G})$ and $\Delta^{\langle j,i \rangle}(z)$ is the determinant of the minor of index j, i of $I - z\mathbf{G}$.

(ii) *The generating function of (rooted) circuits is given by*

$$\sum_i (F^{\langle i,i \rangle}(z) - 1) = -z \frac{\Delta'(z)}{\Delta(z)}.$$

The quantity $\det(I - z\mathbf{G})$ is none other than the reciprocal polynomial of the characteristic polynomial of $\mathbf{G}$.

PROOF. Part (i) results from the discussion above which implies

$$F^{\langle i,j\rangle}(z) = \sum_{n=0}^{\infty} z^n \left(\mathbf{G}^n\right)_{i,j} = \left((I - z\mathbf{G})^{-1}\right)_{i,j},$$

and from the cofactor formula of matrix inversion.

Part (ii) results from Jacobi's trace formula [**208**, p. 11] for square matrices,

$$\det \circ \exp(M) = \exp \circ \operatorname{Tr}(M) \tag{76}$$

(equivalently, with due care paid to determinations: $\log \circ \det(M) = \operatorname{Tr} \circ \log(M)$) which generalizes the scalar identity $e^a e^b = e^{a+b}$ (or $\log ab = \log a + \log b$).

Introduce the quantity known as the *zeta function*,

$$\begin{aligned} \zeta(z) &:= \exp\left(\sum_i \sum_{n=1}^{\infty} F_n^{\langle i,i\rangle} \frac{z^n}{n}\right) &= \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n} \operatorname{Tr} \mathbf{G}^n\right) \\ &= \exp\left(\operatorname{Tr} \log(I - z\mathbf{G})^{-1}\right) &= \det(I - z\mathbf{G})^{-1}, \end{aligned}$$

where the last line results from the Jacobi trace formula. Then, $\zeta(z) = \Delta(z)^{-1}$. On the other hand, differentiation combined with the definition of $\zeta(z)$ yields

$$\begin{aligned} z\frac{\zeta'(z)}{\zeta(z)} &= -z\frac{\Delta'(z)}{\Delta(z)} \\ &= \sum_i \sum_{n=1}^{\infty} F_n^{\langle i,i\rangle} z^n, \end{aligned}$$

and Part (ii) follows. $\square$

$\triangleright$ **26.** *Proof of the Jacobi trace formula.* The Jacobi trace formula is readily verified when the matrix is diagonalizable. It can then be extended to all matrices by a density argument. $\triangleleft$

$\triangleright$ **27.** *Fast computation of the characteristic polynomial.* The following algorithm is due to Leverrier (1811–1877), the astronomer and mathematician who, together with Adams, first predicted the position of the planet Neptune. Observe, for a diagonalizable matrix, that

$$z\frac{\zeta'(z)}{\zeta(z)} = \sum_{n\geq 1} z^n \operatorname{Tr} \mathbf{G}^n = \sum_{\lambda} \frac{\lambda z}{1 - \lambda z}$$

(the sum is over eigenvalues). From this, one deduces an algorithm that determines the characteristic polynomial of a matrix of dimension m in $O(m^4)$ arithmetic operations. [Hint: computing the quantities $\operatorname{Tr} G^j$ for $j = 1, \ldots, m$ requires precisely m matrix multiplications.] $\triangleleft$

$\triangleright$ **28.** *The matrix tree theorem.* Let G be a directed graph without loops and associated matrix $\mathbf{G}$, with $g_{a,b}$ marking edge (a, b). The Laplacian matrix $\mathbf{L}[G]$ is defined by

$$\mathbf{L}[G]_{i,j} = -g_{i,j} + [\![i = j]\!] \delta_i, \qquad \text{where} \quad \delta_i := \sum_k g_{i,k}.$$

Let $\mathbf{L}_1[G]$ be the matrix obtained by deleting the first row and first column of $\mathbf{L}[G]$. Then, the "tree polynomial"

$$T_1[G] := \det \mathbf{L}_1[G]$$

enumerates all (oriented) spanning trees of G rooted at node 1. [This classic result belongs to a circle of ideas initiated by Kirchhoff, Sylvester, Borchardt and others in the 19th century. See, e.g., the discussions by Knuth [**268**, p. 582–583] and Moon [**318**].] $\triangleleft$

Weighted graphs, word models, and finite automata. The numeric substitution $\sigma: \ g_{a,b} \mapsto 1$ transforms the matrix $\mathbf{G}$ into the usual *adjacency matrix*. In particular, the number of paths of length n is obtained, under this substitution, as $[z^n](1-z\mathbf{G})^{-1}$. In a similar vein, it is possible to consider weighted graphs, where the $g_{a,b}$ are assigned *positive real-valued weights*; with the weight of a path being defined by the product of its edges weights. One finds that $[z^n](I-z\mathbf{G})^{-1}$ equals the total weight of all paths of length n. If furthermore the assignment is made in such a way that $\sum_b g_{a,b} = 1$, then the matrix $\mathbf{G}$, which is called a *stochastic matrix*, can be interpreted as the transition matrix of a Markov chain. Naturally, the statement of Proposition V.6 survives such a numerical instantiation of the formal variables.

Word problems corresponding to regular languages can be treated by the theory of regular specifications whenever they have enough structure and an unambiguous regular expression description is of tractable form. (This is the main theme pursued in Sections V. 2 and V. 3.) The dual point of view of automata theory introduced in Section I. 4.2 (p. 53) proves useful whenever no such direct description is in sight. Finite automata resort to the theory of paths in graphs, so that Proposition V.6 remains applicable. Indeed, the language $\mathcal{L}$ accepted by a finite automaton A with set of states Q and with Q_f the set of final states decomposes as

$$\mathcal{L} = \bigcup_{q \in Q_f} \mathcal{L}_q,$$

where $\mathcal{L}_q$ is the language accepted by the automaton A, but with the set of final states reduced to the singleton set $\{q\}$. Then, any such $\mathcal{L}_q$ coincides with set of paths leading from the initial state q_0 to the final state q in A. The corresponding graph G is obtained from A by collapsing multiple edges between any two vertices, i and j, into a single edge equipped with a weight that is the sum of the weights of all the letters leading from i to j.

Profiles. By profile of a set $\mathcal{P}$ of paths is meant here the collection of the m^2 statistics $N = (N_{1,1}, \ldots, N_{m,m})$ where $N_{i,j}$ is the number of times the edge $(i \longrightarrow j)$ is traversed. This notion is for instance consistent with the notion of profile given earlier for lattice paths in Section V. 3. It also contains the information regarding the letter composition of words in a regular language and is thus compatible with the notion of profile introduced in Section V. 2.

Let G be a graph with edge (a, b) weighted by $\gamma_{a,b}$. Then, the BGF of paths with u marking the number of times a particular edge (c, d) is traversed is in matrix form

$$(I - z\widetilde{\mathbf{G}})^{-1}, \qquad \text{with} \quad \widetilde{\mathbf{G}} = \mathbf{G}\left[g_{a,b} \mapsto \gamma_{a,b} u^{[\![a=c \wedge b=d]\!]}\right].$$

The entry (i, j) in this matrix gives the BGF of paths with origin i and destination j. The GF of cumulated values (moments of order 1) is then obtained from there in the usual way, by differentiation followed by the substitution $u = 1$. Higher moments are similarly attainable by successive differentiations.

V. 5.2. Analytic aspects. For rational functions, positivity coupled with some simple ancillary conditions entails a host of important properties, like unicity of the dominant singularity. Such facts result from the classical Perron-Frobenius theory of nonnegative matrices that we summarize in this section. They in turn imply strong statistical regularity of large random structures.

The basic case is that of a d-dimensional column vector $\mathrm{y}(z)$ of generating functions satisfying a linear system of the form

$$\mathrm{y}(z) = \mathrm{a} + zT\,\mathrm{y}(z),$$

for some $(d \times d)$ matrix T and vector a. This system is solved by a quasi-inverse:

$$\mathrm{y}(z) = (I - zT)^{-1}\,\mathrm{a}\,.$$

If T satisfies suitable positivity conditions and a is nonnegative, then any component $\mathrm{y}_j(z)$ closely resembles the extremely simple rational function,

$$\frac{1}{1 - \lambda_1 z},$$

where λ_1 is a well-characterized eigenvalue of T. Accordingly, the asymptotic phenomena associated with such systems are highly predictable. We propose to expose here the general theory and treat in the next section classical applications to paths in graphs and to languages recognized by finite automata.

Perron-Frobenius theory of nonnegative matrices. For an arbitrary square matrix $A \in \mathbb{R}^{m \times m}$, the *spectrum* is the set of its *eigenvalues*, that is, the set of λ such that $\lambda I - A$ is not invertible (i.e., not of full rank), where I is the unit matrix with the appropriate dimension. A *dominant eigenvalue* is one of largest modulus. Finally, the spectral radius of an arbitrary matrix A is defined as

$$\sigma(A) = \max_j \{|\lambda_j|\}, \tag{77}$$

where the set $\{\lambda_j\}$ is the set of eigenvalues of A (also called spectrum). The spectral radius $\sigma(A)$ describes growth properties associated to the powers of A. Indeed, given the Jordan normal form of matrices, it is easy to see that all entries of A^n are bounded from above by a multiple of $\sigma(A)^n \cdot n^{r-1}$, where r is the maximum multiplicity of any dominant eigenvalue. When analysing a family of combinatorial models that admit a matrix formulation, it is then of obvious interest to determine the value of the spectral radius and the multiplicities attached to dominant eigenvalues.

The properties of positive and of nonnegative matrices have been superbly elicited by Perron [**345**] in 1907 and by Frobenius [**189**] in 1908–1912. The corresponding theory has far-reaching implications: it lies at the basis of the theory of finite Markov chains and it extends to positive operators in infinite-dimensional spaces [**280**].

For A a scalar matrix of dimension $m \times m$ with nonnegative entries, a crucial rôle is played by the *dependency graph*; this is the (directed) graph with vertex set $V = \{1\,..\,m\}$ and edge set containing the directed edge $(a \to b)$ iff $A_{a,b} \neq 0$. The reason for this terminology is the following: Let A represent the linear transformation $\left\{y_i^\star = \sum_j A_{i,j} y_j\right\}_i$; then, the fact that an entry $A_{i,j}$ is nonzero means that $y_i^\star$ depends effectively on y_j and is translated by the directed edge $(i \to j)$ in the dependency graph.

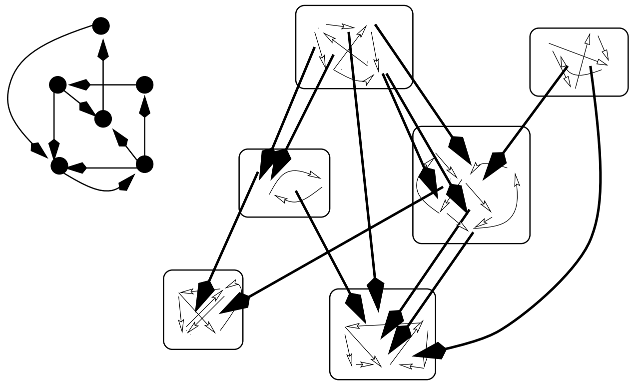


FIGURE 14. The irreducibility conditions of Perron-Frobenius theory. Left: a strongly connected digraph. Right: a weakly connected digraph that is not strongly connected is a collection of strongly connected components linked by a directed acyclic graph.

From this point on, we consider matrices with nonnegative entries. Two notions are essential, irreducibility and aperiodicity (the terms are borrowed from Markov chain theory and matrix theory).

DEFINITION V.5. *The matrix A is called* irreducible *if its dependency graph is strongly connected (i.e., any two vertices are connected by a directed path).*

By considering only simple paths, it is then seen that irreducibility is equivalent to the condition that $(I+A)^m$ has all its entries that are strictly positive. See Figure 14 for a graphical rendering of irreducibility and for the general structure of a (weakly connected) digraph.

DEFINITION V.6. *A strongly connected digraph G is* periodic *with parameter d iff all its cycles have a length that is a multiple of d. In that case, the graph decomposes into cyclically arranged layers: the vertex set V can be partitioned into d classes, $V = V_0 \cup \cdots \cup V_{d-1}$, in such a way that the edge set E satisfies*

$$E \subseteq \bigcup_{i=0}^{d-1} (V_i \times V_{(i+1) \bmod d}). \tag{78}$$

The maximal possible d is called the period*. If no decomposition exists with $d \geq 2$, so that the period has the trivial value 1, then the graph and all the matrices that admit it as their dependency graph are called* aperiodic*.*

As an illustration of periodicity, a directed 10-cycle is periodic with parameter $d = 1, 2, 5, 10$ and the period is 10. See Figure 15 for representations of a periodic and an aperiodic digraph.

Periodicity means that the existence of paths of length n between any two given nodes $i, j\rangle$ is constrained by the congruence class $n \bmod d$. Conversely, aperiodicity entails the existence, for all n sufficiently large, of paths of length n connecting i, j. From the definition, a matrix A with period d has, up to simultaneous permutation of

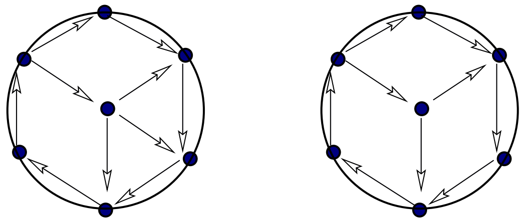


FIGURE 15. The aperiodicity conditions of Perron-Frobenius theory: an aperiodic digraph (left) and a periodic digraph (right).

its rows and columns, a cyclic block structure

$$\begin{pmatrix} 0 & \boxed{A_{0,1}} & 0 & \cdots & 0 \\ 0 & 0 & \boxed{A_{1,2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boxed{A_{d-2,d-1}} \\ \boxed{A_{d-1,0}} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where the blocks $A_{i,i+1}$ are reflexes of the connectivity between V_i and V_{i+1} in (78). In the case of a period d, the matrix A^d admits a diagonal square block decomposition where each of its diagonal block is aperiodic (and of a smaller dimension than the original matrix). Then, the matrices $A^{\nu d}$ can be analysed block by block, and the analysis reduces to the aperiodic case. Similarly for powers $A^{\nu d+r}$ for any fixed r as ν varies. In other words, *the irreducible periodic case with period $d \geq 2$ can always be reduced to a collection of d irreducible aperiodic subproblems.* For this reason, we normally assume in statements both an irreducibility condition and an aperiodicity condition.

Perron-Frobenius Theorem. *Let A be a matrix with* nonnegative elements *that is assumed to be* irreducible. *The eigenvalues of A can be ordered in such a way that*

$$\lambda_1 = |\lambda_2| = \cdots = |\lambda_d| > |\lambda_{d+1}| \geq |\lambda_{d+2}| \geq \cdots ,$$

and each of the dominant eigenvalues is simple, with λ_1 being positive. Furthermore, the quantity d is precisely equal to the period of the dependency graph. In particular, in the aperiodic case $d = 1$, *there is unicity of the dominant eigenvalue. In the* periodic case $d \geq 2$, *the whole spectrum is invariant under the set of transformations*

$$\lambda \mapsto \lambda e^{2ij\pi/d}, \qquad j = 0, 1, \ldots, d-1.$$

For an idea of the proof techniques involved, see APPENDIX B: *Perron-Frobenius theory of nonnegative matrices*, p. 679.

▷ **29.** *Sufficient conditions for unicity of a dominant eigenvalue.* Any one of the following conditions suffices to guarantee aperiodicity:

(i) T has (strictly) positive entries;
(ii) some power T^s has (strictly) positive entries;
(iii) T is irreducible and at least one diagonal element of T is nonzero;
(iv) T is irreducible and the dependency graph of T is such that there exist two circuits (closed paths) rooted at the same vertex that are of relatively prime lengths.

Any such condition guarantees the existence of a unique dominant eigenvalue of$T$. $\lhd$

The importance of Perron-Frobenius theory stems from the fact that uniqueness of the dominant eigenvalue is usually related to a host of analytic properties of generating functions as well as probabilistic properties of structures.

THEOREM V.7 (Asymptotics of paths in graphs). *Consider the matrix*

$$F(z) = (I - zT)^{-1},$$

where T is a scalar nonnegative matrix, in particular, the adjacency matrix $\mathbf{G}$ *of a graph G equipped with positive weights $g_{a,b} = \gamma_{a,b} \in \mathbb{R}_{>0}$.*

Assume that T is irreducible and aperiodic. Then all entries $F_{i,j}(z)$ of $F(z)$ have the same radius of convergence ρ, which can be defined in two equivalent ways:

(i) *as $\rho = \lambda_1^{-1}$ with λ_1 the largest positive eigenvalue of T;*
(ii) *as the smallest root of the determinantal equation:* $\det(I - zT) = 0$.

Furthermore, the point $\rho = \lambda_1^{-1}$ is a simple pole the unique dominant singularity of each $F_{i,j}(z)$. Then, each $F_{i,j}$ satisfies

$$[z^n]F_{i,j}(z) = \varphi_{i.j}\lambda_1^n + O(\Lambda^n), \qquad 0 \leq \Lambda < \lambda_1,$$

for some $\varphi_{i,j} > 0$.

PROOF. Define first $\rho = 1/\lambda_1$, where λ_1 is the eigenvalue of T of largest modulus that is guaranteed to be simple by assumption of irreducibility and by Perron-Frobenius properties. Next, the relations induced by $F = I + zTF$, namely,

$$F_{i,j}(z) = \delta_{i,j} + z \sum_k T_{i,k} F_{k,j}(z),$$

together with positivity and irreducibility entail that the $F_{i,j}(z)$ must all have the same radius of convergence r. Indeed, each F_{ij} depends positively on all the other ones (by irreducibility) so that any infinite value of an entry in the system must propagate to all the other ones.

The characteristic polynomial

$$\Delta(z) = \det(I - zT),$$

has roots that are inverses of the eigenvalues of T and $\rho = 1/\lambda_1$ is smallest in modulus. Thus, since Δ is the common denominator to all the $F_{i,j}(z)$, poles of any $F_{i,j}(z)$ can only be included in the set of zeros of this determinant, so that the inequality $r \geq \rho$ holds.

It remains to exclude the possibility $r > \rho$, which means that no cancellations with the numerator of (75) can occur at $z = \rho$. The argument relies on finding a positive combination of some of the $F_{i,j}$ that *must* be singular at ρ. We offer two proofs, each of interest in its own right: one (a) is conveniently based on the Jacobi trace formula, the other (b) necessitates supplementary Perron–Frobenius properties.

(a) By Jacobi's trace formula (76), p. 324, we have (for z small enough)

$$\begin{aligned}\operatorname{Tr}\log(I-zT)^{-1} &= \sum_i \sum_{n\geq 1} F_n^{\langle i,i\rangle}\frac{z^n}{n} \\ &= \log\det(I-zT)^{-1},\end{aligned}$$

where the first line results from expansion of the logarithm and the second line is an instance of the trace formula. Thus, by differentiation, the sum $\sum_i F_{i,i}(z)$ is seen to have a pole at $\rho = 1/\lambda_1$ and we have established that $r = \rho = \lambda_1^{-1}$.

(b) Alternatively, let v_1 be the eigenvector of T corresponding to λ_1. Perron-Frobenius theory also teaches us that, under the irreducibility and aperiodicity conditions, the vector v_1 has all its coordinates that are nonzero. Then the quantity

$$(1-zT)^{-1}v_1 = \frac{1}{1-z\lambda_1}v_1$$

is certainly singular at $1/\lambda_1$. But it is also a linear combination of the $F_{i,j}$'s. Thus at least one of the entries of F (hence all of them by the discussion above) must be singular at $\rho = 1/\lambda_1$. Therefore, we have again $r = \rho$.

Finally, under the additional assumption that T is aperiodic, Perron-Frobenius theory grants us that $\rho = 1/\lambda_1$ is well-separated in modulus from all other singularities of F. □

Several of these arguments will be recycled when we discuss the harder problem of analysing coefficients of positive algebraic functions in Chapter VII.

Profiles. Let us assume again that positive weights are assigned to the edges of G. In other words, the quantities $g_{a,b}$ in (73) have positive values. If the resulting matrix is irreducible and aperiodic, then Perron-Frobenius theory applies, resulting in Theorem V.7. A host of probabilistic properties of paths follow after a certain "residue matrix" has been calculated.

LEMMA V.1 (Iteration of Perron-Frobenius matrices). *Set $M(z) = (I - z\mathbf{G})^{-1}$ where $\mathbf{G}$ has nonnegative entries, is irreducible, and is aperiodic. Let λ_1 be the dominant eigenvalue of $\mathbf{G}$. Then the "residue" matrix R such that*

$$(I - z\mathbf{G})^{-1} = \frac{R}{1-z\lambda_1} + O(1) \qquad (z \to \lambda_1^{-1}) \tag{79}$$

has entries given by ($\langle x, y\rangle$ represents a scalar product)

$$R_{ij} = \frac{r_i \ell_j}{\langle r, \ell\rangle},$$

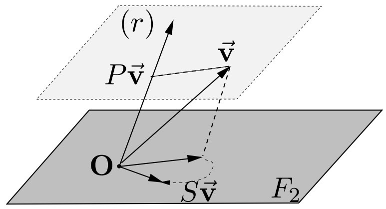
where r and ℓ are respectively right and left eigenvectors of $\mathbf{G}$ corresponding to the eigenvalue λ_1.

▷ **30.** *Proof of Lemma V.1.* Let $\mathcal{E} = \mathbb{C}^m$ be the ambient space, where m is the dimension of $\mathbf{G}$. There exists a direct sum decomposition $\mathcal{E} = \mathcal{F}_1 + \mathcal{F}_2$ where $\mathcal{F}_1$ is the 1-dimensional eigenspace generated by the eigenvector (r) corresponding to eigenvalue λ_1 and $\mathcal{F}_2$ is the supplementary space which is the direct sum of characteristic spaces corresponding to the other eigenvalues $\lambda_2, \ldots$. (For the purposes of the present discussion, one may freely think of the matrix as diagonalizable, with $\mathcal{F}_2$ the union of eigenspaces associated to $\lambda_2, \ldots$.) Then $\mathbf{G}$ as a linear operator acting on $\mathcal{F}$ admits the decomposition

$$\mathbf{G} = \lambda_1 P + S,$$

where P is the projector on $\mathcal{F}_1$ and S acts on $\mathcal{F}_2$ with spectral radius $|\lambda_2|$, as illustrated by the diagram:

(80)



By standard properties of projections, $P^2 = P$ and $PS = SP = 0$ so that

$$\mathbf{G}^n = \lambda_1^n P + S^n.$$

Consequently, there holds,

$$(I - z\mathbf{G})^{-1} = \sum_{n \geq 0} z^n \lambda_1^n P + z^n S^n = \frac{P}{1 - \lambda_1 z} + (I - zS)^{-1}. \tag{81}$$

Thus, the residue matrix R coincides with the projector P.

Now, for any vector w, by general properties of projections, one has ($R \equiv P$):

$$Rw = c(w)r,$$

for some coefficient $c(w)$. Application of this to each of the base vectors e_j (i.e., $e_j = (\delta_{j1}, \ldots, \delta_{jd})^t$) shows that the matrix R has each of its *columns* proportional to the eigenvector r. A similar reasoning with the transpose $\mathbf{G}^t$ of $\mathbf{G}$ and the associated residue matrix R^t shows that the matrix R has each of its *rows* proportional to the eigenvector ℓ. In other words, for some constant γ, one must have

$$R_{i,j} = \gamma \ell_j r_i.$$

The normalization constant γ is itself determined by $\ell R r = \langle \ell, r \rangle$.

We finally observe that a full expansion can be obtained:

$$(I - z\mathbf{G})^{-1} = \frac{R}{1 - \lambda_1 z} + \sum_{k \geq 0} R_k \left(z - \lambda_1^{-1}\right)^k, \quad R_k := S^k (I - \lambda_1^{-1} S)^{-k-1}. \tag{82}$$

The proof also reveals that one needs to solve one polynomial equation for determining λ_1, and then the other quantities in (82) are all obtained by inverting matrices in the field of constants extended by the algebraic quantity λ_1. (Numerical procedures are likely to be used instead for large matrices.) $\lhd$

Equipped with the lemma, we can now state:

THEOREM V.8 (Profiles of paths in graphs). *Let $\mathbf{G}$ be a nonnegative matrix associated to a weighted digraph G, assumed to be irreducible and aperiodic. Let ℓ, r be respectively the left and right vectors corresponding to the dominant (Perron-Frobenius) eigenvalue λ_1. Consider the collection $\mathcal{P}_{a,b}$ of (weighted) paths with fixed origin a and final destination b. Then, the number of traversals of edge (s,t) in a random element of $\mathcal{P}_{a,b}$ has mean*

$$\tau_{s,t} n + O(1) \qquad \textit{where} \quad \tau_{s,t} := \frac{\ell_s g_{s,t} r_t}{\lambda_1 \langle \ell, r \rangle}, \tag{83}$$

where $\langle x, y \rangle$ represents scalar product.

In other words, a long random path tends to spend asymptotically a fixed (nonzero) fraction of its time traversing any given edge. Accordingly, the number of visits to vertex s is also proportional to n and obtained by summing the expression of (83) according to all the possible values of t.

PROOF. First, the total weight ("number") of paths in $\mathcal{P}_{a,b}$ satisfies

$$[z^n]\left[(I-z\mathbf{G})^{-1}\right]_{a,b} \sim \lambda_1 \frac{r_a \ell_b}{\langle \ell, r\rangle}, \tag{84}$$

as follows from Lemma V.1. Next, introduce the modified matrix $\mathbf{H} = (h_{i,j})$ defined by

$$h_{i,j} = g_{i,j}\, u^{[\![i=s \wedge j=t]\!]}.$$

In other words, we mark each traversal of edge i,j by the variable u. Then, the quantity

$$[z^n]\left[\frac{\partial}{\partial u}(I-z\mathbf{H})^{-1}\Big|_{u=1}\right]_{a,b} \tag{85}$$

represents the total number of traversals of edge (s,t), with weights taken into account. Simple algebra[10] shows that

$$\frac{\partial}{\partial u}(I-z\mathbf{H})^{-1}\Big|_{u=1} = (I-z\mathbf{G})^{-1}\,(zH')\,(I-z\mathbf{G}), \tag{86}$$

where $H' := (\partial_u H)_{u=1}$ has all its entries equal to 0, except for the s,t entry whose value is $g_{s,t}$. By the calculation of the residue matrix in Lemma V.1, the coefficient of (85) is then asymptotic to

$$[z^n]\frac{R_{a,s}}{1-\lambda_1 z} g_{s,t} z \frac{R_{t,b}}{1-\lambda_1 z} \sim \upsilon n \lambda_1^{n-1}, \qquad \upsilon := \frac{r_a \ell_s g_{s,t} r_t \ell_b}{\langle \ell, r\rangle^2}. \tag{87}$$

Comparison of (87) and (84) finally yields the result since the relative error terms are $O(n^{-1})$ in each case. □

Another consequence of this last proof and Equation (84) is that the numbers of paths starting at a and ending at either b or c satisfy

$$\lim_{n\to\infty} \frac{P_{a,b,n}}{P_{a,c,n}} = \frac{\ell_b}{\ell_c}. \tag{88}$$

In other words, the quantity

$$\frac{\ell_b}{\sum_j \ell_j}$$

is the asymptotic probability that a random path with origin fixed at some point a but otherwise unconstrained will end up at point b after a large number of steps. Such properties are strongly evocative of Markov chain theory discussed below in Example 13.

[10] If A is an operator depending on u, one has $\partial_u A^{-1} = A^{-1}(\partial_u A)A^{-1}$, which is a noncommutative generalization of the usual differentiation rule for inverses.

▷ **31.** *Concentration of distribution for the number of passages.* Under the conditions of the theorem, the standard deviation of the number of traversals of a designated node or edge is $O(\sqrt{n})$. Thus in a random long path, the distribution of the number of such traversals is concentrated. [Compared to (86), the calculation of the second moment requires taking a further derivative, which leads to a triple pole. The second moment and the square of the mean, which are each $O(n^2)$, are then found to cancel to main asymptotic order.] ◁

Automata and words. By proposition V.6 (p. 323), the OGF of the language defined by a deterministic finite automaton involves a quasi-inverse $(1 - zT)^{-1}$, where the matrix T is a direct encoding of the automaton's transitions. Corollary V.7 and Lemma V.1 have been precisely custom-tailored for this situation. As is by now usual, we shall allow weights on letters of the alphabet, corresponding to a Bernoulli model on words. We say that an automaton is irreducible (resp. aperiodic) if the underlying graph and the associated matrix are irreducible (resp. aperiodic).

PROPOSITION V.7 (Random words and automata). *Let $\mathcal{L}$ be a language recognized by a deterministic finite automaton A whose graph is irreducible and aperiodic. The number of words of $\mathcal{L}$ satisfies*

$$L_n \sim c\lambda_1^n + O(\Lambda^n),$$

where λ_1 is the dominant (Perron-Frobenius) eigenvalue of the transition matrix of A and c, Λ are real constants with $c > 0$ and $0 \leq \Lambda < \lambda_1$.

In a random word of $\mathcal{L}_n$, the number of traversals of a designated vertex or edge has a mean that is asymptotically linear in n and is given by Theorem V.8.

▷ **32.** *Unambiguous automata.* A nondeterministic finite state automaton is said to be unambiguous if the set of accepting paths for any given words comprises at most one element. The translation into generating function as described above also applies to such automata, even though they are nondeterministic. ◁

V. 5.3. Applications. We now provide a few application of Theorems V.7 and V.8.

— First, two simple applications are discussed. Example 11 studies briefly the case of words that are locally constrained in the sense that certain transitions between letters are forbidden. Example 12 revisits walks on an interval and develops an alternative matrix view of a problem otherwise amenable to continued fraction theory.

— Example 13 makes explicit the way the fundamental theorem of finite Markov chain theory can be derived effortlessly as a consequence of the more general Theorem V.8. Example 14 compares on a simple problem, the devil's staircase, the combinatorial and the Markovian approaches.

— Example 15 comes back to words and develops simple consequence of an important combinatorial construction, that of De Bruijn graphs. This graph is precious in predicting in many cases the *shape* of the asymptotic results that are to be expected when confronted with word problems. Example 16 concludes this section with a brief discussion of special case of words with excluded patterns, thereby leading to a quantitative version of Borges' Theorem.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

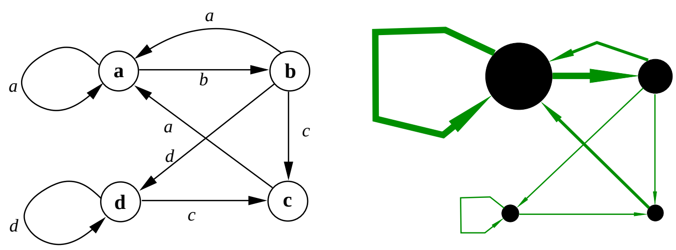


FIGURE 16. Locally constrained words: The transition matrix (T) associated to the forbidden pairs $F = \{ac, ad, bb, cb, cc, cd, da, db\}$, the corresponding automaton, and the graph with widths of vertices and edges drawn in proportion to their asymptotic frequencies.

In all these examples, the counting estimates are of the form $c\lambda_1^n$, while the expectations of parameters of interest have a linear growth.

EXAMPLE 11. *Locally constrained words.* Consider a fixed alphabet $\mathcal{A} = \{a_1, \dots, a_m\}$ and a set $\mathcal{F} \subseteq \mathcal{A}^2$ of forbidden transitions between consecutive letters. The set of words over $\mathcal{A}$ with no such forbidden transition is denoted by $\mathcal{L}$ and is called a locally constrained language. (The particular case where exactly all pairs of equal letters are forbidden corresponds to Smirnov words and has been discussed on p. 249.)

Clearly, the words of $\mathcal{L}$ are recognized by an automaton whose state space is isomorphic to $\mathcal{A}$: state q simply memorizes the fact that the last letter read was a q. The graph of the automaton is then obtained by the collection of allowed transitions $(q, r) \mapsto a$, with $(q, r) \notin \mathcal{F}$. (In other word, the graph of the automaton is the complete graph in which all edges that correspond to forbidden transitions are deleted.) Consequently, the OGF of any locally constrained language is a rational function. Its OGF is given by

$$(1, 1, \dots, 1)(I - zT)^{-1}(1, 1, \dots, 1)^t,$$

where T_{ij} is 0 if $(a_i, a_j) \in \mathcal{F}$ and 1 otherwise. If each letter can follow any other letter in an accepted word, the automaton is irreducible. The graph is aperiodic except in a few degenerate cases (e.g., in the case where the allowed transitions would be $a \to b, c$, $b \to d$, $c \to d$, $d \to a$). Under irreducibility and aperiodicity, the number of words will be $\sim c\lambda_1^{-n}$ and each letter will have on average an asymptotic constant frequency. (See (34) and (35) of Chapter IV for the case of Smirnov words.)

For the example of Figure 16, the alphabet is $\mathcal{A} = \{a, b, c, d\}$. There are eight forbidden transitions and the characteristic polynomial is found to be $\lambda^3(\lambda - 2)$. Thus, one has $\lambda_1 = 2$. The right and left eigenvectors are found to be

$$r = (2, 2, 1, 1)^t, \qquad \ell = (2, 1, 1, 1).$$

Then, the matrix $(\tau_{s,t})$, where $\tau_{s,t}$ represents the asymptotic frequency of transitions from letter s to letter t is found in accordance with Theorem V.8:

$$\Gamma = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{16} & \frac{1}{16} \end{pmatrix}.$$

This means that a random path spends a proportion equal to $\frac{1}{4}$ of its time on a transition between an a and a b, but much less ($\frac{1}{16}$) on transitions between pairs of letters bc, bd, cc, ca. The letter frequencies in a random word of $\mathcal{L}$ are $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$, so that an a is four times more frequent than a c or a d, and so on. See Figure 16 (right) for a rendering.

Various specializations, including multivariate GF's and nonuniform letter models are readily treated by this method. Bertoni *et al.* develop in [**47**] related variance and distribution calculations in the case of the number of occurrences of a symbol in an arbitrary regular language. . . . . END OF EXAMPLE 11. □

EXAMPLE 12. *Walks on the interval revisited.* As a direct illustration, consider the walks associated to the graph $G(5)$ with vertex set $1, \ldots, 5$ and edges being formed of all pairs (i, j) such that $|i - j| \le 1$. The matrix is

$$\mathbf{G}(5) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

The characteristic polynomial factorizes as

$$\chi_{\mathbf{G}(5)}(z) = z(z-1)(z-2)(z^2 - 2z - 2),$$

and its dominant root is $\lambda_1 = 1 + \sqrt{3}$. From there, one finds a left eigenvector (which is also a right eigenvector since the matrix is symmetric):

$$r = \ell^t = (1, \sqrt{3}, 2, \sqrt{3}, 1).$$

Thus a random path (with the uniform distribution over all paths corresponding to the weights being equal to 1) visits nodes $1, \ldots, 5$ with frequencies proportional to

$$1, \quad 1.732, \quad 2, \quad 1.732, \quad 1,$$

implying that the non-extremal nodes are visited more often—such nodes have higher degrees of freedom, so that there tend to be more paths that traverse them.

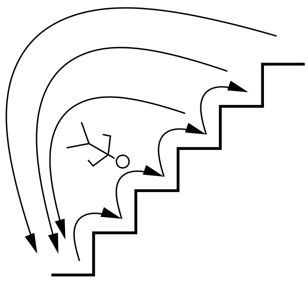
In fact, this example has structure. For instance, the corresponding graph $G(11)$ defined by an interval of length 10, leads to a matrix with a highly factorable characteristic polynomial

$$\chi_{\mathbf{G}(11)} = z\,(z-1)\,(z-2)\left(z^2 - 2\,z - 2\right)\left(z^2 - 2\,z - 1\right)\left(z^4 - 4\,z^3 + 2\,z^2 + 4\,z - 2\right).$$

The reader may have recognized a particular case of lattice paths which resort to the theory exposed in Section V. 3. Indeed, according to Proposition V.3, the OGF of paths from vertex 1 to vertex 1 in the graph $G(k)$ with vertex set $\{1, \ldots, k\}$ is given by the continued fraction

$$\cfrac{1}{1 - z - \cfrac{z^2}{1 - z - \cfrac{z^2}{\ddots \over 1 - z - \cfrac{z^2}{1-z}}}}.$$

(The number of fraction bars is k.) From this it can be shown that the characteristic polynomial of G is an elementary variant of the Fibonacci–Chebyshev polynomial of Example 6. The analysis based on Theorem V.8 is simpler, albeit more rudimentary, as it only provides a first-order asymptotic solution to the problem.



$$\mathbf{G} = \begin{pmatrix} 1&1&0&0&0&0\\ 1&0&1&0&0&0\\ 1&0&0&1&0&0\\ 1&0&0&0&1&0\\ 1&0&0&0&0&1\\ 1&0&0&0&0&0 \end{pmatrix} \qquad \widetilde{\mathbf{G}} = \begin{pmatrix} \frac12&\frac12&0&0&0&0\\ \frac12&0&\frac12&0&0&0\\ \frac12&0&0&\frac12&0&0\\ \frac12&0&0&0&\frac12&0\\ \frac12&0&0&0&0&\frac12\\ 1&0&0&0&0&0 \end{pmatrix}$$

FIGURE 17. The devil's staircase ($m = 6$) and the two matrices that can model it.

This example is typical: in many cases combinatorial problems have some amount of regularity. In such situations, all the resources of linear algebra are available, including the vast body of knowledge gathered over years on calculations of structured determinants; see for instance Krattenthaler's survey [**281**] and the book [**424**]. .... END OF EXAMPLE 12. □

EXAMPLE 13. *Elementary theory of Markov chains.* Consider the case where the row sums of matrix $\mathbf{G}$ are all equal to 1, that is, $\sum_j g_{i,j} = 1$. Such a matrix is called a *stochastic matrix*. The quantity $g_{i,j}$ can then be interpreted as the probability of leaving state i for state j, assuming one is in state i. Assume that the matrix $\mathbf{G}$ is irreducible and aperiodic. Clearly, the matrix $\mathbf{G}$ admits the column vector $r = (1, 1, \ldots, 1)^t$ as a right eigenvector corresponding to the dominant eigenvalue $\lambda_1 = 1$. The left eigenvector ℓ normalized so that its elements sum to 1 is called the (row) vector of stationary probabilities. It must be determined by linear algebra and it involves finding an element of the kernel of matrix $I - \mathbf{G}$, which can be done in a standard way.

Application of Theorem V.8 and Equation (84) shows immediately the following:

PROPOSITION V.8 (Stationary probabilities of Markov chains). *Consider a weighted graph corresponding to a stochastic matrix $\mathbf{G}$ which is irreducible and aperiodic. Let ℓ be the normalized left eigenvector corresponding to the eigenvalue 1. A random (weighted) path of length n with fixed origin and destination visits node s a mean number of times asymptotic to $\ell_s n$ and traverses edge (s,t) a mean number of times asymptotic to $\ell_s g_{s,t} n$. A random path of length n with fixed origin ends at vertex s with probability asymptotic to ℓ_s.*

This first-order asymptotic property certainly constitutes the most fundamental result in the theory of finite Markov chains. .... END OF EXAMPLE 13. □

The next example illustrates an elementary technique often employed in calculations of eigenvalues and eigenvectors. It presupposes that the matrix to be analysed can be reduced to a sparse form and has a regular enough structure.

EXAMPLE 14. *The devil's staircase.* You live in a house that has a staircase with m steps. You come back home a bit loaded and at each second, you can either succeed in climbing a step or fall back all the way down. On the last step, you always stumble and fall back down (Figure 17). Where are you likely to be found at time n?

Precisely, two slightly different models correspond to this informally stated problem. The probabilistic model views it as a Markov chain with equally likely possibilities at each step and is reflected my matrix $\widetilde{\mathbf{G}}$ in Figure 17. The combinatorial model just assumes all possible evolutions ("histories") of the system as equally likely and it corresponds to matrix $\mathbf{G}$. We opt here for the latter, keeping in mind that the same method basically applies to both cases.

We first write down the constraints expressing the joint properties of an eigenvalue λ and its right eigenvector $x = (x_1, \dots, x_m)^t$. The equations corresponding to $(\lambda I - \mathbf{G})x = 0$ are formed of a first batch of $m-1$ relations,

$$(\lambda - 1)x_1 - x_2 = 0, \quad -x_1 + \lambda x_2 - x_3 = 0, \quad \cdots, -x_1 + \lambda x_{m_1} - x_m = 0, \tag{89}$$

together with the additional relation (one cannot go higher than the last step):

$$-x_1 + \lambda x_m = 0. \tag{90}$$

The solution to (89) is readily found by pulling out successively $x_2, \dots, x_m$ as functions of x_1:

$$x_2 = (\lambda - 1)x_1, \quad x_3 = (\lambda^2 - \lambda - 1)x_1, \quad \cdots, \quad x_m = (\lambda^m - \lambda^{m-1} - \cdots - 1)x_1. \tag{91}$$

Combined with the special relation (90), this last relation shows that λ must satisfy the equation

$$1 - 2\lambda^m + \lambda^{m+1}. \tag{92}$$

Let λ_1 be the largest positive root of this equation, existence and dominance being guaranteed by Perron-Frobenius properties. Note that the quantity $\rho := 1/\lambda_1$ satisfies the characteristic equation

$$1 - 2\rho + \rho^{m+1} = 0,$$

already encountered when discussing longest runs in words; the discussion of Example 2 then grants us the existence of an isolated ρ near $\frac{1}{2}$, hence the fact that λ_1 is slightly less than 2.

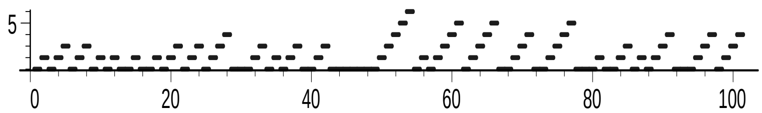
Similar devices yield the left eigenvector $y = (y_1, \dots, y_m)$. It is found easily that y_j must be proportional to λ_1^{-j}. We thus obtain from Theorem V.8 and Equation (88): *The probability of being in state j (i.e., being on step j of the stair) at time n tends to the limit*

$$\varpi_j = \gamma \lambda_1^{-j}$$

where λ_1 is the root near 2 of the polynomial (92) *and the normalization constant γ is determined by $\sum_j \varpi_j = 1$.* In other words, the distribution of the altitude at time n is a truncated geometric distribution with parameter $1/\lambda_1$. For instance, $m = 6$ leads to $\lambda_1 = 1.98358$, and the asymptotic probabilities of being in states $1, \dots, 6$ are

$$0.50413, \quad 0.25415, \quad 0.12812, \quad 0.06459, \quad 0.03256, \quad 0.01641, \tag{93}$$

exhibiting a clear geometric decay. Here is the simulation of a random trajectory for $n = 100$:



In this case, the frequencies observed are $0.44,\ 0.26,\ 0.17,\ 0.08,\ 0.04,\ 0.01$, pretty much in agreement with what is expected.

Finally, the similarity with the longest run problem is easily explained. Let u and d be letters representing steps upwards and downwards respectively. The set of paths from state 1 to state 1 is described by the regular expression

$$\mathcal{P}_{1,1} = \left(d + ud + \cdots + u^{m-1}d\right)^\star,$$

corresponding to the generating function

$$P_{1,1}(z) = \frac{1}{1 - z - z^2 - \cdots - z^m},$$

a variant of the OGF of words without m-runs of the letter u, which also corresponds to the enumeration of compositions with summands $\leq m$. The case of the probabilistic transition matrix $\widetilde{\mathbf{G}}$ is left as an exercise to the reader. .... END OF EXAMPLE 14. □

EXAMPLE 15. *De Bruijn graphs.* Two thieves want to break into a house whose entrance is protected by digital lock with an unknown four-digit code. As soon as the four digits of the code are typed consecutively, the gate opens. The first thief proposes to try in order all the four-digit sequences, resulting in as much as 40,000 key strokes in the worst-case. The second thief, who is a mathematician, says he can try *all* four-digit combinations with only 10,003 key strokes. What is the mathematician's trade secret?

Clearly certain optimizations are possible: for instance, for an alphabet of cardinality 2 and codes of 2 letters, the sequence 00110 is better than the naïve one, 00 01 10 11, which is redundant; a few more attempts will lead to an optimal solution for 3–digit codes that has length 11 (rather than 24), for instance,

$$0001110100.$$

The general question is then: How far can one go and how to construct such sequences?

Fix an alphabet of cardinality m. A sequence that contains as factors (contiguous blocks) *all* the k letter words is called a *de Bruijn sequence*. Clearly, its length must be at least $\delta(m,k) = m^k + k - 1$, as it must have at least m^k positions at distance at least k from the end. A sequence of smallest possible length $\delta(m,k)$ is called a *minimal* de Bruijn sequence. Such sequences were discovered by N. G. de Bruijn [**91**] in 1946, in response to a question coming from electrical engineering, where all possible reactions of a device presented as a black box must be tested at minimal cost. We shall expose here the case of a binary alphabet, $m = 2$, the generalization to $m > 2$ being obvious.

Let $\ell = k - 1$ and consider the automaton $\mathcal{B}_\ell$ that memorizes the last block of length ℓ read when scanning the input text from left to right. A state is thus assimilated to a string of length ℓ and the total number of states is 2^ℓ. The transitions are easily calculated: let $q \in \{0,1\}^\ell$ be a state and let $\sigma(w)$ be the function that shifts all letters of a word w one position to the left, dropping the first letter of w in the process (thus σ maps $\{0,1\}^\ell$ to $\{0,1\}^{\ell-1}$); the transitions are

$$q \overset{0}{\mapsto} \sigma(q)\mathsf{0}, \qquad q \overset{1}{\mapsto} \sigma(q)\mathsf{1}.$$

If one further interprets a state q as the integer in the interval $[0\,..\,2^\ell - 1]$ that it represents, then the transition matrix assumes a remarkably simple form:

$$T_{i,j} = [\![(j \equiv 2i \bmod 2^\ell) \text{ or } (j \equiv 2i + 1 \bmod 2^\ell)]\!].$$

See Figure 18 for a rendering borrowed from [**182**].

Combinatorially, the de Bruijn graph is such that each node has indegree 2 and outdegree 2. By a well known theorem going back to Euler: *A necessary and sufficient condition for an undirected connected graph to have an Eulerian circuit (that is, a closed path that traverses each vertex exactly once) is that every node has even degree. For a strongly connected digraph, the condition is that each node has an outdegree equal to its indegree.* This last condition is obviously satisfied here. Take an Eulerian circuit starting and ending at node 0^ℓ; its length is $2^{\ell+1} = 2^k$. Then, clearly, the sequence of edge labels encountered when prefixed with the word $0^{k-1} = 0^\ell$ constitutes a minimal de Bruijn sequence. In general, the argument gives a de Brujin sequence with minimal length $m^k + k - 1$. Et voilà! The trade secret of the thief-mathematician is exposed.

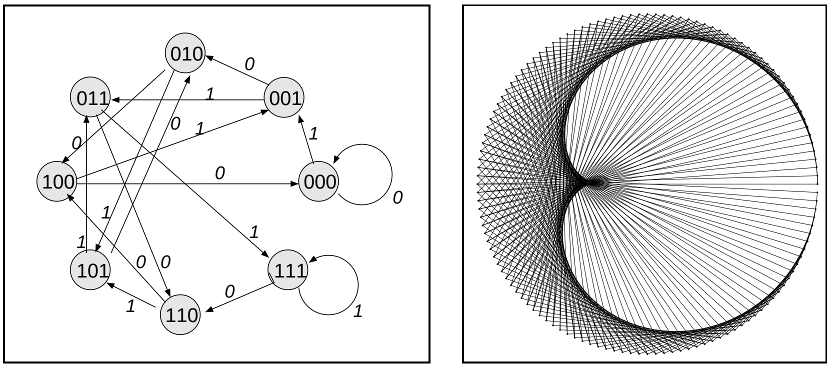


FIGURE 18. The de Bruijn graph: (left) $\ell = 3$; (right) $\ell = 7$.

Back to enumeration. The de Bruijn matrix is irreducible since a path labelled by sufficiently many zeros always leads any state to the state 0^ℓ, while a path ending with the letters of $w \in \{0,1\}^\ell$ leads to state w. The matrix is aperiodic since it has a loop on states 0^ℓ and 1^ℓ. Thus, by Perron Frobenius properties, it has a unique dominant eigenvalue, and it is not hard to check that its value is $\lambda_1 = 2$, corresponding to the right eigenvector $(1, 1, \dots, 1)^t$. If one fixes a *pattern* $w \in \{0,1\}^\ell$, Theorem V.8 yields back the known fact that a random word contains on average $\sim \frac{n}{2^\ell}$ occurrences of pattern w, while Note 31 further implies that the distribution of the number of occurrences is concentrated around the mean, as the variance is $O(n)$. The de Bruijn graph may be used to quantify many properties of occurrences of patterns in random words: see for instance [**33, 161, 182**]. .... END OF EXAMPLE 15. $\square$

EXAMPLE 16. *Words with excluded patterns.* Fix a finite set of patterns $\Omega = \{w_1, \dots, w_r\}$, where each w_j is a word of $\mathcal{A}^\star$. The language $\mathcal{E} \equiv \mathcal{E}^\Omega$ of words that contain no factor in Ω is described by the extended regular expression

$$\mathcal{E} = \mathcal{A}^\star \setminus \bigcup_{j=1}^{r} (A^\star w_j A^\star),$$

which constitutes a concise but highly ambiguous description. By closure properties of regular languages, $\mathcal{E}$ is itself regular and there must exist a deterministic automaton that recognizes it.

An automaton recognizing $\mathcal{E}$ can be constructed starting from the de Bruijn automaton of index $k = -1 + \max |w_j|$ and deleting all the vertices and edges that correspond to a word of Ω. Precisely, vertex q is deleted whenever q contains a factor in Ω; the transition (edge) from q associated with letter α gets deleted whenever the word $q\alpha$ contains a factor in Ω. The pruned de Bruijn automaton, call it $\mathcal{B}_k^\circ$, accepts all words of $0^k\mathcal{E}$, when it is equipped with the initial state 0^k and all states are final. Thus, the OGF $E(z)$ is in all cases a rational function.

The matrix of $\mathcal{B}_k^\circ$ is the matrix of $\mathcal{B}_k$ with some nonzero entries replaced by 0. Assume that $\mathcal{B}_k^\circ$ is irreducible. This assumption only eliminates a few pathological cases (e.g., $\Omega = \{01\}$ on the alphabet $\{0,1\}$). Then, the matrix of $\mathcal{B}_k^\circ$ admits a simple Perron-Frobenius eigenvalue λ_1. By domination properties ($\Omega \neq \emptyset$), we must have $\lambda_1 < m$, where m is the cardinality of the alphabet. Aperiodicity is automatically granted. We then get by a purely qualitative argument: *The number of words of length n excluding patterns from the finite set Ω is, under the assumption of irreducibility, asymptotic to $c\lambda_1^n$, for some $c > 0$ and $\lambda_1 < \|\mathcal{A}\|$.* This gives

Transfer Matrix Method. Let $\mathcal{C}$ be a combinatorial class to be enumerated.
— Determine a collection $\mathcal{C}_1, \mathcal{C}_2, \ldots, \mathcal{C}_m$ of classes, with $\mathcal{C}_1 \equiv \mathcal{C}$ such that the following system of equation holds:

$$\mathcal{C}_j = \sum_{j \in \{1,2,\ldots,m\}} \Omega_{j,k} \mathcal{C}_k + \mathcal{I}_j, \qquad j = 1, 2, \ldots, m, \tag{94}$$

where each $\Omega_{j,k}$ and each $\mathcal{I}_j$ is a finite class.
— The OGF $C(z) \equiv C_0(z)$ is then given by the solution of the linear system

$$C_j(z) = \sum_j \Omega_{j,k}(z) C_k(z) + I_j(z), \qquad j = 0, 1, \ldots, m,$$

where $\Omega_{j,k}(z)$ and $I_j(z)$ are the generating polynomials of $\Omega_{j,k}$ and $\mathcal{I}_j$, respectively. Accordingly, $C(z)$ is a linear combination of entries of the quasi-inverse matrix $(I - \Omega(z))^{-1}$.

FIGURE 19. A Summary of the basic Transfer Matrix Method.

us in a simple manner a strong version of what has been earlier nicknamed "Borges's Theorem" (Note 32, p. 58): *Almost every sufficiently long text contains* all *patterns of some predetermined length* ℓ.

The construction of a pruned automaton is clearly a generalization of the case of words obeying local constraints in Example 11 above. .... END OF EXAMPLE 16. □

▷ **33.** *Words with excluded patterns and digital trees.* Let S be a finite set of words. An automaton recognizing S, considered as a finite language, can be constructed as a tree. The tree obtained is akin to the classical *digital tree* or *trie* that serves as a data structure for maintaining dictionaries [**269**].

A modification of the construction yields an automaton of size linear in the total number of characters that appear in words of S. [Hint. The construction can be based on the Aho–Corasick automaton [**4**]). ◁

V. 6. Transfer matrix models

There exists a cluster of applications of rational functions to problems that are naturally described as paths in digraphs, but with edges that may be of different sizes. In physics, such models lie at the heart of what is known as the "transfer matrix method". Technically, the theory is a simple extension of the standard case of paths in graphs developed in the previous section to which it reduces when all edges have the same length. Its main interest lies in its expressiveness as regards a number of combinatorial problems, including trees of bounded width, models of self-avoiding walks, and certain constrained permutation problems.

V. 6.1. Combinatorial aspects. The transfer matrix method constitutes a variant of the modelling by deterministic automata and by paths in standard graphs. The general framework is summarized in Figure 19.

Usually, when setting up such a system, one has to invent a finite collection of properties ("states") describing the $\mathcal{C}_j$, which are of the same nature as the original class $\mathcal{C}$. The combinatorial system (94) can be visualized as a graph with the objects of the $\Omega_{j,k}$ classes attached to edges ("transitions between states") that are generally of different sizes.

DEFINITION V.7. *Given a directed multigraph G with vertex set V and edge set E, a* size function *on G is any function $\sigma : E \to \mathbb{Z}_{\geq 1}$. A* sized graph *is a pair (G, σ), where σ is a size function.*

Paths are defined in the same way as in Section V. 5. The *length* of a path is, as usual, the number of edges it comprises; the *size* of a path is defined to be the sum of the sizes of its edges. Like in the basic case treated in the previous section, we also allow edges to carry positive weights (multiplicities, probability coefficients), the *weight* of a path being the product of the weights of its edges.

DEFINITION V.8. *A matrix $T(z)$ is a* transfer matrix *if each of its entries is a polynomial in z with nonnegative coefficients. A transfer matrix $M(z)$ is said to be* proper *if $T(0)$ is nilpotent, that is, $T(0)^r = 0$ for some $r \geq 1$.*

Examples of transfer matrices are

$$z \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}, \quad \begin{pmatrix} 0 & 1 \\ z^3 & z + z^2 \end{pmatrix},$$

and both are proper. For the graphs and automata considered in Section V. 5, all edges were taken to be of unit size. In that case, the associated (weighted) adjacency matrices are invariably of the form $T(z) = zS$, with S a scalar matrix having nonnegative entries, and thus are very particular cases of proper transfer matrices.

Given a sized graph G equipped with weight function $w : E \to \mathbb{R}_{>0}$ (with $w(e) \equiv 1$ in the pure enumerative case), we can associate to it a transfer matrix $T(z)$ as follows:

$$T_{a,b}(z) = \sum_{e \in \mathrm{Edge}(a,b)} w(e) z^{|e|}. \tag{95}$$

There, $\mathrm{Edge}(a, b)$ represents the set of all edges connecting a to b; $w(e)$ and $|e| \equiv \sigma(e)$ represent respectively the weight and the size of edge e. The matrix $T(z)$ whose a, b-entry is the polynomial $T_{a,b}(z)$, as given in (95), is called the *transfer matrix* of the (weighted, sized) graph. Clearly, the transfer matrix of a sized graph is always proper. Since $T(z)^m$ describes all paths in the graph with z marking size, the proof techniques of Proposition V.6 (p. 323) immediately provide:

PROPOSITION V.9. *Given a sized graph with associated transfer matrix $T(z)$, the OGF $F^{\langle i,j \rangle}(z)$ of the set of paths from i to j, where z marks size, is the entry i, j of the matrix $(I - T(z))^{-1}$:*

$$F^{\langle i,j \rangle}(z) = \left((I - T(z))^{-1}\right)\Big|_{i,j}.$$

V. 6.2. Analytic aspects. In order to apply the general results from the previous section to transfer matrices, we must first take note of an easy reduction of transfer matrices to the standard case of paths in graphs where all edges have size 1.

Given a sized graph G, one can build as follows a standard graph $\widehat{G}$ where all edges of $\widehat{G}$ have unit size. The set of vertices of $\widehat{G}$ is the set of vertices of G augmented by additional vertices called *relay nodes*. For each edge e of size $\sigma(e) = m$ in G, introduce $m - 1$ additional relay nodes and connect these in $\widehat{G}$ by a simple path from

a to b, with edges all of size 1. Here is for instance the transcription of an edge of length 4 in G by means of three relay nodes in $\widehat{G}$:

Clearly, the vertices of G are a subset of the vertices of $\widehat{G}$ and all paths of G correspond to paths of $\widehat{G}$. Let $\widehat{T}$ be the (scalar) adjacency matrix of G. Then, the quasi-inverse $(I - z\widehat{T})^{-1}$ describes all the paths in G, with size taken into account, in the sense that the entry of index (i, j) in this quasi-inverse is the OGF of paths from node numbered i to node numbered j in the sized graph G.

This construction permits us to apply the main results of Section V. 5 to transfer matrices and sized graphs. Let us say that the sized graph G and its transfer matrix $T(z)$ are *irreducible* (respectively *aperiodic*) if $\widehat{G}$ and $\widehat{T}$ are irreducible (respectively aperiodic). We can then transcribe immediately Theorems V.7 and V.8 as follows:

— *Enumeration*. If the sized graph G is irreducible and aperiodic, then there exist a computable constant λ_1 and numbers $\varphi_{i,j}$ such that the OGF of paths from i to j in G satisfies

$$[z^n]F_{i,j}(z) = \varphi_{i,j}\lambda_1^n + O(\Lambda^n), \qquad 0 \le \Lambda < \lambda_1. \tag{96}$$

— *Profiles.* In a random path from a to b of large size, the number of occurrences of a designated edge (s, t) is asymptotically

$$\varpi_{s,t} n + O(1), \tag{97}$$

for a computable constant $\varpi_{s,t}$.

Thus, on general grounds, the behaviour of paths is predictable. The notes below explore some further properties that make it possible to operate directly with the transfer matrix and the sized graph, without necessitating the explicit construction of $\widehat{T}$ and $\widehat{G}$.

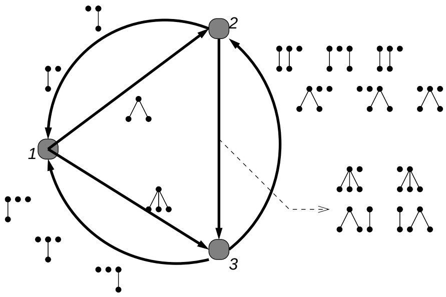
$\triangleright$ **34.** *Irreducibility for sized graphs.* The sized graph G is irreducible (in the sense above) if and only if the graph G_1 where all edges of G are taken to be of size 1 is strongly connected. The transfer matrix $T(z)$ of G is irreducible (in the sense above) if and only if $T(1)$ is irreducible in the usual sense of scalar transfer matrices. $\triangleleft$

$\triangleright$ **35.** *Aperiodicity for sized graphs.* A polynomial $p(z) = \sum_j c_j z^{e_j}$, with every $c_j \neq 0$, is said to be primitive if the quantity $\delta = \gcd(\{e_j\})$ is equal to 1; it is imprimitive otherwise. Equivalently, $p(z)$ is imprimitive iff $p(z) = q(z^\delta)$ for some *bona fide* polynomial q and some $\delta > 1$. An irreducible sized graph is aperiodic (in the sense above) if and only if at least one diagonal entry of some power $T(z)^e$ is a primitive polynomial. Equivalently: there exist two circuits of the same length, rooted at the same vertex, whose sizes, s_1, s_2, satisfy $\gcd(s_1, s_2) = 1$. $\triangleleft$

$\triangleright$ **36.** *Direct determination of the asymptotic growth constant.* Let G be a sized graph assumed to be irreducible and aperiodic. Then, one has $\lambda_1 = 1/\rho$, where ρ is the smallest positive root of

$$\det(I - T(z)) = 0,$$

with $T(z)$ being the transfer matrix of G. $\triangleleft$



$$T^{[4]}(z) = \begin{pmatrix} z & z^2 & z^3 \\ 2z & 3z^2 & 4z^3 \\ 3z & 6z^2 & 10z^3 \end{pmatrix}.$$

FIGURE 20. The sized graph corresponding to general plane trees of width at most 3 and the corresponding transfer matrix. (The transitions from a node to itself are not drawn, for readability.)

V. 6.3. Applications. The quantitative properties summarized by (96) and (97) apply with full strength to classes that are amenable to the transfer matrix method. We shall first illustrate the situation by the width of trees following an early article by Odlyzko and Wilf [**332**], then continue with an example that draws its inspiration from the insightful exposition of domino tilings and generating functions in the book of Graham, Knuth, and Patashnik [**212**], and conclude with an exactly solvable polyomino model.

EXAMPLE 17. *Width of trees.* The width of a tree is defined as the maximal number of nodes that can appear on any layer at a fixed distance from the root. If a tree is drawn in the plane, then width and height can be seen as the horizontal and vertical dimensions of the bounding rectangle. Also, width is an indicator of the complexity of traversing the tree in breadth-first search (by a queue), while height is associated to depth-first search (by a stack).

Transfer matrices are ideally suited to the problem of analysing the number of trees of fixed width. Consider a simple variety of trees $\mathcal{Y}$ corresponding to the equation $Y(z) = z\phi(Y(z))$, where the "generator" ϕ describes the formation of trees. Let $\mathcal{C} := \mathcal{Y}^{[w]}$ be the subclass of trees of width at most w. Such trees are easily built layer by layer. Indeed, with reference to our general description of the transfer matrix method at the beginning of the section, let us introduce a collection of classes $\mathcal{C}_k$, where each $\mathcal{C}_k$ $(k = 1, \ldots, w)$ comprises all trees of width $\leq w$ having exactly k nodes at the deepest level. We then have $\mathcal{C} = \sum_{k=1}^{n} \mathcal{C}_k$ (this is a trivial variant of the case considered in our general description). Thus the states of the transfer matrix model, equivalently the nodes of the size graph, correspond to the number of nodes on the deepest layer of the tree. The transition between configurations $\mathcal{C}_j$ corresponding to state j and configurations $\mathcal{C}_k$ corresponding to state k is effected by grafting in all possible ways a forest of j trees, of total height equal to 1, having k leaves. See Figure 20 for the case of width $w = 3$.

The number of j-forests of depth 1 having k leaves is the quantity

$$t_{j,k} = [u^k]\phi(y)^j.$$

Let T be the $w \times w$ matrix with entry $T_{j,k} = z^k t_{j,k}$. Then, clearly, the quantity $z^i(T^h)_{i,j}$ (with $1 \leq i, j \leq w$) is the number of i-forests of height h and width at most w, having j nodes on level h. Thus, the GF of $\mathcal{Y}$-trees having width at most w is

$$Y^{[w]}(z) = (z, 0, 0, \ldots)(I - T)^{-1}(1, 1, 1, \ldots)^t.$$

For instance, in the case of general Catalan trees, the matrix T has the shape,

$$T^{[w]}(z) = \begin{pmatrix} z\binom{1}{0} & z^2\binom{2}{0} & z^3\binom{3}{0} & z^4\binom{4}{0} \\ z\binom{2}{1} & z^2\binom{3}{1} & z^3\binom{4}{1} & z^4\binom{5}{1} \\ z\binom{3}{2} & z^2\binom{4}{2} & z^3\binom{5}{2} & z^4\binom{6}{2} \\ z\binom{4}{3} & z^2\binom{5}{3} & z^3\binom{6}{3} & z^4\binom{7}{3} \end{pmatrix},$$

for width 4. The analysis of dominant poles provides asymptotic formulae for $[z^n]Y^{[w]}(z)$:

$w = 2$	$w = 3$	$w = 4$	$w = 5$	$w = 6$
$0.0085 \cdot 2.1701^n$	$0.0026 \cdot 2.8050^n$	$0.0012 \cdot 3.1638^n$	$0.0006 \cdot 3.3829^n$	$0.0004 \cdot 3.5259^n$

Irreducibility is granted since all entries in the transfer matrix are nonzero. Aperiodicity derives from aperiodicity of the generator ϕ, as verified by a simple argument (e.g., using Note 35).

PROPOSITION V.10. *The number of trees of width at most w in a simple family of trees satisfies an asymptotic estimate of the form*

$$Y_n^{[w]} = c_w \rho_w^{-n} + O(n),$$

for some computable positive constants c_w, ρ_w.

In addition, the exact distribution of height in trees of size n becomes computable in polynomial time (though with a somewhat high exponent).

The character of these generating functions has not been investigated in detail since the original work [**332**], so that, at the moment, analysis stops there. Fortunately, probability theory can take over. Chassaing and Marckert [**76**] have shown, for Cayley trees, that the width satisfies

$$\mathbb{E}_n(W) = \sqrt{\frac{\pi}{2}} + O\left(n^{1/4}\sqrt{\log n}\right), \qquad \mathbb{P}_n(\sqrt{2}W \le x) \to 1 - \Theta(x),$$

where $\Theta(x)$ is the Theta function defined in (57), p. 307. This answers very precisely an open question of Odlyzko and Wilf [**332**]. The distributional results of [**76**] extend to trees in any simple variety (under mild and natural analytic assumptions on the generator ϕ): see the paper by Chassaing, Marckert, and Yor [**77**], which builds upon earlier results of Drmota and Gittenberger [**112**]. In essence, the conclusion of these works is that the breadth first search traversal of a large tree in a simple variety gives rise to a queue whose size fluctuates asymptotically like a Brownian excursion, and is thus, in a strong sense, of a complexity comparable to depth-first search: trees taken uniformly don't have much of a preference as to the way they may be traversed. . . . . END OF EXAMPLE 17. □

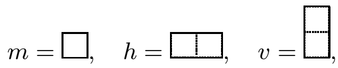
▷ **37.** *A question on width polynomials.* It is unknown whether the following assertion is true. The smallest positive root ρ_k of the denominator of $Y^{[k]}(z)$ satisfies

$$\rho_k = \rho + \frac{c}{k^2} + o(k^{-2}),$$

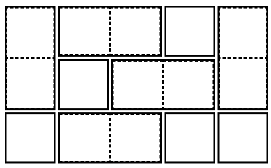
for some $c > 0$. If such an estimate holds together with suitable companion bounds, it would yield a purely analytic proof of the fact that expected width of n–trees is $\Theta(\sqrt{n})$, as well as detailed probability estimates. (The classical theory of Fredholm equations may be useful in this context.) ◁

EXAMPLE 18. *Monomer-dimer tilings of a rectangle.* Suppose one is given pieces that may be one of the three forms: monomers (m) that are 1×1 squares, and dimers that are dominoes, either vertically (v) oriented 1×2, or horizontally (h) oriented 2×1. In how many ways can an $n \times 3$ rectangle be covered completely and without overlap ('tiled') by such pieces?

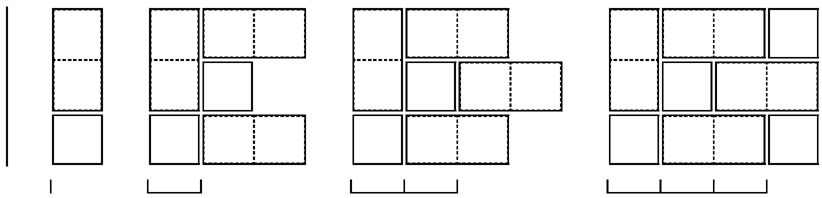
The pieces are thus of the following types,



and here is a particular tiling of a 5×3 rectangle:

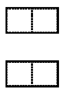


In order to approach this counting problem, one defines a class $\mathcal{C}$ of combinatorial objects called configurations. A configuration relative to an $n \times k$ rectangle is a partial tiling, such that all the first $n - 1$ columns are entirely covered by dominoes while between zero and three unit cells of the last column are covered. Here are for instance, configurations corresponding to the example above.



These diagrams suggest the way configurations can be built by successive addition of dominoes. Starting with the empty rectangle 0×3, one adds at each stage a collection of at most three dominoes in such a way that there is no overlap. This creates a configuration where, like in the example above, the dominoes may not be aligned in a flush-right manner. Continue to add successively dominoes whose left border is at abscissa $1, 2, 3$, etc, in a way that creates no internal "holes".

Depending on the state of filling of their last column, configuration can thus be classified into 8 classes that we may index in binary as $\mathcal{C}_{000}, \ldots, \mathcal{C}_{111}$. For instance $\mathcal{C}_{001}$ represent configurations such that the first two cells (from top to bottom, by convention) are free, while the third one is occupied. Then, a set of rules describes the new type of configuration obtained, when the sweep line is moved one position to the right and dominoes are added. For instance, we have

$\mathcal{C}_{010} \quad \odot$  $\Longrightarrow \quad \mathcal{C}_{101}.$

In this way, one can set up a grammar (resembling a deterministic finite automaton) that expresses all the possible constructions of longer rectangles from shorter ones according to the last layer added. The grammar comprises productions like

$$\begin{aligned} C_{000} \quad = \quad & \epsilon + \underline{mmm}C_{000} + \underline{mv}C_{000} + \underline{vm}C_{000} \\ & + \underline{\cdot mm}C_{100} + \underline{m\cdot m}C_{010} + \underline{mm\cdot}C_{001} + \underline{v\cdot}C_{001} + \underline{\cdot v}C_{100} \\ & + \underline{m\cdot\cdot}C_{011} + \underline{\cdot m\cdot}C_{101} + \underline{\cdot\cdot m}C_{110} + \underline{\cdot\cdot\cdot}C_{111} \,. \end{aligned}$$

In this grammar, a "letter" like $\underline{mv}$ represent the addition of dominoes, in top to bottom order, of types m, v respectively; the letter $\underline{m \cdot m}$ means adding two m-dominoes on the top and on the bottom, etc.

The grammar transforms into a linear system of equations with polynomial coefficients. The substitution,

$$m \mapsto z, \quad h \mapsto z^2, \quad v \mapsto z^2,$$

then gives the generating functions of configurations with z marking the area covered:

$$C_{000}(z) = \frac{(1-2z^3-z^6)(1+z^3-z^6)}{(1+z^3)(1-5z^3-9z^6+9z^9+z^{12}-z^{15})}.$$

In particular, the coefficient $[z^{3n}]C_{000}(z)$ is the number of tilings of an $n \times 3$ rectangle:

$$C_{000}(z) = 1 + 3z^3 + 22z^6 + 131z^9 + 823z^{12} + 5096z^{15} + \cdots .$$

The sequence grows like $c\,\alpha^n$ (for $n \equiv 0 \pmod 3$) where $\alpha \doteq 1.83828$ (α is the cube root of an algebraic number of degree 5). (See [**75**] for a computer algebra session.) On average, for large n, there is a fixed proportion of monomers and the distribution of monomers in a random tiling of a large rectangle is asymptotically normally distributed, as results from the developments of Chapter IX. . . . . END OF EXAMPLE 18. □

The tiling example is a typical illustration of the transfer matrix method as described at the beginning of this section (p. 340). One seeks to enumerate a "special" set of configurations $\mathcal{C}_f$. (In the example above, this is $\mathcal{C}_{000}$ representing complete rectangle coverings.) One determines an extended set of configurations $\mathcal{C}$ (the partial coverings, in the example) such that: (i) $\mathcal{C}$ is partitioned into finitely many classes; (ii) there is a finite set of "actions" that operate on the classes; (iii) size is affected in a well-defined additive way by the actions. The similarity with finite automata is apparent: classes play the rôle of states and actions the rôle of letters.

Often, the method of transfer matrices is used to approximate a hard combinatorial problem that is not known to decompose, the approximation being by means of a family of models of increasing "widths". For instance, the enumeration of the number T_n of tilings of an $n \times n$ square by monomers and dimers remains a famous unsolved problem of statistical physics. Here, transfer matrix methods may be used to solve the $n \times w$ version of the monomer–dimer coverings, in principle at least, for any fixed width w: the result will always be a rational function, though its degree, dictated by the dimension of the transfer matrix, will grow exponentially with w. (The "diagonal" sequence of the $n \times w$ rectangular models corresponds to the square model.) It has been at least determined by computer search that the diagonal sequence T_n starts as (this is *EIS* **A028420**):

$$1,\ 7,\ 131,\ 10012,\ 2810694,\ 2989126727,\ 11945257052321, \ldots .$$

From this and other numerical data, one estimates numerically that $(T_n)^{1/n^2} \to 1.94021\ldots$, but no expression for the constant is known to exist. The difficulty of coping with the finite-width models is that their complexity (as measured , e.g., by the number of states) blows up exponentially with w—such models are best treated by computer algebra; see [**450**]—and no law allowing to take a diagonal is visible. However, the finite width models have the merit of providing at least provable upper and lower bounds on the exponential growth rate of the hard "diagonal problem".

In contrast, for coverings by dimers only, a strong algebraic structure is available and the number of covers of an $n \times n$ square by horizontal and vertical dimers satisfies a beautiful formula originally discovered by Kasteleyn (n even):

$$U_n = 2^{n^2/2} \prod_{j=1}^{n/2} \prod_{k=1}^{n/2} \left(\cos^2 \frac{j\pi}{n+1} + \cos^2 \frac{k\pi}{n+1} \right). \tag{98}$$

This sequence is *EIS* **A004003**,

$$1,\ 2,\ 36,\ 6728,\ 12988816,\ 258584046368,\ 53060477521960000,\ \dots\ .$$

It is elementary to prove from (98) that

$$\lim_{n\to+\infty} (U_n)^{1/n^2} = \exp\left(\frac{1}{\pi} \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)^2} \right) = e^{G/\pi} \doteq 1.33851\dots,$$

where G is Catalan's constant. This means in substance that each cell has a number of degrees of freedoms equivalent to 1.33851. See Percus' monograph [**344**] for proofs of this famous result and Finch's book [**137**, Sec. 5.23] for context and references.

▷ **38.** *Powers of Fibonacci numbers.* Consider the OGFs

$$G(z) := \frac{1}{1-z-z^2} = \sum_{n\ge 0} F_{n+1} z^n, \qquad G^{[k]}(z) := \sum_{n\ge 0} (F_{n+1})^k z^n,$$

where F_n is a Fibonacci number. The OGF of monomer–dimer placements on a $k \times n$ board when only monomers (m) and horizontal dimers (h) are allowed is obviously $G^{[k]}(z)$. On the other hand, it is possible to set up a transfer matrix model with state i $(0 \le i \le k)$ corresponding to i positions of the current column occupied by a previous domino. Consequently,

$$G^{[k]}(z) = \text{coeff}_{k,k} \left((I - zT)^{-1} \right), \qquad \text{where} \quad T_{i,j} = \binom{i}{i+j-k},$$

for $0 \le i, j \le k$. [The denominator of $G^{[k]}(z)$ is otherwise known exactly: see [**268**, Ex. 1.2.8.30].] ◁

▷ **39.** *Tours on chessboards.* The OGF of Hamiltonian tours on an $n \times w$ rectangle is rational (one is allowed to move from any cell to any other vertically or horizontally adjacent cell). The same holds for king's tours and knight's tours. ◁

▷ **40.** *Cover time of graphs.* Given a fixed digraph G assumed to be strongly connected, and a designated start vertex, one travels at random, moving at each time to any neighbour of the current vertex, making choices with equal likelihood. The expectation of the time to visit all the vertices is a rational number that is effectively (though perhaps not efficiently!) computable. [Hint: set up a transfer matrix, a state of which is a subset of vertices representing those vertices that have been already visited. For an interval $[0\,.\,.\,m]$, this can be treated by the dedicated theory of walks on the integer interval, as in Section V. 3; for the complete graph, this is equivalent to the coupon collector problem. Most other cases are "hard" to solve analytically and one has to resort to probabilistic approximations; see Aldous and Fill's forthcoming book [**7**] for a probabilistic approach.] ◁

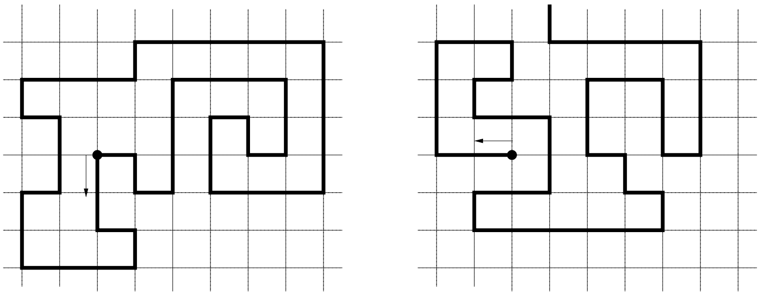
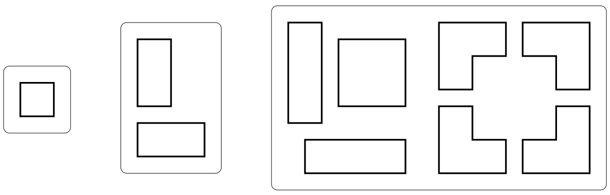


FIGURE 21. A self-avoiding polygon or SAP (left) and a self-avoiding walk or SAW (right).

EXAMPLE 19. *Self-avoiding walks and polygons.* A long standing open problem shared by statistical physics, combinatorics, and probability theory alike is that of quantifying properties of self-avoiding configurations on the square lattice (Figure 21). Here we consider objects that, starting from the origin (the "root"), follow a path, and are solely composed of horizontal and vertical steps of length ± 1. The *self-avoiding walk* or *SAW* can wander but is subject to the condition that it never crosses nor touches itself. The *self-avoiding polygons* or *SAPs*, whose class is denoted by $\mathcal{P}$, are self-avoiding walks, with only an exception at the end, where the end-point must coincide with the origin. We shall focus here on polygons. It proves convenient also to consider *unrooted polygons* (also called simply-connected *polyominoes*), which are polygons where the origin is discarded, so that they plainly represent the possible shapes of SAPs up to translation. For length $2n$, the number p_n of unrooted polygons satisfies $p_n = P_n/(4n)$ since the origin ($2n$ possibilities) and the starting vertex (2 possibilities) of the corresponding SAPs are disregarded in that case. Here is a table, for small values of n, listing polyominoes and the corresponding counting sequences p_n, P_n.



n:	2	3	4	5	6	7	8	9	10
p_n (*EIS* **A002931**):	1	2	7	28	124	588	2938	15268	81826
P_n (*EIS* **A010566**):	8	24	112	560	2976	16464	94016	549648	3273040

Take the (widely open) problem of determining the number P_n of SAPs of perimeter $2n$. This (intractable) problem can be approached as a limit of the (tractable) problem[11] that consists in enumerating the collection $\mathcal{P}^{[w]}$ of SAPs of width w, for increasing values of w. The latter problem is amenable to the transfer matrix method, as first discovered by Entig in 1980; see [**126**]. Indeed, take a polygon and consider a sweepline that moves from its left to its right. Once width is fixed, there are at most 2^{2w+2} possibilities for the ways a vertical sweepline may intersect the polygon's edges at half integer abscissæ. (There are $w+1$ edges and for each of these, one should "remember" whether they connect with the upper or lower boundary.)

[11]In this version of the text, we limit ourselves to a succinct description and refer to the original papers [**126, 246**] for details.

The transitions are then themselves finitely described. In this way, it becomes possible to set up a transfer matrix for any fixed width w. For fixed n, by computing values of $P_n^{[w]}$ with increasing w, one finally determines (in principle) the exact value of any P_n.

The program suggested above has been carried out to record values by the "Melbourne School" under the impulse of Tony Guttmann. For instance, Jensen [**246**] found in 2003 that the number of unrooted polygons of perimeter 100 is

$$p_{50} = 7545649677448506970646886033356862162.$$

Attaining such record values necessitates algorithms that are much more sophisticated than the naïve approach we have just described, as well as a number of highly ingenious programming optimizations.

It is an equally open problem to estimate asymptotically the number of SAPs of perimeter n. Given the exact values till perimeter 100 or more, a battery of fitting tests for asymptotic formula can be applied, leading to highly *convincing* (though still heuristic) formulæ. Thanks to several workers in this area, we can regard the final answer as "known". From the works of Jensen and his predecessors, it results that a reliable empirical estimate is of the form

$$\begin{cases} p_n = B\mu^{2n}(2n)^{-\beta}(1+o(1)), \\ \mu \doteq 2.63815\,85303, \qquad \beta = -\dfrac{5}{2} \pm 3\cdot 10^{-7}, \qquad B \doteq 0.5623013. \end{cases}$$

Thus, the answer is almost certainly of the form $p_n \asymp \mu^{2n} n^{-5/2}$ for unrooted polygons and $P_n \asymp \mu^{2n} n^{-3/2}$ for rooted polygons. It is believed that the same connective constant μ dictates the exponential growth rate of self-avoiding walks. See Finch's book [**137**, Sec. 5.10] for a perspective and numerous references.

There is also great interest in the number $p_{m,n}$ of polyominoes with perimeter $2n$ and area m, with area defined as the number of square cells composing the polyomino. Studies conducted by the Melbourne school yield numerical data that are consistent to an amazing degree (e.g., moments till order ten and small–n corrections are considered) with the following assumption: *The distribution of area in a fixed-perimeter polyomino obeys in the asymptotic limit an "Airy area distribution"*. This distribution is defined as the limit distribution of the area under Dyck paths, a problem that was briefly discussed on p. 309 and to which we propose to return in Chapter IX. See [**246, 363**] and references therein for a discussion of polyomino area. It is finally of great interest to note that the interpretation of data was strongly guided by what is already known for exactly solvable models of the type we are repeatedly considering in this book. .... END OF EXAMPLE 19. □

EXAMPLE 20. *Horizontally convex polyominoes.* Pólya [**348**] and Temperley [**408**] independently discovered an exactly solvable polyomino model. (See also the text by van Rensburg [**423**] for more.) Define as usual a polyomino as a collection of unit squares with vertices in $\mathbb{Z}_{\geq 0} \times \mathbb{Z}_{\geq 0}$ that forms a connected set without articulation points. Such a polyomino is said to be *horizontally convex* (H.C.) if its intersection with any horizontal line is either empty or an interval. An H.C. polyomino is thus a stack of a certain number of rows of squares, where each row has a segment of length ≥ 1 in common with the next row up. (We imagine H.C. polyominoes growing from bottom to top.) The enumeration of such polyominoes, following Temperley [**408**, p. 66] constitutes a nice illustration of the transfer matrix method in the case when the set of states is *infinite*.

Let $\mathcal{T}^{[k]}$ be the class of polyominoes with exactly k square cells on their top row. Size of a polyomino is its number of cells. We wish to enumerate the class $\mathcal{T} := \bigcup_k \mathcal{T}^{[k]}$. In order to

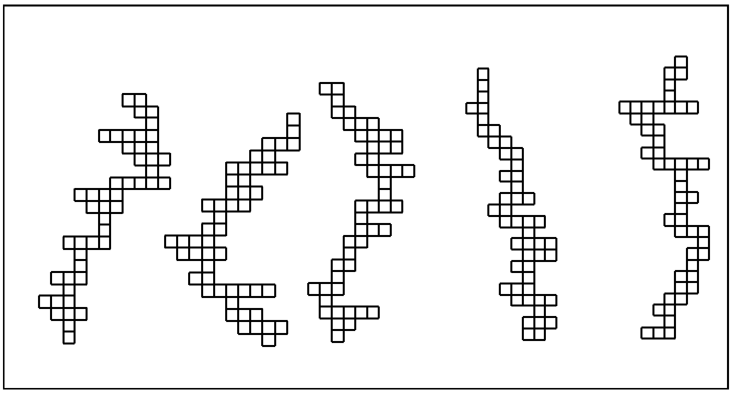


FIGURE 22. Five horizontally convex polyominoes of size $n = 50$ drawn uniformly at random.

do so, according to the transfer matrix method, one needs to relate the $\mathcal{T}^{[k]}$ to one another. Let z be the variable marking size and let x mark the size of the top row. The transition from one $\mathcal{T}^{[k]}$ to a $\mathcal{T}^{[\ell]}$ has a multiplicity equal to $k+\ell-1$. Thus the generating functions $t_k := T^{[k]}(z)$ satisfy the infinite system of equations

$$\begin{array}{rcl} t_1 & = & z + z\,(t_1 + 2t_2 + 3t_3 + \cdots) \\ t_2 & = & z^2 + z^2\,(2t_1 + 3t_2 + 4t_3 + \cdots) \\ t_3 & = & z^3 + z^3\,(3t_1 + 4t_2 + 5t_3 + \cdots) \end{array} \tag{99}$$

This corresponds to an infinite transfer matrix which is highly structured:

$$M(z)_{k,\ell} = (k+\ell-1)z^\ell,$$

and, as shown by Temperley [**408**, p. 66], the system can be solved by elementary manipulations.

In a case like this, it is well worth trying a bivariate generating function. Define

$$T(z,u) = \sum_{n,k} T^{[k]}(z)u^k.$$

The action of "adding a slice" on the top row of a polyomino is reflected by a linear operator $\mathcal{L}$ that transforms u^k representing the top row of the polyomino before addition into a sum of monomials $u^\ell z^\ell$ with the proper multplicites:

$$\mathcal{L}[u^k] = k(uz)^k + (k+1)(uz)^{k+1} + \cdots = (k-1)\frac{uz}{1-uz} + \frac{uz}{(1-uz)^2}.$$

A better formula results if one expresses more generally the quantity $\mathcal{L}[f(u)]$:

$$\mathcal{L}[f(u)] = \frac{uz}{(1-uz)^2}f(1) + \frac{uz}{1-uz}\left(f'(1) - f(1)\right). \tag{100}$$

Treat now the BGF $T(z,u)$ as a function of u, keeping z as a parameter, and write for readability $\tau(u) := T(z,u)$. A horizontally convex polyomino is obtained by starting from a bottom row that can have any number of cells and repeatedly adding a slice[12]. This construction is thus

[12] An earlier instance of the technique of "adding a slice" appears in the context of constrained compositions, Example III.21, p. 188.

reflected by the main functional equation

$$\tau(u) = \frac{zu}{1-zu} + \mathcal{L}[\tau(u)] \tag{101}$$
$$= \frac{zu}{1-zu} + \frac{zu}{1-zu}\tau'(1) + \frac{z^2u^2}{(1-zu)^2}\tau(1),$$

upon making use of (100). Instantiating at $u = 1$ provides the first relation

$$\tau(1) = \frac{z}{1-z} + \frac{z}{1-z}\tau'(1) + \frac{z^2}{(1-z)^2}\tau(1), \tag{102}$$

while differentation of (101) with respect to u followed by the specialization $u = 1$ provides the second relation

$$\tau'(1) = \frac{z}{(1-z)^2} + \frac{z}{(1-z)^2}\tau'(1) + 2\frac{z}{(1-z)^3}\tau(1). \tag{103}$$

We now have a linear system of two equations in two unknowns, resulting in an expression of $\tau(1) = T(z) = T(z,1)$, which enumerates all horizontally convex polyominoes:

$$T(z) = \frac{z(1-z)^3}{1-5z+7z^2-4z^3}. \tag{104}$$

From (101) to (104), the whole calculation is barely three lines of code under a decent computer algebra system. Note that, the original system being infinite, it is far from obvious *a priori* that the generating function should be rational. (In the present context, rationality devolves from the very regular structure of the transfer matrix.)

The counting sequence obtained by expansion,

$$T(z) = z + 2\,z^2 + 6\,z^3 + 19\,z^4 + 61\,z^5 + 196\,z^6 + 629\,z^7 + 2017\,z^8 + \cdots$$

is *EIS* **A001169** (*"Number of board-pile polyominoes with n cells"*). The asymptotic form is also easily obtained: we find

$$T_n \sim CA^n, \qquad c \doteq 0.18091, \quad A \doteq 3.20556,$$

with A a cubic irrational.

An alternative derivation, which is more sophisticated, is due to Klarner and is presented in Stanley's book [**391**, §4.7]. Hickerson [**231**] has found a direct construction, which explains the rationality of the GF by means of a regular language encoding. (The drawings of Figure 22 have been obtained by an application of the recursive method [**183**] to Hickerson's specification.) Louchard [**300**] has conducted an in-depth study of probabilistic properties of several parameters of H.C. polyominoes, using generating functions. ..... END OF EXAMPLE 20. □

▷ **41.** *Height of H.C. polyominoes.* It is possible to introduce an extra variable v to encode height. It is found that height grows on average linearly with n and that the distribution of height is concentrated [**300**]. (This explains the skinny aspects of polyominoes drawn in Figure 22.) ◁

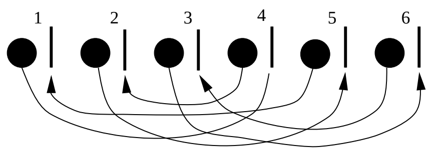
▷ **42.** *A transfer matrix model for lattice paths.* Consider the general context of weighted lattice paths in Section V. 3. Let $\alpha_j, \beta_j, \gamma_j$ be the weights of ascents, descents, and level steps repsectively, when the starting altitude is j. The infinite transfer matrix,

$$T = \begin{pmatrix} \gamma_0 & \alpha_0 & 0 & 0 & 0 & \cdots \\ \beta_1 & \gamma_1 & \alpha_1 & 0 & 0 & \cdots \\ 0 & \beta_2 & \gamma_2 & \alpha_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix},$$

which has a tridiagonal form, "generates" all lattice paths via the quasi-inverse $(I - zT)^{-1}$. In particular, any exactly solvable weighted lattice path model is equivalent to an explicit structured matrix inversion. $\lhd$

V. 6.4. Value-constrained permutations. We conclude this chapter with a discussion of a construction that combines transfer matrix methods with an inclusion-exclusion argument. We treat a collection of constrained permutation problems whose origin lies in nineteenth century recreational mathematics. For instance, the *ménage* problem solved and popularized by Édouard Lucas in 1891, see [**82**], has the following quaint formulation: *What is the number of possible ways one can arrange n married couples (*'ménages'*) around a table in such a way that men and women alternate, but no woman sits next to her husband?*

The ménage problem is equivalent to a permutation enumeration problem. Sit first conventionally the men at places numbered $0, \dots, n-1$, and let σ_i be the position at the right of which the ith wife is placed. Then, a ménage placement imposes the condition $\sigma_i \neq i$ and $\sigma_i \neq i+1$ for each i. We consider here a linearly arranged table (see remarks at the end for the other classical formulation that considers a round table), so that the condition $\sigma_i \neq i+1$ becomes vacuous when $i = n$. Here is a ménage placement for $n = 6$ and its corresponding permutation



$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 2 & 1 & 3 \end{bmatrix}$$

Clearly, this is a generalization of the derangement problem (for which only the weaker condition $\sigma_i \neq i$ is imposed), where the cycle decomposition of permutations suffices to provide a direct solution (see Chapter II).

DEFINITION V.9. *Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, any quantity $\sigma_i - i$ is called an* exceedance *of σ. Given a finite set of integers $\Omega \subset \mathbb{Z}_{\geq 0}$, a permutation is said to be Ω-avoiding if none of its exceedances lies in Ω.*

Inclusion-exclusion. The set Ω being fixed, consider first for all j the class of augmented permutations $\mathcal{P}_{n,j}$ that are permutations of size n such that j of the positions are distinguished and the corresponding exceedances lie in Ω, the remaining positions having arbitrary values (but with the permutation property being satisfied!). Loosely speaking, the objects in $\mathcal{P}_{n,j}$ can be regarded as permutations with "at least" j exceedances in Ω. For instance, with $\Omega = \{1\}$ and

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 3 & 4 & 8 & 6 & 7 & 1 & 5 & 9 \end{pmatrix},$$

there are 5 exceedances that lie in Ω (at positions $1, 2, 3, 5, 6$) and with 3 of these distinguished (say by enclosing them in a box), one obtains an element counted by $\mathcal{P}_{9,3}$ like

$$2\,\boxed{3}\,\boxed{4}\,8\,6\,\boxed{7}\,1\,5\,9.$$

Let $P_{n,j}$ be the cardinality of $\mathcal{P}_{n,j}$. We claim that the number $Q_n = Q_n^\Omega$ of Ω-avoiding permutations of size n satisfies

$$Q_n = \sum_{j=0}^{n} (-1)^j P_{n,j}. \tag{105}$$

Equation (105) is typically an *inclusion-exclusion* relation. To prove it formally[13], define the number $R_{n,k}$ of permutations that have exactly k exceedances in Ω and the generating polynomials

$$P_n(w) = \sum_j P_{n,j} w^j, \qquad R_n(w) = \sum_k R_{n,k} w^k.$$

The GF's are related by

$$P_n(w) = R_n(w+1) \qquad \text{or} \qquad R_n(w) = P_n(w-1)..$$

(The relation $P_n(w) = R_n(w+1)$ simply expresses symbolically the fact that each Ω-exceedance in $\mathcal{R}$ may or may not be taken in when composing an element of $\mathcal{P}$.) In particular, we have $P_n(-1) = R_n(0) = R_{n,0} = Q_n$ as was to be proved.

Transfer matrix model. The preceding discussion shows that everything relies on the enumeration $P_{n,j}$ of permutations with distinguished exceedances in Ω. Introduce the alphabet $\mathcal{A} = \Omega \cup \{\text{‘?’}\}$, where the symbol ‘?’ is called the ‘don’t-care symbol’. A word on $\mathcal{A}$, an instance with $\Omega = \{0, 1, 2\}$ being 20?02?11?, is called a *template*. To an augmented permutation, one associates a template as follows: each exceedance that is not distinguished is represented by a don’t care symbol; each distinguished exceedance (thereby an exceedance with value in Ω) is represented by its value. A template is said to be legal if it arises from an augmented permutation. For instance a template $2\,1\cdots$ cannot be legal since the corresponding constraints, namely $\sigma_1 - 1 = 2$, $\sigma_2 - 2 = 1$, are incompatible with the permutation structure (one would have $\sigma_1 = \sigma_2 = 3$). In contrast, the template 20?02?11? is seen to be legal. Figure 23 is a graphical rendering; there, letters of templates are represented by dominoes, with a cross at the position of a numeric value in Ω, and with the domino being blank in the case of a don’t-care symbol.

Let $T_{n,j}$ be the set of legal templates relative to Ω that have length n and comprise j don’t care symbols. Any such legal template is associated to exactly $j!$ permutations, since $n - j$ position-value pairs are fixed in the permutation, while the j remaining positions and values can be taken arbitrarily. There results that

$$P_{n,n-j} = j!\, T_{n,j} \qquad \text{and} \qquad Q_n = \sum_{j=0}^{n} (-1)^{n-j} j!\, T_{n,j}, \tag{106}$$

by (105). Thus, the enumeration of avoiding permutations rests entirely on the enumeration of legal templates.

The enumeration of legal templates is finally effected by means of a transfer matrix method, or equivalently, by a finite automaton. If a template $\tau = \tau_1 \cdots \tau_n$ is legal,

[13] See also the discussion in Subsection III. 7.4, p. 195.

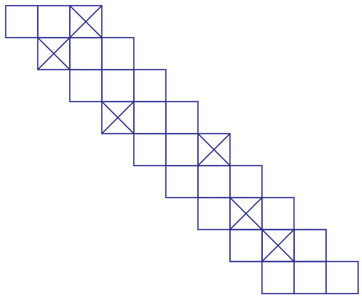


FIGURE 23. A graphical rendering of the legal template **20?02?11?** relative to $\Omega = \{0, 1, 2\}$.

then the following condition is met,

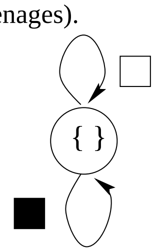
$$\tau_j + j \neq \tau_i + i, \tag{107}$$

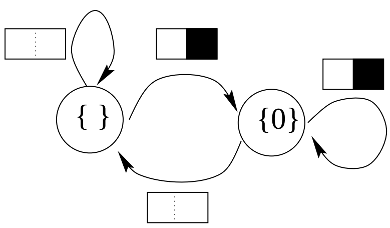
for all pairs (i, j) such that $i < j$ and neither of τ_i, τ_j is the don't-care symbol. (There are additional conditions to characterize templates fully, but these only concern a few letters at the end of templates and we may ignore them in this discussion.) In other words, a τ_i with a numerical value preempts the value $\tau_i + i$. Figure 23 exemplifies the situation in the case $\Omega = \{0, 1, 2\}$. The dominoes are shifted one position each time (since it is the value of $\sigma - i$ that is represented) and the compatibility constraint (107) is that no two crosses should be vertically aligned. More precisely the constraints (107) are recognized by a deterministic finite automaton whose states are indexed by subsets of $\{0, \ldots, b-1\}$ where the "span" b is defined as $b = \max_{\omega\in\Omega} \omega$. The initial state is the one associated with the empty set (no constraint is present initially), the transitions are of the form ($j \in \{0, \ldots, b\}$):

$$\begin{cases} (q_S, j) & \mapsto & q_{S'} & \text{where } S' = ((S-1) \cup \{j-1\}) \cap \{0, \ldots, b-1\} \\ (q_S, ?) & \mapsto & q_{S'} & \text{where } S' = (S-1) \cap \{0, \ldots, b-1\}. \end{cases}$$

The initial state (is $q_{\{\}}$ and it is equal to the final state this translates the fact that no domino can protrude from the right, and is implied by the linear character of the ménage problem under consideration). In essence, the automaton only needs a finite memory since the dominoes slide along the diagonal and, accordingly, constraints older than the span can be forgotten. Notice that the complexity of the automaton, as measured by its number of states, is 2^b.

Here are the automata corresponding to $\Omega = \{0\}$ (derangements) and to $\Omega = \{0, 1\}$ (ménages).





For the ménage problem, there are two states depending on whether or not the currently examined value has been preempted at the preceding step.

From the automaton construction, the bivariate GF $T^{\Omega}(z,u)$ of legal templates, with u marking the position of don't care symbols, is a rational function that can be determined in an automatic fashion from Ω. For the derangement and ménage problems, one finds

$$T^{\{0\}}(z,u) = \frac{1}{1-z(1+u)}, \qquad T^{\{0,1\}}(z,u) = \frac{1-z}{1-z(2+u)+z^2}.$$

In general, this gives access to the OGF of the corresponding permutations. Consider the partial expansion of $T^{\Omega}(z,u)$ with respect to u, taken under the form

$$T^{\Omega}(z,u) = \sum_r \frac{c_r(z)}{1-uu_r(z)}, \tag{108}$$

assuming for simplicity only simple poles. There the sum is finite and it involves algebraic functions c_j and u_j of the variable z. Finally, the OGF of Ω-avoiding permutations is obtained from T^{Ω} by the transformation

$$z^n u^k \quad \mapsto \quad (-z)^n k!,$$

which is the transcription of (106). Define the (divergent) OGF of all permutations,

$$F(y) = \sum_{n=0}^{\infty} n!\, y^n = {}_2F_0[1,1;y],$$

in the terminology of hypergeometric functions. Then, by the remarks above and (108), we find

$$Q^{\Omega}(z) = \sum_r c_r(-z) F(-u_j(-z)).$$

In other words, *the OGF of Ω-avoiding permutations is a composition of the OGF of the factorial series with algebraic functions.*

The expressions simplify much in the case of ménages and derangements where the denominators of T are of degree 1 in u. One has

$$Q^{\{0\}}(z) = \frac{1}{1+z} F(\frac{z}{1+z}) = 1 + z^2 + 2z^3 + 9z^4 + 44z^5 + 265z^6 + 1854z^7 + \cdots,$$

for derangements, whence a new derivation of the known formula,

$$Q_n^{\{0\}} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)!.$$

Similarly, for (linear) ménage placements, one finds

$$Q^{\{0,1\}}(z) = \frac{1}{1+z} F(\frac{z}{(1+z)^2}) = 1 + z^3 + 3z^4 + 16z^5 + 96z^6 + 675z^7 + \cdots,$$

which is *EIS* **A00027** and corresponds to the formula

$$\mathbb{Q}_n^{\{0,1\}} = \sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} (n-k)!.$$

Finally, the same techniques adapts to constraints that "wrap around", that is, constraints taken modulo n. (This corresponds to a round table in the ménage problem.) In that case, what should be considered is the loops in the automaton recognizing templates (see also the discussion of the zeta function of graphs, p. 323). One finds in this way the OGF of the circular (i.e., classical) ménage problem to be *EIS* **A000179**,

$$\widehat{Q}^{\{0,1\}}(z) = \frac{1-z}{1+z}F\Big(\frac{z}{(1+z)^2}\Big)+2z = 1+z+z^3+2z^4+13z^5+80z^6+579z^7+\cdots,$$

which yields the classical solution of the (circular) ménage problem,

$$\widehat{Q}_n^{\{0,1\}} = \sum_{k=0}^{n}(-1)^k\frac{2n}{2n-k}\binom{2n-k}{k}(n-k)!,$$

a formula that is due to Touchard; see [**82**, p. 185] for pointers to the vast classical literature on the subject. The algebraic part of the treatment above is close to the inspiring discussion offered in Stanley's book [**391**]. An application to robustness of interconnections in random graphs is presented in [**159**].

Asymptotic analysis. For asymptotic analysis purposes, the following general property proves useful: *Let F be the OGF of factorial numbers and assume that $y(z)$ is analytic at the origin where it satisfies $y(z) = z - \lambda z^2 + O(z^3)$; then the following estimate holds:*

$$[z^n]F(y(z)) \sim [z^n]F(z(1-\lambda z)) \sim n!e^{-\lambda}. \tag{109}$$

(The proof results from simple manipulations of divergent series in the style of [**28**].) This gives at sight the estimates

$$Q_n^{\{0\}} \sim ne^{-1}, \qquad Q_n^{\{0,1\}} \sim ne^{-2}.$$

More generally, for any set Ω containing λ elements, one has

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}.$$

Furthermore, the number $R_{n,k}^{\Omega}$ of permutations having exactly k occurrences (k fixed) of an exceedance in Ω is asymptotic to

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}\frac{\lambda^k}{k!}.$$

In other words, the rare event that an exceedance belongs to Ω obeys a Poisson distribution with $\lambda = |\Omega|$. These last two results are established by means of probabilistic techniques in the book [**23**, Sec. 4.3]. The relation (109) provides a way of arriving at such estimates by purely analytic-combinatorial techniques.

▷ **43.** *Other constrained permutations.* Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, a *succession gap* is defined as any difference $\sigma_{i+1} - \sigma_i$. Discuss the counting of permutations whose succession gaps are constrained to lie outside of a finite set Ω. In how many ways can a kangaroo pass through all points of the integer interval $[1, n]$ starting at 1 and ending at n while making hops that belong to $\{-2, -1, 1, 2\}$? ◁

V. 7. Perspective

The theorems in this chapter demonstrate the power of the fundamental techniques developed in Chapter IV, which exploit classical theorems in complex analysis to develop coefficient asymptotics. As we start seeing it here, this approach applies to many of the generating functions derived from the formal combinatorial techniques of Part A of this book. By paying careful attention to the types of combinatorial constructions involved, we are able to identify abstract schemas that help us solve whole classes of problems at once. Each schema connects a type of combinatorial construction to a complex asymptotic method. In this way, it becomes possible to discuss properties shared by an infinite collection of combinatorial classes. In this chapter, we have presented the method in detail for classes that involve a sequence construction and classes recursively defined by a linear system of equations (paths in graphs, automata, transfer matrices).

In an ideal world, we might wish to have a direct correspondence between combinatorial constructions and analytic methods—a theory that would carry all the way from combinatorial objects of any description to full analysis of all their properties. The case of paths in graphs and automata, with its strong connectedness condition leading to Perron-Frobenius theory, is an instance of this ideal situation. Reality is however usually a bit more complex: theorems for deriving asymptotic results from combinatorial specifications must often have some sort of analytic side conditions. A typical example is the radius of convergence condition for supercritical sequences. As soon as such conditions are satisfied, the asymptotic properties of large structures become highly predictable. This is the very essence of analytic combinatorics.

In the next two chapters, we investigate generating functions whose singularities are no longer poles—fractional exponents and logarithmic factors become allowed. This first necessitates investing in general methodology, a task undertaken in Chapter VI where the method known as singularity analysis is developed. Then, a chapter parallel to the present one, Chapter VII, will present a number of new schemas based on the set and cyle constructions, as well as on recursion.

Applications of rational functions in discrete and continuous mathematics are in abundance. Many examples are to be found in Goulden and Jackson's book [**208**]. Stanley [**391**] even devotes a full chapter of his book *Enumerative Combinatorics*, vol. I, to rational generating functions. These two books push the theory further than we can do here, but the corresponding asymptotic aspects which we expose lie outside of their scope. The analytic theory of positive rational functions starts with the works of Perron and Frobenius at the beginning of the twentieth century and is explained in books on matrix theory likes those of Bellman [**25**] and Gantmacher [**191**]. Its importance has been long recognized in the theory of finite Markov chains, so that the basic theory of positive matrices is well developed in many elementary treatises on probability theory. For such aspects, we refer for instance to the classic presentations by Feller [**133**] or Karlin and Taylor [**252**].

The supercritical sequence schema is the first in a list of abstract schemas that neatly exemplify the interplay between combinatorial, analytic, and probabilistic properties of large random structures. The origins of this approach are to be traced to early works of Bender [**27, 28**] followed by Soria and Flajolet [**177, 179, 388**].

Turning to more specific topics, we mention in relation to Section V. 3 the first global attempt at a combinatorial theory of continued fractions by Flajolet in [**139**] together with related works of Jackson of which an exposition is to be found in [**208**, Ch. 5] and a synthesis in [**157**] in relation to birth and death processes. Walks on graphs from an algebraic standpoint are well discussed in Godsil's book [**202**]. The discussion of local constraints in permutations based on [**159**] combines some of the combinatorial elements bound in Stanley's book [**391**] with the general philosophy of analytic combinatorics. Our treatment of words and languages largely draws its inspiration from the line of research started by Schützenberger in the early 1960's and on the subsequent account to be found in Lothaire's book [**294**]. A nice review of transfer matrix methods (including a discussion of limit distributions) is offered by Bender, Richmond, and Williamson in [**35**].

VI

Singularity Analysis of Generating Functions

Es ist eine Tatsache, daß die genauere Kenntnis
des Verhaltens einer analytischen Funktion
in der Nähe ihrer singulären Stellen
eine Quelle von arithmetischen Sätzen ist.[1]

— ERICH HECKE [**228**, Kap. VIII]

Contents

A function's singularities are reflected in the function's coefficients. Chapters IV and V have treated in detail rational fractions and meromorphic functions, where the local analysis of polar singularities provides contributions to coefficients in the form of products of polynomials and simple exponentials. In this chapter, we present a general approach to the analysis of coefficients of generating functions that is not restricted to polar singularities and extends to a very large class of functions that have moderate growth or decay at their dominant singularities. The basic principle behind this extension is the existence of a *general correspondence* between

the asymptotic expansion of a function near its dominant singularities
and
the asymptotic expansion of the function's coefficients.

This mapping essentially preserves orders of growth in the sense that larger functions tend to have have larger coefficients. It extends considerably the analysis of meromorphic functions in Chapters IV–V and further justifies the *Principles of Coefficient Asymptotics* enounced in Chapter IV, p. 215.

[1] *"It is a fact that the precise knowledge of the behaviour of an analytic function in the vicinity of its singular points is a source of arithmetic properties."*

Precisely, the method of *singularity analysis* applies to functions whose singular expansions involve fractional powers and logarithms—we refer to such singularities as "*algebraic–logarithmic*" . It principally relies on two types of results.

- — First, it is possible to set up a *catalogue* of asymptotic expansions for coefficients of the standard functions that occur in such singular expansions
- — Second, *transfer theorems* allow us to extract the asymptotic order of coefficients of error terms from singular expansions with error terms.

The developments are based on Cauchy's coefficient formula, used in conjunction with special contours of integration known as *Hankel contours*. The contours come very close to the singularities then steer away: By design, they have the property of capturing essential asymptotic informations contained in the functions' singularities.

The method of singularity analysis is robust, so that functions amenable to it benefit of being closed under a variety of operations, including sum, product, integration, differentiation, and composition. Another important feature of the method is that it only necessitates *local asymptotic properties* of the function to be analysed. In this way, it often proves instrumental in the case of functions that are only indirectly accessible through functional equations.

This chapter is meant to develop the basic technology of singularity analysis and, like Chapter IV, it is largely of a methodological nature. We illustrate the approach by a few combinatorial problems, including simple varieties of trees (e.g, unary-binary trees), combinatorial sums, the supercritical cycle construction, supertrees, Pólya's drunkard walks, and tree recurrences. The next chapter, Chapter VII, will systematically explore combinatorial structures and schemas as well as functional equations that can be asymptotically analysed by means of singularity analysis in a way that parallels Chapter V regarding meromorphic asymptotics.

VI. 1. A glimpse of basic singularity analysis theory

Rational and meromorphic functions involve locally near a singularity elements of the form $(1-z/\omega)^{-k}$. Accordingly their coefficients involve asymptotically exponential polynomials, that is, finite linear combinations of elements of the type $\omega^{-n}n^{k-1}$, with k a positive integer. We examine here an approach that takes into account functions whose singularities are of a richer nature than mere poles found in rational and meromorphic functions. the method, called *singularity analysis*, applies to functions whose expansion at a singularity ω involves elements of the form

$$\left(1-\frac{z}{\omega}\right)^{-\alpha}\left(\log\frac{1}{1-\frac{z}{\omega}}\right)^{\beta}.$$

Under suitable conditions to be discussed in detail in this chapter, any such element contributes a term of the form

$$\omega^{-n}n^{\alpha-1}(\log n)^{\beta}.$$

Here, α and β can be arbitrary real (or even complex) numbers.

Location of singularities and exponential factors. The exponential factor ω^{-n} present in earlier expansions is easily accounted for (see Chapter IV), as the location of the dominant singularities always induces a multiplicative exponential factor for coefficients. Indeed, if $f(z)$ is singular at $z = \omega$, then $g(z) \equiv f(z/\omega)$ satisfies, by the scaling rule of Taylor expansions,

$$[z^n]\, f(z) = \omega^n [z^n]\, f(\frac{z}{\omega}) = \omega^n [z^n]\, g(z),$$

and $g(z)$ itself is singular on the unit circle, but not inside the disc. Consequently, in most of the discussion that follows, we shall examine functions $f(z)$ that are singular at $z = 1$, a condition that entails no loss of generality.

Basic scale. Consider the following table of commonly encountered functions that are singular at 1, together with their coefficients:

(1)

	Function		Coefficient (exact)		Coefficient (asymptotic)
(f_1)	$[z^n]\, 1 - \sqrt{1-z}$	$=$	$\dfrac{2}{n4^n}\dbinom{2n-2}{n-1}$	$\sim$	$\dfrac{1}{2\sqrt{\pi n^3}}$
(f_2)	$[z^n]\, \dfrac{1}{\sqrt{1-z}}$	$=$	$\dfrac{1}{4^n}\dbinom{2n}{n}$	$\sim$	$\dfrac{1}{\sqrt{\pi n}}$
(f_3)	$[z^n]\, \dfrac{1}{1-z}$	$=$	1	$\sim$	1
(f_4)	$[z^n]\, \dfrac{1}{1-z}\log\dfrac{1}{1-z}$	$=$	H_n	$\sim$	$\log n$
(f_5)	$[z^n]\, \dfrac{1}{(1-z)^2}$	$=$	$n+1$	$\sim$	$n.$

Some structure is apparent in this table: a logarithmic factor in the function is reflected by a similar factor in the coefficients, square-roots somehow induce square-roots, and functions involving larger powers have larger coefficients.

It is easy to come up at least with a partial explanation of these observations. Regarding basic functions such as f_1, f_2, f_3, and f_5, the Newton expansion

$$(1-z)^{-\alpha} = \sum_{n=0}^{\infty} \binom{n+\alpha-1}{n} z^n$$

when specialized to an integer k immediately gives the asymptotic form of the coefficients involved,

$$[z^n](1-z)^{-k} \equiv \frac{(n+1)(n+2)\cdots(n+k-1)}{(k-1)!} = \frac{n^{k-1}}{(k-1)!}\left(1 + O(\frac{1}{n})\right). \tag{2}$$

For general α, it is therefore natural to expect

$$[z^n](1-z)^{-\alpha} \equiv \binom{n+\alpha-1}{\alpha-1} = \frac{n^{\alpha-1}}{(\alpha-1)!}\left(1 + O(\frac{1}{n})\right). \tag{3}$$

It turns out that this asymptotic formula is valid for real or complex α, provided we interpret $(\alpha-1)!$ suitably. We shall prove the estimate (see Section VI. 2 and Theorem VI.1)

$$[z^n](1-z)^{-\alpha} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \frac{\alpha(\alpha-1)}{2n} + \cdots\right), \tag{4}$$

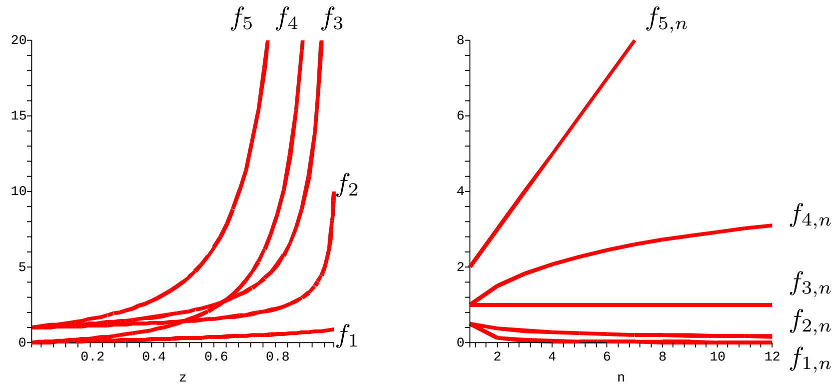


FIGURE 1. The five functions from Eq. (1) and a plot of their coefficient sequences illustrate the tendency of coefficient extraction to be consistent with orders of growth of functions.

where $\Gamma(\alpha)$ is the *Euler Gamma function* defined as

$$\Gamma(\alpha) := \int_0^\infty e^{-t} t^{\alpha-1}\, dt, \tag{5}$$

for $\Re(\alpha) > 0$, which coincides with $(\alpha-1)!$ whenever α is an integer. (Basic properties of this function are recalled in APPENDIX B: *Gamma function*, p. 661.)

We observe from the pair (2)–(3) that functions that are larger at the singularity $z = 1$ have larger coefficients (see Figure 1). The correspondence that this observation suggests is very general as we are going to see repeatedly throughout this chapter. A *catalogue* of exact or asymptotic forms for coefficients of standard singular functions is obtained in Section VI. 2 (see Theorem VI.1).

Transfer of error terms. An asymptotic expansion of a function $f(z)$ that is singular at $z = 1$ is typically of the form

$$f(z) = \sigma(z) + O(\tau(z)) \text{ where } \sigma(z) \gg \tau(z) \text{ as } z \to 1, \tag{6}$$

with σ and τ belonging to an asymptotic scale of standard functions like the collection $\{(1-z)^{-\alpha}\}_{\alpha\in\mathbb{R}}$ in simpler cases. Taking formally Taylor coefficients in the expansion (6), we arrive at

$$f_n \equiv [z^n]f(z) = [z^n]\sigma(z) + [z^n]O(\tau(z)). \tag{7}$$

The term $[z^n]\sigma(z)$ is described asymptotically by (4). Therefore, in order to extract asymptotic informations on the coefficients of $f(z)$, one needs a way of extracting coefficients of functions known only by their order of growth around the singularity. Such a translation of error terms from functions to coefficients is achieved by *transfer theorems*, which, under conditions of analytic continuation, guarantee that

$$[z^n]O(\tau(z)) = O([z^n]\tau(z)).$$

(See Section VI. 3 and Theorem VI.3.) This relation is much less trivial than its symbolic form would seem to imply.

In summary, it is the goal of this chapter to expose the (favorable) conditions under which we have available the correspondence (cf. Section VI. 4 and Theorem VI.4)

$$f(z) = \sigma(z) + O(\tau(z)) \qquad \Longrightarrow \qquad f_n = \sigma_n + O(\tau_n). \tag{8}$$

This process of singularity analysis is then seen to parallel the analysis of coefficients of rational and meromorphic functions presented in the previous two chapters. We describe the method for functions from the scale

$$\frac{1}{(1-z)^\alpha}(\log\frac{1}{1-z})^\beta \qquad (z \to 1),$$

whose coefficients have subexponential factors of the form

$$\theta(n) = n^{\alpha-1}(\log n)^\beta.$$

The range of singular behaviours taken into account by singularity analysis is in fact considerably larger: iterated logarithms ($\log\log$'s) and more exotic functions can be encapsulated in the method.

EXAMPLE 1. *First asymptotics of 2-regular graphs.* As an illustration of the *modus operandi* of singularity analysis, consider the function

$$f(z) = \frac{e^{-z-z^2/2}}{\sqrt{1-z}},$$

which is the EGF of 2–regular graphs (or equivalently, "clouds", see Note II.21, p. 124). Singularity analysis permits us to reason as follows. The function $f(z)$ is only singular at $z = 1$ where it has a branch point. Expanding the numerator around $z = 1$, we have

$$f(z) = \frac{e^{-3/4}}{\sqrt{1-z}} + O((1-z)^{1/2}). \tag{9}$$

Therefore (see Theorems VI.1 and VI.3, as well as the discussion in Example 378, p. 378), upon translating formally and term–by–term, one has

$$[z^n]f(z) = e^{-3/4}\binom{n-1/2}{n} + O\binom{n-3/2}{n} = \frac{e^{-3/4}}{\sqrt{\pi n}} + O(n^{-3/2}). \tag{10}$$

Furthermore, a full asymptotic expansion into descending powers of n can be obtained in the same way from a full expansion of the numerator $e^{-z/2-z^2/4}$. ..... END OF EXAMPLE 1. □

Plan of this chapter. The first part of this chapter, Sections VI. 2–VI. 5, is dedicated to the basic technology of singularity analysis along the lines of our foregoing discussion, and including the case of functions with finitely many singularities on the boundary of their disc of convergence. An "Intermezzo", Section VI. 6, serves a prelude to the second part of the chapter, where we investigate operations on generating functions whose effect on singularities is predictable. The most important of these is inversion, which, under a broad set of conditions, leads to square-root singularity and provides a unified asymptotic theory of simple varieties of trees (Section VI. 7). Polylogarithms are proved to be amenabe to singularity analysis in Section VI. 8, a fact that permits us to take into account weights like $\sqrt{n}$ or $\log n$ in combinatorial sums. Composition of functions is studied in Section VI. 9. Then Section VI. 10 presents several closure properties of functions of singularity analysis class, including dfferentiation, integration, and Hadamard product. The chapter concludes with a brief presentation of two classical alternatives to singularity analysis, Tauberian theory and Darboux's method.

VI. 2. Coefficient asymptotics for the basic scale

This section and the next two present the fundamentals of *singularity analysis*, a theory which was developed by Flajolet and Odlyzko in [**167**]. Technically the theory relies on a systematic use of Hankel contours in Cauchy coefficient integrals. Hankel contours classically serve to express the Gamma function: see APPENDIX B: *Gamma function*, p. 661. Here they are first used to estimate coefficients of a standard scale of functions, and then to prove transfer theorems for error terms in Section VI. 3. With this basic process, an asymptotic expansion of a function near a singularity is directly mapped to a matching asymptotic expansion of its coefficients.

Starting from the binomial expansion, we have for general α,

$$[z^n](1-z)^{-\alpha} = (-1)^n \binom{-\alpha}{n} = \binom{n+\alpha-1}{n} = \frac{\alpha(\alpha+1)\cdots(\alpha+n-1)}{n!}.$$

This quantity is expressible in terms of Gamma factors, and

$$\binom{n+\alpha-1}{n} = \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\Gamma(n+1)}, \tag{11}$$

provided α is neither 0 nor a negative integer. (When $\alpha \in \{0,-1,\ldots\}$, the coefficients $\binom{n+\alpha-1}{n}$ eventually vanish, so that the asymptotic problem of estimating $[z^n](1-z)^{-\alpha}$ becomes void.) The asymptotic analysis of the coefficients $\binom{n+\alpha-1}{n}$ can be carried out elementarily by means of Stirling's formula and real integral estimates: see Notes 1 and 2.

A method far more productive than elementary real analysis techniques consists in analysing coefficients of a function $f(z)$ by means of Cauchy's coefficient formula,

$$[z^n]f(z) = \frac{1}{2i\pi}\int_\gamma f(z)\,\frac{dz}{z^{n+1}}.$$

The basic principle is extremely simple: it consists in choosing a contour of integration γ that comes at distance $\frac{1}{n}$ of the singularity $z=1$. Under the change of variables $z = 1+t/n$, the kernel z^{-n-1} in the integral transforms into an exponential, and the function can be locally expanded, with the differential coefficient only introducing a rescaling factor of $1/n$:

$$\boxed{\begin{array}{rclcrcl} z & \mapsto & \left(1+\dfrac{t}{n}\right), & \qquad & dz & \mapsto & \dfrac{1}{n}dt \\ \dfrac{1}{z^{n+1}} & \mapsto & e^{-t}, & & (1-z)^{-\alpha} & \mapsto & n^\alpha(-t)^{-\alpha}. \end{array}} \tag{12}$$

This gives us for instance (precise justification below):

$$[z^n](1-z)^{-\alpha} \sim g_\alpha n^{\alpha-1}, \qquad g_\alpha := \frac{1}{2i\pi}\int e^{-t}(-t)^{-\alpha}\,dt.$$

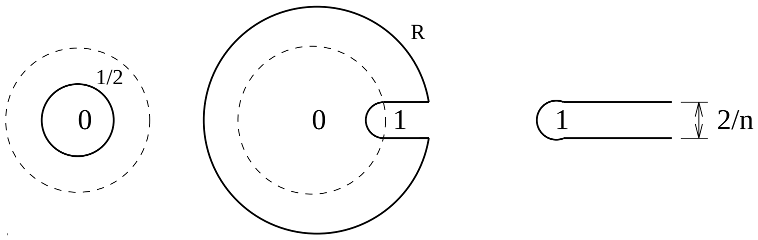


FIGURE 2. The contours $\mathcal{C}_0$, $\mathcal{C}_1$, and $\mathcal{C}_2 \equiv \mathcal{H}(n)$ used for estimating the coefficients of functions from the standard asymptotic scale.

The contour and the associated rescaling "capture" the behaviour of the function near its singularity, thereby enabling coefficient estimation.

THEOREM VI.1 (Standard function scale). *Let α be an arbitrary complex number in $\mathbb{C} \setminus \mathbb{Z}_{\le 0}$. The coefficient of z^n in*

$$f(z) = (1-z)^{-\alpha}$$

admits for large n a full asymptotic expansion in descending powers of n,

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k}\right),$$

where e_k is a polynomial in α of degree $2k$. In particular:

(13)
$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + \frac{\alpha\,(\alpha-1)}{2\,n} + \frac{\alpha\,(\alpha-1)\,(\alpha-2)\,(3\,\alpha-1)}{24\,n^2} + \frac{\alpha^2\,(\alpha-1)^2\,(\alpha-2)\,(\alpha-3)}{48\,n^3} + \cdots\right).$$

The quantity e_k is a polynomial in α that is divisible by $(\alpha-1)\cdots(\alpha-k)$, in accordance with the fact that the asymptotic expansion terminates when $\alpha \in \mathbb{Z}_{\ge 1}$. The factor $1/\Gamma(\alpha)$ vanishes when $\alpha \in \mathbb{Z}_{\le 0}$, in accordance with the fact that coefficients are asymptotically 0 in that case.

PROOF. The first step is to express the coefficient $[z^n](1-z)^{-\alpha}$ as a complex integral by means of Cauchy's coefficient formula,

(14)
$$f_n = \frac{1}{2i\pi}\int_{\mathcal{C}} (1-z)^{-\alpha}\,\frac{dz}{z^{n+1}},$$

where $\mathcal{C}$ is a small enough contour that encircles the origin; see Figure 2. For instance, we can start with $\mathcal{C} \equiv \mathcal{C}_0$, where $\mathcal{C}_0$ is the positively oriented circle $\mathcal{C}_0 = \{z,\ |z| = \frac{1}{2}\}$. The second step is to deform $\mathcal{C}_0$ into another simple closed curve $\mathcal{C}_1$ around the origin that does not cross the half-line $\Re(z) \ge 1$: the contour $\mathcal{C}_1$ consists of a large circle of radius $R > 1$ with a notch that comes back near and to the left of $z = 1$. Since the integrand along large circles decreases as $O(R^{-n-\alpha})$, we can finally let R tend to infinity and are left with an integral representation for f_n where $\mathcal{C}$ has been replaced by a contour $\mathcal{C}_2$ that starts from $-\infty$ in the lower half plane, winds clockwise around

1, and ends at $+\infty$ in the upper half plane. This is a typical case of a *Hankel contour*. A judicious choice of its distance to the half-line $\mathbb{R}_{\geq 1}$ yields the expansion.

To specify precisely the integration path, we particularize $\mathcal{C}_2$ to be the contour $\mathcal{H}(n)$ that passes at a distance $\frac{1}{n}$ from the half line $\mathbb{R}_{\geq 1}$:

$$\mathcal{H}(n) = \mathcal{H}^-(n) + \mathcal{H}^+(n) + \mathcal{H}^\circ(n) \tag{15}$$

where

$$\begin{cases} \mathcal{H}^-(n) &= \{z = w - \frac{i}{n},\ w \geq 1\} \\ \mathcal{H}^+(n) &= \{z = w + \frac{i}{n},\ w \geq 1\} \\ \mathcal{H}^\circ(n) &= \{z = 1 - \frac{e^{i\phi}}{n},\ \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}. \end{cases} \tag{16}$$

Now, a change of variable

$$z = 1 + \frac{t}{n} \tag{17}$$

in the integral (14) gives the form

$$f_n = \frac{n^{\alpha-1}}{2i\pi} \int_{\mathcal{H}} (-t)^{-\alpha} \left(1 + \frac{t}{n}\right)^{-n-1} dt. \tag{18}$$

(The Hankel contour $\mathcal{H}$ is the same as in the proof of Theorem B.1, 663.)

We have the asymptotic expansion

(19)
$$\left(1 + \frac{t}{n}\right)^{-n-1} = e^{-(n+1)\log(1+t/n)} = e^{-t}\left[1 + \frac{t^2 - 2t}{2n} + \frac{3t^4 - 20t^3 + 24t^2}{24n^2} + \cdots\right],$$

which tells us that the integrand in (18) converges pointwise (as well as uniformly in any bounded domain of the t plane) to $(-t)^{-\alpha}e^{-t}$. This quantity is precisely the kernel that appears in Hankel's formula for the Gamma function (p. 663). Substitution of the asymptotic form

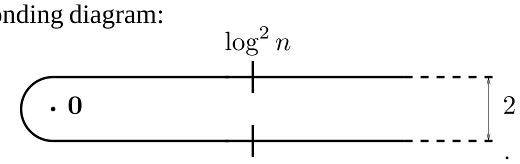
$$\left(1 + \frac{t}{n}\right)^{-n-1} = e^{-t}\left(1 + O(\frac{1}{n})\right),$$

as $n \to \infty$ inside the integral (18) suggests (formally) that

$$[z^n](1-z)^{-\alpha} = \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1 + O(\frac{1}{n})\right).$$

To justify the formal argument outlined in the previous paragraph, we proceed as follows:

(i) Split the contour according to $\Re(t) \leq \log^2 n$ and $\Re(t) \geq \log^2 n$, as in the corresponding diagram:

(20)


	$n=10$	$n=20$	$n=50$
$\frac{4^n}{\sqrt{\pi n^3}}\big(1$	**1** *8708*	**6** *935533866*	*2022877684829178931751713264*
$-\frac{9}{8}N^{-1}$	**16** *603*	**65** *45410086*	**197** *7362936920522405787299715*
$+\frac{145}{128}N^{-2}$	**16** *815*	**656** *5051735*	**19782** *79553371460627490749710*
$-\frac{1155}{1024}N^{-3}$	**1679** *4*	**6564** *073885*	**1978261** *300061101426696482732*
$+\frac{36939}{32768}N^{-4}$	**16796**	**656412** *2750*	**19782616** *64919884629357813591*
$-\frac{295911}{262144}N^{-5}$	**16796**	**6564120** *303*	**1978261657** *612856326190245636*
$+\frac{4735445}{4194304}N^{-6}$	**16796**	**656412042** *6*	**197826165775** *9023715384519184*
$-\frac{37844235}{33554432}N^{-7}\big)$	**16796**	**6564120420**	**19782616577561** *03402179527600*
C_n	**16796**	**6564120420**	**1978261657756160653623774456**

FIGURE 3. Improved approximations to the Catalan numbers obtained by successive terms of their asymptotic expansion.

(ii) Verify that the part corresponding to $\Re(t) \geq \log^2 n$ is negligible in the scale of the problem. For instance, one has

$$\left(1+\frac{t}{n}\right)^{-n} = O(\exp(-\log^2 n)) \qquad \text{for } \Re(t) \geq \log^2 n.$$

(iii) Use a terminating form of (19) to develop an expansion to any predetermined order, with uniform error terms, for the part corresponding to $\Re(t) \leq \log^2 n$. (This is possible because $t/n = O(\log^2 n/n)$ is small.)

These considerations validate term-by-term integration of expansion (19) within the integral of (18), so that the full expansion of f_n is determined as follows: A term of the form t^r/n^s in the expansion (19) induces, by Hankel's formula, a term of the form $n^{-s}/\Gamma(\alpha-r)$. (The expansion so obtained is nondegenerate provided α differs from a negative integer or zero; see also Note 3 for details.) Since

$$\frac{1}{\Gamma(\alpha-k)} = \frac{1}{\Gamma(\alpha)}(\alpha-1)(\alpha-2)\cdots(\alpha-k).$$

the expansion in the statement of the theorem eventually follows. $\square$

The asymptotic approximations obtained from Theorem VI.2 differ from the ones that are associated with meromorphic asymptotics, (Chapter IV), where exponentially small error terms could be derived. However, it is not uncommon to obtain results with about 10^{-6} accuracy, already for values of n in the range 10^1–10^2 with just a few terms of the asymptotic expansion. Figure 3 exemplifies this situation by displaying the approximations obtained for the Catalan numbers,

$$C_n = \frac{4^n}{n+1}[z^n](1-z)^{-1/2},$$

when C_{10}, C_{20}, C_{50} are considered and up to eight asymptotic terms are taken into account.

$\triangleright$ **1.** *Stirling's formula and asymptotics of binomial coefficients.* The Gamma function form (11) of the binomial coefficients yields

$$[z^n](1-z)^{-\alpha} = \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+O(\frac{1}{n})\right),$$

when Stirling's formula is applied to the Gamma factors. $\triangleleft$

$\triangleright$ **2.** *Beta integrals and asymptotics of binomial coefficients.* A direct way of obtaining the general asymptotic form of $\binom{n+\alpha-1}{n}$ bases itself on the Eulerian Beta integral (see [**433**, p.254] and APPENDIX B: *Gamma function*, p. 661). Consider the quantity

$$\phi(n,\alpha) = \int_0^1 t^{\alpha-1}(1-t)^{n-1}\,dt = \frac{(n-1)!}{\alpha(\alpha+1)\cdots(\alpha+n-1)} \equiv \frac{1}{n\binom{n+\alpha-1}{n}},$$

where the second form results elementarily from successive integrations by parts. The change of variables $t = x/n$ yields

$$\phi(n,\alpha) = \frac{1}{n^\alpha}\int_0^n x^{\alpha-1}(1-x/n)^{n-1}\,dt \underset{n\to\infty}{\sim} \frac{1}{n^\alpha}\int_0^\infty x^{\alpha-1}e^{-x}\,dx \equiv \frac{\Gamma(\alpha)}{n^\alpha},$$

where the asymptotic form results from the standard limit formula of the exponential: $\exp(a) = \lim_{n\to\infty}(1+a/n)^n$. $\triangleleft$

$\triangleright$ **3.** *Computability of full expansions.* The coefficients e_k of Theorem VI.1 satisfy

$$e_k = \sum_{\ell=k}^{2k}\lambda_{k,\ell}(\alpha-1)(\alpha-2)\cdots(\alpha-\ell),$$

where $\lambda_{k,\ell} := [v^k t^\ell]e^t(1+vt)^{-1-1/v}$. $\triangleleft$

$\triangleright$ **4.** *Oscillations and complex exponents.* Oscillations occur in the case of singular expansions involving complex exponents. From the consideration of $[z^n](1-z)^{\pm i} \asymp n^{\mp i-1}$, one finds

$$[z^n]\cos\left(\log\frac{1}{1-z}\right) = \frac{P(\log n)}{n} + O(\frac{1}{n^2}),$$

where $P(u)$ is a continuous and 1–periodic function. In general, such oscillations are present in $[z^n](1-z)^{-\alpha}$ for any nonreal α. $\triangleleft$

Logarithmic factors. The basic principle underlying the method of proof of Theorem VI.1 (see also the summary Equation (12)) has the advantage of being easily extended to a wide class of singular functions, most notably the ones that involve logarithmic terms.

THEOREM VI.2 (Standard function scale, logarithms). *Let α be an arbitrary complex number in $\mathbb{C}\setminus\mathbb{Z}_{\le 0}$. The coefficient of z^n in*

$$f(z) = (1-z)^{-\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^\beta$$

admits for large n a full asymptotic expansion in descending powers of $\log n$,

$$f_n = [z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^\beta\left[1+\frac{C_1}{\log n}+\frac{C_2}{\log^2 n}+\cdots\right], \tag{21}$$

where $C_k = (-1)^k\binom{\beta}{k}\Gamma(\alpha)\frac{d^k}{ds^k}\left.\frac{1}{\Gamma(s)}\right|_{s=\alpha}$.

A coefficient of $1/z$ is introduced in front of the logarithm since $\log(1-z)^{-1} = z + O(z^2)$. In this way, $f(z)$ is a *bona fide* power series in z, even in cases when β is not a positive integer.

PROOF. The proof is a simple variant of that of Theorem VI.1 (see [**167**] for details). The basic expansion used is now

$$\begin{aligned} f(1+\frac{t}{n})(1+\frac{t}{n})^{-n-1} &\sim e^{-t}\left(\frac{-n}{t}\right)^{\alpha}\left(\log\left(\frac{-n}{t}\right)\right)^{\beta} \\ &\sim e^{-t}(-t)^{-\alpha}n^{\alpha}(\log n)^{\beta}\left(1-\frac{\log(-t)}{\log n}\right)^{\beta} \\ &\sim e^{-t}(-t)^{-\alpha}n^{\alpha}(\log n)^{\beta}\left(1-\beta\frac{\log(-t)}{\log n}+\frac{\beta(\beta-1)}{2!}\left(\frac{\log(-t)}{\log n}\right)^{2}+\cdots\right). \end{aligned}$$

Again, we are justified in using this expansion inside Cauchy's integral representation of coefficients. What comes out from term by term integration is a collection of Hankel integrals of the form

$$-\frac{1}{2i\pi}\int_{+\infty}^{(0)}(-t)^{-s}e^{-t}(\log(-t))^{k}\,dt$$

which reduce to derivatives of $1/\Gamma(s)$, as is seen by differentiation with respect to s under the integral sign. □

A typical example of application of Theorem VI.2 is the estimate

$$[z^n]\frac{1}{\sqrt{1-z}}\,\frac{1}{\frac{1}{z}\log\frac{1}{1-z}} = \frac{1}{\sqrt{\pi n}\,\log n}\left(1-\frac{\gamma+2\log 2}{\log n}+O(\frac{1}{\log^2 n})\right).$$

(Such singular functions do occur in combinatorics and the analysis of algorithms [**176**].)

▷ **5.** *Singularity analysis of slowly varying functions.* A function $\mathrm{M}(u)$ is said to be *slowly varying* towards infinity (in the complex plane) if for any fixed $\lambda > 0$ and all θ satisfying $|\theta| \le \pi - \phi$ for some $\phi \in (0, \frac{\pi}{2})$, there holds

$$\lim_{u\to+\infty}\frac{\mathrm{M}(\lambda e^{i\theta u})}{\mathrm{M}(u)} = 1.$$

(Powers of logarithms and iterated logarithms are typically slowly varying functions.) Under suitable uniformity assumptions, one has [**167**]

$$[z^n]\frac{1}{(1-z)^{\alpha}}\,\mathrm{M}\left(\frac{1}{1-z}\right) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\,\mathrm{M}(n). \tag{22}$$

For instance: $[z^n]\dfrac{\exp\left(\sqrt{\frac{1}{z}\log\frac{1}{1-z}}\right)}{\sqrt{1-z}} \sim \dfrac{\exp\left(\sqrt{\log n}\right)}{\sqrt{\pi n}}$. See also the discussion of Tauberian theory, p. 416. ◁

▷ **6.** *Iterated logarithms.* For a general $\alpha \notin \mathbb{Z}_{\le 0}$, the relation (22) specializes to

$$[z^n](1-z)^{-\alpha}\left(\frac{1}{z}\log\frac{1}{1-z}\right)^{\beta}\left(\frac{1}{z}\log\left(\frac{1}{z}\log\frac{1}{1-z}\right)\right)^{\delta} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^{\beta}(\log\log n)^{\delta}.$$

A full asymptotic expansion can be derived in this case. ◁

	$\alpha \notin \{0, -1, -2, \ldots\}$	(Eq.)	$\alpha \in \{0, -1, -2, \ldots\}$	(Eq.)
$\beta \notin \mathbb{Z}_{\geq 0}$	$\frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^\beta \sum_{j=0}^{\infty} \frac{C_j}{(\log n)^j}$	(21)	$f_n \sim n^{\alpha-1}(\log n)^\beta \sum_{j=1}^{\infty} \frac{D_j}{(\log n)^j}$	(23)
$\beta \in \mathbb{Z}_{\geq 0}$	$\frac{n^{\alpha-1}}{\Gamma(\alpha)} \sum_{j=0}^{\infty} \frac{E_j(\log n)}{n^j}$	(24)	$n^{\alpha-1} \sum_{j=0}^{\infty} \frac{F_j(\log n)}{n^j}$	(26)

FIGURE 4. The general and special cases of $f_n \equiv [z^n]f(z)$ when $f(z)$ is as in Theorem VI.2.

Special cases. The conditions of Theorems VI.1 and VI.2 exclude explicitly the case when α is a negative integer: the formulæ actually remain valid in this case, provided one interprets them as limit cases, making use of $0 = 1/\Gamma(0) = 1/\Gamma(-1) = \cdots$. Also, when β is a positive integer, the expansion of Theorem VI.2 terminates: in that situation, stronger forms are valid. Such cases are summarized in Figure 4 and discussed below.

The case of integral $\alpha \in \mathbb{Z}_{\leq 0}$ and general $\beta \notin \mathbb{Z}_{\geq 0}$. When α is a negative integer, the coefficients of $f(z) = (1-z)^{-\alpha}$ eventually reduce to zero, so that the asymptotic coefficient expansion becomes trivial: this situation is implicitly covered by the statement of Theorem VI.1 since, in that case, $1/\Gamma(\alpha) = 0$. When logarithms are present (with $\alpha \in \mathbb{Z}_{\leq 0}$ still), the expansion of Theorem VI.2 regarding

$$f(z) = (1-z)^{-\alpha} \left(\frac{1}{z} \log \frac{1}{1-z}\right)^\beta$$

remains valid provided we again take into account the equality $1/\Gamma(\alpha) = 0$ in formula (21) after effecting simplifications by Gamma factors: It is only the first term of (21) that vanishes, and one has

$$[z^n]f(z) \sim n^{\alpha-1} (\log n)^\beta \left[\frac{D_1}{\log n} + \frac{D_2}{\log^2 n} + \cdots\right], \tag{23}$$

where D_k is given by $D_k = (-1)^k \binom{\beta}{k} \frac{d^k}{ds^k} \left.\frac{1}{\Gamma(s)}\right|_{s=\alpha}$. For instance, we find

$$[z^n]\frac{z}{\log(1-z)^{-1}} = -\frac{1}{n\log^2 n} + \frac{2\gamma}{n\log^3 n} + O(\frac{1}{n\log^4 n}).$$

The case of general $\alpha \notin \mathbb{Z}_{\leq 0}$ and integral $\beta \in \mathbb{Z}_{\geq 0}$. When β is a nonnegative integer, the error terms can be further improved with respect to the ones predicted by the general statement of Theorem VI.2. For instance, we have:

$$\begin{aligned} [z^n]\frac{1}{1-z}\log\frac{1}{1-z} &= \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O(\frac{1}{n^4}) \\ [z^n]\frac{1}{\sqrt{1-z}}\log\frac{1}{1-z} &\sim \frac{1}{\sqrt{\pi n}}\left(\log n + \gamma + 2\log 2 + O(\frac{\log n}{n})\right). \end{aligned}$$

(In such a case, the expansion of Theorem VI.2 terminates since only its first $(k+1)$ terms are nonzero.) In fact, in the general case of nonintegral α, there exists an expansion of the form

$$[z^n](1-z)^{-\alpha}\log^k \frac{1}{1-z} \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}\left[E_0(\log n) + \frac{E_1(\log n)}{n} + \cdots\right], \tag{24}$$

where the E_j are polynomials of degree k, as can be proved by adapting the argument employed for general α (see also Note 8).

The joint case of integral $\alpha \in \mathbb{Z}_{\le 0}$ *and integral* $\beta \in \mathbb{Z}_{\ge 0}$. If α is a negative integer, the coefficients appear as finite differences of coefficients of logarithmic powers. Explicit formulæ are then available elementarily from the calculus of finite differences when β is a positive integer. For instance, with $\alpha = -r$ for $r \in \mathbb{Z}_{\ge 0}$, one has

$$[z^n](1-z)^r \log \frac{1}{1-z} = (-1)^r \frac{r!}{n(n-1)\cdots(n-r)}. \tag{25}$$

The case $\alpha = -r$ and $\beta = k$ (with $r, k \in \mathbb{Z}_{\ge 0}$) is covered by (27) in Note 7 below: there is a formula analogous to (24),

$$[z^n](1-z)^r \log^k \frac{1}{1-z} \sim n^{-r-1}\left[F_0(\log n) + \frac{F_1(\log n)}{n} + \cdots\right], \tag{26}$$

but now with $\deg(F_j) = k-1$.

A table of the asymptotic form of coefficients of a few standard functions illustrating Theorems VI.1 and VI.2 as well as some of the “special cases” is given in Figure 5.

$\triangleright$ **7.** *The method of Frobenius and Jungen.* This is an alternative approach to the case $\beta \in \mathbb{Z}_{\ge 0}$ (see [**249**]). Start from the observation that

$$(1-z)^{-\alpha}\left(\log \frac{1}{1-z}\right)^k = \frac{\partial^k}{\partial \alpha^k}(1-z)^{-\alpha}.$$

and allow the operators of differentiation ($\partial/\partial\alpha$) and coefficient extraction ($[z^n]$) to commute—this can be justified by Cauchy’s coefficient formula upon differentiating under the integral sign—, which yields

$$[z^n](1-z)^{-\alpha}\left(\log \frac{1}{1-z}\right)^k = \frac{\partial^k}{\partial \alpha^k}\frac{\Gamma(n+\alpha)}{\Gamma(\alpha)\Gamma(n+1)}, \tag{27}$$

and leads to an “exact” formula (Note 8 below). $\triangleleft$

$\triangleright$ **8.** *Shifted harmonic numbers.* Define the α-shifted harmonic number by

$$h_n(\alpha) := \sum_{j=0}^{n-1} \frac{1}{j+\alpha}.$$

Set $L(z) := -\log(1-z)$. Then, one has

$$\begin{aligned} [z^n](1-z)^{-\alpha}L(z) &= \binom{n+\alpha-1}{n} h_n(\alpha) \\ [z^n](1-z)^{-\alpha}L(z)^2 &= \binom{n+\alpha-1}{n}\left(h_n'(\alpha) + h_n(\alpha)^2\right). \end{aligned}$$

Function	*Coefficients*
$(1-z)^{3/2}$	$\frac{1}{\sqrt{\pi n^5}}(\frac{3}{4}+\frac{45}{32n}+\frac{1155}{512n^2}+O(\frac{1}{n^3}))$
$(1-z)$	(0)
$(1-z)^{1/2}$	$-\frac{1}{\sqrt{\pi n^3}}(\frac{1}{2}+\frac{3}{16n}+\frac{25}{256n^2}+O(\frac{1}{n^3}))$
$(1-z)^{1/2}L(z)$	$-\frac{1}{\sqrt{\pi n^3}}(\frac{1}{2}\log n+\frac{\gamma+2\log 2-2}{2}+O(\frac{\log n}{n}))$
$(1-z)^{1/3}$	$-\frac{1}{3\Gamma(\frac{2}{3})n^{4/3}}(1+\frac{2}{9n}+\frac{7}{81n^2}+O(\frac{1}{n^3}))$
$z/L(z)$	$\frac{1}{n\log^2 n}(-1+\frac{2\gamma}{\log n}+\frac{\pi^2-6\gamma^2}{2\log^2 n}+O(\frac{1}{\log^3 n}))$
1	(0)
$\log(1-z)^{-1}$	$\frac{1}{n}$
$\log^2(1-z)^{-1}$	$\frac{1}{n}(2\log n+2\gamma-\frac{1}{n}-\frac{1}{6n^2}+O(\frac{1}{n^4}))$
$(1-z)^{-1/3}$	$\frac{1}{\Gamma(\frac{1}{3})n^{2/3}}(1+O(\frac{1}{n}))$
$(1-z)^{-1/2}$	$\frac{1}{\sqrt{\pi n}}(1-\frac{1}{8n}+\frac{1}{128n^2}+\frac{5}{1024n^3}+O(\frac{1}{n^4}))$
$(1-z)^{-1/2}L(z)$	$\frac{1}{\sqrt{\pi n}}(\log n+\gamma+2\log 2-\frac{\log n+\gamma+2\log 2}{8n}+O(\frac{\log n}{n^2}))$
$(1-z)^{-1}$	1
$(1-z)^{-1}L(z)$	$\log n+\gamma+\frac{1}{2n}-\frac{1}{12n^2}+\frac{1}{120n^4}+O(\frac{1}{n^6}))$
$(1-z)^{-1}L(z)^2$	$\log^2 n+2\gamma\log n+\gamma^2-\frac{\pi^2}{6}+O(\frac{\log n}{n})$
$(1-z)^{-3/2}$	$\sqrt{\frac{n}{\pi}}(2+\frac{3}{4n}-\frac{7}{64n^2}+O(\frac{1}{n^3}))$
$(1-z)^{-3/2}L(z)$	$\sqrt{\frac{n}{\pi}}(2\log n+2\gamma+4\log 2-4+\frac{3\log n}{4n}+O(\frac{1}{n}))$
$(1-z)^{-2}$	$n+1$
$(1-z)^{-2}L(z)$	$n\log n+(\gamma-1)n+\log n+\frac{1}{2}+\gamma+O(\frac{1}{n})$
$(1-z)^{-2}L(z)^2$	$n(\log^2 n+2(\gamma-1)\log n+\gamma^2-2\gamma+2-\frac{\pi^2}{6}+O(\frac{\log n}{n}))$
$(1-z)^{-3}$	$\frac{1}{2}n^2+\frac{3}{2}n+1$

FIGURE 5. A table of some commonly encountered functions (with $L(z) := \log(1/(1-z)))$ and the asymptotic forms of their coefficients.

(Note: $h_n(\alpha)=\psi(\alpha+n)-\psi(\alpha)$, where $\psi(s):=\partial_s \log\Gamma(s)$.) In particular,

$$[z^n]\frac{1}{\sqrt{1-z}}\log\frac{1}{1-z}=\frac{1}{4^n}\binom{2n}{n}[2H_{2n}-H_n],$$

where $H_n \equiv h_n(1)$ is the usual harmonic number. $\triangleleft$

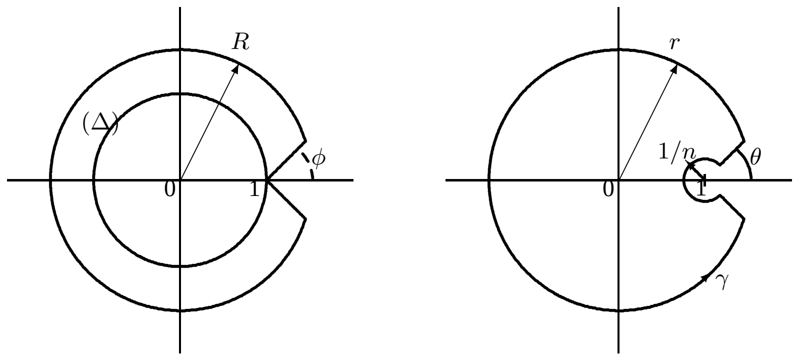


FIGURE 6. A Δ–domain and the contour used to establish Theorem VI.3.

VI. 3. Transfers

Our general objective is to translate an approximation of a function near a singularity into an asymptotic approximation of its coefficients. What is required at this stage is a way to extract coefficients of error terms (known usually in $O(\cdot)$ or $o(\cdot)$ form) in the expansion of a function near a singularity. This task is technically simple as a fairly coarse analysis suffices. Like in the previous section, it relies on contour integration by means of Hankel-type paths; see for instance the summary in Eq. (12) above.

A natural extension of the approach of the previous section is to assume the error terms valid in the complex plane slit along the real half line $\mathbb{R}_{\geq 1}$. In fact weaker conditions suffice and any domain whose boundary makes an acute angle with the half line $\mathbb{R}_{\geq 1}$ appears to be suitable.

DEFINITION VI.1. *Given two numbers ϕ, R with $R > 1$ and $0 < \phi < \frac{\pi}{2}$, the open domain $\Delta(\phi, R)$ is defined as*

$$\Delta(\phi, R) = \{z \mid |z| < R,\ z \neq 1,\ |\operatorname{Arg}(z-1)| > \phi\}.$$

A domain is a Δ–domain *if it is a $\Delta(\phi, R)$ for some R and ϕ. A function is* Δ–analytic *if it is analytic in some Δ–domain.*

Analyticity in a Δ–domain (Figure 6, left) is the basic condition for *transfer* to coefficients of error terms in asymptotic expansions.

THEOREM VI.3 (Transfer, Big-Oh and little-oh). *Let α, β be arbitrary real numbers, $\alpha, \beta \in \mathbb{R}$ and let $f(z)$ be a function that is Δ–analytic.*

(i) *Assume that $f(z)$ satisfies in the intersection of a neighbourhood of 1 with its Δ–domain the condition*

$$f(z) = O\left((1-z)^{-\alpha}(\log \frac{1}{1-z})^{\beta}\right).$$

Then one has: $\qquad [z^n]f(z) = O(n^{\alpha-1}(\log n)^{\beta}).$

(ii) *Assume that $f(z)$ satisfies in the intersection of a neighbourhood of 1 with its Δ–domain the condition*

$$f(z) = o\left((1-z)^{-\alpha}(\log \frac{1}{1-z})^{\beta}\right).$$

Then one has: $\qquad [z^n]f(z) = o(n^{\alpha-1}(\log n)^{\beta})$.

PROOF. (i) The starting point is Cauchy's coefficient formula,

$$f_n \equiv [z^n]f(z) = \frac{1}{2i\pi}\int_\gamma f(z)\,\frac{dz}{z^{n+1}},$$

where γ is any simple loop around the origin which is internal to the Δ–domain of f. We choose the positively oriented contour (Figure 6, right) $\gamma = \gamma_1 \cup \gamma_2 \cup \gamma_3 \cup \gamma_4$, with

$$\begin{cases} \gamma_1 &=& \{\, z \;\mid\; |z-1| = \dfrac{1}{n}, |\operatorname{Arg}(z-1)| \geq \theta] \,\} & \text{(inner circle)} \\ \gamma_2 &=& \{\, z \;\mid\; \dfrac{1}{n} \leq |z-1|,\ |z| \leq r,\ \operatorname{Arg}(z-1) = \theta \,\} & \text{(rectilinear part, top)} \\ \gamma_3 &=& \{\, z \;\mid\; |z-1| = r,\ |\operatorname{Arg}(z-1)| \geq \theta] \,\} & \text{(outer circle)} \\ \gamma_4 &=& \{\, z \;\mid\; \dfrac{1}{n} \leq |z-1|,\ |z| \leq r,\ \operatorname{Arg}(z-1) = -\theta \,\} & \text{(rectilinear part, bottom).} \end{cases}$$

If the Δ domain of f is $\Delta(\phi, R)$, we assume that $1 < r < R$, and $\phi < \theta < \frac{\pi}{2}$, so that the contour γ lies entirely inside the domain of analyticity of f.

For $j = 1, 2, 3, 4$, let

$$f_n^{(j)} = \frac{1}{2i\pi}\int_{\gamma_j} f(z)\,\frac{dz}{z^{n+1}}.$$

The analysis proceeds by bounding the absolute value of the integral along each of the four parts. In order to keep notations simple, we detail the proof in the case where $\beta = 0$.

(1) *Inner circle* (γ_1). From trivial bounds, the contribution from γ_1 satisfies

$$|f_n^{(1)}| = O(\frac{1}{n}) \cdot O\left((\frac{1}{n})^{-\alpha}\right) = O\left(n^{\alpha-1}\right),$$

as the function is $O(n^\alpha)$ (by assumption on $f(z)$), the contour has length $O(n^{-1})$, and z^{-n-1} remains $O(1)$ on this part of the contour.

(2) *Rectilinear parts* (γ_2, γ_4). Consider the contribution $f_N^{(2)}$ arising from the part γ_2 of the contour. Setting $\omega = e^{i\theta}$, and performing the change of variable $z = 1 + \frac{\omega t}{n}$, we find

$$|f_n^{(2)}| \leq \frac{1}{2\pi}\int_1^\infty K\left(\frac{t}{n}\right)^{-\alpha}\left|1 + \frac{\omega t}{n}\right|^{-n-1} dt,$$

for some constant $K > 0$ such that $|f(z)| < K(1-z)^{-\alpha}$ over the Δ–domain, which is granted by the growth assumption on f. From the relation

$$\left|1 + \frac{\omega t}{n}\right| \geq 1 + \Re(\frac{\omega t}{n}) = 1 + \frac{t}{n}\cos\theta,$$

there results the inequality

$$|f_n^{(2)}| \le \frac{K}{2\pi} J_n n^{\alpha-1}, \quad \text{where} \quad J_n = \int_1^\infty t^{-\alpha} \left(1 + \frac{t\cos\theta}{n}\right)^{-n} dt.$$

For a given α, the integrals J_n are all bounded above by some constant since they admit a limit as n tends to infinity:

$$J_n \to \int_1^\infty t^{-\alpha} e^{-t\cos\theta}\, dt.$$

The condition on θ that $0 < \theta < \frac{\pi}{2}$ precisely ensures convergence of the integral. Thus, globally, on the part γ_2 of the contour, we have

$$|f_n^{(2)}| = O(n^{\alpha-1}).$$

A similar bound holds for $f_n^{(4)}$ relative to γ_4.

(3) *Outer circle* (γ_3). There, $f(z)$ is bounded while z^{-n} is of the order of r^{-n}. Thus, the integral $f_n^{(3)}$ is exponentially small.

In summary, each of the four integrals of the split contour contributes $O(n^{\alpha-1})$. The statement of Part (i) of the theorem thus follows.

(ii) An adaptation of the proof shows that $o(.)$ error terms may be translated similarly. All that is required is a further breakup of the rectilinear part at a distance $\log^2 n/n$ from 1 (see Equation (20) or [**167**] for details). □

An immediate corollary of Theorem VI.3 is the possibility of transferring *asymptotic equivalence* from singular forms to coefficients:

COROLLARY VI.1 (sim–transfer). *Assume that $f(z)$ is Δ–analytic and*

$$f(z) \sim (1-z)^{-\alpha}, \qquad \textit{as } z \to 1, \quad z \in \Delta,$$

with $\alpha \notin \{0, -1, -2, \cdots\}$. Then, the coefficients of f satisfy

$$[z^n]f(z) \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)}.$$

PROOF. It suffices to observe that, with $g(z) = (1-z)^{-\alpha}$, one has

$$f(z) \sim g(z) \qquad \text{iff} \qquad f(z) = g(z) + o(g(z)),$$

then apply Theorem VI.1 to the first term, and Theorem VI.3 (little-oh transfer) to the remainder. □

▷ **9.** *Transfer of nearly polynomial functions.* Let $f(z)$ be Δ–singular and satisfy the singular expansion $f(z) \sim (1-z)^r$, where $r \in \mathbb{Z}_{\ge 0}$. Then, $f_n = o(n^{-r-1})$. [This is also a direct consequence of the little-oh transfer.] ◁

▷ **10.** *Transfer of large negative exponents.* The Δ–analyticity condition can be weakened for functions that are large at their singularity. Assume that $f(z)$ is analytic in the open disk $|z| < 1$, and that in the whole of the open disk it satisfies

$$f(z) = O((1-z)^{-\alpha}).$$

Then, provided $\alpha > 1$, one has

$$[z^n]f(z) = O(n^{\alpha-1}).$$

[Hint. Integrate on the circle of radius $1 - \frac{1}{n}$; see also [**167**].] ◁

VI. 4. The process of singularity analysis

In Sections VI. 2 and VI. 3, we have developed a collection of statements granting the existence of *correspondences* between properties of a function $f(z)$ singular at an isolated point ($z = 1$) and the asymptotic behaviour of its coefficients $f_n = [z^n]f(z)$. Using the symbol '$\longmapsto$' to represent such a correspondence[2], we can summarize some of our results relative to the scale $\mathcal{S} = \{(1-z)^{-\alpha},\ \alpha \in \mathbb{C} \setminus \mathbb{Z}_{\le 0}\}$ as follows:

$$\left\{\begin{array}{llll} f(z) = (1-z)^{-\alpha} & \longmapsto & f_n = \dfrac{n^{\alpha-1}}{\Gamma(\alpha)} + \cdots & \text{(Theorem VI.1)} \\ f(z) = O((1-z)^{-\alpha}) & \longmapsto & f_n = O(n^{\alpha-1}) & \text{(Theorem VI.3 } (i)) \\ f(z) = o((1-z)^{-\alpha}) & \longmapsto & f_n = o(n^{\alpha-1}) & \text{(Theorem VI.3 } (ii)) \\ f(z) \sim (1-z)^{-\alpha} & \longmapsto & f_n \sim \dfrac{n^{\alpha-1}}{\Gamma(\alpha)} & \text{(Cor. VI.1).} \end{array}\right.$$

The important requirement is that the function should have an isolated singularity (the condition of Δ–analyticity) and that the asymptotic property of the function near its singularity should be valid in an area of the complex plane extending beyond the disk of convergence of the original series, (in a Δ–domain). Extensions to logarithmic powers and special cases like $\alpha \in \mathbb{Z}_{\le 0}$ are also, as we know, available. We let $\mathcal{S}$ denote the set of such singular functions:

$$\text{(28)} \qquad \mathcal{S} = \left\{(1-z)^{-\alpha}\lambda(z)^{\beta} \ \middle|\ \alpha, \beta \in \mathbb{C}\right\}, \qquad \lambda(z) := \frac{1}{z}\log\frac{1}{1-z}.$$

At this stage, we thus have available tools by which, starting from the expansion of a function at its singularity, also called *singular expansion*, one can justify the term-by-term transfer from an approximation of the function to an asymptotic estimate of the coefficients. We state:

THEOREM VI.4 (Singularity analysis, single singularity). *Let $f(z)$ be function analytic at 0 with a singularity at ζ, such that $f(z)$ can be continued to a domain of the form $\zeta \cdot \Delta_0$, for a Δ-domain Δ_0, where $\zeta \cdot \Delta_0$ is the image of Δ_0 by the mapping $z \mapsto \zeta z$. Assume that there exist two functions σ, τ, where σ is a (finite) linear combination of functions in $\mathcal{S}$ and $\tau \in \mathcal{S}$, so that*

$$f(z) = \sigma(z/\zeta) + O(\tau(z/\zeta)) \qquad \textit{as } z \to \zeta \textit{ in } \zeta\Delta_0.$$

Then, the coefficients of $f(z)$ satisfy the asymptotic estimate

$$f_n = \zeta^{-n}\sigma_n + O(\zeta^{-n}\tau_n^\star),$$

where $\sigma_n = [z^n]\sigma(z)$ has its coefficients determined by Theorems VI.1, VI.2 and $\tau_n^\star = n^{a-1}(\log n)^b$, if $\tau(z) = (1-z)^{-a}\lambda(z)^b$.

We observe that the statement is equivalent to $\tau_n^\star = [z^n]\tau(z)$, except when $a \in \mathbb{Z}_{\le 0}$ (when the $1/\Gamma(a)$ factor should be omitted). Also, generically, we have $\tau_n^\star = o(\sigma_n)$, so that orders of growth of functions at singularities are mapped to orders of growth of coefficients.

[2]The symbol '$\Longrightarrow$' represents an *unconditional* logical implication and is accordingly used in this book to represent the systematic correspondence between combinatorial specifications and generating function equations. In contrast, the symbol '$\longmapsto$' represents a mapping from functions to coefficients, under suitable *analytic conditions* as stated in Theorems VI.1–VI.3.

Let $f(z)$ be a function analytic at 0 whose coefficients are to be asymptotically analysed.
1. *Preparation.* This consists in locating dominant singularities and checking analytic continuation.

- **1a.** *Locate singularities.* Determine the dominant singularities of $f(z)$ (assumed not to be entire). Check that $f(z)$ has a single singularity ζ on its circle of convergence.
- **1b.** *Check continuation.* Establish that $f(z)$ is analytic in some domain of the form $\zeta\Delta_0$.

2. *Singular expansion.* Analyse the function $f(z)$ as $z \to \zeta$ in the domain $\zeta\Delta_0$ and determine in that domain an expansion of the form

$$f(z) \underset{z\to 1}{=} \sigma(z/\zeta) + O(\tau(z/\zeta)) \qquad \text{with} \quad \tau(z) \ll \sigma(z).$$

For the method to succeed, the functions σ and τ should belong to the standard scale of functions $\mathcal{S} = \{(1-z)^{-\alpha}\lambda(z)^{\beta}\}$, with $\lambda(z) := z^{-1}\log(1-z)^{-1}$.
3. *Transfer* Translate the main term term $\sigma(z)$ using the catalogues provided by TheoremsVI.1 and VI.2. Transfer the error term (Theorem VI.3) and conclude that

$$[z^n]f(z) \underset{n\to+\infty}{=} \zeta^{-n}\sigma_n + O\left(\zeta^{-n}\tau_n^{\star}\right),$$

where $\sigma_n = [z^n]\sigma(z)$ and $\tau_n^{\star} = [z^n]\tau(z)$ provided the corresponding exponent $\alpha \notin \mathbb{Z}_{\le 0}$ (otherwise, the factor $1/\Gamma(\alpha) = 0$ should be dropped).

FIGURE 7. A summary of the singularity analysis process (single dominant singularity).

PROOF. The normalized function $g(z) = f(z/\zeta)$ is singular at 1. It is Δ-analytic and satisfies the relation $g(z) = \sigma(z) + O(\tau(z))$ as $z \to 1$ within Δ_0. Theorem VI.3, (i) (the big-Oh transfer) applies to the O-error term. The statement follows finally since $[z^n]f(z) = \zeta^{-n}[z^n]g(z)$. $\square$

The statement of Theorem VI.4 can be concisely expressed by the correspondence:

$$(29)\quad f(z) \underset{z\to 1}{=} \sigma(z/\zeta) + O\left(\tau(z/\zeta)\right) \qquad \bullet\!\!\longrightarrow \qquad f_n \underset{n\to\infty}{=} \zeta^{-n}\sigma_n + O(\zeta^{-n}\tau_n^{\star}).$$

The conditions of analytic continuation and validity of the expansion in a Δ–domain are essential. Similarly, we have

$$(30)\qquad f(z) = \sigma\left(z/\zeta\right)) + o\left(\tau\left(z/\zeta\right)\right) \quad \bullet\!\!\longrightarrow \quad f_n = \zeta^{-n}\sigma_n + O(\zeta^{-n}\tau_n^{\star}),$$

as a simple consequence of Theorem VI.3, (ii) (little-oh transfer). The mappings (29) and (30) supplemented by the accompanying analysis constitute the heart of the *singularity analysis* process summarized in Figure 7.

Many of the functions commonly encountered in analysis are found to be Δ–continuable. This fact results from the property of the elementary functions (like $\sqrt{\ }$, $\log$, $\tan$) to be continuable to larger regions than what their expansions imply, as well as to the rich set of composition properties that analytic functions satisfy. Also, asymptotic expansions at a singularity initially determined along the real axis by elementary real analysis often hold in much wider regions of the complex plane. The singularity analysis process is then likely to be applicable to a large number of generating functions that are provided by the symbolic method—most notably the iterative structures described in Section IV. 4 (p. 236) In such cases, singularity analysis greatly refines the exponential growth estimates obtained in Theorem IV.8 (p. 237). The condition is that singular expansions should be of a suitably moderate[3] growth. We illustrate this situation now by treating combinatorial generating functions obtained by the symbolic methods of Chapters I and II, for which explicit expressions are available.

EXAMPLE 2. *Asymptotics of 2-regular graphs.* This example completes the discussion of Example 1, p. 363. The labelled class $\mathcal{C}$ of 2-regular graphs satisfies

$$\mathcal{C} = \mathrm{SET}(\mathrm{UCYC}_{\geq 3}(\mathcal{Z})) \qquad \Longrightarrow \qquad C(z) = \exp\left(\frac{1}{2}\left(\log(1-z)^{-1} - z - \frac{z}{2}\right)\right),$$

where UCYC is the undirected cycle construction (Note II.21, p. 124). For this example, we follow step by step the singularity analysis process as summarized in Figure 7.

1. *Preparation*. The function $C(z)$ being the product of $e^{-z/2-z^2/4}$ (that is entire) and of $(1-z)^{-1/2}$ (that is analytic in the unit disk) is itself analytic in the unit disk. Also, since $(1-z)^{-1/2}$ is Δ–analytic (it is well-defined and analytic in the complex plane slit along $\mathbb{R}_{\geq 1}$), $C(z)$ is itself Δ–analytic, with a singularity at $z = 1$.

2. *Singular expansion*. The asymptotic expansion of $C(z)$ near $z = 1$ is obtained starting from the standard (analytic) expansion of $e^{-z/2-z^2/4}$ at $z = 1$,

$$e^{-z/2-z^2/4} = e^{-3/4} + e^{-3/4}(1-z) + \frac{e^{-3/4}}{4}(1-z)^2 - \frac{e^{-3/4}}{12}(1-z)^3 + \cdots.$$

The factor $(1-z)^{-1/2}$ is its own asymptotic expansion, clearly valid in any Δ–domain. Performing the *multiplication* yields a complete expansion,

$$C(z) \sim \frac{e^{-3/4}}{\sqrt{1-z}} + e^{-3/4}\sqrt{1-z} + \frac{e^{-3/4}}{4}(1-z)^{3/2} - \frac{e^{-3/4}}{12}(1-z)^{5/2} + \cdots, \tag{31}$$

out of which terminating forms can be extracted.

3. *Transfer.* Take for instance the expansion of (31) limited to two terms plus an error term. The singularity analysis process allows the transfer of (31) to coefficients, which we can present in tabular form as follows:

$C(z)$	$c_n \equiv [z^n]C(z)$
$e^{-3/4}\dfrac{1}{\sqrt{1-z}}$	$e^{-3/4}\dbinom{n-1/2}{-1/2} \sim \dfrac{e^{-3/4}}{\sqrt{\pi n}}\left[1 - \dfrac{1}{8n} + \dfrac{1}{128n^2} + \cdots\right]$
$+\, e^{-3/4}\sqrt{1-z}$	$+e^{-3/4}\dbinom{n-3/2}{-3/2} \sim \dfrac{-e^{-3/4}}{2\sqrt{\pi n^3}}\left[1 + \dfrac{3}{8n} + \cdots\right]$
$+\, O((1-z)^{3/2})$	$+O(\dfrac{1}{n^{5/2}})$.

Terms are then collected with expansions suitably truncated to the coarsest error term, so that here a 3–term expansion results. In the sequel, we shall no longer need to detail such computations and we shall content ourselves with putting in parallel the function's expansion and the coefficient's expansion like in the following correspondence:

$$C(z) = \frac{e^{-3/4}}{\sqrt{1-z}} + e^{-3/4}\sqrt{1-z} + O((1-z)^{3/2}) \quad \longmapsto \quad c_n = \frac{e^{-3/4}}{\sqrt{\pi n}} - \frac{5e^{-3/4}}{8\sqrt{\pi n^3}} + O(\frac{1}{n^{5/2}}).$$

[3]For functions with fast growth at a singularity, the saddle-point method developed in Chapter VIII becomes effectual.

Here is a numerical check. Set $c_n^{(1)} := e^{-3/4}/\sqrt{\pi n}$ and let $c_n^{(2)}$ represent the sum of the first two terms of the expansion of c_n. One finds:

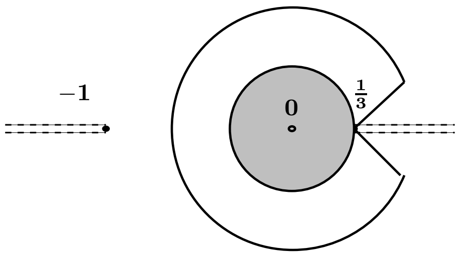
n	5	50	500
$n!c_n^{(1)}$	$\mathbf{14}.30212$	$\mathbf{1.1}462888618 \cdot 10^{63}$	$\mathbf{1.45}42120372 \cdot 10^{1132}$
$n!c_n^{(2)}$	$\mathbf{12}.51435$	$\mathbf{1.13196}02511 \cdot 10^{63}$	$\mathbf{1.4523942}721 \cdot 10^{1132}$
$n!c_n$	$\mathbf{12}$	$\mathbf{1.1319677968} \cdot 10^{63}$	$\mathbf{1.4523943224} \cdot 10^{1132}$

Clearly, a complete asymptotic expansion in descending powers of n can be obtained in this way. . . . . END OF EXAMPLE 2. □

EXAMPLE 3. *Asymptotics of unary–binary Trees and Motzkin numbers.* Unary-binary trees are unlabelled plane trees that admit the specification and OGF:

$$\mathcal{U} = \mathcal{Z}(\mathbf{1} + \mathcal{U} + \mathcal{U} \times \mathcal{U}) \quad \Longrightarrow \quad U(z) = \frac{1 - z - \sqrt{(1+z)(1-3z)}}{2z}.$$

(See Note I.36 (p. 64) and Subsection V. 3 (p. 297) for the lattice path version.) The GF $U(z)$ is singular at $z = -1$ and $z = \frac{1}{3}$, the dominant singularity being at $z = \frac{1}{3}$. By branching properties of the square-root function, $U(z)$ is analytic in a Δ–domain like the one depicted below:



Around the point $\frac{1}{3}$, a singular expansion is obtained by *multiplying* $(1-3z)^{1/2}$ and the analytic expansion of the factor $(1+z)^{1/2}/(2z)$. The singularity analysis process then applies and yields automatically:

$$U(z) = 1 - 3^{1/2}\sqrt{1-3z} + \mathcal{O}((1-3z)) \quad \longrightarrow \quad U_n = \sqrt{\frac{3}{4\pi n^3}}\, 3^n + \mathcal{O}(3^n n^{-2}).$$

Further terms in the singular expansion of $U(z)$ at $z = \frac{1}{3}$ provide additional terms in the asymptotic expression of the Motzkin numbers U_n, for instance,

$$U_n = \sqrt{\frac{3}{4\pi n^3}}\, 3^n \left(1 - \frac{15}{16\,n} + \frac{505}{512\,n^2} - \frac{8085}{8192\,n^3} + \frac{505659}{524288 n^4} + O\left(\frac{1}{n^5}\right)\right)$$

results from an expansion of $U(z)$ till $O((1-3z)^{11/2})$. . . . . END OF EXAMPLE 3. □

EXAMPLE 4. *Asymptotics of children's Rounds.* Stanley [**390**] has introduced certain combinatorial configurations that he has nicknamed "children's rounds": a round is a labelled set of directed cycles, each of which has a center attached. The specification and EGF are

$$\mathcal{R} = \text{SET}(\mathcal{Z} \star \text{CYC}(\mathcal{Z})) \quad \Longrightarrow \quad R(z) = \exp\left(z \log \frac{1}{1-z}\right) = (1-z)^{-z}.$$

The function $R(z)$ is analytic in the $\mathbb{C}$-plane slit alog $\mathbb{R}_{\geq 1}$, as is seen by elementary properties of the *composition* of analytic functions. The singular expansion at $z = 1$ is then mapped to an expansion for the coefficients:

$$R(z) = \frac{1}{1-z} + \log(1-z) + O((1-z)^{1/2}) \quad \longrightarrow \quad [z^n]R(z) = 1 - \frac{1}{n} + \mathcal{O}(n^{-3/2}).$$

A more detailed analysis yields

$$[z^n]R(z) = 1 - \frac{1}{n} - \frac{1}{n^2}(\log n + \gamma - 1) + O\left(\frac{\log^2 n}{n^3}\right),$$

and an expansion to any order can be easily obtained. .... END OF EXAMPLE 4. □

▷ **11.** *The asymptotic shape of the rounds numbers.* A complete asymptotic expansion has the form

$$[z^n]R(z) sim 1 - \sum_{j\geq 1} \frac{P_j(\log n)}{n^j},$$

where P_j is a polynomial of degree $j-1$. (The coefficients of P_j are rational combinations of powers of $\gamma, \zeta(2), \ldots, \zeta(j-1)$.) The successive terms in this expansion are easily obtained by a computer algebra program. ◁

EXAMPLE 5. *Asymptotics of coefficients of an elementary function.* Our final example is meant to show the way rather arbitrary compositions of basic functions can be treated by singularity analysis. Let $\mathcal{C} = \mathcal{Z} \star \mathrm{SEQ}(\mathcal{C})$ be the class of general labelled plane trees. Consider the labelled class defined by substitution

$$\mathcal{F} = \mathcal{C} \circ \mathrm{CYC}(\mathrm{CYC}(\mathcal{Z})) \qquad \Longrightarrow \qquad F(z) = C(L(L(z))).$$

There, $C(z) = \frac{1}{2}(1 - \sqrt{1-4z})$ and $L(z) = \log \frac{1}{1-z}$. Combinatorially, $\mathcal{F}$ is the class of trees in which nodes are replaced by cycles of cycles, a rather artificial combinatorial object, and

$$F(z) = \frac{1}{2}\left[1 - \sqrt{1 - 4\log\frac{1}{1 - \log\frac{1}{1-z}}}\right].$$

The problem is first to locate the dominant singularity of $F(z)$, then to determine its nature, which can be done inductively on the structure of $F(z)$. The dominant positive singularity ρ of $F(z)$ satisfies $L(L(\rho)) = \frac{1}{4}$ and one has

$$\rho = 1 - e^{e^{-1/4}-1} \doteq 0.198443,$$

given that $C(z)$ is singular at $\frac{1}{4}$ and $L(z)$ has positive coefficients. Since $L(L(z))$ is analytic at ρ, a local expansion of $F(z)$ is obtained next by *composition* of the singular expansion of $C(z)$ at $\frac{1}{4}$ with the standard Taylor expansion of $L(L(z))$ at ρ. We find

$$F(z) = \frac{1}{2} - C_1(\rho - z)^{1/2} + O((\rho - z)^{3/2}) \quad \longrightarrow \quad [z^n]F(z) = \frac{C_1 \rho^{-n+1/2}}{2\sqrt{\pi n^3}}\left[1 + O(\frac{1}{n})\right],$$

with $C_1 = e^{\frac{5}{8} - \frac{1}{2}e^{-1/4}} \doteq 1.26566$. .... END OF EXAMPLE 5. □

$\triangleright$ **12.** *The asymptotic number of trains.* Combinatorial trains have been introduced in Section IV. 4 as a way to exemplify the power of complex asymptotic methods. One finds that, at its dominant singularity ρ, the EGF $Tr(z)$ is of the form $Tr(z) \sim C/(1 - z/\rho)$, and, by singularity analysis,

$$[z^n]Tr(z) \sim 0.11768\,31406\,15497 \cdot 2.06131\,73279\,40138^n.$$

(This asymptotic approximation is good to 15 significant digits for $n = 50$, in accordance with the fact that the dominant singularity is a simple pole.) $\triangleleft$

VI. 5. Multiple singularities

The previous section has described in detail the analysis of functions with a single dominant singularity. The extension to functions that have *finitely many* (by necessity isolated) singularities on their circle of convergence follows along entirely similar lines. It parallels the situation of rational and meromorphic functions in Chapter IV (p. 250) and is technically simple, the net result being:

> *In the case of multiple singularities, the separate contributions from each of the singularities, as given by the basic singularity analysis process, must be added up.*

Like in (28), we let $\mathcal{S}$ be the standard scale of functions singular at 1, namely

$$\mathcal{S} = \left\{ (1-z)^{-\alpha}\lambda(z)^{\beta} \;\middle|\; \alpha, \beta \in \mathbb{C} \right\}, \qquad \lambda(z) := \frac{1}{z}\log\frac{1}{1-z}.$$

THEOREM VI.5 (Singularity analysis, multiple singularities). *Let $f(z)$ be analytic in $|z| < \rho$ and have a finite number of singularities on the circle $|z| = \rho$ at points $\zeta_j = \rho e^{i\theta_j}$, for $j = 1\,..\,r$. Assume that there exists a Δ–domain Δ_0 such that $f(z)$ is analytic in the indented disc*

$$\mathbf{D} = \bigcap_{j=1}^{r} (\zeta_j \cdot \Delta_0),$$

with $\zeta \cdot \Delta_0$ the image of Δ_0 by the mapping $z \mapsto \zeta z$.

Assume that there exists r functions $\sigma_1, \ldots, \sigma_r$, each a linear combination of elements from $\mathcal{S}$ and a function $\tau \in \mathcal{S}$ such that

$$f(z) = \sigma_j(z/\zeta_j) + O\left(\tau(z/\zeta_j)\right) \qquad \text{as } z \to \zeta_j \text{ in } \mathbf{D}.$$

Then the coefficients of $f(z)$ satisfy the asymptotic estimate

$$f_n = \sum_{j=1}^{r} \zeta_j^{-n}\sigma_{j,n} + O\left(\rho^{-n}\tau_n^{\star}\right),$$

where each $\sigma_{j,n} = [z^n]\sigma_j(z)$ has its coefficients determined by Theorems VI.1, VI.2 and $\tau_n^* = n^{a-1}(\log n)^b$, if $\tau(z) = (1-z)^{-a}\lambda(z)^b$.

A function analytic in a domain like $\mathbf{D}$ is sometimes said to be star-continuable, a notion that is the natural generalization of Δ–analyticity for functions with several dominant singularities. Also, a similar statement holds with o-error terms replacing O's.

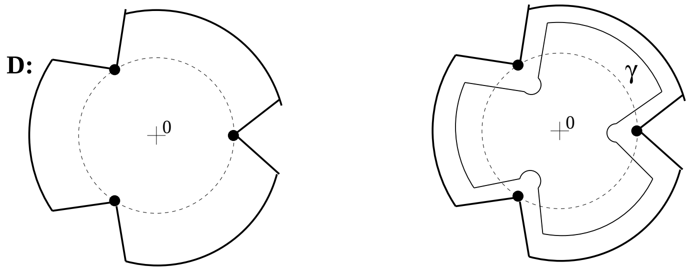


FIGURE 8. Multiple singularities ($r = 3$): analyticity domain ($\mathbf{D}$, left) and composite integration contour (γ, right).

PROOF. Like in the case of a single singularity, the proof bases itself on Cauchy's coefficient formula

$$f_n = [z^n] \int_\gamma f(z) \frac{dz}{z^{n+1}},$$

where a composite contour γ depicted on Figure 8 is used. Estimates on each part of the contour obey exactly the same principles as in the proof of Theorems VI.1–VI.3. Let $\gamma^{(j)}$ be the open loop around ζ_j that comes from the outer circle, winds about ζ_j and joins again the outer circle; let r be the radius of the outer circle.

— The contribution along the arcs of the outer circle is $O(r^{-n})$, that is, exponentially small.
— The contribution along the loop $\gamma^{(1)}$ (say) separates into

$$\frac{1}{2i\pi} \int_{\gamma^{(1)}} f(z) \frac{dz}{z^{n+1}} = I' + I''$$
$$I' := \frac{1}{2i\pi} \int_{\gamma^{(1)}} \sigma_1(z/\zeta_1) \frac{dz}{z^{n+1}}, \quad I'' := \frac{1}{2i\pi} \int_{\gamma^{(1)}} \left(f(z) - \sigma_1(z/\zeta_1)\right) \frac{dz}{z^{n+1}}.$$

The quantity I' is estimated by extending the open loop to infinity by the same method as in the proof of Theorems VI.1 and VI.2: it is found to equal $\zeta_1^{-n}\sigma_{1,n}$ plus an exponentially small term. The quantity I'', corresponding to the error term, is estimated by the same bounding technique as in the proof of Theorem VI.3 and is found to be $O(\rho^n \tau_n^\star)$.

Collecting the various contributions completes the proof of the statement. □

Theorem VI.5 expresses, that in the case of multiple singularities, each dominant singularity can be analysed separately; the singular expansions are then each transferred to coefficients, and the corresponding asymptotic contributions are finally collected. Two examples illustrating the process follow.

EXAMPLE 6. *An artificial example.* Let us demonstrate the *modus operandi* on the simple function

$$g(z) = \frac{e^z}{\sqrt{1-z^2}}. \tag{32}$$

There are two singularities at $z=+1$ and $z=-1$, with

$$g(z) \sim \frac{e}{\sqrt{2}\sqrt{1-z}} \quad z \to +1 \qquad \text{and} \qquad g(z) \sim \frac{e^{-1}}{\sqrt{2}\sqrt{1+z}} \quad z \to -1.$$

The function is clearly star-continuable with the singular expansions valid in the star domain. We have

$$[z^n]\frac{e}{\sqrt{2}\sqrt{1-z}} \sim \frac{e}{\sqrt{2\pi n}} \qquad \text{and} \qquad [z^n]\frac{e^{-1}}{\sqrt{2}\sqrt{1+z}} \sim \frac{e^{-1}(-1)^n}{\sqrt{2\pi n}}.$$

To get the coefficient $[z^n]g(z)$, it suffices to add up these two contributions (by Theorem VI.5), so that

$$[z^n]g(z) \sim \frac{1}{\sqrt{2\pi n}}[e+(-1)^n e^{-1}].$$

If expansions at $+1$ (respectively -1) are written with an error term, which is of the form $O((z-1)^{1/2})$ (respectively, $O((z+1)^{1/2})$, there results an estimate of the coefficients $g_n = [z^n]g(z)$, which can be put under the form

$$g_{2n} = \frac{\cosh(1)}{\sqrt{\pi n}} + O\left(n^{-3/2}\right), \qquad g_{2n+1} = \frac{\sinh(1)}{\sqrt{\pi n}} + O\left(n^{-3/2}\right).$$

This makes explicit the dependency of the asymptotic form of g_n on the parity of the index n. Clearly a full asymptotic expansion can be obtained. .... END OF EXAMPLE 6. □

EXAMPLE 7. *Permutations with cycles of odd length.* Consider the specification and EGF

$$\mathcal{F} = \textsc{Set}(\textsc{Cyc}_{\text{odd}}(\mathcal{Z})) \quad \Longrightarrow \quad F(z) = \exp\left(\frac{1}{2}\log\frac{1+z}{1-z}\right) = \sqrt{\frac{1+z}{1-z}}.$$

The singularities of f are at $z=+1$ and $z=-1$, the function being obviously star-continuable. By singularity analysis (Theorem VI.5), we have automatically:

$$F(z) = \begin{cases} \dfrac{2^{1/2}}{\sqrt{1-z}} + O\left((1-z)^{1/2}\right) & (z \to 1) \\ O\left((1+z)^{1/2}\right) & (z \to -1) \end{cases} \quad \longmapsto \quad [z^n]F(z) = \frac{2^{1/2}}{\sqrt{\pi n}} + O\left(n^{-3/2}\right).$$

For the next asymptotic order, the singular expansions

$$F(z) = \begin{cases} \dfrac{2^{1/2}}{\sqrt{1-z}} - 2^{-3/2}\sqrt{1-z} + O((1-z)^{3/2}) & (z \to 1) \\ 2^{-1/2}\sqrt{1+z} + O((1+z)^{3/2}) & (z \to -1) \end{cases}$$

yield

$$[z^n]F(z) = \frac{2^{1/2}}{\sqrt{\pi n}} - \frac{(-1)^n 2^{-3/2}}{\sqrt{\pi n^3}} + O(n^{-5/2}).$$

This example illustrates the occurrence of singularities that have different weights, in the sense of being associated with different exponents. .... END OF EXAMPLE 7. □

The discussion of multiple dominant singularities ties well with the earlier discussion of Subsection IV. 6.1, p. 250. In the periodic case where the dominant singularities are at roots of unity, different regimes manifest themselves cyclically depending on congruence properties of the index n, like in the two examples above. When the dominant singularities have arguments that are not commensurate to π (a comparatively rare situation), aperiodic fluctuations appear, in which case the situation is similar to what was already discussed, regarding rational and meromorphic functions, in Subsection IV. 6.1.

VI. 6. Intermezzo: functions of singularity analysis class

Let us say that a function is of *singularity analysis class*, or *SA-class* for short, if its satisfies the conditions of singularity analysis, as expressed by Theorem VI.4 (single dominant singularity) or Theorem VI.5 (multiple dominant singularities). The property of being of SA-class is preserved by several basic operations of analysis.: we have already seen this feature in passing, when determining singular expansions of functions obtained by sums, products, or compositions in Examples 2–5.

As a starting example, it is easily recognized that the assumptions of Δ-analyticity for two functions $f(z), g(z)$ accompanied by the singular expansions

$$f(z) \underset{z\to 1}{\sim} c(1-z)^{-\alpha}, \qquad g(z) \underset{z\to 1}{\sim} d(1-z)^{-\delta},$$

and the condition $\alpha, \delta \notin \mathbb{Z}_{\leq 0}$ imply for the coefficients of the sum

$$[z^n]\,(f(z)+g(z)) = \begin{cases} c\dfrac{n^{\alpha-1}}{\Gamma(\alpha)} & \alpha > \delta \\ (c+d)\dfrac{n^{\alpha-1}}{\Gamma(\alpha)} & \alpha = \delta, \quad c+d \neq 0 \\ d\dfrac{n^{\delta-1}}{\Gamma(\delta)} & \alpha < \delta. \end{cases}$$

Similarly, for products, we have

$$[z^n]\,(f(z)g(z)) \sim cd\frac{n^{\alpha+\delta-1}}{\Gamma(\alpha+\delta)},$$

provided $\alpha + \delta \notin \mathbb{Z}_{\geq 0}$.

The simple considerations above illustrate the robustness of singularity analysis. They also indicate that properties are easy to state in the generic case where no negative integral exponents are present. However, if all cases are to be covered, there can easily be an explosion of the number of particular situations, which may render somewhat clumsy the enunciation of complete statements. Accordingly, in what follows, we shall largely confine ourselves to generic cases, as long as these suffice to develop the important mathematical technique at stake for each particular problem.

In the remainder of this chapter, we proceed to enlarge the class of functions recognized to be of SA-class, keeping in mind the needs of analytic combinatorics. The following types of functions are treated in later sections.

— *Inverse functions* (Section VI. 7). The inverse of analytic function is, under mild conditions, of SA-class. In the case of functions attached to simple varieties of trees (corresponding to the inversion of $y/\phi(y)$), the singular expansion invariably has an exponent of $\frac{1}{2}$ attached to it (a square-root singularity). This applies in particular to the Cayley tree function, in terms of which many combinatorial structures and parameters can be analysed.

— *Polylogarithms* (Section VI. 8). These functions are the generating functions of simple arithmetic sequences like (n^θ) for an arbitrary $\theta \in \mathbb{C}$. The fact that polylogarithms are of SA-class opens the possibility of estimating a large number of sums, which involve both combinatorial terms (e.g., binomial coefficients) and elements like $\sqrt{n}$ and $\log n$. Such sums appear recurrently in the analysis of cost functionals of combinatorial structures and algorithms.

— *Composition* (Section VI. 9). The composition of functions of SA-class often proves to be itself of SA-class. This fact has implications for the analysis of composition schemas and makes possible a broad extension of the supercritical sequence schema treated in Section V. 4, (p. 315).

— *Differentiation, integration, and Hadamard products* (Section VI. 10). These are three operations on analytic function that preserve the property for a function to be of SA-class. Applications are given to tree recurrences and to multidimensional walk problems.

A main theme of this book is that elementary combinatorial classes tend to have generating functions whose singularity structure is strongly constrained—in many cases, singularities are isolated. The singularity analysis process is then a prime technique for extracting asymptotic information from such generating functions.

VI. 7. Inverse functions

Recursively defined structures lead to functional equations whose solutions may often be analysed locally near singularities. An important case is the one of functions defined by inversion. It includes the Cayley tree function as well as all generating functions associated to simple varieties of trees (Subsections I. 5.1 (p. 61), II. 5.1 (p. 117), and III. 6.2 (p. 182)). A common pattern in this context is the appearance of singularities of the square-root type, which proves to be universal amongst a broad class of problems involving trees and tree-like structures. Accordingly, by singularity analysis, the square-root singularity induces subexponential factors of the asymptotic form $n^{-3/2}$ in coefficients' expansions.

Inverse functions. Singularities of functions defined by inversion have been located in Subsection IV. 7.1 (p. 262) and our treatment will proceed from there. The goal is to estimate the coefficients of a function defined implicitly by an equation of the form

$$y(z) = z\phi(y(z)) \qquad \text{or equivalently} \qquad z = \frac{y(z)}{\phi(y(z))}. \tag{33}$$

The problem of solving (33) is one of functional inversion: we have seen (Lemmas IV.2 and IV.3, pp. 262–263) that *an analytic function admits locally an analytic inverse if and only if its first derivative is nonzero*. We operate here under the following assumptions:

— **Condition** $(\mathbf{H_1})$. The function $\phi(u)$ is nonlinear[4], analytic at $u = 0$ where it has nonnegative coefficients, and satisfies $\phi(0) \neq 0$:

$$\phi(0) \neq 0, \quad [u^n]\phi(u) \geq 0, \quad \phi(u) \not\equiv \phi_0 + \phi_1 u. \tag{34}$$

(As a consequence, the inversion problem is well defined around 0.)

— **Condition** $(\mathbf{H_2})$. Within the *open* disc of convergence of ϕ at 0, $|z| < R$, there exists a (necessarily unique) positive solution to the *characteristic equation*:

$$\exists \tau,\ 0 < \tau < R, \quad \phi(\tau) - \tau\phi'(\tau) = 0. \tag{35}$$

(By Proposition IV.5, p. 265, existence is granted as soon as $\lim x\phi'(x)/\phi(x) > 1$ as $x \to R^-$,with R the radius of convergence of ϕ at 0.)

Then (by Proposition IV.5, p. 265), the radius of convergence of $y(z)$ is the corresponding positive value ρ of z such that $y(\rho) = \tau$, that is to say,

$$\rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}. \tag{36}$$

We start with a calculation indicating in a plain context the occurrence of a square-root singularity.

EXAMPLE 8. *A simple analysis of the Cayley tree function.* The situation corresponding to the function $\phi(u) = e^u$, so that $y(z) = ze^{y(z)}$ (the Cayley tree function $T(z)$), is typical of general analytic inversion. From (35), the radius of convergence of $y(z)$ is $\rho = e^{-1}$ corresponding to $\tau = 1$. The image of a circle in the y–plane, centered at the origin and having radius $r < 1$, by the function ye^{-y} is a curve of the z–plane that properly contains the circle $|z| = re^{-r}$ (see Figure 9) as $\phi(y) = e^y$, which has nonnegative coefficients, satisfies

$$\left|\phi(re^{i\theta})\right| \leq \phi(r) \qquad \text{for all } \theta \in [-\pi, +\pi],$$

the inequality being strict for all $\theta \neq 0$. The following observation is the key to analytic continuation: *Since the first derivative of $y/\phi(y)$ vanishes at 1, the mapping $y \mapsto y/\phi(y)$ is angle-doubling, so that the image of the circle of radius 1 is a curve $\mathcal{C}$ that has a cusp at* $\rho = e^{-1}$. (See Figure 9; Notes 16 and 17 provide interesting generalizations.)

This geometry shows that the solution of $z = ye^{-y}$ is uniquely defined for z inside $\mathcal{C}$. Thus, $y(z)$ is Δ–analytic. A singular expansion for $y(z)$ is then derived from reversion of the power series expansion of $z = ye^{-y}$. We have

$$ye^{-y} = e^{-1} - \frac{e^{-1}}{2}(y-1)^2 + \frac{e^{-1}}{3}(y-1)^3 - \frac{e^{-1}}{8}(y-1)^4 + \cdots,$$

so that solving for y gives

$$y - 1 = \sqrt{2}(1-ez)^{1/2} + \frac{2}{3}(1-ez) + O((1-ez)^{3/2}),$$

and a full expansion can be obtained. .... END OF EXAMPLE 8. □

[4]The nonlinearity of ϕ means that ϕ'' is not identically 0. This condition only excludes the case $\phi(u) = \phi_0 + \phi_1 u$, corresponding to $y(z) = \phi_0 z/(1 - \phi_1 z)$.

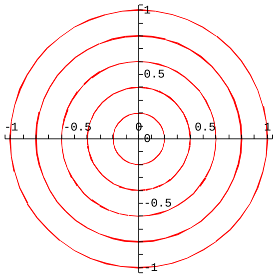
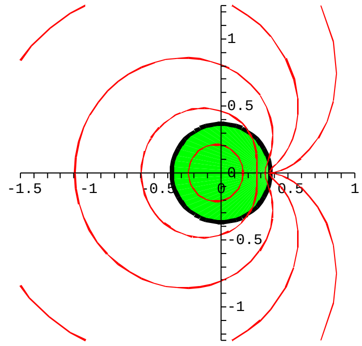
$y:$  $\longrightarrow$ $z:$ 

FIGURE 9. The images of concentric circles by the mapping $y \mapsto z = ye^{-y}$. It is seen that $y \mapsto z = ye^{-y}$ is injective on $|y| \leq 1$ with an image extending beyond the circle $|z| = e^{-1}$ [in grey], so that the inverse function $y(z)$ is analytically continuable in a Δ–domain around $z = e^{-1}$. Since the direct mapping ye^{-y} is quadratic at 1 (with value e^{-1}), the inverse function has a square-root singularity at e^{-1} (with value 1).

Analysis of inverse functions. The calculation of Example 8 now needs to be extended to the general case, $y = z\phi(y)$. This involves three steps: (i) all the dominant singularities are to be located; (ii) analyticity of $y(z)$ in a Δ–domain must be established; (iii) the singular expansion, obtained formally so far and involving a square-root singularity, needs to be determined. Step (i) requires a special discussion and is related to periodicities.

A simple example like $\phi(u) = 1 + u^2$ (binary trees), for which

$$y(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z},$$

shows that $y(z)$ may have several dominant singularities—here, two conjugate singularities at $-\frac{1}{2}$ and $+\frac{1}{2}$. The conditions for this to happen are rather simple. Let us say that a function analytic at 0, $f(u)$, is p*–periodic* if $f(u) = u^r g(u^p)$ for some power series g (see p. 253). A function is called *periodic* if it is p–periodic from some $p \geq 2$ and *aperiodic* otherwise. An elementary argument developed in Note 15 below shows that that periodicity does not occur for $y(z)$ unless $\phi(u)$ is itself periodic, a case which turns out to be easily reducible to the aperiodic situation.

THEOREM VI.6 (Singular Inversion). *Let ϕ be a nonlinear function satisfying the conditions* $(\mathbf{H_1})$ *and* $(\mathbf{H_2})$ *of Equations* (34) *and* (35)*, and let $y(z)$ be the solution of $y = z\phi(y)$ satisfying $y(0) = 0$. Then, the quantity $\rho = \tau/\phi(\tau)$ is the radius of convergence of $y(z)$ at 0 (with τ the root of the characteristic equation), and the singular expansion of $y(z)$ near ρ is of the form*

$$y(z) = \tau - d_1\sqrt{1 - z/\rho} + \sum_{j\geq 2}(-1)^j d_j (1 - z/\rho)^{j/2}, \qquad d_1 := \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}},$$

with the d_j being some computable constants.

Assume that, in addition, ϕ is aperiodic[5]. *Then, one has*

$$[z^n]y(z) \sim \sqrt{\frac{\phi(\tau)}{2\phi''(\tau)}} \frac{\rho^{-n}}{\sqrt{\pi n^3}} \left(1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k}\right),$$

for a family e_k of computable constants.

PROOF. Proposition IV.5, p. 265, shows that ρ is indeed the radius of convergence of $y(z)$. The Singular Inversion Lemma (Lemma IV.3, p. 263) also shows that $y(z)$ can be continued to a neighbourhood of ρ slit along the ray $\mathbb{R}_{\geq \rho}$.

The singular expansion at ρ is determined like in Example 8. Indeed, the relation between z and y, in the vicinity of $(z, y) = (\rho, \tau)$, may be put under the form

$$(37) \qquad \rho - z = H(y), \qquad \text{where} \quad H(y) := \left(\frac{\tau}{\phi(\tau)} - \frac{y}{\phi(y)}\right),$$

the function $H(y)$ in the right hand side being such that $H(\tau) = H'(\tau) = 0$. Thus, the dependency between y and z is locally a quadratic one:

$$\rho - z = \frac{1}{2!}H''(\tau)(y-\tau)^2 + \frac{1}{3!}H'''(\tau)(y-\tau)^3 + \cdots .$$

When this relation is locally inverted: a square-root appears:

$$-\sqrt{\rho - z} = \sqrt{\frac{H''(\tau)}{2}}(y-\tau)\left[1 + c_1(y-\tau) + c_2(y-\tau)^2 + ...\right].$$

The determination with a $-\sqrt{}$ should be chosen there as $y(z)$ increases to τ^- as $z \to \rho^-$. This implies, by solving with respect to $y - \tau$, the relation

$$y - \tau \sim -d_1^\star(\rho - z)^{1/2} + d_2^\star(\rho - z) - d_3^\star(\rho - z)^{3/2} + \cdots ,$$

where $d_1^\star = \sqrt{2/H''(\tau)}$ with $H''(\tau) = \tau\phi''(\tau)/\phi(\tau)^2$. The singular expansion at ρ results.

There now remains to exclude the possibility for $y(z)$ to have singularities other than ρ on the circle $|z| = \rho$. Observe that $y(\rho)$ is well defined (in fact $y(\rho) = \tau$), so that the series representing $y(z)$ converges at ρ as well as on the whole circle (given positivity of the coefficients). If $\phi(z)$ is aperiodic, then so is $y(z)$. Consider any point ζ such that $|\zeta| = \rho$ and $\zeta \neq \rho$ and set $\eta = y(\zeta)$. We then have $|\eta| < \tau$ (by the Daffodil Lemma: Lemma IV.1, p. 253). The function $y(z)$ is analytic at ζ by virtue of the Analytic Inversion Lemma (Lemma IV.2, p. 262) and the property that

$$\left.\frac{d}{dy}\frac{y}{\phi(y)}\right|_{y=\eta} \neq 0.$$

(This last property derives from the fact that the numerator of the quantity on the left,

$$\phi(\eta) - \eta\phi'(\eta) = \phi_0 - \phi_2\eta^2 - 2\phi_3\eta_3 - 3\phi_4\eta^4 - \cdots ,$$

cannot vanish, by the triangle inequality since $|\eta| < \tau$.) Thus, under the aperiodicity assumption, $y(z)$ is analytic on the circle $|z| = \rho$ punctured at ρ. The expansion of the coefficients then results from basic singularity analysis. □

[5]If ϕ has maximal period p, then one must restrict n to $n \equiv 1 \bmod p$; in that case, there is an extra factor of p in the estimate of y_n: see Note 15 and Equation (38).

Type	$\phi(u)$	*Sing. expansion of* $y(z)$	*Coeff.* $[z^n]y(z)$
binary	$(1+u)^2$	$1-4\sqrt{\frac{1}{4}-z}+\cdots$	$\dfrac{4^n}{\sqrt{\pi n^3}}+O(n^{-5/2})$
unary-binary	$1+u+u^2$	$1-3\sqrt{\frac{1}{3}-z}+\cdots$	$\dfrac{3^{n+1/2}}{2\sqrt{\pi n^3}}+O(n^{-5/2})$
general	$(1-u)^{-1}$	$\frac{1}{2}-\sqrt{\frac{1}{4}-z}$	$\dfrac{4^{n-1}}{\sqrt{\pi n^3}}+O(n^{-5/2})$
Cayley	e^u	$1-\sqrt{2e}\sqrt{e^{-1}-z}+\cdots$	$\dfrac{e^n}{\sqrt{2\pi n^3}}+O(n^{-5/2})$

FIGURE 10. Singularity analysis some of simple varieties of trees.

Figure 10 provides a table of the most basic varieties of simple trees and the corresponding asymptotic estimates. With Theorem VI.6, we now have available a powerful method that permits us to analyse not only implicitly defined functions but also expressions built upon them. This fact will be put to good use in Chapter VII, when analysing a number of parameters associated to simple varieties of trees.

$\triangleright$ **13.** *Computability of singular expansions.* Define

$$h(w) := \sqrt{\frac{\tau/\phi(\tau) - w/\phi(w)}{(\tau-w)^2}},$$

so that $y(z)$ satisfies $\sqrt{\rho - z} = (\tau - y)h(y)$. The singular expansion of y can then be deduced by Lagrange inversion from the expansion of the negative powers of $h(w)$ at $w = \tau$. This technique yields for instance explicit forms for coefficients in the singular expansion of $y = ze^y$. $\triangleleft$

$\triangleright$ **14.** *Stirling's formula via singularity analysis.* The solution to $T = ze^T$ analytic at 0 is the Cayley tree function. It satisfies $[z^n] = n^{n-1}/n!$ (by Lagrange inversion) and, at the same time, its singularity is known from Theorem VI.6. As a consequence:

$$\frac{n^{n-1}}{n!} \sim \frac{e^n}{\sqrt{2\pi n^3}}\left(1 - \frac{1}{12}\,n^{-1} + \frac{1}{288}\,n^{-2} + \frac{139}{51840}\,n^{-3} - \cdots\right).$$

Thus Stirling's formula *also* results from singularity analysis. $\triangleleft$

$\triangleright$ **15.** *Periodicities.* Assume that $\phi(u) = \psi(u^p)$with ψ analytic at 0 and $p \geq 2$. Let $y = y(z)$ be the root of $y = z\phi(y)$. Set $Z = z^p$ and let $Y(Z)$ be the root of $Y = Z\psi(Y)^p$. One has by construction $y(z) = Y(z^p)^{1/p}$, given that $y^p = z^p\phi(y)^p$. Since $Y(Z) = Y_1Z + Y_2Z^2 + \cdots$, we verify that the nonzero coefficients of $y(z)$ are amongst those of index $1, 1+p, 1+2p, \ldots$.

If p is chosen maximal, then $\psi(u)^p$ is aperiodic. Then Theorem VI.6 applies to $Y(Z)$: the function $Y(Z)$ is analytically continuable beyond its dominant singularity at $Z = \rho^p$; it has a square root singularity at ρ^p and no other singularity on $|Z| = \rho^p$. Also, since $Y = Z\psi(Y)^p$, the function $Y(Z)$ cannot vanish on $|Z| \leq \rho^p$, $Z \neq 0$. Thus, $Y(Z)^{1/p}$ is analytic in $|Z| \leq \rho^p$, except at ρ^p where it has a $\sqrt{}$ branch point. All computations done, we find that

$$(38) \qquad [z^n]y(z) \sim p \cdot \frac{d_1\rho^{-n}}{2\sqrt{\pi n^3}} \qquad \text{when} \quad n \equiv 1 \pmod{p}.$$

The argument also shows that $y(z)$ has p conjugate roots on its circle of convergence. (This is a kind of Perron-Frobenius property for periodic tree functions.) $\triangleleft$

$\triangleright$ **16.** *Boundary cases I.* The case when τ lies on the boundary of the disc of convergence of ϕ may lead to asymptotic estimates differing from the usual $\rho^{-n}n^{-3/2}$ prototype. Without loss of generality, take ϕ aperiodic to have radius of convergence equal to 1 and assume that ϕ is of the form

$$\phi(u) = u + c(1-u)^\alpha + o((1-u)^\alpha), \qquad \text{with} \quad 1 < \alpha \leq 2, \tag{39}$$

as u tends to 1 with $|u| < 1$. The solution of the characteristic equation $\phi(\tau) - \tau\phi'(\tau) = 0$ is then $\tau = 1$. The function $y(z)$ defined by $y = z\phi(y)$ is Δ–analytic (by a mapping argument similar to the one exemplified by Figure 9 and related to the fact that ϕ "multiplies" angles near 1). The singular expansion of $y(z)$ and the coefficients then satisfy

$$y(z) = 1 - c^{-1/\alpha}(1-z)^{1/\alpha} + o\left((1-z)^{1/\alpha}\right) \quad \bullet\!\!\longrightarrow \quad y_n \sim c^{-1/\alpha}\frac{n^{-1/\alpha-1}}{-\Gamma(-1/\alpha)}. \tag{40}$$

[The case $\alpha = 2$ was first observed by Janson [**243**]. Trees with $\alpha \in (1,2)$ have been investigated in connection with stable Lévy processes [**116**]. The singular exponent $\alpha = \frac{3}{2}$ occurs for instance in planar maps (Chapter VII), so that GFs with coefficients of the form $\rho^{-n}n^{-5/3}$ can arise, when considering trees whose nodes are themselves maps.] $\triangleleft$

$\triangleright$ **17.** *Boundary cases II.* Let $\phi(u)$ be the probability generating function of a random variable X with mean equal to 1 and such that $\phi_n \sim \lambda n^{-\alpha-1}$, with $1 < \alpha < 2$. Then, by a complex version of an Abelian theorem (see, e.g., [**54**, §1.7] and [**152**]), the singular expansion (39) holds when $u \to 1$, $|u| < 1$, within a cone, so that the conclusions of (40) hold in that case. Similarly, if $\phi''(1)$ exists, meaning that X has a second moment, then the estimate (40) holds with $\alpha = 2$, and then coincides with what Theorem VI.6 predicts [**243**]. (In probabilistic terms, the condition of Theorem VI.6 is equivalent to postulating the existence of exponential moments for the one-generation offspring distribution.) $\triangleleft$

VI. 8. Polylogarithms

Expressions involving sequences like $(\sqrt{n})$ or $(\log n)$ can be subjected to singularity analysis. The starting point is the definition of the generalized *polylogarithm* $\operatorname{Li}_{\alpha,r}$, where α is an arbitrary complex number and r a nonnegative integer:

$$\operatorname{Li}_{\alpha,r}(z) := \sum_{n\geq 1} (\log n)^r \frac{z^n}{n^\alpha},$$

The series converges for $|z| < 1$, so that the function $\operatorname{Li}_{\alpha,r}$ is *a priori* analytic in the unit disc. The quantity $\operatorname{Li}_{1,0}(z)$ is the usual logarithm, $\log(1-z)^{-1}$, hence the established name, polylogarithm, assigned to these functions [**290**]. In what follows, we make use of the abbreviation $\operatorname{Li}_{\alpha,0}(z) \equiv \operatorname{Li}_\alpha(z)$, so that $\operatorname{Li}_1(z) \equiv \operatorname{Li}_{1,0}(z) \equiv \log(1-z)^{-1}$ is the GF of the sequence $(1/n)$. Similarly , $\operatorname{Li}_{0,1}$ is the GF of the sequence $(\log n)$ and $\operatorname{Li}_{-1/2}(z)$ is the GF of the sequence $(\sqrt{n})$.

Polylogarithms are continuable to the whole of the complex plane slit along the ray $\mathbb{R}_{\geq 1}$, a fact established early in the twentieth century by Ford [**187**]. They are of SA-class [**147**] and their singular expansions involve the Riemann zeta function defined by

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

for $\Re(s) > 1$, and by analytic continuation elsewhere [**412**].

THEOREM VI.7 (Singularities of polylogarithms). *For all $\alpha \in \mathbb{Z}$ and $r \in \mathbb{Z}_{\geq 0}$, the function $\mathrm{Li}_{\alpha,r}(z)$ is analytic in the slit plane $\mathbb{C} \setminus \mathbb{R}_{\geq 1}$. For $\alpha \notin \{1, 2, \ldots\}$, there exists an infinite singular expansion (with logarithmic terms when as $r > 0$) given by the two rules:*

$$(41)\qquad \begin{cases} \mathrm{Li}_\alpha(z) & \sim \Gamma(1-\alpha)w^{\alpha-1} + \displaystyle\sum_{j\geq 0} \frac{(-1)^j}{j!}\zeta(\alpha - j)w^j, \ w := \sum_{\ell=1}^{\infty} \frac{(1-z)^\ell}{\ell} \\ \mathrm{Li}_{\alpha,r}(z) & = (-1)^r \dfrac{\partial^r}{\partial \alpha^r} \mathrm{Li}_\alpha(z) \qquad (r \geq 0). \end{cases}$$

The expansion of Li_α is conveniently described by the composition of two expansions, with the expansion of $w = \log z$ at $z = 1$, namely $w = (1-z) + \frac{1}{2}(1-z)^2 + \cdots$, to be substituted inside the expansion into powers of w. The exponents of $(1-z)$ involved in the resulting expansion are $\{\alpha - 1, \alpha, \ldots\} \cup \{0, 1, \ldots\}$. For $\alpha < 1$, the main asymptotic term of $\mathrm{Li}_{\alpha,r}$ is

$$\mathrm{Li}_{\alpha,r} \sim \Gamma(1-\alpha)(1-z)^{\alpha-1}L^r(z), \qquad L(z) := \log \frac{1}{1-z},$$

while, for $\alpha > 1$, we have $\mathrm{Li}_{\alpha,r} \sim \zeta^{(r)}(\alpha)$, since the sum converges.

PROOF. The analysis crucially relies on the Mellin transform (see APPENDIX B: *Mellin transform*, p. 674). We start with the case $r = 0$ and consider several ways in which z may approach the singularity 1. Step (i) below describes the main ingredient needed in *obtaining* the expansion, the subsequent steps being only required for *justifying* it in larger regions of the complex plane.

(i) *When $z \to 1^-$ along the real line*: Set $w = -\log z$ and introduce

$$(42)\qquad \Lambda(w) := \mathrm{Li}_\alpha(e^{-w}) = \sum_{n\geq 1} \frac{e^{-nw}}{n^\alpha}.$$

This is a *harmonic sum* in the sense of Mellin transform theory, so that the Mellin transform of Λ satisfies ($\Re(s) > \max(0, 1-\alpha)$)

$$(43)\qquad \Lambda^\star(s) \equiv \int_0^\infty \Lambda(w)w^{s-1}\,dw = \zeta(s+\alpha)\Gamma(s).$$

The function $\Lambda(w)$ can be recovered from the inverse Mellin integral,

$$(44)\qquad \Lambda(w) = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} \zeta(s+\alpha)\Gamma(s)w^{-s}\,ds,$$

with c is taken in the half-plane in which $\Lambda^\star(s)$ is defined. There are poles at $s = 0, -1, -2, \ldots$ due to the Gamma factor and a pole at $s = 1-\alpha$ due to the zeta function. Take d to be of the form $-m - \frac{1}{2}$ and smaller than $1 - \alpha$. Then, a standard residue calculation, taking into account poles to the left of c and based on

$$(45)\qquad \begin{aligned} \Lambda(w) \ &= \sum_{s_0 \in \{0,-1,\ldots,-m\}\cup\{1-\alpha\}} \mathrm{Res}\left(\zeta(s+\alpha)\Gamma(s)w^{-s}\right)_{s=s_0} \\ &\qquad + \frac{1}{2i\pi}\int_{d-i\infty}^{d+i\infty} \zeta(s+\alpha)\Gamma(s)w^{-s}\,ds, \end{aligned}$$

then yields a finite form of the estimate (41) of Li_α (as $w \to 0$, corresponding to $z \to 1^-$).

(ii) *When $z \to 1^-$ in a cone of angle less than π inside the unit disc*: In that case, we observe that the identity in (44) remains valid by analytic continuation, since the integral is still convergent (this property owes to the fast decay of $\Gamma(s)$ towards $\pm i\infty$). Then the residue calculation (45), on which the expansion of $\Lambda(w)$ is based in the real case $w > 0$, still makes sense. The extension of the asymptotic expansion of Li_α from within the unit disc is thus granted.

(iii) *When z tends to 1 vertically.* Details of the proof are given in [**147**]. What is needed is a justification of the validity of expansion (41), when z is allowed to tend to 1 from the exterior of the unit disc. The key to the analysis is a Lindelöf integral representation of the polylogarithm (Notes IV.7 and IV.8, p. 225), which provides analytic continuation. To wit:

$$\mathrm{Li}_\alpha(-z) = -\frac{1}{2i\pi}\int_{1/2-i\infty}^{1/2+i\infty} \frac{z^s}{s^\alpha}\,\frac{\pi}{\sin \pi s}\,ds.$$

The proof then proceeds with the analysis of the polylogarithm when $z = e^{i(w-\pi)}$ and $s = 1/2 + it$, the integral being estimated asymptotically as a *harmonic integral* (a continuous analogue of harmonic sums [**439**]) by means of Mellin transforms. The extension to a cone with vertex at 1, having a vertical symmetry and angle less than π, then follows by an analytic continuation argument. By unicity of asymptotic expansions (the horizontal cone of Parts (i) and (ii) and the vertical cone have a nonempty intersection), the resulting expansion must coincide with the one calculated explicitly in Part (i), above.

To conclude, regarding the general case $r \geq 0$, we may proceed along similar lines, with each $\log n$ factor introducing a derivative of the Riemann zeta function, hence a multiple pole at $s = 1$. It can then be checked that the resulting expansion coincides with what is given by formally differentiating the expansion of Li_α a number of times equal to r. (See also Note 18 below.) □

Figure 11 provides a table of expansions relative to commonly encountered polylogarithms (the function Li_2 is also known as a *dilogarithm*). Example 9 illustrates the use of polylogarithms for establishing a class of asymptotic expansions of which Stirling's formula appears as a special case. Further uses of Theorem VI.7 will appear in the next sections.

EXAMPLE 9. *Stirling's formula, polylogarithms, and superfactorials.* One has

$$\sum_{n\geq 1} \log n! = (1-z)^{-1}\,\mathrm{Li}_{0,1}(z),$$

to which singularity analysis is applicable. Theorem VI.7 then yields the singular expansion

$$\frac{1}{1-z}\,\mathrm{Li}_{0,1}(z) \sim \frac{L(z)-\gamma}{(1-z)^2} + \frac{1}{2}\,\frac{-L(z)+\gamma-1+\log 2\pi}{1-z} + \cdots,$$

from which Stirling's formula reads off:

$$\log n! \sim n\log n - n + \frac{1}{2}\log n + \log\sqrt{2\pi} + \cdots.$$

$$\begin{array}{lclcl}
\mathrm{Li}_{-1/2}(z) & = & \sum_{n\geq 1} \sqrt{n} z^n & = & \frac{\sqrt{\pi}}{2(1-z)^{3/2}} - \frac{3\sqrt{\pi}}{8(1-z)^{1/2}} + \zeta(-\frac{1}{2}) + O\left((1-z)^{1/2}\right) \\
\mathrm{Li}_0(z) & = & \sum_{n\geq 1} z^n & \equiv & \frac{1}{1-z} - 1 \\
\mathrm{Li}_{0,1}(z) & = & \sum_{n\geq 1} \log n\, z^n & = & \frac{L(z)-\gamma}{1-z} - \frac{1}{2}L(z) + \frac{\gamma-1}{2} + \log\sqrt{2\pi} + O\left((1-z)L(z)\right) \\
\mathrm{Li}_{1/2}(z) & = & \sum_{n\geq 1} \frac{z^n}{\sqrt{n}} & = & \sqrt{\frac{\pi}{1-z}} + \zeta(\frac{1}{2}) - \frac{1}{4}\sqrt{\pi}\sqrt{1-z} + O\left((1-z)^{3/2}\right) \\
\mathrm{Li}_{1/2,1}(z) & = & \sum_{n\geq 1} \frac{\log n}{\sqrt{n}} z^n & = & \sqrt{\pi}\frac{L(z)-\gamma-2\log 2}{\sqrt{1-z}} - \zeta(\frac{1}{2})\left(\frac{\gamma}{2} + \frac{\pi}{4} + \log\sqrt{8\pi}\right) + \cdots \\
\mathrm{Li}_1(z) & = & \sum_{n\geq 1} \frac{z^n}{n} & \equiv & L(z) \\
\mathrm{Li}_2(z) & = & \sum_{n\geq 1} \frac{z^n}{n^2} & = & \frac{\pi^2}{6} - (L(z)+1)(1-z) - (\frac{1}{4} + \frac{1}{2}L(z))(1-z)^2 + \cdots
\end{array}$$

FIGURE 11. Sample expansions of polylogarithms ($L(z) := \log(1-z)^{-1}$).

(Stirling's constant $\log\sqrt{2\pi}$ comes out as neatly $-\zeta'(0)$.) Similarly, define the *superfactorial* function: to be $1^1 2^2 \cdots n^n$. One has

$$\sum_{n\geq 1} \log(1^2 2^2 \cdots n^n) z^n = \frac{1}{1-z}\,\mathrm{Li}_{-1,1}(z),$$

to which singularity analysis is mechanically applicable. The analogue of Stirling's formula then reads:

$$\begin{array}{rcl}
1^1 2^2 \cdots n^n & \sim & A n^{\frac{1}{2}n^2 + \frac{1}{2}n + \frac{1}{12}} e^{-\frac{1}{4}n^2}, \\
A & = & \exp\left(\frac{1}{12} - \zeta'(-1)\right) \;=\; \exp\left(-\frac{\zeta'(2)}{2\pi^2} + \frac{\log(2\pi)+\gamma}{12}\right).
\end{array}$$

The constant A is known as the Glaisher-Kinkelin constant [**137**, p. 135]. Higher order factorials can be treated similarly. . . . . END OF EXAMPLE 9. □

▷ **18.** *Polylogarithms of integral index and a general formula.* Let $\alpha = m \in \mathbb{Z}_{\geq 1}$. Then:

$$\mathrm{Li}_m(z) = \frac{(-1)^m}{(m-1)!} w^{m-1}(\log w - H_{m-1}) + \sum_{j\geq 0, j\neq m-1} \frac{(-1)^j}{j!}\zeta(m-j) w^j,$$

where H_m is the harmonic number and $w = -\log z$. [The line of proof is the same as in Theorem VI.7, only the residue calculation at $s = 1$ differs.] The general formula,

$$\mathrm{Li}_{\alpha,r}(z) \underset{z\to 1}{\sim} (-1)^r \frac{\partial^r}{\partial\alpha^r} \sum_{s\in\mathbb{Z}_{\geq 0}\cup\{1-\alpha\}} \mathrm{Res}\left[\zeta(s+\alpha)\Gamma(s)w^{-s}\right], \qquad w := -\log z,$$

holds for all $\alpha \in \mathbb{C}$ and $r \in \mathbb{Z}_{\geq 0}$ and is amenable to symbolic manipulation. ◁

VI. 9. Functional composition

Let f and g be functions analytic at the origin that have nonnegative coefficients. We consider the composition

$$h = f \circ g, \qquad h(z) = f(g(z)),$$

assuming $g(0) = 0$. Let ρ_f, ρ_g, ρ_h be the corresponding radii of convergence, and let $\tau_f = f(\rho_f)$, and so on. We shall assume that f and g are Δ–continuable and that they admit singular expansions in the scale of powers. There are three cases to be distinguished depending on how τ_g compares to ρ_f. Clearly one has:

— *Supercritical case*[6], when $\tau_g > \rho_f$. In that case, when z increases from 0, there is a value r strictly less than ρ_g such that $g(r)$ attains the value ρ_f, which triggers a singularity of $f \circ g$. In other words $r \equiv \rho_h = g^{(-1)}(\rho_f)$. Around this point, g is analytic and a singular expansion of $f \circ g$ is obtained by composing the singular expansion of f with the regular expansion of g at r. *The singularity type is that of the external function* (f).

— *Subcritical case*, when $\tau_g < \rho_f$. In this dual situation, the singularity of $f \circ g$ is driven by that of the inside function g. We have $\rho_h = \rho_g$, $\tau_h = f(\rho_g)$ and the singular expansion of $f \circ g$ is obtained by composing the regular expansion of f with the singular expansion of g at ρ_g. *The singularity type is that of the internal function* (g).

— *Critical case*, when $\tau_g = \rho_f$. In this boundary case, there is a confluence of singularities. We have $\rho_h = \rho_g$, $\tau_h = \tau_f$, and the the singular expansion is obtained by composition rules of the singular expansions. *The singularity type is a mix of the types of the internal and external functions* (f, g).

This terminology extends the notion of supercritical sequence schema introduced in Chapter V, where we considered the case $f(z) = (1 - z)^{-1}$ and discussed some of the probabilistic consequences. Rather than stating general conditions that would be unwieldy, it is better to discuss examples directly, referring to the above guidelines supplemented by the plain algebra of generalized power expansions, whenever necessary.

EXAMPLE 10. "*Supertrees*". Let $\mathcal{G}$ be the class of general Catalan trees:

$$\mathcal{G} = \mathcal{Z} \times \textsc{Seq}(\mathcal{G}) \qquad \Longrightarrow \qquad G(z) = \frac{1}{2}(1 - \sqrt{1 - 4z}).$$

The radius of convergence of $G(z)$ is $\frac{1}{4}$ and the singular value is $G(\frac{1}{4}) = \frac{1}{2}$. The class $\mathcal{Z}\mathcal{G}$ consists of planted trees, which are such that to the root is attached a stem and an extra node, with OGF equal to $zG(z)$. We then introduce two classes of *supertrees* defined by substitution:

$$\begin{array}{lcl} \mathcal{H} = \mathcal{G}[\mathcal{Z}\mathcal{G}] & \Longrightarrow & H(z) = G(zG(z)) \\ \mathcal{K} = \mathcal{G}[(\mathcal{Z} + \mathcal{Z}')\mathcal{G}] & \Longrightarrow & K(z) = G(2zG(z)). \end{array}$$

These are "trees of trees": the class $\mathcal{H}$ is formed of trees such that, on each node there is grafted a planted tree (by the combinatorial substitution of Section I. 6, p. 76); the class $\mathcal{K}$ similarly

[6]This terminology is in accordance with the notion of supercritical sequence schema in Section V. 4 (p. 315), for which the external function is $f(z) = (1 - z)^{-1}$, with $\rho_f = 1$.

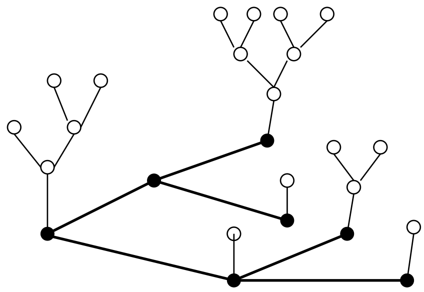


FIGURE 12. A binary supertree is a "tree of trees", with component trees all binary. The number of binary supertrees with $2n$ nodes has the unusual asymptotic form $c4^n n^{-5/4}$.

corresponds to the case when the stems can be of any two colours. Incidentally, combinatorial sum expressions are available for the coefficients,

$$H_n = \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{1}{n-k}\binom{2k-2}{k-1}\binom{2n-3k-1}{n-k-1}, \quad K_n = \sum_{k=1}^{\lfloor n/2 \rfloor} \frac{2^k}{n-k}\binom{2k-2}{k-1}\binom{2n-3k-1}{n-k-1},$$

the initial values being given by

$$H(z) = z^2 + z^3 + 3z^4 + 7z^5 + 21z^6 + \cdots, \quad K(z) = 2z^2 + 2z^3 + 8z^4 + 18z^5 + 64z^6 + \cdots.$$

Since $\rho_G = \frac{1}{4}$ and $\tau_G = \frac{1}{2}$, the composition scheme is subcritical in the case of $\mathcal{H}$ and critical in the case of $\mathcal{K}$. In the first case, the singularity is of square-root type and one finds easily:

$$H(z) \underset{z\to\frac{1}{4}}{\sim} \frac{2-\sqrt{2}}{4} - \frac{1}{\sqrt{8}}\sqrt{\frac{1}{4}-z}, \qquad \longrightarrow \qquad H_n \sim \frac{4^n}{8\sqrt{2\pi}n^{3/2}}.$$

In the second case, the two square-roots combine to produce a fourth root:

$$K(z) \underset{z\to\frac{1}{4}}{\sim} \frac{1}{2} - \frac{1}{\sqrt{2}}(\frac{1}{4}-z)^{1/4} \qquad \longrightarrow \qquad K_n \sim \frac{4^n}{8\Gamma(\frac{3}{4})n^{5/4}}.$$

On a similar register, consider the class $\mathcal{B}$ of complete binary trees:

$$\mathcal{B} = \mathcal{Z} + \mathcal{Z} \times \mathcal{B} \times \mathcal{B} \quad \Longrightarrow \quad B(z) = \frac{1-\sqrt{1-4z^2}}{2z},$$

and define the class of *binary supertrees* (Figure 12) by

$$\mathcal{S} = \mathcal{B}(\mathcal{Z} \times \mathcal{B}) \quad \Longrightarrow \quad S(z) = \frac{1-\sqrt{2\sqrt{1-4z^2}-1+4z^2}}{1-\sqrt{1-4z^2}}.$$

The composition is critical since $zB(z) = \frac{1}{2}$ at the dominant singularity $z = \frac{1}{2}$. It is enough to consider the reduced function

$$\overline{S}(z) = S(\sqrt{z}) = z + z^2 + 3z^3 + 8z^4 + 25z^5 + 80z^6 + 267z^7 + 911z^8 + \cdots,$$

whose coefficients constitute *EIS* **A101490** and occur in Bousquet-Mélou's study of integrated superbrownian excursion [**63**]. We find

$$\overline{S}(z) \sim 1-\sqrt{2}(1-4z)^{1/4}+(1-4z)^{1/2}+\cdots \longrightarrow \overline{S}_n = \frac{4^n}{n^{5/4}}\left(\frac{\sqrt{2}}{4\Gamma(\frac{3}{4})} - \frac{1}{2\sqrt{\pi}n^{1/4}} + \cdots\right).$$

Weights:

(46)

k	$\frac{1}{k}$	$\frac{1}{4^k}\binom{2k}{k}$	1	H_k	k	k^2
f_k	$\log\frac{1}{1-z}$	$\frac{1}{\sqrt{1-z}}$	$\frac{1}{1-z}$	$\frac{1}{1-z}\log\frac{1}{1-z}$	$\frac{z}{(1-z)^2}$	$\frac{z+z^2}{(1-z)^3}$.

Triangular arrays:

(47)

$g(z)$	$\frac{z}{1-z}$	ze^z	$z(1+z)$	$\frac{1-\sqrt{1-4z}}{2}$	$\frac{1-2z-\sqrt{1-4z}}{2z}$	$T(z)$
$g_n^{(k)}$	$\binom{n-1}{k-1}$	$\frac{k^{n-k}}{(n-k)!}$	$\binom{k}{n-k}$	$\frac{k}{n}\binom{2n-k-1}{n-1}$	$\frac{k}{n}\binom{2n}{n-k}$	$k\frac{n^{n-k-1}}{(n-k)!}$

FIGURE 13. Typical weights (top) and triangular arrays (bottom) illustrating the discussion of combinatorial sums $S_n = \sum_{k=1}^n f_k g_n^{(k)}$.

For instance, a seven term expansion yields a relative accuracy of 10^{-4}, already for $n = 100$, so that such approximations are quite usable in practice.

The occurrence of the exponent $-\frac{5}{4}$ in the enumeration of bicoloured and binary supertrees is striking. Related constructions have been considered by Kemp [**253**] who obtained more generally exponents of the form $-1 - 2^{-d}$ by iterating the substitution construction (this, in connection with what he called "multidimensional trees"). It is significant that asymptotic terms of the form $n^{p/q}$ with $q \neq 1, 2$ can appear in elementary combinatorics, even in the context of simple algebraic functions. Such exponents tend to be associated with nonstandard limit laws, akin to the stable distributions of probability theory; see also our discussion at the end of Chapter IX. . . . . END OF EXAMPLE 10. □

▷ **19.** *Supersupertrees.* Define these by

$$S^{[2]}(z) = B(zB(zB(z))).$$

We find automatically (with the help of B. Salvy's program)

$$[z^{2n+1}]S^{[2]}(z) \sim 2^{-13/4}\Gamma\left(\frac{7}{8}\right)^{-1} 4^n n^{-9/8},$$

and further extensions involving an asymptotic term $n^{-1-2^{-d}}$ are possible (see [**253**] for similar cases). ◁

Combinatorial sums. Singularity analysis permits us to discuss the asymptotic behaviour of entire classes of combinatorial sums at a fair level of generality, with asymptotic estimates coming out rather automatically. We consider here combinatorial sums of the form

$$S_n = \sum_{k=0}^n f_k g_n^{(k)},$$

where f_k is a sequence of numbers, usually of a simple form and called the *weights*, while the $g_n^{(k)}$ are a triangular array of numbers, for instance Pascal's triangle.

As weights f_k we shall consider sequences such that $f(z)$ is Δ–analytic with a singular expansion involving functions of the standard scale of Theorems VI.1, VI.2, VI.3. Typical examples for $f(z)$ and (f_k) are[7] are displayed in Figure 13, Equation 46.

[7]Weights like $\log k$ and $\sqrt{k}$, also satisfy these conditions, as seen in Section VI. 8.

The triangular arrays discussed here are taken to be coefficients of the *powers* of some fixed function, namely,

$$g_n^{(k)} = [z^n](g(z))^k \qquad \text{where} \qquad g(z) = \sum_{n=1}^{\infty} g_n z^n,$$

with $g(z)$ an analytic function at the origin having nonnegative coefficients and satisfying $g(0) = 0$. Examples are given in Figure 13, Equation (47). An interesting class of such arrays arises from the Lagrange inversion theorem. Indeed, if $g(z)$ is implicitly defined by $g(z) = zG(g(z))$, one has $g_{n,k} = \frac{k}{n}[w^{n-k}]G(w)^n$; the last three cases of (47) are obtained in this way (by taking $G(w)$ as $1/(1-w), (1+w)^2, e^w$).

By design, the generating function of the S_n is simply

$$S(z) = \sum_{n=0}^{\infty} S_n z^n = f(g(z)) \qquad \text{with} \qquad f(z) = \sum_{k=0}^{\infty} f_k z^k.$$

Consequently, the asymptotic analysis of S_n results by inspection from the way singularities of $f(z)$ and $g(z)$ get transformed by composition.

EXAMPLE 11. *Bernoulli sums.* Let ϕ be a function from $\mathbb{Z}_{\geq 0}$ to $\mathbb{R}$ and write $f_k := \phi(k)$. Consider the sums

$$S_n := \sum_{k=0}^{n} \phi(k) \frac{1}{2^n} \binom{n}{k}.$$

If X_n is a binomial random variable[8], $X_n \in \mathrm{Bin}(n, \frac{1}{2})$, then $S_n = \mathbb{E}(\phi(X_n))$ is exactly the expectation of $\phi(X_n)$. Then, by the binomial theorem, the OGF of the sequence (S_n) is:

$$S(z) = \frac{2}{2-z} f\left(\frac{z}{2-z}\right).$$

Considering weights whose generating function has, like in (46) radius of convergence 1, what we have is a variant of the composition schema, with an additional prefactor. The composition scheme is of the *supercritical type* since the function $g(z) = z/(2-z)$, which has radius of convergence equal to 2, satisfies $\tau_g = \infty$. The singularities of $S(z)$ are then of the same type as those of the weight generating function $f(z)$ and one verifies, in all cases of (46), that, to first asymptotic order, $S_n \sim \phi(n/2)$: this is in agreement with the fact that the binomial distribution is concentrated near its mean $\frac{n}{2}$. Singularity analysis provides complete asymptotic expansions, for instance,

$$\begin{array}{rcl} \mathbb{E}\left(\frac{1}{X_n} \mid X_n > 0\right) & = & \frac{2}{n} + \frac{2}{n^2} + \frac{6}{n^3} + O(n^{-4}) \\ \mathbb{E}\left(H_{X_n}\right) & = & \log\frac{n}{2} + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + O(n^{-3}). \end{array}$$

See [**135, 147**] for more along these lines. .... END OF EXAMPLE 11. □

EXAMPLE 12. *Generalized Knuth–Ramanujan Q-functions.* For reasons motivated by analysis of algorithms, Knuth has encountered repeatedly sums of the form

$$Q_n(\{f_k\}) = f_0 + f_1\frac{n}{n} + f_2\frac{n(n-1)}{n^2} + f_3\frac{n(n-1)(n-2)}{n^3} + \cdots.$$

[8] A binomial random variable is a sum of Bernoulli random variables: $X_n = \sum_{j=1}^{n} Y_j$ where the Y_j are independent and distributed like a Bernoulli variable Y, with $\mathbb{P}(Y=1) = p, \mathbb{P}(Y=0) = q = 1-p$.

(See, e.g., [**274**, pp. 305–307].) There (f_k) is a sequence of coefficients (usually of at most polynomial growth). For instance, the case $f_k \equiv 1$ yields the expected time till the first collision in the birthday paradox problem (Section II. 3, p. 106).

A closer examination shows that the analysis of such Q_n is reducible to singularity analysis. Writing

$$Q_n(\{f_k\}) = f_0 + \frac{n!}{n^{n-1}} \sum_{k \geq 1} f_k \frac{n^{n-k-1}}{(n-k)!}$$

reveals the closeness with the last column of (47). Indeed, setting

$$F(z) = \sum_{k \geq 1} \frac{f_k}{k} z^k,$$

one has $(n \geq 1)$

$$Q_n = f_0 + \frac{n!}{n^{n-1}}[z^n]S(z) \qquad \text{where} \quad S(z) = F(T(z)),$$

and $T(z)$ is the Cayley tree function ($T = ze^T$).

For weights $f_k = \phi(k)$ of polynomial growth, the schema is *critical*. Then, the singular expansion of S is obtained by composing the singular expansion of f with the expansion of T, namely, $T(z) \sim 1 - \sqrt{2}\sqrt{1-ez}$ as $z \to e^{-1}$. For instance, if $\phi(k) = k^r$ for some integer $r \geq 1$ then $F(z)$ has an rth order pole at $z = 1$. Then, the singularity type of $F(T(z))$ is $Z^{-r/2}$ where $Z = (1 - ez)$, which is reflected by $S_n \asymp e^n n^{r/2-1}$ (we use '$\asymp$' to represent order-of-growth information, disregarding multiplicative constants). After the final normalization, we see that $Q_n \asymp n^{(r+1)/2}$. Globally, for many weights of the form $f_k = \phi(k)$, we expect Q_n to be of the form $\sqrt{k}\phi(\sqrt{n})$, in accordance with the fact that the expectation of the first collision in the birthday problem is on average near $\sqrt{\pi n/2}$. . . . . END OF EXAMPLE 12. □

▷ **20.** *General Bernoulli sums.* Let $X_n \in \mathrm{Bin}(n;p)$ be a general binomial random variable,

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k q^{n-k}, \qquad q = 1-p.$$

Then with $f_k = \phi(k)$, one has

$$\mathbb{E}(\phi(X_n)) = [z^n]\frac{1}{1-qz} f\left(\frac{pz}{1-qz}\right),$$

so that the analysis develops as in the case $\mathrm{Bin}(n; \frac{1}{2})$. ◁

▷ **21.** *Higher moments of the birthday problem.* Take the model where there are n days in the year and let B be the random variable representing the first birthday collision. Then $\mathbb{P}_n(B > k) = k!n^{-k}\binom{n}{k}$, and

$$\mathbb{E}_n(\Phi(B)) = \Phi(1) + Q_n(\{\Delta\Phi(k)\}), \qquad \text{where} \quad \Delta\Phi(k) := \Phi(k+1) - \Phi(k).$$

For instance $\mathbb{E}_n(B) = 1 + Q_n(\langle 1, 1, \ldots \rangle)$. We thus get moments of various functionals (here stated to two asymptotic terms):

$\Phi(x)$	x	x^2+x	x^3+x^2	x^4+x^3
$E_n(\Phi(B))$	$\sqrt{\frac{\pi n}{2}} + \frac{2}{3}$	$2n+2$	$3\sqrt{\frac{\pi n^3}{2}} - 2n$	$8n^2 - 7\sqrt{\frac{\pi n^3}{2}}$

via singularity analysis. ◁

$\triangleright$ **22.** *General Bernoulli sums.* Let $X_n \in \mathrm{Bin}(n; p)$ be a binomial random variable with general parameters p, q:

$$\mathbb{P}(X_n = k) = \binom{n}{k} p^k q^{n-k}, \qquad q = 1 - p.$$

Then with $f_k = \phi(k)$, one has

$$\mathbb{E}(\phi(X_n)) = [z^n] \frac{1}{1 - qz} f\left(\frac{pz}{1 - qz}\right),$$

so that the analysis develops as in the case $\mathrm{Bin}(n; \frac{1}{2})$. $\triangleleft$

$\triangleright$ **23.** *How to weigh an urn? The "shake-and-paint" algorithm.* You are given an urn containing an unknown number N of identical looking balls. How to estimate this number in much fewer than $O(N)$ operations? A probabilistic solution due to Brassard and Bratley [**66**] uses a brush and paint. Shake the urn, pull out a ball, then mark it with paint and replace it into the urn. Repeat until you find an already painted ball. Let X be the number of operations. One has $\mathbb{E}(X) \sim \sqrt{\pi N/2}$. Further more the quantity $Y := X^2/2$ constitutes, by the previous note, an asymptotically unbiased estimator of N, in the sense that $\mathbb{E}(Y) \sim N$. In other words, count the time till an already painted ball is first found, and return half of the square of this time. One also has $\sqrt{\mathbb{V}(Y)} \sim N$. By performing the experiment m times (using m different colours of paint) and by taking the arithmetic average of the m estimates, one obtains an unbiased estimator whose typical accuracy is $\sqrt{1/m}$. For instance, $m = 16$ gives an expected accuracy of 25%. (Similar principles are used in the design of data mining algorithms.) $\triangleleft$

$\triangleright$ **24.** *Catalan sums.* These are defined by

$$S_n := \sum_{k \geq 0} f_k \binom{2n}{n-k}, \qquad S(z) = \frac{1}{\sqrt{1-4z}} f\left(\frac{1 - 2z - \sqrt{1-4z}}{2z}\right).$$

The case when $\rho_f = 1$ corresponds to a critical composition, which can be discussed much in the same way as Ramanujan sums. $\triangleleft$

Schemas. Singularity analysis also enables us to discuss at a fair level of generality the behaviour of *schemas*, in a way that parallels the discussion of the sequence schema, based on a meromorphic analysis (Section V. 4, p. 315). We illustrate this point here by means of the supercritical cycle schema. Deeper examples relative to recursively defined structures are developed in Chapter VII.

EXAMPLE 13. *Supercritical cycle schema.* The schema $\mathcal{H} = \mathrm{CYC}\, C(\mathcal{G})$ forms labelled cycles from basic components in $\mathcal{G}$:

$$\mathcal{H} = \mathrm{CYC}(\mathcal{G}) \qquad \Longrightarrow \qquad H(z) = \log \frac{1}{1 - G(z)}.$$

Consider the case where G attains the value 1 before becoming singular, that is, $\tau_G > 1$. This corresponds to a supercritical composition schema, which can be discussed in a way that closely parallels the supercritical sequence schema (Section V. 4, p. 315): a logarithmic singularity replaces a polar singularity.

Let $\sigma := \rho_H$, which is determined by $G(\sigma) = 1$. First, one finds:

$$H(z) \underset{z \to \sigma}{\sim} \log \frac{1}{1 - z/\sigma} - \log(\sigma G'(\sigma)) + A(z),$$

where $A(z)$ is analytic at $z = \sigma$. Thus:

$$[z^n]H(z) \sim \frac{\sigma^{-n}}{n}.$$

(The error term implicit in this estimate is exponentially small).

The BGF $H(z,u) = \log(1 - uG(z))^{-1}$ has the variable u marking the number of components in $\mathcal{H}$-objects. In particular, the mean number of components in a random $\mathcal{H}$-object of size is $\sim \lambda n$, where $\lambda = 1/(\sigma G'(\sigma))$, and the distribution is concentrated around its mean. Similarly, the mean number of components with size k in a random $\mathcal{H}_n$ object is found to be asymptotic to $\lambda g_k \sigma^k$, where $g_k = [z^k]G(z)$. .... END OF EXAMPLE 13. □

VI. 10. Closure properties

At this stage, we have available composition rules for singular expansions under operations like $\pm$, $\times$, $\div$. These are induced by corresponding rules for extended formal power series, where generalized exponents and logarithmic factors are allowed. Also, from Section VI. 7, inversion of analytic functions normally gives rise to square-root singularities and, from Section VI. 9, functions amenable to singularity analysis are essentially closed under composition.

In this section[9] we show that functions of singularity analysis class satisfy explicit closure properties under differentiation, integration, and Hadamard product. In order to keep the developments simple, we shall mostly restrict attention to functions that are Δ–analytic and admit a *simple* singular expansion of the form

$$f(z) = \sum_{j=0}^{J} c_j (1-z)^{\alpha_j} + O((1-z)^A), \tag{48}$$

or a *simple* singular expansion *with logarithmic terms*

$$f(z) = \sum_{j=0}^{J} c_j \left(L(z)\right) \, (1-z)^{\alpha_j} + O((1-z)^A), \qquad L(z) := \log\frac{1}{1-z}. \tag{49}$$

where each c_j is a polynomial. These are the most frequently occurring in applications, and the proof techniques are easily extended to deal with more general situation.

Subsection VI. 10.1 treats differentiation and integration; Subsection VI. 10.2 presents the closure of functions that admit simple expansions under Hadamard product. Finally, in Subsection VI. 10.3, we conclude with an examination of several interesting classes of tree recurrences, where all the closure properties previously established are put to use in order to quantify precisely the asymptotic behaviour of recurrences that are attached to tree models.

---

[9]This section contains supplementary material that may be omitted on a first reading. The contents are liberally borrowed from an article of Fill, Flajolet, and Kapur [**135**].

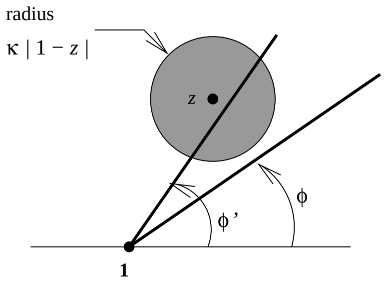


FIGURE 14. The geometry of the contour $\gamma(z)$ used in the proof of the differentiation theorem.

VI. 10.1. Differentiation and integration. Functions of singularity analysis class are closed under differentiation, this is in sharp contrast with real analysis. In the simple cases[10] of (48) and (49), closure under integration is also satisfied. The general principle (Theorems VI.8 and VI.9 below) is the following: *Derivatives and primitives of functions that are of singularity analysis class admit singular expansions that can be obtained term by term via formal differentiation and integration.*

The following statement is a version tuned to our needs of well-known differentiability properties of complex asymptotic expansions (see, e.g., Olver's book [**334**, p. 9]).

THEOREM VI.8 (Singular differentiation). *Let $f(z)$ be Δ-analytic with a singular expansion near its singularity of the simple form*

$$f(z) = \sum_{j=0}^{J} c_j (1-z)^{\alpha_j} + O((1-z)^A).$$

Then, for each integer $r > 0$, the derivative $\frac{d^r}{dz^r} f(z)$ is Δ–analytic. The expansion of the derivative at the singularity is obtained through term-by-term differentiation:

$$\frac{d^r}{dz^r} f(z) = (-1)^r \sum_{j=0}^{J} c_j \frac{\Gamma(\alpha_j+1)}{\Gamma(\alpha_j+1-r)} (1-z)^{\alpha_j - r} + O((1-z)^{A-r}).$$

PROOF. Clearly, all that is required is to establish the effect of differentiation on error terms, which is expressed symbolically as

$$\frac{d}{dz} O((1-z)^A) = O((1-z)^{A-1}).$$

By bootstrapping, only the case of a single differentiation ($r = 1$) needs to be considered.

Let $g(z)$ be a function that is regular in a domain $\Delta(\phi, \eta)$ where it is assumed to satisfy $g(z) = O((1-z)^A)$ for $z \in \Delta$. Choose a subdomain $\Delta' := \Delta(\phi', \eta')$, where $\phi < \phi' < \frac{\pi}{2}$ and $0 < \eta' < \eta$. By elementary geometry, for a sufficiently small $\kappa > 0$, the disc of radius $\kappa(z-1)$ centered at a value $z \in \Delta'$ lies entirely in Δ; see Figure 14.

[10] It would be possible but unwieldy to treat a larger class, which would have to include arbitrarily nested logarithms, since, for instance, $\int dx/x = \log x$, $\int dx/(x \log x) = \log \log x$, and so on.

We fix such a small value κ and let $\gamma(z)$ represent the boundary of that disc oriented positively.

The starting point is Cauchy's integral formula

$$g'(z) = \frac{1}{2\pi i}\int_C g(w)\frac{dw}{(w-z)^2}, \tag{50}$$

a direct consequence of the residue theorem. Here C should encircle z while lying inside the domain of regularity of g, and we opt for the choice $C \equiv \gamma(z)$. Then trivial bounds applied to (50) give

$$\begin{aligned} |g'(z)| &= O\left(\|\gamma(z)\|\cdot(1-z)^A|1-z|^{-2}\right) \\ &= O\left(|1-z|^{A-1}\right). \end{aligned}$$

The estimate involves the length of the contour, $\|\gamma(z)\|$, which is $O(1-z)$ by construction, as well as the bound on g itself, which is $O((1-z)^A)$ since all points of the contour are themselves at a distance exactly of the order of $|1-z|$ from 1. ☐

▷ **25.** *Differentiation and logarithms.* Let $g(z)$ satisfy

$$g(z) = O\left((1-z)^A L(z)^k\right), \qquad L(z) = \log\frac{1}{1-z},$$

for $k \in \mathbb{Z}_{\geq 0}$. Then, one has

$$\frac{d^r}{dz^r}g(z) = O\left((1-z)^{A-r}L(z)^k\right).$$

(The proof follows along the lines of Theorem VI.8.) ◁

It is well known that integration of asymptotic expansions is usually easier than differentiation. Here is a statement custom-tailored to our needs.

THEOREM VI.9 (Singular integration). *Let $f(z)$ be Δ-analytic and admit a Δ-expansion near its singularity of the form*

$$f(z) = \sum_{j=0}^{J} c_j(1-z)^{\alpha_j} + O((1-z)^A).$$

Then $\int_0^z f(t)\,dt$ is Δ–analytic. Assume tfurther hat none of the quantities α_j and A equals -1.

(i) If $A < -1$, then the singular expansion of $\int f$ is

$$\int_0^z f(t)\,dt = -\sum_{j=0}^{J}\frac{c_j}{\alpha_j+1}(1-z)^{\alpha_j+1} + O\left((1-z)^{A+1}\right). \tag{51}$$

(ii) If $A > -1$, then the singular expansion of $\int f$ is

$$\int_0^z f(t)\,dt = -\sum_{j=0}^{J}\frac{c_j}{\alpha_j+1}(1-z)^{\alpha_j+1} + L_0 + O\left((1-z)^{A+1}\right), \tag{52}$$

where the "integration constant" L_0 has the value

$$L_0 := \sum_{\alpha_j<-1}\frac{c_j}{\alpha_j+1} + \int_0^1\left[f(t) - \sum_{\alpha_j<-1}c_j(1-t)^{\alpha_j}\right]dt.$$

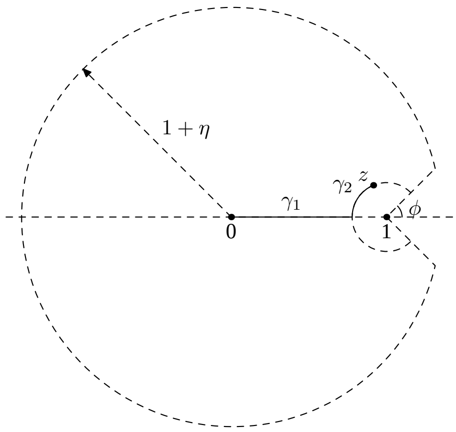


FIGURE 15. The contour used in the proof of the integration theorem.

The case where either some α_j or A is -1 is easily treated by the additional rules

$$\int_0^z (1-t)^{-1}\,dt = L(z), \qquad \int_0^z O((1-t)^{-1})\,dt = O(L(z)).$$

that are consistent with elementary integration, and similar rules are easily derived for powers of logarithms. Furthermore, the corresponding O–transfers hold true. (The proofs are simple modifications of the one given below for the basic case.)

PROOF. The basic technique consists in integrating, term by term, the singular expansion of f. We let $r(z)$ be the remainder term in the expansion of f, that is,

$$r(z) := f(z) - \sum_{j=0}^{J} c_j (1-z)^{\alpha_j}.$$

By assumption, throughout the Δ-domain one has, for some positive constant K,

$$|r(z)| \le K|1-z|^A.$$

(i) *Case* $A < -1$. Straight-line integration between 0 and z, provides (51), as soon as it has been established that

$$\int_0^z r(t)\,dt = O\left(|1-z|^{A+1}\right).$$

By Cauchy's integral formula, we can choose any path of integration that stays within the region of analyticity of r. We choose the contour $\gamma := \gamma_1 \cup \gamma_2$, shown in Figure 15.

Then, one has

$$\begin{aligned}\left|\int_\gamma r(t)\,dt\right| &\le \left|\int_{\gamma_1} r(t)\,dt\right| + \left|\int_{\gamma_2} r(t)\,dt\right| \\ &\le K\int_{\gamma_1} |1-t|^A\,|dt| + K\int_{\gamma_2} |1-t|^A|\,|dt| \\ &= O(|1-z|^{A+1}).\end{aligned}$$

where the symbol $|dt|$ designates the differential line-length element in the corresponding curvilinear integral. Both integrals are $O(|1-z|^{A+1})$: for the integral along γ_1, this results from explicitly carrying out the integration; for the integral along γ_2, this results from the trivial bound $O(\|\gamma_2\|(1-z)^A)$.

(ii) *Case* $A > -1$. We let $f_-(z)$ represent the "divergence part" of f that gives rise to nonintegrability:

$$f_-(z) := \sum_{\alpha_j<-1} c_j(1-z)^{\alpha_j}.$$

Then with the decomposition $f = [f - f_-] + f_-$, integrations can be performed separately. First, one finds

$$\int_0^z f_-(t)\,dt = -\sum_{\alpha_j<-1}\frac{c_j}{\alpha_j+1}(1-z)^{\alpha_j+1} + \sum_{\alpha_j<-1}\frac{c_j}{\alpha_j+1}.$$

Next, observe that the asymptotic condition guarantees the existence of $\int_0^1$ applied to $[f - f_-]$, so that

$$\int_0^z [f(t)-f_-(t)]\;dt = \int_0^1 [f(t)-f_-(t)]\;dt + \int_1^z [f(t)-f_-(t)]\;dt.$$

The first of these two integrals is a constant that contributes to L_0. As to the second integral, term-by-term integration yields

$$\int_1^z [f(t)-f_-(t)]\;dt = -\sum_{\alpha_j>-1}\frac{c_j}{\alpha_j+1}(1-z)^{\alpha_j+1} + \int_1^z r(t)\,dt.$$

The remainder integral is finite, given the growth condition on the remainder term, and, upon carrying out the integration along the rectilinear segment joining 1 to z, trivial bounds show that it is indeed $O(|1-z|^{A+1})$. □

VI. 10.2. Hadamard Products. The *Hadamard product* of two functions $f(z), g(z)$ analytic at the origin is defined as their term-by-term product,

(53)
$$f(z)\odot g(z) = \sum_{n\ge 0} f_n g_n z^n, \qquad \text{where} \quad f(z) = \sum_{n\ge 0} f_n z^n, \quad g(z) = \sum_{n\ge 0} g_n z^n.$$

We are going to see, following an article of Fill, Flajolet, and Kapur [**135**], that simple functions of singularity analysis class are closed under Hadamard product. Establishing such a closure property requires methods for composing functions from the basic scale, namely $(1-z)^a$, as well as error terms of the form $O((1-z)^A)$. We address each problem in turn.

The expansion around the origin,

$$(1-z)^a = 1 + \frac{-a}{1}z + \frac{(-a)(-a+1)}{2!}z^2 + \cdots, \tag{54}$$

gives through term-by-term multiplication

$$(1-z)^a \odot (1-z)^b = {}_2F_1[-a,-b;1;z]. \tag{55}$$

Here ${}_2F_1$ represents the classical *hypergeometric function* of Gauss defined by

$${}_2F_1[\alpha,\beta;\gamma;z] = 1 + \frac{\alpha\beta}{\gamma}\frac{z}{1!} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{\gamma(\gamma+1)}\frac{z^2}{2!} + \cdots. \tag{56}$$

From their transformation theory, see for instance [**433**, Ch XIV], hypergeometric functions can generally be expanded in the vicinity of $z=1$ by means of the $z \mapsto 1-z$ transformation. Instantiation of this transformation with $\gamma = 1$ yields

$$\begin{aligned} {}_2F_1[\alpha,\beta;1;z] &= \frac{\Gamma(1-\alpha-\beta)}{\Gamma(1-\alpha)\Gamma(1-\beta)}\,{}_2F_1[\alpha,\beta;\alpha+\beta;1-z] \\ &\quad + \frac{\Gamma(\alpha+\beta-1)}{\Gamma(\alpha)\Gamma(\beta)}(1-z)^{-\alpha-\beta+1}\,{}_2F_1[1-\alpha,1-\beta;2-\alpha-\beta;1-z]. \end{aligned} \tag{57}$$

From there, we can state:

THEOREM VI.10 (Hadamard Composition). *When neither of a, b, $a+b$ is an integer, the Hadamard product $(1-z)^a \odot (1-z)^b$ has an infinite Δ-expansion with exponent scale $\{0,1,2,\ldots\} \cup \{a+b+1, a+b+2, \ldots\}$, namely,*

$$(1-z)^a \odot (1-z)^b \sim \sum_{k\ge 0} \lambda_k^{(a,b)} \frac{(1-z)^k}{k!} + \sum_{k\ge 0} \mu_k^{(a,b)} \frac{(1-z)^{a+b+1+k}}{k!},$$

where the coefficients λ and μ are given by

$$\lambda_k^{(a,b)} = \frac{\Gamma(1+a+b)}{\Gamma(1+a)\Gamma(1+b)}\frac{(-a)^{\overline{k}}(-b)^{\overline{k}}}{(-a-b)^{\overline{k}}}, \quad \mu_k^{(a,b)} = \frac{\Gamma(-a-b-1)}{\Gamma(-a)\Gamma(-b)}\frac{(1+a)^{\overline{k}}(1+b)^{\overline{k}}}{(2+a+b)^{\overline{k}}}.$$

Here $x^{\overline{k}}$ is defined for $k \in \mathbb{Z}_{\ge 0}$ by $x^{\overline{k}} := x(x+1)\cdots(x+k-1)$.

$\triangleright$ **26.** *Special cases.* The case where either a or b is an integer poses no difficulty, since, for $m \in \mathbb{Z}_{\ge 0}$, the function $(1-z)^m \odot g(z)$ is a polynomial, while, $(1-z)^{-m} \odot g(z)$ is reducible to a derivative of g, to which the Singular Differentiation Theorem can be applied.

The case $a+b \in \mathbb{Z}$ needs transformation formulæ that extend (57): the principles (based on a Lindelöf integral representation and developed by Barnes) are described in [**433**, §14.53], while the formulæ appear explicitly in [**2**, pp. 559–560]. $\triangleleft$

$\triangleright$ **27.** *Simple expansions with logarithmic terms.* The technique of differentiation with respect to a parameter,

$$\left[(1-z)^{-\alpha}L(z)\right] \odot (1-z)^{-\beta} = \frac{\partial}{\partial\alpha}\left[(1-z)^{-\alpha} \odot (1-z)^{-\beta}\right],$$

makes it possible to derive explicit composition rules for expansions involving logarithmic terms. $\triangleleft$

Next, we address the Hadamard composition of error terms in singular expansions. The way Hadamard products preserve Δ–analyticity and compose error terms is summarized by the following statement.

THEOREM VI.11 (Hadamard closure). (i) *Assume that* $f(z)$ *and* $g(z)$ *are* Δ*-analytic in* $\Delta(\psi_0, \eta)$*. Then the Hadamard product* $(f \odot g)(z)$ *is analytic in a (possibly smaller)* Δ*-domain, call it* Δ'.

(ii) *Assume further that*

$$f(z) = O((1-z)^a) \text{ and } g(z) = O((1-z)^b), \quad z \in \Delta(\psi_0, \eta).$$

Then the Hadamard product $(f \odot g)(z)$ *admits in* Δ' *an expansion given by the following rules:*

— If $a + b + 1 < 0$*, then*

$$(f \odot g)(z) = O((1-z)^{a+b+1}).$$

— If $k < a + b + 1 < k + 1$*, for some integer* $k \in \mathbb{Z}_{\geq -1}$*, then*

$$(f \odot g)(z) = \sum_{j=0}^{k} \frac{(-1)^j}{j!} (f \odot g)^{(j)} (1)(1-z)^j + O\left((1-z)^{a+b+1}\right).$$

— If $a + b + 1$ *is a nonnegative integer, then (with* $L(z) = \log(1-z)^{-1}$*)*

$$(f \odot g)(z) = \sum_{j=0}^{k} \frac{(-1)^j}{j!} (f \odot g)^{(j)} (1)(1-z)^j + O\left((1-z)^{a+b+1} L(z)\right).$$

PROOF. (Sketch) The starting point is an important formula due to Hadamard that expresses Hadamard products as a contour integral:

$$f(z) \odot g(z) = \frac{1}{2i\pi} \int_\gamma f(w) g\left(\frac{z}{w}\right) \frac{dw}{w}. \tag{58}$$

The contour γ in the w-plane should be chosen such that both factors, $f(w)$ and $g(z/w)$ are analytic. In other words, given the domain Δ in which both f and g are analytic, one should have $\gamma \subset \Delta \cap (z\Delta^{-1})$.

In the first case $(a + b + 1 < 0)$, the precise geometry of a feasible contour γ is described in [**135**], the principles being similar to those employed in the construction of Hankel contours elsewhere in this chapter. The integral giving the value of the Hadamard product is finally estimated trivially, based on the order of growth assumptions on f and g, as $z \to 1$. This approach extends to the case $a + b + 1 = 0$, where a logarithmic factor comes in,

For the remaining cases, the easy identity

$$\vartheta^{c+d}(f \odot g) = (\vartheta^c f) \odot \left(\vartheta^d g\right), \qquad \text{where} \quad \vartheta \equiv z\frac{d}{dz},$$

reduces the analysis to the situation where $a + b + 1 < 0$. It suffices to differentiate sufficiently many times and finally integrate back, as permitted by the Singular Integration Theorem. $\square$

Globally, Theorems VI.10 and VI.11 establish the closure under Hadamard products of functions amenable to singularity analysis in the sense of (48). In practice, Let $f(z)$ and $g(z)$ be Δ–analytic and admit simple singular expansions of the form (48) or (49). What is sought is the singular expansion of

$$h(z) := f(z) \odot g(z).$$

Step 1. Determine the asymptotic expansions $f_n = [z^n]f(z)$ and $g(z) = [z^n]g(z)$ induced by the singular expansions of f and g in accordance with the singularity analysis process. Given finite singular expansions of f and g, the order C of the error in the expansion of h is known *a priori* by Theorem VI.11.

Step 2. Deduce from Step 1 an asymptotic expansion of $h_n = [z^n]h(z)$ by usual multiplication from the expansions of f_n and g_n.

Step 3. Reconstruct by singularity analysis a function $H(z)$ that is singular at 1 and is such that

$$[z^n]H(z) \sim [z^n]h(z).$$

This can be done by using the expansions of basic functions, as provided by Theorems VI.1 and VI.2 in the reverse direction. By construction, $H(z)$ is a sum of functions of the form $(1-z)^\alpha L(z)^k$, which are all singular at 1.

Step 4. Output the singular expansion of $f \odot g$ as

$$h(z) = H(z) + P(z) + O\left((1-z)^C\right),$$

where P is a polynomial of degree δ, which is the largest integer $< C$. The polynomial $P(z)$ is needed, since polynomials (and more generally functions analytic at 1) do not leave a trace in asymptotic expansions of coefficients. Since $h(z) - H(z)$ is δ times differentiable at 1, one must take

$$P(z) = \sum_{j=0}^{\delta} \frac{(-1)^j}{j!} \partial_z^j \left(h(z) - H(z)\right)_{z=1} (1-z)^j.$$

FIGURE 16. The Zigzag algorithm for computing singular expansions of Hadamard products.

in order to derive the singular expansion of a function at a singularity, one may conveniently appeal to the *Zigzag Algorithm* described in Figure 16, whose validity is ensured by the *a priori* knowledge of the *existence* of an expansion guaranteed by Theorems VI.10 and VI.11. A typical application of this algorithms appears in Equations (61) and (62) below, in the context of Pólya's drunkard problem.

EXAMPLE 14. *Pólya's drunkard problem.* (This example is taken from [**135**].) In the d–dimensional lattice $\mathbb{Z}^d$ of points with integer coordinates, the drunkard performs a random walk starting from the origin with steps in $\{-1,+1\}^d$, each taken with equal likelihood. The probability that the drunkard is back at the origin after $2n$ steps is

$$q_n^{(d)} = \left(\frac{1}{2^{2n}}\binom{2n}{n}\right)^d, \tag{59}$$

since the walk is a product d independent 1–dimensional walks. The probability that $2n$ is the epoch of the *first* return to the origin is the quantity $p_n^{(d)}$, which is determined implicitly by

$$\left(1 - \sum_{n=1}^{\infty} p_n^{(d)} z^n\right)^{-1} = \sum_{n=0}^{\infty} q_n^{(d)} z^n, \tag{60}$$

as results from the convolution equations expressing the decomposition of loops into primitive loops. In terms of the associated ordinary generating functions P and Q, this relation thus reads as $(1-P(z))^{-1} = Q(z)$.

The asymptotic analysis of the q_n's is straightforward; the one of the p_n's is more involved and is of interest in connection with recurrence and transience of the random walk; see, e.g., [**109, 288**]. The Hadamard closure theorem provides a direct access to this problem. Define

$$\lambda(z) := \sum_{n\geq 0} \frac{1}{2^{2n}}\binom{2n}{n} z^n \equiv \frac{1}{\sqrt{1-z}}.$$

Then, Equations (59) and (60) imply:

$$P(z) = 1 - \frac{1}{\lambda(z)^{\odot d}}, \qquad \text{where} \quad \lambda(z)^{\odot d} := \lambda(z)\odot\cdots\odot\lambda(z)\ (d\ \text{times}).$$

The singularities of $P(z)$ are found to be as follows.

Case $d=1$: No Hadamard product is involved and

$$P(z) = 1-\sqrt{1-z}, \qquad \text{implying} \quad p_n^{(1)} = \frac{1}{n2^{2n-1}}\binom{2n-2}{n-1} \sim \frac{1}{2\sqrt{\pi n^3}}.$$

(This agrees with the classical combinatorial solution expressed in terms of Catalan numbers.)

Case $d=2$: By the Hadamard closure theorem, the function $Q(z) = \lambda(z)\odot\lambda(z)$ admits a priori a singular expansion at $z=1$ that is composed solely of elements of the form $(1-z)^\alpha$ possibly multiplied by integral powers of the logarithmic function $L(z)$. From a computational standpoint (cf. the Zigzag Algorithm), it is then best to start from the coefficients themselves,

$$q_n^{(2)} \sim \left(\frac{1}{\sqrt{\pi n}} - \frac{1}{8\sqrt{\pi n^3}} + \cdots\right)^2 \sim \frac{1}{\pi}\left(\frac{1}{n} - \frac{1}{4n^2} + \cdots\right), \tag{61}$$

and reconstruct the only singular expansion that is compatible, namely

$$Q(z) = \frac{1}{\pi}L(z) + K + O((1-z)^{1-\epsilon}), \tag{62}$$

where $\epsilon > 0$ is an arbitrarily small constant and K is fully determined as the limit as $z\to 1$ of $Q(z) - \pi^{-1}L(z)$. Then it can be seen that the function P is Δ–continuable. (Proof: Otherwise, there would be complex poles arising from zeros of the function Q on the unit disc, and this would entail in $p_n^{(2)}$ the presence of terms oscillating around 0, a fact that contradicts the necessary positivity of probabilities.) The singular expansion of $P(z)$ at $z=1$ results immediately from that of $Q(z)$:

$$P(z) \sim 1 - \frac{\pi}{L(z)} + \frac{\pi^2 K}{L^2(z)} + \cdots.$$

so that, by Theorems VI.2 and VI.3, one has

$$\begin{aligned} p_n^{(2)} &= \frac{\pi}{n\log^2 n} - 2\pi\frac{\gamma+\pi K}{n\log^3 n} + O\left(\frac{1}{n\log^4 n}\right) \\ K &= 1 + \sum_{n=1}^{\infty}\left(16^{-n}\binom{2n}{n}^2 - \frac{1}{\pi n}\right) \\ &\doteq 0.88254240061060637358582 57\,. \end{aligned}$$

(See the study by Louchard *et al.* [**302**, Sec. 4] for somewhat similar calculations.)

<u>*Case* $d = 3$</u>: This case is easy since $Q(z)$ remains finite at its singularity $z = 1$ where it admits an expansion in powers of $(1-z)^{1/2}$, to the effect that

$$q_n^{(3)} \sim \left(\frac{1}{\sqrt{\pi n}} - \frac{1}{8\sqrt{\pi n^3}} + \cdots\right)^3 \sim \frac{1}{\pi^{3/2}}\left(\frac{1}{n^{3/2}} - \frac{3}{8n^{5/2}} + \cdots\right).$$

The function $Q(z)$ is a priori Δ-continuable and its singular expansion can be reconstructed from the form of coefficients:

$$Q(z) \underset{z\to 1}{\sim} Q(1) - \frac{2}{\pi}\sqrt{1-z} + O(|1-z|),$$

leading to

$$P(z) = \left(1 - \frac{1}{Q(1)}\right) - \frac{2}{\pi Q^2(1)}\sqrt{1-z} + O(|1-z|).$$

By singularity analysis, the last expansion gives

$$\begin{aligned} p_n^{(3)} &= \frac{1}{\pi^{3/2}Q^2(1)}\,\frac{1}{n^{3/2}} + O\left(\frac{1}{n^2}\right) \\ Q(1) &= \frac{\pi}{\Gamma\left(\frac{3}{4}\right)^4} \doteq 1.3932039296856768591842463. \end{aligned}$$

A complete asymptotic expansion in powers $n^{-3/2}, n^{-5/2}, \ldots$ can be obtained by the same devices. In particular this improves the error term above to $O(n^{-5/2})$. The explicit form of $Q(1)$ results from its expression as the generalized hypergeometric ${}_3F_2[\frac{1}{2}, \frac{1}{2}, \frac{1}{2}; 1, 1; 1]$, which evaluates by Clausen's theorem and Kummer's identity to the square of a complete elliptic integral. (See the papers by Larry Glasser for context, for instance [**200**]; nowadays, several computer algebra systems even provide this value automatically.)

Higher dimensions are treated similarly, with logarithmic terms surfacing in asymptotic expansions for all even dimensions. . . . . END OF EXAMPLE 14. □

VI. 10.3. Tree recurrences. To conclude with singularity analysis theory, we present the general framework of *tree recurrences*, also known as *probabilistic divide-and-conquer recurrences*, which are recurrences of the general form

$$f_n = t_n + \sum_k p_{n,k}(f_k + f_{n-a-k}), \qquad (n \geq n_0). \tag{63}$$

There, (f_n) is the sequence implicitly determined by the recurrence, assuming known initial conditions $f_0, \ldots, f_{n_0-1}$; the sequence (t_n) is known as the sequence of *tolls*; the array $(p_{n,k})$ is a triangular array of numbers that are probabilities in the sense that, for each fixed $n \geq 0$, one has $\sum_k p_{n,k} = 1$; the number a is a small fixed integer (usually 0 or 1).

The interpretation of the recurrence is in the form of a splitting process : a collection of n elements is given; a number a of these is put aside and the rest is partitioned into two subgroups, a "left" subgroup of cardinality K_n and a "right" subgroup of cardinality $n - 1 - K_n$. The quantity K_n is a random variable with probability distribution

$$\mathbb{P}(K_n = k) = p_{n,k}.$$

The splitting is repeated recursively till only groups of size less than the threshold n_0 are obtained. Assuming stochastic independence of all the random variables K involved, it is seen that f_n represents the expectation of the (total) *cost* C_n of a random (recursive) splitting, when a single stage involving n elements incurs a toll equal to t_n. In symbols:

$$f_n = \mathbb{E}(C_n), \qquad C_n = t_n + C_{K_n} + C_{n-a-K_n}.$$

Clearly, a particular realization of the splitting process can be represented by a binary tree. With a suitable choice of probabilities, such processes can be used to analyse cost functional of increasing binary trees, and binary Catalan trees, for instance. A prime motivation is the analysis of divide-and-conquer algorithms in computer science, like quicksort, mergesort, union-find algorithms, and so on [**84, 175, 381**]. Our treatment once more follows the article [**135**].

A general approach to the asymptotic solution of a tree recurrence goes as follows. First, introduce generating functions,

$$f(z) = \sum_n f_n \omega_n z^n, \qquad t(z) = \sum_n t_n \omega'_n z^n,$$

for some normalization sequences (ω_n) and (ω'_n) that are problem-specific. (So, $\omega_n \equiv 1$ gives rise to an OGF, $\omega_n \equiv n!$ to an EGF, with other normalizations being also useful.) Then, by linearity of the original recurrence, there exists a linear operator $\mathfrak{L}$ on series (and functions), such that

$$f(z) = \mathfrak{L}[t(z)].$$

Provided the splitting probabilities $p_{n,k}$ have expressions of a tractable form, it is reasonable to attempt expressing $\mathfrak{L}$ in terms of the usual operations of analysis. One may then investigate the way $\mathfrak{L}$ affects singularities and deduce the asymptotic form of the cost sequence (f_n) from the singularities of its generating function, $f(z)$. An interesting feature of this approach is to allow for a powerful discussion of the relationship between tolls and induced costs, in a way that parallels composition of singularities in Section VI. 9. Closure properties discussed earlier in this section are naturally a crucial ingredient in the intervening singularity analysis process.

The three examples that we present combine closure properties with the singularity analysis of polylogarithms of Section VI. 8. Example 15 is relative to increasing binary trees and binary search trees (Example II.17, p. 133). Example 16 discusses additive costs of random binary Catalan trees in the perspective of tree recurrences. Finally, Example 17 shows the applicability of singularity analysis to a basic coalescence-fragmentation process.

EXAMPLE 15. *The binary search tree recurrence.* One of the simplest random tree models is defined as follows: a random binary tree of size $n \geq 1$ is obtained by taking a root and appending to it a left subtree of size K_n and a right subtree of size $n-1-K_n$, where K_n is uniformly distributed over the set of permissible values $\{0, 1, \ldots, n-1\}$. A tree of size 0 is the empty tree. In earlier notations, this process corresponds to

$$p_{n,k} \equiv \mathbb{P}(K_n = k) = \frac{1}{n}, \qquad 0 \leq k \leq n-1.$$

The associated tree recurrence is then

$$f_n = t_n + \frac{2}{n} \sum_{k=0}^{n-1} f_k, \qquad f_0 = t_0,$$

which translates for OGFs,

$$f(z) := \sum_{n\geq 0} f_n z^n, \qquad t(z) = \sum_{n\geq 0} t_n z^n,$$

into a linear integral equation:

$$f(z) = t(z) + 2\int_0^z f(w)\,\frac{dw}{1-w}. \tag{64}$$

Differentiation yields the ordinary differential equation

$$f'(z) = t'(z) + \frac{2}{1-z} f(z), \qquad f(0) = t_0,$$

which is then solved by the variation-of-constants method. In this way, it is found that an integral transform expresses the relation between the GF of tolls and the GF of total costs. Assuming without loss of generality $t_0 = 0$, we have (with $\partial_w \equiv \frac{d}{dw}$)

$$f(z) = \mathfrak{L}[t(z)], \qquad \text{where} \quad \mathfrak{L}[t(z)] = \frac{1}{(1-z)^2}\int_0^z (\partial_w t(w))\,(1-w)^2\,dw. \tag{65}$$

Simple toll sequences that admit generating functions of a simple form can then be employed to build a *repertoire*[11] that already provides useful indications on the relations between the orders of growth of (t_n) and (f_n). For instance, we find, for the rising-factorial tolls

$$\begin{cases} t_n^{\overline{\alpha}} := \binom{n+\alpha}{\alpha}, & t^{\overline{\alpha}}(z) = (1-z)^{-\alpha-1}, \\ f^{\overline{\alpha}}(z) = \frac{\alpha-1}{\alpha+1}\left[(1-z)^{-\alpha-1} - (1-z)^{-2}\right], & f_n^{\overline{\alpha}} = \frac{\alpha-1}{\alpha+1}\left[\binom{n+\alpha}{\alpha} - n - 1\right], \end{cases}$$

for $\alpha \neq 1$, while $\alpha = 1$ corresponding to $t_n^{\overline{1}} = n+1$ leads to

$$f^{\overline{1}}(z) = \frac{2}{(1-z)^2}\log\frac{1}{1-z}, \qquad f_n^{\overline{1}} = 2(n+1)(H_{n+1}-1) = 2n\log n + O(n),$$

with H_n a harmonic number. The emergence of an extra logarithmic factor for $\alpha = 1$ is to be noted: it corresponds to the fact that path length in a binary search tree or an increasing binary tree of size n is $\sim 2n\log n$. These elementary techniques provide a first set of entries recapitulated in Figure 17.

Singularity analysis furthermore permits us to develop a complete asymptotic expansions for tolls of the form $\sqrt{n}$, $\log n$, and many others. Consider for instance the toll $t_n^{\alpha} = n^{\alpha}$, for which the generating function, a polylogarithm, is known to admit a singular expansions in terms of elements of the form $(1-z)^{\beta}$, with the main term corresponding to $\beta = -\alpha - 1$ when $\alpha > -1$ (Theorem VI.7). The $\mathfrak{L}$ transformation reads as a succession of operations, "differentiate, multiply by $(1-z)^2$, integrate, multiply by $(1-z)^{-2}$", which are covered by Theorems VI.8 and VI.9. The chain on any particular element starts as

$$c(1-z)^{\beta} \quad \xrightarrow{\partial} \quad c\beta(1-z)^{\beta-1} \quad \xrightarrow{\times(1-z)^2} \quad c\beta(1-z)^{\beta+1},$$

at which stage integration intervenes. According to Theorem VI.9, assuming $\beta \neq -2$ and ignoring integration constants, integration gives

$$c\beta(1-z)^{\beta+1} \quad \xrightarrow{\int} \quad -c\frac{\beta}{\beta+2}(1-z)^{\beta+2} \quad \xrightarrow{\times(1-z)^{-2}} \quad -c\frac{\beta}{\beta+2}(1-z)^{\beta}.$$

[11]The repertoire approach is developed in an attractive manner by Green and Knuth in [**214**].

Tolls (t_n)		*Costs* (f_n)
$t_n = n^\alpha$	$(2 < \alpha)$	$f_n = \dfrac{\alpha+1}{\alpha-1}n^\alpha + O(n^{\alpha-1})$
$t_n = n^\alpha$	$(1 < \alpha < 2)$	$f_n = \dfrac{\alpha+1}{\alpha-1}n^\alpha + O(n)$
$t_n = \dbinom{n+\alpha}{\alpha}$	$(\alpha > 1)$	$\dfrac{\alpha-1}{\alpha+1}\left[\dbinom{n+\alpha}{\alpha} - n + 1\right] \sim \dfrac{\alpha+1}{\alpha-1}\dfrac{n^\alpha}{\Gamma(\alpha+1)}$
$t_n = \dbinom{n+\alpha}{\alpha}$	$(\alpha < 1)$	$\dfrac{1-\alpha-1}{1+\alpha}\left[n+1-\dbinom{n+\alpha}{\alpha}\right] \sim \dfrac{1+\alpha}{1-\alpha}n$
$t_n = n^\alpha$	$(0 < \alpha < 1)$	$K_\alpha n + O(n^\alpha)$
$t_n = \log n$		$K_0' n - \log n + O(1)$

FIGURE 17. Tolls and costs for the binary search tree recurrence, assuming $t_0 = 0$.

Thus, the singular element $(1-z)^\beta$ corresponds to a contribution

$$-c\frac{\beta}{\beta+2}\binom{n-\beta-1}{-\beta-1},$$

which is of order $O(n^{-\beta-1})$. It can be verified that this chain of operations suffices to determine the leading order of f_n when $t_n = n^\alpha$ and $\alpha > 1$.

The derivation above is representative of the main lines of the analysis, but it has left aside the determination of integration constants, which play a dominant rôle when $t_n = n^\alpha$ and $\alpha < 1$ (because a term of the form $K/(1-z)^2$ then dominates in $f(z)$). Introduce, in accordance with the statement of the Singular Integration Theorem (Theorem VI.9) the quantity

$$\mathbf{K}[t] := \int_0^1 \left[t'(w)(1-w)^2 - \left(t'(w)(1-w)^2\right)_-\right] dw,$$

where f_- represents the sum of singular terms of exponent < -1 in the singular expansion of $f(z)$. Then, for $t_n = n^\alpha$ with $0 < \alpha < 1$, taking into account the integration constant(which gets multiplied by $(1-z)^{-2}$, given the shape of $\mathfrak{L}$), we find for $\alpha < 1$:

$$f_n \sim K_\alpha n, \qquad K_\alpha = \mathbf{K}[\mathrm{Li}_{-\alpha}] = 2\sum_{n=1}^{\infty}\frac{n^\alpha}{(n+1)(n+2)}.$$

Similarly, the toll $t_n = \log n$ gives rise to

$$f_n \sim K_0' n, \qquad K_0' = 2\sum_{n=1}^{\infty}\frac{\log n}{(n+1)(n+2)} \doteq 1.2035649167.$$

This last estimate quantifies the *entropy* of the distribution of binary search trees, which is studied by Fill in [**136**], and discussed in the reference book by Cover and Thomas on information theory [**85**, p. 74-76]. . . . . END OF EXAMPLE 15. □

EXAMPLE 16. *The binary tree recurrence.* Consider a procedure that on a binary tree performs some calculation (without affecting the tree itself) at a cost of t_n, then recursively calls itself on the left and right subtrees. If the binary tree to which the procedure is applied is drawn

Tolls (t_n)		*Costs* (f_n)
n^α	$(\frac{3}{2} < \alpha)$	$\dfrac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)} n^{\alpha+1/2} + O(n^{\alpha-1/2})$
$n^{3/2}$		$\dfrac{2}{\sqrt{\pi}} n^2 + O(n\log n)$
n^α	$(\frac{1}{2} < \alpha < \frac{3}{2})$	$\dfrac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)} n^{\alpha+1/2} + O(n)$
$n^{1/2}$		$\dfrac{1}{\sqrt{\pi}} n\log n + O(n)$
n^α	$(0 < \alpha < \frac{1}{2})$	$\overline{K}_\alpha n + O(1)$
$\log n$		$\overline{K}'_0 n + O(\sqrt{n}$

FIGURE 18. Tolls and costs for the binary tree recurrence.

uniformly amongst all binary trees of size n the expectation of the total costs of the procedure satisfies the recurrence

$$(66)\qquad f_n = t_n + \sum_{k=0}^{n-1} \frac{C_k C_{n-1-k}}{C_n}(f_k + f_{n-k}) \qquad \text{with} \quad C_n = \frac{1}{n+1}\binom{2n}{n}.$$

Indeed, the quantity

$$p_{n,k} = \frac{C_k C_{n-1-k}}{C_n}$$

represents the probability that a random tree of size n has a left subtree of size k and a right subtree of size $n-k$. It is then natural to introduce the generating functions

$$t(z) = \sum_{n\geq 0} t_n z C_n z^n, \qquad f(z) = \sum_{n\geq 0} f_n z C_n z^n,$$

and the recurrence (66) translates into a linear equation:

$$f(z) = t(z) + 2C(z)f(z).$$

Now, given a toll sequence (t_n) with *ordinary* generation function

$$\tau(z) := \sum_{n\geq 0} t_n z^n,$$

the function $t(z)$ is a Hadamard product: $t(z) = \tau(z) \odot C(z)$. Also, $C(z)$ is well known, so that the fundamental relation is

$$(67)\qquad f(z) = \mathfrak{L}[\tau(z)], \qquad \text{where} \quad \mathfrak{L}[\tau(z)] = \frac{\tau(z)\odot C(z)}{\sqrt{1-4z}}, \quad C(z) = \frac{1-\sqrt{1-4z}}{2z}.$$

This transform relates the ordinary generating function of tolls to the normalized generating function of the total costs via a Hadamard product.

The calculation for simple tolls like n^r with $r \in \mathbb{Z}_{\geq 0}$ can be carried out elementarily. For the tolls $t_n^\alpha = n^\alpha$ what is required is the singular expansion of

$$\tau(z) \odot C\left(\frac{z}{4}\right) = \mathrm{Li}_{-\alpha}(z) \odot C\left(\frac{z}{4}\right) = \sum_{n=1}^{\infty} \frac{n^\alpha}{n+1}\binom{2n}{n}\left(\frac{z}{4}\right)^n.$$

This is precisely covered by Theorems VI.10 and VI.11. The results of Figure 18 follow, after routine calculations. .... END OF EXAMPLE 16. □

EXAMPLE 17. *The Cayley tree recurrence.* Consider n vertices labelled $1, \dots, n$. There are $(n-1)!n^{n-2}$ sequences of edges,

$$\langle u_1, v_1, \rangle, \langle u_2, v_2, \rangle, \cdots \langle u_{n-1}, v_{n-1}, \rangle,$$

that give rise to a tree over $\{1, \dots, n\}$, and the number of such sequences is $(n-1)!n^{n-2}$ since there are n^{n-2} unrooted trees of size n. At each stage k, the edges numbered 1 to k determine a forest. Each addition of an edge connects two trees and reduces the number of trees in the forest by 1, so that the forest evolves from the totally disconnected graph (at time 0) to an unrooted tree (at time $n-1$). If we consider each of the sequences to be equally likely, the probability that u_{n-1} and v_{n-1} belong to components of size k and $(n-k)$ is

$$\frac{1}{2(n-1)}\binom{n}{k}\frac{k^{k-1}n^{n-k-1}}{n^{n-2}}.$$

(The reason is that there are n^{n-2} unrooted trees; the last added edge has $n-1$ possibilities and 2 possible orientations.)

Assume that the aggregation of two trees into a tree of size equal to ℓ incurs a toll of t_ℓ. The total cost of the aggregation process for a final tree of size n satisfies the recurrence

$$(68) \qquad f_n = t_n + \sum_{0<k<n} p_{n,k}(f_k + f_{n-k}), \qquad p_{n,k} = \frac{1}{2(n-1)}\binom{n}{k}\frac{k^{k-1}n^{n-k-1}}{n^{n-2}}.$$

The recurrence (68) has been studied in detail by Knuth and Pittel [**273**], building upon earlier works of Knuth and Schönhage [**274**]. A prime motivation of the cited works is the importance of this recurrence in algorithms that dynamically manage equivalence relations (the so-called union-find algorithm [**274**]).

Given the sequence of tolls (t_n), we introduce the generating function

$$\tau(z) = \sum_{n\geq 1} t_n z^n,$$

and let T be the Cayley tree function ($T = ze^T$). For total costs, the generating function adopted is

$$f(z) = \sum_{n\geq 1} f_n n^{n-1} z^n.$$

The basic recurrence (68) can then be rephrased as an integral transform involving a Hadamard product, namely,

$$(69) \quad f(z) = \mathfrak{L}[\tau(z)], \qquad \text{with} \quad \mathfrak{L}[\tau](z) = \frac{1}{2}\frac{T(z)}{1-T(z)}\int_0^z \partial_w \left(\tau(w) \odot T(w)^2\right) \frac{dw}{T(w)}.$$

Though the expression of the transform looks formidable at first sight, it is really nothing but a short sequence of basic operations, "Hadamard product, multiplication, differentiation, division, integration, multiplication", each of which has a quantifiable effect on functions of singularity analysis class. (The singularity structure of $T(z)$ is itself determined by the Singular Inversion Theorem, Theorem VI.6.)

The net result is that the effect of tolls of the form n^α, $\log n$, and so on, can be analysed: see Figure 19 for a listting of estimates. Details of the proof are left as an exercise to our reader and are otherwise found in [**135**, §5.3]. The analogy of behaviour with the Catalan tree recurrence stands out.

This example is also of interest since it furnishes an analytically tractable model of a coalition-fragmentation process. This is a topic of great interest in several areas of science, for which we refer to Aldous' survey [**8**]. .... END OF EXAMPLE 17. □

Tolls (t_n)		*Costs* (f_n)
n^α	$(\frac{3}{2} < \alpha)$	$\dfrac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{2}\Gamma(\alpha)} n^{\alpha+1/2} + O(n^{\alpha-1/2})$
$n^{3/2}$		$\sqrt{\dfrac{2}{\pi}} n^2 + O(n \log n)$
n^α	$(\frac{1}{2} < \alpha < \frac{3}{2})$	$\dfrac{\Gamma(\alpha-\frac{1}{2})}{\sqrt{2}\Gamma(\alpha)} n^{\alpha+1/2} + O(n)$
$n^{1/2}$		$\dfrac{1}{\sqrt{2\pi}} n \log n + O(n)$
n^α	$(0 < \alpha < \frac{1}{2})$	$K_\alpha n + O(1)$
$\log n$		$\widehat{K}'_0 n + O(\sqrt{n})$

FIGURE 19. Tolls and costs for the Cayley tree recurrence.

VI. 11. Tauberian theory and Darboux's method

There are several alternative approaches to the analysis of coefficients of functions, which are of moderate growth. Naturally, *All methods provide estimates compatible with singularity analysis methods* (Theorems VI.1, VI.2, and VI.3). Each one requires some sort of "*regularity condition*" either on the part of the function or on the part of the coefficient sequence, the regularity condition of singularity analysis being in essence analytic continuation.

The methods briefly surveyed here fall into three broad categories: (i) Elementary real analytic methods; (ii) Tauberian theorems; (iii) Darboux's method.

Elementary real analytic methods assume some *a priori* smoothness conditions on the coefficient sequence; they are included here for the sake of completeness, though properly speaking they do not belong to the galaxy of complex asymptotic methods. Their scope is mostly limited to the analysis of products while the other methods permit to approach more general functional composition patterns. Tauberian theorems belong to the category of advanced real analysis methods; they also needs some *a priori* regularity on the coefficients, typically positivity or monotonicity. Darboux's method requires some smoothness of the function on the closed unit disk, and, by its techniques and scope, it is the closest to singularity analysis.

We content ourselves with a brief discussion of the main results. For more information, the reader is referred to Odlyzko's excellent survey [**330**].

Elementary real analytic methods. An asymptotic equivalent of the coefficients of a function can sometimes be worked out elementarily from simple properties of the component functions. The regularity conditions are a smooth asymptotic behaviour of the coefficients of one of the two factors in a product of generating functions. A good source for these techniques is Bender's survey [**28**].

THEOREM VI.12 (Bender's method). *Let* $a(z) = \sum a_n z^n$ *and* $b(z) = \sum b_n z^n$ *be two power series with radii of convergence* $\alpha > \beta \geq 0$ *respectively. Assume that* $b(z)$ *satisfies the ratio test,*

$$\frac{b_{n-1}}{b_n} \to \beta \qquad \textit{as} \qquad n \to \infty.$$

Then the coefficients of the product $f(z) = a(z) \cdot b(z)$ *satisfy, provided* $a(\beta) \neq 0$,

$$[z^n]f(z) \sim a(\beta) b_n \qquad \textit{as} \qquad n \to \infty.$$

PROOF. (Sketch) The basis of the proof is the following chain:

$$\begin{array}{rcl} f_n & = & a_0 b_n + a_1 b_{n-1} + a_2 b_{n-2} + \cdots + a_n b_0) \\ & = & b_n \left(a_0 + a_1 \dfrac{b_{n-1}}{b_n} + a_2 \dfrac{b_{n-2}}{b_n} + \cdots + a_n \dfrac{b_0}{b_n} \right) \\ & = & b_n \left(a_0 + a_1 (\dfrac{b_{n-1}}{b_n}) + a_2 (\dfrac{b_{n-2}}{b_{n-1}})(\dfrac{b_{n-1}}{b_n}) + \cdots \right) \\ & \sim & b_n (a_0 + a_1 \beta + a_2 \beta^2 + \cdots). \end{array}$$

There, only the last line requires a little elementary analysis that is left as an exercise to the reader. □

This theorem applies for instance to the EGF of 2–regular graphs:

$$f(z) = a(z) \cdot b(z) \qquad \text{with} \qquad a(z) = e^{-z/2 - z^2/4}, \quad b(z) = \frac{1}{\sqrt{1-z}},$$

fow which it gives $f_n \sim e^{-3/4}\binom{n-1/2}{n} \sim \frac{e^{-3/4}}{\sqrt{\pi n}}$, in accordance with Example 2 (p. 378). Clearly, a whole collection of lemmas can be stated in the same vein. Singularity analysis usually provides more complete expansions, though Theorem VI.12 does apply to a few situations not covered by it.

Tauberian theory. Tauberian methods apply to functions whose growth is only known along the positive real line. The regularity conditions are in the form of additional assumptions on the coefficients (positivity or monotonicity) known under the name of Tauberian "side conditions". An insightful introduction to the subject may be found in Titchmarsh's book [**411**], and a detailed exposition in Postnikov's monograph [**351**] and Korevaar's compendium [**279**]. We cite the most famous of all Tauberian theorems due to Hardy, Littlewood, and Karamata. In this section, a function is said to be *slowly varying* at infinity iff, for any $c > 0$, one has $L(cx)/L(x) \to 1$ as $x \to +\infty$; examples of slowly varying functions are provided by powers of logarithms or iterated logarithms.

THEOREM VI.13 (The HLK Tauberian theorem). *Let* $f(z)$ *be a power series with radius of convergence equal to 1, satisfying*

$$f(z) \sim \frac{1}{(1-z)^\alpha} L(\frac{1}{1-z}), \tag{70}$$

for some $\alpha \geq 0$ *with* L *a slowly varying function. Assume that the coefficients* $f_n = [z^n]f(z)$ *are all non–negative (this is the "side condition"). Then*

$$\sum_{k=0}^{n} f_k \sim \frac{n^\alpha}{\Gamma(\alpha+1)} L(n). \tag{71}$$

The conclusion (71) is consistent with what singularity analysis gives: Under the conditions, and if in addition analytic continuation is assumed, then

$$f_n \sim \frac{n^{\alpha-1}}{\Gamma(\alpha)} L(n), \tag{72}$$

which by summation yields the estimate (71).

It must be noted that a Tauberian theorem requires very little on the part of the function. However, it also gives less since the result it provides is valid in the more restrictive sense of mean values, or Cesàro averages. (However, if further regularity conditions on the f_n are injected, for instance monotonicity, then the conclusion of (72) can be deduced from (71) by purely elementary real analysis.) The method applies only to functions that are large enough at their singularity, and despite numerous efforts to improve the conclusions, it is the case that Tauberian theorems have little concrete to offer in terms of error estimates.

Appeal to a Tauberian theorem is justified when a function has, apart from the positive half line, a very irregular behaviour near its circle of convergence, for instance when each point of the unit circle is a singularity. (The function is then said to admit the unit circle as a natural boundary.) An interesting example of this situation is discussed by Greene and Knuth [**213**] who consider the function

$$f(z) = \prod_{k=1}^{\infty}(1 + \frac{z^k}{k}) \tag{73}$$

that is the EGF of permutations having cycles all of different lengths. A little computation shows that

$$\begin{aligned} \log \prod_{k=1}^{\infty}\left(1 + \frac{z^k}{k}\right) &= \sum_{k=1}^{\infty}\frac{z^k}{k} - \frac{1}{2}\sum_{k=1}^{\infty}\frac{z^{2k}}{k^2} + \frac{1}{3}\sum_{k=1}^{\infty}\frac{z^{3k}}{k^3} - \cdots \\ &\sim \log\frac{1}{1-z} - \gamma + o(1). \end{aligned}$$

(Only the last line requires some care, see [**213**].)

Thus, we have

$$f(z) \sim \frac{e^{-\gamma}}{1-z} \qquad \Longrightarrow \qquad \frac{1}{n}(f_0 + f_1 + \cdots + f_n) \sim e^{-\gamma},$$

by virtue of Theorem VI.12. In fact, Greene and Knuth were able to supplement this argument by a "bootstrapping" technique and show a stronger result, namely

$$f_n \to e^{-\gamma}.$$

Darboux's method. The method of Darboux requires, as regularity condition, that functions be smooth enough —i.e., sufficiently differentiable— on their circle of convergence. What lies at the heart of this many–facetted method is a simple relation between the smoothness of a function and the corresponding decrease of its Taylor coefficients.

THEOREM VI.14 (Darboux's method). *Assume that $f(z)$ is continuous in the closed disk $|z| \le 1$, and is in addition k times continuously differentiable $(k \ge 0)$ on $|z| = 1$. Then*

$$[z^n]f(z) = o\left(\frac{1}{n^k}\right). \tag{74}$$

PROOF. Start from Cauchy's coefficient formula

$$f_n = \frac{1}{2i\pi}\int_{\mathcal{C}} f(z)\,\frac{dz}{z^{n+1}}.$$

Because of the continuity assumption, one may take as integration contour $\mathcal{C}$ the unit circle. Setting $z = e^{i\theta}$ yields the Fourier version of Cauchy's coefficient formula,

$$f_n = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\theta})e^{-ni\theta}\,d\theta. \tag{75}$$

The integrand in (75) is strongly oscillating and the Riemann–Lebesgue lemma of classical analysis (see [**411**, p. 403]) shows that the integral giving f_n tends to 0 as $n \to \infty$.

This argument covers the case $k = 0$. The case of a general k is then derived through successive integrations by parts, as

$$[z^n]f(z) = \frac{1}{2\pi(in)^k}\int_0^{2\pi} f^{(k)}(e^{i\theta})e^{-ni\theta}\,d\theta.$$

□

Various consequences of Theorem VI.14 are given in reference texts also under the name of Darboux's method. See for instance [**82, 213, 229, 437**]. We shall only illustrate the mechanism by rederiving in this framework the analysis of the EGF of 2–regular graphs. Clearly, we have

$$f(z) = \frac{e^{-z/2-z^2/4}}{\sqrt{1-z}} = \frac{e^{-3/4}}{\sqrt{1-z}} + e^{-3/4}\sqrt{1-z} + R(z). \tag{76}$$

There $R(z)$ is the product of $(1-z)^{3/2}$ with a function analytic at $z = 1$ that is a rest in the Taylor expansion of $e^{-z/2-z^2/4}$. Thus, $R(z)$ is of class $\mathbf{C}^1$, *i.e.*, continuously differentiable once. By Theorem VI.14, we have

$$[z^n]R(z) = o\left(\frac{1}{n}\right),$$

so that

$$[z^n]f(z) = \frac{e^{-3/4}}{\sqrt{\pi n}} + o\left(\frac{1}{n}\right). \tag{77}$$

Darboux's method bears some resemblance to singularity analysis in that the estimates derive from translating error terms in expansions. However, smoothness conditions, rather than plain order of growth information, are required by it. The method is often applied in situations like in (76)–(77) to functions that are products of the type $h(z)(1-z)^\alpha$ with $h(z)$ analytic at 1, or combinations thereof. In such particular cases, Darboux's method is however subsumed by singularity analysis.

It is inherent to Darboux's method that it cannot be applied to functions whose singular expansion only involves terms that become infinite, while singularity analysis can. A clear example arises in the analysis of the common subexpression problem [**176**] where there occurs a function with a singular expansion of the form

$$\frac{1}{\sqrt{1-z}}\frac{1}{\sqrt{\log\frac{1}{1-z}}}\left[1+\frac{c_1}{\log\frac{1}{1-z}}+\cdots\right].$$

$\triangleright$ **28.** *Darboux versus singularity analysis.* This note provides an instance where Darboux's method applies whereas singularity analysis does not. Let

$$F_r(z)=\sum_{n=0}^{\infty}\frac{z^{2^n}}{(2^n)^r}.$$

The function $F_0(z)$ is singular at every point of the unit circle, and the same property holds for any F_r with $r\in\mathbb{Z}_{\geq 0}$. [Hint: F_0, which satisfies the functional equation $F(z)=z+F(z^2)$, grows unboundedly near 2^nth roots of unity.] Darboux's method can be used to derive

$$[z^n]\frac{1}{\sqrt{1-z}}F_5(z)=\frac{c}{\sqrt{\pi n}}+o\left(\frac{1}{n}\right),\qquad c:=\frac{32}{31}.$$

What is the best error term that can be obtained? $\triangleleft$

VI. 12. Perspective

The method of singularity analysis expands our ability to extract coefficient asymptotics, to a far wider class of functions than the meromorphic and rational functions of Chapters IV and V. This ability is the fundamental tool for analysing many of the generating functions of Chapters I–III, and is applicable at a considerable level of generality.

The basic method is straightforward and appealing: we locate singularities, establish analyticity in a domain around them, expand the functions around the singularities, and apply general transfer theorems to take each term in the function expansion to a term in an asymptotic expansion of its coefficients. The method applies directly to a large variety of explicitly given functions, for instance combinations of rational functions, square roots, and logarithms. as well as to functions that are implicitly defined, like generating functions for tree structures, which are obtained by analytic inversion. Functions amenable to singularity analysis also enjoy rich closure properties, and the corresponding operations mirror the natural operations on generating functions implied by the combinatorial constructions of Chapters I–III.

This approach again sets us in the direction of the ideal situation of having a theory where combinatorial constructions and analytic methods fully correspond, but, again, the very essence of analytic combinatorics is that the theorems that provide asymptotic results cannot be so general as to be free of analytic side conditions. In the case of singularity analysis, these side conditions have to do with establishing analyticity in a domain around singularities. These conditions are automatically satisfied by a large number of functions with moderate (at most polynomial) growth near their dominant singularities (most notably a large subset of the generating functions of combinatorial structures defined by the constructions of Chapters I–III) justifying precisely what we need: a term-by-term transfer from the expansion of a generating function at its singularity to function coefficients, including error terms. The calculations involved in singularity analysis are rather mechanical. Salvy [**373**] has indeed succeeded in automating the analysis of a large class of generating functions in this way.

Again, we can look carefully at specific combinatorial constructions and then apply singularity analysis to general abstract schemas, thereby solving whole classes of combinatorial problems at once. This process (along with several important examples) is the topic of the next chapter. After that, we consider the saddle point method (Chapter VIII), which is appropriate for functions with no singularities at a finite distance (entire functions) as well as those whose growth is rapid (exponential) near their singularities. Singularity analysis will surface again in Chapter IX, given its crucial technical rôle in obtaining uniform expansions of multivariate generating functions near singularities.

General surveys of asymptotic methods in enumeration have been given by Bender [**28**] and more recently Odlyzko [**330**]. A general reference to asymptotic analysis that has a remarkably concrete approach is De Bruijn's book [**93**]. Comtet's [**82**] and Wilf's [**437**] books each devote a chapter to these questions.

This chapter is largely based on the theory developed by Flajolet and Odlyzko in [**167**], where the term "singularity analysis" originates from. That theory itself draws its inspiration from classical analytic number theory, for instance the prime number theorem where similar contours are used (see the discussion in [**167**] for sources). Another area where Hankel contours are used is the inversion theory of integral transforms [**107**], in particular in the case of algebraic and logarithmic singularities. Closure properties developed here are from the articles [**135, 147**] by Flajolet, Fill, and Kapur.

Darboux's method can often be employed as an alternative to singularity analysis. It is still the most widely used technique in the literature, though the direct mapping of asymptotic scales afforded by singularity analysis appears to us much more transparent. Darboux's method is well explained in the books by Comtet [**82**], Henrici [**229**], Olver [**334**], and Wilf [**437**]. Tauberian theory is treated in detail in Postnikov's monograph [**351**] and Korevaar's encyclopedic treatment [**279**], with an excellent introduction to be found in Titchmarsh's book [**411**].

VII

Applications of Singularity Analysis

Mathematics is being lazy. Mathematics is letting the principles do the work for you so that you do not have to do the work for yourself.

— GEORGE PÓLYA[1]

I wish to God these calculations had been executed by steam.

— CHARLES BABBAGE (1792-1871)

Contents

Singularity analysis paves the way to the analysis of a large variety of generating functions. In accordance with Pólya's aphorism, it enables us to "be lazy" and "let the principles work for you". In this chapter we illustrate this situation with singularity analysis developed in Chapter VI being put to use in order to analyse whole classes of generating functions lavishly provided by the symbolic methods of Chapters I–III.

The exp–log schema (Section VII. 1) is a general schema of analytic combinatorics that covers the set construction, either labelled or unlabelled, applied to generators whose singularity is of logarithmic type. This schema parallels in generality the supercritical schema of Chapter V. It applies to permutations, derangements, 2–regular graphs, mappings, and functional graphs. It is even the case that properties relative to the factorization of polynomials with coefficients over a finite field can be attached to it. In particular, one can obtain in a transparent manner a prime number theorem for such polynomials as well as several other characteristics of random polynomials.

The next sections deal with recursively defined structures. In that case, generating functions are accessible by means of an equation or a system that implicitly defines them. A distinctive feature of many such combinatorial generating functions is yet another type of *universality*: square-root singularities are universal, a fact that translates into universality of the exponent $-\frac{3}{2}$ in corresponding asymptotic estimates of coefficients.

Trees are the prototypical recursively defined combinatorial type. For simple varieties, equations merely involve properties of inverses of analytic functions. This applies to simple varieties of trees determined by degree constraints. Universality

[1] Quoted in M Walter, T O'Brien, Memories of George Pólya, Mathematics Teaching 116 (1986)

of square-root singularities entails that several quantities assume the same behaviour across quite different looking tree varieties: the subexponential growth factor of tree counts has a universal $-\frac{3}{2}$ exponent, nodes tend to be found at level about $\sqrt{n}$ and height is of that $\sqrt{n}$ order with high probability, path length grows like $n\sqrt{n}$, and so on. Such results hold for classical tree types (e.g., Catalan, unary-binary, Cayley) and the methods extend to many tree-like classes of combinatorics including functional graphs, mappings, and hierarchies (Section VII. 2). Essentially the same methods apply to functions defined by a single implicit equation: in that case failure cases of the Implicit Function Theorem replace failure of invertiblity in analytic functions, but square-root singularity is still universal. Consequences are found in the general enumeration of nonplane unlabelled, secondary structures of molecular biology, nonplane unlabelled rees, as well as isomers of alkanes in theoretical chemistry (Section VII. 3).

A number of generating functions of combinatorics are algebraic functions (Section VII. 4, meaning that they satisfy either one polynomial equation or are components of a polynomial system. At this level of generality, a whole family of singular behaviours is possible, though singular expansions can only involve fractional exponents. Singularity analysis is invariably applicable. The investigation of algebraic functions requires viewing them as plane algebraic curves and making use of the famous Newton-Puiseux theorem of elementary algebraic geometry, which strongly constrains the allowable types of singularities. For functions given by positive polynomial systems, the general results specialize and one encounters once more square-root singularities (Subsection VII. 4.2).

Algebraic functions manifest themselves first and foremost when dealing with context-free specifications and languages. In that case, under a technical condition of irreducibility, the theory of positive polynomial systems applies. As an example we discuss geometric configurations in the plane satisfying a non-crossing constraint. Algebraic functions may also surface as solutions of various types of functional equations: this is in particular the case for many types of walks generalizing Dyck and Motzkin paths (via the kernel method) and for many types of random maps (via the quadratic method). In all these cases, singular exponents of various forms are bound to occur.

VII. 1. The “exp–log” schema

In this section, we examine a schema that is of a level of generality comparable to the supercritical sequence schema of Chapter V but whose “physics” is rather different. This schema extends what is encountered when constructing of permutations ($\mathcal{P}$) as labelled sets of cycles ($\mathcal{K}$):

$$P(z) = \exp\left(K(z)\right), \qquad K(z) = \log\frac{1}{1-z}.$$

The distinctive feature here is the fact that a logarithmic singularity for the $\mathcal{K}$–components gets composed with an exponential, to the effect that the generating function of the composed $\mathcal{P}$–objects exhibits a singularity of polynomial growth (here $P(z) = (1-z)^{-1}$). In the case of permutations, everything is explicit and we have seen in Chapter V that the distribution of the number of components is concentrated around its

$\mathcal{F}$	κ	$n = 100$	$n = 272$	$n = 739$
Permutations	1	5.18737	6.18485	7.18319
Derangements	1	4.19732	5.18852	6.18454
2–regular	$\frac{1}{2}$	2.53439	3.03466	3.53440
Mappings	$\frac{1}{2}$	2.97898	3.46320	3.95312

FIGURE 1. Some exp–log structures ($\mathcal{F}$) and the mean number of $\mathcal{G}$–components for $n = 100, 272 \equiv \lceil 100 \cdot e \rfloor, 739 \equiv \lceil 100 \cdot e^2 \rfloor$.

mean $H_n = \log n + \gamma + o(1)$. We know from Chapter IV that with a positive probability a large $\mathcal{P}$-structure contains no $\mathcal{K}$–component of size 1; see also the discussion of derangements in Chapter II.

Very similar properties hold true under very general conditions. We start with the definition of a logarithmic function:

DEFINITION VII.1. *A function $G(z)$ with radius of convergence satisfying $0 < \rho < \infty$ is said to be of* logarithmic type *if the following conditions hold:*

(i) *the number ρ is the unique singularity of $G(z)$ on $|z| = \rho$;*
(ii) *the function $G(z)$ is continuable to a Δ–domain;*
(iii) *as $z \to \rho$ in Δ, function $G(z)$ satisfies*

$$G(z) = \kappa \log \frac{1}{1 - z/\rho} + \lambda + O\left(\frac{1}{(\log(1 - z/\rho))^2}\right),$$

for some $\kappa > 0$ and $\lambda \in \mathbb{R}$.

An exp–log schema is then defined as follows:

DEFINITION VII.2. *Let $\mathcal{F} = \mathfrak{P}(\mathcal{G})$ be a labelled set construction. The schema is said to be of the* exp–log type *if the egf $G(z)$ is of logarithmic type.*

Let $\mathcal{F} = \mathfrak{K}(\mathcal{G})$ be an unlabelled set construction: $\mathfrak{K} = \mathfrak{M}$ (multiset) or $\mathfrak{P}$ (powerset). The schema is said to be of the exp–log type *if the OGF $G(z)$ is of logarithmic type and its radius of convergence satisfies $0 < \rho < 1$.*

As we shall see below, beyond permutations, this schema covers mappings, unlabelled functional graphs, polynomials over finite fields, 2–regular graphs, as well as generalized derangements. Singularity analysis gives precise information on the decomposition of large $\mathcal{F}$ objects into $\mathcal{G}$ components.

THEOREM VII.1 (Exp–log schema). *Given a schema $\mathcal{F} = \mathfrak{P}(\mathcal{G})$ of the exp–log type, one has*

$$\begin{aligned}
[z^n]F(z) &= \frac{e^{\lambda + r_0}}{\Gamma(\kappa)} n^{\kappa - 1} \rho^{-n} \left(1 + O\left((\log n)^{-2}\right)\right), \\
[z^n]G(z) &= \frac{\kappa}{n} \rho^{-n} \left(1 + O\left((\log n)^{-2}\right)\right),
\end{aligned}$$

where $r_0 = 0$ in the labelled case and r_0 is given by (1) *in the case of unlabelled multisets.*

Let X be the number of $\mathcal{G}$–components in a random $\mathcal{F}$–object. Then:

$$\begin{aligned}
\mathbb{E}_{\mathcal{F}_n}(X) &= \kappa(\log n - \psi(\kappa)) + \lambda + r_1 + O\left((\log n)^{-2}\right) \quad (\psi(s) \equiv \tfrac{d}{ds}\Gamma(s)) \\
\mathbb{V}_{\mathcal{F}_N}(X) &= (\kappa + r_2) \log n,
\end{aligned}$$

where $r_1 = r_2 = 0$ *in the labelled case and* $r_1 > 0$ *is given by* (2) *in the case of unlabelled multisets. In particular the distribution of* X *in a large* $\mathcal{F}_n$ *object is concentrated around its mean.*

This result is from an article by Flajolet and Soria [**177**] (with a correction to the logarithmic type condition given by Jennie Hansen [**221**]). We shall show in a later chapter that, in addition, the asymptotic distribution of X is invariably Gaussian under such exp-log conditions.

PROOF. We first discuss the labelled case where $F(z) = \exp G(z)$. The estimate for $[z^n]G(z)$ follows directly from singularity analysis with the transfer of error terms of type $\log^{-2}$. For $F(z)$, it is based on

$$F(z) \sim \frac{e^{\lambda}}{(1 - z/\rho)^{\kappa}},$$

(with a $\log^{-2}$ relative error term), to which singularity analysis applies.

The BGF of $\mathcal{F}$ with u marking the number of $\mathcal{G}$–components is $F(z, u) = \exp(uG(z))$, so that the function

$$f_1(z) := \frac{\partial}{\partial u} F(z, u)\bigg|_{u=1} = F(z)G(z),$$

which is an EGF of cumulated values, satisfies near ρ

$$f_1(z) \sim \frac{e^{\lambda}}{(1 - z/\rho)^{\kappa}} \left(\kappa \log \frac{1}{1 - z/\rho} + \lambda\right)$$

to the effect that

$$[z^n]f_1(z) \equiv ([z^n]F(z)) \cdot \mathbb{E}_{\mathcal{F}_n}(X) = \frac{e^{\lambda}}{\Gamma(\kappa)} \rho^{-n} \left(\kappa \log n - \kappa\psi(\kappa) + \lambda + O\left((\log n)^{-2}\right)\right).$$

The variance analysis is conducted in the same way, but using a second derivative.

For the unlabelled case, let us consider the multiset construction. The analysis of $[z^n]G(z)$ obeys the same principles as in the labelled case as it only depends on the logarithmic assumption. The usual translation of multisets can be put under the form

$$F(z) = \exp\left(G(z) + R(z)\right), \qquad R(z) := \sum_{j=2}^{\infty} \frac{G(z^j)}{j},$$

and $R(z)$ involves terms $R(z^2), \ldots$ that are each analytic in $|z| < \rho^{1/2}$. Thus, $R(z)$ is itself analytic (as a uniformly convergent sum of analytic functions) in $|z| < \rho^{1/2}$, which properly contains the disc $|z| < \rho$ since by assumption $\rho < 1$. One thus has Δ-analyticity of F and, as $z \to \rho$,

$$(1) \qquad F(z) \sim \frac{e^{\lambda + r_0}}{(1 - z/\rho)^{\kappa}}, \qquad r_0 := \sum_{j=2}^{\infty} \frac{G(\rho^j)}{j}.$$

The asymptotic expansion of $[z^n]G(z)$ follows.

Regarding the BGF $F(z,u)$ of of $\mathcal{F}$ with u marking the number of $\mathcal{G}$–components, one has

$$F(z,u) = \exp\left(\frac{uG(z)}{1} + \frac{u^2G(z^2)}{2} + \cdots\right).$$

Consequently,

$$f_1(z) := \frac{\partial}{\partial u}F(z,u)\bigg|_{u=1} = F(z)\left(G(z) + R_1(z)\right), \qquad R_1(z) = \sum_{j=2}^{\infty} G(z^j).$$

Again, the singularity type is that of $F(z)$ multiplied by a logarithmic term:

$$(2) \qquad f_1(z) \sim F(z)(G(z) + r_1), \qquad r_1 := \sum_{j=2}^{\infty} G(\rho^j)$$

and the statement for the mean value of X follows.

The variance analysis proceeds along similar lines. For instance, in the labelled case, one has

$$([z^n]F(z))\mathbb{E}_n(X^2) = [z^n](f_1(z)+f_2(z)), \qquad f_2(z) = \frac{\partial^2}{\partial u^2}F(z,u)\bigg|_{u=1} = F(z)G(z)^2,$$

which involves at its singularity a log-squared term. □

EXAMPLE 1. *Direct instances of the exp–log schema.* The case of *permutations* corresponds to $\kappa = 1$, $\lambda = 0$, and it is easily seen to be in agreement with the statement of Theorem VII.1. Let Ω be a finite set of the integers and consider next permutations *without* any cycle of length in Ω. This includes *derangements* ($\Omega = \{1\}$) and their generalizations. Then,

$$G(z) = \log\frac{1}{1-z} - \sum_{\omega\in\Omega}\frac{z^\omega}{\omega}.$$

The theorem applies with $\kappa = 1$ but with now $\lambda := -\sum_{\omega\in\Omega}\omega^{-1}$.

The class of *2–regular graphs* is obtained by the set construction applied to undirected cycles of length ≥ 3. In this case

$$F(z) = \exp(G(z)), \qquad G(z) = \frac{1}{2}\log\frac{1}{1-z} - \frac{z}{2} - \frac{z^2}{4}.$$

This is an exp–log scheme with $\kappa = \frac{1}{2}$ and $\lambda = -\frac{3}{4}$. In particular the mean number of cycles is asymptotic to $\frac{1}{2}\log n$.

In Chapter V, we have encountered the class $\mathcal{F}$ of *mappings* (functions from a finite set to itself) as labelled sets of connected components ($\mathcal{K}$), themselves (directed) cycles of trees ($\mathcal{T}$). The class of all mappings has an EGF given by

$$F(z) = \exp(K(z)), \qquad K(z) = \log\frac{1}{1-T(z)}, \qquad T(z) = ze^{T(z)},$$

with T the Cayley tree function. The analysis of the previous chapter shows that $T(z)$ is singular at $z = e^{-1}$ where it admits the singular expansion $T(z) \sim 1 - \sqrt{2}\sqrt{1-ez}$. Thus $G(z)$ is logarithmic with $\kappa = \frac{1}{2}$ and $\lambda = \log\sqrt{2}$. In particular, the number of connected mappings satisfies

$$K_n \equiv n![z^n]K(z) = n^n\sqrt{\frac{\pi}{2n}}\left(1 + O(n^{-1/2})\right).$$

$$(X+1)\left(X^{10}+X^9+X^8+X^6+X^4+X^3+1\right)\left(X^{14}+X^{11}+X^{10}+X^3+1\right)$$
$$X^3(X+1)\left(X^2+X+1\right)^2\left(X^{17}+X^{16}+X^{15}+X^{11}+X^9+X^6+X^2+X+1\right)$$
$$X^5(X+1)\left(X^5+X^3+X^2+X+1\right)\left(X^{12}+X^8+X^7+X^6+X^5+X^3+X^2+X+1\right)\left(X^2+X+1\right)$$
$$X^2\left(X^2+X+1\right)^2\left(X^3+X^2+1\right)\left(X^8+X^7+X^6+X^4+X^2+X+1\right)\left(X^8+X^7+X^5+X^4+1\right)$$
$$\left(X^7+X^6+X^5+X^3+X^2+X+1\right)\left(X^{18}+X^{17}+X^{13}+X^9+X^8+X^7+X^6+X^4+1\right)$$

FIGURE 2. The factorizations of five random polynomials of degree 25 over $\mathbb{F}_2$. One out of five polynomials in this sample has no root in the base field (the asymptotic probability is $\frac{1}{4}$ by Note 3).

In other words: *the probability for a random mapping of size n to consist of a single component is* $\sim \sqrt{\frac{\pi}{2n}}$. Also, the mean number of components in a random mapping of size n is

$$\frac{1}{2}\log n + \log\sqrt{2e^{\gamma}} + O(n^{-1/2}).$$

Similar properties hold for mappings without fixed points that are analogous to derangements and were discussed in Chapter II. .... END OF EXAMPLE 1. $\square$

We shall see below on page 443 that unlabelled functional graphs, which are the counterpart of (labelled) mappings, also resort to the exp-log schema.

EXAMPLE 2. *Polynomials over finite fields.* Given its importance in various areas of mathematics and in applications to coding theory, symbolic computation, and cryptography, we devote a special item to this factorization properties of random polynomials over finite fields. A preliminary discussion has already been given at the end of Chapter I.

Let $\mathbb{F}_p$ be the finite field with p elements and $\mathcal{P} = \mathbb{F}_p[X]$ the set of monic polynomials with coefficients in the field. We view these polynomials as (unlabelled) combinatorial objects with size identified to degree. Since a polynomial is specified by the sequence of its coefficients, there are p^n monic polynomials of degree n, and the OGF is $P(z) = (1-pz)^{-1}$.

Polynomials are a unique factorization domain, since they can be subjected in the usual way to Euclidean division. A nonconstant polynomial that has no proper nonconstant divisor is termed *irreducible* —irreducibles are the analogues of the primes in the integer realm. The unique factorization property implies that the collection $\mathcal{I}$ of monic irreducible polynomials is implicitly determined by $\mathcal{P} \cong \mathfrak{M}(\mathcal{I})$, which is reflected by a functional relation determining $I(z)$:

$$\log\frac{1}{1-pz} = I(z) + \frac{1}{2}I(z^2) + \frac{1}{3}I(z^3) + \cdots .$$

As seen in Chapter I, one can solve explicitly for $I(z)$ using Möbius inversion, to the effect that

$$\begin{aligned} I(z) &= \sum_{k\geq 1}\frac{\mu(k)}{k}\log\frac{1}{1-pz^k} \\ &= \log\frac{1}{1-pz} + R(z), \end{aligned}$$

where $R(z)$ is analytic in $|z| < p^{-1/2}$. Thus $I(z)$ is logarithmic. There results that $I_n \sim p^n/n$, which constitutes a "Prime Number Theorem" for polynomials over finite fields: *A fraction asymptotic to $\frac{1}{n}$ of the polynomials in $\mathbb{F}_p[X]$ are irreducible.* This says that a polynomial of degree n is roughly comparable to a number written in base p having n digits as the proportion of prime numbers amongst such numbers is about $1/(n \log p)$, by virtue of the classical Prime Number Theorem.

In addition, Theorem VII.1 yields: *The mean number of factors of a random polynomial of degree n is $\sim \log n$ and the distribution is concentrated.* This and similar developments lead to a complete analysis of some of the basic algorithms known for factoring polynomials over finite fields; see [**155**]. . . . . END OF EXAMPLE 2. □

▷ **1.** *The divisor function for polynomials.* Let $\delta(\varpi)$ for $\varpi \in \mathcal{P}$ be the total number of monic polynomials (not necessarily irreducible) dividing ϖ: if $\varpi = \iota_1^{e_1} \cdots \iota_k^{e_k}$, where the ι_j are distinct irreducibles, then $\delta(\varpi) = (e_1 + 1) \cdots (e_k + 1)$. One has

$$\mathbb{E}_{\mathcal{P}_n}(\delta) = \frac{[z^n] \prod_{j \geq 1} (1 + 2z^j + 3z^{2j} + \cdots)}{[z^n] \prod_{j \geq 1} (1 + z^j + z^{2j} + \cdots)} = \frac{[z^n] P(z)^2}{[z^n] P(z)},$$

so that the mean value of δ over $\mathcal{P}_n$ is exactly $(n+1)$. This evaluation is relevant to polynomial factorization over $\mathbb{Z}$ since it gives an upper bound on the number of ireducible factor combinations that need to be considered in order to lift a factorization from $\mathbb{F}_p(X)$ to $\mathbb{Z}(X)$; see [**428, 270**]. ◁

▷ **2.** *The cost of finding irreducible polynomials.* Assume that it takes expected time $t(n)$ to *test* a random polynomial of degree n for irreducibility. Then it takes expected time $\sim nt(n)$ to *find* a random polynomial of degree n: simply draw a polynomial at random and test it for irreducibility. Testing fro ireducibility can be achieved by developing a polynomial factorization algorithm which is stopped as soon as a nontrivial factor is found. See works by Panario *et al.* for details [**336, 337**]. ◁

Under the exp–log conditions, it is also possible (and easy) to analyse the *profile* of structures, that is the number of components of size r for each fixed r. We recall here that the Poisson distribution of parameter ν is the law of a discrete random variable Y such that

$$E(u^Y) = e^{-\nu(1-u)}, \qquad \mathbb{P}(Y = k) = e^{-\nu} \frac{\nu^k}{k!}.$$

A variable Y is said to be a *negative binomial* of parameter (m, α) if its probability generating function and its individual probabilities satisfy:

$$E(u^Y) = \left(\frac{1-\alpha}{1-\alpha u} \right)^m, \qquad \mathbb{P}(Y = k) = \binom{m+k-1}{k} \alpha^k (1-\alpha)^m.$$

(The quantity $\mathbb{P}(Y = k)$ is the probability that the mth success in a sequence of independent trials with individual success probability α occurs at time $m+k$; see [**134**, p. 165].)

PROPOSITION VII.1 (Profiles of exp–log structures). *Assume the conditions of Theorem VII.1 and let the number of $X^{(r)}$ of components of size r in a random $\mathcal{F}_n$ object of size n. In the labelled case, $X^{(r)}$ admits a limit distribution of the Poisson type in the sense that, for any fixed k,*

$$(3) \qquad \lim_{n \to \infty} \mathbb{P}_{\mathcal{F}_n}(X^{(r)} = k) = e^{-\nu} \frac{\nu^k}{k!}, \quad \nu = g_r \rho^r, \quad g_r \equiv [z^r] G(z).$$

In the unlabelled case, $X^{(r)}$ admits a limit distribution of the negative binomial type in the sense that, for any fixed k,

(4)
$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{F}_n}(X^{(r)}=k) = \binom{G_r+k-1}{k}\alpha^k(1-\alpha)^{G_r}, \quad \alpha=\rho^r,\ G_r \equiv [z^r]G(z).$$

PROOF. In the labelled case, the BGF of $\mathcal{F}$ with u marking the number $X^{(r)}$ of r–components is

$$F(z,u) = \exp\left((u-1)g_r z^r\right) F(z).$$

Extracting the coefficient of u^k leads to

$$[u^k]F(z,u) = \exp\left(-g_r z^r\right) \frac{z^{kr}}{k!} F(z),$$

to which singularity analysis applies directly. Observe that the factor of $F(z)$ contributes the probability of (3) as $z \to \rho$ while the singularity type of $F(z)$ remains unaffected.

In the unlabelled case, the starting BGF equation is

$$F(z,u) = \left(\frac{1-z^r}{1-uz^r}\right)^{G_r} F(z),$$

and the analytic reasoning is similar to the labelled case. □

The unlabelled version covers in particular polynomials over finite fields; see [**155, 260**] for related results.

▷ **3.** *Mean profiles.* The mean value of $X^{(r)}$ satisfies

$$\mathbb{E}_{\mathcal{F}_n}(X^{(r)}) \sim g_r\rho^r, \qquad \mathbb{E}_{\mathcal{F}_n}(X^{(r)}) \sim G_r \frac{\rho^r}{1-\rho^r},$$

in the labelled and unlabelled (multiset) case respectively. In particular, the mean number of roots of a random polynomial over $\mathbb{F}_p$ that lie in the base field $\mathbb{F}_p$ is asymptotic to $\frac{p}{p-1}$; the asymptotic probability that a polynomial has no root in the base field is $(1-1/p)^p$. ◁

▷ **4.** *Profiles of powersets.* In case of unlabelled powersets $\mathcal{F} = \mathfrak{P}(\mathcal{G})$ (no repetitions of elements allowed), the distribution of $X^{(r)}$ satisfies

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{F}_n}(X^{(r)}=k) = \binom{G_r}{k}\alpha^k(1-\alpha)^{G_r-k}, \quad \alpha = \frac{\rho^r}{1+\rho^r},$$

i.e., the limit is a binomial law of parameters $(G_r, \rho^r/(1+\rho^r))$. ◁

VII. 2. Simple varieties of trees

A *simple variety of trees* $\mathcal{V}$ is a class of trees determined by a subset Ω of the integers, so that all node degrees of a tree in $\mathcal{V}$ are constrained to belong to $\Omega \ni 0$. Such simple varieties exist in four versions: plane or nonplane, unlabelled or labelled. In three of the four cases, the generating function of $\mathcal{V}$ satisfies an equation of the form

(5)
$$y(z) = z\phi(y(z)),$$

corresponding to the fact that a tree is recursively formed as a root to which is appended a collection (either a sequence or a set) of subtrees. The algebraic situation is then summarized by the following table:

(6)

	Plane	*Non-plane*
Unlabelled (OGF)	$\mathcal{V} = \mathcal{Z} \times \mathfrak{S}_\Omega(\mathcal{V})$ $V(z) = z\phi(V(z))$ $\phi(w) := \sum_{\omega\in\Omega} u^\omega$	$\mathcal{V} = \mathcal{Z} \times \mathfrak{M}_\Omega(\mathcal{V})$ $V(z) = z\Phi(V(z)))$ (Φ a Pólya operator)
Labelled (EGF)	$\mathcal{V} = \mathcal{Z} \star \mathfrak{S}_\Omega(\mathcal{V})$ $\widehat{V}(z) = z\phi(\widehat{V}(z))$ $\phi(w) := \sum_{\omega\in\Omega} u^\omega$	$\mathcal{V} = \mathcal{Z} \star \mathfrak{P}_\Omega(\mathcal{V})$ $\widehat{V}(z) = z\phi(\widehat{V}(z))$ $\phi(w) := \sum_{\omega\in\Omega} \frac{u^\omega}{\omega!}$

The generating functions are ordinary (V) in the unlabelled case, exponential ($\widehat{V}$) otherwise. The nonplane unlabelled trees further involve a Pólya operator Φ, which is a sum of monomials in the quantities $V(z^2), V(z^3), \ldots$.

The relation $y = z\phi(y)$, which prevails in the first three cases of (6) have been treated in Section VI.5. In essence, y is defined by inversion of the relation $z = y/\phi(y)$ and inversion "fails" when the first derivative of the function to be inverted vanishes. At this point, the dependency $y \mapsto z$ becomes quadratic, so that its inverse $z \mapsto y$ gives rise to a square-root singularity. We are going to explore the effect of this situation on the probabilistic behaviour of tree parameters.

VII. 2.1. Basic analyses. In the three cases resorting to Equation (5), the main quantities of interest involve a characteristic equation and a condition on the basic constructor function ϕ of (6), which is invariably assumed to be analytic at the origin. The condition is the existence of the characteristic quantity $\tau > 0$ satisfying

(7) $$\tau\phi'(\tau) - \phi(\tau) = 0, \qquad 0 < \tau < R_{\rm conv}(\phi).$$

We recall also that $\phi(w)$ is said to be *unperiodic* if a decomposition $\psi(w) = w^a h(w^d)$ with h analytic at 0 implies $d = 1$. Paraphrasing the results of Section VI.5, we state:

THEOREM VII.2 (Enumeration of simple trees). *Let $y(z)$ be defined by associated to a simple variety according to* (5)*. Assume that ϕ is unperiodic and such that the characteristic condition* (7) *is satisfied. Then the coefficients of $y(z)$ admit a complete asymptotic expansion*

$$[z^n]y(z) \sim \frac{\gamma\rho^{-n}}{2\sqrt{\pi n^3}}\left[1 + \sum_{k=1}^{\infty}\frac{e_k}{n^k}\right], \qquad \rho := \frac{\tau}{\phi(\tau)}, \quad \gamma = \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}.$$

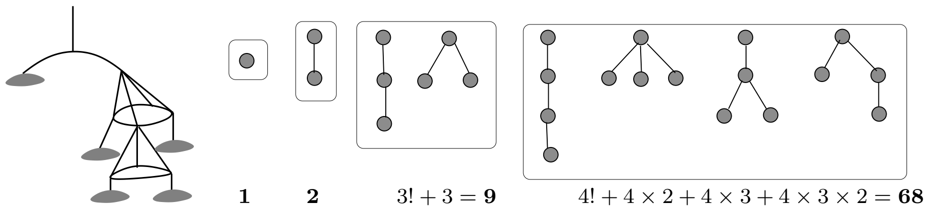
The heart of the matter is, under the conditions of the theorem, the singular expansion of $y(z)$ at $z = \rho$,

(8) $$y(z) = \tau + \sum_{j=1}^{\infty}(-1)^j d_j^\star\left(1 - \frac{z}{\rho}\right)^{j/2},$$

that will prove essential in the analysis of many tree parameters. The developments that follow all make use of the assumptions of Theorem VII.2. We set for convenience of notations:

$$\gamma := d_1^\star = \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}.$$

$\triangleright$ **5.** *Mobiles.* A (labelled) *mobile*, as defined by Bergeron, Labelle, and Leroux [**37**, p. 240], is a (labelled) tree in which subtrees dangling from the root are taken up to cyclic shift:



(Think of Calder's creations.) The EGF satisfies (*EIS* **A038037**)

$$\begin{aligned} M(z) &= z + 2\frac{z^2}{2!} + 9\frac{z^3}{3!} + 68\frac{z^4}{4!} + 730\frac{z^5}{5!} + \cdots \\ &= z\left(1 + \log\frac{1}{1 - M(z)}\right). \end{aligned}$$

The asymptotic formula for the number of of mobiles is $\frac{1}{n!}M_n \sim C \cdot A^n n^{-3/2}$, where $C \doteq 0.46563$, $A \doteq 1.15741$; see [**37**, p. 261]. $\triangleleft$

EXAMPLE 3. *Root degrees in simple varieties.* As an immediate application, we discuss the probability distribution of *root-degree* in simple varieties of trees under the conditions of Theorem VII.2. Let $\mathcal{V}^{[k]}$ be the subset of $\mathcal{V}$ composed of all trees whose root has degree equal to k. The quantity $V_n^{[k]}/V_n$ is the probability that the root of a random tree of size n has degree k. Since a tree in $\mathcal{V}^{[k]}$ is formed by appending a root to a collection of k trees, one has

$$V^{[k]}(z) = \phi_k z V(z)^k, \qquad \phi_k := [w^k]\phi(w).$$

For any *fixed* k, a singular expansion results from raising both members of (8) to the kth power, so that, in particular,

$$V^{[k]}(z) = \tau^k - k\gamma\tau^{k-1}\sqrt{1 - \frac{z}{\rho}} + O\left(1 - \frac{z}{\rho}\right). \tag{9}$$

This is to be compared to the basic estimate (8): the ratio $V_n^{[k]}/V_n$ is then asymptotic to the ratio of the coefficients of $\sqrt{1 - z/\rho}$ in the corresponding generating functions, $V^{[k]}(z)$ and $V(z) \equiv y(z)$. Thus, for any fixed k, we have found that

$$\frac{V_n^{[k]}}{V_n} = \rho k \phi_k \tau^{k-1} + O(n^{-1/2}). \tag{10}$$

(The error term is in fact of the form $O(n^{-1})$, as seen when pushing the expansion one step further.) Since $\rho = 1/\phi'(\tau)$, one can rephrase (10) as follows: *The random variable Δ representing the root-degree admits a discrete limit distribution given by*

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{V}_n}(\Delta) = \frac{k\phi_k\tau^{k-1}}{\phi'(\tau)}. \tag{11}$$

By general principles (Chapter IX), the convergence is uniform and a stronger form can be developed from singularity analysis techniques. The probability generating function of the limit law admits the simple form $u\phi'(\tau u)/\phi'(\tau)$.

For our four pilot examples, this gives:

Tree	$\phi(w)$	τ, ρ	*Probability distr.*	(*type*)
binary	$(1+w)^2$	$1, \frac{1}{4}$	PGF: $\frac{1}{2}u + \frac{1}{2}u^2$	(Bernoulli)
unary-binary	$1+w+w^2$	$1, \frac{1}{3}$	PGF: $\frac{1}{3}u + \frac{2}{3}u^2$	(Bernoulli)
general	$(1-w)^{-1}$	$\frac{1}{2}, \frac{1}{4}$	PGF: $u/(2-u)$	(Sum of two geometric)
Cayley	e^w	$1, e^{-1}$	PGF: ue^{u-1}	(shifted Poisson)

The probability distribution is thus characterized by the fact that its probability generating function is a rescaled version of the *derivative* of the basic tree constructor $\phi(w)$. END OF EXAMPLE 3. □

EXAMPLE 4. *Mean level profile in simple varieties.* The question we address here is that of determining the mean number of nodes at level k (i.e., at distance k from the root) in a random tree of some large size n. An explicit expression for the joint distribution of nodes at all levels has been developed in Section 5 of Chapter III, but this exact multivariate representation is somewhat hard to interpret in concrete terms.

Let $\xi_j(t)$ be the number of nodes at level k in tree t. Define the generating function of cumulated values,

$$X_k(z) := \sum_{t \in \mathcal{V}} \xi_k(t) z^{|t|}.$$

Clearly, $X_0(z) \equiv y(z)$ since each tree has a unique root. Then, since the parameter ξ_k is the sum over subtrees of parameter ξ_{k-1}, we are in the situation of inheritance, as discussed in Chapter III. We find for a tree with root subtrees $t_1, \ldots, t_{\deg(t)}$,

$$\begin{aligned} X_k(z) &= \sum_{t \in \mathcal{V}} \left(\sum_{j=0}^{\deg(t)} \xi_k(t_j) \right) z^{|t|} \\ &= z \sum_r r\phi_r y(z)^{r-1} X_{k-1}(z) \quad = \quad z\phi'(y(z)) X_{k-1}(z), \end{aligned}$$

so that by recurrence:

$$X_k(z) = \left(z\phi'(y(z))\right)^k y(z). \tag{12}$$

Making use of the (analytic) expansion of ϕ' at τ, namely, $\phi'(y) \sim \phi'(\tau) + \phi''(\tau)(y-\tau)$ and of $\rho\phi'(\tau) = 1$, one gets for any fixed k

$$X_k(z) \sim \left(1 - k\gamma\rho\phi''(\tau)\sqrt{1-\frac{z}{\rho}}\right)\left(\tau - \gamma\sqrt{1-\frac{z}{\rho}}\right) \sim \tau - \gamma(\tau\rho\phi''(\tau)k+1)\sqrt{1-\frac{z}{\rho}}.$$

Thus comparing the singular part of $X_k(z)$ to that of $y(z)$, we find: *For* fixed *k, the mean number of nodes at level k in a tree is of the asymptotic form*

$$\mathbb{E}_{\mathcal{V}_n}[\xi_k] \sim Ak + 1, \qquad A := \tau\rho\phi''(\tau).$$

This result was first given by Meir and Moon in an important paper of 1978, which started the general theory of simple families of trees [**312**]. The striking fact is that, although the number of nodes at level k can at least double at each level (in the case of the most shrubby trees), the growth is only linear on average. In figurative terms, the immediate vicinity of the root starts like a "cone". and trees of simple varieties tend to be rather skinny near their base.

When used in conjunction with saddle point bounds, the exact GF expression of (12) additionally provides a probabilistic upper bound on the height of trees of the form $O(n^{1/2+\delta})$ for any $\delta > 0$. Indeed restrict z to the interval $(0, \rho)$ and assume that $k = n^{1/2+\delta}$. Let χ be the height parameter. First, we have

$$\mathbb{P}_{\mathcal{V}_n}(\chi \geq k) \equiv \mathbb{E}_{\mathcal{V}_n}([\![\xi_k \geq 1]\!]) \leq \mathbb{E}_{\mathcal{V}_n}(\xi_k). \tag{13}$$

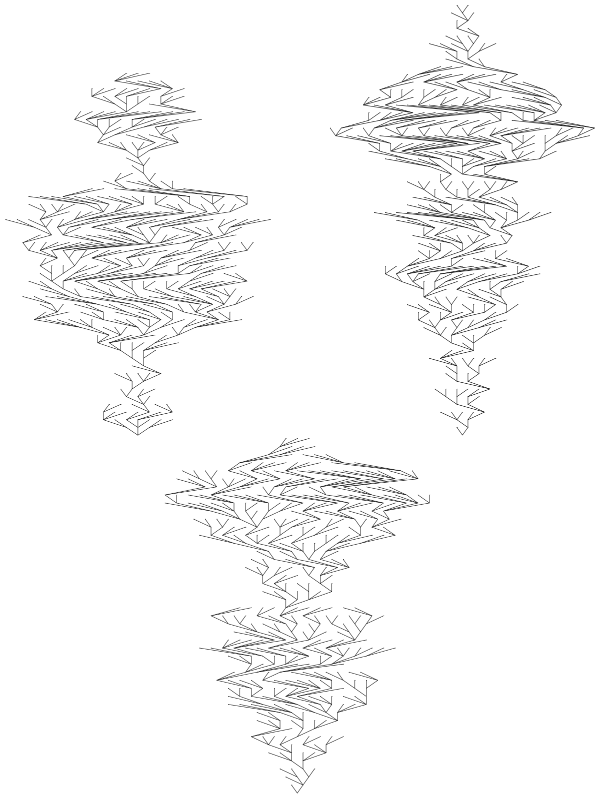


FIGURE 3. Three random 2–3 trees ($\Omega = \{0, 2, 3\}$) of size $n = 500$ have height respectively 48, 57, 47. The skinny aspect of the base, albeit subject to wide variation, is in accordance with the analysis developed in the text.

Next by saddle point bounds, for any legal positive x ($0 < x < R_{\mathrm{conv}}(\phi)$),

$$\mathbb{E}_{\mathcal{V}_n}(\xi_k) \le \left(x\phi'(y(x))\right)^k y(x)x^{-n} \le \tau\left(x\phi'(y(x))\right)^k x^{-n}. \tag{14}$$

Fix now

$$x = \rho - \frac{n^\delta}{n}.$$

Then, local expansions show that

$$\log\left(\left(x\phi'(y(x))\right)^k x^{-n}\right) \sim -\gamma_1 n^{3\delta/2} + \gamma_2 n^\delta, \tag{15}$$

for some positive constants γ_1, γ_2. Thus, by (13) and (14): *In a simple variety of trees, the probability of height exceeding $n^{1/2+\delta}$ is exponentially small, being of the rough form $\exp(-n^{3\delta/2})$. Accordingly, the mean height is $O(n^{1/2+\delta})$ for any $\delta > 0$.* Flajolet and Odlyzko [**165**] have characterized the moments of height, thenmean being in particular asymptotic to $\lambda\sqrt{n}$ and the limit distribution being of the Theta type already encountered in Chapter V in relation to the height of general Catalan trees. Further local limit and large deviation estimates appear in [**150**]. Figure 3 displays three random trees of size $n = 500$. .... END OF EXAMPLE 4. □

▷ **6.** *The variance of level profiles.* The BGF of trees with u marking nodes at level k has an explicit expression, in accordance with the developments of Chapter III. For instance for $k = 3$, this is $z\phi(z\phi(z\phi(uy(z))))$. Double differentiation followed by singularity analysis shows that

$$\mathbb{V}_{\mathcal{V}_n}[\xi_k] \sim \frac{1}{2}A^2k^2 - \frac{1}{2}A(3-4A)k + \tau A - 1,$$

another result of Meir and Moon [**312**]. The precise analysis of the mean and variance in the interesting regime where $k \asymp \sqrt{n}$ is also given in [**312**], but it requires the saddle point method of Chapter VIII or the methods of Chapter IX. ◁

VII. 2.2. Additive functionals. We consider next an important class of recursive parameters of trees, which generalize path length considered in Section 4 of Chapter III. A tree parameter ξ is said to be an *additive functional* if it is defined in terms of a simpler tree parameter η by a recursion of the type:

$$\xi(t) = \eta(t) + \sum_{j=1}^{\deg(t)} \xi(t_j), \tag{16}$$

where $\deg(t)$ is the degree of the root of t and the t_j are the root subtrees of t (whose number is $\deg(t)$). Unwinding the recursion shows that

$$\xi(t) := \sum_{s \preceq t} \eta(s), \tag{17}$$

where the sum is extended to all subtrees s of t (written $s \preceq t$). What is needed is access to moments of ξ viewed as a random variable over the subclass $\mathcal{V}_n$ of all trees of size n.

Expectation of a recursive parameter over trees of a fixed size is of prime relevance to the analysis of algorithms operating on trees [**382**]. For ease of notations, take the case of a simple variety of the *unlabelled plane type*. The generating function of cumulated values are defined in the usual way as

$$X(z) := \sum_{t \in \mathcal{V}} \xi(t) z^{|t|}, \qquad H(z) := \sum_{t \in \mathcal{V}} \eta(t) z^{|t|},$$

and the goal is to determine their relationship. We have

$$X(z) = H(z) + \widetilde{X}(z), \qquad \widetilde{X}(z) := \sum_{t \in \mathcal{V}} \left(z^{|t|} \sum_{j=1}^{\deg(t)} \xi(t_j) \right).$$

Spitting the expression of $\widetilde{X}(z)$ according to the values r of degree, we get

$$\begin{aligned}
\widetilde{X}(z) &= \sum_{r\geq 0}\left(\phi_r z^{1+|t_1|+\cdots+|t_r|}\sum_{j=1}^{r}\xi(t_j)\right)\\
&= \sum_{r\geq 0}\left(\phi_r z^{1+|t_1|+\cdots+|t_r|} r\xi(t_1)\right)\\
&= zX(z)\sum_{r\geq 0}\left(r\phi_r y(z)^{r-1}\right).
\end{aligned}$$

In summary, this leads to a linear equation satisfied by X,

$$X(z) = H(z) + z\phi'(y(z))X(z),$$

which solves to

$$X(z) = \frac{H(z)}{1 - z\phi'(y(z))} = \frac{zy'(z)}{y(z)}H(z). \tag{18}$$

This is our main equation. To get its second form, note that differentiating the fundamental relation $y = z\phi(y))$ yields the identity

$$y'(1 - z\phi'(y)) = \phi(y) = \frac{y}{z}, \qquad \text{i.e.,} \qquad 1 - z\phi'(y) = \frac{y}{zy'}.$$

▷ **7.** *A combinatorial interpretation.* Equation (17) suggest to view $X(z)$ as the gf of trees with one subtree marked to which is attached a weight of η. Then (18) can be read as follows: point to an arbitrary node at a tree in $\mathcal{V}$ (the gf is $zy'(z)$), "subtract" the tree attached to this node (a factor of $y(z)^{-1}$), and replace it by the same tree but now weighted by η (the gf is $H(z)$). ◁

▷ **8.** *Labelled varieties.* Formula (18) holds verbatim for labelled trees (either of the plane or nonplane type), provided we interpret $y(z), X(z), H(z)$ as egf's: $X(z) := \sum_{t\in\mathcal{V}}\xi(t)z^{|t|}/|t|!$, and so on. ◁

Given Equation (18), it is an easy task to churn out a number of mean value estimates for many tree parameters of interest.

EXAMPLE 5. *Mean degree profile.* Let $\xi(t) \equiv \xi_k(t)$ be the number of nodes of degree k in random tree of some variety $\mathcal{V}$. The analysis extends that of the root degree seen earlier. The parameter ξ is an additive functional induced by the basic parameter $\eta(t) \equiv \eta_k(t)$ defined by $\eta_k(t) := [\![\deg(t) = k]\!]$. By the analysis of root degree, we have for the GF of cumulated values associated to η

$$H(z) = \phi_k z y(z)^k, \qquad \phi_k := [w^k]\phi(w),$$

so that, by the fundamental formula (18),

$$X(z) = \phi_k z y(z)^k \frac{zy'(z)}{y(z)} = z^2\phi_k y(z)^{k-1}y'(z).$$

The singular expansion of $zy'(z)$ results from that of $y(z)$ by differentiation (Chapter VI),

$$zy'(z) = \frac{1}{2}\gamma\frac{1}{\sqrt{1 - z/\rho}} + O(1),$$

and the corresponding coefficient is $[z^n](zy') = ny_n$. This gives immediately the singularity type of X, which is of the form of an inverse square root. Thus,

$$X(z) \sim \rho\phi_k\tau^{k-1}(zy'(z))$$

implying ($\rho = \tau/\phi(\tau)$)

$$\frac{X_n}{ny_n} \sim \frac{\phi_k \tau^k}{\phi(\tau)}.$$

Consequently, one has: *In a simple variety, the mean number of nodes of degree k is asymptotic to $\lambda_k n$, where $\lambda_k := \phi_k \tau^k/\phi(\tau)$; in other words, the probability distribution of the degree Δ' of a random node in a random tree of size n satisfies*

$$\lim_{n\to\infty} \mathbb{P}_n(\Delta) = \frac{\phi_k \tau^k}{\phi(\tau)}, \qquad \text{with PGF :} \quad \sum_k \lambda_k u^k = \frac{\phi(u\tau)}{\phi(\tau)}.$$

$\triangleright$ **9.** *Variances.* The variance of the number of k-ary nodes is $\sim \nu n$, so that the distribution of the number of nodes of this type is concentrated, for each fixed k. The starting point is the BGF defined implicitly by

$$Y(z,u) = z\left(\phi(Y(z,u)) + \phi_k(u-1)Y(z,u)^k\right),$$

upon taking a double derivative with respect to u, setting $u = 1$, and finally performing singularity analysis on the resulting GF of cumulated values. $\triangleleft$

For the usual tree varieties this gives:

Tree	$\phi(w)$	τ, ρ	*Probability distr.*	(*type*)
binary	$(1+w)^2$	$1, \frac{1}{4}$	PGF: $\frac{1}{4} + \frac{1}{2}u + \frac{1}{4}u^2$	(Bernoulli)
unary-binary	$1 + w + w^2$	$1, \frac{1}{3}$	PGF: $\frac{1}{3} + \frac{1}{3}u + \frac{1}{3}u^2$	(Bernoulli)
general	$(1-w)^{-1}$	$\frac{1}{2}, \frac{1}{4}$	PGF: $1/(2-u)$	(Geometric)
Cayley	e^w	$1, e^{-1}$	PGF: e^{u-1}	(Poisson)

For instance, asymptotically, a general Catalan tree has on average $\frac{n}{2}$ leaves, $\frac{n}{4}$ nodes of degre 1 $\frac{n}{8}$ of degree 2, and so on; a Cayley tree has $\sim ne^{-1}/k!$ nodes of degree k; for binary (Catalan) trees, the four possible types of nodes each appear each with asymptotic frequency $\frac{1}{4}$. (These data are in agreement with the fact that a random tree under $\mathcal{V}_n$ is distributed like a branching process tree determined by the PGF $\phi(u\tau)/\phi(\tau)$; see the corresponding remarks in Chapter III.)
. . . . END OF EXAMPLE 5. $\square$

$\triangleright$ **10.** *The mother of a random node.* The discrepancy in distributions between the root degree and the the degree of a random node deserves an explanation. Pick up a node distinct from the root at random in a tree and look at the degree of its mother. The PGF of the law is in the limit $u\phi'(u\tau)/\phi'(\tau)$. Thus the degree of the root is asymptotically the same as that of the mother of any non-root node.

More generally, let X have distribution $p_k := \mathbb{P}(X = k)$. Construct a random variable Y such that the probability $q_k := \mathbb{P}(Y = k)$ is proportional both to k and p_k. Then for the associated PGFs, the relation $q(u) = p'(u)/p'(1)$ holds. The law of Y is said to be the *size-biased* version of the law of X. Here, a mother is picked up with an importance proportional to its degree. $\triangleleft$

EXAMPLE 6. *Path length.* Path length in trees can be analysed starting from the bivariate generating function given (see Chapter III, p. 174) by a functional equation of the difference type that involves the transformation $z \mapsto uz$. This is useful for the computation of higher moments, but for mean values, one may as well proceed directly from the additive functional scheme. Path length is definable inductively by

$$\xi(t) = |t| - 1 + \sum_{j=1}^{\deg(t)} \xi(t_j).$$

In this case, we have $\eta(t) = |t| - 1$ corresponding to the GF of cumulated values

$$H(z) = zy'(z) - y(z).$$

Thus, by the fundamental relation (18), one finds:

$$X(z) = (zy'(z) - y(z))\frac{zy'(z)}{y(z)} = \frac{z^2y'(z)^2}{y(z)} - zy'(z).$$

Since the type of $y'(z)$ at its singularity is $Z^{-1/2}$ where $Z := (1 - z/\rho)$, and the formula for $X(z)$ involves the square of y', the singularity of $X(z)$ is of type $Z^{-1/2}$, that is, it resembles a simple pole. This means that $X_n = [z^n]X(z)$ grows like ρ^{-n}, so that the mean value of ξ over $\mathcal{V}_n$ has growth $n^{3/2}$. Determining the dominant terms is done by combining local singular expansions as usual:

$$X(z) + zy'(z) \sim \frac{\gamma^2}{4\tau}\frac{1}{Z} + O(Z^{-1/2}).$$

As a result: *In a random tree of size n in a simple variety, the expectation of path satisfies*

$$\mathbb{E}_{\mathcal{V}_n}(\xi) = \lambda\sqrt{\pi n^3} + O(n), \qquad \lambda := \sqrt{\frac{\phi(\tau)}{2\tau^2\phi''(\tau)}}. \tag{19}$$

For our classical varieties, the main terms of (19) are then:

Binary	Unary-binary	General	Cayley
$\sim \sqrt{\pi n^3}$	$\sim \frac{1}{2}\sqrt{3\pi n^3}$	$\sim \frac{1}{2}\sqrt{\pi n^3}$	$\sim \sqrt{\frac{1}{2}\pi n^3}$

Observe that the quantity $\frac{1}{n}\mathbb{E}_{\mathcal{V}_n}(\xi)$ represents the expected depth of a random node in random tree (the probability model is then $[1\,..\,n] \times \mathcal{V}_n$), which is thus $\sim \lambda\sqrt{n}$. This result is consistent with the previously noted fact that height of a tree is with high probability not much larger than $O(n^{1/2})$. .... END OF EXAMPLE 6. $\Box$

$\triangleright$ **11.** *Variance of path length.* The variance of path length is asymptotic to $\lambda_2 n^{3/2}$ for some computable constant $\lambda_2 > 0$. Hence the distribution is "spread". Louchard [**297**] and Takács [**403**] have additionally worked out the asymptotic form of all moments, leading to a characterization of the limit law of path length that can be described in terms of the Airy function and coincides with the Brownian excursion area. $\triangleleft$

$\triangleright$ **12.** *Generalizations of path length.* Take the elementary cost $\eta(t)$ to be $|t|^\alpha$ for some exponent $\alpha > 0$. Then the results of Chapter VI make it possible to analyse the relevant singularity of the GF $H(z)$ (via Hadamard products and singularities of polylogarithms). This entails for the mean values of ξ precise estimates generalizing those of path length (which corresponds to $\alpha = 1$). There is a quantitative difference depending on whether $\alpha < 1\frac{1}{2}$, $\alpha = \frac{1}{2}$, $\alpha > \frac{1}{2}$.

For instance for binary trees, one finds (K_α and K_0' are explicit constants):

$\eta(t)$ *in the case* $\vert t\vert = n$		$\mathbb{E}_n(\xi)$	
n^α	$(\frac{3}{2} < \alpha)$	$\frac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)} n^{\alpha+\frac{1}{2}}$	$+O(n^{\alpha-\frac{1}{2}})$
$n^{3/2}$		$\frac{1}{\Gamma(3/2)} n^2$	$+O(n\log n)$
n^α	$(\frac{1}{2} < \alpha < \frac{3}{2})$	$\frac{\Gamma(\alpha-\frac{1}{2})}{\Gamma(\alpha)} n^{\alpha+\frac{1}{2}}$	$+O(n)$
$n^{1/2}$		$\frac{1}{\sqrt{\pi}} n\log n$	$+O(n)$
n^α	$(0 < \alpha < \frac{1}{2})$	$K_\alpha n$	$+O(1)$
$\log n$		$K_0' n$	$-2\sqrt{\pi}n^{1/2} + O(1).$

This is based on an article by Fill, Flajolet, and Kapur [**135**]. ◁

In summary, the developments of this section thus provide precise information on the shape of large random trees in simple varieties. The root degree has a special distribution, a random node has degree obeying a distribution (with PGF $\phi(u\tau)/\phi(\tau)$) which directly reflects the nature of the tree constructor ϕ. A typical node is at depth about $\sqrt{n}$, the height of the tree being itself in all likelihood not much larger and path length being $\sim \lambda n^{3/2}$ on average.

VII. 2.3. Enumeration of some non-plane unlabelled trees. We shall discuss here the enumeration of two classes of non-plane trees following Pólya [**347, 349**] and Otter [**335**], which are very important sources for the asymptotic theory of tree enumeration (see also the brief account in [**223**]). These authors used the more traditional method of Darboux instead of singularity analysis, but this distinction is immaterial here as calculations develop under completely parallel lines under both theories. The two classes under consideration are those of "general" and binary non-plane unlabelled trees. Pólya operators are then central to the enumeration, and their treatment in terms of radii of convergence and singularities is typical of the asymptotic theory of unlabelled objects obeying symmetries, i.e., involving the unlabelled $\mathfrak{M}$, $\mathfrak{P}$, $\mathfrak{C}$ constructions.

We prove here:

PROPOSITION VII.2 (Special non-plane unlabelled trees). *Consider the two classes of* non-plane unlabelled trees

$$\mathcal{H} = \mathcal{Z} \times \mathfrak{M}(\mathcal{H}), \quad \mathcal{W} = \mathcal{Z} \times \mathfrak{M}_{\{0,2\}}(\mathcal{W})$$

respectively of the general and binary type. Then, with constants C_H, A_H *and* C_W, A_W *given in* (24) *and* (25)*, one has*

$$H_n \sim \frac{C_H}{2\sqrt{\pi n^3}} A_H^n, \qquad W_{2n-1} \sim \frac{C_W}{2\sqrt{\pi n^3}} A_W^n.$$

PROOF. (i) *General case.* The OGF of nonplane unlabelled trees is the analytic solution to the functional equation

$$H(z) = z\exp\left(\frac{H(z)}{1} + \frac{H(z^2)}{2} + \cdots\right). \tag{20}$$

Let T be the solution to

$$T(z) = ze^{T(z)}, \tag{21}$$

that is to say, the Cayley function. The function $H(z)$ has a radius of convergence ρ strictly less than 1 as its coefficients dominate those of $T(z)$, the radius of convergence of the latter being exactly $e^{-1} \doteq 0.367$. The radius ρ cannot be 0 since the number of trees is bounded from above by the number of plane trees whose OGF has radius $\frac{1}{4}$. Thus, one has $\frac{1}{4} \le \rho \le e^{-1}$.

Rewriting the defining equation of $H(z)$ as

$$H(z) = \zeta e^{H(z)} \qquad \text{with} \qquad \zeta := z \exp\left(\frac{H(z^2)}{2} + \frac{H(z^3)}{3} + \cdots\right),$$

we observe that $\zeta = \zeta(z)$ is analytic for $|z| < \rho^{1/2}$, that is to say in a disk that properly contains the disk of convergence of $H(z)$. We may thus rewrite $H(z)$ as

$$H(z) = T(\zeta(z)).$$

Since $\zeta(z)$ is analytic at $z = \rho$, a singular expansion of $H(z)$ near $z = \rho$ results from composing the singular expansion of T at e^{-1} with the analytic expansion of ζ at ρ. In this way, we get:

$$H(z) = 1 - C(1 - \frac{1}{\rho})^{1/2} + O((1 - \frac{z}{\rho})), C = \sqrt{2e\rho\zeta'(\rho)}. \tag{22}$$

Thus,

$$[z^n]H(z) \sim \frac{C}{2\sqrt{\pi n^3}} \rho^{-n}.$$

(ii) *Binary case.* Consider the functional equation

$$f(z) = z + \frac{1}{2}f(z)^2 + \frac{1}{2}f(z^2). \tag{23}$$

This enumerates non-plane binary trees with size defined as the number of leaves, so that $W(z) = \frac{1}{z}f(z^2)$. Thus, it suffices to analyse $[z^n]f(z)$, which avoids dealing with periodicity phenomena.

The OGF $f(z)$ has a radius of convergence ρ that is at least $\frac{1}{4}$ (since there are fewer non-plane trees than plane ones). It is also at most $\frac{1}{2}$ as results from comparison of f with the solution to the equation $g = z + \frac{1}{2}g^2$. We may then proceed as before: treat the term $\frac{1}{2}f(z^2)$ as a function in $|z| < \rho^{1/2}$, as though it were known, then solve.

To this effect, set

$$\zeta(z) := z + \frac{1}{2}f(z^2),$$

which exists in $|z| < \rho^{1/2}$. Then, the equation (23) becomes a plain quadratic equation, $f = \zeta + \frac{1}{2}f^2$, with solution

$$f(z) = 1 - \sqrt{1 - 2\zeta(z)}.$$

The singularity ρ is the smallest positive solution of $\zeta(\rho) = \frac{1}{2}$. The singular expansion of f is obtained by composing the analytic expansion of ζ at ρ with $\sqrt{1-2\zeta}$. The usual square-root singularity results:

$$f(z) \sim 1 - C\sqrt{1 - z/\rho}, \qquad C := \sqrt{2\rho\zeta'(\rho)}.$$

This induces the $\rho^{-n}n^{-3/2}$ form for the coefficients $[z^n]f(z) \equiv [z^{2n-1}]W(z)$. $\square$

$\triangleright$ **13.** *Full asymptotic expansions for H_n, W_{2n-1}.* These exist since the OGFs admit complete asymptotic expansions in powers of $\sqrt{1-z/\rho}$. $\triangleleft$

The argument used in the proof of the proposition may seem nonconstructive. However, numerically, the values of ρ and C may be determined to great accuracy. See the notes below as well as Finch's section on "Otter's tree enumeration constants" [**137**, Sec. 5.6].

$\triangleright$ **14.** *Numerical evaluation of constants I.* Here is an unoptimized procedure controlled by a parameter $m \geq 0$ for general non-plane unlabelled trees.

Procedure Get_value_of_ρ(m : integer);
1. Set up a procedure to compute and memorize the H_n on demand;
(this can be based on recurrence relations implied by $H'(z)$; see [**326**])
2. Define $f^{[m]}(z) := \sum_{j=1}^{m} H_n z^n$;
3. Define $\zeta^{[m]}(z) := z \exp\left(\sum_{k=2}^{m} \frac{1}{k} f^{[m]}(z^k)\right)$;
4. Solve numerically $\zeta^{[m]}(x) = e^{-1}$ for $x \in (0,1)$ to $\max(m, 10)$ digits of accuracy;
5. Return x as an approximation to ρ.

For instance, here is a conservative estimate of the accuracy attained for $m = 0, 10, \ldots, 50$ (in a few billion machine instructions):

$m=0$	$m=10$	$m=20$	$m=30$	$m=40$	$m=50$
$3\cdot 10^{-2}$	10^{-6}	10^{-11}	10^{-16}	10^{-21}	10^{-26}

Empirically, accuracy appears to be a little better than $10^{-m/2}$. This yields to 25D:
(24)
$$\rho \doteq 0.3383218568992076951961126, \ A_H \equiv \rho^{-1} \doteq 2.9557652856519949747148184$$
$$C_H \doteq 1.559490020374640885542206.$$

The formula of the Proposition correctly estimates H_{100} with a relative error of 10^{-3}. $\triangleleft$

$\triangleright$ **15.** *Numerical evaluation of constants II.* The procedure of the previous note adapts easily to give:
(25)
$$\rho \doteq 0.4026975036714412909690453, \ A_W \equiv \rho^{-1} \doteq 2.4832535361726368585622889$$
$$C_W \doteq 1.130033716398972007144137.$$

The formula of the Proposition correctly estimates $[z^{100}]f(z)$ with a relative error of $7\cdot 10^{-3}$. $\triangleleft$

The two results, general and binary, are thus obtained by a modification of the method used for simple varieties of trees, upon treating the Pólya operator part as a "perturbation" of the corresponding equations of simple varieties of trees. A more general theory is possible for any simple variety of unlabelled non-plane trees $\mathcal{T} = \mathcal{Z}\mathfrak{M}_\Omega(\mathcal{T})$ (for some $\Omega \subset \mathbb{Z}_{\geq 0}$, but it requires more advanced methods to be discussed below.

VII. 2.4. Tree like structures. It is possible to treat many combinatorial structures related to trees by techniques exposed earlier in this section. A square root singularity is usually the norm, as is its corollary, the $\rho^{-n}n^{-3/2}$ asymptotic form for coefficients. As the routine is now well established, we shall content ourselves with a brief discussion of some examples encountered earlier in this book.

1. Trees by leaves and hierarchies. Given a nonempty set $\Omega \subset \mathbb{Z}_{\geq 0}$ that does not contain 0,1, it makes sense to consider the class of labelled trees,

$$\mathcal{C} = Z + \mathfrak{K}_{\Omega}(\mathcal{C}), \qquad \mathfrak{K} = \mathfrak{S} \text{ or } \mathfrak{P}.$$

These are trees (plane or non-plane) with size counted as the number of leaves and with degrees constrained to lie in Ω. The EGF is then of the form

$$C(z) = z + \phi(C(z)).$$

(The fact that $0, 1 \notin \Omega$ ensures well foundedness of the definition.)

The discussion given for simple families of trees adapts easily. Assume for simplicity ϕ here to be entire. Then $C(z)$ is a solution to $y = z + \phi(y)$ that remains analytic, when z increases from 0 along the positive axis, as long as the function $y - \phi(y)$ has a nonzero derivative. Thus, the smallest positive singularity ρ and the corresponding singular value $\tau := y(\rho)$ satisfy the system of two equations:

$$\tau = \rho + \phi(\tau), \qquad 1 = \phi'(\tau). \tag{26}$$

In other words, τ is the smallest root of $\phi'(\tau) = 1$ and ρ is then determined by $\rho = \tau - \phi(\tau)$. Near (ρ, τ), the dependence between z and y is locally a quadratic one:

$$\rho - z \sim \frac{1}{2}\phi''(\tau)(\tau - y),$$

that is,

$$y \sim \tau - \gamma \left(1 - \frac{z}{\rho}\right)^{1/2}, \qquad \gamma := \sqrt{\frac{1}{2}\rho\phi''(\tau)}.$$

(A full expansion can be obtained.) Thus, for coefficients, one has:

$$[z^n]y \sim \frac{\gamma}{2\sqrt{\pi n^3}}\rho^{-n}, \qquad \gamma = \sqrt{\frac{1}{2}\rho\phi''(\tau)}. \tag{27}$$

In Chapter II, we have considered the class $\mathcal{H}$ of labelled hierarchies defined by the choice $\Omega = \{2, 3, \ldots\}$, i.e., $\mathcal{H} = \mathcal{Z} + \mathfrak{P}_{\geq 2}(\mathcal{H})$. These occur in statistical classification theory: given a collection of n distinguished items, H_n is the number of ways of superimposing a nontrivial classification (cf Figure 4). Such abstract classifications usually no planar structure so that $\phi(w) = \sum_{\omega \in \Omega} w^{\omega}/\omega!$ is here $e^w - 1 - w$. Thus, we find mechanically $\tau = \log 2$, $\rho = 2\log 2 - 1$, and

$$\frac{1}{n!}H_n \sim \frac{1}{2\sqrt{\pi n^3}}\left(2\log 2 - 1\right)^{-n+1/2}.$$

For the unlabelled version, $\widetilde{\mathcal{H}}$, a calculation combines the analysis above with the principles employed for nonplane trees to the effect that

$$\widetilde{H}_n \sim \frac{\gamma}{2\sqrt{\pi n^3}}\rho^{-n}, \qquad \rho \doteq 0.29224.$$

```
|===|===|===|German
|   |   |===|English
|   |   |===|Dutch
|   |
|   |===|===|Swedish
|       |===|Danish
|===|Greek
|
|===|===|===|Portuguese
|   |   |===|Spanish
|   |   |===|Italian
|   |   |===|French
|   |
|   |===|Romanian
|
|===|Armenian
```

FIGURE 4. A hierarchy placed on some of the modern Indoeuropean languages.

2. Mappings. The basic decomposition of mappings,

$$F = \exp(K), \qquad K = \log \frac{1}{1-T}, \qquad T = ze^T,$$

lends itself to a number of multivariate extensions. For instance, the parameter $\chi(\phi)$ equal to the number of cyclic points gives rise to the BGF

$$F(z,u) = \exp\left(\log \frac{1}{1-T}\right) = (1-uT)^{-1}.$$

The mean number of a cyclic points in a random mapping of size n is accordingly

$$\mu_n = \frac{n!}{n^n}[z^n]\left(\frac{\partial}{\partial u} F(z,u)\bigg|_{u=1}\right) = \frac{n!}{n^n}[z^n]\frac{T}{(1-T)^2}.$$

Singularity analysis is immediate as

$$\frac{T}{(1-T)^2} \underset{z\to e^{-1}}{\sim} \frac{1}{2}\frac{1}{1-ez} \qquad \text{implying} \qquad [z^n]\frac{T}{(1-T)^2} \underset{n\to\infty}{\sim} \frac{1}{2}e^n.$$

The mean number of cyclic points in a random n–mapping is asymptotic to $\sqrt{\pi n/2}$. A large number of parameters can be analysed in this way systematically as shown in the survey [**166**]: see Figure 5 for a summary of results whose proof we leave as an exercise to the reader. The leftmost table describes global parameters of mappings; the rightmost table is relative to properties of random point in random n-mapping: λ is the distance to its cycle of a random point, μ the length of the cycle to which the point leads, tree size and component size are respectively the size of the largest tree containing the point and the size of its (weakly) connected component. In particular, a random mapping of size n has relatively few components, some of which are expected to be of a fairly large size.

3. Simple varieties of mappings. Let Ω be a subset of the integers and consider mappings $\phi \in \mathcal{F}$ such that the number of preimages of any point is constrained to

# components	$\sim \frac{1}{2}\log n$
# cyclic nodes	$\sim \sqrt{\pi n/2}$
# terminal nodes	$\sim ne^{-1}$

tail length (λ)	$\sim \sqrt{\pi n/8}$
cycle length (μ)	$\sim \sqrt{\pi n/8}$
tree size	$\sim n/3$
component size	$\sim 2n/3$

FIGURE 5. Expectations of the main additive parameters of random mappings of size n.

lie in Ω. Such special mappings may serve to model the behaviour of special classes of functions under iteration. For instance the quadratic functions $\phi(x) = x^2 + a$ over $\mathbb{F}_p$ have the property that each element y has either zero, one, or two preimages (depending on whether $y - a$ is a quadratic nonresidue, 0, or a quadratic residue). Such constrained mappings are of interest in various areas of computational number theory and cryptography.

The basic decomposition of general mappings needs to be amended in this case. Start with the family of trees $\mathcal{T}$ that are the simple variety corresponding to Ω:

$$T = z\phi(T), \qquad \phi(w) := \sum_{\omega\in\Omega} \frac{u^\omega}{\omega 1}.$$

At any point of a cycle, one must graft r trees with the constraint that $r+1 \in \Omega$ (since one arrow "comes from" the cycle itself). Such legal tuples with a root appended are represented by

$$U = z\phi'(T),$$

since ϕ is an exponential generating function and shift corresponds to differentiation. Then connected components and components are formed in the usual way by

$$K = \log\frac{1}{1-U}, \qquad F = \exp(K) = \frac{1}{1-U}.$$

We assume that ϕ (i.e., Ω) satisfies the general conditions of Theorem VII.2, with τ the characteristic value. Then $T(z)$ has a square-root singularity at $\rho = \tau/\phi(\tau)$. The same holds for U which satisfies the singular expansion

$$U(z) \sim 1 - \rho\phi''(\tau)\gamma\sqrt{1-\frac{z}{\rho}}. \tag{28}$$

Thus, eventually,

$$F(z) \sim \frac{\lambda}{\sqrt{1-\frac{z}{\rho}}}.$$

There results the universality of the exponent $-1/2$ in such constrained mappings:

$$[z^n]F(z) \sim \frac{\lambda}{\sqrt{\pi n}}\rho^{-n},$$

which nicely extends what is known to hold for unrestricted mappings. The analysis of additive functionals can then proceed on lines very similar to the case of standard mappings, to the effect that the estimates of Figure 5 hold, albeit with different multiplicative constants. The programme just sketched has been carried out in a thorough way by Arney and Bender in [**14**] to which we refer for a detailed treatment.

4. Unlabelled functional graphs. Unlabelled functional graphs or mapping patterns ($\mathcal{F}$) are unlabelled digraphs in which each vertex has outdegree equal to 1. Equivalently, they can be regarded as multisets of components ($\mathcal{L}$) that are cycles of nonplane unlabelled trees ($\mathcal{H}$),

$$\mathcal{F} = \mathfrak{M}(\mathcal{L}); \quad \mathcal{L} = \mathfrak{C}(\mathcal{U}); \quad \mathcal{H} = \mathcal{Z} \times \mathfrak{M}(\mathcal{H}),$$

a specification that entirely parallels that of mappings.

The OGF $H(z)$ has a square-root singularity by virtue of (22) above, with additionally $H(\rho) = 1$. The translation of the unlabelled cycle construction,

$$L(z) = \sum_{j \geq 1} \frac{\varphi(j)}{j} \log \frac{1}{1 - H(z^j)},$$

implies that $L(z)$ is logarithmic, and $F(z)$ has a singularity of type $1/\sqrt{Z}$ where $Z := 1 - z/\rho$. Thus, *unlabelled functional graphs constitute an exp-log structure* with $\kappa = \frac{1}{2}$. The number of unlabelled functional graphs thus grows like $C\rho^{-n}n^{-1/2}$ and the mean number of components in a random functional graph is $\sim \frac{1}{2}\log n$, like for the labelled mapping counterpart. See [**313**] for more on this topic. Unlabelled functional graphs are sometimes called "mapping patterns" in the literature.

▷ **16.** *Alternative form for $F(z)$.* Arithmetical simplifications associated with the Euler totient function yield:

$$F(z) = \prod_{k=1}^{\infty} \Big(1 - H(z^k)\Big)^{-1}.$$

A similar form applies generally to multisets of unlabelled cycles. ◁

5. Unrooted trees. All the trees considered so far have been rooted and this version is the one most useful in applications. An *unrooted tree*[2] is by definition a connected acyclic (undirected) graph. In that case, the tree is clearly non-plane and no special root node is distinguished.

The counting of the class $\mathcal{U}$ of *unrooted labelled trees* is easy: there are plainly $U_n = n^{n-2}$ of these since all nodes are distinguished by a label, which entails that $nU_n = T_n$ with $T_n = n^{n-1}$ by Cayley's formula. Incidentally, the EGF $U(z)$ satisfies

$$U(z) = \int_0^z T(y)\,\frac{dy}{y} = T(z) - \frac{1}{2}T(z)^2, \tag{29}$$

as already seen when we discussed labelled graphs in Chapter II.

For *unrooted unlabelled trees*, symmetries are in the way and an tree can be rooted in a number of ways that depends on its shape. For instance of star graph leads to a number of different rooted trees that equals 2 (choose either the center or one of the peripheral nodes), while a line graph gives rise to $\lceil n/2 \rceil$ structurally different rooted trees. With $\mathcal{H}$ the class of rooted unlabelled trees and $\mathcal{I}$ the class of unrooted trees, we have at this stage only minor refinements of the general inequality

$$I_n \leq H_n \leq nI_n.$$

[2]Unrooted trees are also called sometimes *free trees*.

A table of values of the ratio H_n/I_n suggests that the answer is closer to the upper bound:

(30)

n	10	20	30	40	50	60
H_n/I_n	6.78	15.58	23.89	32.15	40.39	48.62

The solution is provided by a famous exact formula due to Otter, namely

$$I(z) = H(z) - \frac{1}{2}\left(H(z)^2 - H(z^2)\right). \tag{31}$$

It gives in particular (*EIS* **A000055**)

$$I(z) = z + z^2 + z^3 + 2\,z^4 + 3\,z^5 + 6\,z^6 + 11\,z^7 + 23\,z^8 + 47\,z^9 + 106\,z^{10} + \cdots.$$

Given (31), it is child's play to determine the singular expansion of I knowing that of H. The radius of convergence of I is the same as that of H as the term $H(z^2)$ only introduces exponentially small coefficients. Thus, it suffices to analyse $H - \frac{1}{2}H^2$:

$$H(z) - \frac{1}{2}H(z)^2 \sim \frac{1}{2} - \delta_2\left(1 - \frac{z}{\rho}\right) + \delta_3\left(1 - \frac{z}{\rho}\right)^{3/2} + O\left(\left(1 - \frac{z}{\rho}\right)^2\right).$$

What is noticeable is the cancellation in coefficients for the term $Z^{1/2}$ (since $1 - x - \frac{1}{2}(1-x)^2 = \frac{1}{2} + O(x^2)$), so that $Z^{3/2}$ is the actual singularity type of I. Clearly, the constant δ_3 is computable from the first four terms in the singular expansion of H at ρ. Then singularity analysis yields: *The number of unrooted trees of size n satisfies the formula*

$$I_n \sim \frac{3\delta_3}{4\sqrt{\pi n^5}}\rho^{-n}, \qquad I_n \sim 0.5349496061 \cdot 2.9955765856^n n^{-5/2}. \tag{32}$$

The numerical values are from [**137**] and the result is Otter's original [**335**]. The formula gives an error slighly under 10^{-2} for $n = 100$. An unrooted tree of size n gives rise to about different $0.8n$ rooted trees on average, which agrees well with the observations of (30).

▷ **17.** *Dissimilarity property of trees.* Fix an unrooted tree. Two vertices u, v are *similar* there exists an automorphism of the tree that exchanges u and v. Two edges e, f are similar if there is an automorphism of the tree that exchanges e and f. Let p^* and q^* be the number of equivalence classes of vertices and edges under similarity. Then the following identity, called the dissimilarity theorem, holds,

$$p^* - q^* = 1 - s,$$

where s is 1 if the tree admits a "central symmetric" edge and 0 otherwise. Summing over all unrooted trees gives the basic formula (31). [This is based on Read's account in [**349**, p. 107] and on [**223**, p. 56].] ◁

VII. 3. Positive implicit functions

Our goal here is to show that square root singularity holds in the wide context of generating functions satisfying a single functional equation,

$$y(z) = G(z, y(z)), \qquad G(0,0) = 0,\; G'_w(0,0) = 0, \tag{33}$$

under suitable conditions on G. (By trivial algebra, the initial conditions on G can be further weakened to $G(0,0) > 0$ and $0 \leq G'_w(0,0) < 1$). For many combinatorial applications, it is legitimate to assume that G is *bivariate analytic* around $(0,0)$ and that it has nonnegative Taylor coefficients there, an assumption that is made throughout this section. Thus, we postulate that the representation

$$(34) \qquad G(z,w) = \sum_{m,n\geq 0} g_{m,n} z^m w^n, \qquad \text{with} \quad g_{m,n} \geq 0,\ g_{0,0} = 0,\ g_{0,1} = 0,$$

holds in a domain $|z| < R$, $|y| < S$. Next, in order to avoid trivialities, we consider only $G's$ that are *nonlinear*, that is, one at least of the coefficients $g_{m,n}$ for $n \geq 2$ is strictly positive. (Else, one could plainly solve the linear equation for $y(z)$.) Accordingly, we also postulate that

$$(35) \qquad g_{m,n} > 0 \quad \text{for some } m \text{ and for some } n \geq 2.$$

Finally, for reasons of the same nature as in the discussion of simple families of trees, singularities may only occur from a failure of the implicit function theorem. As we shall see in the proof of the main theorem below, this necessitates the existence of two numbers $r, s > 0$ such that

$$(36) \qquad G(r,s) = s, \quad G_w(r,s) = 1, \qquad \text{with} \quad r < R, \quad s < S.$$

The system (36) is called the *characteristic system*. Finally, we recall that a function $f(z) = \sum_{n\geq 0} f_n z^n$ is *aperiodic* if cannot be put under the form $f(z) = z^a h(z^d)$ with $d > 2$ and h analytic at 0. (Put otherwise, there exist three indices n_1, n_2, n_3 such that $f_{n_1} f_{n_2} f_{n_3} \neq 0$ and $\gcd(n_2 - n_1, n_3 - n_1) = 1$.) The main result here [**28, 314**] is[3]

THEOREM VII.3 (Positive implicit functions). *Let $G(z,w)$ be a positive bivariate analytic function satisfying* (34) *and let the equation* (33) *admit a solution $y(z)$ that is analytic at* 0*, has nonnegative coefficients, and is aperiodic. Assume further the conditions* (35) *and* (36)*. Then, $y(z)$ converges at $z = r$ where it has a square-root singularity:*

$$y(z) \underset{z\to r}{=} s - \gamma\sqrt{1 - z/r} + O(1 - z/r), \qquad \gamma := \sqrt{\frac{2rG'_z(r,s)}{G''_{ww}(r,s)}}.$$

The expansion is valid in a Δ-domain, so that

$$[z^n]y(z) \underset{n\to\infty}{=} \frac{\gamma}{2\sqrt{\pi n^3}} r^{-n} \left(1 + O(n^{-1})\right).$$

Note that statement tacitly implies the existence of at most one root of the characteristic system within the analyticity domain of G.

PROOF. By assumption $y(z)$ is analytic at the origin with nonnegative Taylor coefficients. Let ρ be its radius of convergence, and $\tau = y(\rho)$. By Pringsheim's theorem, ρ is a singularity of y. Meir and Moon [**314**] provide an argument to the effect that

---

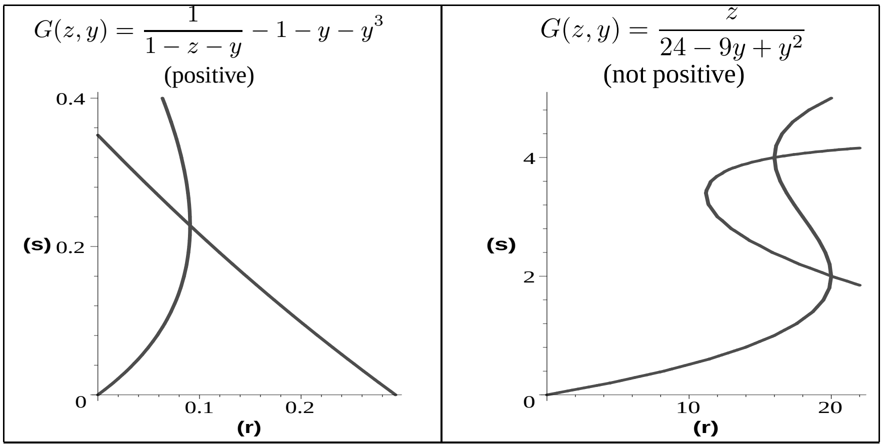
[3] This theorem has an interesting history. A version of it was first stated by Bender in 1974 (Theorem 5 of [**28**]). Canfield [**72**] then pointed out the fact that Bender's conditions were not quite sufficient. A corrected statement was given by Meir and Moon in [**314**] with a further (minor) erratum in [**315**]. We follow heer the form given in Theorem 10.13 by Odlyzko [**330**] with the correction of another minor misprint (regarding $g_{0,1}$ which should read $g_{0,1} \neq 1$ or, better, $g_{0,1} \in [0,1[$).

$\rho = r$ and $\tau = s$. The square-root singularity then follows from the usual argument based on the failure of the implicit function theorem. □

The solutions to the characteristic system (36) can be regarded as the intersection points of two curves, namely,

$$G(r,s) - s = 0, \qquad G'_w(r,s) = 1.$$

Here are plots in the case of two functions G: the first one has nonnegative coefficients while the second one (corresponding to a counterexample of Canfield [**72**]) involves negative coefficients. Positivity of coefficients implies convexity properties that avoid pathological situations.



EXAMPLE 7. *Trees with variable edge lengths and node sizes.* Consider unlabelled plane trees in which nodes can be of different sizes: what is given is a set $\widehat{\Omega}$ of ordered pairs (ω, σ), where a value (ω, σ) means that a node of degree ω and size σ is allowed. Simple varieties in their basic form correspond to $\sigma \equiv 1$; trees enumerated by leaves (including hierarchies) correspond to $\sigma \in \{0, 1\}$ with $\sigma = 1$ iff $\omega = 0$. Figure 6 indicates the way such trees can model the self-bonding of single stranded nucleic acids like RNA, according to Waterman *et al.* [**234, 378, 396**]. Clearly an extremely large number of variations are possible.

The fundamental equation in the case of a finite $\widehat{\Omega}$ is

$$Y(z) = P(z, Y(z)), \qquad P(z,w) := \sum_{(\omega,\sigma)\in\widehat{\Omega}} z^\sigma w^\omega.$$

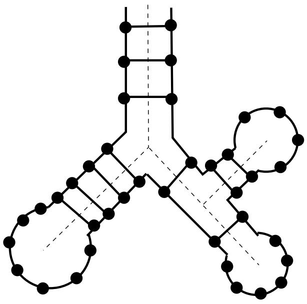
In the aperiodic case, we shall invariably have a formula of the form

$$Y_n \sim C \cdot A^n n^{3/2},$$

with the universal square-root singularity and the universal count exponent $-\frac{3}{2}$. END OF EXAMPLE 7. □

▷ **18.** *Schröder numbers.* Consider the class $\mathcal{Y}$ of unary-binary trees where unary nodes have size 2, while leaves and binary nodes have the usual size 1. The GF satisfies $Y = z + z^2 Y + zY^2$, so that

$$Y(z) = zD(z^2), \qquad D(z) = \frac{1 - z - \sqrt{1 - 6z + z^2}}{2z}.$$



A fragment of RNA is, in first approximation, a tree-like structure with edges corresponding to bases pairs and "loops" corresponding to leaves. There are constraints on the sizes of leaves (taken here between 4 and 7) and length of edges (here between 1 and 4 base pairs). Such a RNA fragment can then be viewed as a planted tree P attached to a binary tree (Y) with equations:

$$\begin{cases} P = AY, \quad Y = AY^2 + B, \\ A = z^2 + z^4 + z^6 + z^8, \ B = z^4 + z^5 + z^6 + z^7. \end{cases}$$

FIGURE 6. A simplified combinatorial model of RNA structures as considered by Waterman *et al.* [**234, 378, 396**].

We have $D(z) = 1+2\,z+6\,z^2+22\,z^3+90\,z^4+394\,z^5+\cdots$, which is *EIS* **A006318** ("Large Schröder numbers"). By the bijective correspondence between trees and lattice paths, $\mathcal{Y}_{2n+1}$ is in bijective correspondence with excursions of length n made of steps $(1,1)$, $(2,0)$, $(1,-1)$. Upon tilting by 45°, this is equivalent to paths connecting the lower left corner to the upper right corner of an $(n \times n)$ square that are made of horizontal, vertical, and diagonal steps, and never going under the diagonal. The series $S = \frac{z}{2}(1+D)$ enumerates Schröder's generalized parenthesis systems (Chapter I): $S := z + S^2/(1-S)$, and the asymptotic formula

$$S_n \sim \frac{1}{2} D_{n-1} \sim \frac{1}{4\sqrt{\pi n^3}} \left(3 - 2\sqrt{2}\right)^{-n+1/2}$$

follows straightforwardly. $\lhd$

EXAMPLE 8. *Nonplane trees and alkanes.* In chemistry, carbon atoms (C) are known to have valency 4 while hydrogen (H) has valency 1. *Alkanes*, also known as paraffins (Figure 7), are are acyclic molecules formed of carbon and hydrogen atoms according to this rule and without multiple bonds; they are thus of the type C_nH_{2n+2}. In combinatorial terms, we are talking of unrooted trees with (total) node degrees in $\{1,4\}$. The rooted version of these trees are determined by the fact that a root is chosen and (out)degrees of nodes lie in the set $\Omega = \{0,3\}$; these are rooted ternary trees and they correspond to alcohols (with the OH group marking one of the carbon atoms). The problem of enumerating isomers of alkanes and alcohols has been at the origin of Pólya paper [**347, 349**].

Alcohols ($\mathcal{A}$) are the simplest to enumerate as they are rooted trees. The OGF starts as (*EIS* **A000598**)

$$A(z) = 1 + z + z^2 + z^3 + 2\,z^4 + 4\,z^5 + 8\,z^6 + 17\,z^7 + 39\,z^8 + 89\,z^9 + \cdots,$$

with size being taken here as the number of internal nodes. The specification is

$$\mathcal{A} = \{\epsilon\} + \mathcal{Z}\mathfrak{M}_3(\mathcal{A}).$$

(Equivalently $\mathcal{A}^+ := \mathcal{A} \setminus \{\epsilon\}$ satisfies $\mathcal{A}^+ = \mathcal{Z}\mathfrak{M}0,1,2,3(\mathcal{A}^+)$.) This implies that $A(z)$ satisfies the functional equation:

$$A(z) = 1 + z\left(\frac{1}{3}A(z^3) + \frac{1}{2}A(z)A(z^2) + \frac{1}{6}A(z)^3\right).$$

```
      H           H   H           H   H   H           H  OH  H
      |           |   |           |   |   |           |   |   |
      |           |   |           |   |   |           |   |   |
   H--C--H     H--C---C--H     H--C---C---C--H     H--C---C---C--H
      |           |   |           |   |   |           |   |   |
      |           |   |           |   |   |           |   |   |
      H           H   H           H   H   H           H   H   H

    Methane      Ethane          Propane             Propanol
```

FIGURE 7. A few examples of alkanes (CH_4, C_2H_6, C_3H_8) and an alcohol.

In order to apply Theorem VII.3, introduce the function

$$G(z,w) = 1 + z\left(\frac{1}{3}A(z^3) + \frac{1}{2}A(z^2)w + \frac{1}{6}w^3\right),$$

which exists in $|z| < |\rho|^{1/2}$ and $|w| < \infty$, with ρ the (yet unknown) radius of convergence of A. Like before, the Pólya terms $A(z^2), A(z^3)$ are teated as known functions. By methods similar to those used in the analysis of binary and general trees (Subsection VII. 2.3), we find that the characteristic system admits a solution,

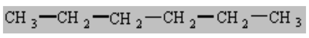
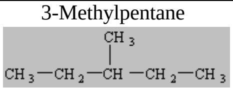
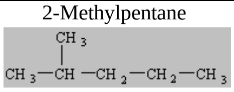
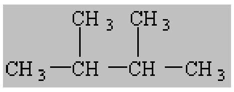
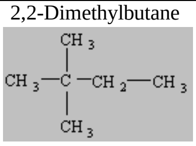
$$r \doteq 0.3551817423143773928, \qquad s \doteq 2.1174207009536310225,$$

so that $\rho = r$ and $y(\rho) = s$. Thus the growth of the number of alcohols is of the form $C \cdot 2.81546^n n^{-3/2}$.

Let $B(z)$ be the OGF of alkanes (*EIS* **A000602**):

$$\begin{aligned} B(z) &= 1 + z + z^2 + 2\,z^3 + 3\,z^4 + 5\,z^5 + 9\,z^6 + 18\,z^7 \\ &\quad + 35\,z^8 + 75\,z^9 + 159\,z^{10} + 355\,z^{11} + 802\,z^{12} + 1858\,z^{13} + \cdots \end{aligned}$$

For instance, $B_6 = 5$ because there are 5 isomers of hexane, C_6H_{14}, for which chemists had to develop a nomenclature system, interestingly enough based on a diameter of the tree:

Hexane	3-Methylpentane	2-Methylpentane
		
2,3-Dimethylbutane	**2,2-Dimethylbutane**	
		

The number of structurally different alkanes can then be found an adaptation of the dissimilarity formula (31) for which we refer to [**37**, p.290]. This problem has served as a powerful motivation for the enumeration of graphical trees and its has a fascinating history which goes back to Cayley. (See Rains and Sloane's article [**358**] and [**349**]). The asymptotic formula of alkanes is of the usual form with an $n^{-5/2}$ term, as these are unrooted molecules. END OF EXAMPLE 8. □

The pattern of analysis should by now be clear, and we state:

THEOREM VII.4 (Nonplane unlabelled trees). *Let $\Omega \ni 0$ be a finite subset of $\mathbb{Z}_{\geq 0}$ and consider the variety $\mathcal{V}$ of (rooted) nonplane unlabelled trees. Assume aperiodicity*

($\gcd(\Omega) = 1$) and the condition that Ω contains at least one element larger than 1. Then the number of trees of size n in $\mathcal{V}$ satisfies an asymptotic formula:

$$V_n \sim C \cdot A^n n^{-3/2}.$$

PROOF. The argument given for alcohols is transposed verbatim. Only the existence of a root of the characteristic system needs to be established.

The radius of convergence of $V(z)$ is a priori ≤ 1. The fact that ρ is strictly less than 1 is established by means of a lower bound, $V_n > B^n$ for some $B > 1$ and infinitely many large enough values of n. To obtain this exponential diversity, first choose an n_0 such that $V_{n_0} > 1$, then build a perfect d-ary tree (for some $d \in \Omega$, $d \neq 0, 1$) tree of height h, and finally graft freely subtrees of size n_0 at $n/(4n_0)$ of the leaves of the perfect tree. Choosing d such that $d^h > n/(4n_0)$ yields the lower bound. That the radius of convergence is nonzero results from the upper bound provided by corresponding plane trees whose growth is at leat exponential. Thus, one has $0 < \rho < 1$.

By the translation of multisets of bounded cardinality, the function G is polynomial in finitely many of the quantities $\{V(z), V(z^2), \ldots\}$. Thus the function $G(z, w)$ constructed like in the case of alcohols converges in $|z| < \rho^{1/2}$, $|w| < \infty$. As $z \to \rho^{-1}$, we must have $\tau := V(\rho)$ finite, since otherwise, there would be a contradiction in orders of growth in the nonlinear equation $V(z) = \cdots + \cdots V(z)^d \cdots$ as $z \to \rho$. Thus (ρ, τ) satisfies $\tau = G(\rho, \tau)$. For the derivative, one must have $G'_w(\rho, \tau) = 1$ since': (i) a smaller value would mean that V is analytic at ρ (by the Implicit Function Theorem); (ii) a larger value would mean that a singularity has been encountered earlier (by the usual argument on failure of the Implicit Function Theorem). Thus, Theorem VII.3 on positive implicit functions is applicable. $\square$

A large number of variations are clearly possible as evidenced by the title of an article [**222**] published by Harary, Robinson, and Schwenk in 1975, namely, "Twenty-step algorithm for determining the asymptotic number of trees of various species".

VII. 4. The analysis of algebraic functions

Algebraic series and algebraic functions are simply defined as solutions of a polynomial equation (Definition VII.3). It is a nontrivial fact established by elimination theory (which can itself be implemented by way of resultants or Groebner bases) that they are equivalently defined as components of solutions of polynomial systems .

The starting point is the following definition of an algebraic function.

DEFINITION VII.3. *A function $f(z)$ analytic in a neighbourhod $\mathcal{V}$ of a point z_0 is said to be* algebraic *if there exists a (nonzero) polynomial $P(z, y) \in \mathbb{C}[z, y]$, such that*

$$P(z, f(z)) = 0, \qquad z \in \mathcal{V}.$$

A power series $f \in \mathbb{C}[\![z]\!]$ is said to be an algebraic power series if it coincides with the expansion of an algebraic function at 0.

$\triangleright$ **19.** *Algebraic definition of algebraic series.* It is customary to define f to be an algebraic series if it satisfies $P(z, f) = 0$ in the sense of formal power series, without *a priori* consideration of convergence issues. Then the technique of majorizing series may be used to prove that the coefficients of f grow at most exponentially. Thus, the new definition is equivalent to Definition —refalg-def. $\lhd$

The *degree* of an algebraic series f is by definition the minimal value of $\deg_y P(z, y) = 1$ over all polynomials that are cancelled by f (so that rational series are algebraic of degree 1). Note that one can always assume P to be irreducible (that is $P = QR$ implies that one of Q or R is a scalar) and of minimal degree.

An algebraic function may also be defined by starting with a polynomial system of the form

$$\left\{ \begin{array}{lcl} P_1(z, y_1, \ldots, y_m) & = & 0 \\ \vdots & & \\ P_m(z, y_1, \ldots, y_m) & = & 0, \end{array} \right. \tag{37}$$

where each P_j is a polynomial. A solution of (37) is by definition an m-tuple $(f_1, \ldots, f_m)$ that cancels each P_j, that is, $P_j(z, f_1, \ldots, f_m) = 0$. Any of the f_j is called a component solution. A basic result of *elimination theory* is that any component solution of a nondegenerate polynomial system is an algebraic series (APPENDIX B: *Algebraic elimination*, p. 657). In other words, one can eliminate the auxiliary variables $y_2, \ldots, y_m$ and construct a single bivariate polynomial Q such that $Q(z, y_1) = 0$.

Algebraic functions have singularities constrained to be *branch points*. By this is meant that the local expansion at such a singularity is a fractional power series known as a Newton–Puiseux expansion. Singularity analysis is systematically applicable to algebraic functions—hence the characteristic form of asymptotic expansions that involve terms of the form $\omega^n n^{p/q}$ (for some algebraic number ω and some rational exponent p/q). In this section, we develop such basic structural results (Subsection VII. 4.1). However, coming up with effective solutions (i.e., decision procedures) is not obvious in the algebraic case. Hence, a number of nontrivial algorithms are also described in order to locate and analyse singularities (Newton's polygon method), and eventually determine the asymptotic form of coefficients. In particular, the multivalued character of algebraic functions creates a need to solve "connection problems". Finally, like in the rational case, positive systems (Subsection VII. 4.2) enjoy special properties that further constrain what can be observed as regards asymptotic behaviours, including a return of the square-root singularity. Our presentation of positive systems is based on an essential result of the theory, the Drmota–Lalley–Woods theorem, that plays for algebraic functions a rôle quite similar to that of Perron-Frobenius theory for rational functions.

VII. 4.1. General algebraic functions. Let $P(z, y)$ be an irreducible polynomial of $\mathbb{C}[z, y]$,

$$P(z, y) = p_0(z)y^d + p_1(z)y^{d-1} + \cdots + p_d(z).$$

The solutions of the polynomial equation $P(z, y) = 0$ define a locus of points (z, y) in $\mathbb{C} \times \mathbb{C}$ that is known as a *complex algebraic curve*. Let d be the y-degree of P. Then, for each z there are at most d possible values of y. In fact, there exist d values of y "almost always", that is except for a finite number of cases:

— If z_0 is such that $p_0(z_0) = 0$, then there is a reduction in the degree in y and hence a reduction in the number of y-solutions for the particular value of $z = z_0$. One can conveniently regard the points that disappear as "points at infinity".
— If z_0 is such that $P(z_0, y)$ has a multiple root, then some of the values of y will coalesce.

Define the *exceptional set* of P as the set ($\mathbf{R}$ is the resultant):

(38) $$\Xi[P] := \{z \mid R(z) = 0\}, \ R(z) := \mathbf{R}(P(z, y), \partial_y P(z, y), y).$$

(The quantity $R(z)$ is also known as the discriminant of $P(z, y)$ taken as a function of y.) If $z \notin \Xi[p]$, then we have a guarantee that there exist d distinct solutions to $P(z, y) = 0$, since $p_0(z) \neq 0$ and $\partial_y P(z, y) \neq 0$. Then, by the implicit function theorem, each of the solutions y_j lifts into a locally analytic function $y_j(z)$. What we call a *branch* of the algebraic curve $P(z, y) = 0$ is the choice of such an $y_j(z)$ together with a connected region of the complex plane throughout which this particular $y_j(z)$ is analytic.

Singularities of an algebraic function can thus only occur if z lies in the exceptional set $\Xi[P]$. At a point z_0 such that $p_0(z_0) = 0$, some of the branches escape to infinity, thereby ceasing to be analytic. At a point z_0 where the resultant polynomial $R(z)$ vanishes but $p_0(z) \neq 0$, then two or more branches collide. This can be either a multiple point (two or more branches happen to assume the same value, but each one exists as an analytic function around z_0) or a branch point (some of the branches actually cease to be analytic). An example of an exceptional point that is not a branch point is provided by the classical lemniscate of Bernoulli: at the origin, two branches meet while each one is analytic there (see Figure 8).

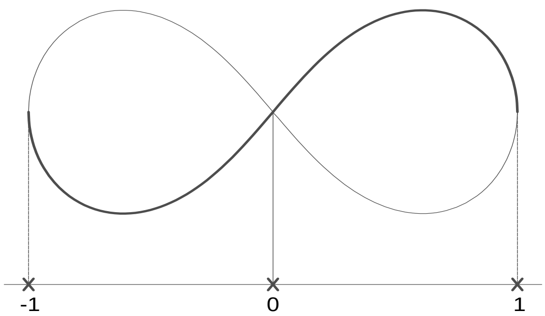


FIGURE 8. The lemniscate of Bernoulli defined by $P(z, y) = (z^2 + y^2)^2 - (z^2 - y^2) = 0$: the origin is a double point where two analytic branches meet.

A partial knowledge of the topology of a complex algebraic curve may be gotten by first looking at its restriction to the reals. Consider for instance the polynomial equation $P(z, y) = 0$, where

$$P(z, y) = y - 1 - zy^2,$$

which defines the OGF of the Catalan numbers. A rendering of the real part of the curve is given in Figure 9. The complex aspect of the curve as given by $\Im(y)$ as a

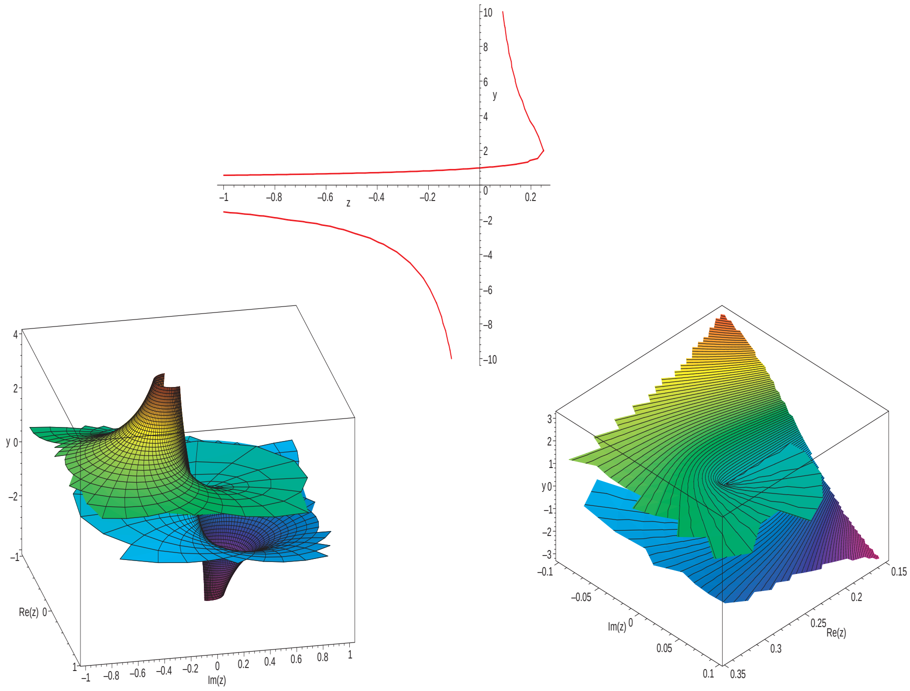


FIGURE 9. The real section of the Catalan curve (top). The complex Catalan curve with a plot of $\Im(y)$ as a function of $z = (\Re(z), \Im(z))$ (bottom left); a blowup of $\Im(y)$ near the branch point at $z = \frac{1}{4}$ (bottom right).

function of z is also displayed there. In accordance with earlier observations, there are normally two sheets (branches) above each each point. The exceptional set is given by the roots of the discriminant,

$$\mathcal{R} = z(1-4z).$$

For $z = 0$, one of the branches escapes at infinity, while for $z = 1/4$, the two branches meet and there is a branch point; see Figure 9.

In summary the exceptional set provides a set of *possible candidates* for the singularities of an algebraic function. This discussion is summarized by the slightly more general lemma that follows.

LEMMA VII.1 (Location of algebraic singularities). *Let $Y(z)$, analytic at the origin, satisfy a polynomial equation $P(z, Y) = 0$. Then, $Y(z)$ can be analytically continued along any half-line emanating from the origin that does not cross any point of the exceptional set* (38).

Nature of singularities. We start the discussion with an exceptional point that is placed at the origin (by a translation $z \mapsto z + z_0$) and assume that the equation $P(0, y) = 0$ has k equal roots $y_1, \ldots, y_k$ where $y = 0$ is this common value (by a translation $y \mapsto y + y_0$ or an inversion $y \mapsto 1/y$, if points at infinity are considered). Consider a punctured disk $|z| < r$ that does not include any other exceptional point relative to P. In the argument that follows, we let $y_1, (z), \ldots, y_k(z)$ be analytic determinations of the root that tend to 0 as $z \to 0$.

Start at at some arbitrary value interior to the real interval $(0, r)$, where the quantity $y_1(z)$ is locally an analytic function of z. By the implicit function theorem, $y_1(z)$ can be continued analytically along a circuit that starts from z and returns to z while simply encircling the origin (and staying within the punctured disk). Then, by permanence of analytic relations, $y_1(z)$ will be taken into another root, say, $y_1^{(1)}(z)$. By repeating the process, we see that after a certain number of times κ with $1 \leq \kappa \leq k$, we will have obtained a collection of roots $y_1(z) = y_1^{(0)}(z), \ldots, y_1^{(\kappa)}(z) = y_1(z)$ that form a set of κ distinct values. Such roots are said to form a *cycle*. In this case, $y_1(t^\kappa)$ is an analytic function of t except possibly at 0 where it is continuous and has value 0. Thus, by general principles (regarding removable singularities), it is in fact analytic at 0. This in turn implies the existence of a convergent expansion near 0:

$$y_1(t^\kappa) = \sum_{n=1}^{\infty} c_n t^n. \tag{39}$$

The parameter t is often called the *local uniformizing parameter*, as it reduces a multivalued function to a single value one. This translates back into the world of z: each determination of $z^{1/\kappa}$ yields one of the branches of the multivalued analytic function as

$$y_1(z) = \sum_{n=1}^{\infty} c_n z^{n/\kappa}. \tag{40}$$

Alternatively, with $\omega = e^{2i\pi/\kappa}$ a root of unity, the κ determinations are obtained as

$$y_1^{(j)} = \sum_{n=1}^{\infty} c_n \omega^n z^{n/\kappa},$$

each being valid in a sector of opening $< 2\pi$. (The case $\kappa = 1$ corresponds to an analytic branch.)

If $r = k$, then the cycle accounts for all the roots which tend to 0. Otherwise, we repeat the process with another root and, in this fashion, eventually exhaust all roots. Thus, all the k roots that have value 0 at $z = 0$ are grouped into cycles of size $\kappa_1, \ldots, \kappa_\ell$. Finally, values of y at infinity are brought to zero by means of the change of variables $y = 1/u$, then leading to negative exponents in the expansion of y.

THEOREM VII.5 (Newton–Puiseux expansions at a singularity). *Let $f(z)$ be a branch of an algebraic function $P(z, f(z)) = 0$. In a circular neighbourhood of a singularity ζ slit along a ray emanating from ζ, $f(z)$ admits a fractional series expansion (Puiseux expansion) that is locally convergent and of the form*

$$f(z) = \sum_{k \geq k_0} c_k (z - \zeta)^{k/\kappa},$$

for a fixed determination of $(z-\zeta)^{1/\kappa}$*, where* $k_0 \in \mathbb{Z}$ *and* κ *is an integer* ≥ 2*, called the "branching type".*

Newton (1643-1727) discovered the algebraic form of Theorem VII.5, published it in his famous treatise *De Methodis Serierum et Fluxionum* (completed in 1671). This method was subsequently developed by Victor Puiseux (1820–1883) so that the name of Puiseux series is customarily attached to fractional series expansions. The argument given above is taken from the neat exposition offered by Hille in [**232**, Ch. 12, vol. II]. It is known as a "monodromy argument", meaning that it consists in following the course of values of an analytic function along paths in the complex plane till it returns to its original value.

Newton polygon. Newton also described a constructive approach to the determination of branching types near a point (z_0, y_0), that by means of the previous discussion can always be taken to be $(0,0)$. In order to introduce the discussion, let us examine the Catalan generating function near $z_0 = 1/4$. Elementary algebra gives the explicit form of the two branches

$$y_1(z) = \frac{1}{2z}\left(1-\sqrt{1-4z}\right), \quad y_2(z) = \frac{1}{2z}\left(1+\sqrt{1-4z}\right),$$

whose forms are consistent with what Theorem VII.5 predicts. If however one starts directly with the equation,

$$P(z,y) \equiv y - 1 - zy^2 = 0$$

then, the translation $z = 1/4 - Z$ (the minus sign is a mere notational convenience), $y = 2 + Y$ yields

$$Q(Z,Y) \equiv -\frac{1}{4}Y^2 + 4Z + 4ZY + ZY^2. \tag{41}$$

Look for solutions of the form $Y = cZ^\alpha(1+o(1))$ with $c \neq 0$ (the existence is granted *a priori* by the Newton-Puiseux Theorem). Each of the monomials in (41) gives rise to a term of a well determined asymptotic order. respectively $Z^{2\alpha}, Z^1, Z^{\alpha+1}, Z^{2\alpha+1}$. If the equation is to be identically satisfied, then the main asymptotic order of $Q(Z,Y)$ should be 0. Since $c \neq 0$, this can only happen if two or more of the exponents in the sequence $(2\alpha, 1, \alpha+1, 2\alpha+1)$ coincide *and* the coefficients of the corresponding monomial in $P(Z,Y)$ is zero, a condition that is an algebraic constraint on the constant c. Furthermore, exponents of all the remaining monomials have to be larger since by assumption they represent terms of lower asymptotic order.

Examination of all the possible combinations of exponents leads one to discover that the only possible combination arises from the cancellation of the first two terms of Q, namely $-\frac{1}{4}Y^2 + 4Z$, which corresponds to the set of constraints

$$2\alpha = 1, \quad -\frac{1}{4}c^2 + 4 = 0,$$

with the supplementary conditions $\alpha + 1 > 1$ and $2\alpha + 1 > 1$ being satisfied by this choice $\alpha = \frac{1}{2}$. We have thus discovered that $Q(Z,Y) = 0$ is consistent asymptotically with

$$Y \sim 4Z^{1/2}, \quad Y \sim -4Z^{1/2}.$$

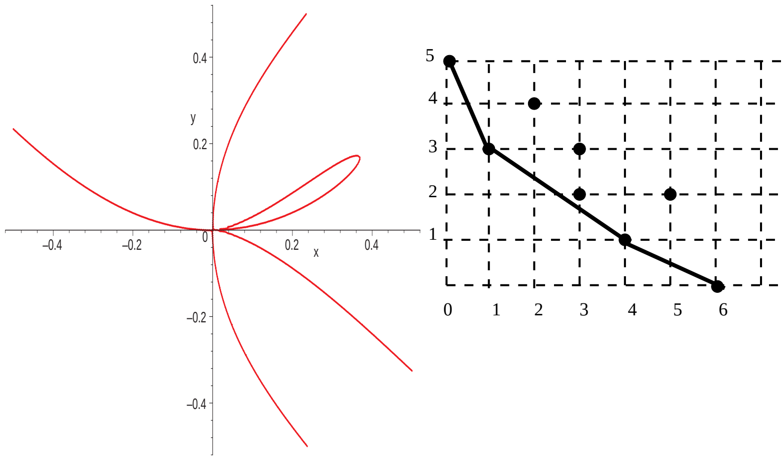


FIGURE 10. The real curve defined by the equation $P = (y - x^2)(y^2 - x)(y^2 - x^3) - x^3y^3$ near $(0, 0)$ (left) and the corresponding Newton diagram (right).

The process can be iterated upon subtracting dominant terms. It invariably gives rise to complete formal asymptotic expansions that satisfy $Q(Z, Y) = 0$ (in the Catalan example, these are series in $\pm Z^{1/2}$). Furthermore, elementary majorizations establish that such formal asymptotic solutions represent indeed convergent series. Thus, local expansions of branches have indeed been determined.

An algorithmic refinement (also due to Newton) can be superimposed on the previous discussion and is known as the method of *Newton polygons*. Consider a general polynomial

$$Q(Z, Y) = \sum_{j \in J} Z^{a_j} Y^{b_j},$$

and associate to it the finite set of points (a_j, b_j) in $\mathbb{N} \times \mathbb{N}$, which is called the Newton diagram. It is easily verified that the only asymptotic solutions of the form $Y \propto Z^t$ correspond to values of t that are inverse slopes (i.e., $\Delta x / \Delta y$) of lines connecting two or more points of the Newton diagram (this expresses the cancellation condition between two monomials of Q) *and* such that all other points of the diagram are on this line or to the right of it. In other words:

> **Newton's polygon method.** *Any possible exponents τ such that $Y \sim cZ^\tau$ is a solution to a polynomial equation corresponds to one of the inverse slopes of the leftmost convex envelope of the Newton diagram. For each viable τ, a polynomial equation constrains the possible values of the corresponding coefficient c. Complete expansions are obtained by repeating the process, which means deflating Y from its main term by way of the substitution $Y \mapsto Y - cZ^\tau$.*

Figure 10 illustrates what goes on in the case of the curve $P = 0$ where

$$\begin{array}{rcl} P(z,y) & = & (y-z^2)(y^2-z)(y^2-z^3)-z^3y^3 \\ & = & y^5-y^3z-y^4z^2+y^2z^3-2\,z^3y^3+z^4y+z^5y^2-z^6, \end{array}$$

considered near the origin. As the partly factored form suggests, we expect the curve to resemble the union of two orthogonal parabolas and of a curve $y = \pm z^{3/2}$ having a cusp, i.e., the union of

$$y = z^2, \quad y = \pm\sqrt{z}, \quad y = \pm z^{3/2},$$

respectively. It is visible on the Newton diagram of the expanded form that the possible exponents $y \propto z^t$ at the origin are the inverse slopes of the segments composing the envelope, that is,

$$\tau = 2, \quad \tau = \frac{1}{2}, \quad \tau = \frac{3}{2}.$$

For computational purposes, once determined the branching type κ, the value of k_0 that dictates where the expansion starts, and the first coefficient, the full expansion can be recovered by deflating the function from its first term and repeating the Newton diagram construction. In fact, after a few initial stages of iteration, the method of indeterminate coefficients can always be eventually applied[4]. Computer algebra systems usually have this routine included as one of the standard packages; see [**376**].

Asymptotic form of coefficients. The Newton–Puiseux theorem describes precisely the local singular structure of an algebraic function. The expansions are valid around a singularity and, in particular, they hold in indented disks of the type required in order to apply the formal translation mechanisms of singularity analysis (Chapter VI).

THEOREM VII.6 (Algebraic asymptotics). *Let $f(z) = \sum_n f_n z^n$ be an algebraic series. Assume that the branch defined by the series at the origin has a unique dominant singularity at $z = \alpha_1$ on its circle of convergence. Then, the coefficient f_n satisfies the asymptotic expansion,*

$$f_n \sim \alpha_1^{-n} \left(\sum_{k \geq k_0} d_k n^{-1-k/\kappa} \right),$$

where $k_0 \in \mathbb{Z}$ and κ is an integer ≥ 2.

If $f(z)$ has several dominant singularities $|\alpha_1| = |\alpha_2| = \cdots = |\alpha_r|$, then there exists an asymptotic decomposition (where ϵ is some small fixed number, $\epsilon > 0$)

$$f_n = \sum_{j=1}^{r} \phi^{(j)}(n) + O((|\alpha_1| + \epsilon))^{-n},$$

where each $\phi^{(j)}(n)$ admits a compleet asymptotic expansion,

$$\phi^{(j)}(n) \sim \alpha_j^{-n} \left(\sum_{k \geq k_0^{(j)}} d_k^{(j)} n^{-1-k/\kappa_j} \right),$$

[4] Bruno Salvy, private communication, August 2000

with $k_0^{(j)}$ in $\mathbb{Z}$, and κ_j an integer ≥ 2.

PROOF. The directional expansions granted by Theorem VII.5 are of the exact type required by singularity analysis (Chapter VI). Composite contours are to be used in the case of multiple singularities, in which case each $\phi^{(j)}(n)$ is the contribution obtained by transfer of a local singular element. □

In the case of multiple singularities, arithmetic cancellations may occur: consider for instance the case of

$$\frac{1}{\sqrt{1-\frac{6}{5}z+z^2}} = 1 + 0.60z + 0.04z^2 - 0.36z^3 - 0.408z^4 - \cdots ,$$

and refer to the corresponding discussion of rational coefficients. Fortunately, such delicate situations tend not to arise in combinatorial situations.

EXAMPLE 9. *Branches of unary-binary trees.* the generating function of unary binary trees is defined by $P(z, f) = 0$ where

$$P(z,y) = y - z - zy - zy^2,$$

so that

$$f(z) = \frac{1-z-\sqrt{1-2z-3z^2}}{2z} = \frac{1-z-\sqrt{(1+z)(1-3z)}}{2z}.$$

There exist only two branches: f and its conjugate $\bar{f}$ that form a 2–cycle at $\frac{1}{3}$. The singularities of all branches are at $0, -1, \frac{1}{3}$ as is apparent from the explicit form of f or from the defining equation. The branch representing $f(z)$ at the origin is analytic there (by a general argument or by the combinatorial origin of the problem). Thus, the dominant singularity of $f(z)$ is at $\frac{1}{3}$ and it is unique in its modulus class. The "easy" case of Theorem VII.7 then applies once $f(z)$ has been expanded ear $\frac{1}{3}$. As a rule, the organization of computations is simpler if one makes use of the local uniformizing parameter with a choice of sign in accordance to the direction along which the singularity is approached. In this case, we set $z = \frac{1}{3} - \delta^2$ and find

$$f(z) = 1 - 3\,\delta + \frac{9}{2}\delta^2 - \frac{63}{8}\delta^3 + \frac{27}{2}\delta^4 - \frac{2997}{128}\delta^5 + \cdots , \quad \delta = (\frac{1}{3} - z)^{1/2}.$$

This translates immediately into

$$f_n \equiv [z^n]f(z) \sim \frac{3^{n+1/2}}{2\sqrt{\pi n^3}} \left(1 - \frac{15}{16n} + \frac{505}{512n^2} - \frac{8085}{8192n^3} + \cdots\right).$$

The approximation provided by the first three terms is quite good: for $n = 10$ already, it estimates $f_{10} = 835.$ with an error less than 1. .... END OF EXAMPLE 9. □

EXAMPLE 10. *Branches of non-crossing forests.* Consider the polynomial equation $P(z, y) = 0$, where

$$P(z,y) = y^3 + (z^2 - z - 3)y^2 + (z+3)y - 1,$$

and the combinatorial GF satisfying $P(z, F) = 0$ determined by the initial conditions (*EIS* **A054727**:

$$F(z) = 1 + 2z + 7z^2 + 33z^3 + 181z^4 + 1083z^5 + \cdots .$$

(Combinatorial aspects are developed below in Section VII. 5.1, and $F_n = [z^n]F(z)$ is the number of non-crossing graphs of size n that are forests.)

The exceptional set is mechanically computed: its elements roots of the discriminant

$$R = -z^3(5z^3 - 8z^2 - 32z + 4).$$

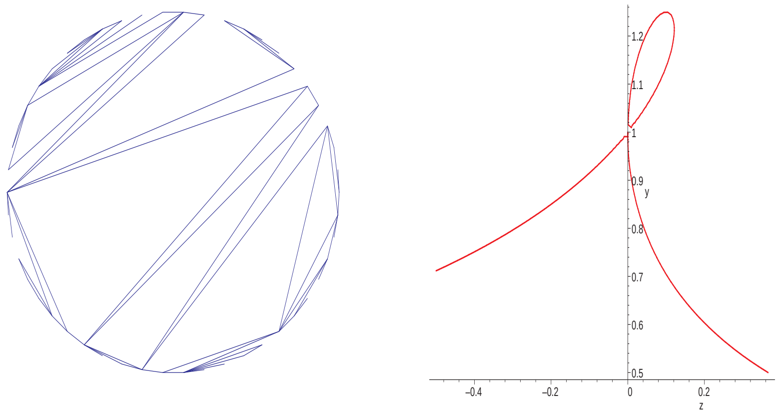


FIGURE 11. Non-crossing graphs: (a) a random connected graph of size 50; (b) the real algebraic curve corresponding to non-crossing forests.

Newton's algorithm shows that two of the branches at 0, say y_0 and y_2, form a cycle of length 2 with $y_0 = 1 - \sqrt{z} + O(z)$, $y_2 = 1 + \sqrt{z} + O(z)$ while it is the "middle branch" $y_1 = 1 + z + O(z^2)$ that corresponds to the combinatorial GF $F(z)$.

The nonzero exceptional points are the roots of the cubic factor of $\mathcal{R}$, namely

$$\Omega = \{-1.93028,\ 0.12158,\ 3.40869\}.$$

Let $\xi \doteq 0.1258$ be the root in $(0,1)$. By Pringsheim's theorem and the fact that the OGF of an infinite combinatorial class must have a positive dominant singularity in $[0,1]$, the only possibility for the dominant singularity of $y_1(z)$ is ξ. (For a more general argument, see below.)

For z near ξ, the three branches of the cubic give rise to one branch that is analytic with value approximately 0.67816 and a cycle of two conjugate branches with value near 1.21429 at $z = \xi$. The expansion of the two conjugate branches is of the singular type,

$$\alpha \pm \beta\sqrt{1 - z/\xi},$$

where

$$\alpha = \frac{43}{37} + \frac{18}{37}\xi - \frac{35}{74}\xi^2 \doteq 1.21429, \quad \beta = \frac{1}{37}\sqrt{228 - 981\xi - 5290\xi^2} \doteq 0.14931.$$

The determination with a minus sign must be adopted for representing the combinatorial GF when $z \to \xi^-$ since otherwise one would get negative asymptotic estimates for the nonnegative coefficients. Alternatively, one may examine the way the three real branches along $(0, \xi)$ match with one another at 0 and at ξ^-, then conclude accordingly.

Collecting partial results, we finally get by singularity analysis the estimate

$$F_n = \frac{\beta}{2\sqrt{\pi n^3}}\omega^n \left(1 + O(\frac{1}{n})\right), \qquad \omega = \frac{1}{\xi} \doteq 8.22469$$

where the cubic algebraic number ξ and the sextic β are as above. . END OF EXAMPLE 10. □

The example above illustrates several important points in the analysis of coefficients of algebraic functions when there are no simple explicit radical forms. First of all a given combinatorial problem determines a unique branch of an algebraic curve at the origin. Next, the dominant singularity has to be identified by "connecting" the

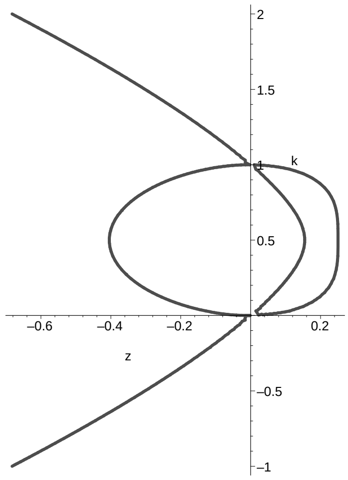


FIGURE 12. The algebraic curve associated to the generating function of supertrees of type K.

combinatorial branch with the branches at every possible singularity of the curve. Finally, computations tend to take place over algebraic numbers and not simply rational numbers.

So far, examples have illustrated the common situation where the exponent at the dominant singularity is $\frac{1}{2}$, which is reflected by a factor of $n^{-3/2}$ in the asymptotic form of coefficients. Our last example shows a case where the exponent assumes a different value, namely $\frac{1}{4}$.

EXAMPLE 11. *Branches of "supertrees"*. Consider the quartic equation

$$y^4 - 2\,y^3 + (1 + 2\,z)\,y^2 - 2\,yz + 4\,z^3 = 0$$

and let K be the branch analytic at 0 determined by the initial conditions:

$$K(z) = 2\,z^2 + 2\,z^3 + 8\,z^4 + 18\,z^5 + +64\,z^6 + 188\,z^7 + \cdots .$$

(This OGF in fact corresponds to bicoloured "supertrees" already studied in Chapter VI, Section 6.)

The discriminant is found to be

$$\mathcal{R} = 16\,z^4\left(16\,z^2 + 4\,z - 1\right)(-1 + 4\,z)^3,$$

with roots at $\frac{1}{4}$ and $(-1 \pm \sqrt{5})/8$. The dominant singularity of the branch of combinatorial interest turns out to be at $z = \frac{1}{4}$ where $K(\frac{1}{4}) = \frac{1}{2}$. The translation $z = \frac{1}{4} + Z, y = \frac{1}{2} + Y$ then transforms the basic equation into

$$4\,Y^4 + 8\,ZY^2 + 16\,Z^3 + 12\,Z^2 + Z = 0.$$

According to Newton's polygon, the main cancellation arises from $\frac{1}{4}Y^4 - Z = 0$: this corresponds to a segment of inverse slope $1/4$ in the Newton diagram and accordingly to a cycle formed with 4 conjugate branches, i.e., a fourth-root singularity. Thus, one has,

$$K(z) \underset{z\to\frac{1}{4}}{\sim} 1/2 - \frac{1}{\sqrt{2}}\left(\frac{1}{4}-z\right)^{1/4} - \frac{1}{\sqrt{2}}\left(\frac{1}{4}-z\right)^{3/4} + \cdots, \quad [z^n]K(z) \underset{n\to\infty}{\sim} \frac{4^n}{8\Gamma(\frac{3}{4})n^{5/4}},$$

which is consistent with earlier found values. Observe that we have started here with the raw algebraic equation satisfies by K. .... END OF EXAMPLE 11. □

Computable coefficient asymptotics. The previous discussion contains the germ of a complete algorithm for deriving an asymptotic expansion of coefficients of any algebraic function. We sketch here the main principles leaving some of the details to the reader. Observe that the problem is a *connection problems*: the "shapes" of the various sheets around each point (including the exceptional points) are known, but it remains to connect them together and see which ones are encountered first when starting from a given branch at the origin.

Algorithm ACA: *Algebraic Coefficient Asymptotics*.
Input: A polynomial $P(z, y)$ with $d = \deg_y P(z, y)$; a series $Y(z)$ such that $P(z, Y) = 0$ and assumed to be specified by sufficiently many initial terms so as to be distinguished from all other branches.
Output: The asymptotic expansion of $[z^n]Y(z)$ whose existence is granted by Theorem VII.6.
The algorithm consists of three main steps: *Preparation*, *Dominant singularities*, and *Translation*.

I. Preparation: Define the discriminant $R(z) = \mathbf{R}(P, P'_y, y)$.

(P_1) Compute the exceptional set $\Xi = \{z \mid R(z) = 0\}$ and the points of infinity $\Xi_0 = \{z \mid p_0(z) = 0\}$, where $p_0(z)$ is the leading coefficient of $P(z, y)$ considered as a function of y.

(P_2) Determine the Puiseux expansions of all the d branches at each of the points of $\Xi \cup \{0\}$ (by Newton diagrams and/or indeterminate coefficients). This includes the expansion of analytic branches as well. Let $\{y_{\alpha,j}(z)\}_{j=1}^{d}$ be the collection of all such expansions at some $\alpha \in \Xi \cup \{0\}$.

(P_3) Identify the branch at 0 that corresponds to $Y(z)$.

II. Dominant singularities (Controlled approximate matching of branches). Let $\Xi_1, \Xi_2, \ldots$ be a partition of the elements of $\Xi \cup \{0\}$ sorted according to the increasing values of their modulus: it is assumed that the numbering is such that if $\alpha \in \Xi_i$ and $\beta \in \Xi_j$, then $|\alpha| < |\beta|$ is equivalent to $i < j$. Geometrically, the elements of Ξ have been grouped in concentric circles. First, a preparation step is needed.

(D_1) Determine a nonzero lower bound δ on the radius of convergence of any local Puiseux expansion of any branch at any point of Ξ. Such a bound can be constructed from the minimal distance between elements of Ξ and from the degree d of the equation.

The sets Ξ_j are to be examined in sequence until it is detected that one of them contains a singularity. At step j, let $\sigma_1, \sigma_2, \ldots, \sigma_s$ be an arbitrary listing of the elements of Ξ_j. The problem is to determine whether any σ_k is a singularity and, in that event, to find the right branch to which it is associated. This part of the algorithm proceeds by controlled numerical approximations of branches and constructive bounds on the minimum separation distance between distinct branches.

(D_2) For each candidate singularity σ_k, with $k \geq 2$, set $\zeta_k = \sigma_k(1 - \delta/2)$. By assumption, each ζ_k is in the domain of convergence of $Y(z)$ and of any $y_{\sigma_k,j}$.

(D_3) Compute a nonzero lower bound η_k on the minimum distance between two roots of $P(\zeta_k, y) = 0$. This separation bound can be obtained from resultant computations.

(D_4) Estimate $Y(\zeta_k)$ and each $y_{\sigma_k,j}(\zeta_k)$ to an accuracy better than $\eta_k/4$. If two elements, $Y(z)$ and $y_{\sigma_k,j}(z)$ are (numerically) found to be at a distance less than η_k for $z = \zeta_k$, then they are matched: σ_k *is* a singularity and the corresponding $y_{\sigma_k,j}$ is the corresponding singular element. Otherwise, σ_k is declared to be a regular point for $Y(z)$ and discarded as candidate singularity.

The main loop on j is repeated until a singularity has been detected., when $j = j_0$, say. The radius of convergence ρ is then equal to the common modulus of elements of Ξ_{j_0}; the corresponding singular elements are retained.

III. Coefficient expansion. Collect the singular elements at all the points σ determined to be a dominant singularity at Phase III. Translate termwise using the singularity analysis rule,

$$(\sigma - z)^{p/\kappa} \mapsto \sigma^{p/\kappa - n} \frac{\Gamma(-p/\kappa + n)}{\Gamma(-p/\kappa)\Gamma(n+1)},$$

and reorganize into descending powers of n, if needed.

This algorithm vindicates the following assertion:

PROPOSITION VII.3 (Decidability of algebraic connections.). *The dominant singularities of a branch of an algebraic function can be determined by the algorithm* ACA *in a finite number of operations.*

VII. 4.2. Positive algebraic systems. The discussion of algebraic singularities specializes nicely to the case of positive functions. We first indicate a procedure that determines the radius of convergence of any algebraic series with *positive* coefficients. The procedure takes advantage of Pringsheim's theorem that allows us to restrict attention to candidate singularities on the positive half-line. It represents a shortcut that is often suitable for human calculation and, in fact, the it systematizes some of the techniques already used implicitly in earlier examples.

Algorithm ROCPAF: *Radius of Convergence of Positive Algebraic Functions.*
Input: A polynomial $P(z, y)$ with $d = \deg_y P(z, y)$; a series $Y(z)$ such that $P(z, Y) = 0$ that is known to have only nonnegative coefficients ($[z^n]Y(z) \geq 0$) and is assumed to be specified by sufficiently many initial terms.
Output: The radius of convergence ρ of $Y(z)$.

Plane-sweep. Let Ξ^+ be the subset of those elements of the exceptional set Ξ which are positive real.

(R_1) Sort the subset of those branches $\{y_{0,j}\}$ at 0^+ that have totally real coefficients. This is essentially a lexicographic sort that only needs the initial parts of each expansion. Set initially $\xi_0 = 0$ and $U(z) = Y(z)$.

(R_2) Sweep over all $\xi \in \Xi^+$ in increasing order. To detect whether a candidate ξ is the dominant positive singularity, proceed as follows:

— Sort the branches $\{y_{\xi,j}\}$ at ξ^- that have totally real coefficients.

— using the orders at ξ_0^+ and ξ^-, match the branch $U(z)$ with its corresponding branch at ξ^-, say $V(z)$; this makes use of the total ordering between real branches at ξ_0^+ and ξ^-. If the branch $V(z)$ is singular, then return $\rho = \xi$ as the radius of convergence of $Y(z)$ and use $V(z)$ as the singular element of $Y(z)$ at $z = \rho$; otherwise continue with the next value of $\xi \in \Xi^+$ while replacing $U(z)$ by $V(z)$ and ξ_0 by ξ.

This algorithm is a plane-sweep that takes advantage of the fact that the real branches near a point can be totally ordered; finding the ordering only requires inspection of a finite number of coefficients. The plane-sweep algorithm enables us to trace at each stage the original branch and keep a record of its order amongst all branches. The method works since no two real branches can cross at a point other than a multiple point, such a point being covered as an element of Ξ^+.

We now turn to positive systems. Most of the combinatorial classes known to admit algebraic generating functions involve singular exponents that are multiples of $\frac{1}{2}$. This empirical observation is supported by the fact, to be proved below, that a wide class of positive systems have solutions with a square-root singularity. Interestingly enough, the corresponding theorem is due to independent research by several authors: Drmota [**111**] developed a version of the theorem in the course of studies relative to limit laws in various families of trees defined by context-free grammars; Woods [**440**], motivated by questions of Boolean complexity and finite model theory, gave a form expressed in terms of colouring rules for trees; finally, Lalley [**285**] came across a similarly general result when quantifying return probabilities for random walks on groups. The statement that follows is a fundamental result in the analysis of algebraic systems arising from combinatorics and is (rightly) called the "Drmota-Lalley-Woods" theorem. Notice that the authors of [**111, 285, 440**] prove more: Drmota and Lalley show how to pull out limit Gaussian laws for simple parameters (e.g., as in [**111**] by a perturbative analysis; see Chapter IX); Woods shows how to deduce estimates of coefficients even in some periodic or non-irreducible cases (see definitions below).

In the treatment that follows we start from a polynomial system of equations,

$$\{y_j = \Phi_j(z, y_1, \ldots, y_m)\}, \quad j = 1, \ldots, m.$$

We shall discuss in the next section a class of combinatorial specifications, the "context-free" specifications, that leads systematically to such fixed-point systems. The case of linear systems has been already dealt with, so that we limit ourselves here to *nonlinear systems* defined by the fact that at least one polynomial Φ_j is nonlinear in some of the indeterminates $y_1, \ldots, y_m$.

First, for combinatorial reasons, we define several possible attributes of a polynomial system.

— *Algebraic positivity* (or a-positivity). A polynomial system is said to be *a-positive* if all the component polynomials Φ_j have nonnegative coefficients.

Next, we want to restrict consideration to systems that determine a unique solution vector $(y_1, \ldots, y_m) \in (\mathbb{C}[\![z]\!])^m$. (This discussion is related to 0-dimensionality in the sense alluded to earlier.) Define the z-*valuation* $\mathrm{val}(\vec{y})$ of a vector $\vec{y} \in \mathbb{C}[\![z]\!]^m$ as the minimum over all j's of the individual valuations[5] $\mathrm{val}(y_j)$. The distance between two vectors is defined as usual by $d(\vec{y}, \vec{y}') = 2^{-\mathrm{val}(\vec{y}-\vec{y}')}$. Then, one has:

— *Algebraic properness* (or a-properness). A polynomial system is said to be *a-proper* if it satisfies a Lipschitz condition

$$d(\Phi(\vec{y}), \Phi(\vec{y}\,')) < K d(\vec{y}, \vec{y}\,') \quad \text{for some } K < 1.$$

In that case, the transformation Φ is a contraction on the complete metric space of formal power series and, by the general fixed point theorem, the equation $y = \Phi(y)$ admits a unique solution. In passing, this solution may be obtained by the iterative scheme,

$$\vec{y}^{(0)} = (0, \ldots, 0)^t, \quad \vec{y}^{(h+1)} = \Phi(y^{(h)}), \quad y = \lim_{h\to\infty} y^{(h)}.$$

The key notion is irreducibility. To a polynomial system, $\vec{y} = \Phi(\vec{y})$, associate its *dependency graph* defined as a graph whose vertices are the numbers $1, \ldots, m$ and the edges ending at a vertex j are $k \to j$, if y_j figures in a monomial of $\Phi_k(j)$. (This notion is reminiscent of the one already introduced for linear system on page **??**.)

— *Algebraic irreducibility* (or a-irreducibility). A polynomial system is said to be *a-irreducible* if its dependency graph is strongly connected.

Finally, one needs a technical notion of periodicity to dispose of cases like

$$y(z) = \frac{1}{2z}\left(1 - \sqrt{1-4z}\right) = z + z^3 + 2z^5 + \cdots,$$

(the OGF of complete binary trees) where coefficients are only nonzero for certain residue classes of their index.

— *Algebraic aperiodicity* (or a-aperiodicity). A power series is said to be *aperiodic* if it contains three monomials (with nonzero coefficients), $z^{e_1}, z^{e_2}, z^{e_3}$, such that $e_2 - e_1$ and $e_3 - e_1$ are relatively prime. A proper polynomial system is said to be aperiodic if each of its component solutions y_j is aperiodic.

THEOREM VII.7 (Positive polynomial systems). *Consider a nonlinear polynomial system $\vec{y} = \Phi(\vec{y})$ that is a-proper, a-positive, and a-irreducible. In that case, all component solutions y_j have the same radius of convergence $\rho < \infty$. Then, there exist functions h_j analytic at the origin such that*

$$y_j = h_j\left(\sqrt{1 - z/\rho}\right) \qquad (z \to \rho^-). \tag{42}$$

In addition, all other dominant singularities are of the form $\rho\omega$ with ω a root of unity. If furthermore the system is a-aperiodic, all y_j have ρ as unique dominant singularity. In that case, the coefficients admit a complete asymptotic expansion of the form

$$[z^n] y_j(z) \sim \rho^{-n}\left(\sum_{k\geq 1} d_k n^{-1-k/2}\right). \tag{43}$$

[5]Let $f = \sum_{n=\beta}^{\infty} f_n z^n$ with $f_\beta \neq 0$; the valuation of f is by definition $\mathrm{val}(f) = \beta$.

PROOF. The proof consists in gathering by stages consequences of the assumptions. It is essentially based on close examination of "failures" of the implicit function theorem and the way these lead to singularities.

(a) As a preliminary observation, we note that each component solution y_j is an algebraic function that has a nonzero radius of convergence. In particular, singularities are constrained to be of the algebraic type with local expansions in accordance with the Newton-Puiseux theorem (Theorem VII.5).

(b) Properness together with the positivity of the system implies that each $y_j(z)$ has nonnegative coefficients in its expansion at 0, since it is a formal limit of approximants that have nonnegative coefficients. In particular, each power series y_j has a certain nonzero radius of convergence ρ_j. Also, by positivity, ρ_j is a singularity of y_j (by virtue of Pringsheim's theorem). From the nature of singularities of algebraic functions, there exists some order $R \geq 0$ such that each Rth derivative $\partial_z^R y_j(z)$ becomes infinite as $z \to \rho_j^-$.

We establish now that $\rho_1 = \cdots = \rho_m$. In effect, differentiation of the equations composing the system implies that a derivative of arbitrary order r, $\partial_z^r y_j(z)$, is a linear form in other derivatives $\partial_z^r y_j(z)$ of the same order (and a polynomial form in lower order derivatives); also the linear combination and the polynomial form have nonnegative coefficients. Assume a contrario that the radii were not all equal, say $\rho_1 = \cdots = \rho_s$, with the other radii $\rho_{s+1}, \ldots$ being strictly greater. Consider the system differentiated a sufficiently large number of times, R. Then, as $z \to \rho_1$, we must have $\partial_z^R y_j$ tending to infinity for $j \leq s$. On the other hand, the quantities y_{s+1}, etc., being analytic, their Rth derivatives that are analytic as well must tend to finite limits. In other words, because of the irreducibility assumption (and again positivity), infinity *has to* propagate and we have reached a contradiction. Thus, all the y_j have the same radius of convergence and we let ρ denote this common value.

(c_1) The key step consists in establishing the existence of a square-root singularity at the common singularity ρ. Consider first the scalar case, that is

$$y - \phi(z, y) = 0, \tag{44}$$

where ϕ is assumed to depend nonlinearly on y and have nonnegative coefficients. The requirement of properness means that z is a factor of all monomials, except the constant term $\phi(0, 0)$.

Let $y(z)$ be the unique branch of the algebraic function that is analytic at 0. Comparison of the asymptotic orders in y inside the equality $y = \phi(z, y)$ shows that (by nonlinearity) we cannot have $y \to \infty$ when z tends to a finite limit. Let now ρ be the radius of convergence of $y(z)$. This argument shows that $y(z)$ is necessarily finite at its singularity ρ. We set $\tau = y(\rho)$ and note that, by continuity $\tau - \phi(\rho, \tau) = 0$.

By the implicit function theorem, a solution (z_0, y_0) of (44) can be continued analytically as $(z, y_0(z))$ in the vicinity of z_0 as long as the derivative with respect to y,

$$J(z_0, y_0) := 1 - \phi'_y(z_0, y_0)$$

remains nonzero. The quantity ρ being a singularity, we must have $J(\rho, \tau) = 0$. (In passing, the system

$$\tau - \phi(\rho, \tau) = 0, \qquad J(\rho, \tau) = 0,$$

determines only finitely many candidates for ρ.) On the other hand, the second derivative $-\phi''_{yy}$ is nonzero at (ρ, τ) (by positivity, since no cancellation can occur); there results by the classical argument on local failures of the implicit function theorem that $y(z)$ has a singularity of the square-root type (see also Chapters IV and VI). More precisely, the local expansion of the defining equation (44) at (ρ, τ) binds (z, y) locally by

$$-(z-\rho)\phi'_z(\rho, \tau) - \frac{1}{2}(y-\tau)^2\phi''_{yy}(\rho, \tau) + \cdots = 0,$$

where the subsequent terms are negligible by Newton's polygon method. Thus, we have

$$y - \tau = -\sqrt{\frac{\phi_z(\rho, \tau)}{\phi''_{yy}(\rho, \tau)}}(\rho - z)^{1/2} + \cdots ,$$

the negative determination of the square-root being chosen to comply with the fact that $y(z)$ increases as $z \to \rho^-$. This proves the first part of the assertion in the scalar case.

(c_2) In the multivariate case, we graft an ingenious argument [**285**] that is based on a linearized version of the system to which Perron-Frobenius theory is applicable. First, irreducibility implies that any component solution y_j depends nonlinearly on itself (by possibly iterating Φ), so that a discrepancy in asymptotic behaviours would result for the implicitly defined y_j in the event that some y_j tends to infinity.

Now, the multivariate version of the implicit function theorem grants locally the analytic continuation of any solution $y_1, y_2, \ldots, y_m$ at z_0 provided there is no vanishing of the Jacobian determinant

$$J(z_0, y_1, \ldots, y_m) := \det\left(\delta_{i,j} - \frac{\partial}{\partial y_j}\Phi_i(z_0, y_1, \ldots, y_m)\right),$$

where $\delta_{i,j}$ is Kronecker's symbol. Thus, we must have

$$J(\rho, \tau_1, \ldots, \tau_m) = 0 \qquad \text{where} \quad \tau_j := y_j(\rho).$$

The next argument (we follow Lalley [**285**]) uses Perron-Frobenius theory and linear algebra. Consider the modified Jacobian matrix

$$K(z_0, y_1, \ldots, y_m) := \left(\frac{\partial}{\partial y_j}\Phi_i(z_0, y_1, \ldots, y_m)\right),$$

which represents the "linear part" of Φ. For $z, y_1, \ldots, y_m$ all nonnegative, the matrix K has positive entries (by positivity of Φ) so that it is amenable to Perron-Frobenius theory. In particular it has a positive eigenvalue $\lambda(z, y_1, \ldots, y_m)$ that dominates all the other in modulus. The quantity

$$\widehat{\lambda}(z) = \lambda(y_1(z), \ldots, y_m(z))$$

is increasing as it is an increasing function of the matrix entries that themselves increase with z for $z \geq 0$.

We propose to prove that $\widehat{\lambda}(\rho) = 1$, In effect, $\widehat{\lambda}(\rho) < 1$ is excluded since otherwise $(I - K)$ would be invertible at $z = \rho$ and this would imply $J \neq 0$, thereby contradicting the singular character of the $y_j(z)$ at ρ. Assume *a contrario* $\widehat{\lambda}(\rho) > 1$ in order to exclude the other case. Then, by the increasing property, there would exists $\rho_1 < \rho$ such that $\widehat{\lambda}(\rho_1) = 1$. Let v_1 be a left eigenvector of $K(\rho_1, y_1(\rho_1), \ldots, y_m(\rho_1))$ corresponding to the eigenvalue $\widehat{\lambda}(\rho_1)$. Perron-Frobenius theory grants that such a vector v_1 has all its coefficients that are positive. Then, upon multiplying on the left by v_1 the column vectors corresponding to y and $\Phi(y)$ (which are equal), one gets an identity; this derived identity upon expanding near ρ_1 gives

$$A(z - \rho_1) = -\sum_{i,j} B_{i,j}(y_i(z) - y_i(\rho_1))(y_j(z) - y_j(\rho_1)) + \cdots, \tag{45}$$

where $\cdots$ hides lower order terms and the coefficients $A, B_{i,j}$ are nonnegative with $A > 0$. There is a contradiction in the orders of growth if each y_i is assumed to be analytic at ρ_1 since the left side of (45) is of exact order $(z - \rho_1)$ while the right side is at least as small as to $(z - \rho_1)^2$. Thus, we must have $\widehat{\lambda}(\rho) = 1$ and $\widehat{\lambda}(x) < 1$ for $x \in (0, \rho)$.

A calculation similar to (45) but with ρ_1 replaced by ρ shows finally that, if

$$y_i(z) - y_i(\rho) \sim \gamma_i(\rho - z)^\alpha,$$

then consistency of asymptotic expansions implies $2\alpha = 1$, that is $\alpha = \frac{1}{2}$. (The argument here is similar to the first stage of a Newton polygon construction.) We have thus proved that the component solutions $y_j(z)$ have a square-root singularity. (The existence of a complete expansion in powers of $(\rho - z)^{1/2}$ results from examination of the Newton diagram.) The proof of the general case (42) is at last completed.

(d) In the aperiodic case, we first observe that each $y_j(z)$ cannot assume an infinite value on its circle of convergence $|z| = \rho$, since this would contradict the boundedness of $|y_j(z)|$ in the open disk $|z| < \rho$ (where $y_j(\rho)$ serves as an upperbound). Consequently, by singularity analysis, the Taylor coefficients of any $y_j(z)$ are $O(n^{-1-\eta})$ for some $\eta > 1$ and the series representing y_j at the origin converges on $|z| = \rho$.

For the rest of the argument, we observe that if $y = \Phi(z, \vec{y})$, then $y = \Phi^{\langle m \rangle}(z, \vec{y})$ where the superscript denotes iteration of the transformation Φ in the variables $\vec{y} = (y_1, \ldots, y_m)$. By irreducibility, $\Phi^{\langle m \rangle}$ is such that *each* of its component polynomials involves *all* the variables.

Assume that there would exists a singularity ρ^* of some $y_j(z)$ on $|z| = \rho$. The triangle inequality yields $|y_j(\rho^*)| < y_j(\rho)$ where strictness is related to the general aperiodicity argument encountered at several other places in this book. But then, the modified Jacobian matrix $K^{\langle m \rangle}$ of $\Phi^{\langle m \rangle}$ taken at the $y_j(\rho^*)$ has entries dominated strictly by the entries of $K^{\langle m \rangle}$ taken at the $y_j(\rho)$. There results that the dominant eigenvalue of $K^{\langle m \rangle}(z, \vec{y}_j(\rho^*))$ must be strictly less than 1. But this would imply that $I - K^{\langle m \rangle}(z, \vec{y}_j(\rho^*))$ is intervertible so that the $y_j(z)$ would be analytic at ρ^*. A contradiction has been reached: ρ is the sole dominant singularity of each y_j and this concludes the argument. $\square$

We observe that the dominant singularity is obtained amongst the positive solutions of the system

$$\vec{\tau} = \Phi(\rho, \vec{\tau}), \quad J(\rho, \vec{\tau}) = 0.$$

For the Catalan GF, this yields for instance

$$\tau - 1 - \rho\tau^2 = 0, \quad 1 - 2\rho\tau = 0,$$

giving back as expected: $\rho = \frac{1}{4}$, $\tau = \frac{1}{2}$.

VII. 5. Combinatorial applications of algebraic functions

In this section, we first present context-free specifications that admit a direct translation into polynomial systems (Section VII. 5.1). When particularized to formal languages, this gives rise to context-free languages that, provided an unambiguity condition is met, lead to algebraic generating functions.

The next two subsections introduce objects whose constructions still lead to algebraic functions, but in a non-obvious way. This includes: walks with a finite number of allowed basic jumps (Section VII. 5.2) and planar maps (Section VII. 5.3). In that case, bivariate functional equations are induced by the combinatorial decompositions. The common form is

$$\Phi(z, u, F(z,u), h_1(z), \ldots, h_r(z)) = 0, \tag{46}$$

where Φ is a known polynomial and the unknowns are F and $h_1, \ldots, h_r$. Specific methods are needed in order to attain solutions to such functional equations that would seem at first glance to be grossly underdetermined. Random walks lead to a linear version of (46) that is treated by the so-called "kernel method". Maps lead to nonlinear versions that are solved by means of Tutte's "quadratic method". In both cases, the strategy consists in binding z and u by forcing them to lie on an algebraic curve (suitably chosen in order to eliminate the dependency on $F(z,u)$), and then pulling out the algebraic consequences of such a specialization.

VII. 5.1. Context-free specifications and languages. A *context-free system* is a collection of combinatorial equations,

$$\left\{ \begin{array}{lcl} \mathcal{C}_1 & = & \Phi_1(\vec{a}, \mathcal{C}_1, \ldots, \mathcal{C}_m) \\ \vdots & & \vdots \\ \mathcal{C}_m & = & \Phi_m(\vec{a}, \mathcal{C}_1, \ldots, \mathcal{C}_m), \end{array} \right. \tag{47}$$

where $\vec{a} = (a_1, \ldots)$ is a vector of atoms and each of the Φ_j only involves the combinatorial constructions of disjoint union and cartesian product. A combinatorial class $\mathcal{C}$ is said to be context-free if it is definable as the first component ($\mathcal{C} = \mathcal{C}_1$) of a well-founded context-free system. The terminology comes from linguistics and it stresses the fact that objects can be "freely" generated by the rules in (47), this without any constraints imposed by an outside context[6].

[6]Formal language theory also defines context-sensitive grammars where each rule (called a production) is applied only if it is enabled by some external context. Context-sensitive grammars have greater expressive power than context-free ones, but they depart significantly from decomposability since they are

For instance the class of plane binary trees defined by

$$\mathcal{B} = e + (i \times \mathcal{B} \times \mathcal{B}) \qquad (e, i \text{ atoms})$$

is a context-free class. The class of general plane trees defined by

$$\mathcal{G} = o \times \mathrm{SEQ}(\mathcal{G}) \qquad (o \text{ an atom})$$

is definable by the system

$$\mathcal{G} = o \times \mathcal{F}, \quad \mathcal{F} = \mathbf{1} + (\mathcal{F} \times \mathcal{G}),$$

with $\mathcal{F}$ defining forests, and so it is also context-free. (This example shows more generally that sequences can always be reduced to polynomial form.)

Context-free specifications may be used to describe all sorts of combinatorial objects. For instance, the class $\mathcal{T}$ of triangulations of convex polygons is specified symbolically by

$$\mathcal{T} = \nabla + (\nabla \times \mathcal{T}) + (\mathcal{T} \times \nabla) + (\mathcal{T} \times \nabla \times \mathcal{T}), \tag{48}$$

where ∇ represents a generic triangle.

The general symbolic rules given in Chapter I apply in all such cases. Therefore the Drmota-Lalley-Woods theorem (Theorem VII.7) provides the asymptotic solution to an important category of problems.

THEOREM VII.8 (Context-free specifications). *A context-free class $\mathcal{C}$ admits an OGF that satisfies a polynomial system obtained from the specification by the translation rules:*

$$\mathcal{A} + \mathcal{B} \mapsto A + B, \quad \mathcal{A} \times \mathcal{B} \mapsto A \cdot B.$$

The OGF $C(z)$ is an algebraic function to which algebraic asymptotics applies. In particular, a context-free class $\mathcal{C}$ that gives rise to an algebraically aperiodic irreducible system has an enumeration sequence satisfying

$$C_n \sim \frac{\gamma}{\sqrt{\pi n^3}} \omega^n,$$

where γ, ω are computable algebraic numbers.

This last result explains the frequently encountered estimates involving a factor of $n^{-3/2}$ (corresponding to a square-root singularity of the OGF) that can be found throughout analytic combinatorics.

$\triangleright$ **20.** *Extended context-free specifications.* If $\mathcal{A}, \mathcal{B}$ are context-free specifications then: (i) the sequence class $\mathcal{C} = \mathrm{SEQ}(\mathcal{A})$ is context-free; (ii) the substitution class $\mathcal{D} = \mathcal{A}[b \mapsto \mathcal{B}]$ is context-free. $\triangleleft$

We detail below an example from combinatorial geometry.

EXAMPLE 12. *Planar non-crossing configurations.* The enumeration of non-crossing planar configurations is discussed here at some level of generality. (An analytic problem in this orbit has been already treated in Example 10.) The purpose is to illustrate the fact that context-free descriptions can model naturally very diverse sorts of objects including particular topological-geometric configurations. The problems considered have their origin in combinatorial musings

surrounded by strong undecidability properties; accordingly context-sensitive grammars cannot be associated with any global generating function formalism.

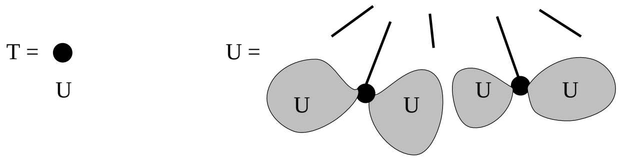
of the Rev. T.P. Kirkman in 1857 and were revisited in 1974 by Domb and Barett [**108**] for the purpose of investigating certain perturbative expansions of statistical physics. Our presentation follows closely the synthesis offered by Flajolet and Noy in [**164**].

Consider for each value of n the regular n-gon built from vertices taken for convenience to be the n complex roots of unity and numbered $0, \ldots, n-1$. A non-crossing graph is a graph on this set of vertices such that no two of its edges cross. From there, one defines non-crossing connected graphs, non-crossing forests (that are acyclic), and non-crossing trees (that are acyclic and connected); see Figure 13. Note that there is a well-defined orientation of the complex plane and also that the various graphs considered can always be rooted in some canonical way (e.g., on the vertex of smallest index) since the placement of vertices is rigidly fixed.

Trees. A non-crossing tree is rooted at 0. To the root vertex, is attached an ordered collection of vertices, each of which has an end-node ν that is the common root of two non-crossing trees, one on the left of the edge $(0, \nu)$ the other on the right of $(0, \nu)$. Let $\mathcal{T}$ denote the class of trees and $\mathcal{U}$ denote the class of trees whose root has been severed. With o denoting a generic node, we then have

$$\mathcal{T} = o \times \mathcal{U}, \quad \mathcal{U} = \mathrm{SEQ}(\mathcal{U} \times o \times \mathcal{U}),$$

which corresponds graphically to the "butterfly decomposition":



In terms of OGF, this gives the system

$$\{T = zU,\ U = (1 - zU^2)^{-1}\} \quad \Longleftrightarrow \quad \{T = zU,\ U = 1 + UV,\ V = zU^2\}, \tag{49}$$

where the latter form corresponds to the expansion of the sequence operator. Consequently, T satisfies $T = T^3 - zT + z^2$, which by Lagrange inversion gives $T_n = \frac{1}{2n-1}\binom{3n-3}{n-1}$.

Forests. A (non-crossing) forest is a non-crossing graph that is acyclic. In the present context, it is not possible to express forests simply as sequences as trees, because of the geometry of the problem.

Starting conventionally from the root vertex 0 and following all connected edges defines a "backbone" tree. To the left of every vertex of the tree, a forest may be placed. There results the decomposition (expressed directly in terms of OGF's),

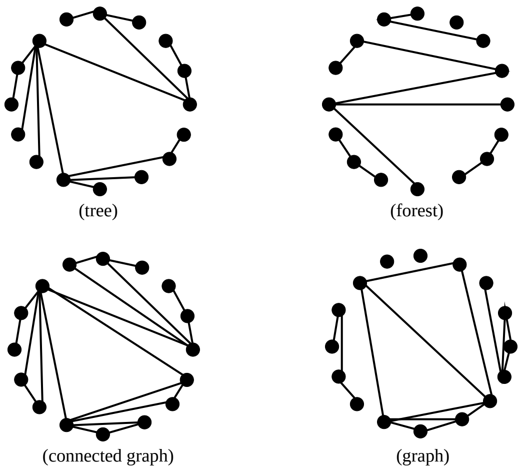
$$F = 1 + T[z \mapsto zF], \tag{50}$$

where T is the OGF of trees and F is the OGF of forests. In (50), the term $T[z \mapsto zF]$ denotes a functional composition. A context-free specification in standard form results mechanically from (49) upon replacing z by zF, namely

$$F = 1 + T,\ T = zFU,\ U = 1 + UV,\ V = zFU^2. \tag{51}$$

This system is irreducible and aperiodic, so that the asymptotic shape of F_n is of the form $\gamma\omega_n n^{-3/2}$, as predicted by Theorem VII.8. This agrees with the precise formula determined in Example 10.

Graphs. Similar constructions (see [**164**]) give the OGF's of connected graphs and general graphs. The results are summarized in Figure 13. Note the common shape of the asymptotic estimates and also the fact that simple binomial terms or sums are available in each case. END OF EXAMPLE 12. □



Configuration / OGF	Coefficients (exact / asymptotic)
Trees (*EIS*: **A001764**)	$z + z^2 + 3z^3 + 12z^4 + 55z^5 + \cdots$
$T^3 - zT + z^2 = 0$	$\frac{1}{2n-1}\binom{3n-3}{n-1}$
	$\sim \frac{\sqrt{3}}{27\sqrt{\pi n^3}}(\frac{27}{4})^n$
Forests (*EIS*: **A054727**)	$1 + z + 2z^2 + 7z^3 + 33z^4 + 181z^5 \cdots$
$F^3 + (z^2 - z - 3)F^2 + (z+3)F - 1 = 0$	$\sum_{j=1}^{n} \frac{1}{2n-j}\binom{n}{j-1}\binom{3n-2j-1}{n-j}$
	$\sim \frac{0.07465}{\sqrt{\pi n^3}}\,(8.22469)^n$
Connected graphs (*EIS*: **A007297**)	$z + z^2 + 4z^3 + 23z^4 + 156z^5 + \cdots$
$C^3 + C^2 - 3zC + 2z^2 = 0$	$\frac{1}{n-1}\sum_{j=n-1}^{2n-3}\binom{3n-3}{n+j}\binom{j-1}{j-n+1}$
	$\sim \frac{2\sqrt{6}-3\sqrt{2}}{18\sqrt{\pi n^3}}\left(6\sqrt{3}\right)^n$
Graphs (*EIS*: **A054726**)	$1 + z + 2z^2 + 8z^3 + 48z^4 + 352z^5 + \cdots$
$G^2 + (2z^2 - 3z - 2)G + 3z + 1 = 0$	$\frac{1}{n}\sum_{j=0}^{n-1}(-1)^j\binom{n}{j}\binom{2n-2-j}{n-1-j}2^{n-1-j}$
	$\sim \frac{\sqrt{140-99\sqrt{2}}}{4\sqrt{\pi n^3}}\left(6+4\sqrt{2}\right)^n$

FIGURE 13. (Top) Non-crossing graphs: a tree, a forest, a connected graph, and a general graph. (Bottom) The enumeration of non-crossing configurations by algebraic functions.

Note on "tree-like" structures. A context-free specification can always be regarded as defining a class of trees. Indeed, if the jth term in the construction Φ_j is "coloured" with the pair (i,j), it is seen that a context-free system yields a class of trees whose nodes are tagged by pairs (i,j) in a way that is consistent with the system's rules (47). However, despite this correspondence, it is often convenient to preserve the possibility of operating directly with objects[7] when the tree aspect is unnatural. By a terminology borrowed from the theory of syntax analysis in computer science, such trees are referred to as "parse trees" or "syntax trees".

Let $\mathcal{A}$ be a fixed finite alphabet whose elements are called letters. A *grammar* G is a collection of equations

$$G \quad : \quad \begin{cases} \mathcal{L}_1 & = \quad \Psi_1(\vec{a}, \mathcal{L}_1, \ldots, \mathcal{L}_m) \\ \vdots & \qquad \vdots \\ \mathcal{L}_m & = \quad \Psi_m(\vec{a}, \mathcal{L}_1, \ldots, \mathcal{L}_m), \end{cases} \tag{52}$$

where each Ψ_j involves only the operations of union ($\cup$) and catenation product ($\cdot$) with $\vec{a}$ the vector of letters in $\mathcal{A}$. For instance,

$$\Psi_1(\vec{a}, \mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3) = a_2 \cdot \mathcal{L}_2 \cdot \mathcal{L}_3 \cup a_3 \cup \mathcal{L}_3 \cdot a_2 \cdot \mathcal{L}_1.$$

A solution to (52) is an m-tuple of languages over the alphabet $\mathcal{A}$ that satisfies the system. By convention, one declares that the grammar G defines the first component, $\mathcal{L}_1$.

To each grammar (52), one can associate a context-free specification (47) by transforming unions into disjoint union, '$\cup$' $\mapsto$ '$+$', and catenation into cartesian products, '$\cdot$' $\mapsto$ '$\times$'. Let $\widehat{G}$ be the specification associated in this way to the grammar G. The objects described by $\widehat{G}$ appear in this perspective to be trees (see the discussion above regarding parse trees). Let h be the transformation from trees of $\widehat{G}$ to languages of G that lists letters in infix (i.e., left-to-right) order: we call such an h the erasing transformation since it "forgets" all the structural information contained in the parse tree and only preserves the succession of letters. Clearly, application of h to the combinatorial specifications determined by $\widehat{G}$ yields languages that obey the grammar G. For a grammar G and a word $w \in \mathcal{A}^\star$, the number of parse trees $t \in \widehat{G}$ such that $h(t) = w$ is called the *ambiguity coefficient* of w with respect to the grammar G; this quantity is denoted by $\kappa_G(w)$.

A grammar G is unambiguous if all the corresponding ambiguity coefficients are either 0 or 1. This means that there is a bijection between parse trees of $\widehat{G}$ and words of the language described by G: each word generated is uniquely "parsable" according to the grammar. From Theorem VII.8, we have immediately:

PROPOSITION VII.4 (Context-free languages). *Given a context-free grammar G, the ordinary generating function of the language $L_G(z)$, counting words* with *multiplicity, is an algebraic function. In particular, a context-free language that admits an unambiguous grammar specification has an ordinary generating function $L(z)$ that is an algebraic function.*

[7] Some authors have even developed a notion of "object grammars"; see for instance [**119**] itself inspired by techniques of polyomino surgery in [**98**].

This theorem originates from early works of Chomsky and Schützenberger [**80**] which have exerted a strong influence on the philosophy of the present book.

For example consider the Łukasiewicz language

$$\mathcal{L} = (a \cdot \mathcal{L} \cdot \mathcal{L} \cdot \mathcal{L}) \cup b.$$

This can be interpreted as the set of functional terms built from the ternary symbol a and the nullary symbol b:

$$\begin{array}{rcl}\mathcal{L} & = & \{b,\ abbb,\ aabbbbb,\ ababbbb, \ldots\} \\ & \simeq & \{b,\ a(b,b,b),\ a(a(b,b,b),b,b),\ a(b,a(b,b,b),b), \ldots\},\end{array}$$

where $\simeq$ denotes combinatorial isomorphism. It is easily seen that the terms are in bijective correspondence with their parse trees, themselves isomorphic to ternary trees. Thus the grammar is unambiguous, so that the OGF equation translates directly from the grammar,

$$L(z) = zL(z)^3 + z. \tag{53}$$

As another example, we revisit Dyck paths that are definable by the grammar,

$$D = 1 \cup (a \cdot D \cdot b \cdot D)\,, \tag{54}$$

where a denotes ascents and b denotes descents. Each word in the language must start with a letter a that has a unique matching letter b and thus it is uniquely parsable according to the grammar (54). Since the grammar is unambiguous, the OGF reads off:

$$D(z) = z + z^2 D(z)^2.$$

VII. 5.2. Walks and the kernel method. Start with a set Ω that is a finite subset of $\mathbb{Z}$ and is called the set of *jumps*. A *walk* (relative to Ω) is a sequence $w = (w_0, w_1, \ldots, w_n)$ such that $w_0 = 0$ and $w_{i+1} - w_i \in \Omega$, for all i, $0 \le i < n$. A *nonnegative walk* (also known as a "meander") satisfies $w_i \ge 0$ and an *excursion* is a nonnegative walk such that, additionally, $w_n = 0$. The quantity n is called the length of the walk or the excursion. For instance, Dyck paths and Motzkin paths analysed in Section V. 3 are excursions that correspond to $\Omega = \{-1, +1\}$ and $\Omega = \{-1, 0, +1\}$ respectively. (Walks and excursions can be viewed as particular cases of paths in a graph in the sense of Section V. 6, with the graph taken to be the infinite set $\mathbb{Z}_{>0}$ of integers.)

We propose to determine f_n, the number of excursions of length n and type Ω, via the corresponding OGF

$$F(z) = \sum_{n=0}^{\infty} f_n z^n.$$

In fact, we shall determine the more general BGF

$$F(z,u) := \sum_{n,k} f_{n,k} u^k z^n,$$

where $f_{n,k}$ is the number of walks of length n and final altitude k (*i.e.*, the value of w_n in the definition of a walk is constrained to equal k). In particular, one has $F(z) = F(z,0)$.

We let $-c$ denote the smallest (negative) value of a jump, and d denote the largest (positive) jump. A fundamental rôle is played in this discussion by the “characteristic polynomial” of the walk,

$$S(y) := \sum_{\omega \in \Omega} y^{\omega} = \sum_{j=-c}^{d} S_j y^j$$

that is a Laurent polynomial[8]. Observe that the bivariate generating function of generalized walks where intermediate values are allowed to be negative, with z marking the length and u marking the final altitude, is rational:

$$G(z,u) = \frac{1}{1 - zS(u)}. \tag{55}$$

Returning to nonnegative walks, the main result to be proved below is the following: *For each finite set $\Omega \in \mathbb{Z}$, the generating function of excursions is an algebraic function that is explicitly computable from Ω.* There are many ways to view this result. The problem is usually treated within probability theory by means of Wiener-Hopf factorizations [**364**]. In contrast, Labelle and Yeh [**282**] show that an unambiguous context-free specification can be systematically constructed, a fact that is sufficient to ensure the algebraicity of the GF $F(z)$. (Their approach is based implicitly on the construction of a finite pushdown automaton itself equivalent, by general principles, to a context-free grammar.) The Labelle-Yeh construction reduces the problem to a large, but somewhat “blind”, combinatorial preprocessing, and, for analysts it has the disadvantage of not extracting a simpler (and noncombinatorial) structure inherent in the problem. The method described below is often known as the “kernel” method. It takes its inspiration from exercises in the 1968 edition of Knuth’s book [**262**] (Ex. 2.2.1.4 and 2.2.1.11) where a new approach was proposed to the enumeration of Catalan and Schröder objects. The technique has since been extended and systematized by several authors; see for instance [**20, 21, 65, 130, 131**].

Let $f_n(u) = [z^n]F(z,u)$ be the generating function of walks of length n with u recording the final altitude. There is a simple recurrence relating $f_{n+1}(u)$ to $f_n(u)$, namely,

$$f_{n+1}(u) = S(u) \cdot f_n(u) - r_n(u), \tag{56}$$

where $r_n(u)$ is a Laurent polynomial consisting of the sum of all the monomials of $S(u)f_n(u)$ that involve negative powers[9] of u:

$$r_n(u) := \sum_{j=-c}^{-1} u^j \left([u^j]\, S(u) f_n(u)\right) = \{u^{<0}\} S(u) f_n(u). \tag{57}$$

---

[8] If Ω is a set, then the coefficients of S lie in $\{0,1\}$. The treatment above applies in all generality to cases where the coefficients are arbitrary positive real numbers. This accounts for probabilistic situations as well as multisets of jump values.

[9] The convenient notation $\{u^{<0}\}$ denotes the singular part of a Laurent expansion: $\{u^{<0}\}f(z) := \sum_{j<0} \left([u^j] f(u)\right) \cdot u^j$.

The idea behind the formula is to subtract the effect of those steps that would take the walk below the horizontal axis. For instance, one has

$$\begin{array}{lcl} S(u) = \dfrac{S_{-1}}{u} + O(1) & : & r_n(u) = \dfrac{S_{-1}}{u} f_n(0) \\ S(u) = \dfrac{S_{-2}}{u^2} + \dfrac{S_{-1}}{u} + O(1) & : & r_n(u) = \left(\dfrac{S_{-2}}{u^2} + \dfrac{S_{-1}}{u}\right) f_n(0) + \dfrac{S_{-2}}{u} f_n'(0) \end{array}$$

and generally:

$$\lambda_j(u) = \frac{1}{j!}\{u^{<0}\}u^j S(u). \tag{58}$$

Thus, from (56) and (57) (multiply by z^{n+1} and sum), the generating function $F(z,u)$ satisfies the fundamental functional equation

$$F(z,u) = 1 + zS(u)F(z,u) - z\{u^{<0}\}\left(S(u)F(z,u)\right). \tag{59}$$

Explicitly, one has

$$F(z,u) = 1 + zS(u)F(z,u) - z\sum_{j=0}^{c-1} \lambda_j(u) \left[\frac{\partial^j}{\partial u^j} F(z,u)\right]_{u=0}, \tag{60}$$

for Laurent polynomials $\lambda_j(u)$ that depend on $S(u)$ in an effective way by (58).

The main equations (59) and (60) involve one unknown bivariate GF, $F(z,u)$ and c univariate GF's, the partial derivatives of F specialized at $u = 0$. It is true, but not at all obvious, that the single functional equation (60) fully determines the $c+1$ unknowns. The basic technique is known as "cancelling the kernel" and it relies on strong analyticity properties; see the book by Fayolle *et al.* [**131**] for deep ramifications. The form of (60) to be employed for this purpose starts by grouping on one side the terms involving $F(z,u)$,

$$F(z,u)(1 - zS(u)) = 1 - z\sum_{j=0}^{c-1} \lambda_j(u) G_j(z), \qquad G_j(z) := \left[\frac{\partial^j}{\partial u^j} F(z,u)\right]. \tag{61}$$

If the right side was not present, then the solution would reduce to (55). In the case at hand, from the combinatorial origin of the problem and implied bounds, the quantity $F(z,u)$ is bivariate analytic at $(z,u) = (0,0)$ (by elementary exponential majorizations on the coefficients). The main principle of the kernel method consists in *coupling* the values of z and u in such a way that $1 - zS(u) = 0$, so that $F(z,u)$ disappears from the picture. A condition is that both z and u should remain small (so that F remains analytic). Relations between the partial derivatives are then obtained from such a specializations, $(z,u) \mapsto (z, u(z))$, which happen to be just in the right quantity.

Consequently, we consider the "kernel equation",

$$1 - zS(u) = 0, \tag{62}$$

which is rewritten as

$$u^c = z \cdot (u^c S(u)).$$

Under this form, it is clear that the kernel equation (62) defines $c+d$ branches of an algebraic function. A local analysis (Newton's polygon method) shows that, amongst these $c+d$ branches, there are c branches that tend to 0 as $z \to 0$ while the other d

tend to infinity as $z \to 0$. Let $u_0(z), \ldots, u_{c-1}(z)$ be the c branches that tend to 0, that we call "small" branches. In addition, we single out $u_0(z)$, the "principal" solution, by the reality condition

$$u_0(z) \sim \gamma z^{1/c}, \quad \gamma := (S_c)^{1/c} \in \mathbb{R}_{>0} \qquad (z \to 0^+).$$

By local uniformization (39), the conjugate branches are given locally by

$$u_\ell(z) = u_0(\mathrm{e}^{2i\ell\pi} z) \qquad (z \to 0^+).$$

Coupling z and u by $u = u_\ell(z)$ produces interesting specializations of Equation (61). In that case, (z, u) is close to $(0, 0)$ where F is bivariate analytic so that the substitution is admissible. By substitution, we get

$$(63) \qquad 1 - z \sum_{j=0}^{c-1} \lambda_j(u_\ell(z)) \left[\frac{\partial^j}{\partial u^j} F(z, u) \right]_{u=0}, \qquad \ell = 0 \,..\, c-1.$$

This is now a linear system of c equation in c unknowns (the partial derivatives) with algebraic coefficients that, in principle, determines $F(z, 0)$.

A convenient approach to the solution of (63) is due to Mireille Bousquet-Mélou. The argument goes as follows. The quantity

$$(64) \qquad M(u) := u^c - z u^c \sum_{j=0}^{c-1} \lambda_j(u) \frac{\partial^j}{\partial u^j} F(z, 0)$$

can be regarded as a polynomial in u. It is monic while it vanishes by construction at the c small branches $u_0, \ldots, u_{c-1}$. Consequently, one has the factorization,

$$(65) \qquad M(u) = \prod_{\ell=0}^{c-1} (u - u_\ell(z)).$$

Now, the constant term of $M(u)$ is otherwise known to equal $-zS_{-c}F(z, 0)$, by the definition (64) of $M(u)$ and by Equation (58) specialized to $\lambda_0(u)$. Thus, the comparison of constant terms between (64) and (65) provides us with an explicit form of the OGF of excursions:

$$F(z, 0) = \frac{(-1)^{c-1}}{S_{-c} z} \prod_{\ell=0}^{c-1} u_\ell(z).$$

One can then finally return to the original functional equation and pull the BGF $F(z, u)$. We can thus state:

PROPOSITION VII.5 (Kernel method for walks). *Let Ω be a finite step of jumps and let $S(u)$ be the characteristic polynomial of Ω. Consider the c small branches of the "kernel" equation,*

$$1 - zS(u) = 0,$$

denoted by $u_0(z), \ldots, u_{c-1}(z)$.

The generating function of excursions is expressible as

$$F(z) = \frac{(-1)^{c-1}}{z S_{-c}} \prod_{\ell=0}^{c-1} u_\ell(z) \quad \textit{where } S_{-c} = [u^{-c}] S(u)$$

is the multiplicity (or weight) of the smallest element $-c \in \Omega$. More generally the bivariate generating function of nonnegative walks (also known as meanders) with u marking final altitude is bivariate algebraic and given by

$$F(z,u) = \frac{1}{u^c - zu^c S(u)} \prod_{\ell=0}^{c-1} (u - u_\ell(z)).$$

Our treatment above is based on an article of Banderier and Flajolet [**21**] where several similar results are established. In particular, bridges are walks (possibly involving negative steps) that return on the horizontal axis, i.e., their final altitude equals 0. The OGF of bridges is expressible in terms of the small branches, by

$$B(z) = z \sum_{j=1}^{c} \frac{u_j'(z)}{u_j(z)} = z \frac{d}{dz} \log (u_1(z) \cdots u_c(z)).$$

(This is easily obtained by a residue calculation of the diagonal of $(1 - zS(u))^{-1}$.)

We give next a few examples illustrating this kernel technique.

EXAMPLE 13. *Trees and Łukasiewicz codes.* A particular class of walks is of special interest; it corresponds to cases where $c = 1$, that is, the largest jump in the negative direction has amplitude 1. Consequently, $\Omega + 1 = \{0, s_1, s_2, \dots, s_d\}$. In that situation, combinatorial theory teaches us the existence of fundamental isomorphisms between walks defined by steps Ω and trees whose degrees are constrained to lie in $1 + \Omega$. The correspondence is by way of Łukasiewicz codes[10], also known as 'Polish" prefix codes, "Polish" prefix notation and introduced in Chapter I. From this, we expect to find tree GF's in such cases.

As regards generating functions, there now exists only *one* small branch, namely the solution $u_0(z)$ to $u_0(z) = z\phi(u_0(z))$ (where $\phi(u) = uS(u)$) that is analytic at the origin. One then has $F(z) = F(z,0) = \frac{1}{z} u_0(z)$, so that the walk GF is determined by

$$F(z,0) = \frac{1}{z} u_0(z), \quad u_0(z) = z\phi(u_0(z)), \quad \phi(u) := uS(u).$$

This form is consistent with what is already known regarding the enumeration of simple families of trees. In addition, one finds

$$F(z,u) = \frac{1 - u^{-1}u_0(z)}{1 - zS(u)} = \frac{u - u_0(z)}{u - z\phi(u)}.$$

Classical specializations are rederived in this way:

— the Catalan walk (Dyck path), defined by $\Omega = \{-1, +1\}$ and $\phi(u) = 1 + u^2$, has

$$u_0(z) = \frac{1}{2z} \left(1 - \sqrt{1 - 4z^2}\right);$$

— the Motzkin walk, defined by $\Omega = \{-1, 0, +1\}$ and $\phi(u) = 1 + u + u^2$ has

$$u_0(z) = \frac{1}{2z} \left(1 - z - \sqrt{1 - 2z - 3z^2}\right);$$

[10] Such a code [**294**] is obtained by a preorder traversal of the tree, recording a jump of $r - 1$ when a node of outdegree r is encountered. The sequence of jumps gives rise to an excursion followed by an extra -1 jump.

— the modified Catalan walk, defined by $\Omega = \{-1, 0, 0+1\}$ (with two steps of type 0) and $\phi(u) = 1 + 2u + u^2$, has

$$u_0(z) = \frac{1}{2z}\left(1 - 2z - \sqrt{1-4z}\right);$$

— the d-ary tree walk (the excursions encode d-ary trees) defined by $\Omega = \{-1, d-1\}$, has $u_0(z)$ that is defined implicitly by

$$u_0(z) = z(1 + u_0(z)^d).$$

This vastly generalizes the enumeration of Dyck paths discussed in Chapter I. END OF EXAMPLE 13. □

EXAMPLE 14. *Walks with amplitude equal to 2.* Take now $\Omega = \{-2, -1, 1, 2\}$ so that

$$S(u) = u^{-2} + u^{-1} + u + u^2.$$

Then, $u_0(z), u_1(z)$ are the two branches that vanish as $z \to 0$ of the curve

$$y^2 = z(1 + y + y^3 + y^4).$$

The linear system that determines $F(z,0)$ and $F'(z,0)$ is

$$\begin{cases} 1 - \left(\dfrac{z}{u_0(z)^2} + \dfrac{z}{u_0(z)}\right) F(z,0) - \dfrac{z}{u_0(z)} F'(z,0) & = & 0 \\ 1 - \left(\dfrac{z}{u_1(z)^2} + \dfrac{z}{u_1(z)}\right) F(z,0) - \dfrac{z}{u_1(z)} F'(z,0) & = & 0 \end{cases}$$

(derivatives are taken with respect to the second argument) and one finds

$$F(z,0) = -\frac{1}{z} u_0(z) u_1(z), \qquad F'(z,0) = \frac{1}{z}(u_0(z) + u_1(z) + u_0(z)u_1(z)).$$

This gives the number of walks, through a combination of series expansions,

$$F(z) = 1 + 2z^2 + 2z^3 + 11z^4 + 24z^5 + 93z^6 + 272z^7 + 971z^8 + 3194z^9 + \cdots.$$

A single algebraic equation for $F(z) = F(z,0)$ is then obtained by elimination (*e.g.*, via Groebner bases) from the system:

$$\begin{cases} u_0^2 - z(1 + u_0 + u_0^3 + u_0^4) & = & 0 \\ u_1^2 - z(1 + u_1 + u_1^3 + u_1^4) & = & 0 \\ zF + u_0 u_1 & = & 0 \end{cases}$$

Elimination shows that $F(z)$ is a root of the equation

$$z^4 y^4 - z^2(1+2z)y^3 + z(2+3z)y^2 - (1+2z)y + 1 = 0.$$

For walks corresponding to $\Omega = \{-2, -1, 0, 1, 2\}$, we find similarly $F(z) = -\frac{1}{z}u_0(z)u_1(z)$, where u_0, u_1 are the small branches of $y^2 = z(1 + y + y^2 + y^3 + y^4)$, the expansion starts as

$$F(z) = 1 + z + 3z^2 + 9z^3 + 32z^4 + 120z^5 + 473z^6 + 1925z^7 + 8034z^8 + \cdots,$$

and $F(z)$ is a root of the equation

$$z^4 y^4 - z^2(1+z)y^3 + z(2+z)y^2 - (1+z)y + 1 = 0.$$

In this case, the GFs are no longer of the simple tree type. .... END OF EXAMPLE 14. □

It is of interest to note that singularities of the branches involved in the statement of Proposition VII.5 can be worked out in all generality [**21**]. The GF of excursions has a dominant singularity of type $\sqrt{Z}$, while that of bridges is of type $1/\sqrt{Z}$. Parameters of walks, excursions, bridges, and meanders can then be analysed in a uniform fashion [**21**].

$\triangleright$ **21.** *Asymptotics of excursions and bridges.* Define the structural constant τ by $S(\tau) = 0$, $\tau > 0$. Then assuming aperiodicity, the number of bridges (B_n) and the number of excursions (E_n) satisfy

$$B_n \sim \beta_0 \frac{P(\tau)^n}{\sqrt{2\pi n}}, \qquad E_n \sim \epsilon_0 \frac{P(\tau)^n}{2\sqrt{\pi n^3}},$$

where

$$\beta_0 = \frac{1}{\tau}\sqrt{\frac{P(\tau)}{P''(\tau)}}, \qquad \epsilon_0 = \frac{(-1)^{c-1}}{S_{-c}}\sqrt{\frac{2P(\tau)^3}{P''(\tau)}}\prod_{j=1}^{c-1} u_j\left(\frac{1}{P(\tau)}\right).$$

There, the u_j represent the small branches and u_0 is the branch that is real positive as $z \to 0$. Details are in [**21**]. $\triangleleft$

VII. 5.3. Maps and the quadratic method. A (planar) map is a connected planar graph together with an embedding into the plane. In all, generality, loops and multiple edges are allowed. A planar map therefore separates the plane into regions called faces (Figure 14). The maps considered here are in addition rooted, meaning that a face, an incident edge, and an incident vertex are distinguished. In this section, only rooted maps are considered[11]. When representing rooted maps, we shall agree to draw the root edge with an arrow pointing away from the root node, and to take the root face as that face lying to the left of the directed edge (represented in grey on Figure 14).

Tutte launched in the 1960's a large census of planar maps, with the intention of attacking the four-colour problem by enumerative techniques[12]; see [**68, 413, 414, 415, 416**]. There exists in fact an entire zoo of maps defined by various degree or connectivity constraints. In this chapter, we shall limit ourselves to conveying a flavour of this vast theory, with the goal of showing how algebraic functions arise. The presentation takes its inspiration from the book of Goulden and Jackson [**208**, Sec. 2.9]

Let $\mathcal{M}$ be the class of all maps where size is taken to be the number of edges. Let $M(z, u)$ be the BGF of maps with u marking the number of edges on the outside face. The basic surgery performed on maps distinguishes two cases based upon the nature of the root edge. A rooted map will be declared to be isthmic if the root edge r of map μ is an "isthmus" whose deletion would disconnect the graph. Clearly, one has,

$$\mathcal{M} = o + \mathcal{M}^{(i)} + \mathcal{M}^{(n)}, \tag{66}$$

[11]Nothing is lost regarding asymptotic properties of random structures when a rooting is imposed. The reason is that a map has, with probability exponentially close to 1, a trivial automorphism group; consequently, almost all maps of m edges can be rooted in $2m$ ways (by choosing an edge, and an orientation of this edge), and there is an almost uniform $2m$-to-1 correspondence between unrooted maps and rooted ones.

[12]The four-colour theorem to the effect that every planar graph can be coloured using only four colours was eventually proved by Appel and Haken in 1976, using structural graph theory methods supplemented by extensive computer search.

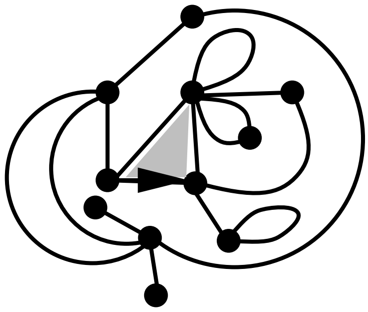
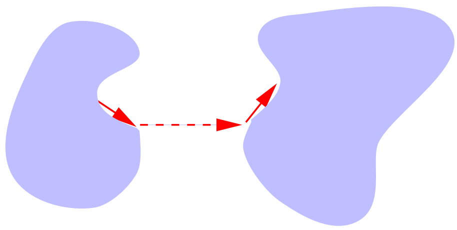


FIGURE 14. A planar map.

where $\mathcal{M}^{(i)}$ (resp. $\mathcal{M}^{(n)}$) represent the class of isthmic (resp. non-isthmic) maps and 'o' is the graph consisting of a single vertex and no edge. There are accordingly two ways to build maps from smaller ones by adding a new edge.

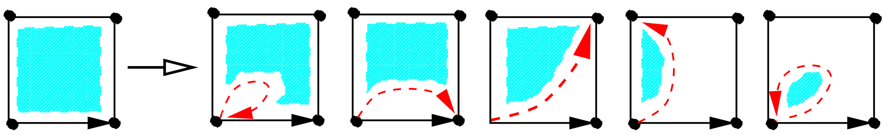
(i) The class of all isthmic maps is constructed by taking two arbitrary maps and joining them together by a new root edge, as shown below:



The effect is to increase the number of edges by 1 (the new root edge) and have the root face degree become 2 (the two sides of the new root edge) plus the sum of the root face degrees of the component maps. The construction is clearly revertible. In other words, the BGF of $\mathcal{M}^{(i)}$ is

$$M^{(i)}(z,u) = zu^2 M(z,u)^2. \tag{67}$$

(ii) The class of non-isthmic maps is obtained by taking an already existing map and adding an edge that preserves its root node and "cuts across" its root face in some unambiguous fashion (so that the construction should be revertible). This operation will therefore result in a new map with an essentially smaller root-face degree. For instance, there are 5 ways to cut across a root face of degree 4, namely,



This corresponds to the linear transformation

$$u^4 \quad \mapsto \quad zu^5 \quad + \quad zu^4 \quad + \quad zu^3 \quad + \quad zu^2 \quad + \quad zu^1.$$

In general the effect on a map with root face of degree k is described by the transformation $u^k \mapsto z(1-u^{k+1})/(1-u)$; equivalently, each monomial $g(u) = u^k$ is transformed into $u(g(1)-ug(u))/(1-u)$. Thus, the OGF of $\mathcal{M}^{(n)}$ involves a discrete difference operator:

$$M^{(n)}(z,u) = zu\frac{M(z,1)-uM(z,u)}{1-u}. \tag{68}$$

Collecting the contributions from (67) and (68) in (66) then yields the basic functional equation,

$$M(z,u) = 1 + u^2 zM(z,u)^2 + uz\frac{M(z,1)-uM(z,u)}{1-z}. \tag{69}$$

The functional equation (69) binds two unknown functions, $M(z,u)$ and $M(z,1)$. Much like in the case of walks, it would seem to be underdetermined. Now, a method due to Tutte and known as the quadratic method provides solutions. Following Tutte and the account in [**208**, p. 138], we consider momentarily the more general equation

$$\left(g_1 F(z,u) + g_2\right)^2 = g_3, \tag{70}$$

where $g_j = G_j(z,u,h(z))$ and the G_j are explicit functions—here the unknown functions are $F(z,u)$ and $h(z)$ (cf. $M(z,u)$ and $M(z,1)$ in (69)). Bind u and z in such a way that the left side of (70) vanishes, that is, substitute $u = u(z)$ (a yet unknown function) so that $g_1F + g_2 = 0$. Since the left-hand side of (70) now has a double root in u, so must the right-hand side, which implies

$$g_3 = 0, \qquad \left.\frac{\partial g_3}{\partial u}\right|_{u=u(z)}. \tag{71}$$

The original equation has become a system of two equations in two unknowns that determines implicitly $h(z)$ and $u(z)$. From there, elimination provides individual equations for $u(z)$ and for $h(z)$. (If needed, $F(z,u)$ can then be recovered by solving a quadratic equation.) It will be recognized that, if the quantities q_1, g_2, g_3 are polynomials, then the process invariably yields solutions that are algebraic functions.

We now carry out this programme in the case of maps and Equation (69). First, isolate $M(z,u)$ by completing the square, giving

$$\left(M(z,u) - \frac{1}{2}\frac{1-u+u^2z}{u^2z(1-u)}\right)^2 = Q(z,u) + \frac{M(z,1)}{u(1-u)}, \tag{72}$$

where

$$Q(z,u) = \frac{z^2u^4 - 2zu^2(u-1)(2u-1) + (1-u^2)}{4u^4z^2(1-u)^2}.$$

Next, the condition expressing the existence of a double root is

$$Q(z,u) + \frac{1}{u(1-u)}M(z,1) = 0, \quad Q'_u(z,u) + \frac{2u-1}{u^2(1-u)^2}M(z,1) = 0.$$

It is now easy to eliminate $M(z,1)$, since the dependency in M is linear, and a straightforward calculation shows that $u = u(z)$ should satisfy

$$\left(u^2 z + (u-1)\right)\left(u^2 z + (u-1)(2u-3)\right) = 0.$$

The first parameterization would lead to $M(z,1) = 1/z$ which is not acceptable. Thus, $u(z)$ is to be taken as the root of the second factor, with $M(z,1)$ being defined parametrically by

$$z = \frac{(1-u)(2u-3)}{u^2}, \quad M(z,1) = -u\frac{3u-4}{(2u-3)^2}.$$

The change of parameter $u = 1 - 1/w$ reduces this further to the "Lagrangean form",

$$z = \frac{w}{1-3w}, \quad M(z,1) = \frac{1-4w}{(1-3w)^2}. \tag{73}$$

To this the Lagrange inversion theorem can be applied. The number of maps with n edges, $M_n = [z^n]M(z,1)$ is then determined as

$$M_n = 2\frac{(2n)!3^n}{n!(n+2)!},$$

and one obtains Sequence **A000168** of the *EIS*:

$$M(z,1) = 1 + 2z + 9z^2 + 54z^3 + 378z^4 + 2916z^5 + 24057z^6 + 208494z^7 + \cdots .$$

We refer to [**208**, Sec. 2.9] for detailed calculations (that are nowadays routinely performed with assistance of a computer algebra system). Currently, there exist many applications of the method to maps satisfying all sorts of combinatorial constraints (*e.g.*, multiconnectivity); see [**377**] for a recent panorama.

The derivation above has purposely stressed a parameterized approach as this constitutes a widely applicable approach in many situations. In a simple case like this, we may also eliminate u and solve explicitly for $M(z,1)$, to wit,

$$M(z) \equiv M(z,1) = -\frac{1}{54\,z^2}\left(1 - 18z - (1-12z)^{3/2}\right).$$

It is interesting to note that the singular exponent here is $\frac{3}{2}$, a fact further reflected by the somewhat atypical factor of $n^{-5/2}$ in the asymptotic form of coefficients:

$$M_n \sim \frac{2}{\sqrt{\pi n^5}} 12^n \qquad (n \to \infty).$$

Accordingly, randomness properties of maps are appreciably different from what is observed in trees and many commonly encountered context-free objects.

VII. 6. Notes

The exp-log schema, like its companion, the supercritical-sequence schema, illustrates the level of generality that can be attained by singularity analysis techniques. Refinements of the results we have given can be found in the book by Arratia, Barbour, and Tavaré [**16**], which develops a stochastic process approach to these questions; see also [**15**] by the same authors for an accessible introduction.

The rest of the chapter deals in an essential manner with recursively defined structures. As noted repeatedly in the course of this chapter, this is very often conducive to square-root singularity and universal behaviours that are quantified by exponents of the form $\frac{1}{2}, \frac{3}{2}, \ldots$. Simple varieties of trees have been introduced in an important paper of Meir and Moon [**312**], that bases itself on methods developed earlier by Pólya [**347, 349**] and Otter [**335**]. One of the merits of [**312**] is to demonstrate that a high level of generality is attainable when discussing properties of trees. A similar treatment can be inflicted more generally to recursively defined structures when their generating function satisfies an implicit equation. In this way, nonplane unlabelled trees are shown to exhibit properties very similar to their plane counterparts. It is of interest to note that such of the enumerative questions in this area had been initially motivated by problems of theoretical chemistry: see the colourful account of Cayley and Sylvester's works in [**52**], the reference books by Harary–Palmer [**223**] and Finch [**137**], as well as PÓlya's original studies [**347, 349**].

Algebraic functions are the modern counterpart of the study of curves by classical Greek mathematicians. They are either approached by algebraic methods (this is the core of algebraic geometry) or by transcendental methods. For our purposes, however, only rudiments of the theory of curves are needed. For this, there exist several excellent introductory books, of which we recommend the ones by Abhyankar [**1**], Fulton [**190**], and Kirwan [**254**]. On the algebraic side, we have striven to provide an introduction to algebraic functions that requires minimal apparatus. At the same time the emphasis has been put somewhat on algorithmic aspects, since most algebraic models are nowadays likely to be treated with the help of computer algebra. As regards symbolic computational aspects, we recommend the treatise by von zur Gathen and Jürgen [**428**] for background, while polynomial systems are excellently reviewed in the book by Cox, Little, and O'Shea [**86**].

In the combinatorial domain, algebraic functions have been used early: in Euler and Segner's enumeration of triangulations (1753) as well as in Schröder's famous "*Vier combinatorische Probleme*" described by Stanley in [**393**, p. 177]. A major advance was the realization by Chomsky and Schützenberger that algebraic functions are the "exact" counterpart of context-free grammars and languages (see their historic paper [**80**]). A masterful summary of the early theory appears in the proceedings edited by Berstel [**43**] while a modern and precise exposition forms the subject of Chapter 6 of Stanley's book [**393**]. On the analytic-asymptotic side, many researchers have long been aware of the power of Puiseux expansions in conjunction with some version of singularity analysis (often in the form of the Darboux–Pólya method: see [**349**] based on Pólya's classic paper [**347**] of 1937). However, there appeared to be difficulties in coping with the fully general problem of algebraic coefficient asymptotics [**72, 316**]. We believe that Section VII. 4.1 sketches the first complete theory. In the case of positive systems, the "Drmota-Lalley-Woods" theorem is the key to most problems encountered in practice—its importance should be clear from the developments of Section VII. 4.2.

The applications of algebraic functions to context-free languages have been known for some time (e.g., [**145**]). Our presentation of 1-dimensional walks of a general type follows a recent article by Banderier and Flajolet [**21**], whch can be regarded as the analytic pendant of algebraic studies by Gessel [**197, 198**]. The kernel method has its origins in problems of queueing theory and random walks [**130, 131**] and is further explored in an article by Bousquet-Mélou and Petkovšek [**65**]. The algebraic treatment of random maps by the quadratic method is due to brilliant studies of Tutte in the 1960's: see for instance his census [**413**] and the account in the book by Jackson and Goulden [**208**]. A combinatorial-analytic treatment of multiconnectivity issues is given in [**22**], where the possibility of treating in a unified manner about a dozen families of maps appears clearly.

VIII

Saddle Point Asymptotics

Like a lazy hiker, the path crosses the ridge at a low point;
but unlike the hiker, the best path takes the steepest ascent to the ridge.
[· · ·] The integral will then be concentrated in a small interval.

— DANIEL GREENE AND DONALD KNUTH [**214**, sec. 4.3.3]

Contents

A *saddle point* of a surface is a point reminiscent of the inner part of a saddle or of a geographical pass between two mountains. If the surface represents the modulus of an analytic function, saddle points are simply determined as the zeros of the derivative of that function.

In order to estimate *complex integrals* with an analytic integrand, it is often a good strategy to take as a contour of integration a line that "crosses" one or several of the saddle points of the function. When applied to integrals depending on a large parameter, as is the case for Cauchy integrals giving coefficients of generating functions, this provides in many cases very accurate asymptotic information.

The saddle point method can lead to asymptotic estimates and even to complete asymptotic expansions. Its principle is to use a saddle point crossing path, then estimate the integrand locally near this saddle point (at which point the integrand achieves its maximum), and deduce by local approximations and termwise integration an asymptotic expansion of the integral itself. Some sort of "localization" or "concentration" property is required to ensure that the contribution near the saddle point captures the essential part of the integral. A simplified form of the method provides what are known as saddle point bounds—these are useful and technically simple upper bounds obtained by applying trivial bounds to a saddle point crossing path

As regards coefficient extraction in the context of analytic combinatorics, the method applies well to rapidly varying functions. Typical instances are entire functions as well as functions with singularities at a finite distance that exhibit some form

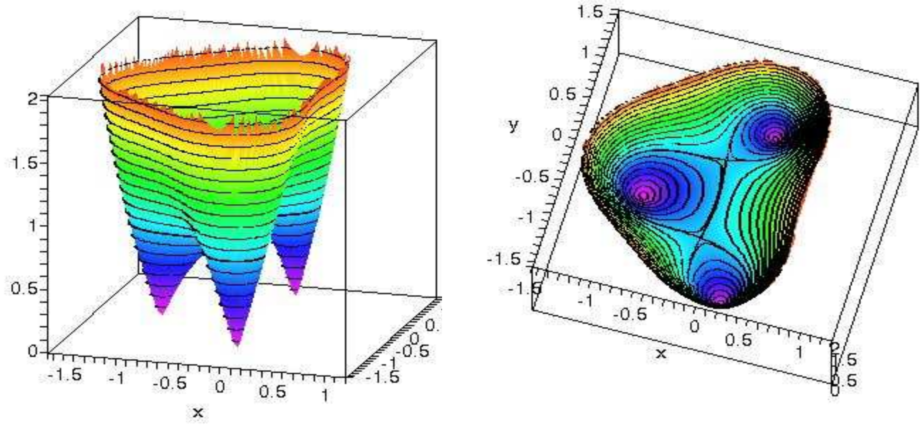


FIGURE 1. The "tripod": two views of $|1+z+z^2+z^3|$ (front, top) with level lines displayed.

of exponential growth. Saddle point analysis then complements singularity analysis whose scope is essentially the category of functions having moderate (polynomial) growth at their singularities. The saddle point method is also a method of choice for the analysis of coefficients of large powers of some fixed function and, in this context, it paves the way to the study of multivariate asymptotics and limiting distributions developed developed in the next chapter.

Applications are given here to Stirling's formula as well as the asymptotics of the involution numbers and Bell numbers enumerating set partitions. The asymptotic enumeration of integer partitions is one of the jewels of classical analysis and we provide an introduction to this rich topic where saddle points give access to effective estimates of an amazingly good quality. Other combinatorial applications include a new derivation of Stirling's formula, balls-in-bins models and capacity, the number of increasing subsequences in permutations, blocks in set partitions, and the counting of acyclic graphs (forests of unrooted trees).

VIII. 1. Preamble: Landscapes of analytic functions and saddle points

Given any function f analytic in an open set Ω, the surface in $\mathbb{R}^3$ generated by its modulus, namely, in (x, y, t) coordinates,

$$t(x,y) = |t(x+iy)|,$$

is far from being arbitrary. Its points can be of only one of three types: ordinary points, (the generic case), zeros, and saddle points.

The function $t(x, y)$ is initially defined from $\mathbb{R}^2$ to $\mathbb{R}$, but it is also convenient to treat it as a function from $\mathbb{C}$ to $\mathbb{R}$ and write $t(z) = t(x, y)$ when $z = x + iy$. Let $z_0 = x_0 + iy_0$ be an interior point of Ω. The local shape of the surface $t(z)$ for z near z_0

depends on which of the initial elements in the sequence $f(z_0), f'(z_0), f''(z_0), \ldots$, vanish. An idea of the typical shape of such a surface can be obtained by examining Figure 1 relative to the modulus of the third degree polynomial $1+z+z^2+z^3$ which has zeros at $-1, i, -i$ while its derivative has zeros at $\frac{1}{3} \pm \frac{i}{3}\sqrt{2}$. The classification of points is conveniently obtained by considering polar coordinates, $z = z_0 + re^{i\theta}$, with r small.

An *ordinary point* is such that $f(z_0) \neq 0$, $f'(z_0) \neq 0$. This is clearly the generic situation as analytic functions have only isolated zeros. In that case, one has for small $r > 0$:

$$(1) \qquad t(z) = \left|f(z_0) + re^{i\theta} f'(z_0) + O(r^2)\right| = t(z_0)\left|1 + \lambda r e^{i(\theta+\phi)} + O(r^2)\right|,$$

where we have set $f'(z_0)/f(z_0) = \lambda e^{i\phi}$. The modulus then satisfies

$$t(z) = t(z_0)\left(1 + \lambda r \cos(\theta+\phi) + O(r^2)\right).$$

Thus, for r taken small enough, as θ varies, $t(z)$ is maximum when $\theta = -\phi$ (where it is $\sim 1+r$), and minimum when $\theta = -\phi+\pi$ (where it is $\sim 1-r$). When $\theta = -\phi \pm \frac{\pi}{2}$, $t(z) = t(z_0) + o(r)$, which means that $t(z)$ is stationary. This is easily interpreted: the line $\theta \equiv -\phi \pmod{\pi}$ is (locally) a steepest descent line; the perpendicular line $\theta \equiv -\phi + \frac{\pi}{2} \pmod{\pi}$ is locally a level line. In particular, near an ordinary point, the surface $t(x, y)$ has neither a minimum nor a maximum. In figurative terms, this is like standing on the flank of a mountain.

A *zero* is by definition a point such that $f(z_0) = 0$. In this case, the function $t(z)$ attains its minimum value 0 at z_0. A zero is thus like a sink or the bottom of a lake, save that, in the landscape of an analytic function, all lakes are at see level.

A *saddle point* is a point such that $f(z_0) \neq 0$, $f'(z_0) = 0$. It is said to be a *simple* saddle point if furthermore $f''(z_0) \neq 0$. In that case, a calculation similar to (1),

$$(2) \quad t(z) = \left|f(z_0) + \frac{1}{2} r^2 e^{2i\theta} f''(z_0) + O(r^3)\right| = t(z_0)\left|1 + \lambda r^2 e^{i(2\theta+\phi)} + O(r^3)\right|,$$

where we have set $\frac{1}{2}f''(z_0)/f(z_0) = \lambda e^{i\phi}$, shows that the modulus satisfies

$$t(z) = t(z_0)\left(1 + \lambda r^2 \cos(2\theta+\phi) + O(r^3)\right).$$

Thus, starting at the direction $\theta = -\phi$ and turning around z_0, the following sequence of events regarding the modulus $t(z) = |f(z)|$ is observed: it is maximal ($\theta = -\phi$), stationary ($\theta = -\phi + \frac{\pi}{2}$), minimal ($\theta = -\phi + \pi$), stationary, ($\theta = -\phi + 3\frac{\pi}{2}$), maximal again ($\theta = -\phi + \pi$), and so on. The pattern, symbolically '+=−=', repeats itself twice. This is superficially similar to an ordinary point, save for the important fact that changes are observed at twice the angular speed. Accordingly, the shape of the surface looks quite different; it is like the central part of a saddle. Two level curves cross at a right angle: one steepest descent line (away from the saddle point) is perpendicular to another steepest descent line (towards the saddle point). In a mountain landscape, this is thus much like a pass between two mountains.

Here is a diagram showing the local structure of level curves (in solid lines), steepest descent lines (dashed with arrows pointing towards the direction of increase) and regions (hashed) where the surface lies below the reference value $t(z_0)$:

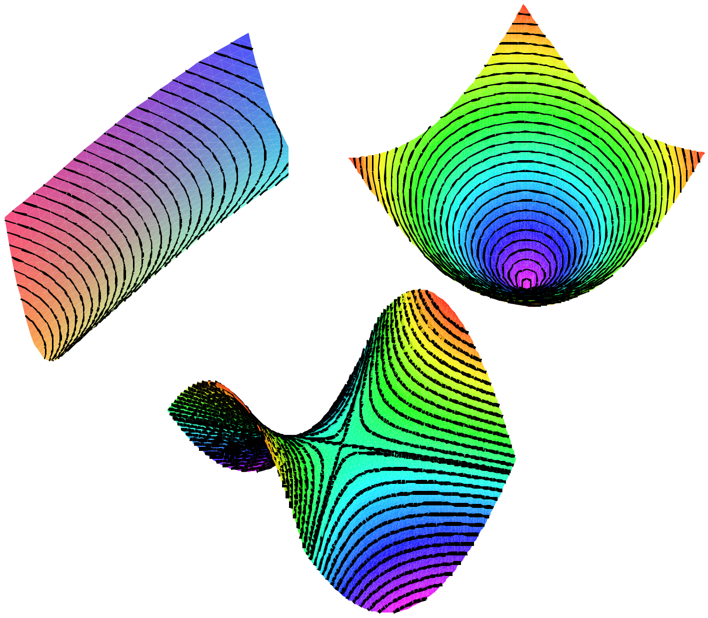
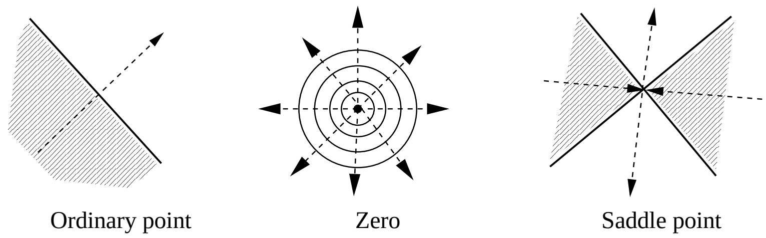


FIGURE 2. The different types of points on a surface $|f(z)|$: an ordinary point, a zero, a simple saddle point. Here $f(z) = \cos z$ and the points are an ordinary point at $\pi/4$ (upper left), a zero at $\pi/2$ (upper right), and a saddle point at 0 (bottom center). Level lines are shown on the surfaces.



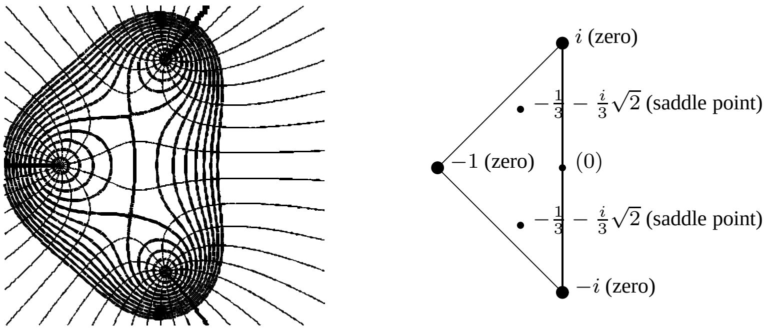
The two regions on each side corresponding to points with an altitude *below* a simple saddle point are often referred to as "valleys".

Generally, a *multiple saddle point* has multiplicity p if $f(z_0) \neq 0$ and all derivatives $f'(z_0), \ldots, f^{(p)}(z_0)$ are equal to zero while $f^{(p+1)}(z_0) \neq 0$. In that case, the basic pattern '$+ = - =$' repeats itself $p+1$ times. A double saddle point is also called a "monkey saddle" since it can be visualized as a saddle having places for the legs and the tail. From such a double saddle point, three roads go down leading to three different valleys.

THEOREM VIII.1 (Classification of points on modulus surfaces). *A surface $|f(z)|$ attached to the modulus of a function analytic over an open set Ω has points of only three possible types:* (i) ordinary points, (ii) zeros, (iii) saddle points. *A simple* *saddle point is locally the common apex of two curvilinear sectors with angle $\frac{\pi}{2}$, also referred to as "valleys".*

As a consequence, the surface defined by the modulus of an analytic function has no maximum: this property is known as the *Maximum Modulus Principle*. It has no minimum either, apart from zeros. It is therefore a peakless landscape in de Bruijn's words [**93**]. The three different types of points are illustrated by Figure 2. Here is a diagram representing for $f(z) = 1 + z + z^2 + z^3$ the network of level curves and the orthogonal network of steepest ascent/descent lines (compare with Figure 1):

(3)



▷ **1.** *The Fundamental Theorem of Algebra.* This theorem asserts that a polynomial has at least one root (hence n roots if its degree is n). Let $P(z) = 1 + a_1 z + \cdots a_n z^n$ be a polynomial of degree n. Consider $f(z) = 1/P(z)$. By basic analysis, one can take R sufficiently large, so that on $|z| = R$, one has $|f(z)| < \frac{1}{2}$. Assume *a contrario* that $P(z)$ has no zero. Then, $f(z)$ which is analytic in $|z| \leq R$ should attain its maximum at an interior point (since $f(0) = 1$), so that a contradiction has been reached. ◁

▷ **2.** *Saddle points of polynomials and the convex hull of zeros.* Let P be a polynomial and $\mathcal{H}$ the convex hull of its zeros. Then any root of $P'(z)$ lies in $\mathcal{H}$. (Proof: assume distinct zeros and consider

$$\phi(z) := \frac{P'(z)}{P(z)} = \sum_{\alpha\,:\,P(\alpha)=0} \frac{1}{z - \alpha}.$$

If z lies outside $\mathcal{H}$, then z sees all zeros α in a half-plane, this by elementary geometry. By projection on the normal to the half plane boundary, there results that, for some θ, one has $\Re(e^{i\theta}\phi(z)) < 0$, so that $P'(z) \neq 0$.) ◁

VIII. 2. Overview of the saddle point method

Saddle point analysis is a general method suited to the estimation of integrals of analytic functions $F(z)$,

(4) $$I = \int_A^B F(z)\, dz,$$

where $F(z) \equiv F_n(z)$ involves some large parameter $n \to +\infty$. The method is usually instrumental when the integrand F is subject to rather violent variations, typically when there occurs in it some exponential or some fixed function raised to a large power (for instance, n). This situation covers a large number of Cauchy coefficient integrals of the form

$$g_n \equiv [z^n]g(z) = \frac{1}{2i\pi} \oint g(z) \frac{dz}{z^{n+1}}. \tag{5}$$

In the last case, the symbol $\oint$ indicates that allowable paths are constrained to encircle the origin (the domain of definition of the integrand is a subset of $\mathbb{C} \setminus \{0\}$; the points A, B can then be seen as coinciding and can be taken somewhere along the negative real line).

VIII. 2.1. Saddle point bounds. Considering the general form (4), we let $\mathcal{C}$ be a contour joining A and B and taken in a domain of the complex plane where $F(z)$ is analytic. By standard inequalities, we have

$$|I| \le ||\mathcal{C}|| \cdot \max_{z\in\mathcal{C}} |F(z)|, \tag{6}$$

with $||\mathcal{C}||$ representing the length of $\mathcal{C}$. This is the usual *trivial bound* from integration theory applied to a fixed contour $\mathcal{C}$.

For an analytic integrand F with A and B inside the domain of analyticity, there is an infinite class $\mathbf{P}$ of acceptable paths to choose from, all in the analyticity domain of F. Thus, we may write

$$|I| \le \min_{\mathcal{C}\in\mathbf{P}} \left[||\mathcal{C}|| \cdot \max_{z\in\mathcal{C}} |F(z)| \right], \tag{7}$$

where the minimum is taken over all paths $\mathcal{C} \in \mathbf{P}$. Broadly speaking, a bound of this type is called a *saddle point bound*[1]. When paths lie in finite regions of the complex plane, the length factor $||\mathcal{C}||$ is normally unimportant for asymptotic bounding purposes.

If there happens to be a path $\mathcal{C}$ from A to B such that no point is at an altitude higher than $\max(|F(A)|, |F(B)|)$, then a simple bound results: $|I| \le ||\mathcal{C}|| \cdot \max(|F(A)|, |F(B)|)$. (This is in a sense the uninteresting case.) However, the usual situation with Cauchy coefficient integrals of combinatorics is that paths have to go at some higher altitude. A path $\mathcal{C}$ that traverses a saddle point by connecting two points at a lower altitude on the surface and by following two steepest descent lines across the saddle point is clearly a local minimum for the path functional

$$\Phi(\mathcal{C}) = \max_{z\in\mathcal{C}} |F(z)|,$$

as neighbouring paths must possess a higher maximum. Such a path is called a *saddle–point path* or a *steepest descent path*. Thus, the search for a path realizing the minimum of

$$\min_{\mathcal{C}} \left[\max_{z\in\mathcal{C}} |F(z)| \right],$$

[1]Notice additionally that the optimization problem need not be solved exactly, as any approximate solution to (7) still furnishes a valid upper bound because of the universal character of the trivial bound (6).

(a simplification of (7) to its essential feature) naturally leads to considering saddle points and saddle-point paths. This leads to the variant of (7),

$$(8)\qquad |I| \le ||\mathcal{C}_0|| \cdot \max_{z\in\mathcal{C}_0} |F(z)|, \qquad \mathcal{C}_0 \text{ minimizes } \max_{z\in\mathcal{C}} |F(z)|,$$

also referred to as a *saddle point bound*.

We can summarize this stage of the discussion by a generic simple statement.

THEOREM VIII.2 (General saddle point bounds). *Let $F(z)$ be a function analytic in a domain Ω. Consider the class of integral $\int_\gamma F(z)\,dz$ where the contour γ connects two points A, B and is constrained to a class $\mathbf{P}$ of allowable paths in Ω (e.g., they may be constrained to encircle 0). Then one has the* saddle point bounds*:*

$$(9)\qquad \begin{array}{rcll} \displaystyle\left|\int_\gamma F(z)\,dz\right| & \le & \displaystyle\min_{\mathcal{C}\in\mathbf{P}} \left[||\mathcal{C}|| \cdot \max_{z\in\mathcal{C}} |F(z)|\right] & \text{(first form)} \\ & \le & \displaystyle ||\mathcal{C}_0|| \cdot \max_{z\in\mathcal{C}_0} |F(z)| & \text{(second form)} \\ & & \text{where } \mathcal{C}_0 \text{ minimizes } \displaystyle\max_{z\in\mathcal{C}} |F(z)|. & \end{array}$$

If A and B lie in opposite valleys of a saddle point z_0, then the best bounds of the second type are obtained as saddle point paths made of arcs connecting A to B through z_0.

The first form is *a priori* better than the second form as it encapsulates the length of the contour itself. The difference is however immaterial in virtually all asymptotic problems. This statement remains silent about topological details regarding the choice of the contour as it is clearly not possible to offer a universally valid criterion and the specific landscape of the modulus surface $|F(z)|$ under consideration has to be investigated. Fortunately, in cases of combinatorial interest some strong positivity is present and the selection of the suitable saddle point contour is normally greatly simplified.

$\triangleright$ **3.** *An integral of powers.* Consider the polynomial $P(z) = 1 + z + z^2 + z^3$ illustrated by Figure 1 and Equation (3). Define the line integral

$$I_n = \int_{-1}^{+i} P(z)^n\,dz.$$

On the segment connecting the end point, the maximum of $|P(z)|$ is 0.63831, giving the weak trivial bound $I_n = O(0.63831^n)$. In contrast, there is a saddle point at $z_0 = \frac{1}{3} + \frac{i}{3}\sqrt{2}$ where $|P(z_0)| = \frac{1}{3}$, resulting in the bound

$$|I_n| \le \lambda \left(\frac{1}{3}\right)^n, \qquad \lambda := |z_0 + 1| + |i - z_0| \doteq 1.44141,$$

as follows from adopting a contour made of two segments connecting -1 to i through z_0. $\triangleleft$

In the particular case of Cauchy coefficient integrals where $F(z) = g(z)z^{-n-1}$ and $g(z)$ is function with nonnegative coefficients, there is usually a saddle point on the positive real axis. We can then reexamine in a new light the bounds of Chapter IV. Assume for simplicity that $g(z)$, which has radius of convergence R with $0 < R \le +\infty$, satisfies $g(x) \to +\infty$ as $x \to R^-$. Then one has $F(0^+) = F(R^-) = +\infty$. This means that there is at least one positive value ζ such that the derivative $F'(x)$ of the real function $F(x)$ vanishes in $(0, R)$. (Actually, there can be only one such point, see below.) But this point ζ is also a derivative of the complex function $F(z)$. Since ζ is a local minimum, we have additionally $F''(\zeta) < 0$, and the saddle point is crossed transversally by a circle of radius ζ. Thus, the saddle point bound of the second kind (with circles centred at the origin constituting the allowable paths) instantiates to

$$[z^n]g(z) \le \frac{g(\zeta)}{\zeta^n}, \qquad \zeta \text{ root of } \zeta\,\frac{g'(\zeta)}{g(\zeta)} = n+1.$$

The bounds of the first kind are very similar and read

$$[z^n]g(z) \le \frac{g(\overline{\zeta})}{\overline{\zeta}^n}, \qquad \overline{\zeta} \text{ root of } \overline{\zeta}\,\frac{g'(\overline{\zeta})}{g(\overline{\zeta})} = n,$$

when optimization is carried out over circles centred at the origin.

We examine below two particular cases related to the central binomial and the inverse factorial. The corresponding landscapes in Figure 3 which bear a surprising resemblance to one another are, by the previous discussion, instances of a general pattern for functions with nonnegative coefficients.

$\triangleright$ **4.** *Upward convexity of* $g(x)x^{-n}$. For $g(z)$ having nonnegative coefficients at the origin, the quantity $g(x)x^{-n}$ is upward convex for $x > 0$, so that the saddle point equation for ζ (or $\overline{\zeta}$) can have at most one root. Indeed, the second derivative

$$\frac{d^2}{dx^2}\frac{g(x)}{x^n} = \frac{x^2 g''(x) - 2nxg'(x) + n(n+1)g(x)}{x^{n+2}}, \tag{10}$$

is positive for $x > 0$ since its numerator,

$$\sum_{k\ge 0}(n+1-k)(n-k)g_k x^k, \qquad g_k := [z^k]g(z),$$

has only nonnegative coefficients. $\triangleleft$

EXAMPLE 1. *Saddle point bounds for central binomials and inverse factorials.* Consider the two contour integrals around the origin,

$$J_n = \frac{1}{2i\pi}\oint (1+z)^{2n}\,\frac{dz}{z^{n+1}}, \qquad K_n = \frac{1}{2i\pi}\oint e^z\,\frac{dz}{z^{n+1}}, \tag{11}$$

whose values are otherwise known, by virtue of Cauchy's coefficient formula:

$$J_n = \binom{2n}{n}, \qquad K_n = \frac{1}{n!}.$$

In that case, with reference to Eq. (4), one can think of the end points A and B as coinciding and taken somewhat arbitrarily on the negative real axis while the contour has to encircle the origin once and counter–clockwise.

The landscapes of the two integrands are represented on Figure 3. The saddle point equations are respectively

$$\frac{2n}{1+z} - \frac{n+1}{z} = 0, \qquad 1 - \frac{n+1}{z} = 0.$$

The saddle points are thus respectively at

$$z_0 = \frac{n+1}{n-1}, \qquad z_0 = n+1.$$

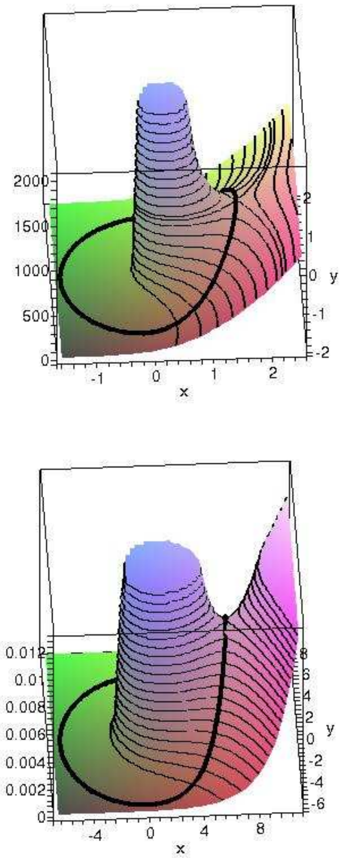


FIGURE 3. The modulus of the integrands of J_n (central binomials) and K_n (inverse factorials) for $n = 5$ and the corresponding saddle point contours.

This provides the upper bounds (of the second kind)

$$J_n \leq \left(\frac{4n^2}{n^2-1}\right)^n, \qquad K_n \leq \frac{e^{n+1}}{(n+1)^n}, \tag{12}$$

which are valid for all values $n \geq 2$. .... END OF EXAMPLE 1. $\square$

It is seen on these two examples that the saddle point bounds catch the proper exponential growths, being off only by a factor of $O(n^{-1/2})$. This is in fact a common phenomenon well explained by the saddle point method.

Borrowing a metaphor of De Bruijn [**93**], the situation may be described as follows. Estimating a path integral is like estimating the difference of altitude between two villages in a mountain range. If the two villages are in different valleys, the best strategy (this is what road networks often do) consists in following paths that cross boundaries between valleys at passes, *i.e.*, through saddle points.

VIII. 2.2. The saddle point method. Given a fixed contour $\mathcal{C}$ traversing a *simple* saddle point along its axis, the saddle point corresponds locally to a maximum of the integrand along the path. It is furthermore natural to expect that a small neighbourhood of the saddle point might provide the dominant contribution to the integral. The saddle point method is applicable precisely when this is the case and when this dominant contribution can be estimated by means of local expansions.

To proceed, it is convenient to set $F(z) = e^{f(z)}$and consider

$$I = \int_A^B e^{f(z)}\, dz,$$

where $f(z) \equiv f_n(z)$ like $F(z) \equiv f_n(z)$ involves some large parameter n. After possibly some preparation based on the use of Cauchy's theorem and suitable contours, we may assume that the contour $\mathcal{C}$ connects the end points A and B and is a path traversing a unique saddle point $z_0 \in \mathcal{C}$ along the steepest descent line. Thus, we have available the saddle point equation $F'(z_0) = 0$ or equivalently

$$f'(z_0) = 0,$$

as well as, by assumption, $|e^{f(A)}| < |e^{f(z_0)}|$ and $|e^{f(B)}| < |e^{f(z_0)}|$, meaning that z_0 is the highest point on the surface of $|F(z)|$.

The saddle point method is based on splitting $\mathcal{C}$ as $\mathcal{C} = \mathcal{C}^{(0)} \cup \mathcal{C}^{(1)}$, where $\mathcal{C}^{(0)}$ contains z_0, and estimating separately the integrals $\int_{\mathcal{C}^{(0)}}$ and $\int_{\mathcal{C}^{(1)}}$. In order for the method to work, conflicting requirements regarding the dimensioning of $\mathcal{C}^{(0)}$ and $\mathcal{C}^{(1)}$ must be satisfied, as we now explain.

— *Neglect the tail integrals.* The contour $\mathcal{C}^{(1)}$ should be taken so that the tail integral $\int_{\mathcal{C}^{(1)}}$ is negligible:

$$\int_{\mathcal{C}^{(1)}} F(z)\, dz = o\left(\int_{\mathcal{C}} F(z)\, dz\right). \tag{13}$$

Usually, this condition, rather than being checked *a priori*, results from global bounds relative to the function under consideration as well as from the choices dictated by the next two steps.

— *Centrally approximate the integrand.* The basic condition upon which the method depends is that locally, along $\mathcal{C}^{(0)}$, the two-term expansion

$$f(z) = f(z_0) = f(z_0) + \frac{1}{2} f''(z_0)(z - z_0) + O(\epsilon_n), \tag{14}$$

should be valid, with $\epsilon_n \to 0$ as $n \to \infty$. This is not automatically granted as F depends on n, as do usually z_0 and $f''(z_0)$. This condition *requires that $\mathcal{C}^{(0)}$ should be sufficiently small.* (Usually, $||\mathcal{C}^{(0)}||/||\mathcal{C}|| \to 0$.) The validity of the local expansion implies:

$$\int_{\mathcal{C}^{(0)}} F(z)\, dz \sim e^{f(z_0)} \int_{\mathcal{C}^{(0)}} e^{\frac{1}{2} f''(z_0)(z-z_0)^2}\, dz$$

— *Complete the tails.* Along the steepest descent line, the quantity $f''(z-z_0)(z-z_0)^2$ is negative. Then, in order to fully capture the contribution from the saddle point region, the Gaussian integral should be asymptotically equivalent to a complete Gaussian integral:

$$\int_{\mathcal{C}^{(0)}} e^{\frac{1}{2}f''(z_0)(z-z_0)^2}\,dz \sim i\int_{-\infty}^{\infty} e^{-\lambda x^2/2}\,dx \equiv i\sqrt{\frac{2\pi}{\lambda}}, \qquad \lambda = |f''(z_0)|\,, \tag{15}$$

where $\lambda := |f''(z_0)|$. This imposes *the conflicting requirement that* $\mathcal{C}^{(0)}$ *should be large enough*: at the end points of $\mathcal{C}^{(0)}$, one should have $f''(z_0)(z-z_0)^2$ tending to infinity. (Note: In (15) we have silently assumed that the change of variables results in the real line being traversed from negative to positive values; else a -1 factor should be inserted.)

The three main steps above are characteristic of the saddle-point method. They represent the complex-analytic counterparts of the method of Laplace (APPENDIX B: *Laplace's method*, p. 667) for the evaluation of real integrals depending on a large parameter. Indeed one can regard the saddle point method as being

Saddle Point Method = Choice of Contour + Laplace's Method.

Like its real-variable counterpart, the saddle point method is a a general strategy rather than a completely deterministic algorithm. Nonetheless, we choose to summarize the previous discussion by a Theorem:

THEOREM VIII.3 (Saddle Point Formula). *Consider an integral* $I = \int_A^B F(z)\,dz$ *where the integrand* F *is an analytic function depending on a large parameter and* A, B *lie in opposite valleys across a saddle point* z_0. *Set* $F(z) = e^{f(z)}$. *Then, if the saddle point contour* $\mathcal{C}$ *connecting* A *to* B *can be split into* $\mathcal{C} = \mathcal{C}^{(0)}) \cup \mathcal{C}^{(1)}$ *in such a way that*

(i) *Tails are negligible, cf. Equation* (13),
(ii) *Central approximations hold, cf. Equation* (14),
(iii) *Tails can be completed back, cf. Equation* (15),

then one has

$$\frac{1}{2i\pi}\int_A^B e^{f(z)}\,dz \sim \frac{e^{f(z_0)}}{\sqrt{2\pi|f''(z_0)|}}.$$

The following remarks may help implement the saddle point method. In many cases, the error in the two-term expansion is likely to be given by the next term, which involves a third derivative. In that case, it is good guess to attempt to dimension $\mathcal{C}^{(0)}$ as being of length $\delta \equiv \delta(n)$ chosen in such a way that

$$f'''(z_0)\delta^3 \to 0, \qquad f''(z_0)\delta^2 \to \infty, \tag{16}$$

so that both local approximation and tail completion can be satisfied. We call this choice the saddle point *dimensioning heuristic*. Also, in many cases, it proves convenient to adopt paths that come very close to the saddle point but need not pass exactly through it. Similarly, the steepest descent line may be followed only approximately. These comments apply for instance for functions $F_n(z)$ of the form $F_n(z) = H_n(z)g(z)$ (e.g., $H_n(z) = h(z)^n$, a large power of a fixed function) where

g does not depend on n and is well-behaved—it is then computationally convenient to treat g as a perturbation and use a contour dictated by H_n alone.

As an illustration, a blind application of the conclusion of Theorem VIII.3 to J_n and K_n of Example 1 gives for $f(z)$ respectively

$$2n\log(1+z)-(n+1)\log z, \qquad z-(n+1)\log z,$$

which results in the correct asymptotic equivalents for J_n and K_n:

$$J_n \sim \frac{4^n}{\sqrt{\pi n}}, \qquad K_n \sim n^n e^{-n}\sqrt{2\pi n}.$$

In the sequel, we make use of the general principles of saddle point analysis but focus on the particular case of Cauchy coefficient integrals for generating functions with positive coefficients. The geometry of the problem is simpler in that case since, as we saw, it usually suffices to consider as integration contour a circles with proper radius centred at the origin and passing through a positive real saddle point.

EXAMPLE 2. *Saddle point analysis of the exponential and the inverse factorial.* The purpose of this example is to provide a concrete illustration of saddle point analysis by working out the problem of estimating $\frac{1}{n!}=[z^n]e^z$. The starting point is the Cauchy coefficient integral

$$K_n=\frac{1}{2i\pi}\int_{|z|=r} e^z\,\frac{dz}{z^{n+1}},$$

where the contour of integration is taken to be a circle of radius r. We carry out the saddle point strategy with its usual three steps: (i) *Neglect the tails*; (ii) *Centrally approximate*; (iii) *Complete the tails*. The landscape of the modulus of the integrand is displayed in Figure 3. There is a saddle point at $z=n+1$ with an axis perpendicular to the real line. We thus expect good bounds to derive from adopting as integration contour a circle centered at the origin with radius $n+1$ (or about) as integration contour. We adopt a circle of radius $r=n$ as integration contour ($n+1$ would do equally well but would slightly complicate calculations). Also, it proves convenient to switch to polar coordinates and set $z=ne^{i\theta}$.

(i) *Neglect the tails.* For $z=ne^{i\theta}$ one has

$$|e^z|=e^{n\cos\theta},$$

and since the cosine function is unimodal on $[-\pi,+\pi]$, the contribution of any part of the contour outside a circular arc defined by its endpoints $ne^{\pm i\theta_0}$ is

$$O\left(ne^n n^{-n} e^{-n\cos\theta_0}\right). \tag{17}$$

Thus this contribution is exponentially small compared to $e^n n^{-n}$, when θ_0 is any fixed number $\theta_0<0$, but also as long as θ_0 is a function of n satisfying

$$n\cos(\theta_0) > C\log n,$$

for any $C>0$. A definite choice is fixed in the next phase.

(ii) *Centrally approximate.* The original integral in polar coordinates becomes

$$K_n=\frac{e^n}{n^n}\cdot\frac{1}{2\pi}\int_{-\pi}^{+\pi} e^{n[e^{i\theta}-1-i\theta]}\,d\theta. \tag{18}$$

Set $h(\theta)=e^{i\theta}-1-i\theta$ with expansion as $\theta\to 0$:

$$h(\theta)=-\frac{\theta^2}{2}-\frac{i\theta^3}{6}+\frac{\theta^4}{24}+\cdots.$$

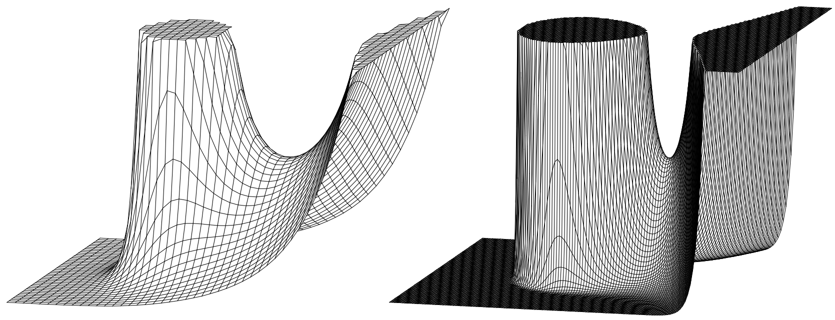


FIGURE 4. Plots of $|e^z z^{-n-1}|$ for $n = 3$ and $n = 30$ (scaled according to the value of the saddle point) illustrate the essential concentration condition as higher values of n produce steeper saddle point paths.

The absence of a linear term in θ indicates a saddle point. The function

$$|e^{h(\theta)}| = e^{\cos\theta - 1},$$

is unimodal with its peak at $\theta = 0$ and the same property holds for $|e^{nh(\theta)}|$ which is even more strongly peaked at $\theta = 0$.

The estimation of K_n proceeds by isolating a small portion of the contour (corresponding to z near the real axis). We thus set

$$K_n^{(0)} = \int_{-\theta_0}^{+\theta_0} e^{nh(\theta)}\, d\theta, \qquad K_n^{(1)} = \int_{\theta_0}^{2\pi-\theta_0} e^{nh(\theta)}\, d\theta,$$

and choose θ_0 in accordance with the general heuristic (16):

$$n\theta_0^2 \to \infty, \qquad n\theta_0^3 \to 0. \tag{19}$$

One way of realizing the compromise is to adopt $\theta_0 = n^a$ where a is any number between $-\frac{1}{2}$ and $-\frac{1}{3}$, for instance

$$\theta_0 \equiv \theta_0(n) = n^{-2/5}.$$

Thus the angle of the central region tends to zero; in the notations of our general discussion in (16), one has $\delta \sim n\theta_0$, that is, $\delta \sim n^{3/5}$. With this choice of θ_0, the bound (17) instantiates to

$$K_n^{(1)} = O\left(\exp\left(-Cn^{1/5}\right)\right), \qquad C > 0. \tag{20}$$

We now turn to the precise evaluation of the central integral $K_n^{(0)}$. Near $\theta = 0$, only the terms till order 2 matter in the expansion of $h(\theta)$ because of the second condition of (19), which ensures $n\theta^3 \to 0$ throughout the central region. One has:

$$\begin{aligned} K_n^{(0)} &\sim \int_{-\theta_0}^{+\theta_0} e^{-n\theta^2/2}\, d\theta \\ &\sim \frac{1}{\sqrt{n}} \int_{-\theta_0\sqrt{n}}^{+\theta_0\sqrt{n}} e^{-t^2/2}\, dt \end{aligned} \tag{21}$$

The first line of (21) uses the fact that $n\theta^3 \to 0$ so that $h(\theta)$ can be reduced to its quadratic approximation, with error terms of order $n\theta_0^3 = n^{-1/5}$; the second line is based on the rescaling $t = \theta\sqrt{n}$.

(iii) *Complete the tails*. We have

$$K_n^{(0)} \sim \frac{1}{\sqrt{n}} \int_{-\infty}^{+\infty} e^{-t^2/2}\, dt \sim \sqrt{\frac{2\pi}{n}}. \tag{22}$$

The first line is justified by the fact that the tails of the Gaussian integral in (21) are exponentially small. ($O(\exp(-Cn^{1/5}))$, again) Accordingly the integral can be completed to the full range $(-\infty, +\infty)$, which induces error terms that are exponentially small anyhow. Finally, the complete Gaussian integral evaluates to closed form and the estimate follows.

Assembling (20) and (22), we have obtained

$$K_n^{(0)} + K_n^{(1)} \sim \sqrt{\frac{2\pi}{n}}.$$

Hence the final result

$$K_n \equiv \frac{1}{n!} \sim \frac{e^n}{n^n\sqrt{2\pi n}}.$$

We have thus established Stirling's formula by the saddle point method. END OF EXAMPLE 2. □

VIII. 2.3. Complete asymptotic expansions. Like Laplace's method, the saddle point method can normally be made to provide full asymptotic expansions. The idea is still to localize the contribution in the same central region as the one that gave rise to the first-order saddle point estimate but take into account the corrections terms to the quadratic approximation. We make explicit here the calculations relative to the inverse factorial.

It suffices to revisit the estimation of $K^{(0)}$ since $K^{(1)}$ is exponentially small. One first rewrites

$$\begin{aligned} K_n^{(0)} &= \int_{-\theta_0}^{\theta_0} e^{-n\theta^2/2} e^{n(\cos\theta - 1 - \frac{1}{2}\theta^2)}\, d\theta \\ &= \frac{1}{\sqrt{n}} \int_{-\theta_0\sqrt{n}}^{\theta_0\sqrt{n}} e^{-w^2/2} e^{n\xi(w/\sqrt{n})}\, dw, \qquad \xi(\theta) := \cos\theta - 1 - \frac{1}{2}\theta^2. \end{aligned}$$

The calculation proceeds exactly in the same way as for the Laplace method (APPENDIX B: *Laplace's method*, p. 667). It suffices to expand $h(\theta)$ to any fixed order, which is legitimate in the central region. In this way, a representation of the form,

$$K_n^{(0)} = \frac{1}{\sqrt{n}} \int_{-\theta_0\sqrt{n}}^{\theta_0\sqrt{n}} e^{-w^2/2} \left(1 + \sum_{k=1}^{M-1} \frac{E_k(w)}{n^{k/2}} + O\left(\frac{1+w^{3M}}{n^{M/2}}\right)\right) dw,$$

is obtained, where the $E_k(w)$ are computable polynomials of degree $3k$. Distributing the integral operator over terms in the asymptotic expansion and completing the tails yields an expansion of the form

$$K_n^{(0)} \sim \frac{1}{\sqrt{n}} \left(\sum_{k=0}^{M-1} \frac{d_k}{n^{k/2}} + O(n^{-M/2})\right),$$

where

$$d_k := \int_{-\infty}^{+\infty} e^{-w^2/2} E_k(w)\, dw.$$

in which all odd terms disappear by parity and $d_0 = \sqrt{2\pi}$. The net result is here:

$$\frac{1}{n!} \sim \frac{e^n n^{-n}}{\sqrt{2\pi n}} \left(1 - \frac{1}{12n} + \frac{1}{288\, n^2} + \frac{139}{51840\, n^3} - \frac{571}{2488320\, n^4} + \cdots\right).$$

(Notice the amazing similarity with the form obtained directly for $n!$ in APPENDIX B: *Laplace's method*, p. 667.)

▷ **5.** *A factorial surprise.* Why is it that the expansion of $n!$ and $1/n!$ involve the same set of coefficients, up to sign? [Hint: the derivations involve similar integrals, but taken along different paths; similarity of the coefficients results from basic properties of Hankel contours.] ◁

In summary the process of saddle point analysis is made possible by a fundamental *split* of the integration contour —here, a circle— into a small arc centered on the real axis. The small arc has to satisfy two conflicting requirement: to be large enough so as to capture the essential contribution of the integral; and to be small enough so as to allow the function to be well approximated locally by its quadratic terms. In addition, the estimation is made possible because the function decays appropriately, away from the real axis, so that the integrand remains small on the noncentral part of the contour. Since only a central region matters (up to exponentially small error terms), compete asymptotic expansions can usually be derived.

VIII. 3. Large powers

The extraction of coefficients in powers of a fixed function and more generally of functions of the form $A(z)B(z)^n$ constitutes a prototypical and easy application of the saddle point method. We wil thus be concerned here with the problem of estimating

$$[z^N]A(z) \cdot B(z)^n = \frac{1}{2i\pi} \oint A(z)B(z)^n \frac{dz}{z^{n+1}}.$$

This situation generalizes directly the example of inverse factorials and the exponential, where we have dealt with $[z^n](e^z)^n$ as well as the case of the central binomial coefficients where an estimate of $[z^n](1+z)^{2n}$ is wanted. On another register, the Lagrange inversion theorem expresses the coefficients of certain implicitly defined functions in terms of coefficients of powers, so that the techniques exposed here are also relevant to the study of trees, forests, functional graphs and maps. Further important applications relate to large deviations and local limit laws explored in Section VIII. 7 as well as in the next chapter.

We consider in this section two fixed functions, $A(z)$ and $B(z)$ satisfying the following conditions:

C_1: The functions $A(z) = \sum_{j\ge 0} a_j z^j$ and $B(z) = \sum_{j\ge 0} b_j z^j$ are analytic at 0 and have nonnegative coefficients; furthermore it is assumed (without loss of generality) that $B(0) \neq 0$.

C_2: The function $B(z)$ is *aperiodic* in the sense that $\gcd\{j \mid b_j > 0\} = 1$. (Thus $B(z)$ is not a function of the form $\beta(z^p)$ for some integer $p > 0$.)

C_3: Let R be the radius of convergence $R \leq \infty$ of $B(z)$; the radius of convergence of $A(z)$ is at least as large as R.

We introduce the following quantity called the *spread*:

$$T := \lim_{x \to R^-} \frac{xB'(x)}{B(x)}.$$

Our purpose is to analyse the coefficients

$$[z^n]\, A(z) \cdot B(z)^n,$$

when N and n are linearly related. The condition $N < Tn$ is both technically needed in our proof and inherent in the nature of the problem. (For B a polynomial of degree d, the spread is $T = d$; for a function B whose derivative at its dominant positive singularity remains bounded, the spread is finite; for $B(z) = e^z$ and more generally for entire functions, the spread is $T = \infty$.)

Saddle-point bounds. First the saddle point bounds come out immediately:

PROPOSITION VIII.1 (Saddle point bounds for large powers). *Consider functions $A(z)$ and $B(z)$ satisfying the conditions C_1, C_2, C_3 above. Let λ be fixed a positive number with $0 < \lambda < T$ and let ζ be the unique positive root of the equation*

$$\zeta \frac{B'(\zeta)}{B(\zeta)} = \lambda.$$

Then, with $N = \lambda n$ an integer, one has

$$[z^N]A(z) \cdot B(z)^n \leq A(\zeta)B(\zeta)^n \zeta^{-N}.$$

PROOF. The existence and unicity of ζ is guaranteed by an argument encountered in Chapter IV in the context of enumerating simple varieties of trees. The rest follows by an immediate application of general saddle point bounds. $\Box$

As a first application, consider the problem of estimating the binomial coefficients $\binom{n}{\lambda n}$ for some λ with $0 < \lambda < 1$. We assume for convenience that λn is an integer and write $N = \lambda n$. Proposition VIII.1 provides

$$\binom{n}{\lambda n} = [z^N](1+z)^n \leq (1+\zeta)^n \zeta^{-N},$$

where $\frac{\zeta}{1+\zeta} = \lambda$, i.e., $\zeta = \frac{\lambda}{1-\lambda}$. A simple computation then shows that

$$2^{-n}\binom{n}{\lambda n} \leq \exp(nH(\lambda)), \qquad \text{where} \qquad H(\lambda) = -\lambda \log \lambda - (1-\lambda)\log(1-\lambda)$$

is the *entropy function*. Thus, for $\lambda \neq \frac{1}{2}$, the binomial coefficients $\binom{n}{\lambda n}$ are exponentially smaller than the central coefficient $\binom{n}{n/2}$, and the entropy function precisely quantifies this exponential gap.

$\triangleright$ **6.** *Anomalous dice games.* The probability of a score equal to λn in n casts of an unbiased die is bounded from above by a quantity of the form e^{-nH} where

$$H = -6 + \log\left(\frac{1-\zeta^6}{1-\zeta}\right) - (\lambda - 1)\log\zeta,$$

and ζ is an algebraic function of λ determined by $\sum_{j=0}^{5}(\lambda - j)\zeta^j = 0$. $\triangleleft$

$\triangleright$ **7.** *Large deviation bounds for sums of random variables.* Let $g(u) = \mathbb{E}(u^X)$ be the probability generating function of a discrete random variable $X \geq 0$ and let $\mu = g'(1)$ be the corresponding mean (assume $\mu < \infty$). Set $N = \lambda n$ and let ζ be the root of $\zeta g'(\zeta)/g(\zeta) = \lambda$ assumed to exist within the domain of analyticity of g. Then, for $\lambda < \mu$, one has

$$\sum_{k \leq N} [u^k] g(u)^n \leq \frac{1}{1-\zeta} g(\zeta)^n \zeta^{-N}.$$

Dually, for $\lambda > \mu$, one finds

$$\sum_{k \geq N} [u^k] g(u)^n \leq \frac{\zeta}{\zeta - 1} g(\zeta)^n \zeta^{-N}.$$

These are exponential bounds on the probability that n copies of the variable X have a sum deviating substantially from the expected value. $\triangleleft$

The saddle point bounds for large powers are technically shallow but still useful if only rough order-of-magnitude estimates are sought. The full saddle point method is in fact applicable under the conditions of Proposition VIII.1.

THEOREM VIII.4 (Saddle point analysis for large powers). *Under the conditions of the preceding proposition, one has*

$$[z^N] A(z) \cdot B(z)^n = A(\zeta) \frac{B(\zeta)^n}{\zeta^{N+1}\sqrt{2\pi n \xi}} (1 + o(1)). \tag{23}$$

where ζ is the unique root of $\zeta B'(\zeta)/B(\zeta) = \lambda$ and

$$\xi = \frac{d^2}{d\zeta^2} \left(\log B(\zeta) - \lambda \log z \right).$$

In addition, a full expansion in descending powers of n exists.

These estimates hold uniformly for λ in any compact interval of $(0, T)$, i.e., any interval $[\lambda', \lambda'']$ with $0 < \lambda' < \lambda'' < T$, where T is the spread.

Note. We have opted for a basic formulation of the theorem with conditions on A and B that are not minimal. It is easily recognized that the estimates of Theorem VIII.4 still hold provided that *the function $|B(re^{i\theta})|$ attains a unique maximum on the positive real axis, when $r \in (0, T)$ is fixed and θ varies on $[-\pi, \pi]$*. Also, in order for the statement to hold true, it is only required that *the function $A(z)$ does not vanish on* $(0, T)$, and $A(z)$ could then well be allowed to have negative coefficients. Finally, if $A(\zeta) = 0$, then a simple modification of the argument still provides precise estimates in this vanishing case; see Note 10 below.

PROOF. We discuss the analysis corresponding to a fixed λ. The function $|B(re^{i\theta})|$ is, by positivity of coefficients *and* aperiodicity, uniquely maximal at $\theta = 0$ for any fixed r. It is also infinitely differentiable at 0. Consequently there exists a (small) angle $\theta_1 \in (0, \pi)$ such that

$$|B(re^{i\theta})| \leq |B(re^{i\theta_1})| \qquad \text{for all } \theta \in [\theta_1, \pi],$$

and at the same time, $|B(re^{i\theta})|$ is strictly decreasing on $[0, \theta_1]$ (it is given by a Taylor expansion without linear term).

We carry out the integration along the saddle point circle, $z = \zeta e^{i\theta}$, where the previous inequalities on $|B(z)|$ hold. The contribution for $|\theta| > \theta_1$ is exponentially negligible. Thus, up to exponentially small terms, the sought coefficient is given asymptotically by $J(\theta_1)$, where

$$J(\phi) = \frac{1}{2\pi}\int_{-\phi_1}^{\phi_1} A(\zeta e^{i\phi})B(\zeta e^{i\phi})e^{ni\phi}\,d\phi.$$

It is then possible to impose a *second* restriction on θ, by introducing θ_0 according to the general heuristic, namely, $n\theta_0^2 \to \infty$, $n\theta_0^3 \to 0$. We fix here

$$\theta_0 \equiv \theta_0(n) = n^{-2/5}.$$

By the decrease of $|B(\zeta e^{i\theta})|$ on $[\theta_0, \theta_1]$ and by local expansions, the quantity $J(\theta_1) - J(\theta_0)$ is of the form $\exp(-cn^{1/5})$ for some $c > 0$, that is, exponentially small.

Finally, local expansion are valid in the central range since θ_0 tends to 0 as $n \to \infty$. One finds for $z = \zeta e^{i\theta}$ and $|\theta| \le \theta_0$,

$$A(z)B(z)^n \sim A(\zeta)B(\zeta)^n\zeta^{-N}\exp(-n\xi\theta^2/2).$$

Then the usual process applies upon completing the tails, resulting in the stated estimate. The existence of a complete expansion in powers of $n^{-1/2}$ results from pushing the expansion of $\log B(z)$ to an arbitrary order (like in the case of Stirling's formula). Finally, by parity all the odd integrals vanish so that the expansion turns out to be in descending powers of n. $\square$

An immediate application of Theorem VIII.4 is to the central binomial coefficient $\binom{2n}{n} = [z^n](1+z)^{2n}$. In the same way, one gets an estimate of the central trinomial number,

$$T_n := [z^n](1+z+z^2)^n \qquad \text{satisfying} \quad T_n \sim \frac{3^{n+1/2}}{2\sqrt{\pi n}}.$$

The Motzkin numbers count unary-binary trees, so that

$$M_n = [z^n]M(z) \qquad \text{where} \qquad M = z(1+M+M^2).$$

The standard approach is the one seen earlier based on singularity analysis as the implicitly defined function $M(z)$ has an algebraic singularity of the $\sqrt{}$-type, but the Lagrange inversion formula provides an equally workable route. It gives

$$M_{n+1} = \frac{1}{n+1}[z^n](1+z+z^2)^{n+1},$$

which is amenable to saddle point analysis via Theorem VIII.4. Hence,

$$M_n \sim \frac{3^{n+1/2}}{2\sqrt{\pi n^3}}.$$

▷ **8.** *Central Stirling numbers.* The central Stirling numbers of both kinds satisfy

$$\frac{n!}{(2n)!}\begin{bmatrix}2n\\n\end{bmatrix} \sim c_1 A_1^n n^{-1/2}\left(1+O(n^{-1})\right), \qquad \frac{n!}{(2n)!}\begin{Bmatrix}2n\\n\end{Bmatrix} \sim c_2 A_2^n n^{-1/2}\left(1+O(n^{-1})\right),$$

where $A_1 \doteq 2.45540$, $A_2 \doteq 1.54413$, and A_1, A_2 are expressible in terms of special values of the Cayley tree function. Similar estimates hold for $\left[{\alpha n \atop \beta n}\right]$ and $\left\{{\alpha n \atop \beta n}\right\}$. ◁

▷ **9.** *Integral points on high-dimensional spheres.* This note is based on an article by Mazo and Odlyzko [**309**]. Let $N(n,\alpha)$ be the number of lattice points (i.e., points with integer coordinates) in n-dimensional space that lie *on* the sphere of radius $N = \sqrt{\alpha n}$ assumed to be an integer. Then,

$$N(n,\alpha) = [z^N]\Theta(z)^n, \qquad \text{where} \quad \Theta(z) := \sum_{n\in\mathbb{Z}} z^{n^2} = 1 + 2\sum_{n=1}^{\infty} z^{n^2}.$$

Thus, there are computable constants K, L depending on a such that $N(n,\alpha) \sim Kn^{-1/2}L^n$. The number of lattice points *inside* the sphere can be similarly estimated. (Such bounds are useful in coding theory, combinatorial optimization, especially the knapsack problem, and cryptography.) ◁

▷ **10.** *Coalescence of a saddle point with roots of multipliers.* Fix ζ and take a multiplier $A(z)$ in Theorem VIII.4 such that $A(\zeta) = 0$. The formula is to be modified,

$$[z^N]A(z)\cdot B(z)^n = \left[A'(\zeta) + \zeta A''(\zeta)\right] \frac{B(\zeta)^n}{\zeta^{N+1}\sqrt{2\pi n^3 \xi^3}}(1+o(1)).$$

Higher order cancellations can also be taken into account. ◁

▷ **11.** *A function with negative coefficients that is minimal along the positive axis.* Take $B(z) = 1 + z - z^{10}$ with $|z| \le \frac{1}{10}$. By design, $B(z)$ has negative Taylor coefficients. On the other hand, $|B(re^{i\theta})|$ for fixed $r \le \frac{1}{10}$ (say) attains its unique maximum at $\theta = 0$. In such a case, the saddle point method applies and an estimate of $[z^n]B(z)^n$ (say) is obtained by (23). ◁

Large powers, saddle points, and singularity analysis. In general, the Lagrange inversion formula establishes an exact correspondence between two problems relative to the estimation of

— coefficients of large order in large powers and
— coefficients of implicitly defined functions.

In one direction, the Lagrange Inversion Theorem has the capacity of bringing the evaluation of coefficients of implicit functions into the orbit of the saddle point method. We obtain in this way a statement that paraphrases Theorem VII.2 on page 429: *Let Y be defined implicitly by $Y = z\phi(Y)$, where ϕ is analytic at 0, aperiodic, and such that the characteristic equation $\phi(\tau) - \tau\phi'(\tau) = 0$ has a positive root within the disc of convergence of ϕ. Then*

$$[z^n]Y(z) \sim \gamma\frac{\rho^{-n}}{2\sqrt{\pi n^3}}, \qquad \rho := \frac{\tau}{\phi(\tau)}, \quad \gamma := \sqrt{\frac{2\phi(\tau)}{\phi''(\tau)}}.$$

(This provides the number of trees in a simple variety, with ϕ being the degree generating function of the variety.)

The saddle point method is in a few cases more convenient to work with than singularity analysis, especially when explicit or uniform upper bounds are required, since constructive bounds tend to be more easily obtained on fixed circles than on variable Hankel contours.

$\triangleright$ **12.** *An assertion of Ramanujan.* In his first letter to Hardy, Ramanujan (1913) announced that

$$\frac{1}{2}e^n = 1 + \frac{n}{1!} + \frac{n^2}{2!} + \cdots + \frac{n^{n-1}}{(n-1)!} + \frac{n^n}{n!}\theta,$$
$$\text{where} \qquad \theta = \frac{1}{3} + \frac{4}{135(n+k)},$$

and k lies between $\frac{8}{45}$ and $\frac{2}{21}$. Ramanujan's assertion indeed holds for all $n \geq 1$; see [**156**] for a proof based on saddle points and effective bounds. $\triangleleft$

Conversely, the Lagrange Inversion Theorem makes it possible to approach large powers problems by means of singularity analysis of an implicitly defined function[2]. This mode of operation can prove very useful when there occurs a coalescence between saddle points and singularities of the integrand.

$\triangleright$ **13.** *Coalescence between a saddle-point and a singularity.* The integral in

$$I_n := [y^n](1+y)^{1n}(1-y)^{-\alpha} = \frac{1}{2i\pi}\int_{0+} \frac{(1+y)^{2n}}{(1-y)^{\alpha}}\,dy,$$

can be treated directly, but this requires a suitable adaptation of the saddle-point method, given the coalescence between a saddle point at 1 [the part without the $(1-y)^{\alpha}$ factor] and a singularity at that same point. Alternatively, it can be subjected to the change of variables $z = y/(1=y)^2$. Then y is defined implicitly by $y = z(1+y)^2$ and subject to singularity analysis, so that

$$I_n = \frac{1}{2i\pi}\int_{0+} \frac{1+y}{(1-y)^{1+\alpha}}\,\frac{dz}{z^{n+1}} = [z^n]\frac{1+y}{(1-y)^{1+\alpha}}.$$

Since $y(z)$ has a square-root singularity at $z = 1/4$, the integrand is of type $Z^{-(1+\alpha)/2}$, and

$$I_n \sim \frac{2^{2n-\alpha}}{\Gamma(\frac{\alpha+1}{2})} n^{(\alpha-1)/2}.$$

In general, for $\phi(y)$ satisfying the assumptions (relative to B) of Theorem VIII.4, one finds $(\tau : \phi(\tau) - \tau\phi'(\tau) = 0)$

$$\frac{1}{2i\pi}\int_{0+} \frac{\phi(y)^n}{(\phi(\tau)-\phi(y))^{\alpha}}\,\frac{dy}{y^n} \sim c\left(\frac{\phi(\tau)}{\tau}\right)^n \frac{n^{(\alpha-1)/2}}{\Gamma(\frac{\alpha+1}{2})}.$$

Van der Waerden discuses this problem systematically in [**420**]. See also Section VIII. 7.2 below for other coalescence situations. $\triangleleft$

VIII. 4. Four combinatorial examples

After a detour through generalities, we are now fully equipped to come back to analytic combinatorics. In this section, we examine four combinatorial examples,

Involutions $(\mathcal{I})$, *Set partitions* $(\mathcal{S})$,
Fragmented permutations $(\mathcal{F})$, *Integer partitions* $(\mathcal{P})$.

---

[2] This is in essence an approach suggested by several sections of the original memoir of Darboux[**88**, §3–§5], in which "Darboux's method" discussed in Chapter VI was first proposed. It is also of interest to note that a Lagrangean change of variables transforms a saddle point contour in a contour approximately of the type used in singularity analysis.

The generating functions I, S, F are EGFs while P is an OGF. They have all been made explicit in Chapters I and II:

$$I(z) = e^{z+z^2/2}, \quad S(z) = e^{e^z-1}, \quad F(z) = e^{\frac{z}{1-z}}, \quad P(z) = \prod_{j=1}^{\infty} \frac{1}{1-z^j}.$$

The first two are entire functions (i.e., they only have a singularity at ∞), the last two have a singularity at $z = 1$, but in each case, they exhibit a violent growth—of an exponential type—near their positive singularity at either a finite or infinite distance.

Each example is treated, starting from the easier saddle point bounds and proceeding with the saddle point method. The allowable paths considered here are invariable circles centred at the origin, so that it is the variation of the generating function along such circles that matters, in accordance with our earlier discussion of saddle points and Cauchy coefficient integrals. Each example illustrates the general principles and at the same time demonstrates some specific twists of the method. For instance, the example of involutions shows what may happen when an approximate saddle point is traversed and linear terms modify the Gaussian integral. Bell numbers illustrate the need of a good asymptotic technology for an implicitly defined saddle point [**375**]. Integer partitions exemplify a situation where the generating function possesses exceptionally rich properties.

EXAMPLE 3. *Involutions.* Let I_n be the number of involutions of $[1..n]$, that is the number of permutations τ such that τ^2 is the identity permutation. The egf of I_n is $I(z) = e^{z+z^2/2}$. The exact value of the positive saddle point of the Cauchy integral

$$\frac{I_n}{n!} = \frac{1}{2i\pi} \oint I(z) \frac{dz}{z^{n+1}}$$

is the positive root ζ of $\zeta(1+\zeta) = n+1$, that is,

$$\zeta = \frac{-1+\sqrt{4n+5}}{2} = \sqrt{n} - \frac{1}{2} + \frac{5}{8\sqrt{n}} + O(n^{-3/2}).$$

By routine asymptotic computations, the saddle point bound becomes

$$\frac{I_n}{n!} \le e^{-1/4} n^{-n/2} e^{n/2+\sqrt{n}} (1+o(1)).$$

Notice that if we use the approximate saddle point value, $\widehat{\zeta}(n) = \sqrt{n}$, we only lose the factor $e^{-1/4} \doteq 0.77880$.

In agreement with the discussion in the introduction these bounds are expected to be quite good, and we shall see later that they only off by a factor of $O(n^{1/2})$ from the true asymptotic form of I_n given by Knuth in [**264**] (where the derivation is carried out by means of the Laplace method for sums). In fact, Moser and Wyman (1955) have shown that this results from the saddle point method, upon integrating along a circle of radius $\zeta(n)$ (or about). We set

$$h(z) := \log I(z) - (n+1)\log z = z + \frac{z^2}{2} - (n+1)\log z.$$

First, some general considerations are in order regarding the behaviour of $|I(z)|$ along large circles, $z = re^{i\theta}$. One has

$$\log |I(re^{i\theta})| = r\cos\theta + \frac{r^2}{2}\cos 2\theta.$$

As a function of θ, this function decreases on $(0, \frac{\pi}{2})$ as it is the sum of two decreasing functions. Thus, $|I(z)|$ attains its maximum $(e^{r+r^2/2})$ at r and its minimum $(e^{-r^2/2})$ at $z = ri$. In the left half plane, first for $\theta \in (\frac{\pi}{2}, \frac{3\pi}{4})$, the modulus $|I(z)|$ is at most e^r since $\cos 2\theta < 0$. Finally, for $\theta \in (\frac{3\pi}{4}, \pi)$ smallness is granted by the fact that $\cos\theta < -1/\sqrt{2}$ resulting in the bound $|I(z)| \le e^{r^2/2 - r/\sqrt{2}}$. The same argument applies to the lower half plane $\Im(z) < 0$ to the effect that $I(z)$ is strongly peaked at $z = r$ and is exponentially small (by a factor of the form e^{-cr}) away from it. This preliminary discussion will enable us to neglect the tails of the Cauchy integral.

We now fix the radius as $r = \sqrt{n}$ as this is convenient for calculations while being extremely close to the actual saddle point. We specify the central region as defined by $|\theta| \le \theta_0$, where $\theta_0 \equiv \theta_0(n)$ is determined according to the general heuristic (16), which, under polar coordinates, reads as

$$r^2\theta_0^2 h''(r) \to +\infty, \qquad r^3\theta_0^3 h'''(r) \to 0.$$

Since one has $r = \sqrt{n}$ and

$$\begin{array}{ccc} h'(z) = 1 + z - \dfrac{n}{z}, & h''(z) = 1 + \dfrac{n}{z^2}, & h'''(z) = -\dfrac{2n}{z^3}, \\ h'(\sqrt{n}) = 1 & h''(\sqrt{n}) = 2, & h'''(\sqrt{n}) = -\dfrac{2}{\sqrt{n}}, \end{array} \tag{24}$$

the heuristic suggests to take

$$\theta_0^2 \gg \frac{1}{n}, \qquad \theta_0^3 \ll \frac{1}{n},$$

i.e., θ_0 should be chosen of an order between $n^{-1/2}$ and $n^{-1/3}$. We fix here

$$\theta_0 = n^{-2/5},$$

from which it is easily checked that tails are exponentially small, i.e., of the form $\exp(-n^\alpha)$ for some $\alpha > 0$, by virtue of the previous paragraph.

We then proceed and consider the central integral ($r = \sqrt{n}$)

$$J_n^{(0)} = \frac{e^{h(r)}}{2\pi} \int_{-\theta_0}^{+\theta_0} \exp\Big(h(re^{i\theta}) - h(r)\Big)\, d\theta.$$

What is required is a Taylor expansion with remainder near the point $\sqrt{n}$. Observe that in the central region, we have $h'''(z) = O(n^{-1/2})$, this by (24. Thus, with still $r = \sqrt{n}$,

$$h(z) - h(r) = h'(r)(z-r) + \frac{1}{2}h''(r)(z-r)^2 + O(n^{-1/2}(z-r)^3).$$

The first order term is present as we are not passing exactly through the saddle point. Since $z - r = r(e^{i\theta} - 1) \sim ir\theta$ and $|z - r| = O(n^{1/10})$, one finds

$$h(z) - h(r) = ir\theta + r\theta^2 + O(n^{-1/5}).$$

This is enough to guarantee that

$$J_n^{(0)} = \frac{e^{h(\sqrt{n})}}{2\pi} \int_{-\theta_0}^{+\theta_0} e^{i\sqrt{n}\theta - n\theta^2}\, d\theta \left(1 + O(n^{-1/5})\right).$$

Then the tails can be completed in the usual way, leading to

$$J_n^{(0)} = \frac{e^{h(\sqrt{n})}}{2\pi} \int_{-\infty}^{+\infty} e^{i\sqrt{n}\theta - n\theta^2/2}\, d\theta \left(1 + O(n^{-1/5})\right).$$

This integral evaluates to closed form by completing the square:

$$\int_{-\infty}^{+\infty} e^{a\theta-n\theta^2}\,d\theta = e^{-a^2/(4n)}\int_{-\infty}^{+\infty} e^{-\eta^2}\,d\eta,$$

and using Cauchy's theorem whenever a is complex. This gives

$$\frac{I_n}{n!} = \frac{e^{-1/4}}{2\sqrt{\pi n}} n^{-n/2} e^{n/2+\sqrt{n}} \left(1 + O(\frac{1}{n^{1/5}})\right), \tag{25}$$

which is our final estimate. The saddle point bound found earlier is thus only off by a factor of $\sqrt{n}$. Here is a table comparing the asymptotic estimate I_n° provided by the right side of (25) to the exact value of I_n:

$n = 10$	$n = 100$	$n = 1000$
$I_{10} = 9496$	$I_{100} = 2.40533 \cdot 10^{82}$	$I_{1000} = 2.14392 \cdot 10^{1296}$
$I_{10}^\circ = 8839$	$I_{100}^\circ = 2.34149 \cdot 10^{82}$	$I_{1000}^\circ = 2.12473 \cdot 10^{1296}$.

The error is empirically close to $0.3/\sqrt{n}$, a fact that could be proved by developing a complete asymptotic expansion along the lines exposed in the previous section. END OF EXAMPLE 3. □

EXAMPLE 4. *Set partitions and Bell numbers.* The number of partitions of a set of n elements defines the Bell number S_n and one has

$$S_n = n! e^{-1}[z^n]F(z) \qquad \text{where} \qquad F(z) = e^{e^z}.$$

The saddle point equation relative to $F(z)z^{-n-1}$ is

$$\zeta e^\zeta = n + 1.$$

This famous equation admits an asymptotic solution obtained by iteration (or "bootstrapping", see [**93**, p. 26]) upon writing $\zeta + \log\zeta = \log(n+1)$:

$$\zeta \equiv \zeta(n) = \log n - \log\log n + \frac{\log\log n}{\log n} + O\left(\frac{\log^2\log n}{\log^2 n}\right). \tag{26}$$

The corresponding saddle point bound reads

$$S_n \le n! \frac{e^{e^\zeta - 1}}{\zeta^n}.$$

With the approximate solution $\widehat{\zeta}(n) = \log n$, this provides the upper bound

$$S_n \le n! \frac{e^{n-1}}{(\log n)^n}.$$

In particular, the last bound is enough to check that there are much fewer set partitions than permutations, the ratio being bounded from abobe by a quantity of the form $e^{-n\log\log n}$.

In order to implement the saddle point strategy, integration will be carried out over a circle of radius r. First, observe that the function $F(z)$ is strongly concentrated near the real axis since, with $z = re^{i\theta}$,

$$|e^z| = e^{r\cos\theta}, \qquad \left|e^{e^z}\right| \le e^{e^{r\cos\theta}}. \tag{27}$$

In particular $F(re^{i\theta})$ is exponentially smaller than $F(r)$ for any fixed $\theta \neq 0$, when r gets large. We then set

$$h(z) = \log\left(\frac{F(z)}{z^{n+1}}\right) = e^z - (n+1)\log z,$$

and proceed to estimate the integral,

$$F_n = \frac{1}{2i\pi} \int F(z) \, \frac{dz}{z^{n+1}},$$

along the circle of radius $r = \zeta$.

The usual saddle point heuristic suggests to define the "range" of the saddle point by a quantity $\theta_0 \equiv \theta_0(n)$ such that the quadratic terms in the expansion of h at r tend to infinity, while the cubic terms tend to zero. In order to carry out the calculations, it is convenient to express all quantities in terms of r alone, which is possible since n can be disposed of by means of the relation $n = 1 = re^r$. We find:

$$h''(r) = e^r(1 + r^{-1}), \qquad h'''(r) = e^r(1 - 2r^2).$$

Thus, θ_0 should be chosen such that $r^2 e^r \theta_0^2 \to \infty$, $r^3 e^r \theta_0^3 \to 0$, and the choice $r\theta_0 = e^{-2r/5}$ is suitable. It is then easily verified from the bounds (27) that tails are exponentially small, so that they can be neglected.

One then considers the central contribution,

$$F_n^{(0)} := \frac{1}{2i\pi} \int_{\gamma_0} F(z) \, \frac{dz}{z^{n+1}},$$

where γ_0 is the part of the circle $z = re^{i\theta}$ such that $|\theta| \le \theta_0 \equiv e^{-2r/5} r^{-1}$. Since on γ_0, the third derivative is uniformly $O(e^r)$, one has there

$$h(re^{i\theta}) = h(r) - \frac{1}{2} r^2 \theta^2 h''(r) + O(r^3 \theta^3 e^r).$$

This approximation can then be transported into the integral $F^{(0)}$, then tails can be completed in the usual way. The net effect is the estimate

$$[z^n]F(z) = \frac{e^{h(r)}}{\sqrt{2\pi h''(r)}} \left(1 + O\left(r^3 \theta^3 e^r\right)\right),$$

which, upon making the error term explicit rephrases, as follows.

PROPOSITION VIII.2. *The number S_n of set partitions of size n satisfies*

$$S_n = n! \frac{e^{e^\zeta - 1}}{\zeta^n \sqrt{2\pi\zeta(\zeta+1)e^\zeta}} \left(1 + O(e^{-r/5})\right), \tag{28}$$

where ζ is defined implicitly by $\zeta e^\zeta = n + 1$, so that $\zeta = \log n - \log\log n + o(1)$.

Here is a numerical table of the exact values S_n compared to the main term S_n° of the approximation (28):

$n = 10$	$n = 100$	$n = 1000$
$S_{10} = 115975$	$S_{100} \doteq 4.75853 \cdot 10^{115}$	$S_{1000} \doteq 2.98990 \cdot 10^{1927}$
$S_{10}^\circ \doteq 114204$	$I_{100}^\circ \doteq 4.75537 \cdot 10^{115}$	$S_{1000}^\circ \doteq 2.99012 \cdot 10^{1927}$.

The error is about 1.5% for $n = 10$, less than 10^{-3} and 10^{-4} for $n = 100$ and $n = 1{,}000$.

This example is probably the most famous application of saddle point techniques to combinatorics, see [**93**, p. 104]. The asymptotic form in terms of ζ itself is the proper one as no back substitution of an asymptotic expansion of ζ (in terms of n and $\log n$) can provide an asymptotic expansion for S_n solely in terms of n. Regarding explicit representations in terms of n, it is only $\log S_n$ that can be expanded as

$$\frac{1}{n} \log S_n = \log n - \log\log n - 1 + \frac{\log\log n}{\log n} + \frac{1}{\log n} + O\left(\left(\frac{\log\log n}{\log n}\right)^2\right).$$

(Saddle point estimates often involve such implicitly defined quantities.) END OF EXAMPLE 4. □

EXAMPLE 5. *Fragmented permutations.* These correspond to $F(z) = \exp(z/(1-z))$. The example illustrates the case of a singularity at a finite distance. We set as usual

$$h(z) = \frac{z}{1-z} - (n+1)\log z,$$

and start with saddle point bounds. The saddle point equation is

$$\frac{\zeta}{(1-\zeta)^2} = n+1, \tag{29}$$

so that ζ comes close to the singularity at 1 as n gets large:

$$\zeta = \frac{2n+3-\sqrt{4n+5}}{2n+2} = 1 - \frac{1}{\sqrt{n}} + \frac{1}{2n} + O(n^{-3/2}).$$

Here it is natural to take

$$\widehat{\zeta}(n) = 1 - \frac{1}{\sqrt{n}},$$

leading to

$$[z^n]F(z) \le e^{-1/2}e^{2\sqrt{n}}(1+o(1)).$$

The saddle point method is then applied with integration along a circle of radius $r = \zeta$. The saddle point heuristic suggests to localize the integral to a small sector of angle $2\theta_0$ and since $h''(r) = O(n^{3/2})$ while $h'''(r) = O(n^2)$, this means taking θ_0 such that $n^{3/4}\theta_0 \to \infty$ and $n^{2/3}\theta_0 \to 0$. For instance, the choice $\theta_0 = n^{-7/10}$ is suitable. Concentration is easily verified: we have

$$\left|e^{1/(1-z)}\right|_{z=re^{i\theta}} = e \cdot \exp\left(\frac{1-r\cos\theta}{1-2r\cos\theta+r^2}\right),$$

which is a unimodal function of θ for $\theta \in (-\pi, \pi)$. (The maximum of this function of θ is of order $\exp((1-r)^{-1})$ attained at $\theta = 0$; the minimum is $O(1)$ attained at $\theta = \pi$.) In particular, along the saddle point circle, one has

$$\left|e^{1/(1-z)}\right|_{z=re^{i\theta}} = O(\exp\left(\sqrt{n} - n^{1/10}\right), \qquad r = 1 - n^{-1/2}, \quad \theta = n^{-7/10}. \tag{30}$$

so that tails are exponentially small. Local expansions then enable us to justify the use of the general saddle point formula (Theorem VIII.3) in this case. The net result is, with $F_n = n![z^n]F(z)$,

$$F_n \sim n!\,\frac{e^{-1/2}e^{2\sqrt{n}}}{2\sqrt{\pi}n^{3/4}}, \tag{31}$$

which is only $O(n^{-3/4})$ of the corresponding saddle point bound. The error of the saddle point approximation is about 4%, 1%, 0.3% for $n = 10, 100, 1000$, respectively. END OF EXAMPLE 5. □

The expansion above has been extended by E. Maitland Wright [**442, 443**] to several classes of functions with a singularity whose type is an exponential of a function of the form $(1-z)^{-\rho}$. (For the case of (31), Wright [**442**] refers to an earlier article of Perron published in 1914.) His interest was due, at least partly, to applications to generalized partition asymptotics, of which the basic cases are discussed below.

▷ **14.** *Wright's expansions.* Here is a special case. Consider the function

$$F(z) = (1-z)^{-\beta}\exp\left(\frac{A}{(1-z)^\rho}\right), \qquad A > 0, \quad \rho > 0.$$

Then, a saddle point analysis yields, e.g., when $\rho < 1$,

$$[z^n]F(z) \sim \frac{N^{\beta-1-\rho/2}\exp\left(A(\rho+1)N^\rho\right)}{\sqrt{2\pi A\rho(\rho+1)}}, \qquad N := \left(\frac{n}{A\rho}\right)^{\frac{1}{\rho+1}}.$$

(The case $\rho \geq 1$ involves more terms of the asymptotic expansion of the saddle point.) The method generalizes to analytic and logarithmic multipliers, as well as to a sum of terms of the form $A(1-z)^{-\rho}$ inside the exponential. See [**443**] for details. ◁

▷ **15.** *Some oscillating coefficients.* Define the function

$$s(z) = \sin\left(\frac{z}{1-z}\right).$$

The coefficients $s_n = [z^n]s(z)$ are seen to change sign at $n = 6, 21, 46, 81, 125, 180, \ldots$. Do signs change infinitely many times? (Hint: Yes. there are two complex conjugate saddle points and their asymptotic form combine a growth of the form $n^a e^{b\sqrt{n}}$ with an oscillating factor similar to $\sin\sqrt{n}$.)

The sum

$$U_n = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!}$$

exhibits similar fluctuations. ◁

EXAMPLE 6. *Integer partitions.* We are dealing here with a famous chapter of both asymptotic combinatorics and additive number theory[3]. A problem similar to that of asymptotically enumerating partitions was first raised by Ramanujan in a letter to Hardy in 1913, and subsequently developed in a famous joint work of Hardy and Ramanujan (see the account in Hardy's *Lectures* [**224**]). The Hardy–Ramanujan expansion was later perfected by Rademacher who, in a sense, gave an "exact" formula for the partition numbers P_n.

Like before, we start with simple saddle point bounds. Let P_n denote the number of integer partitions of n, with OGF

$$P(z) = \prod_{j\geq 1} \frac{1}{1-z^j}.$$

A form amenable to bounding derives from the exp–log reorganization,

$$\begin{aligned} P(z) &= \exp\sum_{n=1}^{\infty} \log(1-z^n)^{-1} \\ &= \exp\left(\frac{z}{1-z} + \frac{1}{2}\frac{z^2}{1-z^2} + \frac{1}{3}\frac{z^3}{1-z^3}\cdots\right) \\ &= \exp\left(\left(\frac{1}{1-z}\right)\cdot\left(\frac{z}{1} + \frac{z^2}{2(1+z)} + \frac{z^3}{3(1+z+z^2)} + \cdots\right)\right). \end{aligned}$$

The denominator of the general term satisfies, for $x \in (0,1)$, $mx^{m-1} < 1+x+\cdots+x^{m-1} < m$, so that

$$\frac{1}{1-x}\sum_{m\geq 1}\frac{x}{m^2} < \log P(x) < \frac{1}{1-x}\sum_{m\geq 1}\frac{x^m}{m^2}. \tag{32}$$

[3]A complete derivation with all details would consume more space than what we can devote to this questions. We outline here the proof strategy in such a way that (hoperfully) the reader can supply details by herself. The cited references provide a complete treatment.

This proves for real $x \to 1$ that

$$P(x) = \exp\left(\frac{\pi^2}{6(1-x)}(1+o(1))\right), \tag{33}$$

given the elementary identity $\sum m^{-2} = \pi^2/6$. The singularity type at $z = 1$ resembles that of fragmented permutations in the previous example, and, at least, the growth along the real axis is similar. An approximate saddle point is then

$$\widehat{\zeta}(n) = 1 - \frac{\pi}{\sqrt{6n}}, \tag{34}$$

which gives a saddle point bound

$$P_n \le \exp\left(K\sqrt{n}(1+o(1))\right), \qquad K = \pi\sqrt{\frac{2}{3}}.$$

Proceeding further involves transforming the saddle point bounds into a saddle point analysis. Based on previous experience, we shall integrate along a circle of radius $r = \widehat{\zeta}(n)$. To do so, two ingredients are needed: (i) an approximation in the central range; (ii) bounds establishing that the function $P(z)$ is small away from the central range so that tails can be neglected and completed. Assuming the expansion (32) to lift to an area of the complex plane near the real axis, the range of the saddle point should be analogous to what was found already for $\exp(z/(1-z))$, so that $\theta_0 = n^{-7/10}$ will be adopted. Accordingly, we choose to integrate along a circle of radius $r = \widehat{\zeta}(n)$ given by (34) and define the central region by $\theta_0 = n^{-7/10}$. Under these conditions, the central region is seen under an angle that is $O(n^{-1/5})$ from the point $z = 1$.

(i) *Central approximation*. This requires a refinement of (32) till $o(1)$ terms as well as an argument establishing a lifting to a region near the real axis. We set $z = e^{-t}$ and start with $t > 0$. The function

$$L(t) := \log P(e^{-t}) = \sum_{m\ge1} \frac{e^{-mt}}{m(1-e^{-mt})}$$

is a harmonic sum which is amenable to Mellin transform techniques (as described in APPENDIX B: *Mellin transform*, p. 674). The base function is $e^{-t}/(1-e^{-t})$, the amplitudes are the coefficients $1/m$ and the frequencies are the quantities m figuring in the exponents. The Mellin transform of the base function, as given in the appendix, is $\Gamma(s)\zeta(s)$. The Dirichlet series associated to the amplitude frequency pairs is $\sum m^{-1}m^{-s} = \zeta(s+1)$, so that

$$L^\star(s) = \zeta(s)\zeta(s+1)\Gamma(s).$$

Thus $L(t)$ is amenable to Mellin asymptotics and one finds

$$L(t) = \frac{\pi^2}{6t} + \frac{1}{2}\log t - \log\sqrt{2\pi} - \frac{1}{24}t + O(t^2), \qquad t \to 0^+, \tag{35}$$

from the poles of $L^\star(s)$ at $s = 1, 0, -1$. This corresponds to an improved form of (33):

$$\log P(z) = \frac{\pi^2}{6(1-z)} + \frac{1}{2}\log(1-z) - \frac{\pi^2}{12} - \log\sqrt{2\pi} + O(1-z). \tag{36}$$

At this stage, we make a crucial observation: *The precise estimate (35) extends when t lies in any sector symmetric about the real axis, situated in the half-plane $\Re(t) > 0$, and with an opening angle of the form $\pi - \delta$ for an arbitrary $\delta > 0$.* This derives from the fact that the Mellin inversion integral and the companion residue calculations giving rise to (35) extend to the complex realm as long as $|\operatorname{Arg}(t)| < \frac{\pi}{2}$. (See the appendix on Mellin or the article [**153**].) Thus, the expansion (36) holds throughout the central region given our choice of the angle θ_0.

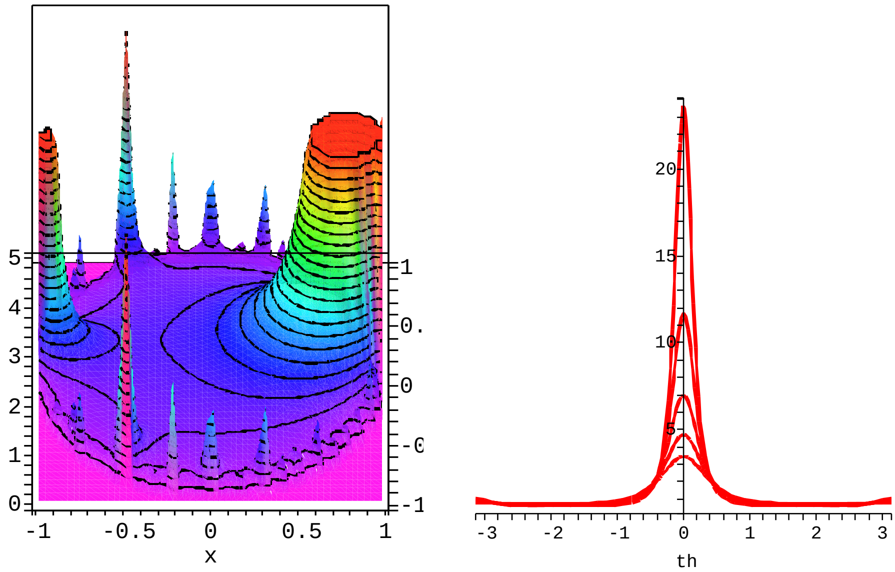


FIGURE 5. Integer partitions. (Left) The surface $|P(z)|$ with $P(z)$ the OGF of integer partitions. The plot shows the major singularity at $z = 1$ and smaller peaks corresponding to singularities at $z = -1, e^{\pm 2i\pi/3}$ and other roots of unity. (Right) A plot of $P(re^{i\theta})$ for varying θ and $r = 0.5, \ldots, 0.75$ illustrates the increasing concentration property of $P(z)$ near the real axis.

The analysis in the central region is then practically isomorphic to the one of $\exp(z/(1-z))$ in the previous example, and it presents no special difficulty.

(ii) *Bounds in the noncentral region.* This is here a nontrivial task since half of the factors entering $P(z)$ are infinite at $z = -1$, one third are infinite at $z = e^{\pm 2i\pi/3}$, and so on. Accordingly, the landscape of $|P(z)|$ along a circle of radius r that tends to 1 is quite chaotic. (See Figure 5 for a rendering.) It is possible to extend the analysis of $\log P(z)$ near the real axis by way of the Mellin transform to the case $z = e^{-t-i\phi}$ as $t \to 0$ and $\phi = 2\pi\frac{p}{q}$ is commensurate to 2π. In that case, one must operate with

$$L_\phi(t) = \sum_{m\geq 1} \frac{1}{m} \frac{e^{-m(t+i\phi)}}{1 - e^{-m(t+i\phi)}} = \sum_{m\geq 1} \sum_{k\geq 1} \frac{1}{m} e^{-mk(t+i\phi)},$$

which is yet another harmonic sum. The net result is that when $|z|$ tends radially towards $e^{2\pi i\frac{p}{q}}$, then $P(z)$ behaves roughly like

$$\exp\left(\frac{\pi^2}{6q^2(1-|z|)}\right), \qquad (37)$$

which is a power $1/q^2$ of the exponential growth as $z \to 1^{-1}$. This analysis extends to a small arc. Considering a complete covering of the circle by arcs whose centres are of argument $2\pi\frac{j}{N}$, $j = 1, \ldots, N-1$, with N chosen large enough, makes it possible to bound the contribution of the noncentral region and prove it to be exponentially small. There are many technical details to be filled in order to justify this approach, so that we switch to a more synthetic one based on transformation properties of $P(z)$, following [**10, 13, 18, 224**]. (Such properties also enter the Hardy-Ramanujan-Rademacher formula for P_n in an essential way.)

The fundamental identity satisfied by $P(z)$ reads

$$P(e^{-2\pi\tau}) = \sqrt{\tau}\exp\left(\frac{\pi}{12}\left(\frac{1}{\tau}-\tau\right)\right)P(e^{-2\pi/\tau}), \tag{38}$$

which is valid when $\Re(\tau) > 0$. The proof is a simple rephrasing of a transformation formula of Dedekind's η (eta) function, summarized in Note 16 below.

$\triangleright$ **16.** *Modular transformation for the Dedekind eta function.* Consider

$$\eta(\tau) := q^{1/24}\prod_{m=1}^{\infty}(1-q^m), \qquad q = e^{2\pi i\tau},$$

with $\Im(\tau) > 0$. Then $\eta(\tau)$ satisfies the "modular transformation" formula,

$$\eta\left(-\frac{1}{\tau}\right)\sqrt{\frac{\tau}{i}}\eta(\tau). \tag{39}$$

This transformation property is first proved when τ is purely imaginary $\tau = it$, then extended by analytic continuation. Its logarithmic form results from a residue evaluation of the integral

$$\frac{1}{2\pi i}\int_\gamma \cot\pi s\cot\pi\frac{s}{\tau}\,\frac{ds}{s},$$

with γ a large contour avoiding poles. (This elementary derivation is due to C. L. Siegel.)
Note. The function $\eta(\tau)$ satisfies transformation formulæ under $S : \tau \mapsto \tau+1$ and $T : \tau \mapsto -1/\tau$, which generate the group of modular (in fact "unimodular") transformations $\tau \mapsto (a\tau+b)/(c\tau+d)$ with $ad-bc=1$. Such functions are called modular forms. $\triangleleft$

Given (38), the behaviour of $P(z)$ away from the positive real axis and near the unit circle can now be quantified. We skip details and content ourselves with a representative special case, the situation when $z \to -1$. Consider thus $P(z)$ with $z = e^{-2\pi t+i\pi}$, where, for our purposes, we may take $t = \frac{1}{\sqrt{24n}}$. Then, Equation (38) relates $P(z)$ to $P(z')$, with $\tau = t - i/2$ and

$$z' = e^{-2\pi/\tau} = \exp\left(-\frac{2\pi t}{t^2+\frac14}\right)e^{i\phi}, \qquad \phi = -\frac{\pi}{t^2+\frac14}.$$

Thus $|z'| \to 1$ as $t \to 0$ with the important condition that $|z'|-1 = O\big((|z|-1)^{1/4}\big)$. In other words, z' has moved *away* from the unit circle. Thus, since $|P(z')| < P(|z'|)$, we may apply the estimate (36) to $P(|z'|)$ to the effect that

$$\log|P(z)| \le \frac{\pi}{24(1-|z|)}(1+o(1), \qquad (z \to -1^+).$$

This is an instance of what was announced in (37) and is in agreement with the surface plot of Figure 5. The extension to an arbitrary angle presents no difficulty.

The two properties developed in (i) and (ii) above guarantee that the approximation (36) can be used and that tails can be completed. We find accordingly that

$$P_n \sim [z^n]\frac{e^{-\pi^2/12}}{\sqrt{1-z}}\exp\left(\frac{\pi^2}{6(1-z)}\right).$$

All computations done, this provides:

PROPOSITION VIII.3. *The number p_n of partitions of integer n satisfies*

$$p_n \equiv [z^n]\prod_{k=1}^{\infty}\frac{1}{1-z^k} \sim \frac{1}{4n\sqrt{3}}e^{\pi\sqrt{2n/3}} \tag{40}$$

The singular behaviour along and near the real line is comparable to that of $\exp((1-z)^{-1})$, which explains a growth like $e^{\sqrt{n}}$ for the number of integer partitions. END OF EXAMPLE 6. $\square$

The asymptotic formula (40) is only the first term of a complete expansion involving decreasing exponentials that was discovered by Hardy and Ramanujan in 1917 and later perfected by Rademacher (see Note 18 below). While the full Hardy–Ramanujan expansion necessitates considering infinitely many saddle–points near the unit circle and requires the modular transformation [**10**], the first term (40) only requires the asymptotic expansion of the partition generating function near $z = 1$.

$\triangleright$ **17.** *A simple yet powerful formula.* Define (cf [**224**, p. 118])

$$P_n^\circ = \frac{1}{2\pi\sqrt{2}}\frac{d}{dn}\left(\frac{e^{K\lambda_n}}{\lambda_n}\right), \qquad K = \pi\sqrt{\frac{2}{3}}, \quad \lambda_n := \sqrt{n - \frac{1}{24}}.$$

Then P_n° approximates P_n with a relative precision of order $e^{-c\sqrt{n}}$ for some $c > 0$. For instance, the error is less than $3 \cdot 10^{-8}$ for $n = 1000$. [Hint: The transformation formula makes it possible to evaluate the central part of the integral very precisely.] $\triangleleft$

$\triangleright$ **18.** *The Hardy–Ramanujan–Rademacher expansion.* The number of integer partitions satisfies the *exact* formula

$$P_n = \frac{1}{\pi\sqrt{2}}\sum_{k=1}^{\infty} A_k(n)\sqrt{k}\frac{d}{dn}\frac{\sinh(\frac{\pi}{k}\sqrt{\frac{2}{3}(n-\frac{1}{24})})}{\sqrt{n-\frac{1}{24}}},$$
$$\text{where} \quad A_k(n) = \sum_{h \bmod k, \gcd(h,k)=1} \omega_{h,k}e^{-2i\pi h/k},$$

$\omega_{h,k}$ is a certain 24th root of unity, $\omega_{h,k} = \exp(\pi i s(h,k))$, and $s_{h,k} = \sum_{\mu=1}^{k-1} \{\{\frac{\mu}{k}\}\}\,\{\{\frac{h\mu}{k}\}\}$ is known as a Dedekind sum, with $\{\{x\}\} = x - \lfloor x \rfloor - \frac{1}{2}$. Proofs are to be found in [**10, 13, 18, 224**]. $\triangleleft$

The principles underlying the partition example have been made into a general method by Meinardus [**311**] in 1954. Meinardus' method abstracts the essential features of the proof and singles out sufficient conditions under which the analysis of an infinite product generating function can be achieved. The conditions, in agreement with the Mellin treatment of harmonic sums, requires analytic continuation of the Dirichlet series involved in $\log P(z)$ (or its analogue), as well as smallness towards infinity of that same Dirichlet series. A summary of Meinardus' method constitutes Chapter 6 of Andrews treatise on partitions [**10**] to which the reader is referred. The method applies to many cases where the summands and their multiplicities have a very regular arithmetic structure.

$\triangleright$ **19.** *Meinardus' theorem.* Consider the infinite product ($a_n \geq 0$)

$$f(z) = \prod_{n=1}^{\infty}(1-z^n)^{-a_n}.$$

The associated Dirichlet series is $\alpha(s) = \sum_{n\geq 1}\frac{a_n}{n^s}$. Assume that $\alpha(s)$ is continuable into a meromorphic function to $\Re(s) \geq -C_0$ for some $C_0 > 0$, with only a simple pole at some $\rho > 0$ and residue A; assume also that $\alpha(s)$ is of moderate growth in the half-plane, namely,

$\alpha(s) = O(|s|^{C_1})$, for some $C_1 > 0$ (as $|s| \to \infty$ in $\Re(s) \geq -C_0$). Let $g(z) = \sum_{n\geq 1} a_n z^n$ and assume a concentration condition of the form

$$\Re g(e^{-t-2i\pi y}) - g(e^{-t}) \leq -C_2 y^{-\epsilon}.$$

Then the coefficient $f_n = [z^n]f(z)$ satisfies

$$f_n = Cn^{\kappa}\exp\left(Kn^{\rho/(\rho+1)}\right), \qquad K = (1+\rho^{-1})\left[A\Gamma(\rho+1)\zeta(\rho+1)\right]^{1/(\rho+1)}.$$

The constants C, κ are:

$$C = e^{D'(0)}(2\pi(1+\rho))^{-1/2}\left[A\Gamma(\rho+1)\zeta(\rho+1)\right]^{(1-2D(0))/(2\rho+2)}, \quad \kappa = \frac{D(0) - 1 - \frac{1}{2}\rho}{1+\rho}.$$

The proof, details of the concentration condition, and error terms are to be found in [**10**, Ch 6]. $\triangleleft$

$\triangleright$ **20.** *Various types of partitions.* The number of partitions into distinct odd summands, squares, cubes, triangular numbers, are cases of application of Meinardus' method. For instance the method provides, for the number Q_n of partitions into *distinct* summands, the asymptotic form

$$Q_n \sim \frac{e^{\pi\sqrt{n/3}}}{4\cdot 3^{1/4}n^{3/4}}.$$

The central approximation is obtained by a Mellin analysis from

$$L(t) := \log Q(e^t) = \sum_{m=1}^{\infty}\frac{(-1)^m}{m}\frac{e^{-mt}}{1-e^{-mt}}, \quad L^\star(s) = -\Gamma(s)\zeta(s)\zeta(s+1)(1-2^{-s}),$$

$$L(t) \sim \frac{\pi^2}{12t} - \log\sqrt{2} + \frac{1}{24}t..$$

(See the already cited references [**10, 13, 18, 224**].) $\triangleleft$

$\triangleright$ **21.** *Plane partitions.* A plane partition of a given number n is a two-dimensional array of integers $n_{i,j}$ that are nonincreasing both from left to right and top to bottom and that add up to n. The first few terms (*EIS* **A000219**) are $1, 1, 3, 6, 13, 24, 48, 86, 160, 282, 500, 859$ and P. A. MacMahon proved that the OGF is

$$R(z) = \prod_{n=1}^{\infty}(1-z^n)^{-n}.$$

Meinardus' method applies to give

$$R_n \sim (\zeta(3)2^{-11})^{1/36}n^{-25/36}\exp\left(3\cdot 2^{-2/3}\zeta(3)^{1/3}n^{2/3} + 2c\right),$$
where $c = -\frac{e}{4\pi^2}(\log(2\pi) + \gamma - 1)$.

See [**10**, p. 199] for this result due to Wright [**441**] in 1931. $\triangleleft$

$\triangleright$ **22.** *Partitions into primes.* Let $P_n^{(\Pi)}$ be the number of partitions of n into summands that are all prime numbers,

$$P^{(\Pi)}(z) = \prod_{n=1}^{\infty}(1-z)^{-p_n},$$

where p_n is the nth prime ($p_1 = 2, p_2 = 3, \dots$). The sequence starts as (*EIS* **A000607**):

$$1, 0, 1, 1, 1, 2, 2, 3, 3, 4, 5, 6, 7, 9, 10, 12, 14, 17, 19, 23, 26, 30, 35, 40.$$

Then

$$\log P_n^{(\Pi)} \sim \left(\frac{2}{3}\right)^{1/2}\pi\left(\frac{n}{\log n}\right)^{1/2}(1+o(1)). \tag{41}$$

An upper bound of a form consistent with (41) can be derived elementarily as a saddle point bound based on the property

$$\sum_{n\geq 1} e^{-tp_n} \sim \frac{t}{\log t}, \qquad t \to 0.$$

This last fact results either from the Prime Number Theorem or from a Mellin analysis based on the fact that $\Pi(s) := \sum p_n^{-s}$ satisfies, with $\mu(m)$ the Möbius function,

$$\Pi(s) = \sum_{m=1}^{\infty} \mu(m) \log \zeta(ms).$$

(See Roth and Szekeres' study [**366**] as well as Yang's article [**449**] for relevant references and recent technology.) This is in sharp contrast with compositions into primes (Chapter V) whose analysis turned out to be especially easy. $\lhd$

VIII. 5. Admissibility

The saddle point method is, as witnessed by the previous section, a very versatile approach to the analysis of coefficients of fast-growing generating functions. However, it is often cumbersome to apply step-by-step. Fortunately, it proves possible to encapsulate the conditions encountered in the analysis of the inverse factorial, involutions, and Bell numbers into a general framework. This leads to the notion of an admissible function. By design, saddle point analysis applies to such functions and asymptotic forms for their coefficients can be systematically determined. Such an approach was initiated by Hayman [**227**] whose steps we closely follow in this section. A crisp account is also given in Section II.7 of Wong's book [**439**] and in Odlyzko's authoritative survey [**330**]. A great merit of abstraction in this context is that admissible functions satisfy useful closure properties, so that an infinite class of admissible functions relevant for combinatorial applications can then be determined.

In this section, we principally base our discussion on H–admissibility (where the prefix H is in recognition of Hayman's original contributions). We consider here a function $f(z)$ that is analytic at the origin and whose coefficient sequence $[z^n]f(z)$ is to be estimated. As was done with previous examples, it proves convenient to switch to polar coordinates and examine the expansion of $f(re^{i\theta})$ when the argument is near the real axis. The fundamental expansion in this context reads

(42)
$$\log f(re^{i\theta}) = \log f(r) + \sum_{\nu=1}^{\infty} \alpha_\nu(r) \frac{(i\theta)^\nu}{\nu!}, \qquad \alpha_0(r) = \log f(r), \quad \alpha_\nu(r) = r\frac{d}{dr}\alpha_{\nu-1}(r).$$

The most important quantities for saddle point analysis are the first two terms, $a(r) = \alpha_1(r)$ and $b(r) = \alpha_2(r)$. It proves convenient to operate with $f(z)$ put into exponential form, $f(z) = e^{h(z)}$, and a simple computation yields

(43)
$$a(r) = rh'(r), \qquad b(r) = r^2 h''(r) + rh'(r).$$

In terms of f, itself, one has

$$a(r) = r\frac{f'(r)}{f(r)}, \qquad b(r) = r\frac{f'(r)}{f(r)} + r^2\frac{f''(r)}{f(r)} - r^2(\frac{f'(r)}{f(r)})^2.$$

Whenever $f(z)$ has nonnegative Taylor coefficients, $a(r)$ and $b(r)$ are positive for $r > 0$. (This follows from an argument already encountered in (10).)

DEFINITION VIII.1 (Hayman–admissibility). *Let $f(z)$ have radius of convergence ρ with $0 < \rho \le \infty$ and be always positive on some subinterval $]R_0, \rho[$ of $]0, \rho[$. The function $f(z)$ is said to be* admissible *if it satisfies the following three conditions.*

H1. [Capture condition] $\lim_{r\to\rho} b(r) = +\infty$.

H2. [Locality condition] *For some function $\delta(r)$ defined over $]R_0, \rho[$ and satisfying $0 < \delta < \pi$, one has*

$$f(re^{i\theta}) \sim f(r)e^{i\theta a(r) - \theta^2 b(r)/2} \qquad \text{as } r \to R_0,$$

uniformly in $|\theta| \le \delta(r)$.

H3. [Decay condition] *Uniformly in $\delta(r) \le |\theta| < \pi$*

$$f(re^{i\theta}) = o\left(\frac{f(r)}{\sqrt{b(r)}}\right).$$

Admissible functions in the above sense are also called Hayman admissible *or H–admissible.*

Coefficients of H–admissible functions can be systematically analysed to first asymptotic order as expressed by the following statement.

THEOREM VIII.5 (Coefficients of admissible functions). *Let $f(z)$ be an H–admissible function and $\zeta \equiv \zeta(n)$ be the unique solution in the interval $]R_0, \rho[$ of the the saddle point equation*

$$\zeta \frac{f'(\zeta)}{f(\zeta)} = n.$$

The Taylor coefficients of $f(z)$ satisfy

$$f_n \equiv [z^n]f(z) \sim \frac{f(\zeta)}{\zeta^n\sqrt{2\pi b(\zeta)}} \qquad \text{as } n \to \infty \tag{44}$$

with $b(z) = z^2 h''(z) + zh'(z)$ and $h(z) = \log f(z)$.

The proof following Hayman's original works bases itself on a general result that describes the shape of the individual terms $f_n r^n$ in the Taylor expansion of $f(z)$ as r gets closer to its limit value ρ. It turns out that the terms $f_n r^n$ exhibit a bell-shaped profile. The asymptotic form (44) will then immediately result from a proper choice of r.

LEMMA VIII.1. *As r tends to ρ, one has*

$$f_n r^n = \frac{f(r)}{\sqrt{2\pi b(r)}}\left[\exp\left(-\frac{(a(r)-n)^2}{b(r)}\right) + \epsilon_n\right], \tag{45}$$

where the error term satisfies $\epsilon_n = o(1)$ as $r \to \rho$ uniformly with respect to integers n, i.e., $\lim_{r\to\rho}\sup_n |\epsilon_n| = 0$.

PROOF. The coefficients f_n are given by Cauchy's formula,

$$f_n r^n = \frac{1}{2\pi}\int_{-\delta}^{2\pi-\delta} f(re^{i\theta})e^{-in\theta}\,d\theta,$$

where $\delta = \delta(n)$ is as specified by the admissibility definition. The estimation of this integral is once more based on a fundamental split

$$f_n r^n = I^{(0)} + I^{(1)} \qquad \text{where} \qquad I^{(0)} = \frac{1}{2\pi}\int_{-\delta}^{+\delta},\ I^{(1)} = \frac{1}{2\pi}\int_{+\delta}^{2\pi-\delta}.$$

From condition H3 (the "decay" condition), uniformly in n:

$$I^{(1)} = o\left(\frac{f(r)}{b(r)^{1/2}}\right). \tag{46}$$

On the other hand, condition H2 (the "locality" condition) gives uniformly in n:

$$\begin{aligned} I^{(0)} &= \frac{f(r)}{2\pi}\int_{-\delta}^{+\delta} e^{i(a(r)-n)\theta-\frac{1}{2}b(r)\theta^2}(1+o(1))\,d\theta \\ &= \frac{f(r)}{2\pi}\left[\int_{-\delta}^{+\delta} e^{i(a(r)-n)\theta-\frac{1}{2}b(r)\theta^2}\,d\theta + o\left(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}b(r)\theta^2}\right)\right]. \end{aligned} \tag{47}$$

The second integral in the last line of (47) is $O(b(r)^{-1/2})$ as $r \to \rho$. Rescaling the first integral and setting $(a(r)-n)(2/b(r))^{1/2} = c$, we obtain

$$I^{(0)} = \frac{f(r)}{\pi\sqrt{2b(r)}}\left[\int_{-\delta\sqrt{b(r)/2}}^{+\delta\sqrt{b(r)/2}} e^{-t^2+ict} + o(1)\right]. \tag{48}$$

Now, it follows from conditions H2 and H3, both taken at $\theta = \delta(r)$ that $b(r)\delta^2 \to \infty$ as $r \to \rho$. Thus the integral in (48) can be extended to a complete Gaussian integral, introducing only $o(1)$ error terms. This entails

$$I^{(0)} = \frac{f(r)}{\pi\sqrt{2b(r)}}\left[\int_{-\infty}^{+\infty} e^{-t^2+ict} + o(1)\right], \tag{49}$$

and the Gaussian integral evaluates to $\sqrt{\pi}e^{-c^2/4}$ (by completing the square and shifting vertically the integration line). Thus, combining the estimate (49) for the central integral $I^{(0)}$ and the estimate (46) for the remainder integral, we obtain the stated estimate (45). □

Theorem VIII.5 has a probabilistic content. There exists a family of discrete random variables $X(r)$ indexed by $r \in (0, R)$ and defined by

$$\mathbb{P}(X(r) = n) = \frac{f_n r^n}{f(r)}.$$

(The model indexed by r in which a random $\mathcal{F}$ structures with GF $f(z)$ is drawn with its size being the random value $X(r)$ defines a Boltzmann model.) These random variables as $r \to R$ thus approach a Gaussian limit; see Figure 6. The choice of a suitable r for a given large n then suffices to complete the proof of the Theorem.

PROOF.[Proof of Theorem VIII.5] To establish the theorem, we first observe that $a(r)$ is increasing (as its derivative $b(r)/r$ is positive) and, in addition tends to infinity (this results from setting $n = 0$ in formula (45)). Thus $\zeta(n)$ is well–defined. Setting $r = \zeta(n)$ in (45) then completes the proof of Theorem VIII.5. □

The rôle of the various conditions should be clear from the preceding discussion and from the study of the exponential function. The choice of the function $\delta(n)$ for a

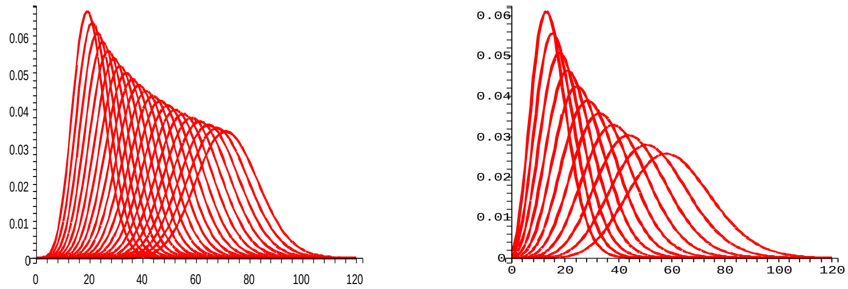


FIGURE 6. The families of Boltzmann distributions associated with involutions ($f(z) = e^{z+z^2/2}$ with $r = 4\,..\,8$) and set partitions ($f(z) = e^{e^z-1}$ with $r = 2\,..\,3$) obey an approximate Gaussian profile.

particular problem is to be guided by consideration of the expansion (42). We must have

$$\alpha_2(r)\delta^2 \to \infty \qquad \text{and} \qquad \alpha_3(r)\delta^3 \to 0.$$

This is because the method requires a nearly complete integral to arise while the error terms after the quadratic part of $\log f(re^{i\theta})$ should be kept small enough. Thus, in order to work, the method necessitates *a priori*

$$\frac{(\alpha_3(r))^2}{(\alpha_2(r))^3} \to 0.$$

Then, δ should be taken in such a way that ($\ll$ still means "much smaller than")

(50) $$\frac{1}{\alpha_2^{1/2}} \ll \delta \ll \frac{1}{\alpha_3^{1/3}},$$

a possible choice being the geometric mean of the two bounds

(51) $$\delta(r) = \alpha_2^{-1/4}\alpha_3^{-1/6}.$$

The note below illustrates the fact that H–admissibility and singularity analysis conditions are in a sense complementary.

▷ **23.** *Non-admissible functions.* The function $f(z) = (1-z)^{-1}$ fails to be be admissible as the asymptotic form that Theorem VIII.5 would imply is the erroneous $[z^n]\dfrac{1}{1-z} \overset{!!}{\sim} \dfrac{e^{-1}}{\sqrt{2\pi}}$, corresponding to a saddle point near $1-n^{-1}$. The explanation of the discrepancy is as follows: Expansion (42) has $\alpha_\nu(r)$ of the order of $(1-r)^{-\nu}$, so that the locality condition and the decay condition cannot be simultaneously satisfied.

Singularity analysis salvages the situation by using a larger part of the contour and by normalizing to a global Hankel Gamma integral instead of a more "local" Gaussian integral. This is also in accordance with the fact that the saddle point formula gives in the case of $[z^n](1-z)^{-1}$ a fraction 0.14676 of the true value, namely, 1. (More generally, functions of the form $(1-z)^{-\beta}$ are typical instances with too slow a growth to be admissible.) ◁

Other functions failing to satisfy the decay condition alone are e^{z^2} and $e^{z^2}+e^z$ as they are also large, away from the central arc and near the negative real axis.

A valuable characteristic of Hayman's work is that it leads to general theorems guaranteeing that large classes of functions are admissible.

THEOREM VIII.6 (Closure of H–admissible functions). *Let $f(z)$ and $g(z)$ be admissible functions and let $P(z)$ be a polynomial with real coefficients. Then:*

- *(i) The product $f(z)g(z)$ and the exponential $e^{f(z)}$ are admissible functions.*
- *(ii) The sum $f(z)+P(z)$ is admissible. If the leading coefficient of $P(z)$ is positive then $f(z)P(z)$ and $P(f(z))$ are admissible.*
- *(iii) If the Taylor coefficients of $e^{P(z)}$ are eventually positive, then $e^{P(z)}$ is admissible.*

PROOF. We refer to Hayman's original paper [**227**] for full proofs that are not difficult. They essentially reduce to making an inspired guess for the choice of the δ function, which may be guided by Equations (50) and (51), and then checking the conditions of the admissibility definition. For instance, in the case of the exponential, $F(z)=e^{f(z)}$, the conditions H1, H2, H3 are satisfied if one takes $\delta(r)=(f(r))^{-2/5}$. $\square$

The closure theorem gives back the known results regarding involutions and set partitions (Bell numbers). It is also to be observed that the OGF of integer partitions is admissible given our discussion of concentration properties in Example 6.

$\triangleright$ **24.** *Idempotent mappings.* Consider functions from a finite set to itself ("mappings" or "functional graphs" in the terminology of Chapter II) that are idempotent, i.e., $\phi\circ\phi=\phi$. The EGF is $I(z)=\exp(ze^z)$ since cycles are constrained to have length 1 exactly. The function $I(z)$ is admissible and

$$I_n \sim \frac{n!}{\sqrt{2\pi n\zeta}}\zeta^{-n}e^{(n+1)/(\zeta+1)},$$

where ζ is the positive solution of $\zeta(\zeta+1)e^{\zeta}=n+1$. This example is due to Harris and Schoenfeld; see [**225**]. $\triangleleft$

$\triangleright$ **25.** *The number of societies.* A society on n distinguished individuals is defined by Sloane and Wieder [**386**] as follows: first partition the n individuals into nonempty subsets and then form an ordered set partition [preferential arrangement] into each subset. The class of societies is thus specified by $\mathcal{S}=\mathfrak{P}\left(\mathfrak{S}_{\ge 1}(\mathfrak{P}(\mathcal{Z}))\right)$, and (labelled) societies are "third-level" structures with EGF

$$S(z)=\exp\left(\frac{1}{2-e^z}\right).$$

The sequence starts as $1,1,4,23,173,1602$ (*EIS* **75729**); asymptotically

$$S_n \sim C\frac{e^{\sqrt{2n/\log 2}}}{n^{3/4}(\log 2)^n},$$

for some computable C. (The singularity is of type "exponential-of-pole" at $\log 2$.) $\triangleleft$

The closure theorem also implies as a very special case that any GF of the form $e^{P(z)}$ with $P(z)$ a polynomial with positive coefficients can be subjected to saddle point analysis resulting in the estimate of Theorem VIII.5. This has been noted by Moser and Wyman [**321, 322**] who list a number of combinatorial applications.

COROLLARY VIII.1 (Exponentials of polynomials). *Let* $P(z) = \sum_{j=1}^{m} a_j z^j$ *have nonnegative coefficients and be aperiodic in the sense that* $\gcd\{j \mid a_j \neq 0\} = 1$. *Let* $f(z) = e^{P(z)}$. *Then, one has*

$$f_n \equiv [z^n]f(z) \sim \frac{1}{\sqrt{2\pi\lambda}} \frac{e^{P(r)}}{r^n}, \qquad \textit{where} \quad \lambda = \left(r\frac{d}{dr}\right)^2 P(r),$$

and r *is a function of* n *given implicitly by* $r\frac{d}{dr}P(r) = n$.

The computations are purely mechanical as they involve the asymptotic expansion (with respect to n) of an algebraic equation. This example covers involutions, permutations of a fixed order in the symmetric group, permutations with cycles of bounded length, as well as set partitions with bounded block sizes. More generally, Corollary VIII.1 applies to any labelled set construction,

$$\mathcal{F} = \mathfrak{P}(\mathcal{G}),$$

corresponding to the EGF equation $F(z) = e^{G(z)}$, when the sizes of $\mathcal{G}$–components are restricted to a finite set. In that case, one has

$$\mathcal{F}^{[m]} = \mathfrak{P}\left(\cup_{j=1}^{r} \mathcal{G}_j\right), \qquad F^{[m]}(z) = \exp\left(\sum_{j=1}^{m} G_j \frac{z^j}{j!}\right),$$

to which the exponential-of-polynomials schemes applies. This covers graphs (plain or functional) whose connected components are of bounded size.

$\triangleright$ **26.** *Applications of "exponentials of polynomials".* Corollary VIII.1 applies to the following combinatorial situations:

Permutations of order p ($\sigma^p = \mathbf{1}$)	$f(z) = \exp\left(\sum_{j \mid p} \frac{z^j}{j}\right)$
Permutations with longest cycle $\leq p$	$f(z) = \exp\left(\sum_{j=1}^{p} \frac{z^j}{j}\right)$
Partitions of sets with largest block $\leq p$	$f(z) = \exp\left(\sum_{j=1}^{p} \frac{z^j}{j!}\right)$.

For instance, the number of solutions of $\sigma^p = \mathbf{1}$ in the symmetric group satisfies

$$f_n \sim \left(\frac{n}{e}\right)^{n(1-1/p)} p^{-1/2} \exp(n^{1/p}),$$

for any fixed prime $p \geq 3$ (Moser and Wyman [**321, 322**]). $\triangleleft$

Complete asymptotic expansions. Harris and Schoenfeld have introduced in [**225**] a technical condition of admissibility that is stronger than Hayman admissibility and is called HS-admissibility. Under such HS-admissibility, a complete asymptotic expansion can be obtained. We omit the definition here due to its technical character but refer instead to the original paper [**225**] and to Odlyzko's survey [**330**]. Odlyzko and Richmond [**331**] later showed that, if $g(z)$ is H–admissible, then $f(z) = e^{g(z)}$ is HS–admissible. Thus, taking H–admissibility to mean at least exponential growth, full asymptotic expansions are to be systematically expected at double exponential growth and beyond.

A range of application is to nested structures generalizing integer partitions and set partitions. For instance, in addition to set partitions, superpartitions (of sets) defined as

$$\mathcal{S} = \mathfrak{P}(\mathfrak{P}_{\geq}1(\mathfrak{P}_{\geq 1}(\mathcal{Z}))), \qquad \text{with EGF} \quad S(z) = e^{e^{e^z-1}-1},$$

can be subjected to this theory and saddle point estimates apply *a priori*.

$\triangleright$ **27.** *Third-level classes.* Consider labelled classes defined from atoms ($\mathcal{Z}$) by *three* nested constructions of the form $\mathfrak{K} \circ \mathfrak{K}'_{\geq 1} \circ \mathfrak{K}''_{\geq 1}$, where each $\mathfrak{K}, \mathfrak{K}', \mathfrak{K}''$ is either a set ($\mathfrak{P}$) or a sequence ($\mathfrak{S}$) construction. All cases can be analysed, either by saddle point and admissibility (SP) or by singularity analysis (SA). Here is a table recapitulating structures, together with their EGF, radius of convergence (ρ), and analytic type.

$\mathfrak{PPP}$	$e^{e^{e^z-1}-1}$	$\rho=\infty$	(SP)	$\mathfrak{SPP}$	$\frac{1}{2-e^{e^z-1}}$	$\rho = 1+\log\log 2$	(SA)
$\mathfrak{PPS}$	$e^{e^{z/(1-z)}-1}$	$\rho=1$	(SP)	$\mathfrak{SPS}$	$\frac{1}{2-e^{z/(1-z)}}$	$\rho=\frac{\log 2}{1+\log 2}$	(SA)
$\mathfrak{PSP}$	$\exp(\frac{1}{2-e^z})$	$\rho=\log 2$	(SP)	$\mathfrak{SSP}$	$\frac{2-e^z}{3-2e^z}$	$\rho=\log\frac{3}{2}$	(SA)
$\mathfrak{PSS}$	$e^{z/(1-2z)}$	$\rho=\frac{1}{2}$	(SP)	$\mathfrak{SSS}$	$\frac{1-2z}{1-3z}$	$\rho=\frac{1}{3}$	(SA)

The outermost construction dictates the analysis type and precise asymptotic equivalents can be developed in all cases. $\triangleleft$

VIII. 6. Combinatorial averages and distributions

Saddle point methods are useful not only for estimating combinatorial counts as we have seen so far, but also for extracting asymptotically probabilistic characteristics of combinatorial structures. This subject may be organized along two major lines dictated by analysis:

— *Univariate problems*: These include the analysis of generating functions that encode information relative to moments of distributions and are obtained by differentiating multivariate generating functions. In the context of saddle point analyses, the dominant asymptotic form of the mean value as well as bounds on the variance usually result, leading in many cases to concentration of distribution (convergence in probability) results.

— *Families of univariate problems*: These correspond to situations where the GF of a structure with parameter value equal to or bounded by some quantity b gives rise to a family (indexed by b) of generating functions that are amenable to either saddle point bounds or even to the complete saddle point method.

Problems a true multivariate nature will be examined in the next chapter dedicated to multivariate asymptotics and limit distributions.

VIII. 6.1. Moment analyses. In cases where a counting generating function $F(z)$ succumbs to saddle point methods, e.g., by being admissible, there are usually a number of associated parameters such that the bivariate GF $F(z,u)$, which is a deformation of $F(z,u)$ when u is close to 1, will also be amenable to a saddle point analysis. In particular the GFs

$$\partial_u F(z,u)\big|_{u=1}, \quad , \partial_u^2 F(z,u)\Big|_{u=1}, \ldots,$$

relative to successive (factorial) moments are likley to be amenable to an analysis that closely resembles that of $F(z)$ itself. In this way, moments can be estimated asymptotically. We illustrate this point by the number of blocks in set partitions.

EXAMPLE 7. *Blocks in random set partitions.* The function

$$f(z,u) = e^{u(e^z-1)}$$

is the bivariate generating function of set partitions with u marking the number of blocks also called parts. We set $f(z) = f(z,1)$ and define

$$g(z) = \frac{\partial}{\partial u} f(z,u)\bigg|_{u=1} = e^{e^z+z-1}.$$

Thus, the quantity

$$\frac{g_n}{f_n} = \frac{[z^n]g(z)}{[z^n]f(z)}$$

represents the mean number of parts in a random partition of $[1\,..\,n]$. We already know that $f(z)$ is admissible and so is $g(z)$ by closure properties. The saddle point for the coefficient integral of $f(z)$ occurs at ζ such that $\zeta e^\zeta = n$, and it is already known that $\zeta = \log n - \log\log n + o(1)$.

It would be possible to analyze $g(z)$ by means of Theorem VIII.5 directly: the analysis then involves a saddle point $\zeta_1 \neq \zeta$ that is relative to $g(z)$. An analysis of the mean would then follow, albeit at some computational effort. It is however more transparent to appeal to Lemma VIII.1 and analyse the coefficients of $g(z)$ at the saddle point of $f(z)$.

Let $a(r), b(r)$ and $a_1(r), b_1(r)$ be the functions of Eq. (43) relative to $f(z)$ and $g(z)$ respectively:

$$\begin{array}{lcllcl}
\log f(z) & = & e^z - 1 & \qquad \log g(z) & = & e^z + z - 1 \\
a(r) & = & re^r & \qquad a_1(r) & = & re^r + r = a(r) + r \\
b(r) & = & (r^2+r)e^r & \qquad b_1(r) & = & (r^2+r)e^r + r = b(r) + r.
\end{array}$$

Thus, estimating g_n by Lemma VIII.1 with the formula taken at $r = \zeta$, one finds

$$g_n = \frac{e^\zeta f(\zeta)}{\sqrt{2\pi b_1(\zeta)}}\left[\exp\left(-\frac{\zeta^2}{b_1(\zeta)}\right) + o(1)\right],$$

while the corresponding estimate for f_n is

$$f_n = \frac{f(\zeta)}{\sqrt{2\pi b_1(\zeta)}}[1 + o(1)].$$

Given that $b_1(\zeta) \sim b(\zeta)$ and that ζ^2 is of much smaller order than $b_1(\zeta)$, one has

$$\frac{g_n}{f_n} = e^\zeta(1+o(1)) = \frac{n}{\log n}(1+o(1)).$$

A similar computation applies to the second moment of the number of parts which is found to be asymptotic to $e^{2\zeta}$ (the computation involves taking a second derivative). Thus, the standard deviation of the number of parts is of an order $o(e^\zeta)$ that is smaller than the mean. This implies a concentration property for the distribution of the number of parts.

PROPOSITION VIII.4. *The variable X_n equal to the number of parts in a random partition of $[1\,..\,n]$ has expectation*

$$\mathbb{E}\{X_n\} = \frac{n}{\log n}(1+o(1)).$$

The distribution satisfies a "concentration" property: for any $\epsilon > 0$, one has

$$\mathbb{P}\left\{\left|\frac{X_n}{\mathbb{E}\{X_n\}} - 1\right| > \epsilon\right\} \to 0 \qquad \text{as } n \to +\infty.$$

The analysis of higher moments is not difficult but it requires care in the manipulation of asymptotic expansions: for instance, Salvy and Shackell [**375**] who "do it right" report that two discrepant estimates (differing by a factor of e^{-1}) had been previously published regarding the value of the mean. . . . . END OF EXAMPLE 7. □

▷ **28.** *Moments of the number of blocks in set partitions.* Let X_n be the number of blocks in a random partition of n elements. Then, one has

$$\mathbb{E}(X_n) = \frac{n}{\log n} + \frac{n \log\log n\,(1+o(1))}{\log^2 n}, \quad \mathbb{V}(X_n) = \frac{n}{\log^2 n} + \frac{n(2\log\log n - 1 + o(1))}{\log^2 n},$$

which proves concentration. The calculation is best performed in terms of the saddle point ζ, then converted in terms of n. [See Salvy's étude [**374**].] ◁

▷ **29.** *The shape of random involutions.* Consider a random involution of size n, the EGF of involutions being $e^{z+z^2/2}$. Then the mean number of 1-cycles and 2-cycles satisfy

$$\mathbb{E}(\#\text{ 1-cycles}) = \sqrt{n} + O(1), \qquad \mathbb{E}(\#\text{ 2-cycles}) = \frac{1}{2}n + O(\sqrt{n}).$$

In addition, the corresponding distributions are concentrated. ◁

▷ **30.** *Mean number of parts in integer partitions.* The mean number of parts (or summands) in a random integer partition of size n is

$$\frac{1}{K}\sqrt{n}\log n + O(n^{1/2}), \qquad K = \pi\sqrt{\frac{2}{3}}.$$

For a partition into distinct part, the mean number of parts is

$$\frac{2\sqrt{3}\log 2}{\pi}\sqrt{n} + o(n^{1/2}).$$

The complex-analytic proof only requires the central estimates of $\log P(e^{-t})$ and $\log Q(e^{-t})$, given the concentration properties, as well as the estimates

$$\sum_{m\ge 1} \frac{e^{-mt}}{1-e^{-mt}} \sim \frac{-\log t + \gamma}{t} + \frac{1}{4}, \quad \sum_{m\ge 1} (-1)^{m-1}\frac{e^{-mt}}{1-e^{-mt}} \sim \frac{\log 2}{t} - \frac{1}{4},$$

which result from a standard Mellin analysis, the respective transforms being

$$\Gamma(s)\zeta(s)^2, \qquad \Gamma(s)(1-2^{1-s})\zeta(s)^2.$$

Full asymptotic expansions of the mean and of moments of any order can be determined. In addition, the distributions are concentrated around their mean. (The first order estimates are due to Erdős and Lehner [**128**] who gave an elementary derivation and also obtained the limit distribution of the number of summands in both cases: they are a double exponential (for P) and a Gaussian (for Q).) ◁

The next example illustrates a simple—but partial—approach based on saddle points to a famous problem of combinatorial theory. In that case, a generating function for the first moments is directly obtained.

EXAMPLE 8. *Increasing subsequences in permutations.* Given a permutation written in linear notation as $\sigma = \sigma_1 \cdots \sigma_n$, an increasing subsequence is a subsequence $\sigma_{i_1} \cdots \sigma_{i_k}$ which is in increasing order, i.e., $i_1 < \cdots < i_k$ and $\sigma_{i_1} < \cdots \sigma_{i_k}$. The question asked is: *What is the mean number of increasing subsequences in a random permutation?*

The problem has a flavour analogous to that of "hidden" patterns in random words, which was tackled in Chapter V, and indeed similar methods are applicable to determine moments of the number of increasing subsequences. Define a *tagged permutation* as a permutation together with one of its increasing subsequence distinguished. (We also consider the null subsequence as an increasing subsequence.) For instance,

$$7\,|\mathbf{3}\;5\;2\;|\mathbf{6}\;4\;1\;|\mathbf{8}\;9$$

is a tagged permutation with the increasing subsequence **3 6 8** that is distinguished. The vertical bars are used to identifty the tagged elements, but they may also be interpreted as decomposing the permutation into subpermutation fragments. We let $\mathcal{T}$ be the class of tagged permutations, $T(z)$ be the corresponding EGF, and set $T_n = n![z^n]T(z)$. The *mean* number of increasing subsequences in a random permutation of size n is clearly $t_n = T_n/n!$.

In order to enumerate $\mathcal{T}$, we let $\mathcal{P}$ be the class of all permutations and $\mathcal{P}^+$ the subclass of non empty permutations. Then, one has up to isomorphism,

$$\mathcal{T} = \mathcal{P} \star \texttt{set}\,(\mathcal{P}^+),$$

since a tagged permutation can be reconstructed from its initial fragment and the *set* of its fragments (by ordering the set according to increasing values of initial elements). This combinatorial argument gives the EGF $T(z)$ as

$$T(z) = \frac{1}{1-z}\exp\left(\frac{z}{1-z}\right).$$

The generating function $T(z)$ can be expanded, so that the quantity T_n admits a closed form,

$$T_n = \sum_{k=0}^{n}\binom{n}{k}\frac{n!}{k!}.$$

From there it is possible to analyse T_n adn t_n asymptotically by means of the Laplace method for sums. However, analytically, the function $T(z)$ is a mere variant of the EGF of fragmented permutations. Saddle point conditions are again easily checked, either directly or via admissibility, to the effect that

$$t_n \equiv \frac{T_n}{n!} \sim \frac{e^{-1/2}e^{2\sqrt{n}}}{2\sqrt{\pi}n^{1/4}}. \tag{52}$$

(Compare with the closely related estimate (31) on p. 509.) The result is originally due to Lifschitz and Pittel [**291**] who obtained it using real analytic methods.

This analysis provides information about an important and much less accessible parameter, $\lambda(\sigma)$, representing the *length of the longest increasing subsequence* in σ If $\iota(\sigma)$ is the number of increasing subsequences, then one has

$$2^{\lambda(\sigma)} \le \iota(\sigma),$$

since the number of increasing subsequences of σ is *at least* as large as the number of subsequences contained in the *longest* increasing subsequence. Let ℓ_n be the expectation of λ over permutations of size n. Then, by convexity of the function 2^x, one has

$$2^{\ell_n} \le t_n, \qquad \text{so that} \qquad \ell_n \le \frac{2}{\log 2}\sqrt{n}(1+o(1)), \tag{53}$$

by (52). In summary:

PROPOSITION VIII.5. *The mean number of increasing subsequences in a random permutation of n elements is asymptotically*

$$\frac{e^{-1/2}e^{2\sqrt{n}}}{2\sqrt{\pi}n^{1/4}}\left(1+o(1)\right).$$

Accordingly, the expected length of the longest increasing subsequence in a random permutation of size n satisfies the inequality

$$\ell_n \leq \frac{2}{\log 2}\sqrt{n}(1 + o(1)).$$

. . . . END OF EXAMPLE 8. □

▷ **31.** *A useful recurrence.* A decomposition according to the location of n yields the recurrence

$$t_n = t_{n-1} + \frac{1}{n}\sum_{k=0}^{n-1} t_k, \qquad t_0 = 1.$$

Hence $T(z)$ satisfies the ordinary differential equation,

$$(1-z)^2 \frac{d}{dz}T(z) = (2-z)T(z), \qquad T(0) = 1,$$

which can be solved explicitly. Also the differential equation gives rise to the recurrence

$$t_{n+1} = 2t_n - \frac{n}{n+1}t_{n-1}, \qquad t_0 = 0, \qquad t_1 = 2,$$

by which t_n can be computed efficiently in a linear number of operations. ◁

▷ **32.** *Related combinatorics.* The sequence of values of T_n starts as $1, 2, 7, 34, 209, 1546, 13327$, and is *EIS* **A002720**. It counts the following equivalent objects: (i) the $n \times n$ binary matrices with at most one entry 1 in each column; (ii) the partial matchings of the complete bipartite graph $K_{n,n}$; (iii) the injective partial mappings of $[1\,..\,n]$ to itself. ◁

▷ **33.** *A simple probabilistic lower bound.* Elementary probability theory provides a simple lower bound on ℓ_n. Let $X_1, \ldots, X_n$ be independent random variables uniformly distributed over $[0, 1]$. Assume $n = m^2$. Partition $[0, 1[$ into m subintervals each of the form $[\frac{j-1}{m}, \frac{j}{m}[$ and $X_1, \ldots, X_n$ into m blocks, each of the form $X_{(k-1)m+1}, \ldots, X_{km}$. There is a probability $1-(1-m^{-1})^m \sim 1-e^{-1}$ that block numbered 1 contains an element of subinterval numbered 1, block numbered 2 contains an element of subinterval numbered 2, and so on. Then, with high probability, at least $\frac{m}{2}$ of the blocks contain an element in their matching subinterval. Consequently, $\ell_n \geq \frac{1}{2}\sqrt{n}$. (The factor $\frac{1}{2}$ can even be improved a little.) The crisp booklet by Steele [**394**] describes many similar as well as more advanced applications to combinatorial optimization. See also the book of Motwani and Raghavan [**323**] for applications to randomized algorithms in computer science. ◁

The upper bound obtained on the expected length ℓ_n of the longest increasing sequence is of the form $2.89\sqrt{n}$ while Note 33 describes a lower bound of the form $\ell_n \geq \frac{1}{2}\sqrt{n}$. In fact, Logan and Shepp [**293**] independently of Vershik and Kerov [**426**] have succeeded in establishing the much more difficult result

$$\ell_n \sim 2\sqrt{n}.$$

Their proof is based on a detailed analysis of the profile of a random Young tableau. (The bound obtained here by a simple mixture of saddle point estimates and combinatorial approximations at least provides the right order of magnitude.) This has led in turn to attempts at characterizing the asymptotic distribution of the length of the longest increasing subsequence. The problem remained unsolved for two decades, despite many tangible progresses. J. Baik, P. A. Deift, and K. Johansson [**19**] eventually obtained a solution (in a publication dated 1999) by relating longest increasing subsequences to eigenvalues of random matrix ensembles. We regretfully redirect the

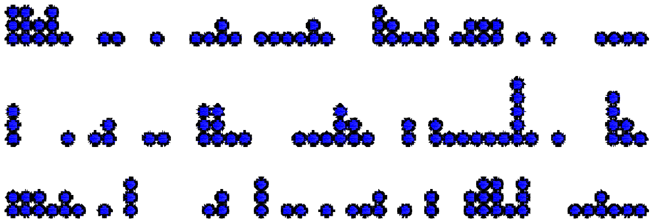


FIGURE 7. Three random allocations of $n = 50$ balls in $m = 50$ bins.

reader to presentations of the beautiful theory surrounding this sensational result, for instance [**6, 96**].

$\triangleright$ **34.** *The Baik–Deift-Johansson Theorem.* Consider the Painlevé II equation $u''(x) = 2u(x)^3 + xu(x)$, and the particular solution $u_0(x)$ that is asymptotic to $-\operatorname{Ai}(x)$ as $x \to +\infty$, with $\operatorname{Ai}(x)$ the Airy function which solves $y'' - xy = 0$. Define the Tracy–Widom distribution (arising in random matrix theory)

$$F(t) = \exp\left(\int_t^\infty (x-t)u_0(x)^2\,dx\right).$$

Then, the distribution of the length of the longest increasing subsequence, L_n satisfies for all fixed t:

$$\lim_{n\to\infty} \mathbb{P}\left(L_n \le 2\sqrt{n} + tn^{1/6}\right) = F(t).$$

Thus the discrete random variable L_n converges to a well-characterized distribution [**19**]. $\triangleleft$

VIII. 6.2. Families of univariate problems. We shall content ourselves with giving a single example related to random allocations, occupancy statistics, and balls-in-bin models that were already discussed in Chapter II. In addition, we limit ourselves to the use of saddle point bounds.

EXAMPLE 9. *Capacity in occupancy problems.* Assume that n balls are thrown into n bins, uniformly at random. How many balls does the most filled bin contain? We shall in fact deal with a generalized version of the problem where n balls are thrown into m bins, in the regime $n = \alpha m$ for some fixed α in $(0, +\infty)$. The size of the most filled bin will be called the *capacity* and we let $C_{n,m}$ denote the random variable, when all m^n tables are taken equally likely. Under our conditions a random bin contains on average a constant number, α, of balls. The proposition below proves that the most filled bin has somewhat more, as exemplified by Figure 7.

PROPOSITION VIII.6. *Let n and m tend simultaneously to infinity, with the constraint that $\frac{n}{m} = \alpha$ remains constant. Then, the expected capacity satisfies*

$$\frac{1}{2}\frac{\log n}{\log\log n}(1+o(1)) \le \mathbb{E}\{C_{n,m}\} \le 2\frac{\log n}{\log\log n}(1+o(1)).$$

In addition, the probability of capacity to lie outside the interval determined by the lower and upper bounds tends to 0 as $m, n \to \infty$.

PROOF. We detail the proof when $\alpha = 1$ and abbreviate $C_n = C_{n,m}$. From Chapter II, we know that

$$(54)\qquad \begin{cases} \mathbb{P}\{C_n \le b\} &=& \dfrac{n!}{n^n}[z^b](e_b(z))^n \\ \mathbb{P}\{C_n > b\} &=& \dfrac{n!}{n^n}\left(e^{nz} - (e_b(z))^n\right), \end{cases}$$

where $e_b(z)$ is the truncated exponential:

$$e_b(z) = \sum_{j=0}^{b} \frac{z^j}{j!}.$$

The two equalities of (54) permit us to bound the left and right tails of the distribution. As suggested by the Poisson approximation of balls-in-bins model, we decide to adopt saddle point bounds based on $z = 1$. This gives

$$(55)\qquad \begin{cases} \mathbb{P}\{C_n \le b\} &\le& \dfrac{n!e^n}{n^n}\left(\dfrac{e_b(1)}{e}\right)^n \\ \mathbb{P}\{C_n > b\} &\le& \dfrac{n!e^n}{n^n}\left(1 - \left(\dfrac{e_b(1)}{e}\right)^n\right). \end{cases}$$

We set

$$(56)\qquad \rho_b(n) = \left(\frac{e_b(1)}{e}\right)^n.$$

This quantity represents the probability that n Poisson variables of rate 1 all have value b or less. (We know for elementary probability theory that this should be a reasonable approximation of the problem at hand.) A weak form of Stirling's formula, namely, $\frac{n!e^n}{n^n} < 2\sqrt{\pi n} \qquad (n \ge 1)$, then yields an alternative version of (55),

$$(57)\qquad \begin{cases} \mathbb{P}\{C_n \le b\} &\le& 2\sqrt{\pi n}\rho_b(n) \\ \mathbb{P}\{C_n > b\} &\le& 2\sqrt{\pi n}\,(1 - \rho_b(n)). \end{cases}$$

For fixed n, the function $\rho_b(n)$ increases steadily from e^{-n} to 1 as b varies from 0 to ∞. In particular, the "transition region" where $\rho_b(n)$ stays away from both 0 and 1 is expected to play a rôle. This suggests defining $b_0 \equiv b_0(n)$ such that

$$b_0! \le n < (b_0 + 1)!,$$

so that

$$b_0(n) = \frac{\log n}{\log\log n}(1 + o(1)).$$

We also observe that, as $n, b \to \infty$, there holds

$$(58)\qquad \begin{aligned} \rho_b(n) &= (e^{-1}e_b(1))^n = \left(1 - \frac{e^{-1}}{(b+1)!} + O(\frac{1}{(b+2)!})\right)^n \\ &= \exp\left(-\frac{ne^{-1}}{(b+1)!} + O(\frac{n}{(b+2)!})\right). \end{aligned}$$

Left tail. We take $b = \lfloor \frac{1}{2}b_0 \rfloor$ and a simple computation from (58) shows that for n large enough, $\rho_b(n) \le \exp(-\sqrt[3]{n})$. Thus, by the first inequality of (57), the probability that the capacity be less than $\frac{1}{2}b_0$ is exponentially small:

$$(59)\qquad \mathbb{P}\{C_n \le \frac{1}{2}b_0(n)\} \le 2\sqrt{\pi n}\exp(-\sqrt[3]{n}).$$

Right tail. Take $b = 2b_0$. Then, again from (58), for n large enough, one has $1 - \rho_b(n) \le 1 - \exp(-\frac{1}{n}) = \frac{1}{n}(1 + o(1))$. Thus, the probability of observing a capacity that exceeds $2b_0$

is vanishingly small, and is $O(n^{-1/2})$. Taking next $b = 2b_0 + r$ with $r > 0$, similarly gives the bound

$$\mathbb{P}\{C_n > 2b_0(n) + r\} \le 2\sqrt{\frac{\pi}{n}} \left(\frac{1}{b_0(n)}\right)^r. \tag{60}$$

The analysis of the left and right tails in Equations (59) and (60) now implies

$$\left\{\begin{array}{lcl} \mathrm{E}\{C_n\} & \le & 2b_0(n) + \displaystyle\sum_{r=0}^{\infty} 2\sqrt{\frac{\pi}{n}}(b_0(n))^{-r} = 2b_0(n)(1+o(1)) \\ \mathrm{E}\{C_n\} & \ge & \displaystyle\sum_{r=0}^{\lfloor \frac{1}{2} b_0(n) \rfloor} \left[1 - 2\sqrt{\pi n}\exp(-\sqrt[3]{n})\right] = \frac{1}{2} b_0(n)(1+o(1)). \end{array}\right. \tag{61}$$

This justifies the claim of the proposition when $\alpha = 1$. The general case ($\alpha \neq 1$) follows similarly from saddle point bounds taken at $z = \alpha$. □

The saddle point bounds described above are obviously not tight, with some care in derivations, one can show by the same means that the distribution is concentrated around $\log n / \log\log n$. In addition, the saddle point method may be used instead of crude bounds. The net effect is: the expected capacity satisfies, for any fixed $\alpha = n/m$:

$$\mathbb{E}\{C_{n,m}\} \sim \frac{\log n}{\log\log n}.$$

This result, in the context of longest probe sequences in hashing, was obtained by Gonnet [**206**] under the Poisson model. Many key estimates regarding random allocations (including capacity) are to be found in the book by Kolchin *et al.* [**278**]. Analyses of the type discussed above are also useful in evaluating various dynamic hashing algorithms by saddle point methods [**142, 360**]. . . . . END OF EXAMPLE 9. □

VIII. 7. Variations on the theme of saddle points

We conclude this chapter with extensions of the basic paradigm in two major directions. First, we discuss perturbations of the basic saddle point paradigm in the case of large powers (Subsection VIII. 7.1): this paves the way for the analysis of Gaussian laws in the next chapter, where the rich framework of "Quasi-Powers" plays a central rôle in so many combinatorial applications. Next we examine the case of higher order saddle points (Subsection VIII. 7.2). We conclude this section with brief indications on what is known as phase transitions or critical phenomena in the applied sciences and as uniform asymptotics in the applied mathematics literature: technically, this involves a combination of multiple saddle points and perturbation theory.

VIII. 7.1. Large powers and Gaussian forms. Saddle point analysis has consequences for multivariate asymptotics and is a direct way of proving that many discrete distributions tend to the Gaussian law in the asymptotic limit. For large powers, this property derives painlessly from our earlier developments, especially Theorem VIII.4, by means of a "perturbation" analysis.

First, let us examine a particularly easy problem: *How do the coefficients of* $[z^N]e^{nz}$ *vary as a function of* N *when* n *is some large but fixed number?* These coefficients are

$$c_N^{(n)} = [z^N]e^{nz} = \frac{n^N}{N!}.$$

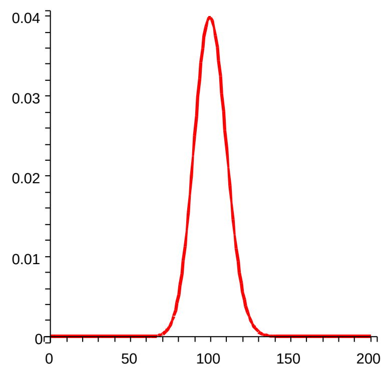


FIGURE 8. The coefficients $[z^N]e^{nz}$ when $n = 100$ and $N = 0\,..\,200$ have a bell-shaped aspect. (The coefficients are normalized by e^{-n}.)

By the ratio test, they have a maximum when $N \approx n$ and are small when N differs significantly from n; see Figure 8. The bell-shaped profile is also apparent on the figure and is easily verified by real analysis. The situation is then similar to what is already known of the binomial coefficients on the nth line of Pascal's triangle, corresponding to $[z^n](1+z)^n$ with N varying.

The asymptotically Gaussian character of coefficients of large powers is actually universal amongst a wide class of analytic functions. We prove this within the framework of large powers already investigated in Section VIII. 4 and consider the general problem of estimating the coefficients $[z^N]\,(A(z)\cdot B(z)^n)$ as N varies. In accordance with Section VIII. 4, we postulate the following: (C_1): $A(z), B(z)$ are analytic at 0 and have nonnegative coefficients; (C_2): $A(z)$ is aperiodic; (C_3) The radius of convergence R of $B(z)$ is a minorant of the radius of convergence of $A(z)$. We also recall that the *spread* has been defined as

$$T := \lim_{x\to R^-} \frac{xB'(x)}{B(x)}.$$

THEOREM VIII.7 (Large powers and Gaussian forms). *Consider the "large powers" coefficients:*

$$c_N^{(n)} := [z^N]\,(A(z)\cdot B(z)^n)\,. \tag{62}$$

Assume that the two analytic functions $A(z), B(z)$ satisfy the conditions (C_1), (C_2), and (C_3) above. Assume also that the radius of convergence of B satisfies $R > 1$. Define the two constants:

$$\mu = \frac{B'(1)}{B(1)}, \qquad \sigma^2 = \frac{B''(1)}{B(1)} + \frac{B'(1)}{B(1)} - \left(\frac{B'(1)}{B(1)}\right)^2 \qquad (\sigma > 0).$$

Then the coefficients $c_N^{(n)}$ for fixed n as N varies have an asymptotically Gaussian profile in the precise sense that for $N = \mu n + x\sqrt{n}$, there holds (as $n \to \infty$)

$$\frac{1}{A(1)B(1)^n} c_N^{(n)} = \frac{1}{\sigma\sqrt{2\pi n}} e^{-x^2/(2\sigma^2)} \left(1 + O(n^{-1/2})\right), \tag{63}$$

uniformly with respect to x, when x belongs to a finite interval of the real line.

PROOF. We start with a few easy observation that shed light on the global behaviour of the coefficients. First, since $R > 1$, we have the exact summation,

$$\sum_{N=0}^{\infty} C_N^{(n)} = A(1)B(1)^n,$$

which explains the normalization factor in the estimate (63). Next, by definition of the spread and since $R > 1$, one has

$$\mu = \frac{B'(1)}{B(1)} < T = \lim_{x \to R^-} \frac{xB'(x)}{B(x)},$$

given the general property that $xB'(x)/B(x)$ is increasing. Thus, the estimation of the coefficients in the range $N = \mu n \pm O(\sqrt{n})$ falls into the orbit of Theorem VIII.4 which expresses the results of the saddle point analysis in the case of large powers.

Referring to the statement of Theorem VIII.4, the saddle point equation is

$$\zeta \frac{B'(\zeta)}{B(\zeta)} = \frac{B'(1)}{B(1)} + \frac{x}{\sqrt{n}},$$

with ζ a function of x and n. For x in a bounded set, we thus have $\zeta \sim 1$ as $n \to \infty$. It then suffices to effect an asymptotic expansion of the quantities $\zeta, A(\zeta), B(\zeta), \xi$ in the saddle point formula of Equation (23). In other words, the fact that N is close to μn induces for ζ a small perturbation with respect to the value 1. With $a_j := A^{(j)}(1)$ and $b_j := B^{(j)}(1)$, one finds mechanically

$$\begin{aligned} \zeta &= 1 + \frac{b_0^2}{b_0 b_2 + b_0 b_1 - {b_1}^2} \frac{x}{\sqrt{n}} + O(n^{-1}) \\ \frac{B(\zeta)}{\zeta^\mu} &= b_0 + \frac{x^2}{2n} \frac{b_0^3}{b_0 b_2 + b_0 b_1 - {b_1}^2} + O(n^{-3/2}), \end{aligned}$$

and so on. The statement follows. □

Take first $A(z) \equiv 1$. In the particular case when $B(z)$ is the probability generating function of a discrete random variable Y, one has $B(1) = 1$, and the coefficient $\mu = B'(1)$ is the mean of the distribution. The function $B(z)^n$ is then the probability generating function (PGF) of a sum of n independent copies of Y. Theorem VIII.7 then describes a Gaussian approximation of the distribution of the sum near the mean. Such an approximation is called a *local limit law*, where the epithet "local" refers to the fact that the estimate applies to the coefficients themselves. (In contrast, an approximation of the partial sums of the coefficients by the Gaussian error function is known as a *central limit law* or as an *integral limit law*.) In the more general case where $A(z)$ is also a probability generating function of a nondegenerate random variable (i.e., $A(z) \neq 1$), similar properties hold and one has:

COROLLARY VIII.2 (Local limit law for sums). *Let X be a random variable with probability generating function (PGF) $A(z)$ and $Y_1, \ldots, Y_n$ be independent variables with PGF $B(z)$, wheer it is assumed that X and the Y_j are supported on $\mathbb{Z}_{\geq 0}$. Assume that $A(z)$ and $B(z)$ are analytic in some disc that contains the unit disc in its interior and that $B(z)$ is aperiodic. Then the sum,*

$$S_n := X + Y_1 + Y_2 + \cdots + Y_n,$$

satisfies a local limit law of the Gaussian type: For x in any finite interval, one has

$$\mathbb{P}\left(S_n = \lfloor \mu n + t\sigma\sqrt{n} \rfloor\right) = \frac{e^{-t^2/2}}{\sqrt{2\pi n}}\left(1 + O(n^{-1/2})\right).$$

PROOF. This is just a restatement of Theorem VIII.7, setting $x = t\sigma$ and taking into account $A(1) = B(1) = 1$. □

▷ **35.** *An alternative proof.* The saddle point ζ is near 1 when N is near the centre $N \approx \mu n$. It is alternatively possible to recover the $c_n^{(N)}$ by Cauchy's formula upon integrating along the circle $|z| = 1$, which is then only an *approximate* saddle point contour. This convenient variant is often used in the literature, but one needs to take care of linear terms in expansions. Its origins go back to Laplace himself in his first proof of the local limit theorem (which was expressed however in the language of Fourier series as Cauchy's theory was yet to be born). See Laplace's treatise *Théorie Analytique des Probabilités* [**287**] first published in 1812 for much fascinating mathematics. ◁

Gaussian forms for large powers admit many variants. As already pointed out in Section VIII. 4, the positivity conditions can be greatly relaxed. Also, estimates for partial sums of the coefficients are possible by similar techniques. The asymptotic expansions can be extended to any order. Finally, suitable adaptations of Theorems VIII.4 and VIII.7 make it possible to allow x to tend slowly to infinity and manage what is known as a "moderate deviation" regime. We do not pursue these aspects here since we shall develop a more general framework, that of "Quasi-Powers" in the next chapter.

VIII. 7.2. Multiple saddle points. All the analyses carried out so far have been in terms of simple saddle points, which is by far the most common situation. In order to get a feel of what to expect in the case of multiple saddle points, consider first the problem of estimating the two *real* integrals,

$$I_n := \int_0^1 (1 - x^2)^n \, dx, \qquad J_n := \int_0^1 (1 - x^3)^n \, dx.$$

(For the purpose of this discussion, we ignore the fact that the integrals can be evaluated in closed form by way of the Beta function.) The contribution of any interval $[x_0, 1]$ is exponentially small, and the ranges to be considered on the right of 0 are a little over $O(n^{-1/2})$ and $O(n^{-1/3})$, respectively. One thus sets

$$x = \frac{t}{\sqrt[2]{n}} \quad \text{for } I_n, \qquad x = \frac{t}{\sqrt[3]{n}} \quad \text{for } J_n.$$

Then local expansions apply, tails can be completed in the usual way, to the effect that

$$I_n \sim \frac{1}{\sqrt[2]{n}} \int_0^\infty e^{-t^2} \, dt, \qquad J_n \sim \frac{1}{\sqrt[3]{n}} \int_0^\infty e^{-t^3} \, dt.$$

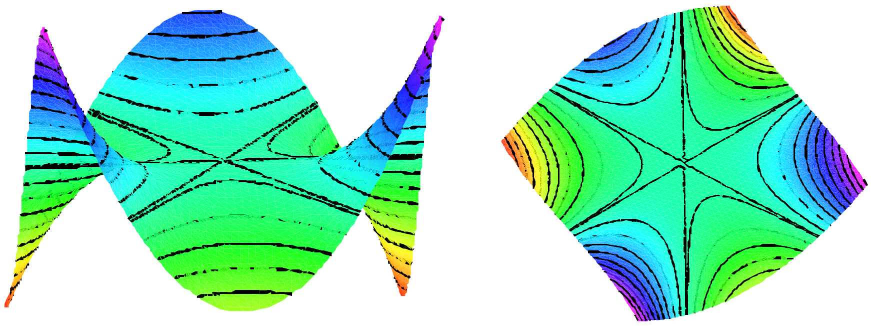


FIGURE 9. Two views of a double saddle point also known as "monkey saddle".

The integrals reduce to the ones defining the Gamma function, which provides the final estimates

$$I_n \sim \frac{1}{2}\frac{\Gamma(\frac{1}{2})}{n^{\frac{1}{2}}}, \qquad J_n \sim \frac{1}{3}\frac{\Gamma(\frac{1}{3})}{n^{\frac{1}{3}}}.$$

The repeated occurrences of $\frac{1}{2}$ in the quadratic case and of $\frac{1}{3}$ in the cubic case stand out. The situation in the cubic case is typical of the Laplace method for integrals when a multiple critical point is present.

What has been just encountered in the case of real integrals is representative of what to expect for *complex* integrals and saddle points of higher orders. Consider, for simplicity, the case of a double saddle point of an analytic function $f(z)$. At such a point z_0, we have $f(z_0) \neq 0$, $f'(z_0) = f''(z_0) = 0$, and $f'''(z_0) \neq 0$. Then, there are three steepest descent lines emanating from the saddle point and three steepest ascent lines. Accordingly, one should think of the landscape of $|f(z)|$ as formed of three "valleys" separated by three mountains and meeting at the common point z_0. The characteristic aspect is that of of a "monkey saddle" (comparable to a saddle with places for two legs and a tail) and is displayed in Figure 9.

In view of the previous discussion, we can enounce a modified form of the saddle point formula of Theorem VIII.3: *Consider an integral $I = \int_A^B F(z)\,dz$, where the integrand $F(z) = e^{f(z)}$ is an analytic function and A, B lie in opposite valleys across a* double *saddle point z_0. Then assuming that tails can be neglected and completed back, and that a fourth-order expansion applies, one has*

$$\frac{1}{2i\pi}\int_A^B e^{f(z)}\,dz \sim \pm\omega\frac{\Gamma(\frac{1}{3})}{2i\pi}\frac{e^{f(z_0)}}{\sqrt[3]{|f'''(z_0)|}}. \tag{64}$$

There, ω is a cube root of unity ($\omega^3 = 1$) dependent upon the position of the contour that results from the original contour linking A and B, while the sign $\pm$ depends on orientation. The approach to formula (64) consists simply in carrying out the integration of

$$\int_C \exp\left(\frac{1}{3!}f'''(z_0)(z-z_0)^3\right)\,dz,$$

with C made of the two rays parametrized by $te^{2ij\pi/3}$, $te^{2ij'\pi/3}$ for $t\in\mathbb{R}_{>0}$.

$\triangleright$ **36.** *Higher-order saddle points.* For a saddle point of order $p+1$, the saddle point formula reads

$$\frac{1}{2i\pi}\int_A^B e^{f(z)}\,dz \sim \pm\omega\frac{\Gamma(\frac{1}{p})}{2i\pi}\frac{e^{f(z_0)}}{\sqrt[p]{|f(p)(z_0)|}},$$

where $\omega^p=1$. $\triangleleft$

$\triangleright$ **37.** *Vanishing multipliers and multiple saddle points.* This note supplements Note 36. For a saddle point of order $p+1$ and an integrand of the form $(z-\zeta)^b\cdot e^{h(z)}$, the saddle point formula must be modified according to

$$\int_0^\infty x^b e^{-ax^p/p!}\,dx = \frac{1}{p}\left(\frac{a}{p!}\right)^{b+1} p!\Gamma\left(\frac{b+1}{p!}\right).$$

Thus, the argument of the Γ factor is changed from $\frac{1}{p}$ to $\frac{b+1}{p}$, as is the exponent of $|f^{(p)}(z_0)|$ and of n^{-1} in the case of large power estimates. $\triangleleft$

We give below an application to the counting of forests of unrooted trees made of a large number of trees. The problem is relevant to the analysis of random graphs during the phase where a giant component has not yet emerged.

EXAMPLE 10. *Forests of unrooted trees.* The problem here consists in determining the number $F_{m,n}$ of ordered forests, i.e., sequences, made of m (labelled, nonplane) unrooted trees and comprised of n nodes in total. The number of unrooted trees is size n is, by virtue of Cayley's formula, n^{n-2} and its EGF is expressed as $U=T-T^2/2$, where T is the Cayley tree function satisfying $T=ze^T$. Consequently, we have

$$\frac{1}{n!}F_{m,n} = [z^n]\left(T(z)-\frac{1}{2}T(z)^2\right)^m = \frac{1}{2i\pi}\int_{0+}(T-\frac{1}{2}T^2)^m\,\frac{dz}{z^{n+1}}.$$

Like in the analytic proof of the Lagrange Inversion Theorem it proves convenient to adopt $t=T$ as an independent variable, so that $z=te^{-t}$ becomes a dependent variable. Since $dz=(1-t)e^{-t}$, this provides the integral representation:

$$\frac{1}{n!}F_{m,n} = \frac{1}{2i\pi}\int_{0+}(t-\frac{1}{2}t^2)^m e^{nt}(1-t)\,\frac{dt}{t^{n+1}}.$$

The case of interest here is when m and n are linearly related. We thus set $m=\alpha n$, where *a priori* $\alpha\in(0,1)$. Then, the integral representation of $F_{m,n}$ becomes

$$\frac{1}{n!}F_{m,n} = \frac{1}{2i\pi}\int_C e^{nh_\alpha(t)}(1-z)\,\frac{dt}{t}, \qquad h_\alpha(t) := \alpha\log(1-\frac{t}{2})+t+(\alpha-1)\log t, \tag{65}$$

where C encircles 0. This has the form of a "large power" integral. Saddle points are found as usual as zeros of the derivative h'_α; there are two of them given by

$$\zeta_0 = 2-2\alpha, \qquad \zeta_1 = 1.$$

For $\alpha<\frac{1}{2}$, one has $\zeta_0>\zeta_1$ while for $\alpha>\frac{1}{2}$ the inequality is reversed and $\zeta_0<\zeta_1$. In both cases, a simple saddle point analysis succeeds, based on the saddle point nearer to the origin; see Note 38 below. In contrast, when $\alpha=\frac{1}{2}$, the points ζ_0 and ζ_1 coalesce to the common

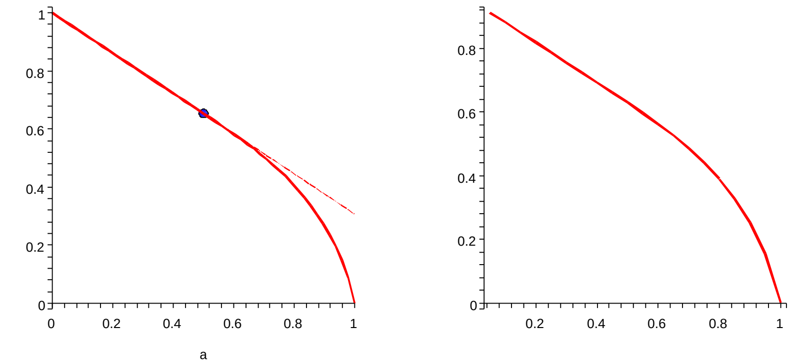


FIGURE 10. The function H governing the exponential rate of the number of forests exhibits a "phase transition" at $\alpha = \frac{1}{2}$ (left); the quantity $\frac{1}{n}\log(F_{m,n}/n!)$ as a function of $\alpha = m/n$ for $n = 200$ (right).

value 1. In this last case, we have $h'_{\frac{1}{2}}(1) = h''_{\frac{1}{2}}(1) = 0$ while $h'''_{\frac{1}{2}}(1) = -2$ is nonzero: there is a *double saddle point* at 1.

▷ **38.** *Forests and simple saddle points.* When $0 < \alpha < \frac{1}{2}$, the number of forests satisfies, for some computable $C_-(\alpha)$, for some computable $C_-(\alpha)$.

$$\frac{1}{n!}F_{n,m} \sim C_-(\alpha)\frac{e^{H_-(\alpha)}}{n^{1/2}}, \qquad H_-(\alpha) = 1 - \alpha \log 2.$$

When $\frac{1}{2} < \alpha < 1$, the number of forests satisfies, for some computable $C_+(\alpha)$,

$$\frac{1}{n!}F_{n,m} \sim C_+(\alpha)\frac{e^{H_+(\alpha)}}{n^{1/2}}, \qquad H_+(\alpha) = \alpha\log\alpha + 2 - 2\alpha + (\alpha - 1)\log(2 - 2\alpha).$$

This results from a routine simple saddle point analysis at ζ_1 and ζ_0 respectively. ◁

The number of forests thus presents two different regimes depending on whether $\alpha < \frac{1}{2}$ or $\alpha > \frac{1}{2}$, and there is a discontinuity of the analytic form of the estimates at $\alpha = \frac{1}{2}$. The situation is reminiscent of "critical phenomena" and phase transitions (e.g., from solid to liquid to gas) in physics, where such discontinuities are encountered. This provides a good motivation to study what happens right at the "critical" value $\alpha = \frac{1}{2}$.

We thus consider the special value $\alpha = \frac{1}{2}$ and set $h \equiv h_{\frac{1}{2}}$. What is to be determined is therefore the number of forests of total size n that are made of $n/2$ trees, assuming naturally n even. Bearing in mind that the double saddle point is at $\zeta = \zeta_0 = \zeta_1 = 1$, one has

$$h(z) = 1 - \frac{1}{3}(z-1)^3 + O((z-1)^4) \qquad (z \to 1).$$

Thus, upon neglecting the tails and localizing the integral to a disc centred at 1 with radius $\delta \equiv \delta(n)$ such that

$$n\delta^3 \to \infty, \qquad n\delta^4 \to 0$$

($\delta = n^{-3/10}$ is suitable), we have the asymptotic equivalence (with y representing $z - 1$)

$$(66) \qquad \frac{1}{n!}F_{m,n} = -\frac{e^{n(1-\frac{1}{2}\log 2)}}{2i\pi}\int_D e^{-ny^3/3}y\,dy + \text{exponentially small},$$

where D is a certain (small) contour containing 0 obtained by transformation from C.

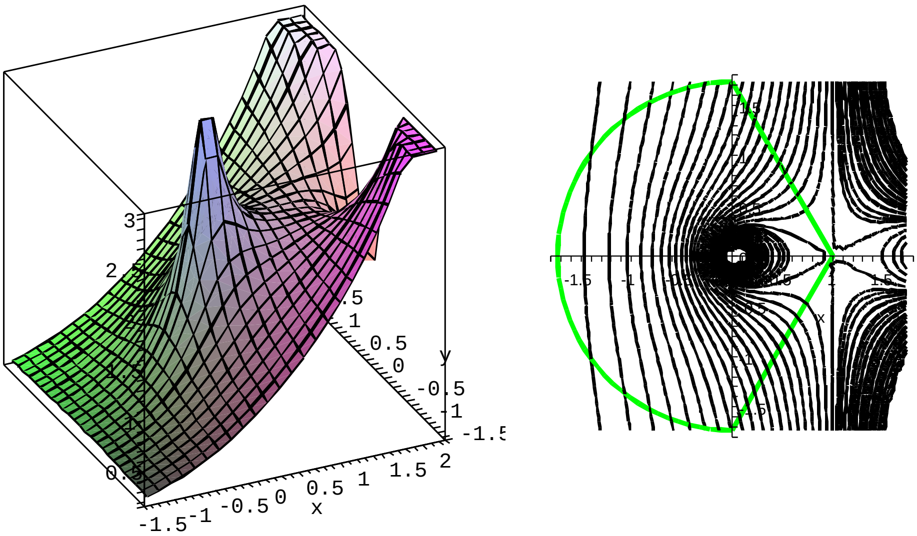


FIGURE 11. A plot of e^h with the double saddle point at 1 (left). The level lines of e^h with valleys (the region higher than $e^{h(1)}$ is darkened) and a legal integration contour (right).

The discussion so far has left aside the choice of the contour C in (65), hence of the geometric aspect of D near 0, which is needed in order to fully specify (66). Because of the minus sign in the third derivative, $h'''(1) = -2$, the three steepest descent half lines stemming from 1 have angles 0, $e^{2i\pi/3}$, $e^{-2i\pi/3}$. This suggest to adopt as original contour C in (65) two symmetric segments stemming from 1 connected by a loop left of 0; see Figure 11. Elementary calculations justify that the contour can be suitably dimensioned so as to remain always below level $h(1)$. See also the right drawing of Figure 11 where the level curves of the valleys *below* the saddle point are drawn together with a legal contour of integration that winds about 0.

Once the original contour of integration has been fixed, the orientation of D in (66) is fully determined. After effecting the further change of variables $y = wn^{-1/3}$ and completing the tails, we find

$$\frac{1}{n!}F_{m,n} \sim \frac{\lambda}{n^{2/3}} e^{n(1-\frac{1}{2}\log 2)}, \qquad \lambda = -\frac{1}{2i\pi}\int_E e^{-y^3/3} y\, dy, \tag{67}$$

where E connects $\infty e^{-2i\pi/3}$ to 0 then to $\infty e^{2i\pi/3}$. The evaluation of the integral giving λ is now straightforward (in terms of the Gamma function), which gives

$$\frac{1}{n!}F_{n/2,n} \sim \frac{2\cdot 3^{-1/3}}{\Gamma(\frac{2}{3})} e^{n(1-\frac{1}{2}\log 2)} n^{-2/3}.$$

The number *three* is characteristically ubiquitous in the formula. (The formula displays the exponent $\frac{2}{3}$ instead of $\frac{1}{3}$ in the general case (66) because of the factor $(1-z)$ present in the integral representation (65), which vanishes at the saddle point 1; see also Note 37.) END OF EXAMPLE 10. □

The problem of analysing random forests composed of a large number of trees has been first addressed by the Russian School, most notably Kolchin and Britikov. We refer the reader to Kolchin's book [**277**, Ch. I] where nearly thirty pages are devoted to a deeper study of the number of forests and of associated parameters. Kolchin's approach is however based on an alternative presentation in terms of sums of independent random variables and stable laws of index $\frac{3}{2}$. As it turns out there is a striking parallel with the analysis of the growth of the random graph in the critical region, when the random graph stops resembling a large collection of disconnected tree components.

An almost sure sign of (hidden or explicit) monkey saddles is the occurrence of $\Gamma(\frac{1}{3})$ factors in the final formulæ and cube roots in exponents of powers of n. It is in fact possible to go much further than we have done here with the analysis of forests (where we have stayed right at the critical point) and provide asymptotic expressions that describe the transition between regimes, here from $A^n n^{-1/2}$, to $B^n n^{-2/3}$, then to $C^n n^{-1/2}$. The analysis then appeals to the theory of coalescent saddle points well developed by applied mathematicians (see, e.g., the exposition in [**55, 334, 439**]) and the already evoked rôle of the Airy function. We do not pursue this thread further since it properly belongs to multivariate asymptotics. It is exposed in a detailed manner in an article of Banderier, Flajolet, Schaeffer, and Soria [**22**] relative to the size of the core in a random map from, on which our presentation of forests has been modelled.

The results of several studies conducted towards the end of the previous millennium do suggest that, amongst threshold phenomena and phase changes, there is a fair amount of universality in descriptions of combinatorial and probabilistic problems by means of multiple and coalescing saddle points. In particular $\Gamma(\frac{1}{3})$ factors and the Airy function surface recurrently in the works of Flajolet, Janson, Knuth, Łuczak and Pittel [**162, 244**], which are relative to the Erdős–Renyi random graph model in its critical phase; see also [**172**] for a partial explanation. The occurrence of the Airy area distribution (in the context of certain polygon models related to random walks) can be related to this orbit of techniques, as first shown by Prellberg [**353**], and strong numerical evidence evoked in Chapter V suggests that this should extend to the difficult problem of self-avoiding walks [**363**]. Airy-related distributions also appear in problems relative to the random satisfiability of boolean expressions [**57**], the path length of trees [**404, 402, 403**], as well as cost functionals of random allocations[**168**]. The reasons are sometimes well understood in separate contexts by probabilists, statistical physcists, combinatorialist, and analysts, but a global framework is still missing.

VIII. 8. Notes

Saddle point methods take their sources in applied mathematics, one of them being the asymptotic analysis by Debye (1909) of Bessel functions of large order. Saddle point analysis is sometimes called steepest descent analysis, especially when integration contours strictly coincide with steepest descent paths. Saddle points themselves are also called critical points (*i.e.*, points where a first derivative vanishes). Because of its roots in applied mathematics, the method is well covered by the literature in this area, and we refer to the books by Olver [**334**], Henrici [**229**], or Wong [**439**] for extensive discussions. A vivid introduction to the subject is to be found in De Bruijn's book [**93**]. We also recommend Odlyzko's impressive survey [**329**].

To a large extent, saddle point methods have made an irruption in combinatorial enumerations in the 1950's. Early combinatorial papers were concerned with permutations (involutions) or set partitions: this includes works by Moser and Wyman [**320, 321, 322**] that are mostly directed towards entire functions.

Hayman's approach [**227**] which we have exposed here (see also [**439**]) is notable in its generality as it envisions saddle point analysis in an abstract perspective, which makes it possible to develop general closure theorems. A similar thread was followed by Harris and Schoenfeld who gave stronger conditions then allowing full asymptotic expansions [**225**]; Odlyzko and Richmond [**331**] were successful in connecting these conditions with Hayman admissibility. Another valuable work is Wyman's extension to nonpositive functions [**448**].

Interestingly enough, developments that parallel the ones in analytic combinatorics have taken place in other regions of mathematics. Erwin Schrödinger introduced saddle point methods in his lectures [**379**] at Dublin in 1944 in order to provide a rigorous foundation to some models of statistical physics that closely resemble balls-in-bins models. Daniels' publication [**87**] of 1954 is a historical source for saddle point techniques in probability and statistics, where refined versions of the central limit theorem can be obtained. (See for instance the description in Greene and Knuth's book [**214**].) Since then, the saddle point method has proved a useful tool for deriving Gaussian limiting distributions. We have given here some idea of this approach which is to be developed further in a later chapter, where we shall discuss some of Canfield's results [**71**]. Analytic number theory also makes a heavy use of saddle point analysis. In additive number theory, the works by Hardy, Littlewood, and Ramanujan relative to integer partitions have been especially influential, see for instance Andrews' book [**10**] and Hardy's *Lectures* on Ramanujan [**224**] for a fascinating perspective. In multiplicative number theory, generating functions take the form of Dirichlet series while Perron's formula replaces Cauchy's formula. For saddle point methods in this context, we refer to Tenenbaum's book [**410**] and his seminar survey [**409**].

A more global perspective on limit probability distributions and saddle point techniques will be given in the next chapter as there are strong relations to the quasi-powers framework developed there, to local limit laws, and to large deviation estimates. General references for some of these aspects of the saddle point method are the articles of Bender–Richmond [**26**], Canfield [**71**], and Gardy [**193, 194, 195**]. Regarding multiple saddle points and phase transitions, we refer the reader to references provided at the end of Section VIII. 7.

Part C

RANDOM STRUCTURES

IX

Multivariate Asymptotics and Limit Distributions

Un problème relatif aux jeux du hasard,
proposé à un austère janseniste par un homme du monde,
a été à l'origine du Calcul des Probabilités[1].

— SIMÉON–DENIS POISSON

Contents

Analytic combinatorics[2] deals with the enumeration of combinatorial structures in relation to algebraic and analytic properties of generating functions. The most basic cases are the enumeration of combinatorial classes and the analysis of moments of combinatorial parameters. These involve generating functions in one (formal or complex) variable as discussed extensively in previous chapters. They are consequently essentially *univariate* problems.

Many applications, in combinatorics as well as in the applied sciences, require quantifying the behaviour of *parameters* of combinatorial structures. It is typically useful to know that a random permutation of size n has a number of runs whose average (mean) equals to $(n+1)/2$, but it may be equally important to know to which

[1]"A problem relative to games of chance proposed to an austere Jansenist by a man of the world has been at the origin of the calculus of probabilities." Poisson refers here to the fact that questions of betting and gambling posed by the Chevalier de Méré (who was both a gambler and a philosopher) led Pascal (an austere religious man) to develop some the first foundations of probability theory.

[2]**Warning**: This chapter is still in a *very* preliminary state (November 2004). It is only included at this stage in order to illsutrate the global architecture of *Analytic Combinatorics*.

extent such an average is *representative* of what occurs in simulations or on actual data that obey the randomness model. As a matter if fact, in a random permutation of size $n = 1,000$, it is found that there are about 70% chances that the number of runs be in the interval $990\,..\,1010$. Even more dramatically, for runs and a permutation of size $n = 1,000$ still, there is probability less than 10^{-6} to observe a case that deviates by more than 10% from the mean value; this probability decreases to about 10^{-65} for $n = 10,000$, and is even less than 10^{-653} for $n = 100,000$. As illustrated by such numeric data, there is obvious interest in analysing the "central" region near the mean, as well as in quantifying the risk of finding instances that deviate appreciably from the expected value. These are now typically *bivariate* problems.

It is frequently observed that the histograms of the distribution of a combinatorial parameter (for varying size values) exhibit a common characteristic "shape". In this case, we say that there exists a *limit law*, which may be of the *discrete* or the *continuous* type. Our aim here is to detect such limit laws, and a few examples have already appeared scattered in this book, in the case where they can be reduced to a collection of univariate analyses. This chapter provides a coherent set of analytic techniques dedicated to extracting coefficients of bivariate analytic functions. The mathematics combine methods of complex asymptotic analysis as previously exposed with a small selection of fundamental theorems from the analytic side of classical probability theory.

In simpler cases, limit laws are *discrete* and, when this happens, they often belong to the geometric or Poisson type. In many other cases, limit laws are *continuous*, a prime example being the Gaussian law associated with the famous bell-shaped curve, which surfaces so frequently in elementary combinatorial structures. The goal of this chapter is to offer a fundamental analytic framework for extracting limit laws from combinatorics.

Symbolic methods provide bivariate generating functions for many natural parameters of combinatorial structures. Analytically, the auxiliary variable marking the combinatorial parameter under study then induces a *deformation* of the (univariate) counting generating function. This deformation may affect the type of singularity that the counting generating function presents in various ways. A *perturbation* of univariate singularity analysis is then often sufficient to derive an asymptotic estimate of the probability generating function of a given parameter, when taken over objects of some large size. Continuity theorems from probability theory finally allow us to conclude on the existence of a limit law.

An especially important component of this paradigm is the framework of "Quasi-Powers". Large powers tend to occur for coefficients of generating functions (think of quantities of the form $\approx \rho^{-n}$ that arise from radius of convergence bounds). The collection of deformations of a single counting generating function is then likely to induce for the corresponding coefficients a collection of approximations that involve large powers together with small error terms—these are referred to as quasi-powers. From there, a Gaussian laws is derived along lines that are somewhat reminiscent of the classical central limit theorem of probability theory (expressing the asymptotically Gaussian character of sums of independent random variables).

The direct relation that can be established between combinatorial specifications and asymptotic properties, in the form of limit laws, is especially striking, and it is a characteristic feature of analytic combinatorics. In fact, almost any classical law of probability theory and statistics is likely to occur somewhere in analytic combinatorics. Conversely, almost any simple combinatorial parameter is likely to be governed by an asymptotic law.

IX. 1. Limit laws and combinatorial structures

What is given is a combinatorial class $\mathcal{C}$, labelled or unlabelled, and an integer valued combinatorial parameter χ. There results both a family of probabilistic models, namely for each n the uniform distribution over $\mathcal{C}_n$ that assigns to any $\gamma \in \mathcal{C}_n$ the probability

$$\mathbb{P}(\gamma) = \frac{1}{C_n}, \qquad \text{with} \quad C_n = \operatorname{card}(\mathcal{C}_n),$$

and a corresponding family of random variables obtained by restricting χ to $\mathcal{C}_n$. Under the uniform distribution over $\mathcal{C}_n$, we then have

$$\mathbb{P}_{\mathcal{C}_n}(\chi = k) = \frac{1}{C_n} \operatorname{card} \left\{ \gamma \in \mathcal{C}_n \ \middle| \ \chi(\gamma) = k \right\}.$$

We write $\mathbb{P}_{\mathcal{C}_n}$ to indicate the probabilistic model relative to $\mathcal{C}_n$, but also freely abbreviate it to $\mathbb{P}_n$ or write $\mathbb{P}(\chi_n)$ whenever $\mathcal{C}$ is clear in context.

As n increases, the histograms of the distributions of χ often share a common profile; see Figure 1 for two characteristic examples that we discuss next. Our purpose is to relate such phenomena to the analysis of bivariate generating functions provided by the symbolic method.

Binary words. Let us start by discussing the case of binary words with two simple parameters, one leading to a discrete law, the other to a continuous limit. The example is purposely chosen simple enough that explicit expressions are available for the probability distributions at stake. Nonetheless, it is typical of the approach taken in this chapter, and, once equipped with suitably general theorems, it is hardly more difficult to discuss the number of leaves in a nonplane unlabelled tree or the number of summands in a composition into prime summands.

Take the class $\mathcal{W}_n$ of binary words of length n over the alphabet $\{a, b\}$ and consider the two parameters for $w \in \mathcal{W}$:

$$\chi(w) := \text{ number of initial } a\text{'s in } w, \qquad \xi(w) := \text{ total number of } a\text{'s in } w.$$

Explicit expressions are available for the counts and

$$\begin{aligned} \mathbb{P}_{\mathcal{W}_n}(\chi = k) &= \frac{1}{2^{k+1}} [\![0 \le k < n]\!] + \frac{1}{2^n} [\![k = n]\!], \\ \mathbb{P}_{\mathcal{W}_n}(\xi = k) &= \frac{1}{2^n} \binom{n}{k}. \end{aligned}$$

The probabilities relative to χ then resemble, in the asymptotic limit of large n, the

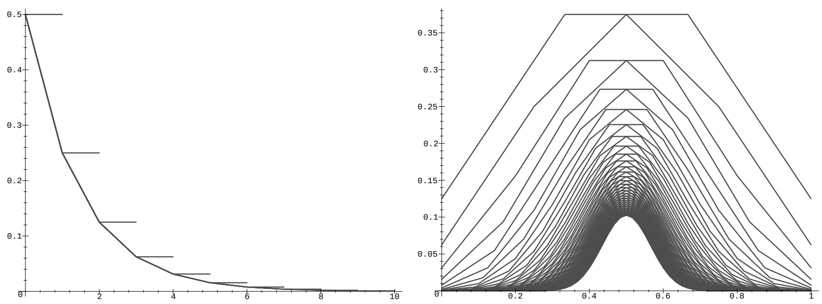


FIGURE 1. Histograms of probability distributions for the number of initial a's in a random binary string (parameter χ, left) and the total number of a's (parameter ξ, right). The histogram corresponding to χ is not normalized and direct convergence to a discrete geometric law is apparent; for ξ, the horizontal axis is scaled to n, and the histograms quickly conform to the bell-shaped curve that is characteristic of a continuous gaussian limit.

geometric distribution. One has, for each k,

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{W}_n}(\chi = k) = \frac{1}{2^k}.$$

(In this simple case, it is even true that the limit is exactly attained as soon as $n > k$.) We say that there is a *limit law of the discrete type* for χ, this limit law being a geometric.

In contrast, the parameter ξ has mean $\mu_n := n/2$ and a standard deviation $\sigma_n := \frac{1}{2}\sqrt{n}$. One should then centre and scale the parameter χ, introducing (over $\mathcal{W}_n$) the "standardized" (or "normalized") random variable

$$X_n^\star := \frac{\xi - n/2}{\frac{1}{2}\sqrt{n}},$$

which can be considered to lie in a fixed scale. It then becomes possible to examine the behaviour of the (cumulative) distribution function $\mathbb{P}(X_n^\star \le y)$ for some fixed y. In terms of χ itself, this means that we are considering $\mathbb{P}(\xi \le \mu_n + y\sigma_n)$ for real values of y. Then, the classical approximation of the binomial coefficients yields the approximation:

$$\lim_{n\to\infty} \mathbb{P}(\xi \le \mu_n + y\sigma_n) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-t^2/2}\, dt, \tag{1}$$

which can be derived by summation from the "local" approximation

$$\frac{1}{2^n} \binom{n}{\frac{1}{2}n + \frac{1}{2}y\sqrt{n}} \sim \frac{e^{-y^2/2}}{\sqrt{\pi n}}. \tag{2}$$

We now say that there is a *limit law of the continuous type* for ξ, this limit law being a Gaussian.

Though cases mixing the discrete and the continuous are theoretically conceivable (a rare instance arises in the theory of map enumerations and "cores", see [**22**]), the discrete–continuous dichotomy applies to most combinatorial cases of interest.

Distributional properties. As illustrated by the previous discussion, there are two major types of convergence that the discrete distribution of a combinatorial parameter may satisfy:

$$\boxed{\text{Discrete} \longrightarrow \text{Discrete}} \qquad \text{and} \qquad \boxed{\text{Discrete} \longrightarrow \text{Continuous}}.$$

In accordance with the general notion of convergence in distribution (or weak convergence, see APPENDIX C: *Convergence in law*, p. 690), we shall say that a *limit law* exists for a parameter if there is convergence of the corresponding family of (cumulative) distribution functions. In the broad context of convergence of probability laws, one also speaks of a *central limit law* when such a convergence holds. In the discrete-to-discrete case, convergence is established without standardizing the random variables involved. In the discrete-to-continuous case, the parameter should be centred at its mean and scaled by its standard deviation, like in (1).

There is also interest in obtaining a *local limit law*, which, when available, quantifies individual probabilities and probability densities, like in (2). The distinction between local and central limits is immaterial in the discrete-to-discrete case, where the existence of one type of law implies the other. In the discrete-to-continuous case, it is technically more demanding to derive a local limit law than a central one, as stronger analytic properties are required.

The *speed of convergence* in a limit law describes the way the finite combinatorial distributions approach their asymptotic limit. It provides useful information on the quality of asymptotic approximations for finite n models.

Finally, quantifying the "risk" of extreme configurations necessitates estimates on the *tails* of the distributions, that is, the behaviour of the probability distribution far away from its mean. Such estimates are also called *large deviation* estimates. Large deviation theory constitutes a useful complement to the study of central and local limits, as exemplified by the discussion of runs in the introduction to this chapter.

In the remainder of the this chapter, we shall first examine the situation of discrete limits. After this, several sections will be dedicated to the case of continuous limits, with special emphasis on limit laws of the Gaussian type. In each of the two cases, the discussion of central laws starts with a *continuity theorem*, which states conditions under which convergence in law can be established from convergence of transforms. (The transforms in question are probability generating functions for the discrete case, characteristic functions or Laplace transforms otherwise). Refinements, known as the Berry-Esseen inequalities when the limit law is continuous, then relate speed of convergence of the combinatorial distributions to their limit on the one hand, a distance between transforms on the other hand. Put otherwise, distributions are close if their transforms are close. Large deviation estimates are often obtained by a technique of "shifting the mean", which is familiar in probability and statistics. The last section gives brief indications on the occurrence of non-Gaussian laws in the discrete-to-continuous scenario.

Limit laws and bivariate generating functions. In this chapter, the starting point of a distributional analysis is invariably a bivariate generating function

$$F(z,u) = \sum_{n,k} f_{n,k} u^k z^n,$$

where $f_{n,k}$ represents (up to a possible normalization factor) the number of structures of size n in some class $\mathcal{F}$. What is sought is asymptotic information relative to the array of coefficients

$$f_{n,k} = [z^n u^k]F(z,u),$$

which could in principle be approached by an iterated use of Cauchy's coefficient formula,

$$[z^n u^k]F(z,u) = \left(\frac{1}{2i\pi}\right)^2 \int_\gamma \int_{\gamma'} F(z,u)\,\frac{dz}{z^{n+1}}\,\frac{du}{u^{k+1}}.$$

Thus, a *double coefficient extraction* is to be effected. It turns out that it is in general arduous if not unfeasible to approach a bivariate counting problem in this way, so that another route is explored throughout this chapter[3].

First, observe that the specialization at $u = 1$ of $F(z,u)$ gives the counting generating function of $\mathcal{F}$, that is, $F(z) = F(z,1)$. Next, as seen repeatedly starting from Chapter III, the moments of the combinatorial distribution $\{f_{n,k}\}$ for fixed n and varying k are attainable through the partial derivatives at $u = 1$, namely

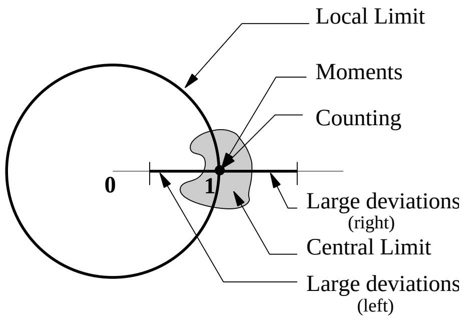
$$\text{first moment} \;\leftrightarrow\; \left.\frac{\partial}{\partial u}F(z,u)\right|_{u=1}, \qquad \text{second moment} \;\leftrightarrow\; \left.\frac{\partial^2}{\partial u^2}F(z,u)\right|_{u=1}, \quad \cdots .$$

In summary: *Counting is provided by the bivariate generating function $F(z,u)$ taken* at $u = 1$*; moments result from the bivariate generating function taken in an* infinitesimal *neighbourhood of* $u = 1$.

Our approach to limit laws will be as follows.

> Estimate the (unormalized) probability generating function
>
> $$f_n(u) := \sum_k f_{n,k} u^k \equiv [z^n]F(z,u).$$
>
> This is viewed a *single coefficient extraction* (extracting the coefficient of z^n) but *parameterized* by u. Thanks to the availability of continuity theorems, the following can be proved for a great many cases of combinatorial interest: *The existence and the shape of the limit law derive from an analysis of the bivariate generating function $F(z,u)$ taken in a* fixed *neighbourhood of* $u = 1$. In addition, thanks to Berry–Esseeen inequalities, *the quality of an asymptotic estimate for $f_n(u)$ translates into a speed of convergence estimate for the corresponding laws.* Also, for the discrete-to-continuous case, *local limit laws derive from consideration of the bivariate generating function $F(z,u)$ taken on the whole of the unit circle,* $|u| = 1$. Finally, *large deviation* estimates are seen to arise from estimates of $f_n(u)$ when u

[3]A collection of recent works by Pemantle and coauthors [**341, 342, 343**] shows however that a well-defined class of bivariate asymptotic problems can be attacked by the theory of functions of several complex variables and a detailed study of the geometry of a singular variety.



Region	*Property*
$u = 1$	Counting
$u = 1 \pm \epsilon$	Moments
$u \in \mathcal{V}(1)$ (neighb.)	Central Limit Law
$\lvert u\rvert = 1$	Local Limit Law
$u \in [\alpha, \beta]$	Large deviations

FIGURE 2. The correspondence between regions of the u–plane and asymptotic properties of combinatorial distributions.

is real and $u < 1$ (left tail) or $u > 1$ (right tail). This is to large extent a reflection of saddle point bounds. In summary: *Large deviations are related to the behaviour of $F(z, u)$ for real values of u in an interval $[\alpha, \beta]$ containing $u = 1$.*

The correspondence between u-domains and properties of laws is summarized in Figure 2.

Singularity perturbation. As seen throughout Chapters IV–VIII, analytic combinatorics approaches the univariate problem of counting objects of size n starting from the Cauchy coefficient integral,

$$[z^n]F(z) = \frac{1}{2i\pi} \int_\gamma F(z) \, \frac{dz}{z^{n+1}}.$$

The singularities of $F(z)$ can be exploited, whether they are of a polar type (Chapters IV and V), algebraic-logarithmic (Chapters VI and VII) or essential and amenable to saddle point methods (Chapter VIII). It is in this way that asymptotic forms of $[z^n]F(z)$ are derived.

From the discussion above, crucial information on combinatorial distributions is accessible from the bivariate generating function $F(z, u)$ when u varies in some domain containing 1. This suggests to consider $F(z, u)$ not so much as an analytic function of two complex variables, where z and u would play a symmetric rôle, but rather as a collection of functions of z indexed by a secondary parameter u. In other words, $F(z, u)$ is considered as a *deformation* of $F() \equiv F(z, 1)$ when u varies in a domain containing $u = 1$. Cauchy's coefficient integral gives

$$f_n(u) \equiv [z^n]F(z, u) = \frac{1}{2i\pi} \int_\gamma F(z, u) \, \frac{dz}{z^{n+1}}.$$

We can then examine the way the parameter u affects the analysis of singularities performed in the aymptotic counting problem of estimating $[z^n]F(z, 1)$. Such an approach is called a *singularity perturbation analysis*. It consists in tracing the effect of a perturbation by u on the standard singularity analysis assocaited to the univariate problem.

The essential feature of the analysis of coefficients by means of complex techniques as seen in Chapters IV–VIII is to be "robust". Being based on explicit estimates of contour integrals, it is usually amenable to smooth perturbations whose effect can be traced throughout calculations. Explicit estimates normally result (though added care in estimations is needed to ensure uniformity). In this chapter, we are going to see many applications of this strategy.

Regarding binary words and the two parameters χ (initial run of a's) and ξ (total number of a's), the general strategy of singularity perturbation instantiates as follows. In the case of W_χ, there are two components in the BGF

$$W_\chi(z,u_0) = \frac{1}{1-u_0 z}\cdot\boxed{\frac{1-z}{1-2z}},$$

and, in essence, the dominant singular part—a simple pole at $z=1/2$—arises from the second component, which does not change when u_0 varies. Accordingly, one has

$$W_\chi(z,u_0) \underset{z\to 1}{\sim} \frac{\frac{1}{2}}{1-\frac{u_0}{2}}W(z), \qquad [z^n]W_\chi(z,u_0)\sim \frac{\frac{1}{2}}{1-\frac{u_0}{2}}\,2^n.$$

The probability generating function of χ over $\mathcal{W}_n$ is then obtained upon dividing by 2^{-n}, and

$$\frac{1}{2^n}\,[z^n]W_\chi(z,u_0)\sim \frac{\frac{1}{2}}{1-\frac{u_0}{2}} = \sum_{k=0}^{\infty}\frac{1}{2^{k+1}}u_0^k,$$

where the last expression is none other than the probability generating function of a discrete law, namely, the geometric distribution of parameter $\frac{1}{2}$. As we shall see in section IX. 2 where we enounce a continuity theorem for probability generating functions, this is enough to conclude that the distribution of χ converges to a geometric law.

In the second case, that of W_ξ, the auxiliary parameter modifies the location of the singularity,

$$W_\xi(z,u_0) = \frac{1}{1-z\,\boxed{(1+u_0)}}.$$

Then, the singular behaviour is strongly dependent upon a singularity at

$$\rho(u_0) = \frac{1}{(1+u_0)}$$

that *moves* as k varies, while the type of singularity (here a simple pole) remains the same. Accordingly, the coefficients obey a "large power law" (here of an exact type) and, as regards the probability generating function of ξ over $\mathcal{W}_n$, one has

$$\frac{1}{2^n}[z^n]W_\xi(z,u_0) = \left(\frac{1}{2\rho(u_0)}\right)^n,$$

This analytical form is reminiscent of the central limit theorem of probability theory after which large powers, corresponding to sums of a large number of independent random variables, entail convergence to a Gaussian law. By continuity theorems for integral transforms exposed in Sections IX. 4, there results a continuous limit law of the Gaussian type in this case.

$F(z,u)$ when $u \approx 1$	*Type of law*	*Method and classes*	
Sing. + expo. fixed	Discrete limit	Subcritical composition	§ xxx
	(Neg. binom, Poisson, ...)	Subcritical Seq., Set, ...	§ xxx
Sing. moves, expo. fixed	Gaussian(n, n)	Supercritical composition	§ xxx
—	—	Meromorphic perturb.	§ xxx
—	—	(Rational fns)	§ xxx
—	—	Sing. Analysis pertub.	§ xxx
—	—	(Alg., implicit fns)	§ xxx
Sing. fixed, expo. moves	Gaussian($\log n, \log n$)	Exp-log struct.	§ xxx
—	—	(Differential eq.)	§ xxx
Sing. + expo. move	Gaussian	[Gao-Richmond [**192**]]	
Essential singularity	often Gaussian	Saddle point perturbation	§ xxx
Discontinuous singular type	non-Gaussian	(Various cases)	§ xxx
—	Stable	Critical composition	§ xxx

FIGURE 3. A rough typology of bivariate generating functions $F(z,u)$ and limit laws.

The foregoing discussion suggests that a "minor" perturbation of bivariate generating function that affects neither the location nor the nature of the singularity could lead to a discrete limit law. A "major" change in exponent or even like here in location is likely to be conducive to a continuous limit law, of which the prime example is the normal distribution. Figure 3 outlines a typology of limit laws in the context of bivariate asymptotics. A bivariate generating function $F(z,u)$ is to be analysed. The deformation induced by u may affect the type of singularity that $F(z,u)$ has in various ways. An adapted complex coefficient extraction then provides various types of limit laws.

IX. 2. Discrete limit laws

Take a class $\mathcal{C}$ on which a parameter χ is defined. This determines for each n a random variable X_n, which is χ restricted to $\mathcal{C}_n$, where $\mathcal{C}_n$ is endowed with the uniform probability distribution. In this section, we give the general definitions and results that are suitable for the discrete-to-discrete situation, where a discrete parameter tends without normalization to a discrete distribution. The corresponding notion of convergence is given in Subsection IX. 2.1. Probability generating functions (PGFs) are important since, by virtue of a continuity theorem stated in Subsection IX. 2.2, convergence in law results from convergence of PGFs. At the same time, the fact that PGFs of two distributions are close entails that the original distribution functions are close. Finally, large deviation estimates for a distribution can be easily related to analytic continuation of its PGFs, a fact introduced in Subsection IX. 2.3. This section organizes some general tools and accordingly we limit ourselves to a single combinatorial application, that of the number of cycles of some small fixed size in a random permutation. The next section will provide a number of deeper applications to random combinatorial structures.

IX. 2.1. Convergence to a discrete law. In order to specify precisely what a limit law is, we base ourselves on the general context described in APPENDIX C: *Convergence in law*, p. 690. The principles exposed there provide for what should be the "right" notion convergence of a family of discrete distributions to a limit discrete distribution. Here is a self-standing definition.

DEFINITION IX.1 (Discrete-to-discrete convergence). *The random variable X_n (suported by $\mathbb{Z}_{\geq 0}$) is said to converge in law to a discrete variable Y supported by $\mathbb{Z}_{\geq 0}$ if for each $k \geq 0$, one has*

$$\lim_{n\to\infty} \mathbb{P}(X_n \leq k) = \mathbb{P}(Y \leq k), \qquad \textit{i.e.}, \quad \lim_{n\to\infty} \sum_{j=k} p_{n,j} = \sum_{j\leq k} q_j, \tag{3}$$

where $p_{n,k} = \mathbb{P}(X_n \leq k)$ and $q_k := \mathbb{P}(X = k)$. One also says that the parameter χ on $\mathcal{C}$ admits a limit law *of type X.*

Convergence is said to take place at speed ϵ_n *if*

$$\sup_k \left| \sum_{j\leq k} \mathbb{P}(X_n = j) - \sum_{j\leq k} q_j \right| \leq \epsilon_n, \tag{4}$$

The condition in (3) can be rewritten in terms of the distribution functions F_n, G of X_n, Y as

$$\lim F_n(k) = G(k),$$

pointwise for each k. When such a property of type (3) relative to distribution functions holds, it is also called a "central" limit law. (One good reason for this terminology is that convergence of distribution functions is principally informative in the "central part" of the distribution, where a fair proportion of the probability mass lies.) By differencing, the condition of (3) is clearly equivalent to the condition that, for each k,

$$\lim_n p_{n,k} = q_k, \tag{5}$$

and δ_n is called a local speed of convergence if

$$\sup_k |p_{n,k} - q_k| \leq \delta_n.$$

The property (5) is said to constitute a *local limit law*, as probabilities $p_{n,k}$ are estimated "locally". Thus: *For the convergence of a discrete law to a discrete law, there is complete equivalence between the existence of central and local limits.* Note 1 below shows elementarily that there always exists a speed of convergence that *tends to* 0 as n tends to infinity. In other words, plain convergence of distribution functions or of individual probabilities implies uniform convergence (this is in fact a general phenomenon).

$\triangleright$ **1.** *Uniform convergence.* Local and central convergences to a discrete limit law are always *uniform*. In other words, there always exists speeds ϵ_n, δ_n tending to 0 as $n \to \infty$.

Assume simply the condition (3) and its equivalent form (5). Fix a small $\epsilon > 0$. First dispose of the tails: there exists a k_0 such that $\sum_{k\geq k_0} q_k \leq \epsilon$, so that $\sum_{k<k_0} q_k > 1-\epsilon$. Now, by simple convergence, there exists an n_0 such that, for all n larger than n_0 and each $k < k_0$, $|p_{n,k} - q_k| < \epsilon/k_0$. Thus, we have $\sum_{k<k_0} p_{n,k} > 1 - 2\epsilon$, hence $\sum_{k\geq k_0} p_{n,k} \leq 2\epsilon$. In other words, $\sum_{k\geq k_0} q_k$ and $\sum_{k\geq k_0} p_{n,k}$ are both in $[0, 2\epsilon]$. There results that convergece of distribution functions is uniform, with speed 5ϵ at most. At the same time, the local speed δ_n is at most 4ϵ. $\lhd$

$\rhd$ **2.** *Speed in local and central estimates.* Let M_n be the spread of χ on $\mathcal{C}_n$ defined as $M_n := \max_{\gamma\in\mathcal{C}_n} \chi(\gamma)$. Then, a speed of convergence in (4) is given by

$$\epsilon_n := M_n\delta_n + \sum_{k>M_n} p_k.$$

(Refinements of these inequalities can obtained from tail estimates detailed below.) $\lhd$

$\rhd$ **3.** *Total variation distance.* The *total variation distance* between X and Y is classically

$$d_{TV}(X,Y) := \sup_{E\subseteq\mathbb{Z}_{\geq 0}} |\mathbb{P}_Y(E) - \mathbb{P}_X(E)| = \frac{1}{2}\sum_{k\geq 0} |\mathbb{P}(Y=k) - \mathbb{P}(X=k)|\,.$$

(Equivalence between the two forms is established elementarily by considering the particular E for which the supremum is attained.) The argument of Note 1 shows that convergence in distribution also implies that the total variation distance between X_n and X tends to 0. In addition, by Note 2, one has $d_{TV}(X_n, X) \leq M_n\delta_n + \sum_{k>M_n} p_k$. $\lhd$

$\rhd$ **4.** *Escape to infinity.* The sequence X_n, where

$$\mathbb{P}\{X_n = 0\} = \frac{1}{3}, \quad \mathbb{P}\{X_n = 1\} = \frac{1}{3}, \quad \mathbb{P}\{X_n = n\} = \frac{1}{3},$$

does not satisfy a discrete limit law in the sense above, although $\lim_n \mathbb{P}\{X_n = k\}$ exists for each k. Some of the probability mass escapes to infinity and, in a way, convergence takes place in $\mathbb{Z}\cup\{+\infty\}$. $\lhd$

A highly plausible indication of the occurrence of a discrete law is the fact that $\mu_n = O(1)$, $\sigma_n = O(1)$. Examination of initial entries in the table of values of the probabilities will then normally permit one to decide whether a limit law holds.

EXAMPLE 1. *Singleton cycles in permutations.* The case of the number of singleton cycles (cycles of length 1) in a random permutation of size n illustrates the basic definitions and it can be analysed with minimal analytic apparatus. The exponential BGF is

$$P(z,u) = \frac{\exp(z(u-1))}{1-z},$$

which determines the mean $\mu_n = 1$ and the standard deviation $\sigma_n = 1$ (for $n \geq 2$). The table of numerical values of the probabilities $p_{n,k} = [z^n u^k]P(z,u)$ immediately tells what goes on.

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$n=4$	0.375	0.333	0.250	0.000	0.041	
$n=5$	0.366	0.375	0.166	0.083	0.000	0.008
$n=10$	0.367	0.367	0.183	0.061	0.015	0.003
$n=20$	0.367	0.367	0.183	0.061	0.015	0.003

The exact distribution is easily extracted from the bivariate GF,

$$p_{n,k} := [z^n u^k]P(z,u) = \frac{1}{k!}[z^{n-k}]\frac{e^{-z}}{1-z} = \frac{d_{n-k}}{k!},$$

where $n!d_n$ is the number of derangements of size n, that is,

$$d_n = [z^n]\frac{e^{-z}}{1-z} = \sum_{j=0}^{n} \frac{(-1)^j}{j!}$$

Asymptotically, one has $d_n \sim e^{-1}$. Thus, for fixed k, we have

$$\lim_{n\to\infty} p_{n,k} = p_k, \qquad p_k = \frac{e^{-1}}{k!}.$$

As a consequence, the distribution of singleton cycles in a random permutation of large size tends to a Poisson law of rate $\lambda = 1$.

Convergence is quite fast. Here is a table of differences, $\delta_{n,k} = p_{n,k} - \frac{e^{-1}}{k!}$:

	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$
$n=10$	$2.3\,10^{-8}$	$-2.5\,10^{-7}$	$1.2\,10^{-6}$	$-3.7\,10^{-6}$	$7.3\,10^{-6}$	$1.0\,10^{-5}$
$n=20$	$1.8\,10^{-20}$	$-3.9\,10^{-19}$	$3.9\,10^{-18}$	$-2.4\,10^{-17}$	$1.1\,10^{-16}$	$-3.7\,10^{-16}$

The speed of convergence is easily bounded. One has $d_n = e^{-1} + O(1/n!)$, by the alternating series property, so that

$$p_{n,k} = \frac{e^{-1}}{k!} + O\left(\frac{1}{k!\,(n-k)!}\right) = \frac{e^{-1}}{k!} + O\left(\frac{1}{n!}\binom{n}{k}\right) = \frac{e^{-1}}{k!} + O\left(\frac{2^n}{n!}\right).$$

As a consequence, one obtains local (δ_n) and central (ϵ_n) speed estimates

$$\delta_n = O\left(\frac{2^n}{n!}\right), \qquad \epsilon_n = O\left(\frac{n2^n}{n!}\right).$$

These bounds are quite tight. For instance one computes that $\delta_{50} \doteq 1.5\,10^{-52}$ while the quantity $2^n/n!$ evaluates to $3.7\,10^{-50}$. . . . . END OF EXAMPLE 1. □

IX. 2.2. Continuity theorem for PGFs. A higher level approach to discrete limit laws in analytic combinatorics is based on asymptotic estimates of $p_n(u)$, the PGF of the random variable X_n. If, for sufficiently many values of u, one has

$$p_n(u) \to q(u) \qquad (n \to +\infty),$$

one can infer that the coefficients $p_{n,k} = [u^k]p_n(u)$ (for any fixed k) tend to the limit q_k with generating function $q(u)$. A *continuity theorem* for characteristic functions describes precisely sets of conditions under which convergence of probability generating functions to a limit entails convergence of coefficients to a limit, that is to say the occurrence of a discrete limit law. We state here a continuity theorem with very general analytic conditions.

THEOREM IX.1 (Continuity Theorem, discrete laws). *Let Ω be an arbitrary set contained in the unit disc and having at least one accumulation point in the* interior *of the disc. Assume that the PGFs $p_n(u) = \sum_{k\ge 0} p_{n,k}u^k$ and $q(u) = \sum_{k\ge 0} q_k u^k$ are such that there is convergence,*

$$\lim_{n\to+\infty} p_n(u) = q(u),$$

pointwise *for each u in Ω. Then a discrete limit law holds in the sense that, for each k,*

$$\lim_{n\to+\infty} \sum_{j\le k} p_{n,j} = \sum_{j\le k} q_j.$$

PROOF. The $p_n(u)$ are *a priori* analytic in $|u| < 1$ and uniformly bounded by 1 in modulus throughout $|u| \le 1$. Vitali's Theorem is a classical result of analysis whose statement (see [**411**, p. 168] or [**229**, p. 566]) is as follows:

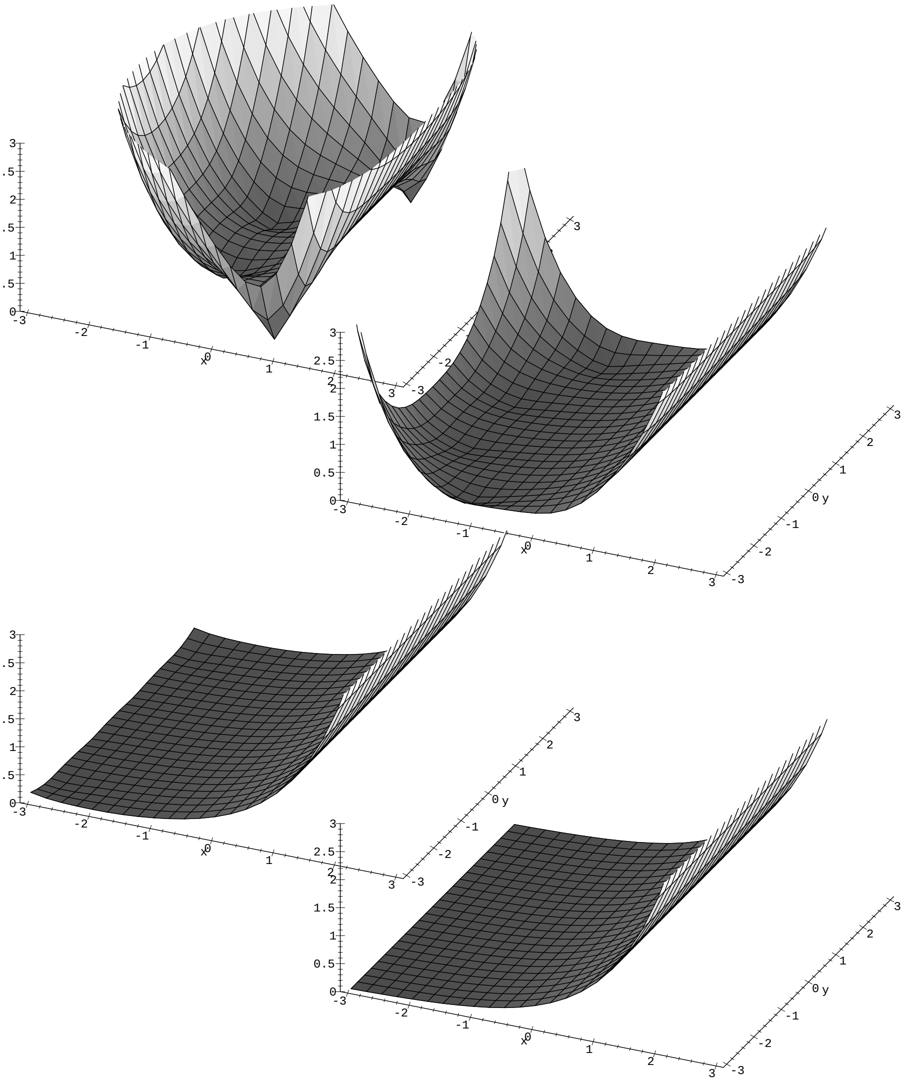


FIGURE 4. The PGFs of singleton cycles in random permutations of size $n = 4, 8, 12$ (left to right and top to bottom) illustrate convergence to the limit PGF of the Poisson(1) distribution (bottom right). Here the modulus of each PGF for $|\Re(u)|, |\Im(u)| \le 3$ is displayed.

Vitali's theorem. *Let $\mathcal{F}$ be a family of analytic functions defined in a region S (i.e., an open connected set) and uniformly bounded on every compact subset of S. Let $\{f_n\}$ be a sequence of functions of $\mathcal{F}$ that converges on a set $\Omega \subset S$ having a point of accumulation $q \in S$. Then $\{f_n\}$ converges in all of S, uniformly on every compact subset $T \subset S$.*

Here, S is the open unit disc on which all the $p_n(u)$ are bounded. The sequence in question is $\{p_n(u)\}$. By assumption, there is convergence of $p_n(u)$ to $q(u)$ on Ω. Vitali's theorem implies that this convergence is uniform in any compact subdisc of the unit disc, for instance, $|u| \le \frac{1}{2}$. Then, Cauchy's coefficient formula provides

$$\begin{aligned} q_k &= \frac{1}{2i\pi}\int_{|u|=1/2} q(u)\,\frac{du}{u^{k+1}} \\ &= \lim_{n\to\infty}\frac{1}{2i\pi}\int_{|u|=1/2} p_n(u)\,\frac{du}{u^{k+1}} \\ &= \lim_{n\to\infty} p_{n,k}. \end{aligned} \tag{6}$$

Uniformity granted by Vitali's theorem combined with continuity of the contour integral (with respect to the integrand) establishes the statement. □

Feller gives the sufficient set of conditions: $p_n(u) \to q(u)$ pointwise for all real $u \in]0,1[$; see [**133**, p. 280] for a proof that only involves elementary real analysis. It is perhaps surprising that very different sets can be taken, for instance,

$$\Omega = \left[-\tfrac{1}{3}, -\tfrac{1}{2}\right], \qquad \Omega = \left\{\tfrac{1}{n}\right\}, \qquad \Omega = \left\{\tfrac{\sqrt{-1}}{2} + \tfrac{1}{2^n}\right\}.$$

The next statement relates a measure of distance between two PGFS, $p(u)$ and $q(u)$ to the distance betwen distributions. It is naturally of interest when quantifying speed of convergence to the limit in the discrete-to-discrete case.

THEOREM IX.2 (Speed of convergence, discrete laws). *Consider two discrete laws supported by $\mathbb{Z}_{\ge 0}$, with corresponding distribution functions $F(x), G(x)$ and probability generating functions $p(u), q(u)$.*

(i) *Assume that the laws have first moments. Then, for any $T \in (0,\pi)$, one has, for some absolute constants $c = \frac{1}{4}$,*

$$\sup_k |F(k) - G(k)| \le c\int_{-T}^{+T} \frac{|p(e^{it}) - q(e^{it})|}{t}\,dt + \frac{c}{T}\sup_{T\le|t|\le\pi}\left|p(e^{it}) - q(e^{it})\right|. \tag{7}$$

(ii) *Assume that $p(u)$ and $q(u)$ are analytic in $|u| < \rho$ for some $\rho > 1$. Then, for any r satisfying $1 < r < \rho$, one has*

$$\sup_k |F(k) - G(k)| \le c(r)\sup_{|u|=r}|p(u) - q(u)|, \qquad c(r) := \frac{1}{r(r-1)}.$$

PROOF. (i) Observe first that $p(1) = q(1) = 1$, so that the integrand is of the form $\frac{0}{0}$ at $u = 1$. By APPENDIX C: *Transforms of distributions*, p. 686, the existence of first moments, say μ and ν, implies that, for small t, one has $p(e^{it}) - q(e^{it}) = (\mu - \nu)t + o(t)$, so that the integral is well defined.

For any given k, Cauchy's coefficient formula provides

$$F(k) - G(k) = \frac{1}{2i\pi}\int_\gamma \frac{p(u) - q(u)}{1-u}\,\frac{du}{u^{k+1}}, \tag{8}$$

where γ is the circle $|u| = 1$. (The factor $(1-u)^{-1}$ sums coefficients.) Set $u = e^{it}$ and split the interval of integration accordingly. For all t, one has

$$\left|\frac{t}{e^{it}-1}\right| \le \frac{\pi}{2}.$$

This makes it possible to replace $(1-u)^{-1}$ by $1/t$, up to a constant multiplier. The statement follows upon splitting the interval of integration according to $|t| \le T$ and $|t| > T$, and then applying trivial bounds.

(ii) Start again from (8), but integrate along $|u| = r$. Trivial bounds provide the statement. ☐ The first form is universal holds with strictly minimal assumptions (existence of expectations); the second form is a priori only usable for distributions that have exponential tails. In the context of limit laws, the first form of the theorem serves to relate the distance on the unit circle between the PGF $p_n(u)$ of a combinatorial parameter and the limit PGF $q(u)$ to the speed of convergence to the limit law. (In this sense, it prefigures the Berry-Esseen inequalities discussed in the continuous context below.)

EXAMPLE 2. *Cycles of length m in permutations.* Let us first revisit the case of singleton cycles, $m = 1$, in this new light. The BGF $P(z,u) = e^{z(u-1)}/(1-z)$ has for each u a simple pole at $z = 1$ and is otherwise analytic in $\mathbb{C} \setminus \{1\}$. Thus, a meromorphic analysis provides instantly, pointwise for any fixed u,

$$[z^n]F(z,u) = e^{(u-1)} + O(R^{-n}),$$

with any $R > 1$. This, by the continuity theorem, Theorem IX.1, implies convergence to a Poisson law.

Next, one should estimate a distance between characteristic functions over the unit circle. One has (for $u = e^{it}$)

$$p_n(u) - q(u) = [z^n]\frac{e^{z(u-1)} - e^{(u-1)}}{1-z}.$$

There is a removable singularity at $z = 1$. Thus, integration over the circle $|z| = 2$ in the z-plane coupled with trivial bounds yields

$$|p_n(u) - q(u)| \le 2^{-n} \sup_{|z|=2} \left|e^{z(u-1)} - e^{(u-1)}\right| = O\left(2^{-n}|1-u|\right).$$

One can then apply Theorem IX.2 with an arbitrary choice of T to the effect that a speed of convergence to the limit is $O(2^{-n})$. (Any $O(R^{-n})$ is possible by the same argument.)

This approach generalizes to the number of m–cycles in a random permutation. The exponential BGF is

$$F(z,u) = \frac{e^{(u-1)z^m/m}}{1-z}.$$

Then, singularity analysis of the meromorphic function of z (for u fixed) gives immediately

$$\lim_{n\to\infty} [z^n]F(z,u) = e^{(u-1)/m}.$$

The right side of this equality is none other than the PGF of a Poisson law of rate $\lambda = \frac{1}{m}$. The continuity theorem and the first form of the speed of convergence theorem then imply:

PROPOSITION IX.1 (m-Cycles in permutations). *The number of m–cycles in a random permutation of large size converges in law to a Poisson distribution of rate $1/m$ with speed of convergence $O(R^{-n})$ for any $R > 1$.*

This vastly generalizes our previous observations on singleton cycles. END OF EXAMPLE 2. □

▷ **5.** *Poisson law for rare events.* Consider the Bernoulli distribution with PGF $(p+qu)^n$. If q depends on n in such a way that $q = \lambda/n$ for some fixed λ, then the limit law of the Bernoulli random variable is Poisson of rate λ. (This "law of small numbers" explains the Poisson character of activity in radioactive decay as well as the probability of accidental deaths of soldiers in the Prussian army resulting from the kick of a horse [Bortkiewicz, 1898].) ◁

IX. 2.3. Large deviations. In the case of discrete limit laws, the study of large deviations is related to saddle-point bounds and is consequently often quite easy. We give with a general statement which is nothing but a rephrasing of saddle point bounds (Chapter IV) in the context of discrete probability distributions.

THEOREM IX.3 (Large deviations, discrete laws). *Let $p(u) = E(u^X)$ be a probability generating function that is analytic for $|u| \le r$ where r is some number satisfying $r > 1$. Then, the following "local" and "central" large deviation bounds hold:*

$$\mathbb{P}(X = k) \le \frac{p(r)}{r^k}, \qquad \mathbb{P}(X > k) \le \frac{p(r)}{r^k(r-1)}.$$

PROOF. The local bound is a direct consequence of saddle point bounds given in Chapter IV. The central bound derives from the equality

$$\mathbb{P}(X > k) = \frac{1}{2i\pi}\int_{|u|=r} p(u)\left(1 + \frac{1}{u} + \frac{1}{u^2} + \cdots\right)\frac{du}{u^{k+2}} = \frac{1}{2i\pi}\int_{|u|=r} p(u)\frac{du}{u^{k+1}(u-1)},$$

upon applying trivial bounds. □

In accordance with this theorem and as is easily checked directly, the geometric and the negative binomial laws have exponential tails; the Poisson law has a "super-exponential" tail, being $O(r^{-k})$ for any $r > 1$, as the PGF is entire. (See definitions in APPENDIX C: *Special distributions*, p. 688.) By their nature, the bounds can be simultaneously applied to a whole family of probability generating functions. Hence their use in obtaining uniform estimates in the context of limit laws. The bound provided always exhibits a geometric decay in the value of k—this is both a stength and a limitation on the method.

IX. 3. Combinatorial instances of discrete laws

In this section, we focus our attention on a general analytic schema based on compositions. The subcritical case of this schema is such that the perturbations induced by the secondary variable (u) affect neither the location nor the nature of the basic singularity involved in the univariate counting problem. The limit laws are then of the discrete type: for sequences, labelled sets, and labelled cycles, theese limit laws are invariably of the negative binomial ($NB[2]$), Poisson, and geometric type, respectively. Additionally, it is easy to describe the profiles of combinatorial objects resulting from such subcritical constructions.

First, we consider the general *composition schema*,

$$F(z, u) = g(uh(z)).$$

This schema expresses over generating functions the combinatorial operation $\mathcal{G}[\mathcal{H}]$ of *substitution* of components $\mathcal{H}$ enumerated by $h(z)$ inside "templates" $\mathcal{G}$ enumerated by $g(z)$. (See Chapters I and II for the unalabelled and labelled versions.) The variable z marks size as usual, and the variable marks the size of the $\mathcal{G}$ template.

We assume globally that g and h have nonnegative coefficients and that $h(0) = 0$ so that the composition $g(h(z))$ is well-defined. We let ρ_g and ρ_h denote the radii of convergence of f and g, and define

$$\tau_g = \lim_{x \to \rho_g^-} g(x) \qquad \text{and} \qquad \tau_h = \lim_{x \to \rho_h^-} h(x). \tag{9}$$

The (possibly infinite) limits exist due to nonnegativity of coefficients. As already seen in Chapter VI, three cases are to be distinguished.

DEFINITION IX.2. *The composition schema $g(uh(z))$ is said to be:* subcritical *if $\tau_h < \rho_g$,* critical *if $\tau_h = \rho_g$,* supercritical *if $\tau_h > \rho_g$.*

In terms of singularities, the behaviour of $g(h(z))$ at its dominant singularity is dictated by the dominant singularity of g (subcritical case), or by the dominant singularity of f (supercritical case), or it should involve a mixture of the two (critical case). This section discusses the *subcritical* case. First, a general statement about subcritical compositions:

PROPOSITION IX.2 (Subcritical composition). *Consider the bivariate composition scheme $F(z, u) = g(uh(z))$. Assume that $g(z)$ and $h(z)$ satisfy the* subcriticality *condition $\tau_h < \rho_g$, and that $h(z)$ has a unique singularity at $\rho = \rho_h$ on its disc of convergence, which is of the algebraic–logarithmic type*

$$h(z) = \tau - c(1 - \frac{z}{\rho})^\lambda + o\left((1 - \frac{z}{\rho})^\lambda\right),$$

where $\tau = \tau_h$, $c \in \mathbb{R}^+$, $0 < \lambda < 1$. Then, a discrete limit law holds,

$$\lim_{n \to \infty} \frac{f_{n,k}}{f_n} = q_k, \qquad q_k = \frac{k g_k \tau^{k-1}}{g'(\tau)},$$

with probability generating function $q(u) = \dfrac{ug'(\tau u)}{g'(\tau)}$.

What stands out is that, via its PGF, *the limit law is a direct reflection of the derivative of the outer function involved in the composition*.

PROOF. First, for the univariate problem, since $g(z)$ is analytic at τ, the function $g(h(z))$ is singular at ρ_h and is analytic in a Δ–domain. Its singular expansion is obtained by composing the regular expansion of $g(z)$ at τ with the singular expansion of $h(z)$ at ρ_h:

$$F(z) \equiv g(h(z)) = g(\tau) - cg'(\tau)(1 - z/\rho)^\lambda(1 + o(1)).$$

Thus, $F(z)$ satisfies the conditions of singularity analysis, and

$$f_n \equiv [z^n]F(z) = -\frac{cg'(\tau)}{\Gamma(-\lambda)} n^{-\lambda-1}(1 + o(1)). \tag{10}$$

Also, the mean and variance of the distribution are clearly $O(1)$.

Next, for the bivariate problem, fix any u with, say, $u \in (0,1)$. The BGF $F(z,u)$ is still singular at $z=\rho$, and its singular expansion obtained from $F(z,u)=g(uh(z))$ by composition, is

$$\begin{array}{rcl} F(z,u)=g(uh(z)) & = & g(u\tau - cu(1-z/\rho)^{\lambda} + o((1-z/\rho)^{\lambda})) \\ & = & g(u\tau) - cug'(u\tau)(1-z/\rho)^{\lambda} + o((1-z/\rho)^{\lambda}). \end{array}$$

Thus, singularity analysis implies immediately:

$$\lim_{n\to\infty} \frac{[z^n]F(z,u)}{[z^n]F(z,1)} = \frac{ug'(u\tau)}{g'(\tau)}.$$

By the continuity theorem for PGFs, this is enough to imply convergence to the discrete limit law with PGF $ug'(\tau u)/g'(\tau)$, and the proposition is established. $\square$

Under the subcritical composition scheme, it is also true that the tails have a uniformly geometric decay. Let u_0 be any number of the interval $(1, \rho_g/\tau_h)$. Then $f(z,u_0)$ a a function of z is analytic near the origin with a dominant singularity at ρ_h obtained by composing the regular expansion of g with the singular expansion of h:

$$f(z,u_0) = h(u_0\tau_h) - ch'(u_0\tau_h)(1-z/\rho)^{\lambda} + o((1-z/\rho)^{\lambda}).$$

There results the asymptotic estimate

$$p_{n,k} = \frac{[z^n]f(z,u_0)}{[z^n]f(z,1)} \sim h'(u_0\tau_h).$$

Thus, for some constant $K \equiv K(u_0)$, one has

$$p_n(u_0) < K.$$

It is easy also to verify that $p_n(u)$ is analytic at u_0, so that, by Theorem IX.3,

$$p_{n,k} < K(u_0)\cdot u_0^{-k}, \qquad \sum_{j>k} p_{j,k} < \frac{K(u_0)}{u_0-1}u_0^{-k}.$$

Thus the combinatorial distributions satisfy uniformly (with respect to n) a large deviations bound. In particular the probability that there are more than a logarithmic number of components satisfies

$$\mathbb{P}_n(\chi > \log n) = O(n^{-\theta}), \qquad \theta = \log u_0. \tag{11}$$

Such tail estimates may additionally serve to evaluate the speed of convergence to the limit law (as well as the total variation distance) in the subcritical composition schema.

$\triangleright$ **6.** *Semi-small powers and singularity analysis.* Let $h(z)$ satisfy the stronger singular expansion

$$h(z) = \tau - c(1-z/\rho)^{\lambda} + O(1-z/\rho)^{\nu},$$

for $0<\lambda<\nu<1$. Then, for $k \le C\log n$ (some $C>0$), the results of singularity analysis can be extended (as stated and proved in Chapter VI, they are only valid for *fixed* k)

$$[z^n]h(z)^k = kc\rho^{-n}n^{-\lambda-1}\left(1+O(n^{-\theta_1})\right),$$

for some $\theta_1>0$, uniformly with respect to k. [The proof recycles all the ideas of Chapter VI and only needs some care in checking uniformity with respect to k of the major steps.] $\triangleleft$

$\triangleright$ **7.** *Speed of convergence in subscritical compositions.* Combining the exponential tail estimate (11) and local estimates deriving from the singularity analysis of "semi-small" powers in the previous notes, one obtains for the distribution functions associated with $p_{n,k}$ and p_k the speed estimate

$$\sup_k |F_n(k) - F(k)| \leq \frac{L}{n^{\theta_2}}.$$

There, L and θ_2 are two positive constants. $\triangleleft$

In the labelled universe, the functional composition schema encompasses the sequence, set, and cycle constructions. It suffices to take for the outer function g in the composition $g \circ h$ the quantities

$$Q(w) = \frac{1}{1-w}, \quad E(w) = e^w, \quad L(w) = \log \frac{1}{1-w}. \tag{12}$$

We state:

PROPOSITION IX.3 (Subcritical constructions). *Consider the constructions of* sequence $\mathfrak{S}(\mathcal{H})$, *whether labelled or not, labelled* set $\mathfrak{P}(\mathcal{H})$ *and labelled* cycle $\mathfrak{C}(\mathcal{H})$ *Assume the subcriticality conditions of the previous proposition, namely $\tau < 1$, $\tau < \infty$, $\tau < 1$, respectively, where τ is the singular value of $h(z)$. Then, the distribution of the number χ of components determined by $f_{n,k}/f_n$, is such that $\chi = 1 + Y$ admits a discrete limit law that is of type, respectively: negative binomial $NB[2]$, Poisson, and geometric. For $k \geq 1$, the limit form for $q_k = \lim_n \mathbb{P}(Y = k)$ are respectively*

$$q_k^{\mathfrak{S}} = (1-\tau)^2(k+1)\tau^k, \qquad q_k^{\mathfrak{P}} = e^{-\tau}\frac{\tau^k}{(k)!}, \qquad q_k^{\mathfrak{C}} = (1-\tau)\tau^k.$$

In an object of positive size, the number of components is always ≥ 1. In terms of the standard definition of the three laws (APPENDIX C: *Special distributions*, p. 688) the distribution of the number of components is $\chi = 1 + Y$ where Y is supported by $\mathbb{Z}_{\geq 0}$.

PROOF. In accordance with Proposition IX.2 and Equation (12), the PGF of the discrete limit law involves the derivatives

$$Q'(w) = \frac{1}{(1-w)^2}, \quad E'(w) = e^w, \quad L'(w) = \frac{1}{1-w}.$$

The last two cases precisely give rise to the classical Poisson and geometric law. The first case gives rise to the negative binomial law $NB[2]$ which also appears in this form as a sum of two geometricly distributed random variables. $\square$

The technical simplicity with which limit laws are pulled out of combinatorics is worthy of note.

EXAMPLE 3. *Root degrees in trees.* Consider first the number of components in a sequence (ordered forest) of general Catalan trees. The bivariate OGF is

$$F(z,u) = \frac{1}{1-uh(z)}, \quad h(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right).$$

We have $\tau_h = 1/2 < \rho_g = 1$, so that the composition schema is subcritical. Thus, for a forest of total size n, the number X_n of tree components satisfies

$$\lim_{n\to\infty} \mathbb{P}\{X_n = k\} = \frac{k}{2^{k+1}} \qquad (k \geq 1).$$

Since a tree is equivalent to a node appended to a forest, this asymptotic estimate also holds for the root degree of a general Catalan tree.

Consider next the number of components in a set (unordered forest) of Cayley trees. The bivariate EGF is

$$F(z,u) = e^{uh(z)}, \quad h(z) = ze^{h(z)}.$$

We have $\tau_h = 1 < \rho_g = +\infty$, again a subcritical composition schema. Thus the number X_n of tree components in a random unordered forest of size n admits the limit distribution

$$\lim_{n\to\infty} \mathbb{P}\{X_n = k\} = e^{-1}/(k-1)!, \qquad (k \geq 1),$$

a shifted Poisson law of parameter 1; asymptotically, the same property also holds for the root degree of a random Cayley tree

The same method applies more generally to a simple variety of trees $\mathcal{V}$ (see Chapter VII) with generator ϕ, under the condition of the existence of a root τ of the characteristic equation $\phi(\tau) - \tau\phi'(\tau) = 0$ at a point interior to the disc of convergence of ϕ. The BGF satisfies

$$V(z,u) = z\phi(uV(z)), \qquad V(z) = 1 - \gamma\sqrt{1 - /z\rho} + O(1 - z/\rho).$$

so that

$$V(z,u) \underset{z\to\rho}{\sim} \rho\phi(u\tau) - \gamma\frac{u\phi'(u\tau)}{\phi'(\tau)}\sqrt{1 - /z\rho}.$$

The PGF of the distribution of root degree is accordingly

$$\frac{u\phi'(\tau u)}{\phi'(\tau)} = \sum_{k\geq 1} \frac{k\phi_k\tau^k}{\phi'(\tau)}u^k.$$

(A limit law was established directly under its local form in Chapter VII.) END OF EXAMPLE 3. □

The root degree in a random labelled nonplane tree (Cayley tree) admits in the asymptotic limit a Poisson law, while the root degree of a large plane tree (a Catalan tree) tends to a negative binomial ($NB[2]$) distribution. Proposition IX.2 shows, in a precise technical sense, that the negative binomial law for Catalan trees is a direct reflection of planarity specified by a sequence construction, while the Poisson law arises from the set construction attached to nonplanarity.

▷ **8.** *Bell number distributions.* Consider the "set-of-sets" schema

$$F(z,u) = \exp(e^{uh(z)} - 1),$$

assuming subcriticality. This corresponds to a scheme $\mathcal{F} = \mathfrak{P}(\mathfrak{P}_{\geq 1}(\mathcal{H}))$. Then the number χ of components satisfies asymptotically a "derivative Bell" law:

$$\mathbb{P}(\chi = k) = \frac{1}{K}\frac{kS_k\tau^k}{k!}, \qquad K = e^{-e^\tau - \tau - 1},$$

where $S_n = n![z^n]e^{e^z - 1}$ is a Bell number. There exists parellel results: for sequence-of-sets, involving the surjection numbers; for set-of-sequences involving the fragmented permutation numbers. ◁

▷ **9.** *High levels in Cayley trees.* The number of nodes at level 5 (i.e., at distance 5 from the root) in a Cayley tree has the nice PGF

$$u\frac{d}{du}\left(e^{-1 + e^{-1 + e^{-1 + e^{-1 + e^{-1+u}}}}}\right),$$

and thus involves "super Bell" numbers. $\triangleleft$

A further direct application of continuity of PGFs is the distribution of the number of $\mathcal{H}$-components of a fixed size m in a composition $\Gamma[\mathcal{H}]$ with GF $g(h(z))$, again under the *subcriticality* condition. In the terminology of Chapter III, we are thus characterizing the *profile* of combinatorial objects, at least as regards components of some fixed size. The bivariate GF is then

$$F(z,u) = g(h(z) + (u-1)h_m z^m),$$

with $h_m = [z^m]h(z)$. The singular expansion at $z = \rho$ is

$$F(z,u) = g(\tau+(u-1)h_m\rho^m) - cg'(\tau+(u-1)h_m\rho^m)(1-z/\rho)^{\lambda}) + o((1-z/\rho)^{\lambda}).$$

Thus, the PGF $p_n(u)$ for objects of size n satisfies

$$\lim_{n\to\infty} p_n(u) = \frac{g'(\tau + (u-1)h_m\rho^m)}{g'(\tau)}. \tag{13}$$

Like before this specializes in the case of sequences, sets, and cycles giving a result analogous to Proposition IX.2.

PROPOSITION IX.4 (Fixed size components). *Under the subcriticality conditions of Propositions IX.2 and IX.3, the number of components of a fixed size m in a random sequence, set, or cycle construction applied to a class with GF $h(z)$ admits a discrete limit law. With $h_m := [z^m]h(z)$, ρ the radius of convergence of $h(z)$, and $\tau := h(\rho)$, the distributions are as follows:*

For sequences, the limit law is a negative binomial ($NB[2]$) of parameter $a = \dfrac{h_m\rho^m}{1-\tau+h_m\rho^m}$. For sets, the limit law is Poisson with parameter $\lambda = h_m\rho^m$. For cycles, the limit is geometric of parameter $a = \dfrac{h_m\rho^m}{1-\tau+h_m\rho^m}$.

EXAMPLE 4. *Root subtrees of size m.* In a Cayley tree, the number of root subtrees of some fixed size m has, in the limit, a Poisson distribution,

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad \lambda = \frac{m^{m-1}e^{-m}}{m!}.$$

In a general Catalan tree, the distribution is a negative binomial $NB[2]$

$$p_k = (1-a)^2(k+1)a^k, \qquad a^{-1} = 1 + \frac{m2^{2m-1}}{\binom{2m-2}{m-1}}.$$

Generally, for a simple variety of trees under the usual conditions of existence of a solution to the characteristic equation, $V = z\phi(V)$, one finds *"en deux coups de cuillère à pot"*,

$$\begin{array}{lcl} V(z,u) & = & z\phi(V(z) + V_m z^m(u-1)) \\ V(z,u) & \sim & \rho\phi(\tau + V_m\rho^m(u-1)) - \rho\gamma\phi'(\tau + V_m\rho^m(u-1))\sqrt{1-z/\rho} \\ \text{limit PGF} & = & \dfrac{\phi'(\tau + V_m\rho^m(u-1))}{\phi'(\tau)}. \end{array}$$

(Notations are the same as in Example 3.) .... END OF EXAMPLE 4. $\square$

Arbitrarily many schemas leading to discrete limit laws could be listed. Roughly, conditions are that the auxiliary variable u does not affect the location nor the nature of the dominant singularity of $F(z, u)$. Such conditions are met by the subcritical schemas, since eventually the auxiliary variable only appears as a multiplicative coefficient in a local singular expansion.

▷ **10.** *The product schema.* Define

$$F(z, u) = A(uz) \cdot B(z),$$

that corresponds to a product construction, $\mathcal{F} = \mathcal{A} \times \mathcal{B}$, with u marking the size of the $\mathcal{A}$–component in the product. Assume that the radii of convergence satisfy $\rho_A > \rho_B$ and that $B(z)$ has a unique dominant singularity of the algebraic-logarithmic type. Then, the size of the $\mathcal{A}$ component in arandom $\mathcal{F}$ structure has a discrete limit law with PGF,

$$p(u) = \frac{A(\rho u)}{A(\rho)}.$$

The proof results directly from singularity analysis. Alternatively, an elementary proof can be given based on the weaker requirement that the coefficient of B satisfy $b_{n+1}/b_n \to \rho^{-1}$. ◁

Regarding the number of components, the case of a supercritical composition leads to continuous limit laws of the Gaussian type, as we shall see in the next sections. The critical case may lead to a variety of probabilistic laws due to the confluence of singularities that then manisfests itself. In the example that follows, we show that a particular critical composition scheme already studied in Chapter VII leads to a collection of Poisson laws describing the small component profile of composite structures.

EXAMPLE 5. *Small components in sets of logarithmic structures.* Consider first the exp–log schema in the simpler labelled case: it is corresponds to the construction $\mathcal{F} = \mathfrak{P}(\mathcal{G})$, that is, $F(z, u) = \exp(uG(z))$ under the assumption that $G(z)$ is logarithmic. This means (Chapter VII) that $G(z)$ is Δ-singular and satisfies locally

$$G(z) = \kappa L(z/\rho) + \lambda + \eta(z), \qquad \text{where} \quad L(z) := \log(1 - z)^{-1},$$

and $\eta(z) = O(1/L(z/\rho)^2)$ as $z \to \rho$ in a Δ domain. We already know from Chapter VII that the number of components has mean and variance each of the order of $\log n$, so that a discrete limit law is not to be expected for the total number of components. However, the situation becomes quite different if *fixed* size components are considered. A limit distribution has already been obtained in Chapter VII under its local form and it may be revisited in the light of methods of the present chapter as follows. Let m be a fixed integer larger than 1. The BGF of $\mathcal{F}$ objects with u marking the number of m components is

$$F(z, u) = \exp\left((u - 1)g_r z^r\right).$$

Under the logarithmic assumption, one has for any u in a small neighbourhood of 1 as $z \to \rho$ in a Δ-domain:

$$F(z, u) \sim e^{\lambda} w(u)(1 - z/\rho)^{-\kappa}, \qquad w(u) = \exp\left((u - 1)g_r \rho^r\right).$$

By singularity analysis, this tells us that *the number of m–components in a random $\mathcal{F}$–structure of large size tends to a Poisson distribution with parameter $\lambda := g_r\rho^r$.*

This result applies for any m less than some arbitrary fixed bound B. In addition, truely multivariate methods discused at the end of this chapter enable one to prove that the the number of components of sizes $1, 2, \ldots, B$ are *asymptotically independent*. This gives a very precise model of the probabilistic profile of small components in random $\mathcal{F}$–objects as

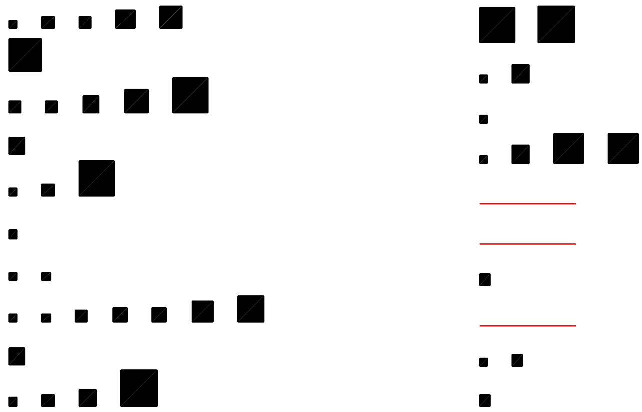


FIGURE 5. Small components of size ≤ 20 in random permutations (left) and random mappings (right) of size 1,000: each object corresponds to a line and each component is represented by a square of proportional area.

a product of independent Poisson laws of parameter $g_r \rho^r$ for $r = 1, \ldots, B$. Similar results hold for unlabelled multisets, but with the negative binomial law replacing the Poisson law. END OF EXAMPLE 5. □

The previous example covers well known exp-log structures introduced by Flajolet and Soria in [**177**]. In the labelled case, we have permutations (as sets of cycles), random mappings and 2–regular graphs (as sets of connected components). A rendering of the cycle structure of random permutations already appears in Chapter III; see also Figure 5. In the unlabelled case, the prime example is that of polynomials over finite fields to which we return later in this chapter.

In contrast, *large* component sizes cannot be independently distributed. (E.g., a permutation can have only cycle one larger than $n/2$, two cycles larger than $n/3$, etc.) A general probabilistic theory of the joint distribution of largest components in exp–log structures has been developed by Arratia, Barbour, and Tavaré [**16**] and some of its developments draw their inspiration from earlier studies conducted under the analytic combinatorial angle. This joint distribution of large components can be characterized in terms of what is known as the Poisson–Dirichlet process. For instance, as shown by Gourdon [**210**], the largest component itself involves the Dickmann function otherwise known to describe the distribution of the largest prime divisor of a random integer over a large interval of the form $[1\,..\,N]$.

▷ **11.** *Random mappings.* The number of components of some fixed size m in a large random mapping (functional graph) is asymptotically Poisson(λ) where $\lambda = K_m e^{-m}/m!$ and $K_m = m![z^m]\log(1-T)^{-1}$ enumerates connected mappings. (There T is the Cayley tree function.) The fact that $K_m e^{-m}/m! \approx 1/(2m)$ explains the fact that small compoents are somewhat sparser for mappings than for permutations. ◁

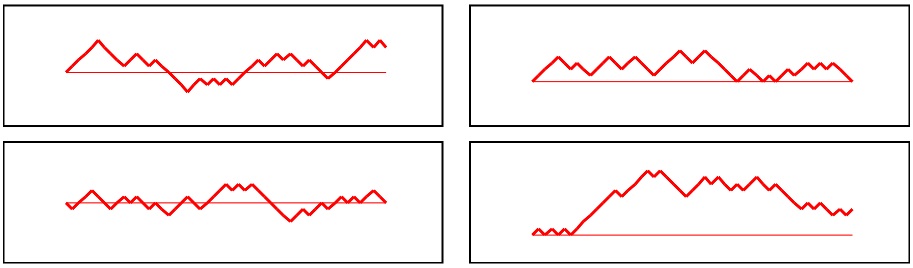


FIGURE 6. Walks, excursions, bridges, and meanders: random samples of length 50.

As a last example here, we discuss the length of the longest initial run of a's in random binary words satisfying various types of constraints. This discussion completes the informal presentation of Section IX. 1. The basic combinatorial objects are the set $\mathcal{W} = \{a, b\}^\star$ of binary words. A word $w \in \{\mathcal{W}\}$ can also be viewed as describing a walk in the plane, provided one interprets a and b as the vectors $(+1, +1)$ and $(+1, -1)$ respectively. Such walks in turn describe fluctuations in coin tossing games, as described by Feller [**133**]. What is especially interesting here is to observe the complete chain where a specific constraint leads in succession to a combinatorial decomposition, a specific analytic type of BGF, and a local singular structure that is then reflected by a particular limit law.

EXAMPLE 6. *Initial runs in random walks.* We consider here walks in the right half plane that start from the origin and are made of steps $a = (1, 1)$, $b = (1, -1)$. According to the discussion of Chapters V and VII, one can distinguish four major types of walks (Figure 6).

- *Unconstrained walks* ($\mathcal{W}$) corresponding to words and freely described by $\mathcal{W} = \mathfrak{S}(a, b)$;
- *Dyck paths* ($\mathcal{D}$) that always have a nonnegative ordinate and end at level 0; the closely related class $\mathcal{G} = \mathcal{D}b$ represents the collection of gambler's ruin sequences. In probability theory, Dyck paths are also refereed to as *excursions*.
- *Bridges* ($\mathcal{B}$) that are walks that may have negative ordinates but must finish at level 0.
- *Meanders* ($\mathcal{M}$) which have have always a nonnegative altitude and may end at an arbitrary nonnegative altitude.

The parameter χ of interest is in all cases the length of the longest initial run of a's.

First, the unconstrained walks obey the decomposition

$$\mathcal{W} = \mathfrak{S}(a)\mathfrak{S}(b\mathfrak{S}(a)),$$

already employed in Chapters I and IV. Thus, the BGF is

$$W(z, u) = \frac{1}{1 - zu} \frac{1}{1 - z(1 - z)^{-1}}.$$

By singularity analysis of the pole at $\rho = 1/2$, the PGF of χ on random words of $\mathcal{W}_n$ satisfies

$$p_n(u) \sim \frac{\frac{1}{2}}{1 - \frac{u}{2}},$$

and, as expected, this corresponds to a limit geometric law of parameter $\frac{1}{2}$. This is the first example presented (Section IX. 1) in order to introduce discrete limit laws.

As it is well-known, Dyck paths $\mathcal{D}$ play an important rôle in combinatorial constructions related to lattice paths (Chapters I and V). A sequence decomposes into "arches" that are themselves Dyck paths encapsulated by a pair a, b,

$$\mathcal{D} = \mathfrak{S}(a\mathcal{D}b),$$

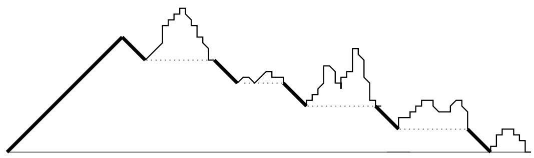
which yields a GF of the Catalan domain,

$$D(z) = \frac{1}{1 - z^2 D(z)}, \qquad D(z) = \frac{1 - \sqrt{1 - 4z^2}}{2z^2}.$$

In order to extract the initial run of a's, we observe that a word whose initial a-run is a^k contains k components of the form $b\mathcal{D}$. This corresponds to a decomposition in terms of the first traversals of altitudes $k - 1, \ldots, 1, 0$,

$$\mathcal{D} = \sum_{k \geq 0} a^k (b\mathcal{D})^k,$$

illustrated by the following diagram:



Thus, the BGF is

$$D(z, u) = \frac{1}{1 - z^2 u D(z)}.$$

This is an even function of z. In terms of the singular element, $\delta = (1 - 4z)^{1/2}$, one finds

$$F_1(z^{1/2}, u) = \frac{2}{2 - u} - \frac{2u}{(2 - u)^2}\delta + O(\delta^2),$$

as $z \to 1/4$. Thus, the PGF of χ on random words of $\mathcal{D}_{2n}$ satisfies

$$p_{2n}(u) \sim \frac{u}{(2 - u)^2},$$

which is the PGF of a negative binomial $NB[2]$ of parameter $\frac{1}{2}$ shifted by 1. (Naturally, in this case, explicit expressions for the combinatorial distribution are available, as this is equivalent to the classical ballot problem.)

A bridge decomposes into a sequence of arches, either positive or negative,

$$\mathcal{B} = \mathfrak{S}(a\mathcal{D}b + b\overline{\mathcal{D}}a),$$

where $\overline{\mathcal{D}}$ is like $\mathcal{D}$, but with the rôles of a and b interchanged. In terms of OGFs, this gives

$$B(z) = \frac{1}{1 - 2z^2 D(z)} = \frac{1}{\sqrt{1 - 4z^2}}.$$

The set $\mathcal{B}^+$ of nonempty walks that start with at least one a admits a decomposition similar to that $\mathcal{D}$,

$$B^+(z) = \left(\sum_{k \geq 1} a^k b (\mathcal{D}b)^{k-1} \right) \cdot \mathcal{B},$$

since the paths factor uniquely as a $\mathcal{D}$ component that hits 0 for the first time followed by a $\mathcal{B}$ oscillation. Thus,

$$B^+(z) = \frac{z^2}{1 - z^2 D(z)} B(z).$$

The remaining cases $\mathcal{B}^- = \mathcal{B} \setminus \mathcal{B}^+$ consist of either the empty word or of a sequence of positive or negative arches starting with a negative arch, so that

$$B^-(z) = 1 + \frac{z^2 D(z)}{1 - 2z^2 D(z)}.$$

The BGF results from these decompositions:

$$B(z,u) = \frac{z^2 u}{1 - z^2 u D(z)} B(z) + 1 + \frac{z^2 D(z)}{1 - 2z^2 D(z)}.$$

Again, the singular expansion is obtained mechanically,

$$B(z^{1/2}, u) = \frac{1}{2-u} \frac{1}{\delta} + O(1),$$

where $\delta = (1-4z)^{1/2}$. Thus, the PGF of χ on random words of $\mathcal{B}_{2n}$ satisfies

$$p_{2n}(u) \sim \frac{1}{2-u}.$$

The limit law is geometric of parameter $1/2$.

A meander decomposes into an initial run a^k, a succession of descents with their companion (positive) arches in some number $\ell \leq k$, and a succession of ascents with their corresponding (positive) arches. The computations are similar to the previous cases, more intricate, but still "automatic". One finds that

$$M(z,u) = \left(\frac{XY}{(1-X)(1-Y)} - \frac{XY^2}{(1-XY)(1-Y)} \right) \frac{1}{1-Y} + \frac{1}{1-X},$$

with $X = zu$, $Y = zW_1(z)$, so that

$$M(z,u) = 2\, \frac{1 - u - 2\,z + 2\,uz^2 + (u-1)\sqrt{1-4\,z^2}}{(1-zu)\left(1 - 2\,z - \sqrt{1-4\,z^2}\right)\left(2 - u + u\sqrt{1-4\,z^2}\right)}.$$

There are now two singularities at $z = \pm\frac{1}{2}$, with singular expansions,

$$M(z,u) \underset{z \to 1/2}{=} \frac{u\sqrt{2}}{(2-u)^2} \frac{1}{\sqrt{1-2z}} + O(1), \qquad M(z,u) \underset{z \to -1/2}{=} \frac{4-u}{4-u^2} + o(1),$$

so that only the singularity at $1/2$ matters asymptotically. Then, we have

$$p_n(u) \sim \frac{u}{(2-u)^2},$$

and the limit law is a shifted negative binomial $NB[2]$ of parameter $1/2$. In summary:

PROPOSITION IX.5. *The length of the initial run of a's in unconstrained walks and bridges is asymptotically distributed like a geometric; in Dyck excursions and meanders like a negative binomial $NB[2]$.*

Similar analyses can be applied to walks with a finite set of steps [**21**]. END OF EXAMPLE 6. □

▷ **12.** *The number of meanders.* A meander uniquely decomposes into an excursion followed by a (possibly empty) sequence of elements of the form $a\mathcal{D}$. There results that $M(z) = D(z)/(1-zD(z))$, and

$$M(z) = \frac{\sqrt{1-4z^2}-1+2z}{2z(1-2z)},$$

so that $M_n = \binom{n}{\lfloor n/2 \rfloor}$. ◁

▷ **13.** *Leftmost branch of a unary-binary (Motzkin) tree.* The class of unary-binary trees (or Motkzkin trees) is defined as the class of unlabelled rooted plane trees where (out)degrees of nodes are restricted to the set $\{0,1,2\}$. The parameter χ under consideration is the length of the lefmost branch measured by the number of nodes it contains. A tree can be viewed as a leftmost branch at each node of which is grafted either nothing (the node has degree 1) or a tree, except for the last node on the branch. Hence the decomposition and the BGF:

$$M(z) = \sum_{k \ge 1} z^k M(z)^{k-1}, \qquad M(z,u) = \frac{z}{1-zuM(z)}.$$

The first equation corresponds to $M = z(1+M+M^2)$ as it should. The dominant singularity is at $z = 1/3$ where $M(\frac{1}{3}) = 1$. There results that the limit PGF of χ is $4u/(3-u)^2$. The limit distribution is a negative binomial $NB[2]$ with parameter $\frac{1}{3}$, shifted by 1. ◁

IX. 4. Continuous limit laws

Throughout this chapter, our goal is to quantify sequences of random variables X_n that arise from an integer valued combinatorial parameter χ defined on a combinatorial class $\mathcal{F}$. It is a fact that, when the mean μ_n and the standard deviation σ_n of χ on $\mathcal{F}_n$ tend to infinity as n gets large, then a continuous limit law usually holds. That limit law arises not from the X_n themselves (as was the case for discrete-to-discrete convergence in the previous section) but from their standardized versions:

$$X_n^\star = \frac{X_n - \mu_n}{\sigma_n}.$$

In this section, we provide definitions and major theorems needed to deal with the discrete-to-continuous situation.

A random variable Y specified by its *distribution function*,

$$\mathbb{P}\{Y \le x\} = F(x),$$

is said to be *continuous* if $F(x)$ is continuous (see APPENDIX C: *Random variables*, p. 685). In that case, $F(x)$ has no jump, and there is no single value in the range of Y that bears a nonzero probability mass. If in addition $F(x)$ is differentiable, the random variable Y is said to have a *density*, $g(x) = F'(x)$, so that

$$\mathbb{P}(Y \le x) = \int_{-\infty}^{x} g(x)\,dx, \qquad \mathbb{P}\{x < Y \le x+dx\} = g(x)\,dx.$$

A particularly important case for us here is the standard *Gaussian* or *normal* distribution function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\,dw,$$

also called the *error function* (erf), the corresponding density being

$$\xi(x) \equiv \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

This section and the next ones are relative to the existence of limit laws of the continuous type, with Gaussian limits playing a prominent rôle. The general definitions of convergence in law (or in distribution) and of weak convergence (see APPENDIX C: *Convergence in law*, p. 690) instantiate as follows.

DEFINITION IX.3 (Discrete-to-continuous convergence). *Let Y be a* continuous *random variable with distribution function $F_Y(x)$. A sequence of random variables Y_n with distribution functions $F_{Y_n}(x)$ is said to converge in distribution to Y if,* pointwise, *for each x,*

$$\lim_{n\to\infty} F_{Y_n}(x) = F_Y(x).$$

In that case, one writes $Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y$ and $F_{Y_n} \stackrel{\mathcal{D}}{\Longrightarrow} F_Y$.

Convergence is said to take place at speed ϵ_n if

$$\sup_{x\in\mathbb{R}} |F_{Y_n}(x) - F_Y(x)| \le \epsilon_n.$$

The definition does not *a priori* require uniform convergence. It is a known fact that convergence to a continuous limit is always uniform. This uniformity means that there always exists a speed ϵ_n that tends to 0 as $n \to \infty$.

Discrete limit laws can be established via convergence of probability generating functions to a common limit, as asserted by the continuity theorem for PGFs, Theorem IX.1. In the case of continuous limit laws, one has to resort to integral transforms (see APPENDIX C: *Transforms of distributions*, p. 686), whose definitions we now recall.

— The *Laplace transform* —also called the *moment generating function*— $\lambda_Y(s)$ is defined by

$$\lambda_Y(s) := E\{e^{sY}\} = \int_{-\infty}^{+\infty} e^{sx}\, dF(x).$$

— the *Fourier transform* —also called the *characteristic function*— $\phi_Y(t)$ is defined by

$$\phi_Y(t) := E\{e^{itY}\} = \int_{-\infty}^{+\infty} e^{itx}\, dF(x).$$

(Integrals are taken in the sense of Lebesgue-Stieltjes or Riemann-Stieltjes; cf APPENDIX C: *Probability spaces and measure*, p. 683.)

There are two classical versions of the continuity theorem, one for characteristic functions, the other for Laplace transforms. Both may be viewed as extensions of the continuity theorem for PGF's. Characteristic functions always exist and the corresponding continuity theorem gives a necessary and sufficient condition for convergence of distributions. As they are a universal tool, characteristic functions are therefore often favoured in the probabilistic literature. In the context of this book, strong analyticity properties go along with combinatorial constructions so that both transforms usually exist and can be put to good use.

THEOREM IX.4 (Continuity of integral transforms). *Let Y, Y_n be random variables with Fourier transforms (characteristic functions) $\phi(t), \phi_n(t)$, and assume that Y has a continuous distribution function. A necessary and sufficient condition for the convergence in distribution, $Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y$, is that,* pointwise, *for each real t,*

$$\lim_{n\to\infty} \phi_n(t) = \phi(t).$$

Let Y, Y_n be random variables with Laplace transforms $\lambda(s), \lambda_n(s)$ that exist in a common interval $[-s_0, s_0]$. If, pointwise *for each real $s \in [-s_0, s_0]$,*

$$\lim_{n\to\infty} \lambda_n(s) = \lambda(s),$$

then the Y_n converge in distribution to Y: $Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y$.

The first part of this thorem is also known as *Lévy's continuity theorem* for characteristic functions.

PROOF. See Billingsley's book [**53**, Sec. 26], for Fourier transforms, and [**53**, p. 408], for Laplace transforms. □

▷ **14.** *Laplace transforms need not exists.* Let Y_n be a mixture of a Gaussian and a Cauchy distribution:

$$\mathbb{P}(Y_n \le x) = \left(1 - \frac{1}{n}\right) \int_{-\infty}^{x} \frac{e^{-w^2/2}}{\sqrt{2\pi}}\, dw + \frac{1}{\pi n} \int_{-\infty}^{x} \frac{dw}{1+w^2}.$$

Then Y_n convergences in distribution to a standard Gaussian limit Y, though $\lambda_n(s)$ only exists for $\Re(s) = 0$. ◁

The continuity theorem for PGFs eventually relies on continuity of the Cauchy coefficient formula that realizes the inversion needed in recovering coefficients from PGFs. Similarly, the continuity theorem for integral transforms may be viewed as expressing the continuity of inverse Laplace or Fourier transforms, this in the specific context of probability distribution functions.

The next theorem is an effective version of the Fourier inversion theorem that proves especially useful for characterizing speeds of convergence. It bounds in a constructive manner the sup-norm distance between two distribution functions by a special metric distance between their characteristic functions. Recall that $||f||_\infty := \sup_{x\in\mathbb{R}} |f(x)|$.

THEOREM IX.5 (Berry-Esseen inequality). *Let F, G be distribution functions with characteristic functions $\phi(t), \gamma(t)$. Assume that G has a bounded derivative. There exist absolute constants c_1, c_2 such that for any $T > 0$,*

$$||F - G||_\infty \le c_1 \int_{-T}^{+T} \left| \frac{\phi(t) - \gamma(t)}{t} \right| dt + c_2 \frac{||G'||_\infty}{T}.$$

PROOF. See Feller [**134**, p. 538] who gives

$$c_1 = \frac{1}{\pi}, \qquad c_2 = \frac{24}{\pi}$$

as possible values for the constants. □

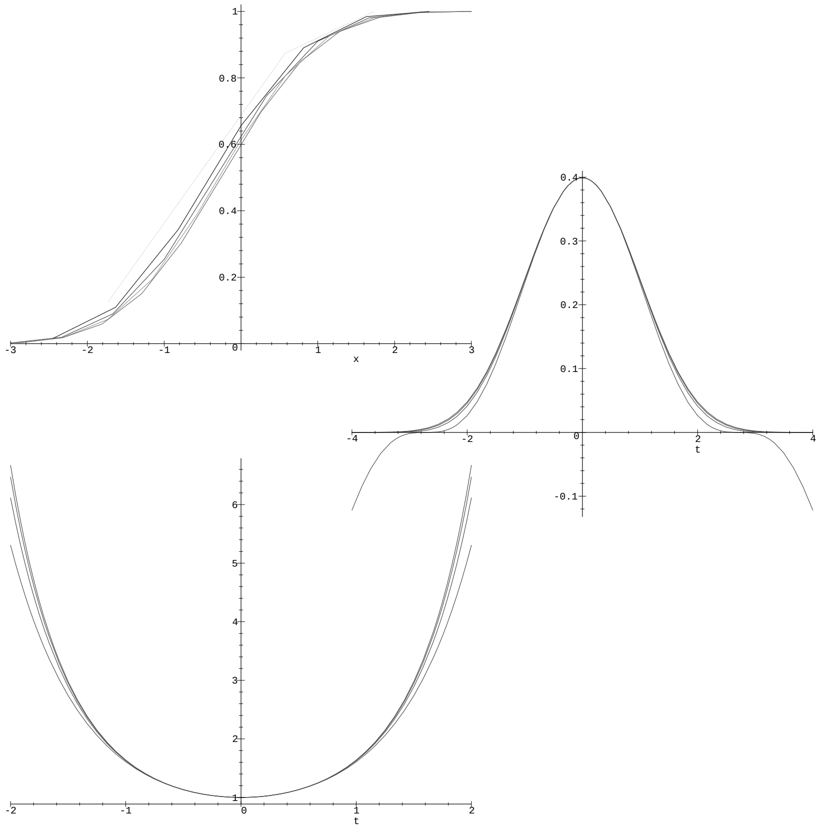


FIGURE 7. The standardized distribution functions of the binomial law (top), the corresponding Fourier transforms (middle), and the Laplace transforms (bottom), for $n = 3, 6, 9, 12, 15$. The distribution functions centred around the mean $\mu_n = n/2$ and scaled according to the standard deviation $\sigma_n = n^{1/2}/2$ converge to a limit which is the Gaussian error function, $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw$. Accordingly, the corresponding Fourier transforms —or characteristic functions— converge to $\phi(t) = e^{-t^2/2}$, while the Laplace transforms —or moment generating functions— converge to $\lambda(s) = e^{s^2/2}$.

This theorem is typically used with G being the limit distribution function (often a Gaussian for which $||G'||_\infty = (2\pi)^{-1/2}$) and $F = F_n$ a distribution that belongs to a sequence converging to G. The quantity T may be assigned an arbitrary value; the one giving the best bound in a specific application context is then normally chosen.

$\triangleright$ **15.** *A general version of Berry–Esseen.* Let F, G be two distributions functions. Define Lévy's "concentration function", $Q_G(h) := \sup_x (G(x+h) - G(x)), \qquad h > 0$. There exists an absolute constant C such that

$$||F-G||_\infty \le CQ_G(\frac{1}{T}) + C\int_{-T}^{+T}\left|\frac{\phi(t)-\gamma(t)}{t}\right| dt.$$

See Elliott's book [**125**, Lemma 1.47] and the article by Stef and Tenenbaum for a discussion [**395**]. The latter provides inequalities analogous to Berry-Esseen, but relative to Laplace transforms on the real line (bounds tend to be much weaker due to the smoothing nature of the Laplace transform). $\lhd$

Large powers and the central limit theorem. The binomial distribution is defined as the distribution of a random variable X_n with PGF

$$p_n(u) = \left(\frac{1}{2}+\frac{u}{2}\right)^n,$$

and characteristic function, $\phi_n(t) = p_n(e^{it})$. The mean is $\mu_n = n/2$ and the variance is $\sigma_n^2 = n/4$. Therefore, the standardized variable $X_n^* = (X_n - \mu_n)/\sigma_n$ has characteristic function

$$\phi_n^*(t) \equiv \mathbb{E}(e^{itX_n^\star}) = \left(\cosh(\frac{it}{\sqrt{n}}\right)^n. \tag{14}$$

The asymptotic form is directly found by taking logarithms, and one finds

$$\log\phi_n^*(t) = n\log\left(1-\frac{t^2}{2n}+\frac{t^4}{6n^2}+\cdots\right) = -\frac{t^2}{2}+O(\frac{1}{n}), \tag{15}$$

pointwise, for any fixed t, as $n\to\infty$. This establishes convergence to the Gaussian limit. In addition, the Berry-Esseen inequalities show that the speed of convergence is $O(n^{-1/2})$, a fact that is otherwise easily verified directly using Stirling's formula.

$\rhd$ **16.** *De Moivre's Central Limit Theorem.* Characteristic functions extend the normal limit law for unbiased Bernoulli distributions to the general case with PGF $(p+qu)^n$, for fixed p,q with $p+q=1$. (The result is accessible directly from Stirling's formula, which constitutes De Moivre's original derivation.) $\lhd$

The *central limit theorem* (CLT, then term was coined by Pólya in 1920, originally because of its "*zentralle Rolle*" in probability theory) of probability theory expresses the Gaussian character of sums of random variables. It was first discovered[4] in the particular case of Bernoulli variables by De Moivre. The general version is due to Gauss (who, around 1809, had realized from his works on geodesy and astronomy the universality of the "Gaussian" law but had only unsatisfactory arguments) and to Laplace (in the period 1812–1820). Laplace in particular uses Fourier methods and his formulation of the CLT is fully general, though some of the precise validity conditions of his arguments only became apparent a century later.

THEOREM IX.6 (Basic CLT). *Let T_j be independent random variables supported by $\mathbb{Z}_{\ge 0}$ with a common distribution of (finite) mean μ and (finite) standard deviation σ.*

[4]For a perspective on historical aspects of CLT, we refer to Hans Fischer's well-informed monograph [**138**].

Let $S_n := T_1 + \cdots + T_n$. Then the standardized sum $S_n^\star$ converges to the standard normal distribution,

$$S_n^\star \equiv \frac{S_n - \mu n}{\sigma\sqrt{n}} \stackrel{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,1).$$

PROOF. The proof is based on local expansions of characteristic functions. First, by a general theorem, the existence of the first two moments implies that

$$\phi_T(t) = 1 + i\mu t - \frac{1}{2}(\mu^2+\sigma^2)t^2 + o(t^2), \qquad t \to 0.$$

By shifting, it suffices to consider the case of zero-mean variables ($\mu = 0$). We then have, pointwise for each t as $n \to \infty$,

$$\phi_T\left(\frac{t}{\sigma\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{2n})\right)^n \to e^{-t^2/2},$$

like in Equations (14) and (15). The conclusion follows from the continuity theorem. (This theorem is in virtually any basic book on probability theory, e.g., [**134**, p. 259] or [**53**, Sec. 27].) □

The central limit theorem in the independent case is the subject of Petrov's comprehensive monograph [**346**]. There are many extensions of the CLT, to variables that are independent but not necessarily identically distributed (the Lindeberg–Lyapunov conditions) or variables that are only dependent in some weak sense (mixing conditions); see the discussion by Billingsley [**53**, Sec. 27]. In the particular case where the T's are discrete, a stronger "local" form of the Theorem results from the saddle point method; see our discussion in Chapter VIII, the classic treatment by Gnedenko and Kolmogorov [**201**], and extensions in Section IX. 9.

▷ **17.** *Poisson distributions of large parameter.* Let X_λ be Poisson with rate λ. As λ tends to infinity, Stirling's formula provides easily convergence to a Gaussian limit. The error terms can then be compared to what the Berry-Esseen bounds provide. (In terms of speed of convergence, such Poisson approximations to combinatorial distributions are sometimes of a better quality than the standard Gaussian law; see Hwang's comprehensive study [**239**] for a general analytic approach.) ◁

IX. 5. Quasi-powers and Gaussian limits

The central limit theorem of probability theory admits a fruitful extension in the context of analytic combinatorics. As we now show, it suffices that the PGF of a combinatorial parameter behaves nearly like a large power of a fixed function to ensure convergence to a Gaussian limit. We first illustrate this point by considering the Stirling cycle distribution.

EXAMPLE 7. *The Stirling cycle distribution.* Consider the Stirling cycle numbers $\left[{n \atop k}\right]$, and let X_n be the corresponding random variable with probability distribution $\left[{\frac{1}{n!} n \atop k}\right]$, with PGF,

$$p_n(u) = \binom{n+u-1}{n} = \frac{u(u+1)(u+2)\cdots(u+n-1)}{n!} = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)}.$$

We have for fixed u near 1,

$$p_n(u) = \frac{n^{u-1}}{\Gamma(u)} \left(1 + O(\frac{1}{n})\right) = \frac{1}{\Gamma(u)} \left(e^{(u-1)}\right)^{\log n} \left(1 + O(\frac{1}{n})\right). \tag{16}$$

As results from Stirling's formula for the Gamma function (or from singularity analysis of $[z^n](1-z)^{-u}$, Chapter VI), the error term in (16) is $O(n^{-1})$ when u stays in a small enough neighbourhhod of 1, for instance $|u-1| \le \frac{1}{2}$. Thus, as $n \to +\infty$, $p_n(u)$ is approximately a "large power" of e^{u-1} taken with exponent $\log n$, multiplied by a fixed function, $(\Gamma(u))^{-1}$. By analogy to the central limit theorem, we may expect a Gaussian law.

The mean satisfies $\mu_n = \log n + \gamma + o(1)$, the standard deviation satisfies $\sigma_n = \sqrt{\log n} + o(1)$. We thus consider the standardized random variable,

$$X_n^* = \frac{X_n - L - \gamma}{\sqrt{L}}, \qquad L = \log n,$$

whose characteristic function is

$$\phi_n^*(t) = \frac{e^{-it(L^{1/2} + \gamma L^{-1/2})}}{\Gamma(e^{it/\sqrt{L}})} \exp\left(L(e^{it/\sqrt{L}} - 1)\right) \left(1 + O(\frac{1}{n})\right).$$

For fixed t, with $L \to \infty$, the logarithm is then found mechanically to satisfy

$$\log \phi_n^*(t) = -\frac{t^2}{2} + O((\log n)^{-1/2}).$$

This is sufficient to establish a Gaussian limit law,

$$\lim_{n\to\infty} \mathbb{P}\left\{X_n \le \log n + \gamma + x\sqrt{\log n}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw. \tag{17}$$

PROPOSITION IX.6 (Goncharov's Theorem). *The Stirling cycle distribution,* $\mathbb{P}(X_n = k) = \frac{1}{n!}\left[{n \atop k}\right]$, *describing the number of cycles and the number of records in a random permutation of size* n *is asymptotically normal.*

This result was obtained by Goncharov as early as 1944, see [**204**], albeit without an error term as his investigations predate the Berry-Esseen inequalities. ... END OF EXAMPLE 7. $\square$

The cycle example is characteristic of the occurrence of Gaussian laws in analytic combinatorics. What happens is that the approximation (16) by a power with "large" exponent $\beta_n = \log n$ leads after normalization, to the characteristic function of a Gaussian variable, namely $e^{-t^2/2}$. From there, the limit distribution (17) results by the continuity theorem. This is in fact a very general phenomenon, as demonstrated by a theorem of Hsien-Kuei Hwang [**235, 238**] that we state next and that builds upon earlier statements of Bender and Richmond [**34**].

The following notations prove especially convenient: given a function $f(u)$ analytic at $u = 1$, we set

$$\mathfrak{m}(f) = \frac{f'(1)}{f(1)}, \qquad \mathfrak{v}(f) = \frac{f''(1)}{f(1)} + \frac{f'(1)}{f(1)} - \left(\frac{f'(1)}{f(1)}\right)^2. \tag{18}$$

The notations $\mathfrak{m}, \mathfrak{v}$ suggest their probabilistic counterparts while neatly distinguishing between the analytic and probabilistic realms: If f is the PGF of a random variable X, then $f(1) = 1$ and $\mathfrak{m}(f)$, the *mean*, coincides with the expectation $\mathbb{E}(X)$; the quantity $\mathfrak{v}(f)$ then coincides with the variance $\mathbb{V}(X)$.

THEOREM IX.7 (Quasi-Powers, Central law). *Let the X_n be nonnegative discrete random variables with probability generating function $p_n(u)$. Assume that,* uniformly *in a fixed complex neighbourhood of $u = 1$, for sequences $\beta_n, \kappa_n \to +\infty$, there holds*

$$p_n(u) = A(u)\,(B(u))^{\beta_n}\left(1 + O(\frac{1}{\kappa_n})\right), \tag{19}$$

where $A(u), B(u)$ are analytic at $u = 1$ and $A(1) = B(1) = 1$. Assume finally that $B(u)$ satisfies the so-called "variability condition",

$$\mathfrak{v}(B(u)) \equiv B''(1) + B'(1) - B'(1) \neq 0.$$

Under these conditions, the distribution of X_n is asymptotically Gaussian, and the speed of convergence to the Gaussian limit is $O(\kappa_n^{-1} + \beta_n^{-1/2})$:

$$\mathbb{P}\left\{\frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}} \leq x\right\} = \Phi(x) + O\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right).$$

The mean and variance of X_n satisfy

$$\begin{array}{lclcl} \mu_n & \equiv & \mathbb{E}(X_n) & = & \beta_n\,\mathfrak{m}(B(u)) + \mathfrak{m}(A(u)) + O(\frac{1}{\kappa_n}) \\ \sigma_n^2 & \equiv & \mathbb{V}(X_n) & = & \beta_n\,\mathfrak{v}(B(u)) + \mathfrak{v}(A(u)) + O(\frac{1}{\kappa_n}) \end{array} \tag{20}$$

This theorem is a direct application of the following lemma, also due to Hwang, that applies more generally to arbitrary discrete or continuous distributions, and is thus entirely phrased in terms of integral transforms.

LEMMA IX.1 (Quasi-Powers, general distributions). *Assume that the Laplace transforms $\lambda_n(s) = E\{e^{sX_n}\}$ of a sequence of random variables X_n are analytic in a disc $|s| < \rho$, for some $\rho > 0$, and satisfy there an expansion of the form*

$$\lambda_n(s) = e^{\beta_n U(s) + V(s)}\left(1 + O(\frac{1}{\kappa_n})\right), \tag{21}$$

with $\beta_n, \kappa_n \to +\infty$ as $n \to +\infty$, and $U(s), V(s)$ analytic in $|s| \leq \rho$. Assume also the variability condition,

$$U''(0) \neq 0.$$

Under these assumptions, the mean and variance of X_n satisfy

$$\begin{array}{lcl} E\{X_n\} & = & \beta_n U'(0) + V'(0) + O(\kappa_n^{-1}), \\ \mathbb{V}\{X_n\} & = & \beta_n U''(0) + V''(0) + O(\kappa_n^{-1}). \end{array} \tag{22}$$

The distribution of X_n is asymptotically Gaussian and the speed of convergence to the Gaussian limit is $O(\kappa_n^{-1} + \beta_n^{-1/2})$.

PROOF. This closely follows the lines of Hwang's works [**235, 238**]. First, we estimate the mean and variance. The variable s is *a priori* restricted to a small neighbourhood of 0. By assumption, the function $\log\lambda_n(s)$ is analytic at 0 and it satisfies

$$\log\lambda_n(s) = \beta_n U(s) + V(s) + O(\frac{1}{\kappa_n})$$

This asymptotic expansion carries over, with the same type of error term, to derivatives at 0 because of analyticity: this can be checked directly from Cauchy integral representations,

$$\frac{1}{k!}\,\frac{d^r}{ds^r}\log\lambda_n(s)\bigg|_{s=0} = \frac{1}{2i\pi}\int_\gamma \log\lambda_n(s)\,\frac{ds}{s^{r+1}},$$

upon using a small but fixed integration contour γ and taking advantage of the basic expansion of $\log\lambda_n(s)$. In particular, the mean and variance satisfy the estimates of (22).

Next, we consider the standardized variable,

$$X_n^\star = \frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}}, \qquad \lambda_n^\star(s) = E\{e^{sX_n^\star}\}.$$

We have

$$\log\lambda_n^\star(s) = -\frac{\beta_n U'(0)}{\sqrt{\beta_n U''(0)}}s + \log\lambda_n(\frac{s}{\sqrt{\beta_n U''(0)}}).$$

Local expansions to third order based on the assumption (21) show that

$$\log\lambda_n^\star(s) = \frac{s^2}{2} + O(\frac{|s|+|s|^3}{\beta_n^{1/2}}) + O(\frac{1}{\kappa_n}), \tag{23}$$

uniformly with respect to s in a disc of radius $O(\beta_n^{1/2})$, and in particular in any fixed neighbourhood of 0. This is enough to conclude as regards convergence in distribution to a Gaussian limit, by the continuity theorem of either Laplace transforms (restricting s to be real) or of Fourier transforms (taking $s = it$).

Finally, the speed of convergence results from the Berry-Esseen inequalities. Take $T \equiv T_n = c\beta_n^{1/2}$, where c is taken sufficiently small but nonzero, in such a way that the local expansion of $\lambda_n(s)$ at 0 applies. Then, the expansion (23) instantiated at $s = it$ entails that the quantity

$$\Delta_n := \int_{-T_n}^{T_n}\left|\frac{\lambda_n^\star(it) - e^{-t^2/2}}{t}\right| dt + \frac{1}{T_n}$$

satisfies

$$\Delta_n = O(\beta_n^{-1/2} + \kappa_n^{-1}),$$

and the statement follows by the Berry-Esseen theorem. □

Theorem IX.7 applies immediately to the Stirling cycle distribution for which the estimate (16) was derived. It shows in addition that the speed of convergence is $O((\log n)^{-1/2})$ for this distribution.

The Quasi-Powers Theorem under either form (19) or (21) can be read *formally* as expressing the distribution of a (pseudo)random variable

$$Z = Y_0 + W_1 + W_2 + \cdots + W_{\beta_n},$$

where Y_0 "corresponds" to $e^{V(s)}$ (or $A(u)$) and each W_j to $e^{U(s)}$ (or $B(u)$). However, there is no *a priori* requirement that β_n should be an integer, nor that $e^{U(s)}, e^{V(s)}$ be Laplace transforms of probability distribution functions. In a way, the theorem recycles the intuition that underlies the central limit theorem and makes use of the analytic machinery behind it. But, in applications, functions like $e^{U(s)}, e^{V(s)}$ do not necessarily admit a direct probabilistic interpretation.

It is of particular importance to note that the conditions of Theorem IX.7 and Lemma IX.1 are purely local: *what is required is* local *analyticity of the quasi-power approximation at* $u = 1$ *for PGF's or, equivalently,* $s = 0$ *for Laplace-Fourier transforms*. This important feature is ultimately due the normalization of random variables and transforms that goes along with continuous limit laws

▷ **18.** *Higher moments under Quasi-powers.* Following Hwang [**238**], one has under the conditions of the Quasi-Powers Theorem, Lemma IX.1, and for each fixed k,

$$\mathbb{E}(X_n^k) = k!\varpi_k(\beta_n) + O\left(\frac{\beta_n^k}{\kappa_n}\right), \qquad \varpi_k(s) := [s^k]e^{\beta_n U(s)+V(s)}.$$

(ϖ_k is a polynomial of degree k, which describes precisely the behaviour of higher moments.) ◁

Singularity perturbation and Gaussian laws. The main thread of this chapter is *bivariate generating functions*. In general, we are given a BGF $F(z,u)$ and aim at extracting a limit distribution from it. The quasi-power paradigm in the form (19) is what one should look for, in the case where the mean and the standard deviation both tend to infinity with the size n of the model.

We proceed heuristically in this informal discussion. Start from the BGF and consider u as a parameter. If singularity analysis applies to the counting generating function $F(z,1)$, it leads to an approximation,

$$f_n \approx C \cdot \rho^{-n} n^{\alpha},$$

where ρ is the dominant singularity of $F(z,1)$ and α is related to the critical exponent of $F(z,1)$ at ρ. A similar type of analysis is often applicable to $F(z,u)$ for u near 1. Then, it is reasonable to expect an approximation for the z-coefficients of the bivariate GF,

$$f_n(u) \approx C(u)\rho(u)^{-n} n^{\alpha(u)}.$$

In this perspective, the corresponding PGF is of the form

$$p_n(u) \approx \frac{C(u)}{C(1)} \left(\frac{\rho(u)}{\rho(1)}\right)^{-n} n^{\alpha(u)-\alpha(1)}.$$

The strategy envisioned here is thus a perturbation analysis of singular expansions with the auxiliary parameter u being restricted to a small neighbourhood of 1.

In particular if only the dominant singularity moves with u, we have a rough form

$$p_n(u) \approx \frac{C(u)}{C(1)} \left(\frac{\rho(u)}{\rho(1)}\right)^{-n},$$

suggesting a Gaussian law with mean and variance that are both $O(n)$. If only the exponent moves, then

$$p_n(u) \approx \frac{C(u)}{C(1)} n^{\alpha(u)-\alpha(1)},$$

suggests again a Gaussian law, but with mean and variance that are both $O(\log n)$.

These cases point to the fact that a rather simple perturbation of a univariate analysis may yield limiting Gaussian distributions. Each major coefficient extraction method of Chapters IV–VIII plays a rôle, and the present chapter illustrates this important point in the following contexts:

— *meromorphic analysis* for functions with polar singularities (Section IX. 6 below, based on a perturbation of methods of Chapters IV and V);
— *singularity analysis* for functions with algebraic–logarithmic singularity (Section IX. 7 below, based on a perturbation of methods of Chapters VI and VII);
— *saddle point analysis* for functions with fast growth at their singularity (Section IX. 8 below, based on a perturbation of methods of Chapters VIII).

Roughly, the decomposable character of many elementary combinatorial structures is reflected by strong analyticity properties of bivariate GF's that, after perturbation analysis, lead, via the Quasi-Powers Theorem (Theorem IX.7), to Gausssain laws. The coefficient extraction methods being based on contour integration supply the necessary uniformity conditions. (In contrast, Darboux's method or Tauberian theorems, being nonconstructive, are *not* normally applicable in this context.)

IX. 6. Perturbation of meromorphic asymptotics

This section discusses schemas that rely on the analysis of coefficients of meromorphic functions, as discussed in Chapters IV and V. It is largely based on works of Bender who, starting with his seminal article [**27**], was the first to propose abstract analytic schemas leading to Gaussian laws in analytic combinatorics. Our presentation also follows subsequent works of Bender, Flajolet, Hwang, Richmond, and Soria [**34, 177, 179, 235, 236, 237, 238, 388**].

EXAMPLE 8. *The surjection distribution.* We revisit the distribution of image cardinality in surjections for which the concentration property has been established in Chapter V. This example serves to introduce bivariate asymptotics in the meromorphic case. Consider the distribution of image cardinality in surjections, with BGF

$$F(z,u)=\frac{1}{1-u(e^z-1)}.$$

Restrict u near 1, for instance $|u-1|\le\frac{1}{10}$. The function $F(z,u)$, as a function of z, is meromorphic with singularities at

$$\rho(u)+2ik\pi,\qquad \rho(u)=\log(1+\frac{1}{u}).$$

The principal determination of the logarithm is used (with $\rho(u)$ near $\log 2$ when u is near 1). It is then seen that $\rho(u)$ stays within 0.06 from $\log 2$, for $|u-1|\le\frac{1}{10}$. Thus $\rho(u)$ is the unique dominant singularity of F, the next nearest one being $\rho(u)\pm 2i\pi$ with modulus certainly larger than 6.

From the coefficient analysis of meromorphic functions (Chapter IV), the quantities $f_n(u) = [z^n]F(z,u)$ are estimated as follows,

$$\begin{aligned} f_n(u) &= \operatorname{Res}\left(F(z,u)z^{-n-1}\right)_{z=\rho(u)} + \frac{1}{2i\pi}\int_{|z|=5} F(z,u)\,\frac{dz}{z^{n+1}} \\ &= \frac{1}{u\rho(u)e^{\rho(u)}}\,\rho(u)^{-n} + O(5^{-n}). \end{aligned} \tag{24}$$

It is important to note that the error term is *uniform* with respect to u, once u has been constrained to satisfy $|u-1| \le 0.1$. This fact derives from the coefficient extraction method, since, in the remainder Cauchy integral of (24), the denominator of $F(z,u)$ stays bounded away from 0.

The second estimate in Equation (24), constitutes a prototypical case of application of the quasi-power schema. Thus, the number X_n of image points in a random surjection of size n obeys in the limit a Gaussian law. The local expansion of $\rho(u)$,

$$\rho(u) \equiv \log(1+u^{-1}) = \log 2 - \frac{1}{2}(u-1) + \frac{3}{8}(u-1)^2 + \cdots,$$

yields

$$\frac{\rho(1)}{\rho(u)} = 1 + \frac{1}{2\log 2}\,(u-1) - \frac{3\ln(2)-2}{8(\log 2)^2}\,(u-1)^2 + O\left((u-1)^3\right),$$

so that the mean and standard deviation satisfy

$$\mu_n \sim C_1 n, \quad \sigma_n \sim \sqrt{C_2\, n}, \qquad C_1 := \frac{1}{2\log 2}, \quad C_2 := \frac{1-\log 2}{4(\log 2)^2}.$$

In particular, the variability condition is satisfied. Finally, one obtains, with Φ the Gaussian error function,

$$\mathbb{P}\{X_n \le C_1 n + x\sqrt{C_2 n}\} = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right).$$

This estimate can alternatively be viewed as a purely asymptotic statement regarding Stirling partition numbers.

PROPOSITION IX.7. *The surjection distribution defined as* $\frac{1}{S_n}\left\{ {n \atop k} \right\}$, *with* $S_n = \sum_k k!\left\{ {n \atop k} \right\}$ *the normalizing factor (the surjection number), satisfies* uniformly for all real x,

$$\frac{1}{S_n}\sum_{k \le C_1 n + x\sqrt{C_2 n}} k!\left\{ {n \atop k} \right\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-w^2/2}\,dw + O\left(\frac{1}{\sqrt{n}}\right).$$

This result already appears in Bender's foundational study [**27**], ... END OF EXAMPLE 8. □

The following analytic schema vastly generalizes the case of surjections. It is again strongly inspired by the works of Bender [**27**].

THEOREM IX.8 (Meromorphic schema). *Let $F(z,u)$ be a bivariate function that is bivariate analytic at $(z,u) = (0,0)$ and has nonnegative coefficients there. Assume that $F(z,1)$ is meromorphic in $z \le r$ with only a simple pole at $z = \rho$ for some positive $\rho < r$. Assume also the following conditions.*

(i) Meromorphic perturbation: *there exists $\epsilon > 0$ and $r > \rho$ such that in the domain, $\mathcal{D} = \{|z| \le r\} \times \{|u-1| < \epsilon\}$, the function $F(z,u)$ admits the representation*

$$F(z,u) = \frac{B(z,u)}{C(z,u)},$$

where $B(z,u), C(z,u)$ are analytic for $(z,u) \in \mathcal{D}$ with $B(\rho,1) \neq 0$. (Thus ρ is a simple zero of $C(z,1)$.)

(ii) Nondegeneracy: *one has $\partial_z C(\rho,1) \cdot \partial_u C(\rho,1) \neq 0$, ensuring the existence of a nonconstant $\rho(u)$ analytic at $u=1$, such that $C(\rho(u),u) =$ and $\rho(1) = \rho$.*

(iii) Variability: *one has*

$$\mathfrak{v}\left(\frac{\rho}{\rho(u)}\right) \neq 0.$$

Then, the random variable with probability generating function

$$p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

converges in distribution to a Gaussian variable with a speed of convergence that is $O(n^{-1/2})$. The mean and the standard deviation of X_n are asymptotically linear in n.

First we offer a few comments. Given the analytic solution $\rho(u)$ of the implicit equation $C(\rho(u),u) = 0$, the PGF $\mathbb{E}(u^{X_n})$ satisfies a quasi-powers approximation of the form $A(u)(\rho(1)/\rho(u))^n$, as we prove below. The mean μ_n and variance σ_n^2 are then of the form

$$\mu_n = \mathfrak{m}\left(\frac{\rho(1)}{\rho(u)}\right) n + O(1), \qquad \sigma_n^2 = \mathfrak{v}\left(\frac{\rho(1)}{\rho(u)}\right) n + O(1). \tag{25}$$

The variability condition of the Quasi-Powers Theorem is precisely ensured by condition (iii). Set

$$c_{i,j} := \left.\frac{\partial^{i+j}}{\partial z^i \partial u^j} C(z,u)\right|_{(\rho,1)}.$$

The numerical coefficients in (25) can themselves be solely expressed in terms of partial derivatives of $C(z,u)$ by series reversion,

(26)

$$\rho(u) = \rho - \frac{c_{0,1}}{c_{1,0}}(u-1) - \frac{c_{1,0}^2 c_{0,2} - 2c_{1,0}c_{1,1}c_{0,1} + c_{2,0}c_{0,1}^2}{2c_{1,0}^3}(u-1)^2 + O((u-1)^3).$$

In particular the fact that $\rho(u)$ is nonconstant, analytic, and a simple root corresponds to $c_{0,1}c_{1,0} \neq 0$ (by the analytic Implicit Function Theorem). The variance condition is then computed to be equivalent to the cubic inequality in the $c_{i,j}$:

$$\rho\, {c_{1,0}}^2 c_{0,2} - \rho\, c_{1,0}c_{1,1}c_{0,1} + \rho\, c_{2,0}{c_{0,1}}^2 + {c_{0,1}}^2 c_{1,0} + c_{0,1}{c_{1,0}}^2 \rho \neq 0. \tag{27}$$

PROOF. We can now proceed with asymptotic estimates. Consider a domain $|u-1| \leq \delta$ inside the region of analyticity of B, C. Then, one has

$$f_n(u) := [z^n]F(z,u) = \frac{1}{2i\pi} \oint F(z,u) \frac{dz}{z^{n+1}},$$

where the integral is taken along a small enough contour encircling the origin. We use the analysis of polar singularities described in Chapter IV, exactly like in (24). As $F(z,u)$ has at most one (simple) pole in $|z| \leq r$, we have

$$f_n(u) = \text{Res}\left(\frac{B(z,u)}{C(z,u)} z^{-n-1}\right)_{z=\rho(u)} + \frac{1}{2i\pi} \int_{|z|=r} F(z,u) \frac{dz}{z^{n+1}}, \tag{28}$$

where we may assume u suitably restricted by $|u-1|<\delta$ in such a way that $|r-\rho(u)|<\frac{1}{2}(r-\rho)$.

The modulus of the second term in (28) is bounded from above by

$$\frac{K}{r^n} \qquad \text{where} \qquad K=\frac{\sup_{|z|=r,|u-1|\le\delta}|B(z,u)|}{\inf_{|z|=r,|u-1|\le\delta}|C(z,u)|}. \tag{29}$$

Since the domain $|z|=r, |u-1|\le\delta$ is closed, $C(z,u)$ attains its minimum that must be nonzero, given the unicity of the zero of C. At the same time, $B(z,u)$ being analytic, its modulus is bounded from above. Thus, the constant K in (29) is finite.

A residue computation of the first term, in accordance with the analysis of meromorphic functions, then yields

$$f_n(u)=\frac{B(\rho(u),u)}{C'(\rho(u),u)}\rho(u)^{-n-1}+O(r^{-n}),$$

uniformly for u in a small enough fixed neighbourhood of 1. The mean and variance then satisfy (25), with the coefficient in the leading term of the variance term that is, by assumption, nonzero. Thus, the conditions of the Quasi-Powers Theorem in the form (19) are satisfied, and the law is Gaussian in the asymptotic limit. $\square$

Some form of condition regarding nondegeneracy is a necessity. For instance, the functions

$$\frac{1}{1-z},\quad \frac{1}{1-zu},\quad \frac{1}{1-zu^2},\quad \frac{1}{1-z^2u},$$

each fail to satisfy the nondegeneracy and the variability condition, and the variance of the corresponding discrete distribution is identically 0. The combinatorial variance is $O(1)$ for a related function like

$$F(z,u)=\frac{1}{1-z(u+2)+2z^2u}=\frac{1}{(1-2z)(1-zu)},$$

which is excluded by the variability condition of the theorem—there a discrete limit law, a geometric, is known to hold; see page 544. Yet another situation arises when considering

$$F(z,u)=\frac{1}{(1-z)(1-zu)}.$$

There is now a double pole at 1 when $u=1$ that arises from “confluence” at $u=1$ of two analytic branches $\rho_1(u)=1$ and $\rho_2(u)=1/u$. In this particular case, the limit law is continuous but non-Gaussian; in fact, this limit is the uniform distribution over the interval $[0,1]$, since

$$F(z,u)=1+z(1+u)+z^2(1+u+u^2)+z^3(1+u+u^2+u^3)+\cdots.$$

In addition, for this case, the mean is $O(n)$ but the variance is $O(n^2)$. Such situations are briefly examined in Section IX. 11 at the end of this Chapter.

$\triangleright$ **19.** *Higher order poles.* Under the conditions of Theorem IX.8, a limit Gaussian law holds for the distributions generated by the BGF $F(z,u)^m$, which has an mth order pole. See [**27**]. $\triangleleft$

EXAMPLE 9. *The Central Limit Theorem and discrete renewal theory.* Let $g(u)$ be any PGF ($g(1)=1$) of a random variable supported by $\mathbb{Z}_{\geq 0}$ that is analytic at 1 and nondegenerate (i.e., $\mathfrak{v}(g)>0$). Then

$$F(z,u)=\frac{1}{1-zg(u)}$$

has a singularity at $1/g(u)$ that is a simple pole,

$$\rho(u)=\frac{1}{g(u)}.$$

Theorem IX.8 then applies to give a weak form of the central limit theorem for discrete probability distributions with PGFs that are analytic at 1. (In such a case, a refined Gaussian convergence property—a local limit law, see Chapter VIII and Section IX. 9 below—also derives from the saddle point method.)

Under the same analytic assumptions on g, consider now the "dual" BGF,

$$G(z,u)=\frac{1}{1-ug(z)},$$

where the rôles of z and u have been interchanged. In addition, we must impose for consistency that $g(0)=0$. There is a simple probabilistic interpretation in terms of *renewal processes* of classical probability theory. Assume a light bulb has a lifetime of m days with probability $g_m=[z^m]g(z)$ and is replaced as soon as it ceases to function. Let X_n be the number of light bulbs consumed in n days assuming independence, conditioned upon the fact that a replacement takes place on the nth day. Then the PGF of X_n is $[z^n]G(z,u)/[z^n]G(z,1)$. (The normalizing quantity $[z^n]G(z,1)$ is precisely the probability that a renewal takes place on day n.) Theorem IX.8 applies. The function G has a simple dominant pole at $z=\rho(u)$ such that $g(\rho(u))=1/u$, with $\rho(1)=1$ since g is by asumption a PGF. One finds

$$\frac{1}{\rho(u)}=1+\frac{1}{g'(1)}(u-1)+\frac{1}{2}\frac{g''(1)+2g'(1)-2g'(1)^2}{g'(1)^3}(u-1)^2+\cdots.$$

Thus the limit distribution of X_n is normal with mean and variance satisfying

$$\mathbb{E}(X_n)\sim\frac{n}{\mu},\qquad \mathbb{V}(X_n)\sim n\frac{\sigma^2}{\mu^3},$$

where $\mu:=\mathfrak{m}(g)$ and $\sigma^2:=\mathfrak{v}(g)$ are the mean and variance attached to g. (This calculation checks the variability condition *en passant*.) The mean value result certainly conforms to probabilistic intuition. .... END OF EXAMPLE 9. $\square$

$\triangleright$ **20.** *Renewals every day.* In the renewal scenario, no longer condition on the fact that a bulb breaks down on day n. Let Y_n be the number of bulbs consumed so far. Then the BGF of Y_n is found by expressing that there is a sequence of renewals followed by a last renewal that is to be credited to all intermediate epochs:

$$\sum_{n\geq 1}\mathbb{E}(u_n^Y)z^n=\frac{1}{1-ug(z)}\frac{g(u)-g(zu)}{1-z}.$$

A Gaussian limit also holds for Y_n. $\triangleleft$

$\triangleright$ **21.** *A mixed CLT–renewal scenario.* Consider $G(z,u)=1/(1-g(z,u))$ where g has nonnegative coefficients, satisfies $g(1,1)=1$, and is analytic at $(z,u)=(1,1)$. This models the situation where bulbs are replaced but a random cost is incurred, depending on the duration of the bulb. Under general conditions, a limit law holds and it is Gaussian. This applies for instance to $H(z,u)=1/(1-a(z)b(u))$, where a and b are nondegenerate PGFs (a random repairman is called). $\triangleleft$

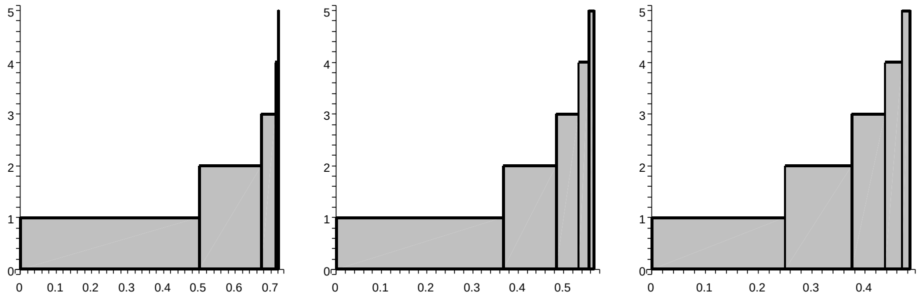


FIGURE 8. When components are sorted by size and represented by vertical segment of corresponding length, supercritical sequences present various profiles described by Proposition IX.8. The diagrams display the mean profiles of large surjections, alignments, and compositions for component sizes ≤ 5.

The preceding discussion of renewal processes also brings us extremely close analytically to a sequence schema $\mathcal{F} = \mathfrak{S}(\mathcal{G})$ and

$$F(z,u) = \frac{1}{1-ug(z)},$$

in the case where the schema is *critical*. It is then possible to refine the moment estinmates of Chapter V and obtain the probabilistic profile of supercritical sequences.

PROPOSITION IX.8 (Supercritical sequences). *Consider a sequence schema that is supercritical, i.e., the value of g at its dominant positive singularity satisfies $\tau_g > 1$. Assuming g to be aperiodic and $g(0) = 0$, the number X_n of $\mathcal{G}$–components in a random $\mathcal{F}_n$ structure of some large size n is asymptotically Gaussian with*

$$\mathbb{E}(X_n) \sim \frac{n}{g'(\sigma)}, \qquad \mathbb{V}(X_n) \sim n\frac{g''(\sigma)+g'(\sigma)-g'(\sigma)^2}{g'(\sigma)^3},$$

where σ is the radius of convergence of g. The number $X_n^{(m)}$ of components of some fixed size m is asymptotically normal with mean $\sim\ \theta_m n$, where $\theta_m = g_m\sigma^m/(\sigma g'(\sigma))$.

PROOF. The first part is a direct consequence of Theorem IX.8 and of the previous calculations with ρ replacing 1. The second part results from the BGF

$$\frac{1}{1-(u-1)g_m z^m - g(z)},$$

and from the fact that $u = 1$ induces a smooth perturbation of the pole at ρ corresponding to $u = 1$. □

This proposition aplies to alignments, surjections, compositions of various sorts—including compositions into prime summands. The profile of supercritical sequences is then appreciably different from what was obtained in the subcritical case, where discrete limit laws prevail. Fundamentally, the proportion of fixed size components is close to θ_m, up to Gaussian fluctuations. The diagrams of Chapter V and Figure 8 clearly illustrate this situation.

$\triangleright$ **22.** *Alignments and Stirling cycle numbers.* Alignments are sequences of cycles (Chapter II), corresponding to $\mathfrak{S}(\mathfrak{C}_{\geq 1}(\mathcal{Z}))$, with exponential BGF

$$F(z,u) = \frac{1}{1 - u\log(1-z)^{-1}}.$$

The function $\rho(u)$ is explicit, $\rho(u) = 1 - e^{-1/u}$, and the number of cycles in a random alignment is asymptotically Gaussian. This yields an asymptotic statement on Stirling cycle numbers: *Uniformly for all real x, with $O_n = \sum_k k!\left[{n \atop k}\right]$ the alignment number, there holds*

$$\frac{1}{O_n} \sum_{k \leq C_1 n + x\sqrt{C_2 n}} k! \begin{bmatrix} n \\ k \end{bmatrix} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw + O\left(\frac{1}{\sqrt{n}}\right),$$

where the two constants C_1, C_2 are $C_1 = \dfrac{1}{e-1}$, $C_2 = \dfrac{1}{(e-1)^2}$. $\triangleleft$

$\triangleright$ **23.** *Summands in constrained integer compositions.* Consider integer compositions where the summands are constrained to belong to a set $\Gamma \subseteq \mathbb{N}^+$, and let X_n be the number of summands in a random composition of integer n. The ordinary BGF is

$$F(z,u) = \frac{1}{1-ug(z)}, \qquad g(z) = \sum_{\gamma\in\Gamma} z^{\gamma}.$$

Assume that Γ contains at least two relatively prime elements, so that $g(z)$ is aperiodic. The radius of convergence of $g(z)$ can only be ∞ (when $g(z)$ is a polynomial) or 1 (when $g(z)$ comprises infinitely many terms but is dominated by $(1-z)^{-1}$). At any rate, the sequence construction is supercritical, so that the distribution of X_n is asymptotically normal. For instance, a Gaussian limit holds for compositions into prime or even twin-prime summands of Chapter V. $\triangleleft$

The next two examples are relative to runs in permutations and patterns in words. They do not resort to a supercritical sequence but their analytic structure is very much similar. It is of interest to note that the BGFs were each deduced in Chapter III by an inclusion-exclusion argument that involves sequences in an essential way.

EXAMPLE 10. *Ascending runs in permutations and Eulerian numbers.* The exponential BGF of Eulerian numbers (that count runs in permutations) is

$$F(z,u) = \frac{u(1-u)}{e^{(u-1)z} - u},$$

where, for $u = 1$, we have $F(z,1) = (1-z)^{-1}$. The roots of the denominator are then

$$\rho_k(u) := \rho(u) + \frac{2ik\pi}{u-1}, \qquad \rho(u) = \frac{\log u}{u-1},$$

where k is an arbitrary element of $\mathbb{Z}$. As u is close to 1, $\rho(u)$ is close to 1, while the other poles $\rho_k(u)$ with $k \neq 0$ escape to infinity. This fact is also consistent with the limit form $F(z,1) = (1-z)^{-1}$ which has only one pole at 1. If one restricts u to $|u| \leq 2$, there is clearly at most one root of the denominator in $|z| \leq 2$ that is given by $\rho(u)$. Thus, we have for u close enough to 1,

$$F(z,u) = \frac{1}{\rho(u) - z} + R(z,u),$$

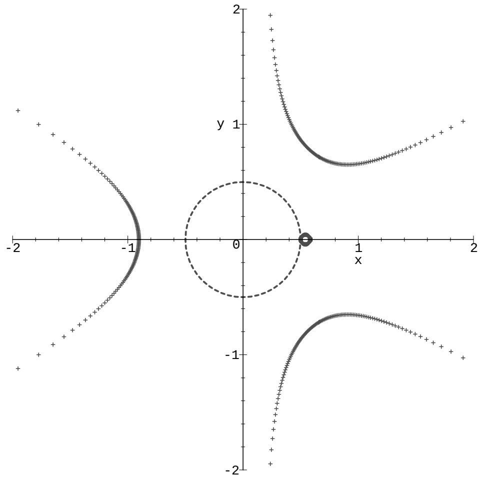


FIGURE 9. The diagrams of poles of the BGF $F(z,u)$ associated to the pattern $abaa$ with correlation polynomial $c(z) = 1 + z^3$ when u varies on the unit circle. The denominator is of degree 4 in z: one branch, $\rho(u)$ clusters near the dominant singularity $\rho = \frac{1}{2}$ of $F(z,1)$ while three other singularities stay away from the disc $|z| \le \frac{1}{2}$ and escape to infinity as $u \to 1$.

with $R(z,u)$ analytic in $|z| \le 2$, and

$$[z^n]F(z,u) = \rho(u)^{-n-1} + O(2^{-n}).$$

The variability conditions are satisfied since

$$\rho(u) = \frac{\log u}{(u-1)} = 1 - \frac{1}{2}(u-1) + \frac{1}{3}(u-1)^2 + \cdots,$$

so that $\mathfrak{v}(1/\rho(u)) = \frac{1}{12}$ is nonzero.

PROPOSITION IX.9. *The Eulerian distribution is asymptotically Gaussian, with mean and variance given by* $\mu_n = \frac{n+1}{2}$, $\sigma_n^2 = \frac{n+1}{12}$.

This example is a famous one and our derivation follows Bender's paper [**27**]. The Gaussian character of the distribution has been known for a long time; it is for instance to be found in David and Barton's *Combinatorial Chance* [**90**] published in 1962. There are in this case interesting connections with elementary probability theory: if U_j are independent random variables that are uniformly distributed over the interval $[0,1]$, then one has

$$[z^n u^k]F(z,u) = \mathbb{P}\{\lfloor U_1 + \cdots + U_n \rfloor < k\}.$$

Because of this fact, the normal limit is thus often derived a consequence of the central limit theorem of probability theory, after one takes care of unimportant details relative to the integer part $\lfloor\cdot\rfloor$ function; see [**90, 371**]. . . . . END OF EXAMPLE 10. $\square$

EXAMPLE 11. *Patterns in strings.* Consider the class $\mathcal{F}$ of binary strings (the "texts"), and fix a "pattern" w of length k. Let χ be the number of (possibly overlapping) occurrences of

w. (The pattern w occurs if it is a factor, *i.e.*, if its letters occur contiguously in the text.) Let $F(z,u)$ be the BGF relative to the pair $(\mathcal{F},\chi)$. The Guibas-Odlyzko correlation polynomial[5] $c(z) \equiv c_w(z)$ relative to w is defined for instance in [**382**], where it is shown that the OGF of words with pattern w excluded is

$$F(z,0) = \frac{c(z)}{z^k + (1-2z)c(z)}.$$

By similar string decompositions, the full BGF is found to be [**161**, p. 145]

$$F(z,u) = \frac{1-(c(z)-1)(u-1)}{1-2z-(u-1)(z^k+(1-2z)(c(z)-1))}.$$

Let $D(z,u)$ be the denominator. Then $D(z,u)$ depends analytically on z, for u near 1 and z near $1/2$. In addition, the partial derivative $D'_z(\frac{1}{2},1)$ is nonzero. Thus, $\rho(u)$ is analytic at $u=1$, with $\rho(1)=1/2$. The local expansion of the root $\rho(u)$ of $D(\rho(u),u)$ follows from local series reversion,

$$2\rho(u) = (1-2^{-k}(u-1) + (k2^{-2k} - 2^{-k}c(\frac{1}{2}))\,(u-1)^2 + O((u-1)^3).$$

Theorem IX.8 applies.

PROPOSITION IX.10. The number of occurrences of a fixed pattern in a random large string is asymptotically normal. *The number of occurrences has mean and variance σ_n^2 that satisfy*

$$\frac{n}{2^k} + O(1), \qquad \sigma_n^2 = \left(2^{-k}(1+2c(\frac{1}{2})) + 2^{-2k}(1-2k)\right) n + O(1).$$

The mean does not depend on the order of letters, only on the length of the pattern. END OF EXAMPLE 11. □

▷ **24.** *Patterns in Bernoulli texts.* Asymptotic normality also holds when letters in strings are chosen independently but with an arbitrary probability distribution. It suffices to use the weighted correlation polynomial described in a note of Chapter III. ◁

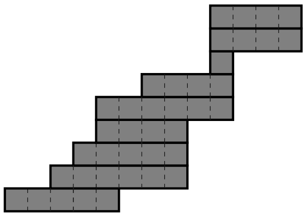
EXAMPLE 12. *Parallelogram polyominoes.* Polyominoes are plane diagrams that are closely related to models of statistical physics, while having been the subject of a vast combinatorial literature. This example has the merit of illustrating a level of difficulty somewhat higher than in previous examples and typical of many "real–life" applications. Our presentation follows an early article of [**29**] and a more recent paper of Louchard [**299**]. We consider here the variety of polyominoes called *parallelograms*. A parallelogram is a sequence of segments,

$$[a_1,b_1],\ [a_2,b_2],\ \ldots,[a_m,b_m], \qquad a_1 \le a_2 \cdots \le a_m,\ b_1 \le b_2 \le \cdots \le b_m,$$

where the a_j and b_j are integers with $b_j - a_j \ge 1$, and one takes $a_1 = 0$ for definiteness. A parallelogram can thus be viewed as a stack of segments (with $[a_{j+1},b_{j+1}]$ placed on top of

[5] The correlation polynomial, as defined in Chapter I, has coefficients in $\{0,1\}$, with $[z^j]c(z)=1$ iff w matches its left shifted image by j positions.

$[a_j, b_j]$) that leans smoothly to the right:



(This instance has area 39, width 13, height 9, and perimeter $13 + 9 = 22$.)

The quantity m is called the height, the quantity $b_m - a_1$ the width, their sum is called the (semi)perimeter, and the grand total $\sum_j (b_j - a_j)$ is called the area. We examine parallelograms of fixed area and investigate the distribution of the perimeter. The ordinary BGF of parallelograms, with z marking area and u marking perimeter turns out to be

$$F(z, u) = u \frac{J_1(z, u)}{J_0(z, u)}, \tag{30}$$

where J_0, J_1 belong to the realm of "q–analogues" and generalize the classical Bessel functions,

$$J_0(q, u) := \sum_{n \geq 0} \frac{(-1)^n u^n q^{n(n+1)/2}}{(q; q)_n (uq; q)_n}, \qquad J_1(q, u) := \sum_{n \geq 1} \frac{(-1)^{n-1} u^n q^{n(n+1)/2}}{(q; q)_{n-1} (uq; q)_n},$$

with the "q–factorial" notation being used:

$$(a; q)_n = (1 - a)(1 - aq) \cdots (1 - aq^{n-1}).$$

The expression (30) of the BGF results from a simple construction: a parallelogram is either an interval, or it is derived from an existing parallelogram by stacking on top a new interval. Let $G(w) \equiv G(x, y, z, w)$ be the OGF with x, y, z, w marking width, height, area, and length of top segment, respectively. The GF of a parallelogram made of a single nonzero interval is

$$a(w) \equiv a(x, y, z, w) = \frac{xyzw}{1 - xzw}.$$

The operation of piling up a new segment on top of a segment of length m that is represented by a term w^m is described by

$$y \left(\frac{z^m w^m}{1 - xzw} + \cdots + \frac{zw}{1 - xzw} \right) = xyzw \frac{1 - x^m w^m}{(1 - zw)(1 - xzw)}.$$

Thus, G satisfies the functional equation,

$$G(w) = \frac{xyzw}{1 - xzw} + \frac{xyzw}{(1 - zw)(1 - xzw)} \left[G(1) - G(xzw) \right]. \tag{31}$$

This is the method of "adding a slice" already employed in Chapter III. and reflected by the relation (31). Now, an equation of the form,

$$G(w) = a(w) + b(w)[G(1) - G(\lambda w)],$$

is solved by iteration:

$$\begin{aligned} G(w) &= a(w) + b(w)G(1) - b(w)G(\lambda(w)) \\ &= \big(a(w) - b(w)a(\lambda w) + b(w)b(\lambda w)a(\lambda^2 w) - \cdots\big) \\ &\quad + G(1)\big(b(w) - b(w)b(\lambda w) + b(w)b(\lambda w)b(\lambda^2 w) - \cdots\big). \end{aligned}$$

One then isolates $G(1)$ by setting $w = 1$. This expresses $G(1)$ as the quotient of two similar looking series (formed with sums of products of b–values). Here, this gives $G(x, y, z, 1)$, from which the form (30) of $F(z, u)$ derives, since $F(z, u) = G(u, u, z, 1)$.

In such a seemingly difficult situation, one should first estimate $[z^n]F(z,1)$, the number of parallelogram of "size" (*i.e.*, area) equal to n. We have $F(z,1) = J_1(z,1)/J_0(z,1)$, where the denominator is

$$J_0(z,1) = 1 - \frac{z}{(1-z)^2} + \frac{z^3}{(1-z)^2(1-z^2)^2} - \frac{z^6}{(1-z)^2(1-z^2)^2(1-z^3)^2} + \cdots .$$

Clearly, $J_0(z,1)$ and $J_1(z,1)$ are analytic in $|z| < 1$, and it is not hard to see that $J_0(z,1)$ decreases from 1 to about -0.24 when z varies between 0 and $\frac{1}{2}$, with a root at

$$\rho \doteq 0.43306\,19231\,29252,$$

where $J_0'(\rho,1) \doteq -3.76 \neq 0$, so that the zero is simple[6]. Since $F(z,1)$ is by construction meromorphic in the unit disc, and $J_1(\rho,1) \doteq 0.48 \neq 0$, the number of parallelograms satisfies

$$[z^n]F(z,1) \sim \frac{J_1(\rho,1)}{\rho J_0'(\rho,1)}\left(\frac{1}{\rho}\right)^n = \alpha_1 \cdot \alpha_2^n,$$

where

$$\alpha_1 \doteq 0.29745\,35058\,07786, \qquad \alpha_2 \doteq 2.30913\,85933\,31230.$$

As is common in meromorphic analyses, the approximation of coefficients is quite good; for instance, the relative error is only about 10^{-8} for $n = 35$.

We are now ready for bivariate asymptotics. Take $|z| \le r = \frac{7}{10}$ and $|u| \le \frac{11}{10}$. Because of the form of their general terms that involve $z^{n^2/2}u^n$ in the numerators while the denominators stay bounded away from 0, the functions $J_0(z,u)$ and $J_1(z,u)$ remain analytic there. Thus, $\rho(u)$ exists and is analytic for u in a sufficiently small neighbourhood of 1 (by Weierstrass preparation or implicit functions). The nondegeneracy conditions are easily verified by numerical computations. There results that Theorem IX.8 applies.

PROPOSITION IX.11. *The perimeter of a random parallelogram polyomino of area n admits a limit law that is Gaussian with mean and variance that satisfy: $\mu_n \sim \mu n$, $\sigma_n \sim \sigma\sqrt{n}$, with*

$$\mu \doteq 0.84176\,20156, \qquad \sigma \doteq 0.42420\,65326.$$

This indicates that a random parallelogram is most likely to resemble a slanted stack of fairly short segments. . . . . END OF EXAMPLE 12. □

▷ **25.** *Width and height of parallelogram polyominoes are normal.* Similar perturbation methods show that the expected height and width are each $O(n)$ on average, again with Gaussian limits. ◁

▷ **26.** *The base of a coin fountain.* A coin fountain (Chapter IV) is defined as a vector $v = (v_0, v_1, \ldots, v_\ell)$, such that $v_0 = 0$, $v_j \ge 0$ is an integer, $v_\ell = 0$ and $|v_{j+1} - v_j| = 1$. Take as size the *area*, $n = \sum v_j$. Then the distribution of the base length ℓ in a random coin fountain of size n is asymptotically normal. (This amounts to considering all ruin sequences of a fixed area as equally likely, and considering the number of steps in the game as a random variable.) Similarly the number of vector entries equal to 0 is asymptotically Gaussian. ◁

Perturbation of systems of linear equations. There is usually a fairly transparent approach to the analysis of BGFs defined implicitly as solutions of functional equations. One should start with the analysis at $u = 1$ and then examine the effect on singularities when u varies in a very small neighbourhood of 1. In accordance with

[6] As usual, such computations can be easily validated by carefully controlled numerical evaluations coupled with Rouché's theorem (see Chapter IV).

what we have already seen many times, the process is a perturbation analysis of the solution to a functional equation near a singularity, here one that *moves*.

We illustrate, mostly by way of examples, the application of Theorem to functions defined implicitly by a linear system of positive equations. Positive rational functions arise in connection with problems that can be equivalently described by finite state devices, by paths in graphs, and by Markov chains. The bivariate problem is then expressed by a linear equation

$$Y(z,u) = V(z,u) + T(z,u) \cdot Y(z,u), \tag{32}$$

where $T(z,u)$ is an $m \times m$ matrix with polynomial entries in z, u having nonnegative coefficients, $Y(z,u)$ is an $m \times 1$ column vector of unknowns, and $V(z,u)$ is a column vector of nonnegative initial conditions.

Regarding the univariate problem,

$$Y(z) = V(z) + T(z) \cdot Y(z). \tag{33}$$

, where $Y(z) = Y(z,1)$ and so on, we place ourselves under the assumptions of Theorem **??** of Chapter V. This means that properness, positivity, irreducibility, and aperiodicity are assumed throughout. In this case (see the developments of Chapter V), Perron-Frobenius theory applies to the univariate matrix $T(z)$. In other words, the function

$$C(z) = \det(I - T(z))$$

has a unique dominant root $\rho > 0$ that is a simple zero. Accordingly, any component $F(z) = Y_i(z)$ of a solution to the system (32) has a unique dominant singularity at $z = \rho$ that is a simple pole,

$$F(z) = \frac{B(z)}{C(z)},$$

with $B(\rho) \neq 0$.

In the bivariate case, each component of the solution to the system (32) can be put under the form

$$F(z,u) = \frac{B(z,u)}{C(z,u)}, \qquad C(z,u) = \det(I - T(z,u)).$$

Since $B(z,u)$ is a polynomial, it does not vanish for (z,u) in a sufficiently small neighbourhood of $(\rho,1)$. Similarly, by the analytic Implicit Function Theorem, there exists a function $\rho(u)$ locally analytic near $u = 1$, such that

$$C(\rho(u),u) = 0, \qquad \rho(1) = \rho.$$

Thus, it is sufficient that the variability conditions (26) be satisfied to infer a limit Gaussian distribution.

THEOREM IX.9 (Positive rational systems). *Let $F(z,u)$ be a bivariate function that is analytic at $(0,0)$ and has nonnegative coefficients. Assume that $F(z,u)$ coincides with the component Y_1 of a system of linear equations in $Y = (Y_1, \ldots, Y_m)^T$,*

$$Y = V + T \cdot Y,$$

where $V = (V_1(z,u),\dots,V_m(z,u))$, $T = (T_{i,j}(z,u))_{i,j=1}^m$, *and each of* $V_j, T_{i,j}$ *is a polynomial in* z, u *with nonnegative coefficients. Assume also that* $T(z,1)$ *is transitive, proper, and primitive, and let* $\rho(u)$ *be the unique solution of*

$$\det(I - T(\rho(u),u)) = 0,$$

assumed to be analytic at 1, such that $\rho(1) = \rho$*. Then, provided the variability condition,*

$$\mathfrak{v}\left(\frac{\rho(1)}{\rho(u)}\right) > 0,$$

is satisfied, a Gaussian Limit Law holds for the coefficients of $F(z,u)$ *with mean and variance that are* $O(n)$ *and speed of convergence that is* $O(n^{-1/2})$.

The constants μ, σ involved in estimates of the mean and standard deviation, $\mu_n \sim \mu n$, $\sigma_n \sim \sigma\sqrt{n}$, are then determined from $C(z,u) = \det(I - T(z,u))$ by Eq. (26). Thus, in any particular application, one can determine by computation whether the variability condition is satisfied. It may be however more difficult to check these conditions for a whole classes of problems.

EXAMPLE 13. *Limit theorem for Markov chains.* Assume that M is the transition matrix of an irreducible aperiodic Markov chain, and consider the parameter χ that records the number of passages through state 1 in a path of length n that starts in state 1. Then, the theorem applies with

$$V = (1,0,\dots,0)^T, \; T_{i,j}(z,u) = zM_{i,j} + z(u-1)M_{i,0}\delta_{j,0}.$$

We therefore derive a classical limit theorem for Markov chains:

PROPOSITION IX.12. *In an irreducible and aperiodic (finite) Markov chain, the number of times that a designated state is reached when* n *transitions are effected is asymptotically Gaussian.*

The conclusion also applies to paths in any strongly connected aperiodic digraph as well as to paths conditioned by their source and/or destination. .... END OF EXAMPLE 13. □

▷ **27.** *Sets of patterns in words.* This note extends Example 11 relative to the occurrence of a *single* pattern in a random text. Given the class $\mathcal{W} = \mathfrak{S}(\mathcal{A})$ of words over a finite alphabet $\mathcal{A}$, fix a finite set of "patterns" $S \subset \mathcal{W}$ and define the parameter $\chi(w)$ as the total number of occurrences of members of S in the word $w \in \mathcal{W}$. It is possible to build finite automaton (essentially a digital tree built on S equipped with return edges) that records simultaneously the number of partial occurrences of each pattern. Then, the limit law of χ is Gaussian; see Bender and Kochman's paper [**33**] and [**158, 161**] for an approach based on the de Bruin graph. ◁

Virtually all of the combinatorial classes that resort to transfer matrix methods exposed in Chapter V lead to Gaussian laws in the asymptotic limit.

EXAMPLE 14. *Tilings.* (See Bender [**35**].) Take an $(2 \times n)$ chessboard of 2 rows and n columns, and consider coverings with "monomer tiles" that are (1×1)-pieces, and "dimer tiles" that are either of the horizontal (1×2) or vertical (2×1) type. The parameter of interest is here the number of tiles. Consider next the collection of all "partial coverings" in which each column is covered exactly, except possibly for the last one. The partial coverings are of one of 4 types and the legal transitions are described by a compatibility graph. For instance, if the previous column started with one horizontal dimer and contained one monomer, the current column has one occupied cell, and one free cell that may then be occupied either by a monomer or a dimer. This finite state description corresponds to a set of linear equations over BGFs (with z marking the area covered and u marking the total number of tiles), with the transition matrix found to be

$$T(z,u) = z\begin{pmatrix} u & u^2 & u^2 & u^2 \\ 1 & 0 & 0 & 0 \\ u & 0 & 0 & 0 \\ u & 0 & 0 & 0 \end{pmatrix}.$$

In particular, we have

$$\det(I - T(z,u)) = 1 - zu - z^2(u^2 + u^3).$$

Then, Theorem IX.9 applies: the number of tiles is asymptotically normal. The method clearly extends to $(k \times n)$ chessboards, for any fixed k. .... END OF EXAMPLE 14. □

▷ **28.** *Succession-constrained integer compositions.* Consider integer compositions where consecutive summands add up to at least 4. The number of summands in such a composition of large size is asymptotically normal. [Hint: see Bender and Richmond [**35**]] ◁

▷ **29.** *Height in trees of bounded width.* Consider general Catalan trees of width less than a fixed bound w. (The width is the maximum number of nodes at any level in the tree.) In such trees, the distribution of height is asymptotically Gaussian. ◁

IX. 7. Perturbation of singularity analysis asymptotics

In this section, we examine schemes that arises when generating functions contain algebraic-logarithmic singularities. For instance, trees often lead to singularities that are of the square-root type and such a singular behaviour persists for a number of bivariate generating functions associated to aditively inherited parameters. In such cases, the underlying machinery is the method of singularity analysis detailed in Chapter VI, on which suitable perturbative developments are applied.

An especially important feature of the method of singularity analysis and of the associated Hankel contours is the fact that it preserves uniformity of expansions[7]. This feature is crucial in translating bivariate expansion, where we need to estimate *uniformly* a coefficient $f_n(u) = [z^n]F(z,u)$ that depends on the parameter u, given some (uniform) knowledge on the singular structure of $F(z,u)$ in terms of z. We state here an easy but crucial lemma that takes care of remainder terms in expansions and hence enables the use of singularity analysis in a perturbed context.

LEMMA IX.2 (Uniformity lemma, singularity analysis). *Let $f_u(z)$ be a family of functions analytic in a common Δ-domain Δ, with u a parameter taken in a bounded set U. Suppose that there holds*

$$|f_u(z)| < K(u)|1 - z|^{-\alpha(u)},$$

[7]For instance, Darboux's method only provides *non-constructive* error terms, as it is based on the Riemann-Lebesgue lemma; it cannot be employed for bivariate asymptotics. A similar comment applies to most Tauberian theorems.

where $K(u)$ is uniformly bounded, $K(u) < K$ for $u \in U$, and $\alpha(u)$ is such that $-\Re(\alpha(u) > B$ for some finite real B. Then, there exists a constant $\widetilde{K}$ (computable from Δ, K, B such that

$$|[z^n]f_u(z)| < \widetilde{K}n^{B-1}.$$

PROOF. It suffices to revisit the proof of the Big-Oh transfer (O-transfer) theorem of Chapter VI, paying due attention to uniformity. The proof proceeds by Cauchy's formula,

$$f_u, n \equiv [z^n]f_u(z) = \frac{1}{2i\pi}\int \gamma f_u(z)\,\frac{dz}{z^{n+1}},$$

where $\gamma = \cup_j \gamma_j$ is the contour used earlier. Accordingly, we let $f_{u,n}^{(j)}$ be the contribution in Cauchy's integral arising from part γ_j of the contour. Let r be the radius of the circular part of the contour, corresponding in earlier notations to γ_3. Without loss of generality, we may assume $|r-1| < 1$. Trivial bounds imply when $B > 0$ that that

$$|f_{u,n}^{(3)}| \le \frac{K}{(r-1)^{B+1}}r^{-n},$$

with an analogous formula if $B < 0$. The part γ_1 corresponding to the small circular arc at distance $1/n$ from 1 is similarly dealt with by trivial bounds to the effect that

$$|f_{u,n}^{(1)}| \le Kn^{B-1}.$$

The two conjugate rectlinear parts corresponding to γ_2, γ_4 each lead to

$$|f_{u,n}(2)| = |f_{u,n}^{(4)}| \le \frac{K}{2\pi}J_n n^{B-1}, \qquad J_n := \int_1^\infty t^{-B}\left(1 + \frac{1}{n}t\cos\theta\right)^n.$$

Combining the four majorizations yields the result. ☐ What this lemma expresses is more general than the meromorphic scheme; only the error terms in estimates of PGFs tend to be naturally less good as we replace an exponentially small error term inherent to meromorphic functions by a term that is usually $O(n^{-\beta})$ in the context of singularity analysis. (Note that the proof above also supplies the uniformity estimates needed in the proof of the little-oh transfer (o-transfer) of Chapter VI.)

▷ **30.** *Uniformity in the presence of lagarithmic multipliers.* Similar estimates hold when $f(z)$ is multiplied by a power of $L(z) = -\log(1-z)$. ◁

EXAMPLE 15. *Leaves in general Catalan trees.* As an introductory example, let us briefly revisit the analysis of the number of leaves in general Catalan trees, a problem already treated in Chapter III. where an explicit expression (a product of two binomial coefficients) has been derived. The computations are a little simpler if we adopt as BGF

$$G(z,u) = F(z,u^2) = \frac{1}{2}\left(1 + (u^2-1)z - \sqrt{1 - 2(u^2+1)z + (u^2-1)^2z^2}\right),$$

so that we consider a parameter equal to twice the number of leaves. In this case, the discriminant factors nicely:

$$1 - 2(u^2+1)z + (u^2-1)^2z^2 = (1 - z(1+u)^2)(1 - z(1-u)^2),$$

which leads to the expression

$$G(z,u) = A(z,u) + B(z,u)\sqrt{C(z,u)}, \tag{34}$$

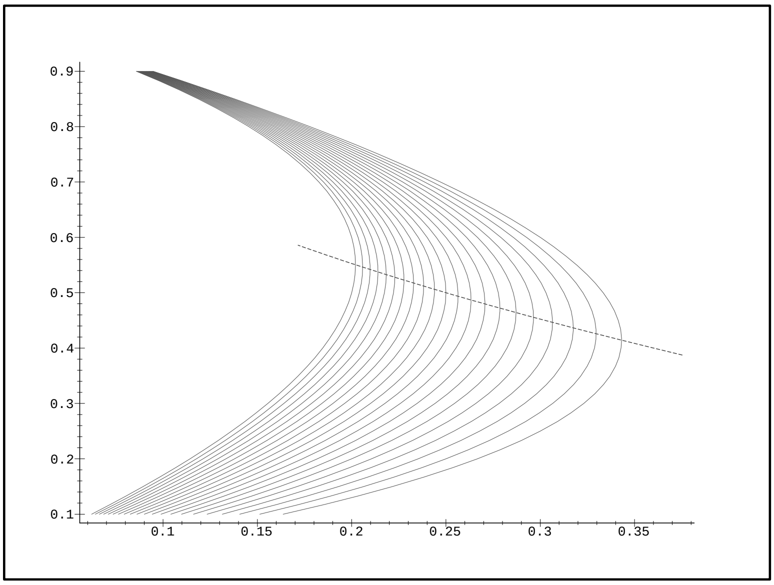


FIGURE 10. A display of the family of GF's $F(z, u_0)$ corresponding to leaves in general Catalan trees when $u_0 \in [\frac{1}{2}, \frac{3}{2}]$. It is seen that the singularities are all of the square root type (dashed line), with a movable singularity at $\widetilde{\rho}(u) = (1 + u^{1/2})^{-2}$.

with

$$A(z,u) = \frac{1}{2}(1 + (u^2 - 1)z), \quad B(z,u) = -\frac{1}{2}\sqrt{1 - z(1-u)^2},$$
$$C(z,u) = \frac{1}{2}\left(1 - z(1+u)^2\right).$$

This decomposition clearly shows that, when u is close enough to 1, the function $G(z,u)$ has a dominant singularity of the square-root type at

$$\rho(u) = \frac{1}{(1+u)^2}.$$

At the same time, if u is kept such that $|1-u| \le \frac{1}{2}$, then $B(z,u)$ remains analytic in both of its arguments for $|z| < 2$. For any such fixed u, we have for the BGF, by (34),

$$G(z,u) = a_0(u) + b(u)\sqrt{1 - z/\rho(u)} + a_1(u)(1 - z/\rho(u)) + O((1 - z/\rho(u))^{3/2}), \tag{35}$$

for some computable coefficients a_0, a_1, b, c that depend on u and are in fact analytic in u near $u = 1$. Singularity analysis then provides, pointwise for each u,

$$[z^n]G(z,u) = \frac{-2}{\sqrt{\pi}}\, B(\rho(u), u)\, \rho(u)^{-n} n^{-3/2}\left(1 + O(\frac{1}{n})\right). \tag{36}$$

The expansion (35) is uniform when u lies in a sufficiently small complex neighbourhood of 1. It can be seen (details below) that the expansion of the coefficient in (36) is also uniform by virtue of of the general uniformity preserving property of the singularity analysis process, as expressed by Lemma IX.2. We are thus exactly in a case of application of the Quasi-Powers Theorem, so that the limit law for the number of leaves is asymptotically Gaussian. END OF EXAMPLE 15. □

IX. 7.1. General algebraic-logarithmic conditions. The example of leaves in tres leads to simple computations, but is is characteristic of the machinery needed in more general cases. The theorem that follows is relative to any singular exponent α not in $\mathbb{Z}_{\leq 0}$.

THEOREM IX.10 (Algebraic singularity schema). *Let $F(z,u)$ be a bivariate function that is bivariate analytic at $(z,u) = (0,0)$ and has nonnegative coefficients there. Assume the following conditions:*

- (i) Algebraic perturbation: *there exist three functions A, B, C, analytic in a domain $\mathcal{D} = \{|z| \leq r\} \times \{|u-1| < \epsilon\}$, for some $r > 0$ and $\epsilon > 0$, such that the following representation holds,*

$$F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}, \tag{37}$$

  that $\rho < r$ is the unique (simple) root in $|z| \leq r$ of the equation $C(z,1) = 0$, and that $B(\rho,1) \neq 0$.
- (ii) Nondegeneracy*: one has $\partial_z C(\rho,1) \cdot \partial_u C(\rho,1) \neq 0$, ensuring the existence of a nonconstant $\rho(u)$ analytic at $u = 1$, such that $C(\rho(u),u) = 0$ and $\rho(1) = \rho$.*
- (iii) Variability: *one has*

$$\mathfrak{v}\left(\frac{\rho}{\rho(1)}\right) \neq 0.$$

Then, the random variable with probability generating function

$$p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

converges in distribution to a Gaussian variable with a speed of convergence that is $O(n^{-1/2})$. The mean μ_n and the standard deviation σ_n are asymptotically linear in n.

The remarks following the statement of Theorem IX.8 apply. Accordingly, the mean μ_n and variance σ_n^2 are computable by the general formula (25), and the variability condition is expressible in terms of the values of C and its derivatives at $(\rho,1)$ by means of Equation (27).

PROOF. Observe first that one does not need to worry about the *a priori* domain of existence of $F(z,u)$ since Equation (37) provides automatically analytic continuation to a collection of Δ–domains at $\rho(u)$ when u varies. Thus, it suffices that the representation (37) be established initially in some open domain of $\{|z| < \rho\} \times \{|u| < 1\}$, by unicity of analytic continuation.

By the assumptions made, the function $F(z,1)$ admits a singular expansion of the form

$$\begin{aligned} F(z,1) \quad &= \quad (a_0 + a_1(z-\rho) + \cdots) \\ &\qquad + (b_0 + b_1(z-\rho) + \cdots)\left(c_1(z-\rho) + c_2(z-\rho)^2 + \cdots\right)^{-\alpha}. \end{aligned} \tag{38}$$

There, the a_j, b_j, c_j represent the coefficients of the expansion in z of A, B, C for z near ρ when u is instantiated at 1. (We may consider $C(z,u)$ normalized by the condition that c_1 is positive real, and take, e.g., $c_1 = 1$.) Singularity analysis then implies the estimate

$$[z^n]F(z,1) = b_0(-c_1\rho)^{-\alpha}\rho^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+O(\frac{1}{n})\right). \tag{39}$$

All that is needed now is a "lifting" of relations (38) and (39), for u in a small neighbourhood of 1. First, we observe that by the analyticity assumption on A, the coefficient $[z^n]A(z,u)$ is exponentially small compared to ρ^{-n}, for u close enough to 1. Thus, for our purposes, we may freely restrict attention to $B(z,u)C(z,u)^{-\alpha}$. (The function A is only needed in some cases so as to ensure nonnegativity of the first few coefficients of F.) Next, it is convenient to operate with a *fixed* rather than movable singularity. This is simply achieved by considering the normalized function

$$\Phi(z,u) := B\left(\frac{z}{\rho(u)},u\right)C\left(\frac{z}{\rho(u)},u\right)^{-\alpha}.$$

Provided u is restricted to a suitably small neighbourhood of 1 and z to $|z| < R$ for some $R >'$, the functions $B(z/\rho(u),u)$ and $C(z/\rho(u),u)$ are analytic in both z and u, with $C(z,u)$ having a fixed simple zero at $z=1$. There results that the function

$$\frac{1}{1-z}C\left(\frac{z}{\rho(u)},u\right)$$

has a removable singularity at $z=1$ and is in fact analytic in $|z| < r$, $|u-1| < \delta$. Thus, Φ satisfies an expansion of the form

$$\Phi(z,u) = (1-z)^{-\alpha}\sum_{n\geq 0}\phi_n(u)(1-z)^n,$$

that is convergent and such that each coefficient $\phi_j(u)$ is an analytic function of u for $|u-1| < \delta$.

We may restrict this neighbourhood as we please, with $|u-1| \leq \delta$ provided we keep $\epsilon \geq \delta > 0$. First, by Weierstrass preparation, there is for u sufficiently near to 1, a unique simple root $\rho(u)$ near ρ of the equation

$$C(\rho(u),u) = 0.$$

We have $\rho(1) = \rho$ with $\rho(u)$ being locally analytic at 1. One can then expand A, B, C near $(\rho(u),u)$. This gives the bivariate expansion

$$\begin{aligned} F(z,u) &= (a_0(u) + a_1(z-\rho(u)) + \cdots) \\ &+ (b_0(u) + b_1(u)(z-\rho(u)) + \cdots)\left(c_1(u)(z-\rho(u)) + c_2(u)(z-\rho(u))^2 + \cdots\right)^{-\alpha}. \end{aligned} \tag{40}$$

There, by assumption, we have that $a_j(u), b_j(u), c_j(u)$ are analytic in $|u-1| \leq \epsilon$, and are each $O(r^{-n})$. In addition, $\rho(u)^\alpha$ and $(-c_1(u))^\alpha$ are well-defined by principal values, since their specializations at $u=1$ are positive. Thus, we have a singular expansion for $F(z,u)$; for instance, when $\alpha \in]-1,0[$,

$$\begin{aligned}F(z,u) &= a_0(u) + a_1(u)(z-\rho(u)) \\ &\quad + b_0(u)(-c_1(u)\rho(u))^{-\alpha}(1-z/\rho(u))^{-\alpha} + R(z),\end{aligned} \tag{41}$$

where

$$R(z) = O\left((1-z/\rho(u))^{\alpha+1}\right),$$

and the O–error term is uniform for $|u-1|<\delta$:

$$|R(z)| \le K \cdot |1-z/\rho(u)|,$$

for some absolute constant K. We thus have

$$[z^n]F(z,u) = b_0(u)(-c_1(u)\rho(u))^{-\alpha}\rho(u)^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+O(\frac{1}{n})\right). \tag{42}$$

where the error term is again uniform. An especially important fact for this argument is the following: *the singularity analysis process is a uniform coefficient extraction method.* This is precisely provided by Lemma IX.2.

Equation (42) shows that $f_n(u) = [z^n]F(z,u)$ satisfies precisely the conditions of the Quasi-Powers Theorem. Therefore, the law with PGF $f_n(u)/f_n(1)$ is asymptotically normal with a mean and a standard deviation that are both $O(n)$. Since the error term in (42) is $O(1/n)$, the speed of convergence to the Gaussian limit is $O(1/\sqrt{n})$. $\square$

$\triangleright$ **31.** *Logarithmic multipliers.* The conclusions of Theorem IX.10 extend to functions representable under the more general form

$$F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}\left(\log C(z,u)\right)^k.$$

(The proof follows exactly the same pattern.) $\triangleleft$

EXAMPLE 16. *Leaves in classical varieties of trees.* We start with binary Catalan trees and with the BGF

$$F(z,u) = z(u + 2zF(z,u) + F(z,u)^2),$$

so that

$$F(z,u^2) = \frac{1}{2z}\left(1-2z-\sqrt{(1-2z(1+u))(1-2z(1-u))}\right).$$

This is almost the same as the BGF of leaves in general Catalan trees. The dominant singularity is at $\rho(u) = \frac{1}{2(1+u)}$, and the limit law is Gaussian. The asymptotic form of the mean and variance are immediately derived from ρ, and we find that the number of leaves X_n in a binary Catalan tree satisfies

$$E\{X_n\} = \frac{n}{4} + O(1), \qquad \sigma\{X_n\} = \frac{\sqrt{n}}{4} + O(n^{-1/2}).$$

In the case of Cayley trees, the BGF equation[8] is

$$F(z,u) = z(u-1+e^{F(z,u)}).$$

By Lagrange inversion, the distribution is related to the Stirling partition numbers. The functional equation admits an explicit solution in terms of Lambert's "W-function", which is such

[8]This example constitutes a typical application of symbolic manipulation systems.

that $z = We^W$, with the branch choice that $W = 0$ when $z = 0$. Thus, $W(z) = -T(-z)$, where $T = ze^T$ is the classical "Cayley tree function". Here, we have

$$F(z,u) = z(u-1) - W(-ze^{z(u-1)}).$$

The function W has a dominant singularity of the square-root type at $-e^{-1}$. Thus, one can solve for $\rho(u)$, again in terms of the W function. Here, we find

$$\rho(u) = \frac{1}{u-1} W(e^{-1}(u-1)).$$

In particular, we get $\rho(1) = e^{-1}$, as we should. The expansion near $u = 1$ then comes automatically

$$\frac{\rho(u)}{\rho(1)} = 1 - e^{-1}(u-1) + \frac{3}{2}e^{-2}(u-1)^2 + O((u-1)^3).$$

Hence the mean and the variance of the number X_n of leaves in a random tree of size n satisfy:

$$E\{X_n\} \sim e^{-1}\, n \approx 0.36787\, n, \qquad \sigma^2\{X_n\} \sim e^{-2}(e-2)\, n \approx 0.09720\, n,$$

and the limit law is a Gaussian. . . . . END OF EXAMPLE 16. $\square$

$\triangleright$ **32.** *Leaves in Motzkin trees.* The number of leaves in a unary-binary (Motzkin) tree is asymptotically Gaussian. $\triangleleft$

EXAMPLE 17. *Patterns in binary Catalan trees.* We develop here a more sophisticated example coming from the analysis of pattern matching in trees [**399, 176**] that generalizes the problem of leaves. Fix a nonempty binary tree w and let $\omega[t] \equiv \omega_w[t]$ be the number of occurrences of pattern w in tree t. By this, we mean the number of internal nodes ν in t such that the subtree of t rooted at ν is isomorphic to w. The problem is of interest in the analysis of some symbolic manipulation algorithms and of "sharing" strategies; see [**399, 176**] for the algorithmic context.

A pattern occurs either in the left root subtree t_0 or in the right root subtree t_1 or at the root iself if t coincides with w. This gives rise to the recursive definition

$$\omega[t] = \omega[t_0] + \omega[t_1] + [\![t = w]\!], \qquad \omega[\emptyset] = 0,$$

where $[\![P]\!]$ denotes the indicator function of P whose value is 1 if P is true, and 0 otherwise. The function $u^{\omega[t]}$ is almost multiplicative, and

$$u^{\omega[t]} = u^{[\![t=w]\!]} u^{\omega[t_0]} u^{\omega[t_1]} = u^{\omega[t_0]} u^{\omega[t_1]} + [\![t = w]\!] \cdot (u-1).$$

Thus, the bivariate generating function $F(z,u)$ where z marks internal nodes and u marks the number of occurrences of w,

$$F(z,u) := \sum_t z^{|t|} u^{\omega[t]},$$

satisfies the algebraic equation,

$$F(z,u) = 1 + (u-1)z^m + zF(z,u)^2,$$

with $m = |w|$ the number of internal nodes of w.

The quadratic equation for F leads to

$$F(z,u) = \frac{1}{2z}\left(1 - \sqrt{1 - 4z - 4z^{m+1}(u-1)}\right).$$

The discriminant has a unique root $\rho = 1/4$ when $u = 1$, while it has $m+1$ roots for $u \neq 1$. By general properties of implicit and algebraic functions (implicit function theorem, Weierstrass preparation), as u tends to 1, one of these roots, call it $\rho(u)$ tends to $1/4$ while all the other ones $\{\rho_j(u)\}_{j=1}^m$ escape to infinity. We have

$$H(z,u) := \frac{1-4z-4z^{m+1}(u-1)}{1-z/\rho(u)} = \prod_{j=1}^{m}(1-z/\rho_j(u)),$$

which is an analytic function in (z,u) for (z,u) in a complex neighbourhood of $(1/4,1)$. This results from the fact that the algebraic function $1/\rho(u)$ is analytic at $u=1$. It gives the singular expansion of $G(z,u)=zF(z,u)$:

$$G(z,u) = \frac{1}{2} - \frac{1}{2}\sqrt{H(z,u)}\sqrt{1-z/\rho(u)}.$$

Thus, we are exactly under the conditions of the theorem. The quantity ω taken over a random binary tree of size $n+1$ has mean and variance given asymptotically by

$$\mathfrak{m}\left(\frac{1}{4\rho(u)}\right)n, \qquad \mathfrak{v}\left(\frac{1}{4\rho(u)}\right)n.$$

The expansion of $\rho(u)$ at 1 is computed easily by iteration of the defining equation:

$$z = \frac{1}{4} - z^{m+1}(u-1) = \frac{1}{4} - (\frac{1}{4} - z^{m+1}(u-1))^{m+1}(u-1) + \cdots.$$

Thus,

$$\rho(u) = \frac{1}{4} - \frac{1}{4^{m+1}}(u-1) + \frac{m+1}{4^{2m+1}}(u-1)^2 + \cdots.$$

This shows that the mean μ_n and the variance σ_n^2 of the number of occurrences of a pattern of size m in a random binary tree of size n satisfy

$$\mu_n \sim \frac{n}{4^m}, \qquad \sigma_n^2 \sim n(\frac{1}{4^m} - \frac{2m+1}{4^{2m}});$$

also, the distribution is asymptotically Gaussian. In particular, the probability of occurrence of a pattern at a random node of a random trees decreases fast (the factor of 4^{-m} in the estimate of averages) with the size of the pattern, a property that was to be expected and that also holds for strings. The paper of Steyaert and Flajolet [**399**] shows that similar properties (equivalent to the mean value analysis) hold for any simply generated family. The expression of the BGF $F(z,u)$ is given by Flajolet, Sipala, and Steyaert in [**176**], where similar developments are used to show that the minimal "dag representation" of a random tree —identical subtrees are "shared" and represented only once— is of average size $O(n(\log n)^{-1/2})$. ..... END OF EXAMPLE 17. □

▷ **33.** *Patterns in classical varieties of trees.* Patterns in general Catalan trees and Cayley trees can be similarly analysed. ◁

We shall see later that such laws, established here via explicit representations of the BGFs, extend to varieties of trees whose generating functions are only accessible implicitly via functional equations (Subsection IX. 7.3).

IX. 7.2. The exponential–logarithmic schema. So far, the occurrence of a Gaussian law has been related to a *movable singularity* that causes coefficients of a bivariate generating function $F(z,u)$ to obey a rough power law of the form

$$f_n(u) = [z^n]F(z,u) \approx \rho(u)^{-n},$$

so that the Quasi-Powers Theorem applies with a scaling factor $\beta_n = n$. In this section, we discuss the situation of a fixed singularity and *variable exponent* in singular expansions. This means a somewhat stronger decomposition property for a BGF as the singularity remains constant when the auxiliary parameter u varies, as in $F(z,u) = C(z)^{-\alpha(u)}$. Typical cases of application are to the set constructions, where the analysis of number of components can be rephrased as the estimation of coefficients in

$$F(z,u) = \exp\left(uG(z)\right),$$

when $G(z)$ is, roughly speaking, logarithmic. In this case, we have parameters whose mean and variance grow logarithmically, a typical instance being the number of cycles in permutations. Analytically, this comes from an approximate form

$$F(z,u) \approx (1 - z/\rho)^{-\alpha(u)},$$

so that

$$f_n(u) = [z^n]F(z,u) \approx \rho^{-n} n^{\alpha(u)-1} \equiv \frac{\rho^{-n}}{n} \exp(\alpha(u)\log n).$$

This is again a case of application of the Quasi-Powers Theorem, but now with a scaling factor $\beta_n = \log n$. The developments in this section are inspired by a paper of Flajolet and Soria [**177**] who first extracted certain universally valid laws for such assemblies of logarithmic structures.

THEOREM IX.11 (General variable exponent schema). *Let $F(z,u)$ be a bivariate function that is analytic at $(z,u) = (0,0)$ and has nonnegative coefficients there. Assume the following conditions.*

(i) Exponent perturbation. *Assume that there exist $\epsilon > 0$ and $r > \rho$ such that in the domain,*

$$\mathcal{D} = \left\{ (z,u) \;\middle|\; |z| \le r, |u-1| \le \epsilon \right\},$$

the function $F(z,u)$ admits the representation

$$F(z,u) = A(z,u) + B(z,u)C(z)^{-\alpha(u)} \tag{43}$$

where $A(z,u), B(z,u)$ are analytic for $(z,u) \in \mathcal{D}$, the function $\alpha(u)$ is analytic in $|u-1| \le \epsilon$ with $\alpha(1) \notin \{0,-1,-2,\ldots\}$, and $C(z)$ is analytic for $|z| \le r$, the equation $C(\zeta) = 0$ having a unique root $\zeta = \rho$ in $|z| \le r$ that is simple, with $B(\rho,1) \neq 0$.

(ii) Variability: *one has*

$$\alpha'(1) + \alpha''(1) \neq 0.$$

Then the variable with probability generating function

$$p_n(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

converges in distribution to a Gaussian variable and the speed of convergence is $O((\log n)^{-1/2})$. The corresponding mean μ_n and variance σ_n^2 satisfy

$$\mu_n \sim \alpha'(1)\log n, \qquad \sigma_n^2 \sim \alpha''(1)\log n.$$

PROOF. Clearly, for the univariate problem, by singularity analysis, one has

$$[z^n]F(z,1) = B(\rho,1)(-\rho C'(\rho))^{-\alpha(1)}\rho^{-n}\frac{n^{\alpha(1)-1}}{\Gamma(\alpha(1))}\left(1+O(\frac{1}{n}\right).$$

For the bivariate problem, the contribution arising from $[z^n]A(z,u)$ is exponentially small, since $A(z,u)$ is z-analytic in $|z| \le r$.

Write next

$$B(z,u) = (B(z,u) - B(\rho,u)) + B(\rho,u).$$

The first term satisfies

$$B(z,u) - B(\rho,u) = O((z-\rho)),$$

uniformly with respect to u, since

$$\frac{B(z,u) - B(\rho,u)}{z-\rho}$$

is analytic in z and u, by division of power series representations. Let A be an upper bound on $\alpha(u)$ on $|u-1| \le \epsilon$. Then, by singularity analysis and its companion uniformity,

$$[z^n](B(z,u) - B(\rho,u))C(z)^{-\alpha(u)} = O(\rho^{-n}n^{A-2}).$$

By suitably restricting the domain of u to $|u-1| \le \delta$, one may freely assume that $A - 2 < \alpha(1) - \frac{7}{4}$. Thus, the contribution from this part is small.

It only remains to analyse

$$[z^n]B(\rho,u)C(z)^{-\alpha(u)}.$$

This is done exactly like in the univariate case, again taking advantage of the uniformity afforded by singularity analysis. We find, uniformly for u in a small neighbourhood of 1,

$$[z^n]F(z,u) = \frac{B(\rho,u)\rho^{-n}}{n\Gamma(\alpha(u))}(-\rho C'(\rho))^{-\alpha(u)}e^{\alpha(u)\log n}\left(1+O(n^{-1/2})\right).$$

Thus, the Quasi-Powers Theorem applies and the law is Gaussian in the limit. □

The next proposition covers a scheme closely related to the exponential logarithmic setting. Its proof only requires a slight modification of the calculations involved in the error terms. It complements Example 5 where the number of small components has been found to be Poisson.

PROPOSITION IX.13 (Sets of labelled logarithmic structures). *Consider the labelled set construction $\mathcal{F} = \mathfrak{P}(\mathcal{G})$. Assume that $G(z)$ has radius of convergence ρ and is Δ-continuable with a singular expansion of the form*

$$G(z) = \kappa \log \frac{1}{1-z/\rho} + \lambda + O\left(\frac{1}{\log^2(1-z/rho)}\right).$$

Then, the limit law of the number of $\mathcal{G}$-components in a large $\mathcal{F}$-structure is asymptotically Gaussian with mean and variance both asymptotic to $\kappa \log n$.

The bivariate EGF for *permutations* with u marking the number of cycles is

$$F(z,u)=\sum \begin{bmatrix} n \\ k \end{bmatrix} u^k \frac{z^n}{n!} = (1-z)^{-u} = \exp\left(u \log \frac{1}{1-z}\right)\cdot,$$

so that we are in the simplest case of an exponential-logarithmic schema. Theorem IX.11 implies that *the number of cycles in a random permutation of size n converges to a Gaussian limiting distribution.* This classical result stating the asymptotically normal distribution of the Stirling numbers (of the first kind) constitutes Goncharov's Theorem. It has already been stated with a direct proof in Proposition IX.6, thanks to the explicit character of the "horizontal" generating functions (the Stirling polynomials) in this particular case.

EXAMPLE 18. *Cycles in derangements.* The number of cycles is asymptotically normal in generalized derangements where a finite set S of cycle lengths are forbidden. This results immediately from the BGF

$$F(z,u)=\exp(uG(z)), \qquad G(z)=\log\frac{1}{1-z}-\sum_{s\in S}\frac{z^s}{s}.$$

The classical derangement problem corresponds to $S = \{1\}$; see [**82**]. END OF EXAMPLE 18. □

EXAMPLE 19. *Clouds and 2-regular graphs.* "Clouds" are defined in [**82**, p. 274] and they have already been encountered in Chapters II and VI: let n straight lines in the plane be given in general position, so that there are $\binom{n}{2}$ intersecting points; a cloud of size n is a (maximal) set of n intersection points, no three of which are collinear. By duality, there is a one–to–one correspondence between clouds and *2–regular graphs*. A 2–regular graph of size n is an undirected graph with n edges, such that each vertex has degree exactly 2. Any 2–regular graph may be decomposed into a product of connected components that are (undirected) cycles of length at least 3. Hence the bivariate EGF for 2–regular graphs, with u marking the number of connected components, is:

$$F(z,u)=\exp\left(u(\frac{1}{2}\log\frac{1}{1-z}-\frac{z}{2}-\frac{z^2}{4})\right)=\frac{e^{-uz/2-uz^2/4}}{(1-z)^{u/2}}.$$

The function $\exp(u(z/2+z^2/4))$ is entire, so that the conditions of Theorem IX.11 are satisfied. Thus, the number of connected components in a 2–regular graph, (this is equivalent to the number of polygons in a cloud) has a Gaussian limiting distribution. END OF EXAMPLE 19. □

EXAMPLE 20. *Random mappings.* Let f denote a function that maps the set $N = \{1,2,\cdots,n\}$ into itself. Such a function f may be represented by a directed graph G_f with vertex set N and edge set $\{(i,f(i)); i\in N\}$. Such graphs, in which every point has out–degree one, are called *functional digraphs*; see [**223**, p. 68]. A functional digraph may be viewed as a set of components that are themselves cycles of rooted labelled trees. The bivariate EGF for functional digraphs with u marking connected components is

$$F(z,u)=\exp\left(u(\log\frac{1}{1-T(z)})\right),$$

where the generating function of rooted labelled trees $T(z)$ is the Cayley tree function defined implicitly by the relation $T(z) = z\exp(T(z))$. By the inversion theorem for implicit functions we have

$$T(z) = 1 - \sqrt{2(1-ez)} + \sum_{k\geq 2} c_k (1-ez)^{k/2}.$$

Thus,

$$F(z,u) = \exp\left\{u\left(\frac{1}{2}\log\frac{1}{1-ez} + H((1-ez)^{1/2})\right)\right\},$$

where $H(v)$ is analytic at $v = 0$. From this form and Theorem IX.11, we obtain a theorem of Stepanov [**397**]: *The number of components in functional digraphs has a limiting Gaussian distribution.*

This approach extends to functional digraphs satisfying various degree constraints as considered in [**14**]. This analysis and similar ones are relevant to integer factorization, using Pollard's "rho" method [**166, 266, 382**]. . . . . END OF EXAMPLE 20. □

Unlabelled constructions. In the case of unlabelled structures, the class $\mathcal{F}$ of multisets over a class $\mathcal{G}$ have OGF,

$$\sum_{n\geq 0} F_n z^n = \prod_{n\geq 1} (1-z^n)^{-G_n}.$$

By taking logarithms and reorganizing the corresponding series, we get the alternative form

$$F(z) = \exp\left(\frac{G(z)}{1} + \frac{G(z^2)}{2} + \frac{G(z^3)}{3} + \cdots\right).$$

Similarly, in the bivariate case, where u marks the number of components, the bivariate GF is (see Chapter III),

$$F(z,u) = \sum_{n,k\geq 0} F_{n,k} u^k z^n = \exp\left(\frac{u}{1}G(z) + \frac{u^2}{2}G(z^2) + \frac{u^3}{3}G(z^3) + \cdots\right),$$

which is of the form $\exp(G(z))^u \cdot B(z,u)$. Here, we are interested in structures such that $G(z)$ has a logarithmic singularity, in which case Theorem IX.11 applies, as soon as $G(z)$ has radius of convergence $\rho < 1$.

EXAMPLE 21. *Polynomial factorization.* Fix a finite field $K = GF(q)$ and consider the class $\mathcal{P}$ of monic polynomials (having leading coefficient 1) in $K[z]$, with $\mathcal{I}$ the subclass of irreducible polynomials. Obviously, $P_n = q^n$, so that

$$P(z) = (1-qz)^{-1}.$$

Because of the unique factorization property, a polynomial is a multiset of irreducible polynomial, whence the relation

$$P(z) = \exp\left(\frac{I(z)}{1} + \frac{I(z^2)}{2} + \frac{I(z^3)}{3} + \cdots\right).$$

The preceding relation can be inverted using Möbius inversion. If we set $L(z) = \log P(z)$, then we have

$$I(z) = \sum_{k\geq 1} \mu(k)\frac{L(z^k)}{k} = \log\frac{1}{1-qz} + \sum_{k\geq 2}\mu(k)\frac{L(z^k)}{k},$$

where μ is the M$\ddot{\text{b}}$ius function.

Since $L(z^k)$ is analytic for $|z| < q^{-1/2}$ whenever $k \geq 2$, and $|L(z^k)| < c^{st}|z|^k$, the sum $\sum_{k\geq 2} \mu(k)L(z^k)/k$ is analytic for $|z| \leq \tau$, with $q^{-1} < \tau < q^{-1/2}$. Hence $I(z)$ has an isolated singularity of logarithmic type at $z = q^{-1} < 1$.

Thus the average number of irreducible factors in a polynomial, and its variance, are both asymptotically $\log n + O(1)$ (this result appears in [**266**, Ex. 4.6.2.5]). Let Ω_n be the random variable representing the number of irreducible factors of a random polynomial of degree n over $GF(q)$, each factor being counted with its order of multiplicity. Then as n tends to infinity, for any two real constants $\lambda < \mu$, we have

$$\mathbb{P}\{\log n + \lambda\sqrt{\log n} < \Omega_n < \log n + \mu\sqrt{\log n}\} \to \frac{1}{\sqrt{2\pi}}\int_\lambda^\mu e^{-t^2/2}\,dt.$$

This statement [**177**] is a counterpart of the famous Erdös–Kac Theorem (1940) for the number of prime divisors of natural numbers (with here $\log n$ that replaces $\log\log n$ when dealing with integers at most n). A similar result holds for the parameter ω_n that represents the number of *distinct* irreducible factors in a random polynomial of degree n. ... END OF EXAMPLE 21. □

It is perhaps instructive to re-examine this last example at an abstract level, in the light of general principles of analytic combinatorics.

> *A polynomial over a finite field is determined by the* sequence *of its coefficients. Hence, the class of all polynomials, as a sequence class, has a polar singularity. On the other hand, unique factorization entails that a polynomial is also a* multiset *of irreducible factors ("primes"). Thus, the class of irreducible polynomials, that is implicitly determined, is logarithmic, since the multiset construction to be inverted is in essence an exponential operator. Consequently, the number of irreducible factors obeys the exponential-logarithmic scheme, so that it is asymptotically Gaussian.*

Eventually, the limit law arises because of the purely analytic character of the generating functions involved, together with permanence of analytic relations implied by combinatorial constructions.

EXAMPLE 22. *Mapping patterns.* Let f and g be two functions mapping the set $\{1, 2, \cdots, n\}$ into itself. Mappings f and g are said to be equivalent if there exists a permutation π of $\{1, 2, \cdots, n\}$ such that $f(i) = j$ iff $g(\pi(i)) = \pi(j)$. Mapping patterns are thus equivalence classes of mapping functions, or equivalently functional digraphs on unlabelled points. They correspond to multisets of cycles of rooted unlabelled trees. The OGF for rooted unlabelled trees satisfies the implicit relation $A(z) = z\exp(\sum \frac{1}{k}A(z^k))$, and Otter [**335**] proved that

$$A(z) = 1 - c_1\sqrt{(1 - z/\eta)} + \sum_{k\geq 2} c_k(1 - z/\eta)^k.$$

for some $\eta < 1$: see our detailed account in Chapter VII.

On the other hand, by the translation of the cycle construction, if $\mathcal{G}$ is the unlabelled cycle construction applied to $\mathcal{A}$, then (see Chapter III),

$$G(z) = \sum_{k\geq 1} \frac{\phi(k)}{k}\log\frac{1}{1 - A(z^k)},$$

where $\phi(k)$ is the Euler totient function. In the present context, since $A(z)$ has radius of convergence η strictly less than 1,

$$G(z) = \log \frac{1}{1 - A(z)} + S(z),$$

where $S(z)$ is analytic at η. Finally the bivariate OGF for random mapping patterns satisfies

$$\begin{aligned} F(z,u) &= \exp\big(\sum_{k\geq 1} u^k \frac{G(z^k)}{k}\big) \\ &= \exp\big(u \log \frac{1}{1-A(z)} + uS(z) + T(z,u)\big) \\ &= \exp\left(\frac{u}{2} \log \frac{1}{1 - z/\eta} + u\big((1 - z/\eta)^{1/2}\big) + uS(z) + T(z,u)\right), \end{aligned}$$

where $S(z)$ is analytic at η, $T(z,u)$ is analytic for $z = \eta$ and $u = 1$, and H is analytic around 0, with $H(0) = 0$. Thus conditions for applying Theorem IX.11 are satisfied and the number of components in random mapping patterns has a Gaussian limiting distribution. The mean value is asymptotic to $\frac{1}{2}\log n$ (this result appears in [**313**] and the variance is $\frac{1}{2}\log n + O(1)$.
END OF EXAMPLE 22. □

EXAMPLE 23. *Arithmetical semigroups.* Knopfmacher [**259**] defines an arithmetical semigroup as a semigroup with unique factorization, and a size function (or degree) such that

$$|xy| = |x| + |y|,$$

where the number of elements of a fixed size is finite. If $\mathcal{P}$ is an arithmetical semigroup and $\mathcal{I}$ its set of 'primes' (irreducible elements), axiom $A^{\#}$ of Knopfmacher asserts the condition

$$\text{card}\{x \in \mathcal{P} \ / \ |x| = n\} = cq^n + O(q^{\alpha n}) \quad (\alpha < 1).$$

It is shown by Knopfmacher that several algebraic structures forming arithmetical semigroups satisfy axiom $A^{\#}$, and thus the conditions of Theorem IX.11 are automatically satisfied. Therefore, the results deriving from Theorem IX.11 fit into the framework of Knopfmacher's "abstract analytic number theory", since they provide general conditions under which theorems of the Erdös–Kac type must hold true. Examples of application are Galois polynomial rings (the example of polynomial factorization), finite modules or semisimple finite algebras over a finite field $K = GF(q)$, integral divisors in algebraic function fields, ideals in the principal order of a algebraic function field, finite modules, or semisimple finite algebras over a ring of integral functions. . . . . END OF EXAMPLE 23. □

IX. 7.3. Algebraic and implicit functions. Many combinatorial problems, especially as regards paths and trees, lead to descriptions by context–free languages. Accordingly, the GF's are algebraic functions. The most frequent situation is that of univariate GF's having singularities of the square-root type.

COROLLARY IX.1 (Algebraic functions). *Let $F(z,u)$ be a bivariate function that is analytic at $(0,0)$ and has nonnegative coefficients. Assume that $F(z,u)$ is one of the solutions y of a polynomial equation*

$$\Phi(z,u,y) = 0,$$

where Φ is an irreducible polynomial of total degree m, of degree $d \geq 2$ in y. Assume that $F(z,1)$ is has a unique dominant singularity at $\rho > 0$, with a singular behaviour of the square-root type there. Define the resultant polynomial,

$$\Delta(z,u) = \text{result}_{\,y}\left(\Phi(z,u,y), \frac{\partial}{\partial y}\Phi(z,u,y)\right),$$

and assume that ρ is a simple root of $\Delta(z,1)$. Let $\rho(u)$ be the unique root of the equation

$$\Delta(\rho(u),u),$$

analytic at 1, such that $\rho(1) = \rho$. Then, provided the variability condition

$$\mathfrak{v}\left(\frac{\rho(1)}{\rho(u)}\right) > 0,$$

is satisfied, a Gaussian Limit Law holds for the coefficients of $F(z,u)$.

PROOF. The assumption of a square-root singularity (see Chapters VI and VII) means that the polynomial $\Phi(\rho,1,y)$ has a double zero at $y = \tau$, where $\tau = \lim_{z\to\rho^-} F(z,1)$. Equivalently, we have

$$\left(\frac{\partial}{\partial y}\Phi(\rho,1,y)\right)_{y=\tau} = 0, \qquad \left(\frac{\partial^2}{\partial y^2}\Phi(\rho,1,y)\right)_{y=\tau} \neq 0.$$

Thus, Weierstrass preparation gives the local factorization

$$\Phi(z,u,y) = (y^2 + c_1(z,u)y + c_2(z,u))H(z,u,y),$$

where $H(z,u,y)$ is analytic and nonzero at $(\rho,1,\tau)$ while $c_1(z,u), c_2(z,u)$ are analytic at $(z,u) = (\rho,\tau)$.

From the solution of the quadratic equation, we must have locally

$$y = \frac{1}{2}\left(-c_1(z,u) \pm \sqrt{c_1(z,u)^2 - 4c_2(z,u)}\right).$$

Consider first (z,u) restricted by $0 \leq z < \rho$ and $0 \leq u < 1$. Since $F(z,u)$ is real there, we must have $c_1(z,u)^2 - 4c_2(a,u)$ also real and nonnegative. Since $F(z,u)$ is continuous and increasing with z for fixed u, and since the discriminant $c_1(z,u)^2 - 4c_2(a,u)$ vanishes at 0, the determination with the minus sign has to be constantly taken. In summary, we have

$$F(z,u) = \frac{1}{2}\left(-c_1(z,u) - \sqrt{c_1(z,u)^2 - 4c_2(z,u)}\right).$$

The function $C(z,u) = c_1^2(z,u) - 4c_2(z,u)$ has a simple real zero at $(\rho,1)$. Thus there is locally a unique analytic branch of the solution to $C(\rho(u),u) = 0$ such that $\rho(1) = \rho$.. This branch is also by necessity a root of the resultant equation $\Delta(\rho(u),u) = 0$. The conditions of Theorem IX.10 therefore apply and the Gaussian law follows. □

This theorem asserts that, under suitable conditions, the only possible dominant singularity of the BGF is a "lifting" of the singularity of the univariate GF $F(z,1)$ and the nature of the singularity —the square-root type— does not change. The result generalizes to the case of a function Φ that is analytic in sufficiently large bounded domains, *e.g.*, an entire function. The condition is that the analytic curves

$$\Phi(z,u,y)=0, \qquad \frac{\partial}{\partial y}\Phi(z,u,y)=0$$

have an intersection that "moves analytically" and nontrivially for u near 1, and a sufficient condition for this is the nonvanishing of the Jacobian determinant

$$J(z,u,y):=\begin{vmatrix} \frac{\partial}{\partial z}\Phi(z,u,y) & \frac{\partial}{\partial y}\Phi(z,u,y) \\ \frac{\partial^2}{\partial zy}\Phi(z,u,y) & \frac{\partial^2}{\partial y^2}\Phi(z,u,y) \end{vmatrix} \tag{44}$$

and its first derivative with respect to u at $(\rho,1,\tau)$,

$$J(\rho,1,\tau)\neq 0, \qquad \left.\frac{\partial}{\partial u}J(z,u,y)\right|_{(\rho,1,\tau)}\neq 0. \tag{45}$$

In the case of Corollary IX.1 and of these extensions, the expansion of $\rho(u)$ at $u=1$, hence the mean and variance of the distribution, are computable explicitly from Φ, its derivatives, and the quantities ρ and $\tau=F(\rho,1)$.

The corollary applies to a great variety of decomposable parameters of context–free languages, tree like objects, and more generally many recursively defined combinatorial types. Examples of parameters covered are leaves, node types, and various sorts of patterns in combinatorial tree models. Drmota has worked out a different set of conditions for asymptotic normality. In particular, one of Drmota's important results [**111**] yields asymptotic normality, under minor technical restrictions, for a polynomial *system* with positive coefficients that is "irreducible", meaning that the dependency graph between nonterminals is strongly connected.

▷ **34.** *Nodes of degree k in simple varieties of trees.* Their distribution is asymptotically Gaussian. ◁

▷ **35.** *Leaves in nonplane unlabelled trees.* Their distribution is asymptotically Gaussian. ◁

IX. 7.4. Differential equations. *Ordinary differential equations* (ODE's, for short) in one variable, when *linear* and with analytic coefficients, have solutions whose singularities occur at well-defined places, namely those that entail a reduction of order. The possible singular exponents of solutions are then obtained as roots of a polynomial equation, the indicial equation. Such ordinary differential equations are usually a reflection of a combinatorial decomposition and suitably parametrized versions then open access to a number of combinatorial parameters. In this case, the ODE normally remains an ODE in the main variable z that records size, while the auxiliary variable u only affects the coefficients but not the global shape of the original ODE.

Three cases may then occur for a linear ODE parametrized by u.

- *Movable singularity:* the location of the dominant singularity $\rho(u)$ changes with u but the singular exponent does not change; the analysis is then similar to that of algebraic-logarithmic singularities.
- *Movable exponent:* the dominant singularity does not move but the singular exponent $\alpha(u)$ changes; the analysis then resorts to the exponential-logarithmic schema.

- *Movable singularity and movable exponent:* in this case, the singular behaviour is essentially dictated by the movable singularity but with an auxiliary contribution arising from the movable exponent; the analysis of this mixed case then requires an extension of the quasi-power framework, as developed by Gao in Richmond in [**192**].

Here, we focus on the important case of a fixed singularity and a movable exponent. The required singularity perturbation analysis is inspired by the treatment of Flajolet and Lafforgue in [**163**]. The corresponding univariate problems resort to holonomic asymptotics.

Linear differential equations. The example of the distribution of levels of nodes in random binary search trees or heap–ordered trees illustrates well the situation of a fixed singularity and movable exponent. A heap–ordered tree (HOT) is a plane binary increasing tree. HOTs constitute an unambiguous tree representation of permutations [**382**]. The EGF of HOTs is

$$F(z) = \frac{1}{1-z} = \sum_{n\geq 0} n!\,\frac{z^n}{n!},$$

as results either from the combinatorial bijection with permutations or from the root decomposition of increasing trees that translates into the functional equation,

$$F(z) = 1 + \int_0^z F^2(t)\,dt, \tag{46}$$

a Riccati equation in disguise. Let $F(z,u)$ be the BGF of HOT's where u records the depth of external nodes. In other words, $f_{n,k} = [z^n u^k]F(z,u)$ is such that $\frac{1}{n}f_n n,k$ represents the probability that a random external node in a random tree of size n is at depth k in a random tree. The probability space is then a product set of cardinality $(n+1)\cdot n!$, as there are $n!$ trees each containing $(n+1)$ external nodes. By a standard equivalence principle, the quantities $\frac{1}{n}f_n n,k$ also give the probability that a random unsuccessful search in a random binary search tree of size n necessitates k comparisons.

Since the depth of a node is inherited from subtrees, the function $F(z,u)$ satisfies the *linear* integral equation,

$$F(z,u) = 1 + 2u\int_0^z F(t,u)\,\frac{dt}{1-t}, \tag{47}$$

or, after differentiation,

$$\frac{\partial}{\partial z}F(z,u) = \frac{2u}{1-z}F(z,u), \qquad F(0,u) = 1.$$

This equation is in fact a linear ODE with u entering as a parameter,

$$\frac{d}{dz}y(z) - \frac{2u}{1-z}y(z) = 0, \qquad y(0) = 0.$$

The solution of any separable first-oder ODE is obtained by quadratures, here,

$$F(z,u) = \frac{1}{(1-z)^{2u}}.$$

From singularity analysis, provided u avoids $\{0, -\frac{1}{2}, -1, \ldots\}$, we have

$$f_n(u) := [z^n]F(z,u) = \frac{n^{2u-1}}{\Gamma(2u)}\left(1 + O(\frac{1}{n})\right),$$

and the error term is uniform in u provided, say, $|u-1| \leq \frac{1}{4}$. Thus, Theorem IX.11 applies, and the law with PGF $f_n(u)/f_n(1)$ converges to a Gaussian limit.

A similar result holds for levels of internal nodes, and is proved by similar devices. The Gaussian profile is even perceptible on single instance (see the particular figure in Chapter III), which actually suggests a stronger "functional limit theorem" for these objects: this has been proved by Chauvin and Jabbour [**78**] using martingale theory.

Naturally, explicit expressions are available in such a simple case,

$$\frac{f_n(u)}{f_n(1)} = \frac{2u \cdot (2u+1) \cdots (2u+n-1)}{(n+1)!},$$

so a direct proof of the Gaussian limit in the line of Goncharov's theorem is clearly possible; see Mahmoud's book [**307**, Ch. 2], for this result originally due to Louchard. What is interesting here is the fact that $F(z,u)$ viewed as a function of z has a singularity at $z=1$ that does not move and, in a way, originates in the combinatorics of the problem—the EGF $(1-z)^{-1}$ of permutations. The auxiliary parameter u appears here directly in the exponent, so that the application of singularity analysis or of the more sophisticated Theorem IX.11 is immediate.

COROLLARY IX.2 (Linear differential equations). *Let $F(z,u)$ be a bivariate generating function with nonnegative coefficients that satisfies a linear differential equation*

$$a_0(z,u)\frac{\partial^r F}{\partial z^r} + \frac{a_1(z,u)}{(\rho - z)}\frac{\partial^{r-1}F}{\partial z^{r-1}} + \cdots + \frac{a_r(z,u)}{(\rho-z)^r}F = 0,$$

with $a_j(z,u)$ analytic at ρ, and $a_0(\rho,1) \neq 0$. Let $f_n(u) = [z^n]F(z,u)$, and assume the following conditions:

- [Nonconfluence] *The indicial polynomial*

(48) $$J(\alpha) = a_0(\rho,1)(\alpha)_{(r)} + a_1(\rho,1)(\alpha)_{(r-1)} + \cdots + a_r(\rho,1)$$

  has a unique root $\sigma > 0$ which is simple and such that all other roots $\alpha \neq \sigma$ satisfy $\Re(\alpha) < \sigma$;
- [Dominant growth] $f_n(1) \sim C \cdot \rho^{-n} n^{\sigma-1}$, *for some $C > 0$.*
- [Variability condition]

$$\sup \frac{\mathfrak{v}(f_n(u))}{\log n} > 0.$$

Then the coefficients of $F(z,u)$ admit a limit Gaussian law.

PROOF. (See the paper by Flajolet and Lafforgue [**163**] for a detailed example or the books by Henrici [**229**] and Wasow [**431**] for a general treatment of singularities of linear ODEs.) We assume in this proof that no two roots of the indicial polynomial (48) differ by an integer. Consider first the univariate problem. A differential equation,

$$a_0(z)\frac{d^rF}{dz^r} + \frac{a_1(z)}{(\rho - z)}\frac{d^{r-1}F}{dz^{r-1}} + \cdots + \frac{a_r(z)}{(\rho - z)^r}F = 0, \tag{49}$$

with the $a_j(z)$ analytic at ρ and $a_1(\rho) \neq 0$ has a basis of local singular solutions obtained by substituting $(\rho - z)^{-\alpha}$ and cancelling the terms of maximum order of growth. The candidate exponents are thus roots of the *indicial equation*,

$$J(\alpha) \equiv a_0(\rho)(\alpha)_{(r)} + a_1(\rho)(\alpha)_{(r-1)} + \cdots + a_r(\rho) = 0.$$

If there is a unique (simple) root of maximum real part, α_1, then there exists a solution to (49) of the form

$$Y_1(z) = (\rho - z)^{-\alpha_1} h_1(\rho - z),$$

where $h_1(w)$ is analytic at 0 and $h_1(0) = 1$. (This results easily from a solution by indeterminate coefficients.) All other solutions are then of smaller growth and of the form

$$Y_j(z) = (\rho - z)^{-\alpha_j} h_j(\rho - z)\left(\log(z - \rho)\right)^{k_j},$$

for some integers k_j and some functions $h_j(w)$ analytic and nonzero at $w = 0$. Then, $F(z)$ has the form

$$F(z) = \sum_{j=1}^{r} c_j Y_j(z).$$

Then, provided $c_1 \neq 0$,

$$[z^n]F(z) = \frac{c_1}{\Gamma(\sigma)}\rho^{-n} n^{\alpha_1 - 1}(1 + o(1)).$$

Under the assumptions of the theorem, we must have $\sigma = \alpha_1$, and $c_1 \neq 0$. The reality assumption is natural for a series $F(z)$ that has real coefficients.

When u varies in a neighbourhood of 1, we have a uniform expansion

$$F(z, u) = c_1(u)(\rho - z)^{-\sigma(u)} H_1(\rho - z, u)(1 + o(1)), \tag{50}$$

for some bivariate analytic function $H_1(w, u)$ with $H_1(0, u) = 1$, where $\sigma(u)$ is the algebraic branch that is a root of

$$J(\alpha, u) \equiv a_0(\rho, u)(\alpha)_{(r)} + a_1(\rho, u)(\alpha)_{(r-1)} + \cdots + a_r(\rho, u) = 0,$$

and coincides with σ at $u = 1$. By singularity analysis, this entails

$$[z^n]F(z, u) = \frac{c_1(u)}{\Gamma(\sigma)}\rho^{-n} n^{\alpha_1(u) - 1}(1 + o(1)), \tag{51}$$

uniformly for u in a small neighbourhood of 1, with the error term being $O(n^{-a})$ for some $a > 0$. Thus Theorem IX.11 applies and the limit law is Gaussian.

The crucial point in (50,51) is the uniform character of expansions with respect to u. This results from two facts: (i) the solution to (49) may be specified by analytic conditions at a point z_0 such that $z_0 < \rho$ and there are no singularities of the equation between z_0 and ρ. (ii) there is a suitable set of solutions with an analytic component in z and u and singular parts of the form $(\rho - z)^{-\alpha_j(u)}$, as results from the matrix theory of differential systems and majorant series. (This last point is easily verified if no two roots of the indicial equation differ by an integer; otherwise, see [**163**] for an alternative basis of solutions for u near 1, $u \neq 1$.) □

EXAMPLE 24. *Node levels in quadtrees.* This example is taken from [**163**]. Quadtrees are one of the most versatile data structure for managing a collection of points in multidimensional space. They are based on a recursive decomposition similar to that of BSTs.

Here d is the dimension of the data space. Let $f_{n,k}$ be the number of external nodes at level k in a quadtree of size n grown by random insertions, and let $F(z,u)$ be the corresponding BGF. Two integral operators play an essential rôle,

$$\mathbf{I}\, g(z) = \int_0^z g(t)\, \frac{dt}{1-t} \qquad \mathbf{J}\, g(z) = \int_0^z g(t)\, \frac{dt}{t(1-t)}.$$

The basic equation that reflects the recursive splitting process of quadtrees is then

$$F(z,u) = 1 + 2^d u \mathbf{J}^{d-1} \mathbf{I}\, F(z,u). \tag{52}$$

The integral equation (52) satisfied by F then transforms into a differential equation of order d,

$$\mathbf{I}^{-1}\mathbf{J}^{1-d}\, F(z,u) = 2^d u F(z,u),$$

where

$$\mathbf{I}^{-1} g(z) = (1-z) g'(z), \qquad \mathbf{J}^{-1} g(z) = z(1-z) g'(z).$$

The linear ODE version of (52) has an indicial polynomial that is easily determined by examination of the reduced form of the ODE (52) at $z = 1$. There, one has

$$\mathbf{J}^{-1} g(z) = \mathbf{I}^{-1} g(z) - (z-1)^2 g'(z) \approx (1-z) g'(z).$$

Thus,

$$\mathbf{I}^{-1}\mathbf{J}^{1-d} (1-z)^{-\theta} = \theta^d (1-z)^{-\theta} + O((1-z)^{-\theta+1}),$$

and the indicial polynomial is

$$J(\alpha, u) = \alpha^d - 2^d u.$$

In the univariate case, the root of largest real part is $\alpha_1 = 2$; in the bivariate case, we have

$$\alpha_1(u) = 2u^{1/d},$$

where the principal branch is chosen. Thus,

$$f_n(u) = \gamma(u) n^{\alpha_1(u)} (1 + o(1)).$$

By the combinatorial origin of the problem, $F(z,1) = (1-z)^{-2}$, so that the coefficient $\gamma(1)$ is nonzero. Thus, the conditions of the corollary are satisfied. The law is Gaussian in the limit, with mean and variance

$$\mu_n \sim \frac{2}{d} \log n, \qquad \sigma_n^2 \sim \frac{2}{d} \log n.$$

The same result applies to the cost of a random search, either successful or not, as shown in [**163**] by an easy combinatorial argument. .... END OF EXAMPLE 24. □

Nonlinear differential equations. Though nonlinear differential equations do not obey a simple classification of singularities, there are a few examples in analytic combinatorics that can be treated by singularity perturbation methods. We detail here typical analysis of properties of binary search trees (BSTs), equivalently HOTs, that is taken from [**154**]. The Riccati equation involved reduces, by classical techniques, to a linear second order equation whose perturbation analysis is particularly transparent and akin to earlier analyses of ODEs. In this problem, the auxiliary parameter induces a movable singularity that directly resorts to the Quasi-Powers Theorem.

EXAMPLE 25. *Paging of binary search trees.* Fix a "bucket size" parameter $b \geq 2$. Given a binary search tree t, its b-*index* is a tree that is constructed by retaining only those internal nodes of t which correspond to subtrees of size $> b$. As a data structure, such an index is well-suited to "paging", where one has a two-level hierarchical memory structure: the index resides in main memory and the rest of the tree is kept in pages of capacity b on peripheral storage, see for instance [**307**]. We let $\iota[t] = \iota_b[t]$ denote the size —number of nodes— of the b-index of t.

Like in Eq. (46), the bivariate generating function

$$F(z,u) := \sum_t \lambda(t) u^{\iota[t]} z^{|t|}$$

satisfies a Riccati equation that reflects the root decomposition of trees,

$$\text{(53)} \qquad \frac{\partial}{\partial z} F(z,u) = uF^2(z,u) + (1-u)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right), \qquad F(0,u) = 1,$$

where the general quadratic relation (46) has to be corrected in its low order terms.

The GFs of moments are rational functions with a denominator that is a power of $(1-z)$, as results from differentiation at $u = 1$. Mean and variance follow:

$$\mu_n = \frac{2(n+1)}{b+2} - 1, \qquad \sigma_n^2 = \frac{2}{3}\,\frac{(b-1)b(b+1)}{(b+2)^2}\,(n+1).$$

(The result for the mean is well-known, refer to quantity A_n in the analysis of quicksort on p. 122 of [**264**].)

Multiplying both sides of (53) by u now gives an equation satisfied by $H(z,u) := uF(z,u)$,

$$\frac{\partial}{\partial z} H(z,u) = H^2(z,u) + u(1-u)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right),$$

that may as well be taken as a starting point since $H(z,u)$ is the bivariate GF of parameter $1+\iota_b$ (a quantity also equal to the number of external pages). The classical linearization transformation of Riccati equations,

$$H(z,u) = -\frac{X_z'(z,u)}{X(z,u)},$$

yields

$$\text{(54)} \qquad \frac{\partial^2}{\partial z^2} X(z,u) + u(u-1)A(z)X(z,u) = 0, \qquad A(z) = \frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right),$$

with $X(0,u) = 1, X_z'(0,u) = -u$. By the classical existence theorem of Cauchy, the solution of (54) is an entire function of z for each fixed u, as the linear differential equation has no singularity at a finite distance. Furthermore, the dependency of X on u is also everywhere analytic; see the remarks of [**431**, Sec. 24], for which a proof derives by inspection of the classical existence proof based on indeterminate coefficients and majorant series. Thus, $X(z,u)$ is actually an entire function of *both* complex variables z and u. As a consequence, for any fixed $u = u_0$, the function $H(z,u_0)$ is a meromorphic function of z whose coefficients are amenable to singularity analysis.

In order to proceed further, we need to prove that, in a sufficiently small neighbourhood of $u = 1$, $X(z,u)$ has only one simple root, corresponding for $H(z,u)$ to a unique dominant and simple pole. This fact itself derives from general considerations surrounding the Preparation Theorem of Weierstrass: *in the vicinity of any point* (z_0,u_0) *with* $X(z_0,u_0) = 0$, *the roots* *of the bivariate analytic equation $X(z,u)=0$ are locally branches of an algebraic function.* Here, we have $X(z,1)\equiv 1-z$. Thus, as u tends to 1, all solutions of $X(z,u)$ must escape to infinity except for one branch $\rho(u)$ that satisfies $\rho(1)=1$. By the nonvanishing of $X'_u(z,1)$ and the implicit function theorem, the function $\rho(u)$ is additionally an analytic function about $u=1$.

The argument is now complete: for u in a sufficiently complex neighbourhood of 1, we have a Quasi-Powers approximation,

$$[z^n]H(z,u)=\rho(u)^{-n-1}\left(1+O(K^{-n})\right),$$

for some fixed constant $K>0$. The Gaussian limit results. .... END OF EXAMPLE 25. □

As shown in [**154**], a similar analysis applies to patterns in binary search trees and heap–ordered trees. This is related to the analysis of local order patterns in permutations, for which gaussian limit laws have been obtained by Devroye [**104**] using extensions of the central limit theorem to weakly dependent random variables.

Similar displacements of singularity arise for node types in varieties of increasing trees, extending the case of HOTs that are binary. This is discussed in [**38**]. For instance, if $\phi(w)$ is the degree generator a family of increasing trees, the nonlinear ODE satisfied by the BGF of leaves is

$$\frac{\partial}{\partial z}F(z,u)=(u-1)\phi(0)+\phi(F(z,u)).$$

Whenever ϕ is a polynomial, there is a spontaneous singularity at some $\rho(u)$ that depends analytically on u. Thus, again the Quasi-Powers Theorem applies; see [**38**].

IX. 8. Perturbation of saddle point asymptotics

We shall be brief here, as the subject is excellently covered in Sachkov's book to which we refer for details. Entire functions and functions with a fast growth at their singularity do not in general lead to quasi-power expansions. As we known from univariate asymptotics (Chapter VIII), the coefficient expansions involve a combination of large powers (that arise from the Cauchy kernel) and of the very fast singular behaviour of the function under consideration. Accordingly, bivariate asymptotic studies necessitate a perturbation of saddle point expansions. A framework more flexible than the Quasi-Powers Theorem is then needed.

Here, we base our brief discussion on a theorem taken from Sachkov's book [**371**].

THEOREM IX.12 (Generalized quasi-powers). *Assume that the generating function $p_n(u)$ of a discrete random variable X_n has a representation of the form*

$$p_n(u)=\exp\left(h_n(u)\right)(1+o(1)),$$

that holds uniformly, where each $h_n(u)$ is analytic in a fixed neighbourhood Ω of 1. Assume also the condition,

$$\frac{h_n'''(u)}{(h_n'(1)+h_n''(1))^{3/2}}\to 0, \tag{55}$$

uniformly for $u\in\Omega$. Then, the random variable

$$X_n^*=\frac{X_n-h_n'(1)}{(h_n'(1)+h_n''(1))^{1/2}}$$

converges in distribution to a normal law with parameters $(0, 1)$.

PROOF. See [**371**, Sec. 1.4] for details. Set $\sigma^2 = h_n'(1) + h_n''(1)$, and expand the Laplace transform of X_n at t/σ. This gives

$$h_n(e^{t/\sigma}) = h_n'(1)\frac{t}{\sigma} + (h_n'(1) + h_n''(1))\frac{t^2}{2\sigma} + o(1).$$

Thus, the Laplace transform of X_n^* converges to the transform of a standard Gaussian. □

This theorem extends the quasi-power scheme. In effect, if

$$h_n(u) = \beta_n \log B(u) + A(u),$$

then the quantity (55) is $O(\beta_n^{-1/2})$, uniformly. The application of this theorem to saddle point integrals is in principle routine, though the manipulation of asymptotic scales associated with expressions involving the saddle point value may become cumbersome. We detail here the case of singletons in random involutions for which the saddle point is an algebraic function of n and u.

▷ **36.** *Effective speed bounds.* A metric version of the theorem, with error terms, cane be developed assuming suitable error bounds. ◁

EXAMPLE 26. *Singletons in random involutions.* This example is again borrowed from Sachkov's book [**371**]. The BGF is

$$F(z, u) = \exp\left(zu + \frac{z^2}{2}\right).$$

The saddle point equation (see Chapter VIII) is then

$$\left(\frac{d}{dz}uz + \frac{z^2}{2} - (n+1)\log z\right)_{z=\zeta} = 0.$$

This defines the saddle point $\zeta \equiv \zeta(n, u)$,

$$\begin{aligned}\zeta(n,u) &= -\frac{u}{2} + \frac{1}{2}\sqrt{4n+4+u^2} \\ &= \sqrt{n} - \frac{u}{2} + \frac{u^2+4}{8}\,\frac{1}{\sqrt{n}} + O(n^{-1}),\end{aligned}$$

where the error term is uniform for u near 1. By the saddle point formula, one has

$$[z^n]F(z,u) = \frac{1}{\sqrt{2\pi D(n,u)}} F(\zeta(n,u),u)\zeta(n,u)^{-n}.$$

The denominator is determined in terms of second derivatives, according to the classical saddle point formula (Chapter VIII),

$$D(n,u) = \left(z^2\frac{\partial^2}{\partial z^2} + z\frac{\partial^2}{\partial z^2}\left[uz + \frac{z^2}{2}\right]\right)_{z=\rho},$$

and its main asymptotic order does not change when u varies in a sufficiently small neighbourhood of 1,

$$D(n,u) = 2n - u\sqrt{n} + O(1),$$

again uniformly. Thus, the PGF of the number of singleton cycles satisfies

$$p_n(u) = \frac{F(\zeta(n,u),u)}{F(\zeta(n,1),1)}\left(\frac{\zeta(n,u)}{\zeta(n,1)}\right)^{-n}(1+o(1)),$$

uniformly, for u near 1. This is of the form

$$p_n(u) = \exp\left(h_n(u)\right)\,(1 + o(1)),$$

and local expansions then yield the centering constants

$$a_n := h_n'(1) = \sqrt{n} - \frac{1}{2} + O(n^{-1/2}), \qquad b_n^2 := h_n'(1) + h_n''(1) = \sqrt{n} - 1 + O(n^{-1/2}).$$

The theorem applies directly to this case and the variable

$$\frac{1}{b_n}(X_n - a_n)$$

is asymptotic to a standard normal.

A little care with the error terms in the asymptotic expansions shows that the mean and standard deviation μ_n, σ_n are asymptotic to a_n, b_n, respectively. Therefore, the number of singletons in a random involution of size n has mean μ_n and standard deviation σ_n that satisfy

$$\mu_n \sim n^{1/2}, \qquad \sigma_n \sim n^{1/4}.$$

This computation also determines the law of doubleton cycles and of all cycles, that are given by

$$\frac{1}{2}(n - X_n), \qquad \frac{1}{2}(n + X_n),$$

respectively. In particular, the number of doubleton cycles has average $\frac{1}{2}n - \frac{1}{2}n^{1/2}$. Thus, a random involution has a relatively small number of singleton cycles. END OF EXAMPLE 26. □

EXAMPLE 27. *The Stirling partition numbers.* The numbers $\left\{ {n \atop k} \right\}$ correspond to the BGF

$$F(z, u) = \exp\left(u(e^z - 1)\right).$$

The saddle point $\zeta \equiv \zeta(n, u)$ is the positive root near $n/\log n$ of the equation

$$\zeta e^{\zeta} = \frac{n+1}{u}.$$

The theorem applies:

PROPOSITION IX.14. *The Stirling partition distribution defined by $\frac{1}{S_n}\left\{ {n \atop 2} \right\}$, with S_n a Bell number, is asymptotically normal, with mean and variance that satisfy*

$$\mu_n \sim \frac{n}{\log n}, \qquad \sigma_n^2 \sim \frac{n}{(\log n)^2}.$$

We refer once more to Sachkov's book for computational details. END OF EXAMPLE 27. □

Summarizing the last example as well as earlier results, we now have the fact that all four Stirling-related distributions,

$$\frac{1}{n!}\begin{bmatrix} n \\ k \end{bmatrix}, \quad \frac{k!}{O_n}\begin{bmatrix} n \\ k \end{bmatrix}, \quad \frac{1}{S_n}\begin{Bmatrix} n \\ k \end{Bmatrix}, \quad \frac{k!}{R_n}\begin{Bmatrix} n \\ k \end{Bmatrix},$$

associated to permutations, alignments, set partitions, and surjections are asymptotically Gaussian.

Saddle point and functional equations. The average-case analysis of the number of nodes in random digital trees or "tries" can be carried out using the Mellin transform technology. The corresponding distributional analysis is appreciably harder and due to Jacquet and Régnier [**241**]. A complete description is offered in Section 5.4 of Mahmoud's book which we follow. What is required is to analyse the BGF

$$F(z,u) = e^z T(z,u),$$

where the Poisson generating function $T(z,u)$ satisfies the nonlinear difference equation,

$$T(z,u) = uT^2(\frac{z}{2},u) + (1-u)(1+z)e^{-z}.$$

This equation is a direct reflection of the problem specification. At $u = 1$, one has $T(z,1) = 1$, $F(z,1) = e^z$. The idea is thus to analyse $[z^n]F(z,u)$ by the saddle point method.

The saddle point analysis of F requires asymptotic information on $T(z,u)$ for $u = e^{it}$ (the original treatment of [**241**] is based on characteristic functions). There, the main idea is to "quais-linearize" the problem, setting

$$L(z,u) = \log T(z,u),$$

with u a parameter. This function satisfies the approximate relation $L(z,u) \approx 2L(z/2,u)$, and a bootstrapping argument shows that, in suitable regions of the complex plane, $L(z,u) = O(|z|)$, uniformly with respect to u. The function $L(z,u)$ is then expanded with respect to $u = e^{it}$ at $u = 1$, *i.e.*, $t = 0$, using a Taylor expansion, its companion integral representation, and the bootstrapping bounds. The moment-like quantities,

$$L_j(z) = \left. \frac{\partial^j}{\partial t^j} L(z, e^{it}) \right|_{t=0},$$

can be subjected to Mellin analysis for $j = 1, 2$ and bounded for $j \geq 3$. In this way, there results that

$$L(z, e^{it}) = L_1(z)t + \frac{1}{2}L_2(z)t^2 + O(zt^3),$$

uniformly. The Gaussian law under a Poisson model immediately results from the continuity theorem of characteristic functions. Under the original Bernoulli model, the Gaussian limit follows from a saddle point analysis of

$$F(z, e^{it}) = e^z e^{L(z,e^{it})}.$$

An even more delicate analysis has been carried out by Jacquet and Szpankowski in [**242**]. It is relative to path length in digital search trees and involves the formidable non-linear bivariate difference-differential equation

$$\frac{\partial}{\partial z}F(z,u) = F^2(\frac{z}{2},u).$$

IX. 9. Local limit laws

Under conditions similar to those of the Quasi-Powers Theorem, a cluster of conclusions may be drawn regarding densities of distributions and probabilities of large deviations from the mean. We examine here the occurrence of local limit laws, which corresponds to convergence of a discrete probability distribution to the *Gaussian density function* rather than convergence of distribution functions to the *Gaussian error* *function*, as we have seen so far. Such local laws hold very frequently, but their proofs require some sort of additional "smoothness" assumptions, either a combinatorial or analytic. Under assumptions of the Quasi-Powers Theorem, it is also possible to quantify precisely the exponential rate of decay for probabilities of rare events, far away from the center of the distribution. This section explores both aspects that fit well withing the general framework of quasi-powers. One aspects provides precise asymptotic information on values of the individual probabilities, especially near the mean; the other aspect quantifies the smallness of probabilities far away from the mean and, when conditions apply, it provides sharp quantitative versions of the concentration of distribution discussed at the beginning of this chapter.

So far, we have examined the occurrence of continuous limit laws in the sense of convergence of distribution functions. Thus, a standardized Y_n converges in distribution to Y, if

$$\mathbb{P}\{Y_n \leq x\} \to \mathbb{P}\{Y \leq x\}.$$

In the case of a Gaussian limit that arises from a sequence of discrete distributions of variables X_n with mean and variance μ_n, σ_n^2, such a property quantifies the probabilities over any nonempty interval scaled according to σ_n,

$$Pr\{\mu_n + a\sigma_n < X_n \leq \mu_n + b\sigma_n\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx + o(1), \tag{56}$$

for any a, b with $a < b$. From there, it is however in general not possible to draw information on any individual probability,

$$p_{n,k} = \mathbb{P}\{X_n = k\},$$

by differencing, since the error terms in (56) will usually hide any nontrivial asymptotic information on individual $p_{n,k}$.

On the other hand, numerical examination of discrete probability distributions reveals that the histograms of the $p_{n,k}$ often assume a bell-shape profile in the asymptotic limit. For instance Figure 11, borrowed from our book [**382**], displays the $p_{n,k}$ that correspond to the Eulerian numbers. For a given value of n, the maximum probability $p_{n,k}$ is seen to occur "in the middle", near the mean, and to obey an approximate law,

$$p_{2n,n} \approx \frac{1.35}{\sqrt{2n}},$$

for values near $n = 60$. The standard deviation of the distribution is otherwise known to be $\sim \sqrt{n/12}$. Thus, the we expect an approximate formula of the form

$$p_{n,n/2+x\sqrt{n/12}} \approx \frac{C}{\sqrt{n}} e^{-x^2/2},$$

for integral values of the argument $k = n/2 + x\sqrt{n/12}$, with some constant C about 1.35.

DEFINITION IX.4. *A sequence of discrete probability distributions, $p_{n,k} = \mathbb{P}\{X_n = k\}$, with mean μ_n and standard deviation σ_n is said to obey a* local limit law of the

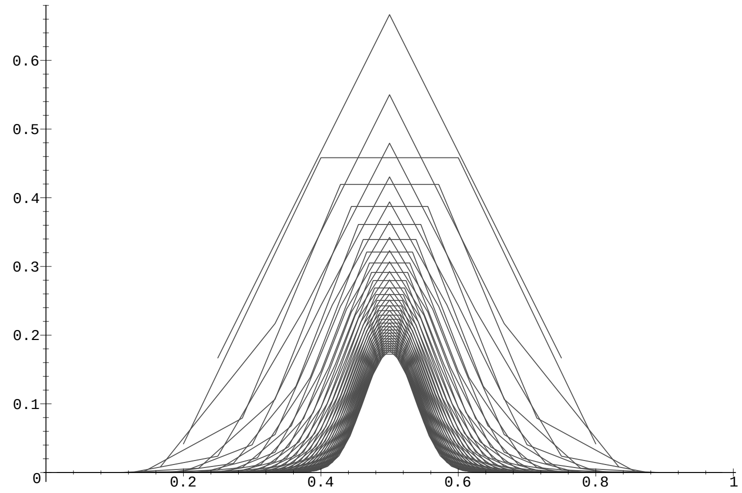


FIGURE 11. The histogram of the Eulerian distribution scaled to $(n+1)$ on the horizontal axis, for $n = 3\,..\,60$. The distribution is seen to quickly converge to a bell-shaped curve corresponding to the Gaussian density $e^{-x^2/2}/(2\pi)^{1/2}$.

Gaussian type *if, for some set* S *of real numbers, and a sequence* $\epsilon \to 0$,

$$\sup \left| \sigma_n p_{n,\lfloor \mu_n + x\sigma_n \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \le \epsilon_n.$$

The local limit law is said to hold on S *and the law is said to hold with* relative speed of convergence ϵ_n.

When such a local limit law exists, it usually holds on arbitrary bounded intervals of the real line.

THEOREM IX.13 (Local limit law). *Let* X_n *be a sequence of nonnegative discrete random variables with probability generating function* $p_n(u)$. *Assume that* uniformly *in an annulus,*

$$1 - \epsilon \le u \le 1 + \epsilon, \qquad \epsilon > 0$$

the PGFs satisfy

$$p_n(u) = A(u)\,(B(u))^{\beta_n} \left(1 + O(\frac{1}{\kappa_n})\right), \tag{57}$$

where $A(u), B(u)$ *are analytic in the annulus and* $A(1) = B(1) = 1$, $\mathfrak{v}(B(u)) = B''(1) + B'(1) - B'(1) \neq 0$. *Assume also that* $B(u)$ *attains uniquely in maximum on* $|u| = 1$ *at* $u = 1$*: for all* v*, with* $|v| = 1$ *and* $v \neq 1$*, one has* $|B(v)| < 1$.

Under these conditions, the distribution of X_n *satisfies a local limit law of the Gaussian type on arbitrary bounded intervals of the real line.*

Note that the mean and variance of X_n are given by Eq. (20).

PROOF. A direct application of the saddle point method, as developed in Chapter VIII. □

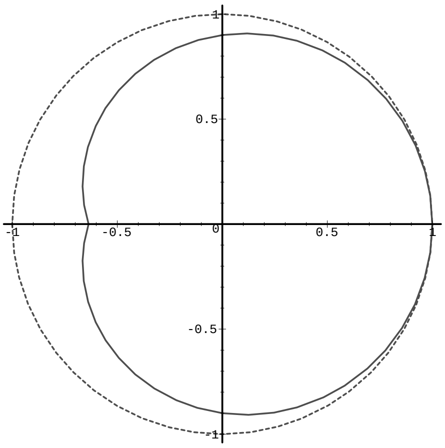


FIGURE 12. The values of the function $B(u)$ for the Eulerian distribution when $|u| = 1$, represented by a polar plot of $|B(e^{i\theta})|$ on the ray of angle θ (right). (The dashed contours represent the relevant parts of the unit circle, for comparison.) The maximum is uniquely attained at $u = 1$, where $B(1) = 1$. This entails a local limit law for the Eulerian distribution.

This theorem applies in particular to quasi-power expansions, whenever the dominant singularity $\rho(u)$, that is a perturbation of the dominant singularity ρ of the univariate problem, is analytic at all points of $|u| = 1$ and uniquely attains its minimum at $u = 1$.

EXAMPLE 28. *Local laws for sums of RV's.* The simplest application is to the binomial distribution, for which

$$B(u) = \frac{1+u}{2}.$$

In a precise technical sense, the local limit arises in the BGF,

$$F(z,u) = \frac{1}{1 - z(1+u)/2},$$

because the dominant singularity $\rho(u) = 2/(1+u)$ exists on the whole of the unit circle, $|u| = 1$, and it attains uniquely its minimum modulus at $u = 1$; accordingly, $B(u) = \rho(1)/\rho(u)$ is uniquely maximal at $u = 1$.

More generally, the theorem applies to any sum $S_n = T_1 + \cdots + T_n$ of independent, identically, random variables whose maximal span is equal to 1 and whose PGF is analytic on the unit circle. In that case, the BGF is

$$F(z,u) = \frac{1}{1 - zB(u)},$$

the PGF of S_n is a pure power,

$$p_n(u) = B(u)^n,$$

and the fact that the minimal span of the X_j is 1 entails that $B(u)$ attains uniquely its maximum at 1. Such cases have been known for a long time in probability theory. See Chapter 9 of [**201**]. END OF EXAMPLE 28. □

At this stage, it is worth pointing an example *not* leading to a local law. Consider the binomial distribution restricted to even values,

$$p_{n,2k} = \frac{2}{2^n}\binom{n}{2k}, \qquad p_{n,2k+1} = 0.$$

The BGF is

$$F(z,u) = \frac{1}{1 - z(1+u)/2} + \frac{1}{1 - z(1-u)/2} - 1.$$

This has two poles,

$$\rho_1(u) = \frac{2}{1+u}, \qquad \rho_2(u) = \frac{2}{1-u},$$

and it is clearly not true that a single one dominates throughout the domain $|u| = 1$. Accordingly, the PGF satisfies

$$p_n(u) = (1+u)^n + (1-u)^n,$$

and no quasi-power law, with a unique analytic $B(u)$, holds uniformly for u on the unit circle. In essence, a local limit law will be likely to hold when a PGF has a sharp peak near 1 and stays much smaller in modulus along the rest of the unit circle. In contrast, for the even binomial distribution, one has $p_n(1) = p_n(-1)$.

EXAMPLE 29. *Local law for the Eulerian distribution.* For Eulerian numbers, we have derived the approximate expression,

$$p_n(u) = B(u)^{-n-1} + O(2^{-n}),$$

when u is close enough to 1, with

$$B(u) = \rho(u)^{-1} = \frac{u-1}{\log u}.$$

The plot of the function $B(u)$ when u varies over $|u| = 1$ is then displayed in Fig. 12.

This case requires in fact a minor extension of Theorem IX.13 since the principal determination of the logarithm cannot be extended to the whole of the unit circle, in particular at $u = -1$. However, it is easily realized that the quasi-power expansion holds with the possible exception of a small segment of the integration contour near $u = -1$. However, there, the integrand is anyway exponentially smaller than on the rest of the contour, and the proof of Theorem IX.13 is easily adjusted to cover such case.

From this enhanced argument, there results that a local limit law of the Gaussian type holds for the Eulerian distribution on any compact subset of the real line. END OF EXAMPLE 29. □

With a similar care to be exercised regarding principal determinations and dominant singularities, many of our earlier analyses can be turned into local limit laws. What is needed is a dominant singularity $\rho(u)$ that yields the main asymptotic form of the PGF's on most of the unit disc and that achieves uniquely its minimum at 1, while the rest of the unit disc contributes negligibly. For instance, this covers the surjection distribution, for which

$$\rho(u) = \log(1 + u^{-1}), \qquad B(u) = \frac{\log 2}{\log(1+u^{-1})},$$

leaves in general Catalan trees, where

$$B(u) = \frac{(1+\sqrt{u})^2}{4},$$

or in binary Catalan trees.

The Stirling cycle distribution satisfies

$$p_n(u) = \frac{e^{(u-1)\log n}}{\Gamma(u)}\left(1 + O(\frac{1}{n})\right).$$

This approximation remains uniform as long as u avoids -1, but, there, $p_n(u)$ is small anyway (being $O(n^{-2})$), so that again an extended form of Theorem IX.13 applies and a local limit law holds. The same argument applies to node levels in quadtrees of Example 24.

$\triangleright$ **37.** *Peaks of distributions.* It is possible to analyse asymptotically in detail the values of the peak of the Eulerian and Stirling cycle distributions. (For the Eulerian distribution, see, e.g., the study of Lesieur and Nicolas [**289**].) $\triangleleft$

IX. 10. Large deviations

Moment inequalities constrain the shape of a distribution given its mean and variance. In particular, if $\sigma_n/\mu_n \to 1$, the concentration property holds. This property comes from Chebyshev's inequality according to which the probability of observing a value that deviates by more than x standard deviations from the mean is $O(x^{-2})$. Such general bounds, though sufficient to establish a concentration property, are much weaker than what holds under conditions of the quasi-power type, where the probabilities of deviation are in fact exponentially decreasing with in x.

Figure 13 displays the logarithms of the Eulerian distribution. As logarithms of probabilities are plotted, the distribution is seen to decay very rapidly away from the mean $\mu_n \sim n/2$. Consider for instance extreme cases. Clearly, there is a unique permutation that has a minimal number of rises, namely the fully sorted permutation with probability

$$p_{n,1} = \frac{1}{n!}.$$

In contrast, since $\mu_n \sim n/2$ and $\sigma_n^2 \sim n/12$, this extreme case is roughly at $x = \sqrt{3n}$ from the mean; thus, the Chebyshev inequalities only provides the very weak upper bound of $\sim \frac{1}{3n}$ for this extreme case. For $n = 40$, the Chebyshev upper bound on the probability is thus about 0.008 while the exact value $1/40!$ is of the order of 10^{-48}.

Extensions of the quasi-power framework are once more well-suited to prove such exponentially small tails, as we now explain. It turns out that the ubiquitous functions $\rho(u), B(u)$ are directly related to *large deviation* estimates. Such estimates nicely supplement the already known limit laws, either central or local.

DEFINITION IX.5. *A sequence of discrete random variables* $\{X_n\}$ *with* $p_{n,k} = \mathbb{P}\{X_n = k\}$*, satisfies a* local large deviation property *of type* $(\beta_n, W(x))$ *over the interval* $[x_0, x_1]$*, if for any* $x \in [x_0, x_1]$*,*

$$\frac{1}{\beta_n}\log p_{n,x\beta_n} \le W(x) + O(\beta_n^{-1}). \qquad (58)$$

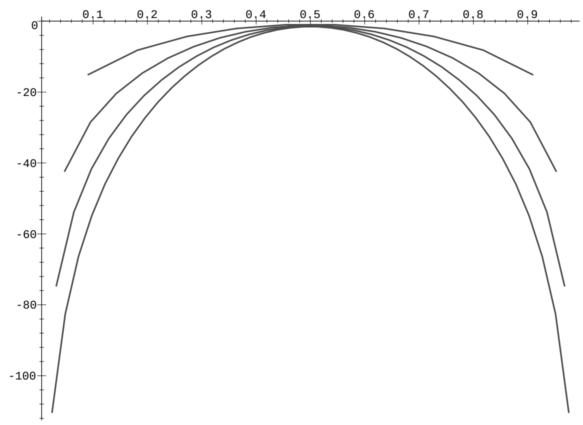


FIGURE 13. The quantities $\log p_{n,k}$ relative to the Eulerian numbers illustrate an extremely fast decay of the distribution away from the mean. Here, the diagrams corresponding to $n = 10, 20, 30, 40$ (top to bottom) are plotted. The common shape of the curves indicates a large deviation property.

The function $W(x)$ *is called a* large deviation function *and* β_n *is the* scaling factor.

The inequality (58) is *a priori* only meaningful if $x\beta_n$ is an integer, but it makes sense as well if it is understood that $p_{n,w} = 0$ for nonintegral values of w and $\log 0 = -\infty$. Of course, the large deviation property is nontrivial only when $W(x) \le 0$, with $W(x)$ not identically 0. A global (and marginally stronger) form of large deviations can also be defined when local probabilities are replaced by corresponding values of the cumulative distribution function. Large deviation theory is introduced nicely in the book of den Hollander [**101**].

THEOREM IX.14 (Quasi-powers, large deviations). *Consider a sequence of discrete random variables* $\{X_n\}$ *with PGF* $p_n(u)$. *Assume that there exist a functions* $A(u), B(u)$, *analytic in some interval* $[u_0, u_1]$ *with* $0 < u_0 < 1 < u_1$, *such that a quasi-power expansion holds,*

$$p_n(u) = A(u)B(u)^{\beta_n}\left(1 + O(\kappa_n^{-1}\right), \tag{59}$$

uniformly. Then the X_n *satisfy a large deviation property,*

$$\frac{1}{\beta_n}\log p_{n,x\beta_n} \le W(x) + O(\beta_n^{-1}), \tag{60}$$

where the large deviation function $W(x)$ *is given by*

$$W(x) = \min_{u\in[u_0,u_1]} \log\left(\frac{B(u)}{u^x}\right). \tag{61}$$

PROOF. The basic observation is that if $f(u) = \sum_n f_n u^n$ is an analytic function with nonnegative coefficients, then, for positive u,

$$f_k := [u^k]f(u) \le \frac{f(u)}{u^k} \le \min_{u>0}\frac{f(u)}{u^k}. \tag{62}$$

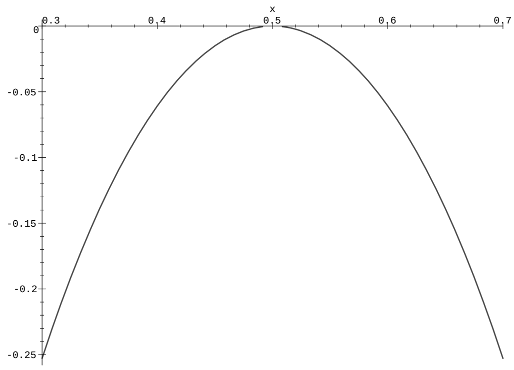


FIGURE 14. The large deviation function relative to the Eulerian distribution, for $u \in [0.3, 0.7]$.

The first inequality holds for any positive u in the disc of analyticity of $f(u)$; the second bound, with a similar condition, consists in taking the best possible value of u. See our earlier discussion of saddle point bounds.

The combination of the principle (62) applied to $f(u) = p_n(u)$, and of the assumption of the theorem (59) yields

$$\log p_{n,x\beta_n} \le \beta_n \min_{u\in[u_0,u_1]} \log\left(\frac{B(u)}{u^x}\right) + O(1).$$

Thus, a large deviation property holds with $W(x)$ given by (61). $\square$

In general, the function $W(x)$ is computable from $B(u)$ and its derivatives. The minimum is attained at either an end-point or a point such that

$$\frac{d}{du}(\log B(u) - x\log u) = 0.$$

Let $\eta(x)$ be a value of $u \in [u_0, u_1]$ that cancels this derivative. Thus, η is an inverse function of $uB'(u)/B(u)$,

$$\eta(x)\frac{B'(\eta(x))}{B(\eta(x)} = x.$$

Then, a large deviation function is

$$W(x) = \log B(\eta(x)) - x\log\eta(x). \tag{63}$$

$\triangleright$ **38.** Prove similar types of bounds for the cumulative quantities

$$P_{n,k} = \sum_{j\le k} p_{n,j}, \qquad Q_{n,k} = \sum_{j\ge k} p_{n,j}.$$

$\triangleleft$

EXAMPLE 30. *Large deviations for the Eulerian distribution.* In this case, the BGF has a unique dominant singularity for u with $\epsilon < u < 1/\epsilon$, and any $\epsilon > 0$. Thus, there is a quasi-power expansion with

$$B(u) = \frac{(u-1)}{\log u},$$

on any interval $[\epsilon, 1/\epsilon]$. Then $\eta(x)$ is computable as the inverse function of

$$h(u) = \frac{u}{u-1} - \frac{1}{\log u}.$$

This function increases from 0 to 1 as u increases from 0 to 1, so that the inverse function is well defined over any closed interval $[\epsilon, 1-\epsilon]$. The function $W(x)$ is then determined by (63); see Figure 14 for a plot of $W(x)$ that "explains" the data of Figure 13.

We find that

$$W(0.3) = W(0.7) = -0.252,\ W(0.4) = W(0.6) = -0.061,$$

$$W(0.45) = W(0.55) = -0.015,$$

and $W(0.5) = 0$, as expected. For instance, the probability of deviating by 20% from the mean value $\mu_n \sim 0.5n$ is approximately $\exp(-0.061\, n)$. For $n = 100$, this upper bound is about $e^{-6.07}$, while the exact value of the probability gives $p_{100,60} \doteq e^{-8.58}$. In the same vein, there is probability less than 10^{-6} of deviating by 10% from the mean, when $n = 1,000$; the upper bound becomes less than 10^{-65}, for $n = 10,000$, less than 10^{-653}, for $n = 100,000$. (These are the estimates stated at the very beginning of this chapter.) ..... END OF EXAMPLE 30. □

▷ **39.** *Quasi-Powers and large deviations.* Under the Quasi-Powers assumption, it is usually possible to convert the upperbound into an equality. This has been done by Hwang [**235, 236, 237**], who bases himself on a technique of Cramér. Roughly, by shiting the mean, the main Quasi-Powers Theorem can be applied at some $u = u_0$ with $u_0 \neq 1$. ◁

IX. 11. Non-Gaussian continuous limits

Previous sections of this chapter have developed two basic paradigms for bivariate asymptotics (see also Figure 3):

— a "minor" singularity perturbation mode leading to discrete laws,
— a "major" singularity perturbation mode leading to continuous laws.

However, in both cases, the assumption has been made so far that the collection of singular expansions parameterized by the auxiliary variable all belong to a common analytic class and exhibit no sharp discontinuity when the secondary parameter traverses the value $u = 1$. In this section we briefly explore by means of examples the way discontinuities in singular behaviour induce no-Gaussian laws (Subsection IX. 11.1), then conclude with a fairly general discussion of the critical composition schema (Subsection IX. 11.2), thereby completing the classification of analytic composition schemes. The discontuities observed in the cases discussed here are reminiscent of what is known as phase transition phenomena in statistical physics, and we find it suggestive to borrow this terminology here.

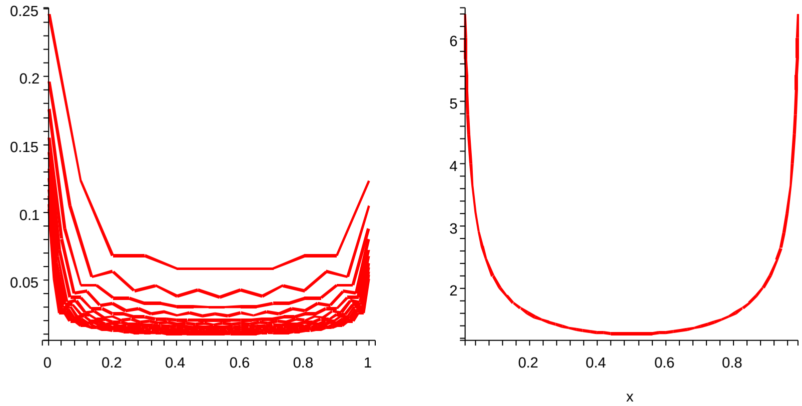


FIGURE 15. Histograms of the distribution of the maximum of a random walk for $n = 10\,..\,60$ (left) and the density of the arcsine law (right).

IX. 11.1. Phase transition diagrams. Perhaps the simplest case of discontinuity in singular behaviour is the already discussed BGF,

$$F(z,u) = \frac{1}{(1-z)(1-zu)},$$

where u records the number of a's in a random word of $a^\star b^\star$. The limit law is clearly the continuous uniform distribution over the interval $[0,1]$. From the point of view of the singular structure of $F(z,u)$, as a function of z, three distinct cases arise depending on the values of u:

- $u < 1$: simple pole at $\rho(u) = 1$;
- $u = 1$: double pole at $\rho(1) = 1$;
- $u > 1$: simple pole at $\rho(u) = 1/u$.

Thus both the singularity location at $\rho(u)$ and the singular exponent $\alpha(u)$ experience a nonanalytic transition at $u = 1$. This arises from a "confluence" of two singular terms when $u = 1$.

To visualize such cases, it is useful to introduce a simplified diagram representation, called a *phase transition diagram* and defined as follows. Write $Z = \rho(u) - z$ and reduce the singular expansion to its dominant singular term $Z^{\alpha(u)}$. Then, the diagram representing $F(z,u)$ above is

$u = 1 - \epsilon$	$u = 1$	$u = 1 + \epsilon$
$\rho(u) = 1$	$\rho(1) = 1$	$\rho(u) = 1/u$
Z^{-1}	Z^{-2}	Z^{-1}

A complete classification of such confluences and discontinuities is still lacking (see however Marianne Durand's thesis [**117**] for interesting fragments), and is perhaps beyond reach given the vast diversity of situations encountered in a combinatorialist's practice.

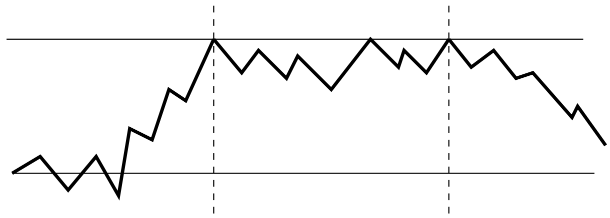
EXAMPLE 31. *Arcsine law for unbiased random walks.* This problem is studied in detail by Feller [**133**, p. 94] who notes: "Contrary to intuition, the maximum accumulated gain is much more likely to occur towards the very beginning or the very end of a coin-tossing game than somewhere in the middle." See Figure 15. In fact, if X_n is the time of the first occurrence of the maximum in a random game (walk with ± 1 steps) of duration n, one has

$$\mathbb{P}\{X_n < xn\} \sim \frac{2}{\pi} \arcsin \sqrt{x},$$

a distribution function with density

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

The BGF results from the standard decomposition of positive walks. Roughly, there is a sequence of steps ascending to the (nonnegative) maximum accompanied by "arches" (the left factor) followed by an excursion below than back to the maximum, followed by a sequence of descending steps with their companion arches. This translates directly into an equation satisfied by the BGF $F(z,u)$ of the location of the first maximum.



$$F(z,u) \quad = \quad \frac{1}{1-zuD(zu)} \quad \cdot \quad D(z) \quad \cdot \quad \frac{1}{1-zD(z)}, \tag{64}$$

which involves the GF of gambler's ruin sequences (Example 6),

$$D(z) = \frac{1-\sqrt{1-4z^2}}{2z}.$$

In such a simple case, explicit expressions are available from (64), as it suffices to expand first with respect to u, then to z. We obtain in this way the ultra-classical result:

PROPOSITION IX.15 (Arc sine law). *Set* $u_{2\nu} := 2^{-2\nu}\binom{2\nu}{\nu}$. *The probability that the first maximum in a random walk of length* $n = 2\nu$ *occurs at* $k = 2\rho$ *or* $k = 2\rho + 1$ *is* $\frac{1}{2}u_{2\rho}u_{2\nu-2\rho}$, *for* $0 < k < 2\nu$. *For any* $x \in (0,1)$, *the position* T_n *of the first maximum satisfies*

$$\lim_{n\to\infty} \mathbb{P}_n(X_n < xn) = \frac{2}{\pi} \arcsin \sqrt{x}.$$

(The asymptotic form reflects by summation that of $u_{2\nu}$ since $u_{2\nu} \sim (\pi n u)^{-1/2}$.)

It is instructive to compare this to the way singularities evolve as u crosses the value 1. The dominant positive singularity is at $\rho(u) = 1/2$ if $u < 1$, while $\rho(u) = 1/(2u)$, if $u > 1$. Local expansions show that, with $c_{<}(u), c_{(}u) >$ two computable functions, there holds:

$$F(z,u) \sim c_{<}(u)\frac{1}{\sqrt{1-2z}}, \qquad F(z,u) \sim c_{>}(u)\frac{1}{\sqrt{1-2z}}.$$

Naturally, at $u = 1$, all words are counted and

$$F(z,1) = \frac{1}{1-2z}.$$

Thus, the corresponding phase transition diagram is (see Figure 16):

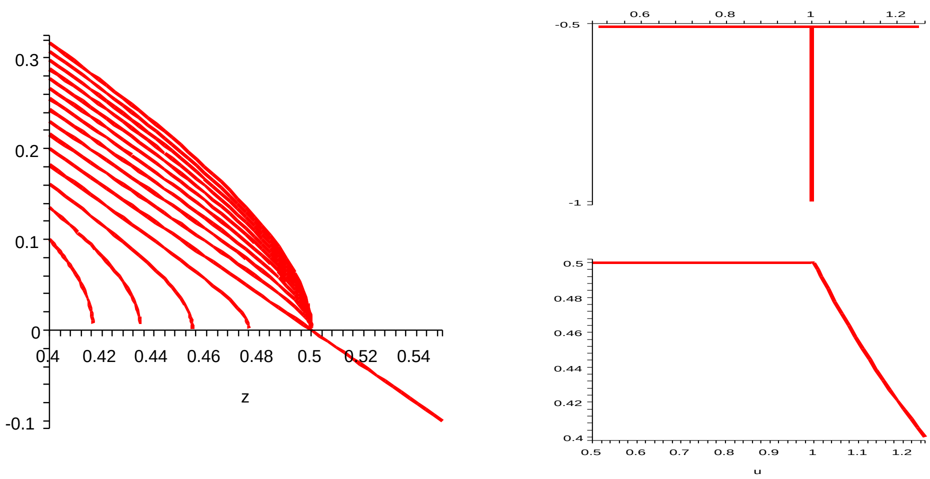


FIGURE 16. A plot of $1/F(z,u)$ for $z \in [0.4, 0.55]$ when u is assigned values between $\frac{1}{2}$ and $\frac{5}{4}$ (left). The exponent function $\alpha(u)$ and the singular value $\rho(u)$ for $u \in [1/2, 3/2]$ (right).

$u = 1 - \epsilon$	$u = 1$	$u = 1 + \epsilon$
$\rho(u) = \frac{1}{2}$	$\rho(1) = \frac{1}{2}$	$\rho(u) = \frac{1}{2u}$
$Z^{-1/2}$	Z^{-1}	$Z^{-1/2}$

(Negative singularities have a smaller weight and may be discarded.)
END OF EXAMPLE 31. □

In this particular case, elementary combinatorics yields the arcsine distribution without the need of a recourse to singularities. The point to be made here is that the arcsine law could be expected when a similar phase transition diagram occurs. There is of course universality in this singular view of the arcsine law, which can be extended to walks with zero drift (Chapter VII). This kind of universality is a parallel to the universality of Brownian motion, which is otherwise familiar to probabilists.

▷ **40.** *Number of maxima and other stories.* The construction underlying (64) also serves to analyse; (i) the number of times the maximum is attained. (ii) the difference between the maximum and the final altitude of the walk; (iii) the duration of the period following the last occurrence of the maximum. ◁

EXAMPLE 32. *Path length in trees.* A final example is the distribution of path length in trees, which has been studied by Louchard, Takacs and others [**297, 296, 404, 405**]. The distribution is known *not* to be Gaussian as results from computation of the first few moments. In the case of general Catalan trees, the analysis reduces to that of the functional equation

$$F(z,u) = \frac{1}{1 - zF(zu,u)}.$$

This defines $F(z,u)$ as a formal continued fraction, which suggests setting (*cf* Chapters III and V as well as our discussion of coin fountains and polyomino models)

$$F(z,u) = \frac{A(z)}{B(z)},$$

the variable u being viewed as a parameter. From the basic functional equation, there results

$$A(z) = B(zu), \qquad B(z) = B(zu) - zB(zu^2).$$

The functional equation for B may now be solved by indeterminate coefficients:

$$B(z) = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{u^{n(n-1)} z^n}{(1-u)(1-u^2)\cdots(1-u^n)}.$$

Because of the quadratic exponents involved, the functions $B(z)$ and $F(z,u)$ have radius of convergence 0 when $u > 1$, and are thus nonanalytic. In contrast, when $u < 1$, then $B(z,u)$ is an entire function of z, so that $F(z,u)$ is meromorphic in z. Hence the singularity diagram:

$u = 1 - \epsilon$	$u = 1$	$u = 1 + \epsilon$
$\rho(u) > \frac{1}{4}$	$\rho(1) = \frac{1}{4}$	$\rho(u) = 0$
Z^{-1}	$Z^{1/2}$	—

The limit law is the *Airy area distribution*, that is related to the Airy function [**296, 297, 404, 405**]. By an analytical *tour de force*, Prellberg [**353**] has developed a method based on cintegral representations and oalescing saddle points (Chapter VIII) that permits us to extract the phase transition diagram above, together with precise uniform asymptotic expansions. As similar problems occur in relation to connectivity of random graphs [**172**], future years should see more applications of Prellberg's method. .... END OF EXAMPLE 32. □

IX. 11.2. Semi-large powers, critical compositions, and stable laws. We conclude this section by a discussion of critical compositions that typically involve confluences of singularities and lead to a general class of continuous distributions closely related to *stable laws* of probability theory. We start with an example where everything is explicit, that of zero contacts in random bridges, then state a general theorem on "semi-large" powers of functions of singularity analysis type, and finally discuss combinatorial applications.

EXAMPLE 33. *Zero-contacts in bridges.* Consider once more fluctuations in coin tossings, and specifically bridges, corresponding to a conditioning of the game by the fact that the final gain is 0 (negative capitals are allowed). These are sequences of arbitrary positive or negative "arches", and the number of arches in a bridge is exactly equal to the number of intermedaite steps at which the capital is 0. From the arch decomposition, theer results that the ordinary BGF of bridges with z marking length and u marking zero-contacts is

$$B(z,u) = \frac{1}{1 - 2uz^2 D(z)}.$$

Analysing this function is conveniently done by introducing

$$F(z,u) \equiv B\left(\frac{1}{2}\sqrt{z}, u\right) = \frac{1}{1 - u(1 - \sqrt{1-z})}.$$

The phase transition diagram is then easily found to be:

$u = 1-\epsilon$	$u=1$	$u = 1+\epsilon$
$\rho(u)=1$	$\rho(1)=1$	$\rho(u) = 1-(1-u^{-1})^2$
$Z^{1/2}$	$Z^{-1/2}$	Z^{-1}

Thus, there are discontinuities, both in the location of the singularity and the exponent. But these are of a type different from what gave rise to the arcsine law of random walks.

The problem of the limit law is here easily solved since explicit expressions are provided by the Lagrange Inversion Theorem. One finds:

$$\begin{array}{rcl} [u^k][z^n]F(z,u) & = & [z^n]\left(1-\sqrt{1-z}\right)^k \\ & = & \dfrac{k}{n}[w^{n-k}](2-w)^{-n} \quad = \quad 2^{k-2n}\dfrac{k}{n}\dbinom{2n-k-1}{n-1}. \end{array}$$

Then Stirling's formula provides:

PROPOSITION IX.16. *The number X_n of zero-contacts of a random bridge of size $2n$ satisfies, as $\to\infty$ the local limit law,*

$$\lim_{n\to\infty}\mathbb{P}(X_n = x\sqrt{n}) = \frac{x}{2\sqrt{n}}e^{-x^2/4},$$

for x in any compact set of $[0,+\infty[$.

A random variable with density and distribution function given by

$$r(x) = \frac{x}{2}e^{-x^2/4}, \qquad R(x) = 1-e^{x^2/4}, \tag{65}$$

is called a Rayleigh law. Thus the number of zero contacts obeys a Rayleigh law in the asymptotic limit. . . . . END OF EXAMPLE 33. □

▷ **41.** *Cyclic points in mappings.* The number of cyclic points in mappings has exponential BGF $(1-uT(z))^{-1}$, with T the Cayley tree function. The singularity diagram is of the same form as in Example 33. Explicit forms are available by Lagrange inversion, and the limit law is again Rayleigh. (Note: This has been vastly generalized by Drmota and Soria [**113, 114**].) ◁

Both Example 33 and Note 41 exemplify the situation of an analytic composition scheme of the form $(1-uf(z))^{-1}$ which is critical, since in each case f assumes value 1 at its singularity. Both can be treated elementarily since they involve powers that are amenable to Lagrange inversion, eventually resulting in a Rayleigh law. As we now explain, there is a family of functions that appear to play a universal rôle in problems sharing such singular types. What follows is taken from an article by Banderier *et al.* [**22**].

We first introduce a function G that otherwise naturally surfaces in the study of *stable*[9] distributions in probability theory. For any parameter $\lambda \in (0,2)$, define the entire function

$$G(x,\lambda) := \begin{cases} \dfrac{1}{\pi}\displaystyle\sum_{k\ge1}(-1)^{k-1}x^k\frac{\Gamma(1+\lambda k)}{\Gamma(1+k)}\ \sin(\pi k\lambda) & (0<\lambda<1) \\ \dfrac{1}{\pi x}\displaystyle\sum_{k\ge1}(-1)^{k-1}x^k\frac{\Gamma(1+k/\lambda)}{\Gamma(1+k)}\ \sin(\pi k/\lambda) & (1<\lambda<2) \end{cases} \tag{66}$$

---

[9] In probability theory, stable laws are defined as the possible limit laws of sums of independent identically distributed random variables. The function G is a trivial variant of the density of the stable law of index λ; see Feller's book [**134**, p. 581–583]. Valuable informations regarding stable laws may be found in the books by Breiman [**67**, Sec. 9.8], Durett [**118**, Sec. 2.7], and Zolotarev [**452**].

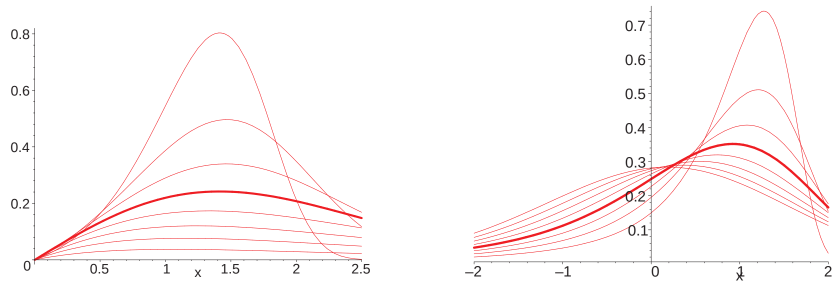


FIGURE 17. The G-functions for $\lambda = 0.1\,..\,0.8$ (left; from bottom to top) and for $\lambda = 1.2\,..\,1.9$ (right; from top to bottom); the thicker curves represent the Rayleigh law (left, $\lambda = \frac{1}{2}$) and the Airy law (right, $\lambda = \frac{3}{2}$).

The function $G(x; \frac{1}{2})$ is a normalized variant of the Rayleigh distribution (65). The function $G(x; \frac{3}{2})$ constitutes the density of the "Airy map" distribution found in random maps as well as in other colascence phenomena and discussed in detail below, see (73).

THEOREM IX.15 (Semi-large powers). *The coefficient of z^n in a power $H(z)^k$ of a Δ-continuable function $H(z)$ with singular exponent λ admits the following asymptotic estimates.*

(i) For $0 < \lambda < 1$, that is, $H(z) = \sigma - h_\lambda(1 - z/\rho)^\lambda + O(1 - z/\rho)$, and when $k = xn^\lambda$, with x in any compact subinterval of $(0, +\infty)$, there holds

$$[z^n]H^k(z) \;\sim\; \sigma^k \rho^{-n} \frac{1}{n} G\left(\frac{xh_\lambda}{\sigma}, \lambda\right). \tag{67}$$

(ii) For $1 < \lambda < 2$, that is, $H(z) = \sigma - h_1(1 - z/\rho) + h_\lambda(1 - z/\rho)^\lambda + O((1 - z/\rho)^2)$, when $k = \frac{\sigma}{h_1}n + xn^{1/\lambda}$, with x in any compact subinterval of $(-\infty, +\infty)$, there holds

$$[z^n]H^k(z) \;\sim\; \sigma^k \rho^{-n} \frac{1}{n^{1/\lambda}} (h_1/h_\lambda)^{1/\lambda} G\left(\frac{xh_1^{1+1/\lambda}}{\sigma h_\lambda^{1/\lambda}}, \lambda\right). \tag{68}$$

(iii) For $\lambda > 2$, a Gaussian approximation holds. In particular, for $2 < \lambda < 3$, that is, $H(z) = \sigma - h_1(1 - z/\rho) + h_2(1 - z/\rho)^2 - h_\lambda(1 - z/\rho)^\lambda + O((1 - z/\rho)^3)$, when $k = \frac{\sigma}{h_1}n + x\sqrt{n}$, with x in any compact subinterval of $(-\infty, +\infty)$, there holds

$$[z^n]H^k(z) \;\sim\; \sigma^k \rho^{-n} \frac{1}{\sqrt{n}} \frac{\sigma/h_1}{a\sqrt{2\pi}}\, e^{-x^2/2a^2} \qquad \textit{with } a = 2\left(\tfrac{h_2}{h_1} - \tfrac{h_1}{2\sigma}\right)\sigma^2/h_1^2. \tag{69}$$

The term "semi-large" refers to the fact that the exponents k in case (i) are of the form $O(n^\theta)$ for some $\theta < 1$ chosen in accordance with the region where an "interesting" renormalization takes place and dependent on each particular singular exponent.

When the interesting region reaches the $O(n)$ range in case (iii), the analysis of large powers, as detailed in Chapter IX, starts to apply and Gaussian forms results.

PROOF. The proofs are somewhat similar to the basic ones in singularity analysis, but they require a suitable adjustment of the geometry of the Hankel contour and of the corresponding scaling.

Case (i). A classical Hankel contour, with the change of variable $z = \rho(1-t/n)$, yields the approximation

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{2i\pi n} \int e^{t-\frac{h_\lambda x}{\sigma}t^\lambda}\, dt$$

The integral is then simply estimated by expanding $\exp(-\frac{h_\lambda x}{\sigma}t^\lambda)$ and integrating termwise

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{n} \sum_{k\geq 1} \frac{(-x)^k}{k!} \left(\frac{h_\lambda}{\sigma}\right)^k \frac{1}{\Gamma(-\lambda k)}, \tag{70}$$

which is equivalent to Equation (67), by virtue of the complement formula for the Gamma function.

Case (ii). When $1 < \lambda < 2$, the contour of integration in the z-plane is chosen to be a positively oriented loop, made of two rays of angle $\pi/(2\lambda)$ and $-\pi/(2\lambda)$ that intersect on the real axis at a distance $1/n^{1/\lambda}$ left of the singularity. The coefficient integral of H^k is rescaled by setting $z = \rho(1 - t/n^{1/\lambda})$, and one has

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{2i\pi n^{1/\lambda}} \int e^{\frac{h_\lambda}{h_1}t^\lambda} e^{-\frac{xh_1}{\sigma}t}\, dt.$$

There, the contour of integration in the t-plane comprises two rays of angle π/λ and $-\pi/\lambda$, intersecting at -1. Setting $u = t^\lambda h_\lambda/h_1$, the contour transforms into a classical Hankel contour, starting from $-\infty$ over the real axis, winding about the origin, and returning to $-\infty$. So, with $\alpha = 1/\lambda$, one has

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{2i\pi n^\alpha}\, \alpha \left(\frac{h_1}{h_\lambda}\right)^\alpha \int e^u\, e^{-\frac{xh_1^{\alpha+1}}{\sigma h_\lambda^\alpha}u^\alpha}\, u^{\alpha-1}\, du\,.$$

Expanding the exponential, integrating termwise, and appealing to the complement formula for the Gamma function finally reduces this last form to (68).

Case (iii). This case is only included here for comparison purposes, but, as recalled before the proof, it is essentially implied by the developments of Chapter IX based on the saddle point method. When $2 < \lambda < 3$, the angle ϕ of the contour of integration in the z–plane is chosen to be $\pi/2$, and the scaling is $\sqrt{n}$: under the change of variable $z = \rho(1 - t/\sqrt{n})$, the contour is transformed into two rays of angle $\pi/2$ and $-\pi/2$ (*i.e.*, a vertical line), intersecting at -1, and

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{2i\pi\sqrt{n}} \int e^{pt^2 - \frac{h_1 x}{\sigma}t}\, dt\,,$$

with $p = \frac{h_2}{h_1} - \frac{h_1}{2\sigma}$. Complementing the square, and letting $u = t - \frac{h_1 x}{2p\sigma}$, we get

$$[z^n]H^k(z) \sim -\frac{\sigma^k \rho^{-n}}{2i\pi\sqrt{n}} e^{-\frac{h_1^2}{4p\sigma^2}x^2} \int e^{pu^2}\, du\,,$$

which gives Equation (69). By similar means, such a Gaussian approximation can be shown to hold for any non-integral singular exponent $\lambda > 2$. $\square$

$\triangleright$ **42.** *Zipf laws.* Zipf's law, named after the Harvard linguistic professor George Kingsley Zipf (1902–1950), is the observation that, in a language like English, the frequency with which a word occurs is roughly inversely proportional to its rank—the kth most frequent word has frequency proportional to $1/k$. The *generalized Zipf distribution* of parameter $\alpha > 1$ is the law of a variable Z such that

$$\mathbb{P}(Z = k) = \frac{1}{\zeta(\alpha)} \frac{1}{k^\alpha}.$$

It has infinite mean for $\alpha \leq 2$ and infinite variance for $\alpha \leq 3$. It was proved in Chapter VI that polylogarithms are amenable to singularity analysis. Consequently, the sum of a large number of independent Zipf variables satisfies a local limit law of the stable type with index $\alpha - 1$ $(\alpha \neq 2)$. $\triangleleft$

EXAMPLE 34. *Mean level profiles of trees.* Consider the depth of a random node in a random tree taken from a simple variety $\mathcal{Y}$ that satisfies the usual analytic assumptions of Chapter VII. The problem of quantifying this distribution is equivalent to that of determining the *mean* level profile, that is the sequence of numbers $M_{n,k}$ representing the mean number of nodes at distance k from the root. (The probability that a random node lies at level k is then $M_{n,k}/n$.) The first few levels have been characterized in Chapter VII, and the analysis of that chapter can now be completed thanks to Theorem IX.15. The problem was solved by Meir and Moon [**312**] in an important article that launched the analytic study of simple varieties of trees. As usual, we let $\phi(w)$ be the generator of the simple variety $\mathcal{Y}$, with $Y(z)$ satisfying $Y = z\phi(Y)$, and we designate by τ the positive root of the characteristic equation:

$$\tau\phi'(\tau) - \phi(\tau) = 0.$$

It is known from Chapter VII that the GF $Y(z)$ has a square root singularity at $\rho = \tau/\phi(\tau)$. We also assume aperiodicity of ϕ. Then Meir and Moon's major result (Theorem 4.3 of [**312**]) is as follows

PROPOSITION IX.17 (Mean level profiles). *The mean profile of a large tree in a simple variety obeys a Rayleigh law in the asymptotic limit: for $k/\sqrt{n}$ in any bounded interval of $\mathbb{R}_{\geq 0}$, the mean number of nodes at altitude k satisfies asymptotically*

$$M_{n,k} \sim Ake^{Ak^2/(2n)},$$

where $A = \tau\phi''(\tau)$.

(Note: Meir and Moon base their analysis on a Lagrangean change of variable and on the saddle point method.)

PROOF. For each k, define $Y_k(z,u)$ to be the BGF with u marking the number of nodes at depth k. Then, the root decomposition of trees translates into the recurrence:

$$Y_k(z,u) = z\phi(Y_{k-1}(z,u)), \qquad Y_0(z,u) = zu\phi(Y(z)) = uY(z).$$

By construction, we have

$$M_{n,k} = \frac{1}{Y_n}[z^n]\left(\frac{\partial}{\partial u} Y_k(z,u)\right)_{u=1}.$$

On the other hand, the fundamental recurrence yields

$$\left(\frac{\partial}{\partial u} Y_k(z,u)\right)_{u=1} = \left(z\phi'(Y(z))\right)^k Y(z).$$

Now, $\phi'(Y)$ has, like Y, a square root singularity. The semi-large powers theorem applies with $\lambda = \frac{1}{2}$, and the result follows. $\square$ The same method of gives access to the variance of the number of nodes at any depth k. The variance of the altitude of a random node is also easily computed [**312**]. .... END OF EXAMPLE 34. $\square$

$\triangleright$ **43.** *The number of cyclic points in mappings.* In the basic case of random mapping, we are dealing with $F(z,u) = (1 - uT(z))^{-1}$, and a Rayleigh law holds. This extends to the number of cyclic points in a simple variety of mappings (e.g., mappings defined by a finite constraint on degrees). $\triangleleft$

$\triangleright$ **44.** *The width of trees.* The expectation of the width W of a tree in a simple variety satisfies

$$C_1\sqrt{n} \le \mathbb{E}_{\mathcal{Y}_n}(W) \le C_2\sqrt{n \log n},$$

for some $C_1, C_2 > 0$. This is due to Odlyzko and Wilf [**332**] in 1987. (Better bounds are now known, since $W_n/\sqrt{n}$ has been later recognized to be related to Brownian excursion. In particular, the expected width is $\sim c\sqrt{n}$.) $\triangleleft$

The results of Theorem IX.15 provide in addition useful information on composition schemas of the form

$$M(z,u) = C(uH(z)),$$

provided C and H are algebraic-logarithmic in the sense above. Combinatorially, this represents a substitution between structures, $\mathcal{M} = \mathcal{C} \circ \mathcal{H}$, and the coefficient $[z^n u^k]M(z,u)$ counts the number of $\mathcal{M}$-structures of size n whose $\mathcal{C}$-componnet, also called *core* in what follows, has size k. Then the probability distribution of core-size X_n in $\mathcal{M}$-structures of size n is given by

$$\mathbb{P}(X_n = k) = \frac{[z^k]C(z)}{[z^n]C(H(z))}\,[z^n]H(z)^k.$$

The case where the schema is critical, in the sense that $H(r_H) = r_G$ with r_H, r_C the radii of convergence of H, G, follows as a direct consequence of Theorem IX.15. What comes out is the following informally stated general principle (details would closely mimic the statement of Theorem IX.15 and are omitted).

PROPOSITION IX.18 (Critical compositions). *In a composition schema $G(uH(z))$ where H and G have singular exponents λ, λ' with $\lambda' \le \lambda$:*

(i) *for $0 < \lambda < 1$, the normalized core-size X_n/n^λ is spread over $(0, +\infty)$ and it satisfies a local limit law whose density involves the stable law of index λ; in particular, $\lambda = \frac{1}{2}$ corresponds to a Rayleigh law.*

(ii) *for $1 < \lambda < 2$, the distribution of X_n is bimodal and the "large region" $X_n = cn + xn^{1/\lambda}$ leads to a stable law of index λ;*

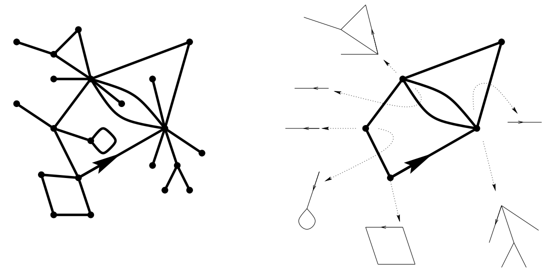
(iii) *for $2 < \lambda$, the standardized version of X_n admits a local limit law that is of Gaussian type.*

Similar phenomena occur when $\lambda' > \lambda$, but with a greater preponderance of the "small" region. Many instances have already appeared scattered in the literature. especially in connection with rooted trees. For instance, this proposition explains well the occurrence of the Rayleigh law ($\lambda = \frac{1}{2}$) as the distribution of cyclic points in random mappings and of zero-contacts in random bridges. The case $\lambda = 3/2$ appears in forests of unrooted trees (see the discussion in Chapter VIII for a complementary approach based on coalescing saddle points) and it is ubiquitous in planar maps, as attested by the article of Banderier *et al.* on which this subsection is largely based [**22**]. We detail one of the cases in the following example, which explains the meaning of the term "large region" in Proposition IX.18.

EXAMPLE 35. *Biconnected cores of planar maps.* The OGF of rooted planar maps, with size determined by the number of *edges*, is by Chapter VII,

$$M(z) = -\frac{1}{54z^2}\left(1 - 18z - (1-12z)^{3/2}\right), \tag{71}$$

with a characteristic $\frac{3}{2}$ exponent. Define a separating vertex or *articulation point* in a map to be a vertex whose removal disconnects the graph. Let $\mathcal{C}$ denote the class of nonseparable maps, that is, maps without an articulation point (also known as biconnected maps). Starting from the root edge, any map decomposes into a nonseparable map, called the "core" on which are grafted arbitrary maps, as illustrated by the following diagram:



There results the equation:

$$M(z) = C(H(z)), \qquad H(z) = z(1+M(z))^2. \tag{72}$$

This gives in passing the OGF of nonseparable maps as the algebraic function of degree 3 specified implicitly by the equation

$$C^3 + 2C^2 + (1-18z)C + 27z^2 - 2z = 0,$$

with expansion at the origin (*EIS* **A000139**):

$$C(z) = 2\,z + z^2 + 2\,z^3 + 6\,z^4 + 22\,z^5 + 91\,z^6 + \cdots, \quad C_k = 2\frac{(3k)!}{(k+1)!(2k+1)!}.$$

(The closed form results from a Lagrangean parameterization.) Now the singularity of C is also of the $Z^{3/2}$ type as seen by inversion of (72) or from the Newton diagram attached to the cubic equation. We find in particular

$$C(z) = \frac{1}{3} - \frac{4}{9}(1-27z/4) + \frac{8\sqrt{3}}{81}(1-27z/4)^{3/2} + O((1-27z/4)^2),$$

which is reflected by the asymptotic estimate,

$$C_k \sim \frac{2}{27}\frac{\sqrt{3}}{\pi}\left(\frac{17}{4}\right)^k k^{-5/2}.$$

The parameter considered here is the distribution of the size X_n of the core (containing the root) in a random map of size n. The composition relation is $\mathcal{M} = \mathcal{C} \circ H$, where $\mathcal{H} = \mathcal{Z}(\mathbf{1}+\mathcal{M})^2$. The BGF is thus $M(z,u) = C(uH(z))$ where the composition $C \circ H$ is of the singular type $Z^{3/2} \circ Z^{3/2}$. What is peculiar here is the "bimodal" caracter of the distribution of core-size (see Figure 18 borrowed from [**22**]), which we now detail.

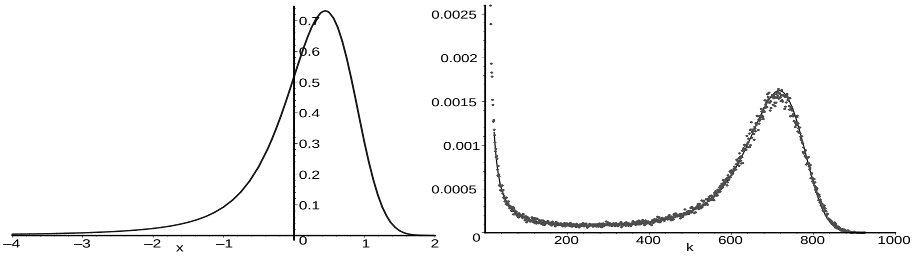


FIGURE 18. Left: The standard Airy distribution. Right: Observed frequencies of core-sizes $k \in [20; 1000]$ in 50,000 random maps of size 2,000, showing the bimodal character of the distribution.

First straight singularity analysis shows that, for *fixed* k,

$$\mathbb{P}(X_n = k) = C_k \frac{[z^n]H(z)^k}{M_n} \underset{n\to\infty}{\sim} kC_k h_0^{k-1},$$

where $h_0 = \frac{4}{27}$ is the value of $H(z)$ at its singularity. In other words, there is local convergence of the probabilities to a fixed *discrete* law. The estimate above can be proved to remain uniform as long as k tends to infinity sufficiently slowly. We shall call this the "small range" of k values. Now, summing the probabilities associated to this small range gives the value $C(h_0) = \frac{1}{3}$. Thus, *one-third of the probability mass of core-size arises from the small range, where a discrete limit law is observed.*

The other part of the distribution constitutes the "large range" to which Theorem IX.15 applies. This contains asymptotically $\frac{2}{3}$ of the probability mass of the distribution of X_n. In that case, the limit law is given by a $G(x; \frac{3}{2})$ law, also known as "map Airy" law and one finds for $k = \frac{1}{3}n + xn^{2/3}$, the continuous local limit:

$$(73) \qquad \mathbb{P}(X_n) \sim \frac{1}{3}\mathcal{A}(\frac{3}{4}2^{2/3}x), \qquad \mathcal{A}(x) = 2e^{-2x^2/3}\left(x\,\mathrm{Ai}(x^2) - \mathrm{Ai}'(x^2)\right).$$

There $\mathrm{Ai}(x)$ is the Airy function , and $\mathcal{A}(x)$ defines the map Airy distribution displayed in Figure 18, a variant of the stable law of index $\frac{3}{2}$. . . . . END OF EXAMPLE 35. □

The bimodal character of the law can now be bettler understood following [**22**]. A random maps decomposes completely into biconnected components and the largest biconnected component has, with high probability, a size that is $O(n)$. There are also a large number ($O(n)$) "dangling" biconnected components. In a rooted map, the root is in a sense placed "at random". Then, with a fixed probability is either lies in the large compoent (in which case, the distribution of that large component is observed, this is the continuous part of the distribution given by the Airy map law), or else one of the small components is picked up by the root (this is the discrete part of the distribution).

▷ **45.** *Critical cycles.* The theory adapts to logarithmic factors. For instance the critical composition $F(z,u) = -\log(1 - ug(z))$ leads to developments similar to those of the critical sequence. In this way, it becomes possible for instance to analyse the number of cyclic points in a random connected mapping. ◁

$\triangleright$ **46.** *The base of supertrees.* Supertrees defined in Chapter VI are trees rooted on trees. Here we consider the bicoloured variant $\mathcal{K} = \mathcal{G}(2\mathcal{Z}\mathcal{G})$, with $\mathcal{G}$ the class of general Catalan trees. Then, the law of the external $\mathcal{G}$-component is related to a stable law of index $\frac{1}{4}$. $\triangleleft$

IX. 12. Multivariate limit laws

There exist natural extensions of continuity theorems, both for PGFs and for integral transforms. Consider for instance the joint distribution of the numbers χ_1, χ_2 of singletons and doubletons in random permutations. Then, the parameter $\chi = (\chi_1, \chi_2)$ has a trivariate EGF

$$F(z, u_1, u_2) = \frac{\exp((u_1 - 1)z + (u_2 - 1)z^2/2)}{1 - z}.$$

Thus, the bivariate PGF satisfies, by meromorphic analysis,

$$p_n(u_1, u_2) = [z^n]F(z, u_1, u_2) \sim e^{(u_1 - 1)}\, e^{(u_2 - 1)/2}.$$

The joint distribution of (χ_1, χ_2) is then a product of a Poisson(1) and a Poisson$(1/2)$ distribution; in particular χ_1 and χ_2 are asymptotically independent. Such a fact results from an extension of the continuity theorem (Theorem IX.1) to multivariate PGF's that is proved by multiple Cauchy integration.

Consider next the joint distribution of $\chi = (\chi_1, \chi_2)$, where χ_j is the number of j-summands in a random integer composition. Each parameter individually obeys a limit Gaussian law, since the sequence construction is supercritical. The trivariate GF is

$$F(z, u_1, u_2) = \frac{1}{1 - z(1 - z)^{-1} - (u_1 - 1)z - (u_2 - 1)z^2}.$$

By meromorphic analysis, a higher dimensional quasi-power approximation may be derived:

$$[z^n]F(z, u_1, u_2) \sim c(u_1, u_2)\rho(u_1, u_2)^{-n},$$

for some 3rd degree algebraic function $\rho(u_1, u_2)$. In such cases, multivariate versions of the continuity theorem for integral transforms can be applied. See the book by Gnedenko and Kolmogorov [**201**], and especially the treatment of Bender and Richmond in [**34**]. As a result, the joint distribution is, in the asymptotic limit, a bivariate Gaussian distribution. Such generalizations are typical and involve essentially no radically new concept, just natural technical adaptations.

A highly interesting approach to multivariate problems is that of *functional limit theorems.* There the goal is to characterize the joint distribution of a potentially infinite collections of parameters. The limit process is then a stochastic process. For instance, the joint distribution of all altitudes in random walks gives rise to Brownian motion. The joint distribution of all cycle lengths in random permutations is described explicitly by Cauchy's formula (Chapter III), and DeLaurentis and Pittel [**97**] have also shown convergence to the standard Brownian motion process. A rather spectacular application of this context of ideas was provided in 1977 by Logan, Shepp, Vershik and Kerov [**293, 426**]. These authors show that the shape of the pair of Young tableaux [**264**] associated to a random permutation conforms, in the asymptotic limit and with high probability, to a deterministic trajectory defined as the solution to a variational problem. In particular, the width of a Young tableau associated to a permutation gives the length of the longest increasing sequence of the permutation. By specializing their results, the authors were able to show that the expected length in a random permutation of size n is asymptotic to $2\sqrt{n}$, a long standing conjecture at the time.

IX. 13. Notes

This chapter is primarily inspired by the works of Bender and Richmond [**27, 34, 35**], Canfield [**71**], Flajolet, Soria, and Drmota [**110, 111, 113, 114, 177, 179, 388**] as well as Hwang [**235**].

Bender's seminal paper [**27**] initiated the study of bivariate analytic schemes that lead to Gaussian laws and the paper [**27**] may rightly be considered to be at the origin of the field. Canfield [**71**], building upon earlier works showed the approach to extend to saddle point schemas.

Tangible progress was next made possible by the development of the singularity analysis method [**167**]. Earlier works were mostly restricted to methods based on subtraction of singularities, as in [**27**], which is in particular effective for meromorphic cases. The extension to algebraic–logarithmic singularities was however difficult given that the classical method of Darboux does not provide for uniform error terms. In contrast, singularity analysis *does* apply to classes of analytic functions, since it allows for uniformity of estimates. The papers by Flajolet and Soria [**177, 179**] were the first to make clear the impact of singularity analysis on bivariate asymptotics. Gao and Richmond [**192**] were then able to extend the theory to cases where both a singularity and its singular exponent are allowed to vary.

From there, Soria developed considerably the framework of schemas in her doctorate [**388**]. Hwang extracted the very important concept of "quasi-powers" in his thesis [**235**] together with a wealth of properties like full asymptotic expansions, speed of convergence, and large deviations. Drmota established general existence conditions leading to Gaussian laws in the case of implicit, especially algebraic, functions [**110, 111**]. The "singularity perturbation" framework for solutions of linear differential equations first appears under that name in [**163**]. The presentation in this chapter is very liberally based on the survey paper [**160**]. Finally, the books by Sachkov, see [**369**] and especially [**371**], offer a modern perspective on bivariate asymptotics applied to classical combinatorial structures.

As pointed out in the introduction, the way combinatorial constructions induce limit laws via schemas based on a purely local perturbation of a singular structure is quite striking. Take for instance the principle that any fixed pattern occurs almost surely in a large random object and its number of occurrences is governed by Gaussian fluctuations. We have shown this property to hold true for strings, uniform tree models, and search trees. In a context that involves either a rational function, an algebraic function, or a solution to a nonlinear differential equation, it eventually reduces to a very simple property, a singularity that smoothly moves. . .

I can see looming ahead one of those terrible exercises in probability where six men have white hats and six men have black hats and you have to *work it out by mathematics how likely it is that the hats will get mixed up and in what proportion. If you start thinking about things like that, you would go round the bend. Let me assure you of that!*

—AGATHA CHRISTIE
(*The Mirror Crack'd*. Toronto, Bantam Books, 1962.)

Part D

APPENDICES

APPENDIX A

Auxiliary Elementary Notions

This appendix contains entries arranged in alphabetical order regarding the following topics:

> *Arithmetical functions; Asymptotic Notations; Combinatorial probability; Cycle construction; Formal power series; Lagrange Inversion; Regular languages; Stirling numbers; Tree concepts.*

The corresponding notions and results are used throughout the book, and in particular in Part A relative to *Symbolic Methods*.

1. Arithmetical functions. A general reference for this section is Apostol's book [**12**]. First, the *Euler totient function* $\varphi(k)$ intervenes in the unlabelled cycle construction. It is defined as the number of integers in $[1, k]$ that are relatively prime to k. Thus, one has $\varphi(p) = p - 1$ if $p \in \{2, 3, 5, \ldots\}$ is a prime. More generally when the prime number decomposition of k is $k = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, then

$$\varphi(k) = p_1^{\alpha_1 - 1}(p_1 - 1) \cdots p_r^{\alpha_r}(p_r - 1).$$

A number is squarefree if it is not divisible by the square of a prime. The *Möbius function* $\mu(n)$ is defined to be 0 if n is not squarefree and otherwise is $(-1)^r$ if $n = p_1 \cdots p_r$ is a product of r distinct primes.

Many elementary properties of arithmetical functions are easily established by means of a *Dirichlet generating functions* (DGF). Let $(a_n)_{n \geq 1}$ be a sequence; its DGF formally defined by

$$\alpha(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}.$$

In particular, the DGF of the sequence $a_n = 1$ is the Riemann zeta function, $\zeta(s) = \sum_{n \geq 1} n^{-s}$. The fact that every number uniquely decomposes into primes is reflected by Euler's formula,

$$\zeta(s) = \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p^s}\right)^{-1}, \tag{1}$$

where p ranges over the set $\mathcal{P}$ of all primes. (As observed by Euler, the fact that $\zeta(1) = \infty$ in conjunction with (1) provides a simple analytic proof that there are infinitely many primes! See Note IV.1, p. 215)

Equation (1) implies elementarily that

$$M(s) := \sum_{n \geq 1} \frac{\mu(n)}{n^s} = \prod_{p \in \mathcal{P}} \left(1 - \frac{1}{p^s}\right) = \frac{1}{\zeta(s)}, \tag{2}$$

where $\mu(n)$ is the Möbius coefficient defined above.

Finally, if $(a_n), (b_n), (c_n)$ have DGF $\alpha(s), \beta(s), \gamma(s)$, then one has the equivalence

$$\alpha(s) = \beta(s)\gamma(s) \quad \Longleftrightarrow \quad a_n = \sum_{d \mid n} b_d c_{n/d}.$$

In particular, taking $c_n = 1$ ($\gamma(s) = \zeta(s)$) and solving for $\beta(s)$ shows (using (2)) the implication

$$a_n = \sum_{d \mid n} b_d \quad \Longleftrightarrow \quad b_n = \sum_{d \mid n} \mu(d) a_{n/d},$$

which is known as *Möbius inversion*. This relation is used in the enumeration of irreducible polynomials (Section I. 6.3).

2. Asymptotic Notations. Let $\mathbb{S}$ be a set and $s_0 \in \mathbb{S}$ a particular element of $\mathbb{S}$. We assume a notion of neighbourhood to exist on $\mathbb{S}$. Examples are $\mathbb{S} = \mathbb{Z}_{>0} \cup \{+\infty\}$ with $s_0 = +\infty$, $\mathbb{S} = \mathbb{R}$ with s_0 any point in $\mathbb{R}$, and $\mathbb{S} = \mathbb{C}$ or a subset of $\mathbb{C}$ with $s_0 = 0$, and so on. Two functions ϕ and g from $\mathbb{S} \setminus \{s_0\}$ to $\mathbb{C}$ are given.

— $\mathcal{O}$*–notation*: write

$$\phi(s) \underset{s \to s_0}{=} O(g(s))$$

if the ratio $\phi(s)/g(s)$ stays bounded as $s \to s_0$ in $\mathbb{S}$. In other words, there exists a neighbourhood $\mathcal{V}$ of s_0 and a constant $C > 0$ such that

$$|\phi(s)| \le C\,|g(s)|\,, \qquad s \in \mathcal{V}, \quad s \ne s_0.$$

One also says that *"ϕ is of order at most g, or ϕ is big–Oh of g (as s tends to s_0)"*.

— $\sim$*–notation*: write

$$\phi(s) \underset{s \to s_0}{\sim} g(s)$$

if the ratio $\phi(s)/g(s)$ tends to 1 as $s \to s_0$ in $\mathbb{S}$. One also says that "*ϕ and g are asymptotically equivalent (as s tends to s_0)*".

— o*–notation*: write

$$\phi(s) \underset{s \to s_0}{=} o(g(s))$$

if the ratio $\phi(s)/g(s)$ tends to 0 as $s \to s_0$ in $\mathbb{S}$. In other words, for any (arbitrarily small) $\varepsilon > 0$, there exists a neighbourhood $\mathcal{V}_\varepsilon$ of s_0 (depending on ε), such that

$$|\phi(s)| \le \varepsilon\,|g(s)|\,, \qquad s \in \mathcal{V}_\varepsilon, \quad s \ne s_0.$$

One also says that *"ϕ is of order smaller than g, or ϕ is little–oh of g (as s tends to s_0)"*.

These notations are due to Bachmann and Landau towards the end of the nineteenth century. See Knuth's note for a historical discussion [**271**, Ch. 4].

Related notations, of which however we only make scanty use, are

— Ω-*notation*: write

$$\phi(s) \underset{s \to s_0}{=} \Omega(g(s))$$

if the ratio $\phi(s)/g(s)$ stays bounded from below in modulus by a nonzero quantity, as $s \to s_0$ in $\mathbb{S}$. One then says that ϕ is of order at least g.

— Θ-*notation*: write
$$\phi(s) \underset{s\to s_0}{=} \Theta(g(s))$$
if $\phi(s) = O(s)$ and $\phi(s) = \Omega(s)$. One then says that ϕ is of order exactly g.

For instance, one has as $n \to +\infty$ in $\mathbb{Z}_{>0}$:
$$\begin{array}{l} \sin n = o(\log n); \quad \log n = O(\sqrt{n}); \quad \log n = o(\sqrt{n}); \\ \binom{n}{2} = \Omega(n\sqrt{n}); \quad \pi n + \sqrt{n} = \Theta(n). \end{array}$$
As $x \to 1$ in $\mathbb{R}_{\le 1}$, one has
$$\sqrt{1-x} = o(1); \quad e^x = O(\sin x); \quad \log x = \Theta(x-1).$$

We take as granted in this book the elementary asymptotic calculus with such notations (see, e.g., [**382**, Ch. 4] for a smooth introduction close to the needs of analytic combinatorics and de Bruijn's classic [**93**] for a beautiful presentation.). We shall retain here in particular the fact that Taylor expansions imply asymptotic expansions; for instance, the convergent expansions valid for $|u| < 1$,
$$\log(1+u) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} u^k, \ \exp(u) = \sum_{k\ge 0} \frac{1}{k!} u^k, \ (1-u)^{-\alpha} = \sum_{k\ge 0} \binom{k+\alpha-1}{k} u^k,$$
imply (as $u \to 0$)
$$\log(1+u) = u + O(u^2), \ \exp(u) = 1 + u + \frac{u^2}{2} + O(u^3), \ (1-u)^{1/2} = 1 - \frac{u}{2} + O(u^2),$$
and, in turn, (as $n \to +\infty$)
$$\log\left(1+\frac{1}{n}\right) = \frac{1}{n} + O\left(\frac{1}{n^2}\right), \quad \left(1 - \frac{1}{\log n}\right)^{1/2} = 1 - \frac{1}{2\log n} + o\left(\frac{1}{\log n}\right).$$

Two important asymptotic expansions are Stirling's formula for factorials and the harmonic number approximation, valid for $n \ge 1$,
$$(3) \qquad \begin{array}{rcll} n! & = & n^n e^{-n}\sqrt{2\pi n}\,(1+\epsilon_n), & 0 < \epsilon_n < \frac{1}{12n} \\ \mathrm{H}_n & = & \log n + \gamma + \dfrac{1}{2n} - \dfrac{1}{12n^2} + \eta_n & \eta_n = O\left(n^{-4}\right), \quad \gamma \doteq 0.57721, \end{array}$$
that are best established as consequences of the Euler–Maclaurin summation formula (see [**93, 382**] as well as APPENDIX B: *Mellin transform*, p. 674).

Asymptotic scales. An important notion due to Henri Poincaré is that of an *asymptotic scale*. A sequence of functions $\omega_0, \omega_1, \ldots$ is said to constitute an asymptotic scale if all functions ω_j exist in a common neighbourhood of $s_0 \in \mathbb{S}$ and if they satisfy there, for all $j \ge 0$:
$$\omega_{j+1}(s) = o(\omega_j(s)), \qquad \text{i.e.,} \quad \lim_{s\to s_0} \frac{\omega_{j+1}(s)}{\omega_j(s)} = 0.$$
Examples at 0 are the scales: $u_j(x) = x^j$; $v_{2j}(x) = x^j \log x$ and $v_{2j+1}(x) = x^j$; $w_j(x) = x^{j/2}$. Examples at infinity are $t_j(n) = n^{-j}$, and so on. Given a scale $\Phi = (\omega_j(s))_{j\ge 0}$, a function f is said to admit an *asymptotic expansion* in the scale Φ

if there exists a family of complex coefficients (λ_j) (the family is then necessarily unique) such that, for each integer m:

$$f(s) = \sum_{j=0}^{m} \lambda_j \omega_j(s) + O(\omega_{m+1}(s)) \qquad (s \to s_0). \tag{4}$$

In this case, one writes

$$f(s) \sim \sum_{j=0}^{\infty} \lambda_j \omega_j(s), \qquad (s \to s_0) \tag{5}$$

with an extension of the symbol '$\sim$'. (Some authors prefer the notation '$\approx$".) The scale may be finite and in most cases, we do not need to specify it as it clear from context. For instance, one can write

$$\mathrm{H}_n \sim \log n + \gamma + \frac{1}{12n}, \qquad \tan x \sim x + \frac{1}{3}x^3 + \frac{2}{15}x^5.$$

In the first case, it is understood that $n \to \infty$ and the scale is $\log n, 1, n^{-1}, n^{-2}, \ldots$. In the second case, $x \to 0$ and the scale is $x, x^3, x^5, \ldots$. Note that in the case of an infinite expansion, convergence of the infinite sum is not implied in (5): the relation is to be interpreted literaly in the sense of (4) as a collection of more and more precise descriptions of f as s becomes closer and closer to s_0.

$\triangleright$ **1.** *Simplification rules for the asymptotic calculus.* Some of them are

$$\begin{array}{llll} O(\lambda f) & \longrightarrow & O(f) & (\lambda \neq 0) \\ O(f) \pm O(g) & \longrightarrow & O(|f| + |g|) & \\ & \longrightarrow & O(f) & \text{if } g = O(f) \\ O(f \cdot g) & \longrightarrow & O(f)O(g). & \end{array}$$

Similar rules apply for $o(\cdot)$. $\triangleleft$

$\triangleright$ **2.** *Harmonics of harmonics.* The harmonic numbers are readily extended to non-integral index by (cf also the ψ function p. 664)

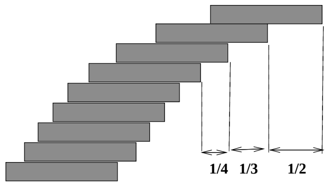
$$\mathrm{H}_x := \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+x} \right).$$

For instance, $\mathrm{H}_{1/2} = 2 - 2\log 2$. This extension is related to the Gamma function [**433**], and it can be proved that the asymptotic estimate (3), with x replacing n, remains valid as $x \to +\infty$. A typical asymptotic calculation shows that

$$\mathrm{H}_{\mathrm{H}_n} = \log\log n + \gamma + \frac{\gamma + \frac{1}{2}}{\log n} + O\left(\frac{1}{\log^2 n} \right).$$

What is the shape of an asymptotic expansion of $\mathrm{H}_{\mathrm{H}_{\mathrm{H}_n}}$? $\triangleleft$

$\triangleright$ **3.** *Stackings of dominos*. A stock of dominos of length 1cm is given. It is well known that one can stack up dominos in a harmonic mode:



Estimate within 1% the minimal number of dominos needed to achieve a horizontal span of 1m (=100cm). [Hint: about 1.50926 10^{43} dominos!] Set up a scheme to evaluate this integer exactly, and do it! $\triangleleft$

$\triangleright$ **4.** *High precision fraud.* Why is it that, to forty decimal places, one finds

$$\begin{array}{rcl} 4\sum_{k=1}^{500,000} \dfrac{(-1)^{k-1}}{2k-1} & \doteq & 3.14159\underline{0}65358979324\underline{0}\underline{4}6264338326\underline{9}502884197 \\ \pi & \doteq & 3.14159265358979323846264338327950288419 7, \end{array}$$

with only four "wrong" digits in the first sum? (Hint: consider the simpler problem

$$\frac{1}{9801} \doteq 0.00\,01\,02\,03\,04\,05\,06\,07\,08\,09\,10\,11\,12\,13\,14\,15\,16\,17\,18\,19\,20\,21\,22\,23\,24\,25\cdots.)$$

Many fascinating facts of this kind are found in works by Jon and Peter Borwein [**59, 60**]. $\triangleleft$

Uniform asymptotic expansions. The notions previously introduced admit of *uniform* versions in the case of families dependent on a secondary parameter [**93**, pp. 7–9]. Let $\{f_u(s)\}_{u\in U}$ be a family of functions indexed by U. An asymptotic equivalence like

$$f_u(s) = O(g(s)) \qquad (s \to s_0),$$

is said to be *uniform with respect to* u if there exists an absolute constant K (independent of $u \in U$) and a fixed neighbourhood $\mathcal{V}$ of s_0 such that

$$\forall u \in U,\ \forall s \in \mathcal{V}: \qquad |f_u(s)| \le K|g(s)|.$$

This definition in turn gives rise to the notion of a uniform asymptotic expansion: it suffices that, for each m, the O error term in (4) be uniform in the sense above. Such notions are central for the determination of limit laws in Chapter IX, where a uniform expansion of a class of generating functions near a singularity is usually required.

$\triangleright$ **5.** *Examples of uniform asymptotics.* One has *uniformly*, for $u \in \mathbb{R}$ and $u \in [0,1]$ respectively:

$$\sin(ux) \underset{x\to\infty}{=} O(1), \qquad \left(1+\frac{1}{n}\right)^u \underset{n\to\infty}{=} 1 + \frac{u}{n} + O\left(\frac{1}{n^2}\right).$$

However, the second expansion no longer holds uniformly with respect to u when $u \in \mathbb{R}$ (take $u = \pm n$), though it holds *pointwise* (non-uniformly) for any fixed $u \in \mathbb{R}$. What about the assertion $\left(1+\frac{1}{n}\right)^u \underset{n\to\infty}{=} 1 + \frac{u}{n} + O\left(\frac{u^2}{n^2}\right)$ for $u \in \mathbb{R}$? $\triangleleft$

3. Combinatorial probability. This entry gathers elementary concepts from probability theory specialized to the discrete case and used in Chapter III. A more elaborate discussion of probability theory forms the subject of Appendix C.

Given a finite set $\mathcal{S}$, the *uniform probability measure* assigns to any $\sigma \in \mathcal{S}$ the probability mass

$$\mathbb{P}(\sigma) = \frac{1}{\operatorname{card}(\mathcal{S})}.$$

The probability of any set, also known as *event*, $\mathcal{E} \subseteq \mathcal{S}$, is then measured by

$$\mathbb{P}\{\mathcal{E}\} := \frac{\operatorname{card}(\mathcal{E})}{\operatorname{card}(\mathcal{S})} = \sum_{\sigma \in \mathcal{E}} \mathbb{P}(\sigma)$$

("the number of favorable cases over the total number of cases").

Given a combinatorial class $\mathcal{A}$, we make extensive use of this notion with the choice of $\mathcal{S} = \mathcal{A}_n$. This defines a probability model (indexed by n), in which of elements of the size n in $\mathcal{A}$ are taken with equal likelihood. For this uniform probabilistic model, we write

$$\mathbb{P}_n \qquad \text{and} \qquad \mathbb{P}_{\mathcal{A}_n},$$

whenever the size and the type of combinatorial structure considered need to be emphasized.

Next consider a parameter χ, which is a function from $\mathcal{S}$ to $\mathbb{Z}_{\geq 0}$. We regard such a parameter as a *random variable*, determined by its probability distribution,

$$\mathbb{P}(\chi = k) = \frac{\operatorname{card}\left(\{\sigma \mid \chi(\sigma) = k\}\right)}{\operatorname{card}(\mathcal{S})}.$$

The notions above extend gracefully to nonuniform probability models that are determined by a family of nonnegative numbers $(p_\sigma)_{\sigma \in \mathcal{S}}$ which add up to 1:

$$\mathbb{P}(\sigma) = p_\sigma, \quad \mathbb{P}(\mathcal{E}) := \sum_{\sigma \in \mathcal{E}} p_\sigma, \quad \mathbb{P}(\chi = k) = \sum_{\chi(\sigma) = k} p_\sigma.$$

Moments. An important information on a distribution is provided by its *moments*. We state here the definitions for an arbitrary discrete random variable supported by $\mathbb{Z}$ and determined by its probability distribution, $\mathbb{P}(X = k) = p_k$ where the $(p_k)_{k \in \mathbb{Z}}$ are nonnegative numbers that add up to 1. The *expectation* of $f(X)$ is defined as the linear functional

$$\mathbb{E}(f(X)) = \sum_k \mathbb{P}\{X = k\} \cdot f(k).$$

In particular, the (power) *moment* of order r is defined as the expectation:

$$\mathbb{E}(X^r) = \sum_k \mathbb{P}\{X = k\} \cdot k^r.$$

Of special importance are the first two moments of the random variable X. The expectation (also mean or average) $\mathbb{E}(X)$ is

$$\mathbb{E}(X) = \sum_k \mathbb{P}\{X = k\} \cdot k.$$

The second moment $\mathbb{E}(X^2)$ gives rise to the *variance*,

$$\mathbb{V}(X) = \mathbb{E}\left((X - \mathbb{E}(X))^2\right) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

and, in turn, to the *standard deviation*

$$\sigma(X) = \sqrt{\mathbb{V}(X)}.$$

The mean deserves its name as first observed by Galileo Galilei (1564–1642): if a large number of draws are effected and values of X are observed, then the arithmetical mean of the observed values will normally be close to the expectation $\mathbb{E}(X)$. The standard deviation measures in a mean quadratic sense the dispersion of values around the expectation $\mathbb{E}(X)$.

$\triangleright$ **6.** *The weak law of large numbers.* Let (X_k) be a sequence of mutually independent random variables with a common distribution. If the expectation $\mu = \mathbb{E}(X_k)$ exists, then for every ϵ:

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\frac{1}{n}(X_1 + \cdots + X_n) - \mu\right| > \epsilon\right) = 0.$$

(See [**133**, Ch X] for a proof.) Note that the property does not require finite variance. $\triangleleft$

Probability generating function. The *probability generating function* (PGF) of X is by definition:

$$p(u) := \sum_k \mathbb{P}(X = k)u^k,$$

and an alternative expression is $p_n(u) = \mathbb{E}(u^X)$. Moments can be recovered from the PGF by differentiation at 1, for instance:

$$\mathbb{E}(X) = \left.\frac{d}{du}p(u)\right|_{u=1}, \qquad \mathbb{E}(X(X-1)) = \left.\frac{d^2}{du^2}p(u)\right|_{u=1}.$$

More generally, the quantity,

$$\mathbb{E}(X(X-1)\cdots(X-k+1)) = \left.\frac{d^k}{du^k}p(u)\right|_{u=1},$$

is known as the kth factorial moment.

$\triangleright$ **7.** *Relations between factorial and power moments.* Let X be a discrete random variable with PGF $p(u)$; denote by $\mu_r = \mathbb{E}(X^r)$ its rth moment and by ϕ_r its factorial moment. One has

$$\mu_r = \partial_t^r p(e^t)\big|_{t=0}, \qquad \phi_r = \partial_u^r p(u)|_{u=1}.$$

Consequently, with $\left\{ {n \atop k} \right\}$ and $\left[{n \atop k} \right]$ the Stirling numbers of both kinds (APPENDIX A: *Stirling numbers*, p. 652),

$$\phi_r = \sum_j (-1)^{r-j} \begin{bmatrix} r \\ j \end{bmatrix} \mu_j; \qquad \mu_r = \sum_j \begin{Bmatrix} r \\ j \end{Bmatrix} \phi_j.$$

(Hint: for $\phi_r \to \mu_r$, expand the Stirling polynomial defined in (12) below; in the converse direction, write $p(e^t) = p(1 + (e^t - 1))$.) $\triangleleft$

Markov-Chebyshev inequalities. These are fundamental inequalities that apply equally well to discrete and to continuous random variables (see Appendix C for the latter).

THEOREM A.1 (Markov-Chebyshev inequalities). *Let X be a* nonnegative *random variable and Y an* arbitrary *real random variable. One has for an arbitrary $t > 0$:*

$$\begin{array}{lcll} \mathbb{P}\{X \geq t\mathbb{E}(X)\} & \leq & \dfrac{1}{t} & \textit{(Markov inequality)} \\ \mathbb{P}\{|Y - \mathbb{E}(Y)| \geq t\sigma(Y)\} & \leq & \dfrac{1}{t^2} & \textit{(Chebyshev inequality)}. \end{array}$$

PROOF. Without loss of generality, one may assume that x has been scaled in such a way that $\mathbb{E}(X) = 1$. Define the function $f(x)$ whose value is 1 if $x \geq t$, and 0 otherwise. Then

$$\mathbb{P}\{X \geq t\} = \mathbb{E}(f(X)).$$

Since $f(x) \leq x/t$, the expectation on the right is less than $1/t$. Markov's inequality follows. Chebyshev's inequality then results from Markov's inequality applied to $X = |Y - \mathbb{E}(Y)|^2$. □

Theorem A.1 informs us that the probability of being much larger than the mean must decay (Markov) and that an upperbound on the decay is measured in units given by the standard deviation (Chebyshev).

Moment inequalities are discussed for instance in Billingsley's reference treatise [**53**, p. 74]. They are of great importance in discrete mathematics where they have been put to use in order to show the *existence* of surprising configurations. This field was pioneered by Erdős and is often known as the "probabilistic method" [in combinatorics]; see the book by Alon and Spencer [**9**] for many examples. Moment inequalities can also be used to estimate the probabilities of complex events by reducing the problems to moment estimates for occurrences of simpler configurations—this is one of the bases of the "first and second moment methods", again pioneered by Erdős, which are central in the theory of random graphs [**56, 245**]. Finally, moment inequalities serve to design, analyse, and optimize randomized algorithms, a theme excellently covered in the book by Motwani and Raghavan [**323**].

4. Cycle construction. The unlabelled cycle construction is introduced in Chapter I and is classically obtained within the framework of Pólya theory [**82, 347, 349**]. The derivation given here is based on an elementary use of symbolic methods that follows [**178**]. It relies on bivariate GF's developed in Chapter III, with z marking size and u marking the number of components. Consider a class $\mathcal{A}$ and the sequence class $\mathcal{S} = \mathrm{SEQ}_{\geq 1}(\mathcal{A})$. A sequence $\sigma \in \mathcal{S}$ is primitive (or aperiodic) if it is not the repetition of another sequence (e.g., $\alpha\beta\beta\alpha\alpha$ is primitive, but $\alpha\beta\alpha\beta = (\alpha\beta)^2$ is not). The class $\mathcal{PS}$ of primitive sequences is determined implicitly,

$$S(z, u) \equiv \frac{uA(z)}{1 - uA(z)} = \sum_{k \geq 1} PS(z^k, u^k),$$

which expresses that every sequence possesses a “root” that is primitive. Möbius inversion then gives

$$PS(z,u) = \sum_{k\geq 1} \mu(k) S(z^k, u^k) = \sum_{k\geq 1} \mu(k) \frac{u^k A(z^k)}{1-u^k A(z^k)}.$$

A cycle is primitive if all of its linear representations are primitive. There is an exact one-to-ℓ correspondence between primitive ℓ-cycles and primitive ℓ-sequences. Thus, the BGF $PC(z,u)$ of primitive cycles is obtained by effecting the transformation $u^\ell \mapsto \frac{1}{\ell} u^\ell$ on $PS(z,u)$, which means

$$PC(z,u) = \int_0^u P(z,v)\, \frac{dv}{v},$$

giving after term-wise integration,

$$PC(z,u) = \sum_{k\geq 1} \frac{\mu(k)}{k} \log \frac{1}{1-u^k A(z^k)}.$$

Finally, cycles can be composed from arbitrary repetitions of primitive cycles (each cycle has a primitive “root”), which yields for $\mathcal{C} = \text{CYC}(\mathcal{A})$:

$$C(z,u) = \sum_{k\geq 1} PC(z^k, u^k).$$

The arithmetical identity $\sum_{d\,|\,k} \mu(d)/d = \varphi(k)/k$ gives eventually

$$C(z,u) = \sum_{k\geq 1} \frac{\varphi(k)}{k} \log \frac{1}{1-u^k A(z^k)}. \tag{6}$$

Formula (6) specializes to the one that appears in the translation of the cycle construction in the unlabelled case (Theorem I.1), upon setting $u = 1$; this formula also coincides the statement of Proposition III.5 regarding the number of components in cycles, and it yields the general multivariate version (Theorem III.1) by a simple adaptation of the argument.

▷ **8.** *Around the cycle construction.* Similar methods yield the BGFs of multisets of cycles and multisets of aperiodic cycles as

$$\prod_{k\geq 1} \frac{1}{1-u^k A(z^k)} \quad \text{and} \quad \frac{1}{1-uA(z)},$$

respectively [**94**]. (The latter fact corresponds to the property that any word can be written as a decreasing product of Lyndon words. Notably, it serves to construct bases of free Lie algebras [**294**, Ch. 5].) ◁

▷ **9.** *Aperiodic words.* An aperiodic word is a primitive sequence of letters. The number of aperiodic words of length n over an m-ary alphabet corresponds to primitive sequences with $A(z) = mz$ and is

$$PW_n^{(m)} = \sum_{d\,|\,n} \mu(d) m^{n/d}.$$

For $m = 2$, the sequence starts as $2, 2, 6, 12, 30, 54, 126, 240, 504, 990$ (*EIS* **A027375**). ◁

5. Formal power series. Formal power series extend the usual operations on polynomials to infinite series of the form

$$f = \sum_{n \geq 0} f_n z^n, \tag{7}$$

where z is a formal indeterminate. The notation $f(z)$ is also employed. Let $\mathbb{K}$ be a ring of coefficients (usually we shall take one of the fields $\mathbb{Q}, \mathbb{R}, \mathbb{C}$); the ring of formal power series is denoted by $\mathbb{K}[[z]]$ and it is the set $\mathbb{K}^{\mathbb{N}}$ (of infinite sequences of elements of $\mathbb{K}$) written as infinite power series (7) and endowed with the operations of sum and product,

$$\begin{array}{rcl} \left(\sum_n f_n z^n\right) + \left(\sum_n g_n z^n\right) & := & \sum_n (f_n + g_n)\, z^n \\ \left(\sum_n f_n z^n\right) \times \left(\sum_n g_n z^n\right) & := & \sum_n \left(\sum_{k=0}^{n} f_k g_{n-k}\right) z^n. \end{array}$$

A topology, known as the *formal topology*, is put on $\mathbb{K}[[z]]$ by which two series f, g are "close" if they coincide to a large number terms. First, the valuation of a formal power series $f = \sum_n f_n z^n$ is the smallest r such that $f_r \neq 0$ and is denoted by $\operatorname{val}(f)$. (One sets $\operatorname{val}(0) = +\infty$.) Given two power series f and g, their distance $d(f,g)$ is then defined as $2^{-\operatorname{val}(f-g)}$. With this distance (in fact an ultrametric distance), the space of all formal power series becomes a *complete metric space*. Roughly, the limit of a sequence of series $\{f^{(j)}\}$ exists if, for each n, the coefficient of order n in $f^{(j)}$ eventually stabilizes to a fixed value as $j \to \infty$. In this way *formal convergence* can be defined for infinite sums: it suffices that the general term of the sum should tend to 0 in the formal topology, *i.e.*, the valuation of the general term should tend to ∞. Similarly for infinite products, where $\prod (1 + u^{(j)})$ converges as soon as $u^{(j)}$ tends to 0 in the topology of formal power series.

It is then a simple exercise to prove that the sum $Q(f) := \sum_{k \geq 0} f^k$ exists (the sum convergerges in the formal topology) whenever $f_0 = 0$; the quantity then defines the *quasi-inverse* written $(1-f)^{-1}$, with the implied properties with respect to multiplication (namely, $Q(f)(1-f) = 1$). In the same way one defines formally logarithms and exponentials, primitives and derivatives, etc. Also, the composition $f \circ g$ is defined whenever $g_0 = 0$ by substitution of formal power series. More generally, any (possibly infinitary) process on series that involves at each coefficient only finitely many operations is well-defined and is accordingly a continuous functional in the formal topology.

▷ **10.** *The OGF of permutations.* The ordinary generating function of permutations,

$$P(z) := \sum_{n=0}^{\infty} n! z^n = 1 + z + 2z^2 + 6z^3 + 24z^4 + 120z^5 + 720z^6 + 5040z^7 + \cdots$$

exists as an element of $\mathbb{C}[[z]]$, although the series has radius of convergence 0. The quantity $1/P(z)$ is for instance well-defined (via the quasi-inverse) and one can compute legitimately and effectively $1 - 1/P(z)$ whose coefficients enumerate indecomposable permutations (p. 81). The formal series $P(z)$ can even be made sense of analytically as an asymptotic series (Euler), since

$$\int_0^\infty \frac{e^{-t}}{1+tz}\, dt \sim 1 - z + 2!z^2 - 3!z^3 + 4!z^4 - \cdots \qquad (z \to 0+).$$

Thus, the OGF of permutations is also representable as the (formal, divergent) asymptotic series associated to an integral. $\lhd$

It can be proved that the usual functional properties of analysis extend to formal power series provided they make sense formally. The extension to multivariate formal power series follows along entirely similar lines.

6. Lagrange Inversion. Lagrange inversion (Lagrange, 1770) relates the coefficients of the inverse of a function to coefficients of the powers of the function itself. It thus establishes a fundamental correspondence between functional composition and standard multiplication of series. Although the proof is technically simple, the result altogether non-elementary.

The inversion problem $z = h(y)$ is solved by the Lagrange series given below. It is assumed that $[y^0]h(z) = 0$, so that inversion is formally well defined and analytically local, and $[y^1]h(y) \neq 0$. The problem is is then conveniently standardized by setting $h(y) = y/\phi(y)$.

THEOREM A.2. *Let $\phi(u) = \sum_{k\geq 0} \phi_k u^k$ be a power series of $\mathbb{C}[[z]]$ with $\phi_0 \neq 0$. Then, the equation $y = z\phi(y)$ admits a unique solution in $\mathbb{C}[[z]]$ whose coefficients are given by (Lagrange form)*

$$y(z) = \sum_{n=1}^{\infty} y_n z^n, \qquad \text{where} \quad y_n = \frac{1}{n}\,[u^{n-1}]\,\phi(u)^n. \tag{8}$$

Furthermore, one has for $k > 0$ (Bürmann form)

$$y(z)^k = \sum_{n=1}^{\infty} y_n^{(k)} z^n, \qquad \text{where} \quad y_n^{(k)} = \frac{k}{n}\,[u^{n-k}]\,\phi(u)^n. \tag{9}$$

By linearity, a form equivalent to Burmann's (9), with H an arbitrary function, is

$$[z^n]H(y(z)) = \frac{1}{n}[u^{n-1}]\left(H'(u)\phi(u)^n\right).$$

PROOF. The method of indeterminates coefficients provides a system of polynomial equations for $\{y_n\}$ that is seen to admit a unique solution:

$$y_1 = \phi_0, \quad y_2 = \phi_0\phi_1, \quad y_3 = \phi_0\phi_1^2 + \phi_0^2\phi_2, \; \ldots \; .$$

Since y_n only depends polynomially on the coefficients of $\phi(u)$ till order n, one may assume without loss of generality, in order to establish (8) and (9) that ϕ is a polynomial. Then, by general properties of analytic functions, $y(z)$ is analytic at 0 (see Chapter IV and APPENDIX B: *Equivalent definitions of analyticity*, p. 659 for definitions) and it maps conformally a neighborhood of 0 into another neighbourhood of 0. Accordingly, the quantity $ny_n = [z^{n-1}]y'(z)$ can be estimated by Cauchy's coefficient formula:

$$
(10)\qquad
\begin{aligned}
ny_n &= \frac{1}{2i\pi}\int_{0+} y'(z)\,\frac{dz}{z^n} && \text{(Direct coefficient formula for } y'(z)\text{)}\\
&= \frac{1}{2i\pi}\int_{0+} \frac{dy}{(y/\phi(y))^n} && \text{(Change of variable } z\mapsto y\text{)}\\
&= [y^{n-1}]\,\phi(y)^n && \text{(Reverse coefficient formula for } \phi(y)^n\text{).}
\end{aligned}
$$

In the context of complex analysis, this useful result appears as nothing but an avatar of the change-of-variable formula. The proof of Bürmann's form is similar. □

There exist instructive (but longer) combinatorial proofs based on what is known as the "cyclic lemma" or "conjugacy principle" [**359**] for Łukasiewicz words. (See also Note 44 in Chapter I.) Another classical proof due to Henrici relies on properties of iteration matrices [**230**, §1.9]; see also Comtet's book for related formulations [**82**].

Lagrange inversion serves most notably to develop explicit formulæ for simple families of trees (Chapters I and II), random mappings (Chapter II), and more generally for problems involving coefficients of powers of functions.

▷ **11.** *Lagrange–Bürmann inversion for fractional powers.* The formula

$$[z^n]\left(\frac{y(z)}{z}\right)^\alpha = \frac{\alpha}{n+\alpha}[u^n]\phi(u)^{n+\alpha}$$

holds for any real or complex exponent α, and hence generalizes Bürmann's form. One can similarly expand $\log(y(z)/z)$. ◁

▷ **12.** *Abel's identity.* By computing in two different ways the coefficient

$$[z^n]e^{(\alpha+\beta)y} = [z^n]e^{\alpha y}\cdot e^{\beta y},$$

where $y = ze^y$ is the Cayley tree function, one derives *Abel's identity*

$$(\alpha+\beta)(n+\alpha+\beta)^{n-1} = \alpha\beta\sum_{k=0}^{n}\binom{n}{k}(k+\alpha)^{k-1}(n-k+\beta)^{n-k-1}.$$

◁

7. Regular languages. A *language* is a set of words over some fixed alphabet $\mathcal{A}$. The structurally simplest (yet nontrivial) languages are the *regular languages* that, as asserted on p. 54, can be defined in a variety of equivalent ways (see [**3**, Ch. 3] or [**123**]): by regular expressions, either ambiguous or not, and by finite automata, either deterministic or nondeterministic. Our definitions of S-regularity (S as in specification) and A-regularity (A as in automaton) from Chapter I correspond to definability by *unambiguous* regular expression and *deterministic* automaton, respectively.

Regular expressions and ambiguity. Here is first the classical definition of a regular language in formal language theory.

DEFINITION A.1. *The category* RegExp *of* regular expressions *is defined inductively by the property that it contains all the letters of the alphabet* $(a\in\mathcal{A})$ *as well as the empty symbol* ϵ, *and is such that, if* $R_1, R_2 \in$ RegExp, *then the formal expressions* $R_1\cup R_2$, $R_1\cdot R_2$ *and* $R_1^\star$ *are regular expressions.*

Regular expressions are meant to denote *languages*. The language $\mathbf{L}(R)$ associated to R is obtained by interpreting '$\cup$' as set-theoretic union, '$\cdot$' as catenation product extended to sets and '*' as the star operation: $\mathbf{L}(R^\star) := \{\epsilon\} \cup \mathbf{L}(R) \cup (\mathbf{L}(R) \cdot \mathbf{L}(R)) \cup \cdots$. These operations rely on set-theoretic operations and place no condition on multiplicities (a word may be obtained in several different ways). Accordingly, the notions of a regular expression and a regular language are useful when studying structural properties of languages, but they must be adapted for enumeration purposes, where unambiguous specifications are needed.

A word $w \in \mathbf{L}(R)$ may be parsable in several ways according to R: the *ambiguity coefficient* (or *multiplicity*) of w with respect to the regular expression R is defined[1] as the number of parsings and written $\kappa(w) = \kappa_R(w)$.

A regular expression R is said to be *unambiguous* if for all w, we have $\kappa_R(w) \in \{0, 1\}$, ambiguous otherwise. In the unambiguous case, if $\mathcal{L} = \mathbf{L}(R)$, then $\mathcal{L}$ is S-regular in the sense of Chapter I, a specification being obtained by the translation rules:

$$\cup \mapsto +, \qquad \cdot \mapsto \times, \qquad (\,)^\star \mapsto \mathrm{SEQ}, \tag{11}$$

and the translation mechanism afforded by Proposition I.2 p. 48 applies. (Use of the general mechanism (11) in the ambiguous case would imply that we enumerate words with multiplicity (ambiguity) coefficients taken into account.)

A-regularity implies S-regularity. This construction is due to Kleene [**256**] whose interest had its origin in the formal expressive power of nerve nets. Within the classical framework of the theory of regular languages, it produces from an automaton (possibly nondeterministic) a regular expression (possibly ambiguous).

For our purposes, let a deterministic automaton $\mathfrak{a}$ be given, with alphabet $\mathcal{A}$, set of states Q, with q_0 and $\overline{Q}$ the initial state and the set of final states respectively. The idea consists in constructing inductively the family of languages $\mathcal{L}_{i,j}^{(r)}$ of words that connect state q_i to state q_j passing only through states $q_0, \ldots, q_r$ in between q_i and q_j. We initialize the data with $\mathcal{L}_{i,j}^{(-1)}$ to be the singleton set $\{a\}$ if the transition $(q_i \circ a) = q_j$ exists, and the emptyset ($\emptyset$) otherwise. The fundamental recursion

$$\mathcal{L}_{i,j}^{(r)} = \mathcal{L}_{i,j}^{(r-1)} + \mathcal{L}_{i,r}^{(r-1)}\, \mathrm{SEQ}(S)\{\mathcal{L}_{r,r}^{(r-1)}\}\mathcal{L}_{r,j}^{(r-1)},$$

incrementally takes into account the possibility of traversing the "new" state q_r. (The unions are clearly disjoint and the segmentation of words according to passages through state q_r is unambiguously defined, hence the validity of the sequence construction.) The language $\mathcal{L}$ accepted by $\mathfrak{a}$ is then given by the regular specification

$$\mathcal{L} = \sum_{q_j \in \overline{Q}} \mathcal{L}_{0,j}^{\|Q\|},$$

that describes the set of all words leading from the initial state q_0 to any of the final states while passing freely through any intermediate state of the automaton.

[1] For instance if $R = (a \cup aa)^\star$ and $w = aaaa$, then $\kappa(w) = 5$ corresponding to the five parsings: $a \cdot a \cdot a \cdot a$, $a \cdot a \cdot aa$, $a \cdot aa \cdot a$, $aa \cdot a \cdot a$, $aa \cdot aa$.

$$\begin{array}{lcccc}
S\text{-regularity} & \equiv & \begin{array}{c}\text{Unambiguous}\\ \text{RegExp}\end{array} & \longrightarrow & \begin{array}{c}\text{General}\\ \text{RegExp}\end{array}\\
 & & \uparrow \mathbf{K} & & \downarrow \mathbf{I}\\
A\text{-regularity} & \equiv & \begin{array}{c}\text{Deterministic}\\ \text{FA}\end{array} & \xleftarrow{\mathbf{RS}} & \begin{array}{c}\text{Nondeterministic}\\ \text{FA}\end{array}
\end{array}$$

FIGURE 1. Equivalence between various notions of regularity: **K** is Kleene's construction; **RS** is Rabin-Scott's reduction; **I** is the inductive construction of the text.

S-regularity implies A-regularity. An object described by a regular specification $\mathfrak{r}$ can be first encoded as a word, with separators indicating the way the word should be parsed unambiguously. These encodings are then describable by a regular expression using the correspondence of (11). Next any language described by a regular expression is recognizable by an automaton (possibly nondeterministic) as shown by an inductive construction. (We only state the principles informally here.) Let $\to\bullet\boxed{\mathfrak{r}}\bullet\to$ represent symbolically the automaton recognizing the regular expression $\mathfrak{r}$, with the initial state on the left and the final state(s) on the right. Then, the rules are schematically

$$\begin{array}{lcl}
\to\bullet\boxed{\mathfrak{r}+\mathfrak{s}}\bullet\to & = & \begin{array}{l}\nearrow\!\to\bullet\boxed{\mathfrak{r}}\bullet\to\\ \searrow\!\to\bullet\boxed{\mathfrak{s}}\bullet\to\end{array}\\
\to\bullet\boxed{\mathfrak{r}\times\mathfrak{s}}\bullet\to & = & \to\bullet\boxed{\mathfrak{r}}\bullet\to\to\bullet\boxed{\mathfrak{s}}\bullet\to\\
\to\bullet\boxed{\mathfrak{r}^\star}\bullet\to & = & \overline{\downarrow\to\bullet\boxed{\mathfrak{r}}\bullet\to\uparrow}\,.
\end{array}$$

Finally, a standard result of the theory, the Rabin-Scott theorem, asserts that any nondeterministic finite automaton can be emulated by a deterministic one. (Note: this general reduction produces a deterministic automaton whose set of states is the powerset of the set of states of the original automaton; it may consequently involve an exponential blow-up in the size of descriptions.)

8. Stirling numbers.. These numbers count amongst the most famous ones of combinatorial analysis. They appear in two kinds:

- the *Stirling cycle number* (also called 'of the first kind') $\left[{n \atop k}\right]$ enumerates permutations of size n having k cycles;
- the *Stirling partition number* (also called 'of the second kind') $\left\{{n \atop k}\right\}$ enumerates partitions of an n-set into k nonempty equivalence classes.

The notations $\left[{n \atop k}\right]$ and $\left\{{n \atop k}\right\}$ proposed by Knuth (himself anticipated by Karamata) are nowadays most widespread; see [**212**].

The most natural way to define Stirling numbers is in terms of the "vertical" EGFs when the value of k is kept fixed:

$$\begin{aligned}
\sum_{n\geq 0}\left[{n \atop k}\right]\frac{z^n}{n!} &= \frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k\\
\sum_{n\geq 0}\left\{{n \atop k}\right\}\frac{z^n}{n!} &= \frac{1}{k!}\left(e^z-1\right)^k.
\end{aligned}$$

From there, the bivariate EGFs follow straightforwardly:

$$\begin{aligned}\sum_{n,k\geq 0} \left[{n \atop k}\right] u^k \frac{z^n}{n!} &= \exp\left(u \log \frac{1}{1-z}\right) = (1-z)^{-u} \\ \sum_{n,k\geq 0} \left\{{n \atop k}\right\} u^k \frac{z^n}{n!} &= \exp\left(u(e^z-1)\right).\end{aligned}$$

Stirling numbers and their cognates satisfy a host of algebraic relations. For instance, the differential relations of the EGFs imply recurrences reminiscent of the binomial recurrence

$$\left[{n \atop k}\right] = \left[{n-1 \atop k-1}\right] + (n-1)\left[{n-1 \atop k}\right], \qquad \left\{{n \atop k}\right\} = \left\{{n-1 \atop k-1}\right\} + k\left\{{n-1 \atop k}\right\}.$$

By expanding the powers in the vertical EGF of the Stirling partition numbers or by techniques akin to Lagrange inversion, one finds explicit forms

$$\begin{aligned}\left[{n \atop k}\right] &= \sum_{0\leq j\leq h\leq n-k} (-1)^{j+h} \binom{h}{j} \binom{n-1+h}{n-k+h} \binom{2n-k}{n-k-h} \frac{(h-j)^{n-k+h}}{h!} \\ \left\{{n \atop k}\right\} &= \frac{1}{k!} \sum_{j=0}^{r} \binom{k}{j} (-1)^j (k-j)^n.\end{aligned}$$

Though comforting, these forms are not too useful in general. (The one relative to Stirling cycle numbers was obtained by Schlömilch in 1852 [**82**, p. 216].)

A more important relation is that of the generating polynomials of the $\left[{n \atop r}\right]$ for fixed n,

$$(12) \qquad P_n(u) \equiv \sum_{r=0}^{n} \left[{n \atop r}\right] u^r = u \cdot (u+1) \cdot (u+2) \cdots (u+n-1).$$

This nicely parallels the OGF for the $\left\{{n \atop r}\right\}$ for fixed r

$$\sum_{n=0}^{\infty} \left\{{n \atop r}\right\} z^n = \frac{z^r}{(1-z)(1-2z)\cdots(1-kz)}.$$

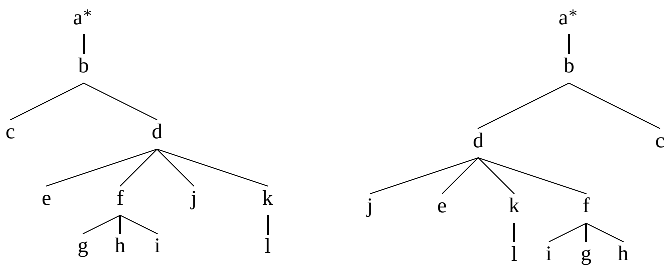
▷ **13.** *Schlömilch's formula.* It is established starting from

$$\frac{k!}{n!}\left[{n \atop k}\right] = \frac{1}{2i\pi} \oint \log^k \frac{1}{1-z} \, \frac{dz}{z^{n+1}},$$

via the change of variable *a la* Lagrange: $z = 1 - e^{-t}$. See [**82**, p.216] and [**170**]. ◁

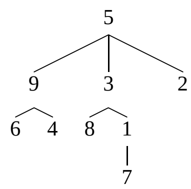
9. Tree concepts. In the abstract graph-theoretic sense, a *forest* is an acyclic (undirected) graph and a *tree* is a forest that consists of just one connected component. A *rooted tree* is a tree in which a specific node is distinguished, the *root*. Rooted trees are drawn with the root either below (the mathematician's and botanists's convention) or on top (the genealogist's and computer scientist's convention), and in this book, we employ both conventions indifferently. Here are then two planar representations of the same rooted tree

(13)



where the star distinguishes the root. (Tags on nodes, a, b, c, etc, are not part of the tree structure but only meant to discriminate nodes here.) A tree whose nodes are labelled by distinct integers then becomes a *labelled tree*, this in the precise technical sense of Chapter II. Size is defined by the number of nodes (vertices). Here is for instance a labelled tree of size 9:

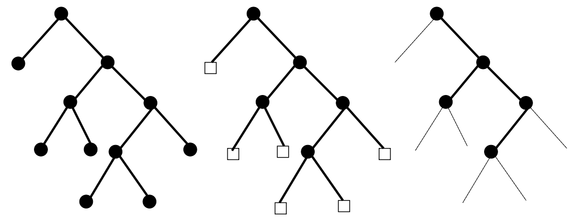
(14)



In a rooted tree, the *outdegree* of a node is the number of its descendants; with the sole exception of the root, outdeegree is thus equal to degree (in the graph-theoretic sense, i.e., the number of neighbours) minus 1. Once this convention is clear, one usually abbreviates “outdegree” by “degree” when speaking of rooted trees. A *leaf* is a node without descendant, that is, a node of (out)degree equal to 0. For instance the tree in (14) has 5 leaves. Non-leaf nodes are also called internal nodes.

Many applications from genealogy to computer science require superimposing an additional structure on a graph-theoretic tree. A *plane tree* (sometimes also called a *planar tree*) is defined as a tree in which subtrees dangling from a common node are ordered between themselves and represented from left to right in order. Thus, the two representations in (13) are equivalent as graph-theoretic trees, but they become distinct objects when regarded as plane trees.

Binary trees play a special role in combinatorics. These are rooted trees in which every nonleaf node has degree 2 exactly as, for instance, in the first two drawings below:



In the second case, the leaves have been distinguished by '□'. The *pruned binary tree* (third representation) is obtained from a regular binary tree by removing the leaves. A binary tree can be fully reconstructed from its pruned version, and a tree of size $2n+1$ always expands a pruned tree of size n.

A few major classes are encountered throughout this book. Here is a summary[2].

General plane trees (Catalan trees)	$\mathcal{G} = \mathcal{Z} \times \mathrm{SEQ}\{\mathcal{G}\}$	(unlabelled)
Binary trees	$\mathcal{A} = \mathcal{Z} + (\mathcal{Z} \times \mathcal{A} \times \mathcal{A})$	(unlabelled)
Nonempty pruned binary trees	$\mathcal{B} = \mathcal{Z} + 2(\mathcal{Z} \times \mathcal{B}) + (\mathcal{Z} \times \mathcal{B} \times \mathcal{B})$	(unlabelled)
Pruned binary trees	$\mathcal{C} = \mathbf{1} + (\mathcal{Z} \times \mathcal{B} \times \mathcal{B})$	(unlabelled)
General nonplane trees (Cayley trees)	$\mathcal{T} = \mathcal{Z} \times \mathrm{SET}\{\mathcal{T}\}$	(labelled)

The corresponding GFs are respectively

$$G(z) = \frac{1-\sqrt{1-4z}}{2}, \quad A(z) = \frac{1-\sqrt{1-4z^2}}{2z}, \quad B(z) = \frac{1-2z-\sqrt{1-4z}}{2z},$$
$$C(z) = \frac{1-\sqrt{1-4z}}{2z}, \quad T(z) = ze^{T(z)},$$

being respectively of type OGF for the first four and EGF for the last one. The corresponding counts are

$$G_n = \frac{1}{n}\binom{2n-2}{n-1}, \quad A_{2\nu+1} = \frac{1}{\nu+1}\binom{2\nu}{\nu}, \quad B_n = \frac{1}{n+1}\binom{2n}{n} \quad (n \geq 1),$$
$$C_n = \frac{1}{n+1}\binom{2n}{n}, \quad T_n = n^{n-1}.$$

The common occurrence of the Catalan numbers, ($C_n = B_n = A_{2n+1} = G_{n+1}$) is explained by pruning and by the rotation correspondence described on p. 69.

[2] The term "general" refers to the fact that no degree constraint is imposed.

APPENDIX B

Basic Complex Analysis

This appendix contains entries arranged in alphabetical order regarding the following topics:

Algebraic elimination; Equivalent definitions of analyticity; Gamma function; Implicit Function Theorem; Laplace's method; Mellin transform; Perron-Frobenius theory of nonnegative matrices; Several complex variables.

The corresponding notions and results are used in particular starting with Part B, which is relative to *Complex Asymptotics*.

1. Algebraic elimination. Auxiliary quantities can be eliminated from systems of polynomial equations. In essence, elimination is achieved by suitable combinations of the equations themselves. One of the best strategies is based on Gröbner bases and is presented in the excellent book of Cox, Little, and O'Shea [**86**]. This entry develops a more elementary approach based on *resultants*.

Resultants. Consider a field of coefficients $\mathbb{K}$ which may be specialized as $\mathbb{Q}, \mathbb{C}, \mathbb{C}(z), \ldots$, as the need arises. A polynomial of degree d in $\mathbb{K}[x]$ has at most d roots in $\mathbb{K}$ and exactly d roots in the algebraic closure $\bar{\mathbb{K}}$ of $\mathbb{K}$. Given two polynomials,

$$P(x) = \sum_{j=0}^{\ell} a_j x^{\ell-j}, \qquad Q(x) = \sum_{k=0}^{m} b_k x^{m-k},$$

their *resultant* (with respect to the variable x) is the determinant of order $(\ell + m)$,

$$\mathbf{R}(P, Q, x) = \det \begin{vmatrix} a_0 & a_1 & a_2 & \cdots & 0 & 0 \\ 0 & a_0 & a_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{\ell-1} & a_\ell \\ b_0 & b_1 & b_2 & \cdots & 0 & 0 \\ 0 & b_0 & b_1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & b_{m-1} & b_m \end{vmatrix}, \tag{1}$$

also called the Sylvester determinant. By its definition, the resultant is a polynomial form in the coefficients of P and Q. The main property of resultants is the following: (i) *If* $P(x), Q(x) \in \mathbb{K}[x]$ *have a common root in the algebraic closure* $\bar{\mathbb{K}}$ *of* $\mathbb{K}$*, then* $\mathbf{R}(P(x), Q(x), x) = 0$. (ii) *Conversely, if* $\mathbf{R}(P(x), Q(x), x) = 0$ *holds, then either* $a_0 = b_0 = 0$ *or else* $P(x), Q(x)$ *have a common root in* $\bar{\mathbb{K}}$. [The idea of the proof of (i) is as follows. Let S be the matrix in (1). Then the homogeneous linear system $Sw = 0$ admits a solution $w = (\xi^{\ell+m-1}, \ldots, \xi^2, \xi, 1)$ where ξ is a common root of P and Q; this is only possible if $\det(S) \equiv \mathbf{R}$ vanishes.] See especially van der Waerden's crips treatmenent in [**421**] and Lang's treatise [**286**, V.10] for a detailed presentation of resultants

Equating the resultant to 0 thus provides a *necessary* condition for the existence of common roots, but not always a sufficient one. This has implications in situations where the coefficients a_j, b_k depend on one or several parameters. In that case, the condition $\mathbf{R}(P, Q, x) = 0$ will certainly capture all the situations where P and Q have a common root, but it may also include some situations where there is a reduction in degree, although the polynomials have no common root. For instance, take $P(x) = tx - 2$ and $Q(x) = tx^2 - 4$ (with t a parameter); the resultant with respect to x is found to be

$$\mathbf{R} = 4t(1-t).$$

Indeed, the condition $\mathbf{R} = 0$ corresponds to either a common root ($t = 1$ for which $P(2) = Q(2) = 0$) or to some degeneracy in degree ($t = 0$ for which $P(x) = -2$ and $Q(x) = -4$ have no common zero).

Systems of equations. Given a system

$$\{P_j(z, y_1, y_2, \ldots, y_m) = 0\}, \qquad j = 1 \,.\,.\, m, \tag{2}$$

defining an algebraic curve, we can then proceed as follows in order to extract a single equation satisfied by one of the indeterminates. By taking resultants with P_m, eliminate all occurrences of the variable y_m from the first $m - 1$ equations, thereby obtaining a new system of $m - 1$ equations in $m - 1$ variables (with z kept as a parameter, so that the base field is $\mathbb{C}(z)$). Repeat the process and successively eliminate $y_{m-1}, \ldots, y_2$. The strategy (in the simpler case where variables are eliminated in succession exactly one at a time) is summarized in the skeletton procedure Eliminate:

procedure Eliminate $(P_1, \ldots, P_m, y_1, y_2, \ldots y_m)$;
$\{$*Elimination of* $y_2, \ldots, y_m$ *by resultants*$\}$
$(A_1, \ldots, A_m) \;:=\; (P_1, \ldots, P_m)$;
for j **from** m **by** -1 to 2 **do**
for k **from** $j-1$ **by** -1 to 1 **do**
    $A_k \;:=\; \mathbf{R}(A_k, A_j, y_j)$;
return(A_1).

The polynomials obtained need not be minimal, in which case, one should appeal to multivariate polynomial factorization in order to select the relevant factors at each stage. (Groebner bases provide a neater alternative to these questions, see [**86**].)

Computer algebra systems usually provide implementations of both resultants and Groebner bases. The complexity of elimination is however exponential in the worst-case: degrees essentially multiply, which is somewhat intrinsic as y_0 in the quadratic system of k equations

$$y_0 - z - y_k = 0,\; y_k - y_{k-1}^2 = 0,\; \ldots\; , y_1 - y_0^2 = 0$$

(determining the OGF of regular trees of degree 2^k) represents an algebraic function of degree 2^k and no less.

$\triangleright$ **1.** *Resultant and roots.* Let $P, Q \in \mathbb{C}[x]$ have sets of roots $\{\alpha_j\}$ and $\{\beta_k\}$ respectively. Then

$$\mathbf{R}(P, Q, x) = a_0^\ell b_0^m \prod_{i=1}^{\ell} \prod_{j=1}^{m} (\alpha_i - \beta_j) = a_0^\ell \prod_{i=1}^{m} Q(\alpha_i).$$

The *discriminant* of P classically defined by $D(P) := a_0^{-1}\,\mathbf{R}(P(x), P'(x), x)$ satisfies

$$D(P) \equiv a_0^{-1}\,\mathbf{R}(P(x), P'(x), x) = a_0^{2\ell-2}\prod_{i\neq j}(\alpha_i - \alpha_j).$$

Given the coefficients of P and the value of $D(P)$, there results an effectively computable bound on the minimal separation distance δ between any two roots of P. [Hint. Let $A = 1 + \max_j(|a_j/a_0|)$. Then each α_j satisfies $|\alpha_j| < mA$. Set $L = \binom{\ell}{2}$. Then $\delta \geq |a_0|^{2-2\ell}|D(P)|(2A)^{L-1}$.] $\lhd$

2. Equivalent definitions of analyticity. Two parallel notions are introduced at the beginning of Chapter IV: analyticity (defined by power series expansions) and holomorphy (defined as complex differentiability). As is known from any textbook on complex analysis, these notions are equivalent. Given their importance for analytic combinatorics, this appendix entry sketches a proof of the equivalence, which is summarized by the following diagram:

$$\begin{array}{ccc} \text{Analyticity} & \begin{array}{c}\xrightarrow{[A]}\\ \xleftarrow[{[C]}]{}\end{array} & \mathbb{C}\text{-differentiability} \\ & & \downarrow\ [B] \\ & & \text{Null integral Property}\end{array}$$

A. Analyticity implies complex-differentiability. Let $f(z)$ be analytic in the disc $D(z_0; R)$. We may assume without loss of generality that $z_0 = 0$ and $R = 1$ (else effect a linear transformation on the argument z). According to the definition of analyticity, the series representation

$$f(z) = \sum_{n=0}^{\infty} f_n z^n, \tag{3}$$

converges for all z with $|z| < 1$. Elementary series rearrangements first entail that $f(z)$ given by this representation is analytic at any z_1 interior to $D(0; 1)$. Similar techniques then show the existence of the derivative as well as the fact that the derivative can be obtained by term-wise differentiation of (3).

$\rhd$ **2.** *Proof of* $[A]$*: Analyticity implies differentiability.* First, formally, the binomial theorem provides

$$\begin{aligned} f(z) &= \sum_{n\geq 0} f_n z^n \quad = \quad \sum_{n\geq 0} f_n(z_1 + z - z_1)^n \\ &= \sum_{n\geq 0}\sum_{k=0}^{n}\binom{n}{k} f_n z_1^k (z-z_1)^{n-k} \\ &= \sum_{m\geq 0} c_m (z-z_1)^m, \qquad c_m := \sum_{k\geq 0}\binom{m+k}{k} f_{m+k} z_1^k. \end{aligned} \tag{4}$$

Let r_1 be any number smaller than $1 - |z_1|$. We observe that (4) makes analytic sense. Indeed, one has the bound $|f_n| \leq CA^n$, valid for any $A > 1$ and some $C > 0$. Thus, the terms in (4) are dominated in absolute value by those of the double series

$$\sum_{n\geq 0}\sum_{k=0}^{n}\binom{n}{k} CA^n |z_1|^k r_1^{n-k} = C\sum_{n\geq 0} A^n(|z_1| + r_1)^n = \frac{C}{1 - A(|z_1| + r_1)}, \tag{5}$$

which is absolutely convergent as soon as A is chosen such that $A < (|z_1| + r_1)^{-1}$.

Complex differentiability of at any $z_1 \in D(0;1)$ derives from the analogous calculation, valid for small enough δ,

$$\begin{aligned}\frac{1}{\delta}(f(z_1+\delta)-f(z_1))) &= \sum_{n\geq 0} n f_n z_1^{n-1} + \delta \sum_{n\geq 0}\sum_{k=2}^{n}\binom{n}{k} f_n z_1^k \delta^{n-k-2} \\ &= \sum_{n\geq 0} n f_n z_1^{n-1} + O(\delta),\end{aligned} \tag{6}$$

where boundedness of the coefficient of δ results from an argument analogous to (5). ◁

The argument of Note 2 has shown that the derivative of f at z_1 is obtained by differentiating termwise the series representing f. More generally derivatives of all orders exist and can be obtained in a similar fashion. In view of this fact, the equalities of (4) can also be interpreted as the *Taylor expansion* (by grouping terms according to values of k first):

$$f(z_1+\delta) = f(z_1) + \delta f'(z_1) + \frac{\delta^2}{2!} f''(z_1) + \cdots, \tag{7}$$

which is thus generally valid for analytic functions.

B. Complex differentiability implies the "Null Integral" Property. The Null Integral Property relative to a domain Ω is the property:

$$\int_\lambda f = 0 \qquad \text{for any loop } \lambda \subset \Omega.$$

(A loop is a closed path that can be contracted to a single point in the domain Ω, cf Chapter IV). Its proof results simply from the Cauchy-Riemann equations and from Green's formula.

▷ **3.** *Proof of* $[B]$*: the Null Integral Property.* This starts from the *Cauchy–Riemann equations*. Let $P(x,y) = \Re f(x+iy)$ and $Q(x,y) = \Im f(x+iy)$. By adopting successively in the definition of complex differentiability $\delta = h$ and $\delta = ih$, one finds $P'_x + iQ'_x = Q'_y - iP'_y$, implying

$$\frac{\partial P}{\partial x} = \frac{\partial Q}{\partial y} \qquad \text{and} \qquad \frac{\partial P}{\partial y} = -\frac{\partial Q}{\partial x}, \tag{8}$$

known as the Cauchy–Riemann equations. (The functions P and Q satisfy the partial differential equations $\Delta f = 0$, where Δ is the 2-dimensional *Laplacian* $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$; such functions are known as *harmonic functions*.) The Null Integral Property, given differentiability, results from the Cauchy–Riemann equations, upon taking into account Green's theorem of multivariate calculus,

$$\int_{\partial K} A dx + B dy = \int\int_K \left(\frac{\partial B}{\partial x} - \frac{\partial A}{\partial y}\right) dx\, dy,$$

which is valid for any (compact) domain K enclosed by a simple curve ∂K. ◁

C. Complex differentiability implies analyticity. The starting point is the formula

$$f(a) = \frac{1}{2i\pi}\int_\gamma \frac{f(z)}{z-a}\, dz, \tag{9}$$

knowing only differentiability of f and its consequence, the Null Integral Property (but precisely *not* postulating the existence of an analytic expansion). There γ is a simple positive loop encircling a inside a region where f is analytic.

$\triangleright$ **4.** *Proof of* $[C]$*: the integral representation.* The proof of (9) is obtained by decomposing $f(z)$ in the original integral as $f(z) = f(z) - f(a) + f(a)$. Define accordingly

$$g(z) = \begin{cases} \frac{f(z)-f(a)}{z-a} & \text{for} \quad z \neq a \\ f'(a) & \text{for} \quad z = a. \end{cases}$$

By the differentiability assumption, g is continuous and holomorphic (differentiable) at any point other than a. Its integral is thus 0 along γ. On the other hand, we have

$$\int_\gamma \frac{1}{z-a}\, dz = 2i\pi,$$

by a simple computation: deform γ to a small circle along a and evaluate the integral directly by setting $z - a = re^{i\theta}$. $\triangleleft$

Once (9) is granted, it suffices to write, e.g., for an expansion at 0,

$$\begin{aligned} f(z) &= \frac{1}{2i\pi}\int_\gamma f(t)\,\frac{dt}{t-z} \\ &= \frac{1}{2i\pi}\int_\gamma f(t)\left(1 + \frac{z}{t} + \frac{z^2}{t^2} + \cdots\right)\frac{dt}{t} \\ &= \sum_{n\geq 0} f_n z^n, \qquad f_n := \frac{1}{2i\pi}\int_\gamma f(t)\,\frac{dt}{t^{n+1}}. \end{aligned}$$

(Exchanges of integration and summation are justified by normal convergence.) Analyticity is thus proved from complex-differentiability and its consequence the Null Integral Property.

$\triangleright$ **5.** *Cauchy's formula for derivatives.* One has

$$f^{(n)}(a) = \frac{n!}{2i\pi}\int_\gamma \frac{f(z)}{(z-a)^{n+1}}\, dz.$$

This follows from (9) by differentiation under the integral sign. $\triangleleft$

$\triangleright$ **6.** *Morera's Theorem.* Suppose that f is continuous [but not *a priori* known to be differentiable] in an open set Ω and that its integral along any triangle in Ω is 0. Then, f is analytic (hence holomorphic) in Ω. [For a proof, see, e.g, [**354**, p. 68].] $\triangleleft$

3. Gamma function. The formulæ of singularity analysis in Chapter IV involve the *Gamma function* in an essential manner. The Gamma function extends to nonintegral arguments the factorial function and we collect in this appendix a few classical facts regarding it. Proofs may be found in classic treatises like Henrici's [**229**] or Whittaker and Watson's [**433**].

Basic properties. Euler introduced the Gamma function as

$$\Gamma(s) = \int_0^\infty e^{-t}t^{s-1}\, dt, \tag{10}$$

where the integral converges provided $\Re(s) > 0$. Through integration by parts, one immediately derives the basic functional equation of the Gamma function,

$$\Gamma(s+1) = s\Gamma(s). \tag{11}$$

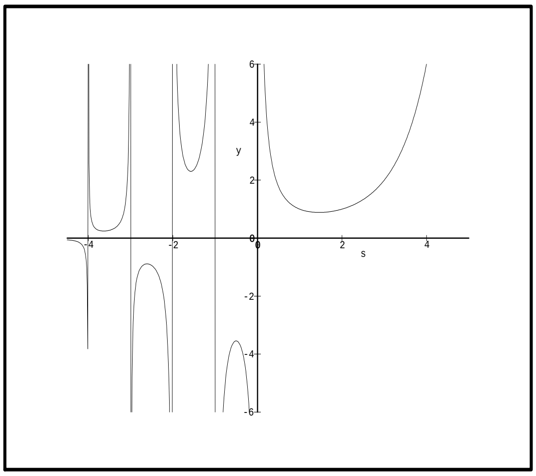


FIGURE 1. A plot of $\Gamma(s)$ for real s.

Since $\Gamma(1) = 1$, one has $\Gamma(n + 1) = n!$, so that the Gamma function serves to extend the factorial function for nonintegral arguments. For combinatorial purposes, the special value,

$$\Gamma\left(\frac{1}{2}\right) := \int_0^\infty e^{-t} \frac{dt}{\sqrt{t}} = 2\int_0^\infty e^{-x^2}\,dx = \sqrt{\pi}, \tag{12}$$

proves to be quite important. It implies in turn $\Gamma(-\frac{1}{2}) = -2\sqrt{\pi}$.

From (11), the Gamma function can be analytically continued to the whole of $\mathbb{C}$ with the exception of poles at $0, -1, -2, \ldots$. The functional equation used backwards yields

$$\Gamma(s) \sim \frac{(-1)^m}{m!}\frac{1}{s+m} \qquad (s \to -m)\,,$$

so that the residue of $\Gamma(s)$ at $s = -m$ is $(-1)^m/m!$. Figure 1 depicts the graph of $\Gamma(s)$ for real values of s.

$\triangleright$ **7.** *Evaluation of the Gaussian integral.* Define $J := \int_0^\infty e^{-x^2}\,dx$. The idea is to evaluate J^2:

$$J^2 = \int_0^\infty \int_0^\infty e^{-(x^2+y^2)}\,dxdy.$$

Going to polar coordinates, $(x^2+y^2)^{1/2} = \rho$, $x = \rho\cos\theta$, $y = \rho\sin\theta$ yields, via the standard change of variables formula:

$$J^2 = \int_0^\infty \int_0^{\frac{\pi}{2}} e^{-\rho^2}\,\rho d\rho d\theta.$$

The equality $J^2 = \pi/4$ results. $\triangleleft$

Hankel contour representation. Euler's integral representation of $\Gamma(s)$ used in conjunction with the functional equation permits us to continue $\Gamma(s)$ to the whole of the complex plane. A direct approach due to Hankel provides an alternative integral representation valid for all values of s.

THEOREM B.1 (Hankel's contour integral). *Let $\int_{+\infty}^{(0)}$ denote an integral taken along a contour starting at $+\infty$ in the upper plane, winding counterclockwise around the origin, and proceeding towards $+\infty$ in the lower half plane. Then, for all $s \in \mathbb{C}$,*

$$\frac{1}{\pi}\sin(\pi s)\Gamma(1-s) = \frac{1}{\Gamma(s)} = -\frac{1}{2i\pi}\int_{+\infty}^{(0)}(-t)^{-s}e^{-t}\,dt. \tag{13}$$

In (13), $(-t)^{-s}$ is assumed to have its principal determination when t is negative real, and this determination is then extended uniquely by continuity throughout the contour. The integral then closely resembles the definition of $\Gamma(1-s)$. The first form of (13) can also be rewritten as $\frac{1}{\Gamma(s)}$, by virtue of the complement formula given below.

▷ **8.** *Proof of Hankel's representation.* We refer to volume 2 of Henrici's book [**229**, p. 35] or Whittaker and Watson's treatise [**433**, p. 245] for a detailed proof.

A contour of integration that fulfills the conditions of the theorem is typically the contour $\mathcal{H}$ that is at distance 1 of the positive real axis comprising three parts: a line parallel to the positive real axis in the upper half–plane; a connecting semi–circle centered at the origin; a line parallel to the positive real axis in the lower half–plane. More precisely, $\mathcal{H} = \mathcal{H}^- \cup \mathcal{H}^+ \cup \mathcal{H}^\circ$, where

$$\left\{\begin{array}{lcl} \mathcal{H}^- & = & \{z = w - i,\ w \geq 0\} \\ \mathcal{H}^+ & = & \{z = w + i,\ w \geq 0\} \\ \mathcal{H}^\circ & = & \{z = -e^{i\phi},\ \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}. \end{array}\right. \tag{14}$$

Let ϵ be a small positive real number, and denote by $\epsilon \cdot \mathcal{H}$ the image of $\mathcal{H}$ by the transformation $z \mapsto \epsilon z$. By analyticity, for the integral representation, we can equally well adopt as integration path the contour $\epsilon \cdot \mathcal{H}$, for any $\epsilon > 0$. The main idea is then to let ϵ tend to 0.

Assume momentarily that $s < 0$. (The extension to arbitrary s then follows by analytic continuation.) The integral along $\epsilon \cdot \mathcal{H}$ decomposes into three parts:

> The integral along the semi–circle is 0 if we take the circle of a vanishing small radius, since $-s > 0$.
> The contributions from the upper and lower lines give, as $\epsilon \to 0$
>
> $$\int_{+\infty}^{(0)}(-t)^{-s}e^{-t}\,dt = (-U+L)\int_0^\infty t^{-s}e^{-t}\,dt$$
>
> where U and L denote the determinations of $(-1)^{-s}$ on the half-lines lying in the upper and lower half planes respectively.

By continuity of determinations, $U = (e^{-i\pi})^{-s}$ and $L = (e^{+i\pi})^{-s}$. Therefore, the right hand side of (13) is equal to

$$-\frac{(-e^{i\pi s} + e^{-i\pi s})}{2i\pi}\Gamma(1-s) = \frac{\sin(\pi s)}{\pi}\Gamma(1-s),$$

which completes the proof of the theorem. ◁

Expansions. The Gamma function has poles at the nonpositive integers but has no zeros. Accordingly, $1/\Gamma(s)$ is an entire function with zeros at $0, -1, \ldots$, and the position of the zeros is reflected by the product decomposition,

$$\frac{1}{\Gamma(s)} = se^{\gamma s} \prod_{n=1}^{\infty} \left[(1 + \frac{s}{n})e^{-s/n} \right] \tag{15}$$

(of the so–called Weierstraß type). There $\gamma = 0.57721$ denotes Euler's constant

$$\gamma = \lim_{n\to\infty} (H_n - \log n) \equiv \sum_{n=1}^{\infty} \left[\frac{1}{n} - \log(1 + \frac{1}{n}) \right].$$

The logarithmic derivative of the Gamma function is classically known as the psi function and is denoted by $\psi(s)$:

$$\psi(s) := \frac{d}{ds} \log \Gamma(s) = \frac{\Gamma'(s)}{\Gamma(s)}.$$

In accordance with (15), $\psi(s)$ admits a partial fraction decomposition

$$\psi(s+1) = -\gamma - \sum_{n=1}^{\infty} \left[\frac{1}{n+s} - \frac{1}{n} \right]. \tag{16}$$

From (16), there results that the Taylor expansion of $\psi(s+1)$, hence of $\Gamma(s+1)$, involves values of the Riemann zeta function,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

at the positive integers: for $|s| < 1$,

$$\psi(s+1) = -\gamma + \sum_{n=2}^{\infty} (-1)^n \zeta(n) s^{n-1}.$$

so that the coefficients in the expansion of $\Gamma(s)$ around any integer are polynomially expressible in terms of Euler's constant γ and values of the zeta function at the integers. For instance, as $s \to 0$,

$$\Gamma(s+1) = 1 - \gamma\, s + \left(\frac{\pi^2}{12} + \frac{\gamma^2}{2} \right) s^2 + \left(-\frac{\zeta(3)}{3} - \frac{\pi^2\gamma}{12} - \frac{\gamma^3}{6} \right) s^3 + O(s^4).$$

Another direct consequence of the infinite product formulæ for $\Gamma(s)$ and $\sin \pi s$ is the complement formula for the Gamma function,

$$\Gamma(s)\Gamma(-s) = -\frac{\pi}{s \sin \pi s}, \tag{17}$$

which directly results from the factorization of the sine function (due to Euler),

$$\sin s = s \prod_{n=1}^{\infty} \left(1 - \frac{s^2}{n^2\pi^2} \right).$$

In particular, Equation (17) gives back the special value (cf (12)): $\Gamma(\frac{1}{2}) = \sqrt{\pi}$.

$\triangleright$ **9.** *The duplication formula.* This is

$$2^{2s-1}\Gamma(s)\Gamma(s+\frac{1}{2}) = \pi^{1/2}\Gamma(2s),$$

which provides the expansion of Γ near $1/2$:

$$\Gamma(s+\frac{1}{2}) = \pi^{1/2} - (\gamma + 2\,\log 2)\,\pi^{1/2}s + \left(\frac{\pi^{5/2}}{4} + \frac{(\gamma+2\,\log 2)^2\,\pi^{1/2}}{2}\right)s^2 + O(s^3).$$

The coefficients now involve $\log 2$ as well as zeta values. $\triangleleft$

Finally, a famous and absolutely fundamental asymptotic formula is Stirling's approximation, familiarly known as "Stirling's formula":

$$\Gamma(s+1) = s\Gamma(s) \sim s^s e^{-s}\sqrt{2\pi s}\left[1 + \frac{1}{12s} + \frac{1}{288s^2} - \frac{139}{51840s^3} + \cdots\right].$$

It is valid for (large) real $s > 0$, and more generally for all $s \to \infty$ in $|\operatorname{Arg}(s)| < \pi - \delta$ (any $\delta > 0$). For the purpose of obtaining effective bounds, the following quantitative relation [**433**, p. 253] often proves useful,

$$\Gamma(s+1) = s^s e^{-s}(2\pi s)^{1/2} e^{\theta/(12s)}, \qquad \text{where } 0 < \theta \equiv \theta(s) < 1,$$

an equality that holds now for all $s \geq 1$. Stirling's formula is usually proved by appealing to the method of Laplace applied to the integral representation for $\Gamma(s+1)$, see APPENDIX B: *Laplace's method*, p. 667, or by Euler-Maclaurin summation (Note 10). It is derived by different means in APPENDIX B: *Mellin transform*, p. 674.

$\triangleright$ **10.** *Stirling's formula via Euler-Maclaurin summation.* Stirling's formula can be derived from Euler–Maclaurin summation applied to $\log\Gamma(s)$. [See: [**212**, Sec. 9.6].] $\triangleleft$

$\triangleright$ **11.** *The Eulerian Beta function.* It is defined for $\Re(p), \Re(q) > 0$ by any of the following integrals,

$$B(p,q) := \int_0^1 x^{p-1}(1-x)^{q-1}\,dx = \int_0^\infty \frac{y^{p-1}}{(1+y)^{p+q}}\,dy = 2\int_0^{\frac{\pi}{2}} \cos^{2p-1}\theta\,\sin^{2q-1}\theta\,d\theta,$$

where the last form is known as a Wallis integral. It satisfies:

$$B(p,q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}.$$

[See [**433**, p. 254] for a proof generalizing that of Note 7.] $\triangleleft$

4. Implicit Function Theorem. In its real variable version, the implicit function theorem asserts that, for a sufficiently smooth function $F(z,w)$ of two variables, a solution to the the equation $F(z,w) = 0$ exists in the vicinity of a solution point (z_0, w_0) (therefore satisfying $F(z_0,w_0) = 0$) provided the partial derivative satisfies $F'_w(z_0,w_0) \neq 0$. This theorem admits a complex-analytic extension, which is essential for the analysis of recursive structures.

Without loss of generality, one restricts attention to $(z_0,w_0) = (0,0)$. We consider here a function $F(z,w)$ that is analytic in two complex variables in the sense that it admits a convergent representation valid in a polydisc,

$$F(z,w) = \sum_{m,n\geq 0} f_{m,n} z^m w^n, \qquad |z| < R, \quad |w| < S. \tag{18}$$

for some $R, S > 0$ (cf APPENDIX B: *Several complex variables.*, p. 680).

THEOREM B.2 (Analytic Implicit Functions). *Let F be bivariate analytic near $(0,0)$. Assume that $F(0,0) \equiv f_{0,0} = 0$ and $F'_w(0,0) \equiv f_{0,1} \neq 0$. Then, there exists a unique function $f(z)$ analytic in a neighbourhood $|z| < \rho$ of 0 such that $f(0) = 0$ and*

$$F(z, f(z)) = 0, \qquad |z| < \rho.$$

▷ **12.** *Proofs of the Implicit Function Theorem.* See Hille's book [**232**] for details.

(i) Proof by residues. Make use of the principle of the argument and Rouché's Theorem to see that the equation $F(z,w)$ has a unique solution near 0 for $|z|$ small enough. Appeal then to the related result of Chapter IV (based on the residue theorem) that expresses the sum of the solutions to an equation as a contour integral. Here, this expresses the solution as (C a small enough contour around 0 in the w–plane)

$$f(z) = \frac{1}{2i\pi} \int_C w \frac{F'_w(z,w)}{F(z,w)}\, dw,$$

which is checked to represent an analytic function of z.

(ii) Proof by majorant series. Set $G(z,w) := w - f_{0,1}^{-1} F(z,w)$. The equation $F(z,w) = 0$ becomes the fixed-point equation $w = G(z,w)$. The bivariate series G has its coefficients dominated termwise by those of

$$\widehat{G}(z,w) = \frac{A}{(1-z/R)(1-w/S)} - A - A\frac{w}{S}.$$

The equation $w = \widehat{G}(z,w)$ is quadratic. It admits a solution $\widehat{f}(z)$ analytic at 0,

$$\widehat{f}(z) = A\frac{z}{R} + \frac{A(A^2 + AS + S^2)}{S^2}\frac{z^2}{R^2} + \cdots,$$

whose coefficients dominate termwise those of f.

(iii) Proof by Picard's method of successive approximants. With G like before, define the sequence of functions

$$\phi_0(z) := 0; \quad \phi_{j+1}(z) = G(z, \phi_j(z)),$$

each analytic in a small neighbourhood of 0. Then $f(z)$ can be obtained as

$$f(z) = \lim_{j\to\infty} \phi_j(z) \equiv \phi_0(z) - \sum_{j=0}^{\infty} \left(\phi_j(z) - \phi_{j+1}(z)\right),$$

which is itself checked to be analytic near 0 by the geometric convergence of the series. ◁

Weierstrass Preparation. The Weierstrass Preparation Theorem (WPT) also known as *Vorbereitungssatz* is a useful complement to the Implicit Function Theorem.

Given a collection $Z = (z_1, \ldots, z_m)$ of variables, we designate as usual by $\mathbb{C}[[Z]]$ the ring of formal power series in indeterminates Z. We let $C\{Z\}$ denote the subset of these that are convergent in a neighbourhood of $(0, \ldots, 0)$, i.e., analytic (cf APPENDIX B: *Several complex variables.*, p. 680).

THEOREM B.3 (Weierstraß Preparation). *Let $f = f(z_1, \ldots, z_m)$ in $\mathbb{C}[[z_1, \ldots, z_m]]$ (respectively $\mathbb{C}\{Z\}$) be such that $f(0, \ldots, 0) = 0$. A Weierstraß polynomial is a polynomial of the form*

$$z^d + g_1 z^{d-1} + \cdots + g_d,$$

where $g_j \in \mathbb{C}[[z_2, \ldots, z_m]]$ *(respectively* $g_j \in \mathbb{C}\{z_2, \ldots, z_m\}$*) and* $gj(0, \ldots, 0) = 0$*. Assume that* f *depends on at least one of the* z_j *with* $j \geq 2$ *(i.e.,* $f(0, z_2, \ldots, z_m)$ *is not identically 0). Then,* f *admits a unique factorization*

$$f(z_1, z_2, \ldots, z_m) = W(z_1) \cdot F(z_1, \ldots, z_m),$$

where $W(z)$ *is a Weierstraß polynomial and* F *is an element of* $\mathbb{C}[[z_1, \ldots, z_m]]$ *(respectively* $\mathbb{C}\{z_1, \ldots, z_m\}$*) satisfying* $F(0, 0 \ldots, 0) \neq 0$*.*

An accessible proof and a discussion are found in Abhyankar's lecture notes [**1**, Ch. 16]. In essence, Theorem B.3 implies that functions implicitly defined by a transcendental equation (an equation $f = 0$) are locally of the same nature as algebraic functions (corresponding to the equation $W = 0$). In particular, for $m = 2$, when the solutions have singularities, these singularities can only be branch points and companion Puiseux expansions hold (Chapter VII). The theorem acquires even greater importance when perturbative singular expansions (corresponding to $m \geq 3$) become required for the purpose of extracting limit laws (Chapter IX).

5. Laplace's method. The method of Laplace serves to estimate asymptotically *real* integrals depending on a large parameter n (which may be a positive integer or real number). Though it is primarily a real analysis technique, we present it in detail in this appendix given its relevance to the saddle point method, which deals instead with *complex* contour integrals.

Case study: a Wallis integral. In order to demonstrate the essence of the method, consider first the problem of estimating asymptotically the Wallis integral

$$I_n := \int_{-\pi/2}^{\pi/2} (\cos x)^n \, dx, \tag{19}$$

as $n \to +\infty$. The cosine attains its maximum at $x = 0$ (where its value is 1), and since the integrand of I_n is a large power, the contribution to the integral outside any fixed segment containing 0 is exponentially small and can consequently be discarded for all asymptotic purposes. A glance at the plot of $\cos^n x$ as n varies (Figure 2) also

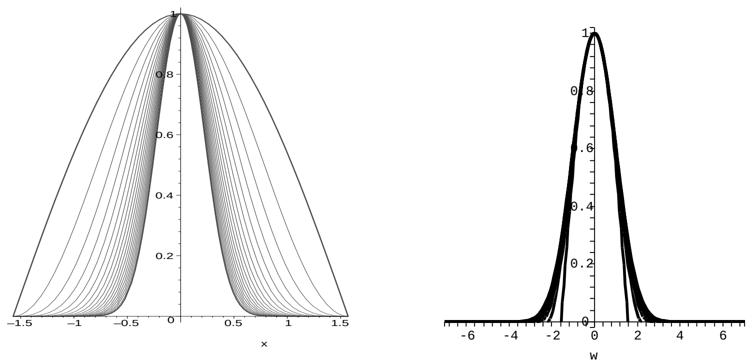


FIGURE 2. Plots of $\cos^n x$ [left] and $\cos^n(w/\sqrt{n})$ [right], for $n = 1 \ldots 20$.

$$\begin{array}{lcll}
\displaystyle\int_{-\pi/2}^{\pi/2} \cos^n x\,dx & = & \displaystyle\frac{1}{\sqrt{n}}\int_{-\frac{\pi}{2}\sqrt{n}}^{\frac{\pi}{2}\sqrt{n}} \left(\cos\frac{w}{\sqrt{n}}\right)^n dw & \text{Set } x = w/\sqrt{n}\text{; choose } \kappa_n = n^{1/10} \\
 & \sim & \displaystyle\frac{1}{\sqrt{n}}\int_{-\kappa_n}^{\kappa_n} \left(\cos\frac{w}{\sqrt{n}}\right)^n dw & [\textit{Neglect the tails}] \\
 & \sim & \displaystyle\frac{1}{\sqrt{n}}\int_{-\kappa_n}^{\kappa_n} e^{-w^2/2} dw & [\textit{Central approxim.}] \\
 & \sim & \displaystyle\frac{1}{\sqrt{n}}\int_{-\infty}^{\infty} e^{-w^2/2} dw & [\textit{Complete the tails}] \\
 & \sim & \displaystyle\sqrt{\frac{2\pi}{n}}. &
\end{array}$$

FIGURE 3. A typical application of the Laplace method.

suggests that the integrand tends to conform to a bell-shaped profile near the centre as n increases. This is not hard to verify: set $x = w/\sqrt{n}$, then a local expansion yields

$$\cos^n x \equiv \exp(n \log\cos(x)) = \exp\left(-\frac{w^2}{2} + O(n^{-1}w^4)\right), \tag{20}$$

the approximation being valid as long as $w = O(n^{1/4})$. Accordingly, we choose (somewhat arbitrarily)

$$\kappa_n := n^{1/10},$$

and define the central range by $|w| \le \kappa_n$. These considerations suggest to rewrite the integral I_n as

$$I_n = \frac{1}{\sqrt{n}}\int_{-\pi\sqrt{n}/2}^{+\pi\sqrt{n}/2} \left(\cos\frac{w}{\sqrt{n}}\right)^n dw,$$

and expect under this new form an approximation by a Gaussian integral arising from the central range.

Laplace's method proceeds in three steps:

(i) *Neglect the tails of the original integral*;
(ii) *Centrally approximate the integrand by a Gaussian*;
(iii) *Complete the tails of the Gaussian integral*.

In the case of the cosine integral (19), the chain is summarized in Figure 3. Details of the analysis follow.

(i) *Neglect the tails of the original integral*: By (20), we have

$$\cos^n\left(\frac{\kappa_n}{\sqrt{n}}\right) \sim \exp\left(-\frac{1}{2}n^{1/5}\right),$$

and, as the integrand is unimodal, this exponentially small quantity bounds the integrand throughout $|w| > \kappa_n$, that is, on a large part of the integration interval. This gives

$$I_n = \frac{1}{\sqrt{n}}\int_{-\kappa_n/\sqrt{n}}^{+\kappa_n/\sqrt{n}} \cos^n x\,dx + O\left(\exp\left(-\frac{1}{2}\kappa_n^2\right)\right), \tag{21}$$

and the error term is of the order of $\exp(-\frac{1}{2}n^{1/5})$.

(ii) *Centrally approximate the integrand by a Gaussian*: In the central region, we have

$$
\begin{aligned}
I_n^{(1)} &:= \int_{-\kappa_n/\sqrt{n}}^{+\kappa_n/\sqrt{n}} \cos^n x \, dx \\
&= \frac{1}{\sqrt{n}} \int_{-\kappa_n}^{+\kappa_n} e^{-w^2/2} \exp\left(O(n^{-1}w^4)\right) \, dw \\
&= \frac{1}{\sqrt{n}} \int_{-\kappa_n}^{+\kappa_n} e^{-w^2/2} \left(1 + O(n^{-1}w^4)\right) \, dw \\
&= \frac{1}{\sqrt{n}} \int_{-\kappa_n}^{+\kappa_n} e^{-w^2/2} \, dw + O(n^{-3/2}),
\end{aligned} \tag{22}
$$

given the uniformity of approximation (20) for w in the integration interval.

(iii) *Complete the tails of the Gaussian integral*: The incomplete Gaussian integral in (22) can be easily estimated once it is observed that its tails are small. Precisely, one has, for $W \geq 0$,

$$
\int_W^\infty e^{-w^2/2} \, dw \leq e^{-W^2/2} \int_0^\infty e^{-h^2/2} \, dh \equiv \sqrt{\frac{\pi}{2}} e^{-W^2/2}
$$

(by the change of variable $w = W + h$). Thus,

$$
\int_{-\kappa_n}^{+\kappa_n} e^{-w^2/2} \, dw = \int_{-\infty}^{+\infty} e^{-w^2/2} \, dw + O\left(\exp\left(-\frac{1}{2}\kappa_n^2\right)\right). \tag{23}
$$

It now suffices to collect the three approximations, (21), (22), and (23): we have obtained in this way.

$$
I_n = \frac{1}{\sqrt{n}} \int_{-\infty}^{+\infty} e^{-w^2/2} \, dw + O(n^{-3/2}) \equiv \sqrt{\frac{2\pi}{n}} + O(n^{-3/2}). \tag{24}
$$

These three steps are the heart of Laplace's method.

In the asymptotic scale of the problem, the exponentially small errors in the tails can be completely neglected. The error in (24) then arises from the central approximation (20), and its companion $O(w^4 n^{-1})$ term. This can easily be improved and it suffices to appeal to further terms in the expansion of $\log \cos x$ near 0. For instance, one has ($x = w/\sqrt{n}$):

$$
\cos^n x = e^{-w^2/2} \left(1 - \frac{w^4}{12n} + O(n^{-2}w^8)\right). \tag{25}
$$

Proceeding like before, we find that a further term in the expansion of I_n is obtained by considering the additive correction

$$
\epsilon_n := -\frac{1}{\sqrt{n}} \int_{-\infty}^{+\infty} e^{-w^2/2} \left(\frac{w^4}{12n}\right) \, dw \equiv -\sqrt{\frac{\pi}{8n^3}},
$$

so that

$$
I_n = \sqrt{\frac{2\pi}{n}} - \sqrt{\frac{\pi}{8n^3}} + O(n^{-5/2}).
$$

Clearly, a full asymptotic expansion can be obtained in this manner.

$\triangleright$ **13.** *Wallis integrals and central binomials.* The integral I_n is an integral considered by John Wallis (1616–1703). It can be evaluated through partial integration or by its relation to the Beta integral (Note 11) as $I_n = \Gamma(\frac{1}{2})\Gamma(\frac{n}{2}+\frac{1}{2})/\Gamma(\frac{n}{2}+1)$. There results ($n \mapsto 2n$)

$$\binom{2n}{n} \sim \frac{2^{2n}}{\sqrt{\pi n}}\left(1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} - \cdots\right),$$

which is yet another avatar of Stirling's formula. $\triangleleft$

General case of large powers. Laplace's method applies under very general conditions to integrals involving large powers of a fixed function.

THEOREM B.4 (Laplace's method). *Let f and g be indefinitely differentiable real valued functions defined over some compact interval I of the real line. Assume that $|g(x)|$ attains its maximum at a unique point x_0 interior to I and that $f(x_0), g(x_0), g''(x_0) \neq 0$. Then, the integral*

$$I_n := \int_I f(x)g(x)^n\,dx$$

admits a full asymptotic expansion:

$$I_n \sim \sqrt{\frac{2\pi}{\lambda n}} f(x_0)g(x_0)^n \left(1 + \sum_{j\geq 1}\frac{\delta_j}{n^j}\right), \qquad \lambda := -\frac{g''(x_0)}{g(x_0)}. \tag{26}$$

$\triangleright$ **14.** *Proof of Laplace's Theorem.* It follows exactly the steps explained above. Let us asume first that $f(x) \equiv 1$. Then, one chooses κ_n as a function tending slowly to infinity like before ($\kappa_n = n^{1/10}$ is suitable). It suffices to expand

$$I_n^{(1)} := \int_{x_0-\kappa_n/\sqrt{n}}^{x_0+\kappa_n/\sqrt{n}} e^{n\log g(x)}\,dx,$$

as the difference $I_n - I_n^{(1)}$ is exponentially small. Set first $x = x_0 + X$ and

$$L(X) := \log g(x_0 + X) - \log g(x_0) + \lambda\frac{X^2}{2},$$

so that, with $w = X\sqrt{n}$, the central contribution becomes:

$$I_n^{(1)} = \frac{g(x_0)^n}{\sqrt{n}}\int_{-\kappa_n}^{\kappa_n} e^{-\lambda w^2/2} e^{nL(w/\sqrt{n})}\,dw.$$

Then, it is possible to expand $L(X)$ to any order M,

$$L(X) = \sum_{j=3}^{M-1} \ell_j X^j + O(X^M),$$

and $e^{nL(w/\sqrt{n})}$ admits a full expansion in descending powers of $\sqrt{n}$:

$$e^{nL(w/\sqrt{n})} \sim 1 + \frac{\ell_3 w^3}{\sqrt{n}} + \frac{2\ell_4 w^4 + \ell_3^2 w^6}{2n} + \cdots.$$

There, by construction, the coefficient of $n^{-k/2}$ is a polynomial $E_k(w)$ of degree $3k$. This expression can be truncated to any order, resulting in

$$I_n^{(1)} = \frac{g(x_0)^n}{\sqrt{n}}\int_{-\kappa_n}^{\kappa_n} e^{-\lambda w^2/2}\left(1 + \sum_{k=1}^{M-1}\frac{E_k(w)}{n^{k/2}} + O\left(\frac{1+w^{3M}}{n^{M/2}}\right)\right)dw.$$

One can then complete the tails at the expense of exponentially small terms since the Gaussian tails are exponentially small.

The full asymptotic expansion is revealed by the following device: for any power series $h(w)$, introduce the Gaussian transform,

$$\mathfrak{G}[f] := \int_0^\infty e^{-w^2/2} f(w)\, dw,$$

which is understood to operate by linearity on integral powers of w,

$$\mathfrak{G}[w^{2r}] = 1 \cdot 3 \cdots (2r-1)\sqrt{2\pi}, \qquad \mathfrak{G}[w^{2r+1}] = 0.$$

Then, the complete asymptotic expansion of I_n is obtained by the formal expansion

$$(27)\quad \frac{g(x_0)^n}{\sqrt{n\lambda}} \cdot \mathfrak{G}\left[\exp\left(\lambda^{-3/2} w^3 y \widetilde{L}(\lambda^{-1/2} wy)\right)\right], \qquad \widetilde{L}(X) := \frac{1}{X^3} L(X), \quad y \mapsto \frac{1}{\sqrt{n}}.$$

The addition of the prefactor $f(x)$ (omitted so far) induces a factor $f(x_0)$ in the in the main term of the final result and it affects the coefficients in the smaller order terms in a computable manner. Details are left as an exercise to the reader. $\lhd$

$\rhd$ **15.** *The next term?* One has (with $f_j := f^{(j)}(x_0)$, etc):

$$\frac{I_n\sqrt{\lambda n}}{\sqrt{2\pi}g(x_0)^n} = f_0 + \frac{-9\lambda^3 f_0 + 12\lambda^2 f_2 + 12\lambda f_1 g_3 + 3\lambda f_0 g_4 + 5g_3^2 f_0}{24\lambda^3 n} + O(n^{-2}),$$

which is best determined using a symbolic manipulation system. $\lhd$

The method is susceptible of a large number of extensions. Roughly it requires a point where the integrand is maximized, which induces some sort of exponential behaviour, local expansions then allowing for a replacement by standard integrals.

$\rhd$ **16.** *Special cases of Laplace's method.* When $f(x_0) = 0$, the integral normalizes to an integral of the form $\int w^2 e^{-w^2/2}$. If $g''(x_0) = 0$ but $g^{(iv)}(x_0) \neq 0$ then a factor $\Gamma(\frac{1}{4})$ replaces the characteristic $\sqrt{\pi} \equiv \Gamma(\frac{1}{2})$. [Hint: $\int_0^\infty \exp(-w^\beta) w^\alpha\, dw = \beta^{-1}\Gamma((\alpha+1)\beta^{-1})$.] If the maximum is attained at one end of the interval $I = [a, b]$ while $g'(x_0) = 0$, $g''(x_0) \neq 0$, then the estimate (26) must be multiplied by a factor of $\frac{1}{2}$. If the maximum is attained at one end of the interval I while $g'(x_0) \neq 0$, then the right normalization is $w = x/n$ and the integrand is reducible to an exponential e^{-w}. Here are some dominant asymptotic terms:

$x_0 \neq a, b$	$g''(x_0) \neq 0, f(x_0) = 0$	$\sqrt{\frac{\pi}{2\lambda^5 n^3}} g(x_0)^n (\lambda f''(x_0) + f'(x_0) g'''(x_0))$
$x_0 \neq a, b$	$g''(x_0) = 0, g^{(iv)}(x_0) \neq 0$	$\Gamma(\frac{1}{4}) \sqrt[4]{\frac{3}{2\lambda^\star n}} f(x_0) g(x_0)^n \quad \left(\lambda^\star = -\frac{g^{(iv)}(x_0)}{g(x_0)}\right)$
$x_0 = a$	$f(x_0) \neq 0, g'(x_0) \neq 0$	$-\frac{1}{ng'(x_0)} f(x_0) g(x_0)^{n+1}$.

A similar analysis is employed in Chapter VIII, when we discuss coalscence cases of the saddle-point method. $\lhd$

EXAMPLE 1. *Stirling's formula via Laplace's method.* Start from an integral representation involving $n!$, namely,

$$I_n := \int_0^\infty e^{-nx} x^n\, dx = \frac{n!}{n^{n+1}}\,.$$

This is a direct case of application of the theorem, except for the fact that the integration interval is not compact. The integrand attains its maximum at $x_0 = 1$ and the remainder integral $\int_2^\infty$ is accordingly exponentially small as proved by the chain

$$\begin{aligned}\int_2^\infty e^{-nx}x^n\,dx &= (2e^{-2})^n\int_0^\infty \left(1+\frac{x}{2}\right)^n e^{-nx}\,dx & [x\mapsto x+2]\\ &< (2e^{-2})^n\int_0^\infty e^{nx/2}e^{-nx}\,dx=\frac{2}{n}(2e^{-2})^n & [\log(1+x/2)<x/2].\end{aligned}$$

Then the integral from 0 to 2 is amenable to the standard version of Laplace's method as stated in Theorem B.4 to the effect that

$$n! = n^n e^{-n}\sqrt{2\pi n}\left(1+O\left(\frac{1}{n}\right)\right).$$

The asymptotic expansion of I_n derives from (27) and involves the combinatorial GF

$$H(z,u) := \exp\left(u\left(\log(1-z)^{-1}-z-\frac{z^2}{2}\right)\right). \tag{28}$$

The noticeable fact is that $H(z,u)$ is the exponential BGF of permutations that are generalized derangements involving no cycles of length 1 or 2, with z marking size and u marking the number of cycles:

$$H(z,u)=\sum_{n,k\geq 0} h_{n,k}u^k\frac{z^n}{n!} = 1+\tfrac{1}{3}uz^3+\tfrac{1}{4}\,uz^4+\tfrac{1}{5}\,uz^5+(\tfrac{1}{6}u+\tfrac{1}{18}u^2)z^6+(\tfrac{1}{7}u+\tfrac{1}{12}u^2)z^7+\cdots.$$

Then, a full asymptotic expansion of I_n is obtained by applying the Gaussian transform $\mathfrak{G}$ to $H(wy,-y^{-2})$ (with $y=n^{-1/2}$), resulting in

$$n! \sim n^n e^{-n}\sqrt{2\pi n}\left(1+\frac{1}{12n}+\frac{1}{288n^2}-\frac{139}{51840n^3}-\cdots\right).$$

PROPOSITION B.1 (Stirling's formula). *The factorial function admits the complete asymptotic expansion as $x\to+\infty$:*

$$x! \equiv \Gamma(x+1) \sim x^x e^{-x}\sqrt{2\pi x}\left(1+\sum_{q\geq 1}\frac{c_q}{x^q}\right).$$

The coefficients satisfy $c_q = \displaystyle\sum_{k=1}^{2q}\frac{(-1)^k}{2^{q+k}(q+k)!}h_{2q+2k,k}$, *where $h_{n,k}$ counts the number of permutations of size n having k cycles, all of length ≥ 3.*

The derivation above is due to Wrench (see [**82**, p. 267]). .... END OF EXAMPLE 1. □

The scope of the method goes much beyond the case of integrals of large powers. Roughly, what is needed is a localization of the main contribution of an integral to a smaller range ("Neglect the tails") where local approximations can be applied ("Centrally approximate") . The approximate integral is then finally estimated by completing back the tails ("Complete the tails").

The Laplace method is excellently described in books by de Bruijn [**93**] and Henrici [**229**]. A thorough discussion of special cases and multidimensional integrals is found in the book by Bleistein and Handelsman [**55**]. Its principles are fundamental to the development of the saddle point method in Chapter VIII.

▷ **17.** *The classical proof of Stirling's formula.* This proceeds from the integral

$$J_n := \int_0^\infty e^{-x} x^n \, dx \quad (= n!)$$

The maximum of the integrand is at $x_0 = n$ and the central range is now now $n \pm \kappa_n \sqrt{n}$. Reduction to a Gaussian integral follows, though the estimate is no longer an immediate case of application of Theorem B.4. ◁

Laplace's method for sums. The basic principles of the method of Laplace (for integrals) can are often be recycled for the asymptotic evaluation of discrete sums. Take a finite or infinite sum S_n defined by

$$S_n := \sum_k t(n,k).$$

A preliminary task consists in working out the general aspect of the family of numbers $\{t(n,k)\}$ for fixed (but large) n as k varies. In particular, one should locate the value $k_0 \equiv k_0(n)$ of k for which $t(n,k)$ is maximal. In a vast number of cases, tails can be neglected; a central approximation $\widehat{t}(n,k)$ of $t(n,k)$ for k in the "central" region near k_0 can be determined, frequently under the form [remember that we use in this book '$\approx$' in the loose sense of 'approximately equal']

$$\widehat{t}(n,k) \approx s(n)\phi\left(\frac{k-k_0}{\sigma_n}\right).$$

There ϕ is some simple smooth function while $s(n)$ and σ_n are scaling constants. The quantity σ_n indicates the range of the asymptotically significant terms. One may then expect

$$S_n \approx s(n) \sum_k \phi\left(\frac{k-k_0}{\sigma_n}\right).$$

Then provided $\sigma_n \to \infty$, one may further expect to approximate the sum by an integral, which after completing the tails, gives

$$S_n \approx s(n)\sigma_n \int_{-\infty}^{\infty} \phi(t)\, dt.$$

Case study: Sums of powers of binomial coefficients. Here is, in telegraphic style, an application to sums of powers of binomial coefficients:

$$S_n^{(r)} = \sum_{k=-n}^{+n} \binom{2n}{n+k}^r.$$

The largest term arises at $k_0 = 0$. Also, one has elementarily

$$\frac{\binom{2n}{n+k}}{\binom{2n}{n}} = \frac{\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{k-1}{n}\right)}{\left(1+\frac{1}{n}\right)\cdots\left(1+\frac{k}{n}\right)}.$$

Upon taking logarithms, using approximations of $\log(1 \pm x)$, and exponentiating back, one finds

$$\frac{\binom{2n}{n+k}}{\binom{2n}{n}} = \exp\left(-\frac{k^2}{n} + O(k^3 n^{-2})\right). \tag{29}$$

This approximation holds for $k = o(n^{2/3})$, where it provides a gaussian approximation ($\phi(x) = e^{-rx^2}$) with a span of $\sigma_n = \sqrt{n}$. Tails can be neglected to the effect that

$$\frac{1}{\binom{2n}{n}^r} S_n^{(r)} \sim \sum_k \exp\left(-r\frac{k^2}{n}\right),$$

say with $|k| < n^{1/2}\kappa_n$ where $\kappa_n = n^{1/10}$. Then approximating the Riemann sum by an integral and completing the tails, one gets

$$S_n^r \sim \binom{2n}{n}^r \sqrt{n} \int_{-\infty}^{\infty} e^{-rw^2}\, dw, \quad \text{that is,} \quad S_n^r \sim \frac{2^{2rn}}{\sqrt{r}} (\pi n)^{-(r-1)/2},$$

which is our final estimate. l

▷ **18.** *Elementary approximation of Bell numbers.* The Bell numbers counting set partitions are

$$B_n = n![z^n]e^{e^z-1} = e^{-1}\sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

The largest term occurs for k near e^u where u is the positive root of the equation $ue^u = n+1$; the central terms are approximately Gaussian. There results the estimate,

$$B_n = n!e^{-1}(2\pi)^{-1/2}(1+u^{-1})^{-1/2}\exp\left(e^u(1-u\log u) - \frac{1}{2}u\right)\left(1+O(e^{-u})\right). \tag{30}$$

This example is taken from de Bruijn's book [**93**, p. 108]. ◁

6. Mellin transform. The Mellin transform of a function f defined over $\mathbb{R}_{>0}$ is the complex-variable function $f^\star(s)$ defined by the integral

$$f^\star(s) := \int_0^\infty f(x)x^{s-1}\, dx. \tag{31}$$

This transform is also occasionally denoted by $\mathcal{M}[f]$ or $\mathcal{M}[f(x);s]$. Its importance devolves from two properties: (i) it *maps* asymptotic expansions of a function at 0 and $+\infty$ to singularities of the transform; (ii) it *factorizes* harmonic sums (defined below). The conjunction of the mapping property and the harmonic sum property makes it possible to analyse asymptotically rather complicated sums arising from a linear superposition of models taken at different scales. Major properties are summarized in Figure 4. In this brief review, detailed analytic conditions must be omitted: see [**153**] as well as comments and references at the end of this entry.

It is assumed that f is locally integrable. Then, the two conditions,

$$f(x) \underset{x\to 0^+}{=} O(x^u), \qquad f(x) \underset{x\to+\infty}{=} O(x^v),$$

guarantee that f^* exists for s in a *strip*,

$$s \in \langle -u, -v\rangle, \qquad \text{i.e.,} \qquad -u < \Re(s) < -v.$$

Thus existence of the transform is granted provided $v < u$. The prototypical Mellin transform is the Gamma function discussed earlier in this appendix:

$$\Gamma(s) := \int_0^\infty e^{-x}x^{s-1}\, dx = \mathcal{M}[e^{-x};s], \qquad 0 < \Re(s) < \infty.$$

Similarly $f(x) = (1+x)^{-1}$ is $O(x^0)$ at 0 and $O(x^{-1})$ at infinity, and hence its transform exists in the strip $\langle 0, 1\rangle$; it is in fact $\pi/\sin\pi s$, as a consequence of the Eulerian Beta integral. The Heaviside function defined by $H(x) := [\![0 \le x < 1]\!]$ exists in $\langle 0, +\infty\rangle$ and has transform $1/s$.

Harmonic sum propery. The Mellin transform is a linear transform. In addition, it satisfies the simple but important rescaling rule:

$$f(x) \stackrel{\mathcal{M}}{\mapsto} f^\star(s) \qquad \text{implies} \qquad f(\mu x) \stackrel{\mathcal{M}}{\mapsto} \mu^{-s} f^\star(s),$$

for any $\mu > 0$. Linearity then entails the derived rule

$$\sum_k \lambda_k f(\mu_k x) \stackrel{\mathcal{M}}{\mapsto} \left(\lambda_k \mu_k^{-s}\right) \cdot f^\star(s), \tag{32}$$

valid a priori for any finite set of pairs (λ_k, μ_k) and extending to infinite sums whenever the interchange of $\int$ and $\sum$ is permissible. A sum of the form (32) is called a *harmonic sum*, the function f is the "base function", the λ's are the "amplitudes" and the μ's the "frequencies". Equation (32) then yields the "harmonic sum rule": *The Mellin transform of a harmonic sum* factorizes *as the product of the transform of the base function and a generalized Dirichlet series associated to amplitudes and frequencies*. Harmonic sums surface recurrently in the context of analytic combinatorics and Mellin transforms are a method of choice for coping with them.

Here are a few examples of application of the rule (32):

$$\begin{array}{llll} \displaystyle\sum_{k\ge1} e^{-k^2x^2} \underset{\Re(s)>1}{\mapsto} \frac12\Gamma(s/2)\zeta(s) & & \displaystyle\sum_{k\ge0} e^{-x2^k} \underset{\Re(s)>0}{\mapsto} \frac{\Gamma(s)}{1-2^{-s}} \\ \displaystyle\sum_{k\ge0} (\log k) e^{-\sqrt{k}x} \underset{\Re(s)>2}{\mapsto} -\zeta'(s/2)\Gamma(s) & & \displaystyle\sum_{k\ge1} \frac{1}{k(k+x)} \underset{0<\Re(s)<1}{\mapsto} \zeta(2-s)\frac{\pi}{\sin\pi s}. \end{array}$$

$\triangleright$ **19.** *Connection between power series and Dirichlet series.* Let (f_n) be a sequence of numbers with at most polynomial growth, $f_n = O(n^r)$, and with OGF $f(z)$. Then, one has

$$\sum_{n\ge1} \frac{f_n}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty f\left(e^{-x}\right) x^{s-1}\, dx, \qquad \Re(s) > r+1.$$

For instance, one obtains the Mellin pairs

$$\frac{e^{-x}}{1-e^{-x}} \stackrel{\mathcal{M}}{\mapsto} \zeta(s)\Gamma(s) \quad (\Re(s)>1), \qquad \log\frac{1}{1-e^{-x}} \stackrel{\mathcal{M}}{\mapsto} \zeta(s+1)\Gamma(s) \quad (\Re(s)>0).$$

These serve to analyse sums or, conversely, deduce analytic properties of Dirichlet series. $\triangleleft$

Mapping properties. Mellin transfoms map asymptotic terms in the expansions of a function f at 0 and $+\infty$ onto singular terms of the transform $f^\star$. This property stems from the basic identities

$$H(x)x^\alpha \stackrel{\mathcal{M}}{\mapsto} \frac{1}{s+\alpha}\,(s\in\langle-\alpha,+\infty\rangle), \qquad (1-H(x))x^\beta \stackrel{\mathcal{M}}{\mapsto} \frac{1}{s+\beta}\,(s\in\langle-\infty,-\beta\rangle),$$

as well as what one obtains by differentiation with respect to α, β.

Function ($f(x)$)	Mellin transform ($f^\star(s)$)	
$f(x)$	$\int_0^\infty f(x)x^{s-1}\,dx$	**definition**, $s \in \langle -u, -v \rangle$
$\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} f^\star(s)x^{-s}\,ds$	$f^\star(s)$	**inversion th.**, $-u < c < -v$
$\sum_i \lambda_i f_i(x)$	$\sum_i \lambda_i f_i^\star(s)$	linearity
$f(\mu x)$	$\mu^{-s} f^\star(s)$	scaling rule ($\mu > 0$)
$x^\rho f(x^\theta)$	$\frac{1}{\theta} f^\star\left(\frac{s+\rho}{\theta}\right)$	power rule
$\sum_i \lambda_i f(\mu_i x)$	$\left(\sum_i \lambda_i \mu_i^{-s}\right) \cdot f^\star(s)$	**harmonic sum rule** ($\mu_i > 0$)
$\int_0^\infty \lambda(t) f(tx)\,dt$	$\int_0^\infty \lambda(t)t^{-s}\,dt \cdot f^\star(s)$	harmonic integral rule
$f(x)\log^k x$	$\partial_s^k f^\star(s)$	diff. I, $k \in \mathbb{Z}_{\geq 0}$, $\partial_s := \frac{d}{ds}$
$\partial_x^k f(x)$	$\frac{(-1)^k\Gamma(s)}{\Gamma(s-k)} f^\star(s-k)$	diff. II, $k \in \mathbb{Z}_{\geq 0}$, $\partial_x := \frac{d}{dx}$
$\underset{x\to 0}{\sim} x^\alpha(\log x)^k$	$\underset{s\to -\alpha}{\sim} \frac{(-1)^k k!}{(s+\alpha)^{k+1}}$	**mapping**: $x \to 0$, left poles
$\underset{x\to +\infty}{\sim} x^\beta(\log x)^k$	$\underset{s\to -\beta}{\sim} \frac{(-1)^{k-1} k!}{(s+\beta)^{k+1}}$	**mapping**: $x \to \infty$, right poles

FIGURE 4. A summary of major properties of Mellin transforms.

The converse mapping property also holds. Like for other integral transforms, there is an inversion formula: if f is continuous in an interval containing x, then

$$f(x) = \frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} f^\star(s)x^{-s}\,ds, \tag{33}$$

where the abscissa c should be chosen in the "fundamental strip" of f; for instance any c satisfying $-u < c < -v$ with u, v as above is suitable.

In many cases of practical interest, $f^\star$ is continuable as a meromorphic function to the whole of $\mathbb{C}$. If the continuation of $f^\star$ does not grow too fast along vertical lines, then one can estimate the inverse Mellin integral of (33) by residues. This corresponds to shifting the line of integration to some $d \neq c$ and taking poles into account by the residue theorem. Since the residue at a pole s_0 of $f^\star$ involves a factor of x^{-s_0}, the contribution of s_0 will give useful information on $f(x)$ as $x \to \infty$ if s_0 lies to the right of c, and on $f(x)$ as $x \to 0$ if s_0 lies to the left. Higher order poles introduce additional logarithmic factors. The "dictionary" is simply

$$\frac{1}{(s-s_0)^{k+1}} \quad \overset{\mathcal{M}^{-1}}{\longrightarrow} \quad \pm\frac{(-1)^k}{k!}x^{-s_0}(\log x)^k, \tag{34}$$

where the sign is '$+$' for a pole on the left of the fundamental strip and '$-$' for a pole on the right.

Mellin asymptotic summation. The combination of mapping properties and the harmonic sum property constitutes a powerful tool of asymptotic analysis. As an example, let us first investigate the pair

$$F(x) := \sum_{k\geq 1} \frac{1}{1+k^2x^2}, \qquad F^\star(s) = \frac{1}{2}\frac{\pi}{\sin\frac{1}{2}\pi s}\zeta(s),$$

where $F^\star$ results from the harmonic sum rule and is is originally defined in the strip $\langle 1,2\rangle$. The function is meromorphically continuable to the whole of $\mathbb{C}$ with poles at the points $0,1,2$ and $4,6,8,\ldots$. The transform $F^\star$ is small towards infinity, so that application of the dictionary (34) is justified. One then finds mechanically:

$$F(x) \underset{x\to 0^+}{\sim} \frac{\pi}{2x} - \frac{1}{2} + O(x^M), \qquad F(x) \underset{x\to +\infty}{\sim} \frac{\pi^2}{6x^2} - \frac{\pi^4}{90x^4} + \cdots,$$

for any $M>0$.

A particularly important quantity in analytic combinatorics is the harmonic sum

$$\Phi(x) := \sum_{k=0}^{\infty}\left(1-e^{-x/2^k}\right).$$

It occurs for instance in the analysis of longest runs in words (p. 291). By the harmonic sum rule, one finds

$$\Phi^\star(s) = -\frac{\Gamma(s)}{1-2^s}, \qquad s\in\langle -1,0\rangle$$

(The transform of $e^{-x}-1$ is also $\Gamma(s)$, but in the shifted strip $\langle -1,0\rangle$.) The singularities of $\Phi^\star$ are at $s=0$, where there is a double pole, at $s=-1,-2,\ldots$ which are simple poles, but also at the complex points

$$\chi_k = \frac{2ik\pi}{\log 2}.$$

The Mellin dictionary (34) can still be applied provided one integrates along a long rectangular contour that passes in-between poles. The salient feature is here the presence of fluctuations induced by the imaginary poles, since

$$x^{-\chi_k} = \exp\left(-2ik\pi\log_2 x\right),$$

and each pole induces a Fourier element. All in all, one finds (any $M>0$):

$$(35)\qquad \begin{cases} \Phi(x) \underset{x\to+\infty}{\sim} \log_2 x + \dfrac{\gamma}{\log 2} + \dfrac{1}{2} + P(x) + O(x^M) \\ P(x) := \dfrac{1}{\log 2}\displaystyle\sum_{k\in\mathbb{Z}\setminus\{0\}} \Gamma\left(\frac{2ik\pi}{\log 2}\right) e^{-2ik\pi\log_2 x}. \end{cases}$$

The analysis for $x\to 0$ is also possible: in this particular case, it yields

$$\Phi(x) \underset{x\to 0}{\sim} \sum_{n\geq 1} \frac{(-1)^{n-1}}{1-2^{-n}}\frac{x^n}{n!},$$

which is what would result from expanding the exponential in $\Phi(x)$ and reorganizing the terms, and consequently constitutes an exact representation (i.e., '$\sim$' can be replaced by '$=$').

$\triangleright$ **20.** *Mellin-type derivation of Stirling's formula.* One has the Mellin pair

$$L(x) = \sum_{k\geq 1} \log\left(1+\frac{x}{k}\right) - \frac{x}{k}, \qquad L^\star(s) = \frac{\pi}{s\sin \pi s}\zeta(-s), \quad s\in\langle -2,-1\rangle.$$

Note that $L(x) = \log(e^{-\gamma x}/\Gamma(1+x))$. Mellin asymptotics provides

$$L(x) \underset{x\to+\infty}{\sim} -x\log x - (\gamma-1)x - \frac{1}{2}\log x - \log\sqrt{2\pi} - \frac{1}{12x} + \frac{1}{360x^3} - \frac{1}{1260x^5} + \cdots,$$

where one recognizes Stirling's expansion of $x!$,

$$\log x! \underset{x\to+\infty}{\sim} \log\left(x^x e^{-x}\sqrt{2\pi x}\right) + \sum_{n\geq 1}\frac{B_{2n}}{2n(2n-1)}x^{1-2n},$$

with B_n the Bernoulli numbers. $\triangleleft$

$\triangleright$ **21.** *Mellin-type analysis of the harmonic numbers.* For a parameter $\alpha > 0$, one has the Mellin pair:

$$K_\alpha(x) = \sum_{k\geq 1}\left(\frac{1}{k^\alpha} - \frac{1}{(k+x)^\alpha}\right), \qquad K_\alpha^\star(s) = -\zeta(\alpha - s)\frac{\Gamma(s)\Gamma(\alpha-s)}{\Gamma(\alpha)}.$$

This serves to estimate harmonic numbers and their generalisations, for instance

$$H_n \underset{n\to\infty}{\sim} \log n + \gamma - \frac{1}{2n} - \sum_{k\geq 2}\frac{B_k}{k}n^{-k} \sim \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \cdots,$$

since $K_1(n) = H_n$. $\triangleleft$

EXAMPLE 2. *Euler-Maclaurin summation via Mellin analysis.* Let f be continuous on $(0,+\infty)$ and satisfy $f(x) =_{x\to+\infty} O(x^{-1-\delta})$, for some $\delta > 0$, and

$$f(x) \underset{x\to 0^+}{\sim} \sum_{k=0}^{\infty} f_k x^k.$$

The summatory function $F(x)$ satisfies

$$F(x) := \sum_{n\geq 1} f(nx), \qquad F^\star(s) = \zeta(s) f^\star(s),$$

by the harmonic sum rule. The collection of (trimmed) singular expansions of $f^\star$ at $s = 0, -1, -2, \ldots$ is summarized by the formal sum

$$f^\star(s) \asymp \left(\frac{f_0}{s}\right)_{s=0} + \left(\frac{f_1}{s+1}\right)_{s=1} + \left(\frac{f_2}{s+2}\right)_{s=1} + \cdots.$$

Thus, by the mapping properties, provided $F^\star(s)$ is small towards $\pm i\infty$ in finite strips, one has

$$F(x) \underset{x\to 0}{\sim} \frac{1}{x}\int_0^\infty f(t)\,dt + \sum_{j=0}^{\infty} f_j \zeta(-j) x^j,$$

where the main term is associated to the singularity of $F^\star$ at 1 and arises from the pole of $\zeta(s)$, with $f^\star(1)$ giving the integral of f. The interest of this approach is that it is very versatile and allows for various forms of asymptotic expansions of f at 0 as well as multipliers like $(-1)^k$,

$\log k$, and so on; see [**153**] for details and Gonnet's note [**205**] for alternative approaches. END OF EXAMPLE 2. □

General references on Mellin transforms are the books by Doetsch [**107**] and Widder [**434**]. The term "harmonic sum" and some of the corresponding technology originates with the abstract [**171**]. This brief presentation is based on the survey article [**153**] to which we refer for a detailed treatment. Mellin analysis of "harmonic integrals" is a classical topic of applied mathematics for which we refer to the books by Wong [**439**] and Paris–Kaminski [**339**]. Useful treatments of properties of use in discrete mathematics and analysis of algorithms appear in the books by Hofri [**233**], Mahmoud [**307**], and Szpankowski [**401**].

7. Perron-Frobenius theory of nonnegative matrices. Perron-Frobenius theory gives access to growth properties associated to nonnegative matrices and hence to the dominant singularities of generating functions that satisfy linear systems of equations with nonnegative coefficients. Applications to rational asymptotics, paths, graphs, and automata are detailed in Chapter V. The purpose here is only to sketch the main techniques from elementary matrix analysis that intervene in this theory.

THEOREM B.5 (Basic Perron-Frobenius Theorem). *Let A be a matrix whose entries are all positive. Then, A has a unique eigenvalue $\lambda(A)$ which has greatest modulus. This eigenvalue is positive and simple.*

▷ **22.** *Proof of the Basic Perron-Frobenius Theorem B.5.* The main idea consists in investigating the set of possible "expansion factors"

$$S := \left\{ \lambda \;\middle|\; \exists\, \mathrm{v} \geq 0, \quad A\,\mathrm{v} \geq \lambda\,\mathrm{v} \right\}. \tag{36}$$

(There $\mathrm{v} \geq 0$ means that all components of v are nonnegative and $\mathrm{v} \geq \mathrm{w}$ means that $\mathrm{v} - \mathrm{w} \geq 0$.) The largest of the expansion factors,

$$\mu := \sup(S),$$

plays a vital rôle in the argument. The proof relies on establishing that it coincides with the dominant eigenvalue $\lambda(A)$. We set $d = \dim(A)$.

Simple inequalities show that S contains at least the interval $[0, d\min_{i,j} a_{i,j}]$. Inequalities relative to the norm $\|\cdot\|_1$ show that $S \subseteq [0, \sum_{i,j} a_{i,j}]$. Thus, μ is finite and nonzero. That the supremum value μ is actually attained (i.e., $\mu \in S$) results from a simple topological argument detailed in [**25**]: take a bounded family $\mathrm{v}^{(j)}$ corresponding to a sequence $\lambda^{(j)}$ tending to μ; extract a convergent subsequence tending to a vector $\mathrm{v}^{(\infty)}$, which must then satisfy $A\,\mathrm{v}^{(\infty)} \geq \mu\,\mathrm{v}^{(\infty)}$. We let w be such a vector of $\mathbb{R}^d_{\geq 0}$ satisfying $A\,\mathrm{w} \geq \mu\,\mathrm{w}$.

Next, one has $A\,\mathrm{w} = \mu\,\mathrm{w}$. Indeed, suppose a contrario that this is not the case and that (without loss of generality)

$$\sum_j A_{1,j} w_j - \mu\,\mathrm{w}_1 = \eta, \qquad \sum_j A_{i,j} w_j - \mu\,\mathrm{w}_1 \geq 0 \quad (i = 2, \ldots, d), \tag{37}$$

for $\eta > 0$. Then, given the slack afforded by η, one could construct a small perturbation $\mathrm{w}^\star$ of w (by $\mathrm{w}^\star_j = \mathrm{w}_j$ for $j = 2, \ldots, d$ and $\mathrm{w}^\star_1 = w_1 + e_1/(2\mu)$) as well as a value $\mu^\star$ such that $A\,\mathrm{w}^\star \geq \mu^\star\,\mathrm{w}$ with $\mu^\star > \mu$, a contradiction. Thus, μ is an eigenvalue of A and w is an eigenvector corresponding to this eigenvalue.

Furthermore, all eigenvalues are dominated in modulus by μ. Let indeed ν and x be such that $A\,\mathrm{x} = \nu\,\mathrm{x}$. One has $A|\,\mathrm{x}\,| \geq |\nu|\,|\,\mathrm{x}\,|$, where $|\,\mathrm{x}\,|$ designates the vector whose entries are the absolute values of the corresponding entries of x. Thus, by the maximality property defining μ, one must have $|\nu| \leq \mu$. If $|\nu| = \mu$ and x is a corresponding eigenvector, then $A|\,\mathrm{x}\,| \geq \mu|\,\mathrm{x}\,|$, and by the same argument as in (37), one must have $A|\,\mathrm{x}\,| = \mu|x|$. Thus $|x|$ is also an eigenvector corresponding to μ. Then, by the triangle inequality, one has $|A\,\mathrm{x}\,| \geq A|\,\mathrm{x}\,|$, so that in fact $A|\,\mathrm{x}\,| = |A\,\mathrm{x}\,|$, which by the converse triangle inequality implies that $\mathrm{x} = \omega\,\mathrm{y}$, where $\omega \in \mathbb{C}$ and y has nonnegative entries. From this observation and the fact that $A\,\mathrm{y} = \nu\,\mathrm{y}$, it results that ν is positive real, so that $\nu = \mu$. Unicity of the dominant eigenvalue is therefore established.

Finally, simplicity of the eigenvalue μ results from a specific argument based on submatrices. If B_k is obtained from A by deleting the kth row and the kth column, then, on general grounds, one has $\lambda(A) > \lambda(B_k)$. From there, through the equality

$$-\frac{d}{d\lambda}|A - \lambda I| = |B_1 - \lambda I| + \cdots + |B_d - \lambda I|$$

(here $|A| = \det(A)$), it can be verified that the derivative of the characteristic polynomial of A at μ is strictly negative, and in particular nonzero; hence simplicity of the eigenvalue μ. (See [**25**] for details of this argument.) $\lhd$

Extensions of this basic theorem are discussed in the text (Chapter V). Excellent treatments of Perron-Frobenius theory are to be found in the books of Bellman [**25**, Ch. 16], Gantmacher [**191**, Ch. 13], as well as Karlin and Taylor [**252**, p. 536–551].

8. Several complex variables.. The theory of analytic (or holomorphic) functions of one complex variables extends nontrivially to several complex variables. This profound theory has been largely developed in the course of the twentieth century. Here we shall only need the most basic *concepts*, not the deeper results, of the theory.

Consider the space $\mathbb{C}^m$ endowed with the metric

$$|z| = |(z_1, \ldots, z_m)| = \sum_{j=1}^{m} |z_j|^2,$$

under which it is isomorphic to the Euclidean space $\mathbb{R}^{2m}$. A function f from $\mathbb{C}^m$ to $\mathbb{C}$ is said to be analytic at some point a if in a neighbourhood of a it can be represented by a convergent power series,

$$f(z) \equiv f(z_1, \cdots, z_m) = \sum_n f_n (z-a)^n \equiv \sum_{n_1,\ldots,n_m} f_{n_1,\ldots,n_m} (z_1-a_1)^{n_1} \cdots (z_m-a_m)^{n_m}. \tag{38}$$

There and throughout the theory extensive use is made of multi-index conventions, as encountered in Chapter III.

An expansion (38) converges in a polydisc $\prod_j \{|z_j - a_j| < r_j\}$, for some $r_j > 0$. A convergent expansion at $(0, \ldots, 0)$ has its coefficients majorized in absolute value by those of a series of the form

$$\prod_{j=1}^{m} \frac{1}{1 - z_j/R_j} = \sum_n R^{-n} z^n \equiv \sum_{n_1,\ldots,n_m} R_1^{-n_1} \cdots R_m^{-n_m}\, z_1^{n_1} \cdots z_m^{n_m}.$$

From there, closure of analytic functions under sums, products, and compositions result from standard manipulations of majorant series (see Chapter IV for the univariate case). Finally, a function is analytic in an open set $\Omega \subseteq \mathbb{C}^m$ iff it is analytic at each $a \in \Omega$.

A remarkable theorem of Hartogs asserts that $f(z)$ with $z \in \mathbb{C}^m$ is analytic *jointly* in all the z_j (in the sense of (38)) if it is analytic *separately* in each variable z_j. (The version of the theorem that postulates *a priori* continuity is elementary.)

Like in the one-dimensional case, analytic functions can be equivalently defined by means of differentiability conditions. A function is $\mathbb{C}$-differentiable or holomorphic at a if as $\Delta z \to 0$ in $\mathbb{C}^m$, one has

$$f(a + \Delta z) - f(a) = \sum_{j=1}^{m} c_j \Delta z_j + o\left(|\Delta z|\right).$$

The coefficients c_j are the partial derivatives, $c_j = \partial_{z_j} f(a)$. The fact that this relation does not depend on the way Δz tends to 0 implies the Cauchy-Riemann equations. In a way that parallels the single variable case, it is proved that two conditions are equivalent: f is analytic; f is complex differentiable.

Iterated integrals are defined in the natural way and one finds, by a repeated use of calculus in a single variable,

$$f(z) = \frac{1}{(2i\pi)^n} \int_{C_1} \cdots \int_{C_m} \frac{f(\zeta)}{(\zeta_1 - z_1)\cdots(\zeta_m - z_m)}\, d\zeta_1 \cdots d\zeta_m, \tag{39}$$

where C_j is a small circle surrounding z_j in the z_j–plane. By differentiation under the integral sign, Equation (39) also provides an integral formula for the partial derivatives of f, which is the analogue of Cauchy's coefficient formula. Iterated integrals are independent of details of the "polypath" on which they are taken, and uniqueness of analytic continuation holds.

The theory of functions of several complex variables develops in the direction of an integral calculus that is much more powerful than the iterated integrals mentioned above; see for instance the book by Aĭzenberg and Yuzhakov [**5**] for a multidimensional residue approach. Egorychev's monograph [**121**] develops systematic applications of the theory of functions of one or several complex variables to the evaluation of combinatorial sums. Pemantle [**341, 342, 343**] has launched an ambitious research programme meant to extract the coefficients of meromorphic multivariate generating functions by means of this theory, with the ultimate goal of obtaining systematically asymptotics from multivariate generating functions. In contrast, see especially Chapter IX, we can limit ourselves to developing a perturbative theory of one-variable complex function theory.

In the context of this book, the basic notion of analyticity in several complex variables serves to confer a *bona fide* analytic meaning to multivariate generating functions. Basic definitions are also needed in the context of functions f defined implicitly by functional relations of the form $H(z, f) = 0$ or $H(z, u, f) = 0$, where analytic functions of two or three complex variables (like H) make an appearance. (See in particular the discussion of the analytic Implicit Function Theorem in this Appendix.)

APPENDIX C

Complements of Probability Theory

This appendix contains entries arranged in logical order regarding the following topics:

Probability spaces and measure; Random variables; Transforms of distributions; Special distributions; Convergence in law.

In this book we start from probability spaces that are finite, since they arise from objects of a fixed size in some combinatorial class (see Chapter III of Part A and APPENDIX A: *Combinatorial probability*, p. 644 for elementary aspects), then need basic properties of continuous distributions in order to characterize asymptotic limit laws. The entries in this appendix are used principally in Chapter IX of Part C relative to *Random Structures*. They present a unified framework that encompasses discrete and continuous probability distributions alike.

1. Probability spaces and measure. An axiomatization of probability theory[1] was discovered in the 1930s by Kolmogorov. A *measurable space* consists of a set Ω, called the set of elementary events or the sample set and a σ-algebra $\mathcal{A}$ of subsets of Ω called events (that is, a collection of sets containing $\emptyset$ and closed under complement and denumerable unions). A *measure space* is a measurable space endowed with a measure $\mu : \mathcal{A} \mapsto \mathbb{R}_{\geq 0}$ that is additive over finite or denumerable unions of disjoint sets; in that case, elements of $\mathcal{A}$ are called measurable sets. A *probability space* is a measure space for which the measure satisfies the further normalization $\mu(\Omega) = 1$; in that case, we also write $\mathbb{P}$ for μ. Any set $S \subseteq \Omega$ such that $\mu(S) = 1$ is called a *support* of the probability measure.

The definitions given above cover several important cases.

(i) *Finite sets with the uniform measure* also known as "counting" measure. In this case, Ω is finite, all sets are in $\mathcal{A}$ (i.e., are measurable), and ($\|\cdot\|$ denotes cardinality)

$$\mu(E) := \frac{\|E\|}{\|S\|}.$$

Nonuniform measures over a finite set Ω are determined by assigning a nonnegative weight $p(\omega)$ to each element of Ω (with $\sum_{\omega\in\Omega} p(\omega) = 1$) and setting

$$\mu(E) := \sum_{e\in E} p(e).$$

(We also write $\mathbb{P}(e)$ for $\mathbb{P}(\{e\}) \equiv \mu(\{e\}) = p(e)$.) In this book, Ω is usually the subclass $\mathcal{C}_n$ formed by the objects of size n in some combinatorial class $\mathcal{C}$. The uniform probability is normally assumed, although sometimes weighted models are considered: see for instance in Chapter III the discussion of weighted word models and Bernoulli trials as well as the case of weighted tree models and branching processes.

[1]For this entry we refer to the vivid and well motivated presentation in Williams' book [**438**] or to many classical treatises like the ones by Billingley [**53**] and Feller [**133**].

(ii) *Discrete probability measures over the integers* (supported by $\mathbb{Z}$ or $\mathbb{Z}_{\geq 0}$). In this case the measure is determined by a function $p : \mathbb{Z} \mapsto \mathbb{R}_{\geq 0}$ and

$$\mu(E) := \sum_{e \in E} p(e),$$

with $\mu(\mathbb{Z}) = 1$. (All sets are measurable.) More general discrete measures supported by denumerable sets of $\mathbb{R}$ can be similarly defined.

(iii) *The real line* $\mathbb{R}$ equipped with the σ-algebra generated by the open intervals constitutes a standard example of a measurable space; in that case, any member of the σ-algebra is known as a Borel set. The measure, denoted by λ, that assigns to an interval (a, b) the value $\lambda(a, b) = b - a$ (and is extended nontrivially to all Borel sets by additivity) is known as the Lebesgue measure. The interval $[0, 1]$ endowed with λ is a probability space. The line $\mathbb{R}$ itself is not a probability space since $\lambda(\mathbb{R}) = +\infty$.

In the measure-theoretic framework, a *random variable* is a mapping X from a probability space Ω (equipped with its σ-algebra $\mathcal{A}$ and its measure $\mathbb{P}_\Omega$) to $\mathbb{R}$ (equipped with its Borel sets $\mathcal{B}$) such that the preimage $X^{-1}(B)$ of any $B \in \mathcal{B}$ lies in $\mathcal{A}$. For $B \in \mathcal{B}$, the probability that X lies in B is then defined as

$$\mathbb{P}(X \in B) := \mathbb{P}_\Omega(X^{-1}(B)).$$

Since the Borel sets can be generated by the semi-infinite intervals $(-\infty, x]$, this probability is equivalently determined by the function

$$F(x) := \mathbb{P}(X \leq x),$$

which is called the *distribution function* or *cumulative distribution function* of X. It is then possible to introduce random variables directly by means of distribution functions, see the next entry below, *Random variables*.

The next step is to go from measures of sets to integrals of (real valued) functions. Lebesgue integrals are constructed, first for indicator functions of intervals, then for simple (staircase) functions, then for nonnegative functions, finally for integrable functions. One defines in this way, for an arbitrary measure μ, the Lebesgue integral

$$\int f d\mu, \qquad \text{also written} \quad \int f(x) d\mu(x) \quad \text{or} \quad \int f(x) \mu(dx), \tag{1}$$

where the last notation is often preferred by probabilists. The basic idea is to decompose the domain of *values* of f into finitely many measurable sets (A_i) and, for a positive function f, consider the supremum over all finite decompositions (A_i)

$$\int f \, d\mu := \sup_{(A_i)} \sum_i \left[\inf_{\omega \in A_i} f(\omega) \right] \mu(A_i). \tag{2}$$

(Thus Riemman integration proceeds by decomposing the domain of the function's *arguments* while Lebesgue integrals decomposes the domain of *values* and appeals to a richer notion of measure.)

In (1) and (2), the possibility exists that μ assigns a nonzero measure to certain individual points. In such a context, the integral is sometimes referred to as the *Lebesgue-Stieltjes* integral. It suitably generalizes the *Riemann-Stieltjes* integral which, given a real valued function M, defines the following extension of the standard Riemann integral:

$$\int f(x)\, dM(x) = \lim_{(B_k)} \sum_k f(x_k)\Delta_{B_k}(M). \tag{3}$$

There the B_k form a finite partition of the domain in which the argument of f ranges, the limit is taken as the largest B_k tends to 0, each x_k lies in B_k, and $\Delta_{B_k}(M)$ is the variation of M on B_k. The great advantage of Stieltjes (hence automatically of Lebesgue) integrals is to unify many of the formulæ relative to discrete and continuous probability distributions while providing a simple framework adapted to mixed cases.

2. Random variables. A real random variable X is fully characterized by its (cumulative) distribution function

$$F_X(x) := \mathbb{P}(X \le x),$$

which is a nondecreasing right-continuous function satisfying $F(-\infty) = 0$, $F(+\infty) = 1$.

A variable is *discrete* if it is supported by a finite or denumerable set. Almost all discrete distributions in this book are supported by $\mathbb{Z}$ or $\mathbb{Z}_{\ge 0}$. (An interesting exception is the collection of limit distributions occurring in longest runs of words; see Chapter IV.)

A variable X is *continuous* if it assigns zero probability mass to any finite or denumerable set. In particular, it has no jump. An easy theorem states that any distribution function can be decomposed into a discrete and a continuous part,

$$F(x) = c_1 F^{\mathrm{d}}(x) + c_2 F^{\mathrm{c}}(x), \qquad c_1 + c_2 = 1.$$

(The jumps must sum to at most 1, hence their set is at most denumerable.) A variable is *absolutely continuous* if it assigns zero probability mass to any Borel set of measure 0. In that case, the Radon Nikodym Theorem asserts that there exists a function w such that

$$F_X(x) = \int_{-\infty}^{x} w(y)\, dy.$$

(There, in all generality, the Lebesgue integral is required but the Riemann integral is sufficient for all practical purposes in this book.) The function $w(x)$ is called a *density* of the random variable X (or of its distribution function). When F_X is differentiable everywhere it admits the density

$$w(x) = \frac{d}{dx} F_X(x),$$

by the Fundamental Theorem of Calculus.

$\triangleright$ **1.** *The Lebesgue decomposition theorem.* It states that any distribution function $F(x)$ decomposes as

$$F(x) = c_1 F^{\mathrm{d}}(x) + c_2 F^{\mathrm{ac}} + c_3 F^{\mathrm{s}}(x), \qquad c_1 + c_2 + c_3 = 1,$$

where F^{d} is discrete, F^{ac} is absolutely continuous, and F^{s} is continuous but *singular*, i.e., it is supported by a Borel set of Lebesgue measure 0. Singular random variables are constructed, e.g., from the Cantor set. $\lhd$

In this book, all combinatorial distributions are discrete (and then usually supported by $\mathbb{Z}_{\geq 0}$). All continuous distributions obtained as limits of discrete ones are, in our context, absolutely continuous and the qualifier "absolutely" is globally understood when discussing continuous distributions.

If X is a random variable, the *expectation* of a function $g(X)$ is defined

$$\mathbb{E}\left(g(X)\right) = \int_\Omega g(X)dP = \int_{\mathbb{R}} g(x)dF(x),$$

where the latter form involves the distribution function F of X. In particular the *expectation* or *mean* of X is $\mathbb{E}(X)$, and generally its *moment* of order r is

$$\mu^{(r)} = \mathbb{E}(X^r).$$

(These quantities may not exist for $r \neq 0$.)

$\rhd$ **2.** *Alternative formulæ for expectations.* If X is supported by $\mathbb{R}_{\geq 0}$ and has a density:

$$\mathbb{E}(X) = \int_0^\infty (1 - F(x))\ dx.$$

If X is supported by $\mathbb{Z}_{\geq 0}$:

$$\mathbb{E}(X) = \sum_{k\geq 0} \mathbb{P}(X > k).$$

Prrofs are by partial integration and summation: for instance with $p_k = \mathbb{P}(X = k)$,

$$\mathbb{E}(X) = \sum_{k\geq 1} kp_k = (p_1 + p_2 + p_3 + \cdots) + (p_2 + p_3 + \cdots) + (p_3 + \cdots) + \cdots .$$

Similar formulæ hold for higher moments. $\lhd$

3. Transforms of distributions. The Laplace transform of X (or of its distribution function F) is defined[2] by

$$\lambda_X(s) := \mathbb{E}\left(e^{sX}\right) = \int_{-\infty}^{+\infty} e^{sx}\, dF(x),$$

and is also known as the moment generating function (see below for an existential discussion). The characteristic function is defined by

$$\phi_X(t) = \mathbb{E}\left(e^{itX}\right) = \int_{-\infty}^{+\infty} e^{itx}\, dF(x),$$

and it is a Fourier transform Both transforms are formal variants of one another and $\phi_X(t) = \lambda_X(it)$.

If X is discrete and supported by $\mathbb{Z}$, then its *probability generating function* (PGF) is defined as

$$P_X(u) := \mathbb{E}(u^X) = \sum_{k\in\mathbb{Z}} \mathbb{P}(X = k)u^k.$$

[2]If F has a discrete component, then integration is to be taken in the sense of Lebesgue-Stieltjes or Riemann-Stieltjes.

As an analytic object this always exists when X is nonnegative (supported by $\mathbb{Z}_{\geq 0}$), in which case the PGF is analytic at least in the open disc $|u| < 1$. If X assumes arbitrarily large negative values, then the PGF certainly exists on the unit circle, but sometimes not on a larger domain. The precise domain of existence of the PGF as an analytic function depends on the geometric rate of decay of the left and right tails of the distribution, that is, of $\mathbb{P}(X = k)$ as $k \to \pm\infty$. The characteristic function of the variable X (and of its distribution function F_X) is

$$\phi_X(t) := \mathbb{E}(e^{itX}) = P_X(e^{it}) = \sum_{k\in\mathbb{Z}} \mathbb{P}(X = k)e^{ikt}.$$

It always exists for *real* values of t. The Laplace transform of a discrete distribution is

$$\lambda_X(s) := \mathbb{E}(e^{sX}) = P_X(e^s) = \sum_{k\in\mathbb{Z}} \mathbb{P}(X = k)e^{ks}.$$

If X is a continuous random variable with distribution function $F(x)$ and density $w(x)$, then the characteristic function is expressed as

$$\phi_X(t) := \mathbb{E}(e^{itX}) = \int_{\mathbb{R}} e^{itx} w(x)\, dx.$$

and the Laplace transform is

$$\lambda_X(s) := \mathbb{E}(e^{sX}) = \int_{\mathbb{R}} e^{sx} w(x)\, dx.$$

The Fourier transform always exists for real arguments (by integrability of the Fourier kernel e^{it} whose modulus is 1). The Laplace transform, when it exists in a strip, extends analytically the characteristic function via the equality $\phi_X(t) = \lambda_X(it)$. The Laplace transform is also called the *moment generating function* since an alternative formulation of its definition, valid for discrete and continuous cases alike, is

$$\lambda_X(s) := \sum_{k\geq 0} \mathbb{E}(X^k)\frac{s^k}{k!},$$

which indeed represents the exponential generating function of moments. (We prefer not to use this terminology so as to avoid a possible confusion with the many other types generating functions employed in this book.)

▷ **3.** *Centring, scaling, and standardization.* Let X be a random variable. Define $Y = \frac{X-\mu}{\sigma}$. The representations as expectations of the Laplace transform of the characteristic function make it obvious that

$$\phi_Y(t) = e^{-\mu it}\phi_X\left(\frac{t}{\sigma}\right), \qquad \lambda_Y(s) = e^{-\mu s}\lambda_X\left(\frac{s}{\sigma}\right).$$

One says that Y is obtained from X by centring (by a shift of μ) and scaling (by a factor of σ). If μ and σ are the mean and standard deviation of X, then one says that Y is a standardized version of X. ◁

▷ **4.** *Moments and transforms.* The moments are accessible from either transform,

$$\mu^{(r)} := E\{Y^r\} = \left.\frac{d^r}{ds^r}\lambda(s)\right|_{s=0} = (-i)^r \left.\frac{d^r}{dt^r}\phi(t)\right|_{t=0}.$$

In particular, we have

$$(4) \qquad \begin{aligned} \mu &= \frac{d}{ds}\lambda(s)\Big|_{s=0} = -i\,\frac{d}{dt}\phi(t)\Big|_{t=0} \\ \mu^{(2)} &= \frac{d^2}{ds^2}\lambda(s)\Big|_{s=0} = -\,\frac{d}{dt}\phi(t)\Big|_{t=0} \\ \sigma^2 &= \frac{d^2}{ds^2}\log\lambda(s)\Big|_{s=0} = -\,\frac{d^2}{dt^2}\log\phi(t)\Big|_{t=0}. \end{aligned}$$

The direct expression of the standard deviation in terms of $\log\lambda(s)$, called the *cumulant generating function*, often proves computationally handy. ◁

▷ **5.** *Mellin transforms of distributions.* The quantity $M(s) := \mathbb{E}(X^{s-1})$ is called the Mellin transform of X (or of its distribution function F), when X is supported by $\mathbb{R}_{\geq 0}$. In particular, if X admits a density, then this notion coincides with the usual definition of a Mellin transform. When it exists, the value of the Mellin transform at an integer $s = k$ provides the moment of order $k-1$. At other points, the Mellin transform provides moments of fractional order. ◁

▷ **6.** *A "symbolic" fragment of probability theory.* Consider discrete random variables supported by $Z_{\geq 0}$. Let $X, X_1, \ldots$ be random variables with PGF $p(u)$ and let Y have PGF $q(u)$. Then, certain natural operations admit a translation into PGFs:

Operation		PGF
Switch	$(\mathrm{Bern}(\lambda) \Rightarrow X \mid Y)$	$\lambda p(u) + (1-\lambda)q(u)$
Sum	$X + Y$	$p(u)\cdot q(u)$
	$X_1 + \cdots + X_n$	$p(u)^n$
Random sum	$X + 1 + \cdots + X_Y$	$q(p(u))$
Size bias	∂_X	$\dfrac{up'(u)}{p'(1)}$

("Bern" means a Bernoulli $\{0,1\}$ variable B and the switch is interpreted as $BX + (1-B)Y$. Size-biased distributions occur in Chapter VII.) ◁

The importance of these transforms derives from the existence of *continuity theorem* by which convergence of distributions can be established via convergence of transforms.

4. Special distributions. A compendium of special distribution is provided by Figure 1.

A *Bernoulli trial* of parameter q is an event that has probability q of having value 0 (interpreted as "failure") and probability p of having value 1 (interpreted as "success"), with $p+q=1$. Formally, this is the set $\Omega = \{0,1\}$ endowed with the probability measure $\mathbb{P}(0) = q$, $\mathbb{P}(1) = p$. The binomial distribution (also called Bernoulli distribution) of parameters n, q is the random variable that represents the number of successes in n independent Bernoulli trials. This is the probability distribution associated with the game of heads-and-tails. The geometric distribution is the distribution of a random variable X that records the number of failures till the first success is encountered in a potentially arbitrarily long sequence of Bernoulli trials. By extension, one also refers to independent experiments with finitely many possible outcomes as Bernoulli trials. In that sense, the model of words of some fixed length over a finite alphabet and nonuniform letter weights (or probabilities) belongs to the category of Bernoulli models; see Chapter III. The negative binomial distribution of index m (written $NB[m]$)

	Distrib.	Prob. (D), density (C)	PGF(D), Char. function (C)
D	Binomial (n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	$(q+pz)^n$
D	Geometric (q)	$(1-q)q^k$	$\frac{1-q}{1-qz}$
D	Neg. binomial$[m]$ (q)	$\binom{m+k-1}{k}q^k(1-q)^m$	$\left(\frac{1-q}{1-qz}\right)^m$
D	Log. series (λ)	$\frac{1}{-\log(1-\lambda)}\frac{\lambda^k}{k!}$	$\frac{\log(1-\lambda z)}{\log(1-\lambda)}$
D	Poisson (λ)	$e^{-\lambda}\frac{\lambda^k}{k!}$	$e^{\lambda(1-z)}$
C	Gaussian or Normal, $\mathcal{N}(0,1)$	$\frac{e^{-x^2/2}}{\sqrt{2\pi}}$	$e^{-t^2/2}$
C	Exponential	e^{-x}	$\frac{1}{1-it}$
C	Uniform $[-\frac{1}{2},+\frac{1}{2}]$	$[\![-\frac{1}{2}\le x\le+\frac{1}{2}]\!]$	$\frac{\sin(t/2)}{(t/2)}$

FIGURE 1. A list of commonly encountered discrete (D) and continuous (C) probability distributions: type, name, probabilities or density, probability generating function or characteristic function.

and parameter q corresponds to the number of failures before m successes are encountered. We have found in Chapter VII that it is systematically associated with the number of r–components in an unlabelled multiset schema $\mathcal{F}=\mathfrak{M}(\mathcal{G})$ whose composition of singularities is of the exp-log type. The geometric distribution appears in several schemas related to sequences while the logarithmic series distribution is closely tied to cycles (Chapter V).

The Poisson distribution counts amongst the most important distributions of probability theory. Its essential properties are recalled in Figure 1. It occurs for instance in the distribution of singleton cycles and of r-cycles in a random permutation and more generally in labelled composition schemes (Chapter IX).

In this book all probability distributions arising directly from combinatorics are *a priori* discrete as they are defined on finite sets—typically a certain subclass $\mathcal{C}_n$ of a combinatorial class $\mathcal{C}$. However, as the size n of the objects considered grows, these finite distributions may approach a continuous limit. In this context, by far the most important law is the *Gaussian law* also known as *normal law*, which is defined by its density and its distribution function:

$$g(x)=\frac{e^{-x^2/2}}{\sqrt{2\pi}},\qquad \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-y^2/2}\,dy. \tag{5}$$

The corresponding Laplace transform is then evaluated by completing the square:

$$\lambda(s)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty}e^{-y^2/2+sy}\,dy.=e^{s^2/2},$$

Characteristic function ($\phi(t)$)	*Distribution function* ($F(x)$)
$\phi(0) = 1$	$F(-\infty) = 0$, $F(+\infty) = 1$
$\lvert\phi(t_0)\rvert = 1$ for some $t_0 \neq 0$	Lattice distribution, span $\frac{2\pi}{t_0}$
$\phi(t) \underset{t\to 0}{=} 1 + i\mu t + o(t)$	$\mathbb{E}(X) = \mu < \infty$
$\phi(t) \underset{t\to 0}{=} 1 + i\mu t - \nu\dfrac{t^2}{2} + o(t^2)$	$\mathbb{E}(X^2) = \nu < \infty$
$\log\phi(t) = -\frac{t^2}{2}$	$X \overset{d}{=} \mathcal{N}(0,1)$
$\phi(t) \to 0$ as $t \to \infty$	X is continuous
$\phi(t)$ integrable (is in $\mathcal{L}_1$)	X is absolutely continuous density is $w(x) = \dfrac{1}{2\pi}\displaystyle\int_{-\infty}^{+\infty} e^{-itx}\phi(t)\,dt$.
$\lambda(s) := \phi(-is)$ exists in $\alpha < \Re(s) < \beta$	Exponential tails
$\lim_{T\to\infty} \frac{1}{2T}\int_{-T}^{+T} \lvert\phi(t)\rvert^2 \, dt$	equals $\sum_i (p_i)^2$; the p_i are the jumps
$\phi_n(t) \to \phi(t)$ (point conv.)	$F_n \overset{\mathcal{D}}{\Longrightarrow} F$ (weak conv.) $X_n \overset{\mathcal{D}}{\Longrightarrow} X$ (conv. in distribution)
ϕ_n "close" to ϕ	F_n "close" to F (Berry-Esseen)

FIGURE 2. The correspondence between properties of the distribution function (F) of a random variable (X) and properties of the corresponding characteristic functions (ϕ).

and, similarly, the characteristic function is $\phi(t) = e^{-t^2/2}$. The distribution of (5) is referred to as the *standard* normal distribution, $\mathcal{N}(0,1)$; if X is $\mathcal{N}(0,1)$, the variable $Y = \mu + \sigma X$ defines the normal distribution with mean μ and standard deviation σ, denoted $\mathcal{N}(\mu,\sigma)$.

Amongst other continuous distributions appearing in this book, we mention the theta distributions associated to the height of trees and Dyck paths (Chapter V) and the stable laws alluded to in Chapter VI.

5. Convergence in law. Let F_n be a family of distribution functions F_n. We say generally that the F_n *converge weakly* to a distribution function F if pointwise

$$\lim_n F_n(x) = F(x), \tag{6}$$

for every continuity point x of F. This is expressed by writing $F_n \Rightarrow F$ as well as $X_n \overset{\mathcal{D}}{\Longrightarrow} X$, if X_n, X are random variables corresponding to F_n, F. We say that X_n *converges in distribution* or *converges in law* to X. For discrete distributions supported by $\mathbb{Z}$, and equivalent form of (6) is $\lim_n F_n(k) = F(k)$ for each $k \in \mathbb{Z}$; for continuous distributions, Equation (6) just means that $\lim_n F_n(x) = F(x)$ for all $x \in \mathbb{R}$. Although in all generality anything can tend to anything else, due to the finite nature of combinatorics, we shall only need in this book the convergences

Discrete $\Rightarrow$ Discrete, Discrete $\Rightarrow$ Continuous (after standardization).

Properties of random variables are reflected by probabilities of characteristic functions and Figure 2 offers an aperçu. Most important for us is the *Continuity Theorem* of characteristic functions due to Lévy and stated in Chapter IX. The Berry–Esseen inequalities also stated in Chapter IX lie at the origin of precise speed of convergence estimates to asymptotic limits.

Bibliography

1. S.-S. Abhyankar, *Algebraic geometry for scientists and engineers*, American Mathematical Society, 1990.
2. Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions*, Dover, 1973, A reprint of the tenth National Bureau of Standards edition, 1964.
3. A. V. Aho and J. D. Ullman, *Principles of compiler design*, Addison-Wesley, 1977.
4. Alfred V. Aho and Margaret J. Corasick, *Efficient string matching: an aid to bibliographic search*, Communications of the ACM **18** (1975), 333–340.
5. I. A. Aĭzenberg and A. P. Yuzhakov, *Integral representations and residues in multidimensional complex analysis*, Translations of Mathematical Monographs, vol. 58, American Mathematical Society, Providence, RI, 1983.
6. David Aldous and Persi Diaconis, *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem*, American Mathematical Society. Bulletin. New Series **36** (1999), no. 4, 413–432.
7. David Aldous and James A. Fill, *Reversible Markov chains and random walks on graphs*, 2003, Book in preparation; mansucript available electronically.
8. David J. Aldous, *Deterministic and stochastic models for coalescence (aggregation and coagulation): a review of the mean-field theory for probabilists*, Bernoulli **5** (1999), no. 1, 3–48.
9. Noga Alon and Joel H. Spencer, *The probabilistic method*, John Wiley & Sons Inc., New York, 1992.
10. George E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and its Applications, vol. 2, Addison–Wesley, 1976.
11. George E. Andrews, Richard Askey, and Ranjan Roy, *Special functions*, Encyclopedia of Mathematics and its Applications, vol. 71, Cambridge University Press, 1999.
12. Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, 1976.
13. ———, *Modular functions and Dirichlet series in number theory*, Springer-Verlag, New York, 1976, Graduate Texts in Mathematics, No. 41.
14. J. Arney and E. D. Bender, *Random mappings with constraints on coalescence and number of origins*, Pacific Journal of Mathematics **103** (1982), 269–294.
15. Richard Arratia, A. D. Barbour, and Simon Tavaré, *Random combinatorial structures and prime factorizations*, Notices of the American Mathematical Society **44** (1997), no. 8, 903–910.
16. ———, *Logarithmic combinatorial structures: a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003.
17. Krishna B. Athreya and Peter E. Ney, *Branching processes*, Springer-Verlag, New York, 1972, Die Grundlehren der mathematischen Wissenschaften, Band 196.
18. Raymond Ayoub, *An introduction to the analytic theory of numbers*, Mathematical Surveys, No. 10, American Mathematical Society, Providence, R.I., 1963.
19. Jinho Baik, Percy Deift, and Kurt Johansson, *On the distribution of the length of the longest increasing subsequence of random permutations*, Journal of the American Mathematical Society **12** (1999), no. 4, 1119–1178.
20. Cyril Banderier, Mireille Bousquet-Mélou, Alain Denise, Philippe Flajolet, Danièle Gardy, and Dominique Gouyou-Beauchamps, *Generating functions of generating trees*, Discrete Mathematics **246** (2002), no. 1-3, 29–55.
21. Cyril Banderier and Philippe Flajolet, *Basic analytic combinatorics of directed lattice paths*, Theoretical Computer Science **281** (2002), no. 1-2, 37–80.
22. Cyril Banderier, Philippe Flajolet, Gilles Schaeffer, and Michèle Soria, *Random maps, coalescing saddles, singularity analysis, and Airy phenomena*, Random Structures & Algorithms **19** (2001), no. 3/4, 194–246.

23. A. D. Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, The Clarendon Press Oxford University Press, New York, 1992, Oxford Science Publications.
24. Alan F. Beardon, *Iteration of rational functions*, Graduate Texts in Mathematics, Springer Verlag, 1991.
25. Richard Bellman, *Matrix analysis*, S.I.A.M. Press, Philadelphia, Pa, 1997, A reprint of the second edition, first published by McGraw-Hill, New York, 1970.
26. E. A. Bender and L. B. Richmond, *Multivariate asymptotics for products of large powers with application to Lagrange inversion*, Electronic Journal of Combinatorics **6** (1999), R8, 21pp.
27. Edward A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, Journal of Combinatorial Theory **15** (1973), 91–111.
28. ______, *Asymptotic methods in enumeration*, SIAM Review **16** (1974), no. 4, 485–515.
29. ______, *Convex n–ominoes*, Discrete Mathematics **8** (1974), 219–226.
30. Edward A. Bender and E. Rodney Canfield, *The asymptotic number of labeled graphs with given degree sequences*, Journal of Combinatorial Theory, Series A **24** (1978), 296–307.
31. Edward A. Bender, E. Rodney Canfield, and Brendan D. McKay, *Asymptotic properties of labeled connected graphs*, Random Structures & Algorithms **3** (1992), no. 2, 183–202.
32. Edward A. Bender and Jay R. Goldman, *Enumerative uses of generating functions*, Indiana University Mathematical Journal (1971), 753–765.
33. Edward A. Bender and Fred Kochman, *The distribution of subword counts is usually normal*, European Journal of Combinatorics **14** (1993), 265–275.
34. Edward A. Bender and L. Bruce Richmond, *Central and local limit theorems applied to asymptotic enumeration II: Multivariate generating functions*, Journal of Combinatorial Theory, Series A **34** (1983), 255–265.
35. Edward A. Bender, L. Bruce Richmond, and S. G. Williamson, *Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions*, Journal of Combinatorial Theory, Series A **35** (1983), no. 3, 264–278.
36. Jon Bentley and Robert Sedgewick, *Fast algorithms for sorting and searching strings*, Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM Press, 1997.
37. F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, Cambridge University Press, Cambridge, 1998.
38. François Bergeron, Philippe Flajolet, and Bruno Salvy, *Varieties of increasing trees*, CAAP'92 (J.-C. Raoult, ed.), Lecture Notes in Computer Science, vol. 581, 1992, Proceedings of the 17th Colloquium on Trees in Algebra and Programming, Rennes, France, February 1992., pp. 24–48.
39. Elwyn R. Berlekamp, *Algebraic coding theory*, Mc Graw-Hill, 1968, Revised edition, 1984.
40. Bruce C. Berndt, *Ramanujan's notebooks, part I*, Springer Verlag, 1985.
41. J. Berstel, *Sur les pôles et le quotient de Hadamard de séries n-rationnelles*, Comptes–Rendus de l'Académie des Sciences **272** (1971), no. Série A, 1079–1081.
42. J. Berstel and C. Reutenauer, *Recognizable formal power series on trees*, Theoretical Computer Science **18** (1982), 115–148.
43. Jean Berstel (ed.), *Séries formelles*, LITP, University of Paris, 1978, (Proceedings of a School, Vieux–Boucau, France, 1977).
44. Jean Berstel and Dominique Perrin, *Theory of codes*, Academic Press Inc., Orlando, Fla., 1985.
45. Jean Berstel and Christophe Reutenauer, *Les séries rationnelles et leurs langages*, Masson, Paris, 1984.
46. Jean Bertoin, Philippe Biane, and Marc Yor, *Poissonian exponential functionals, q-series, q-integrals, and the moment problem for log-normal distributions*, Tech. Report PMA-705, Laboratoire de Probabilitś et Modèles Aléatoires, Université Paris VI, 2002.
47. Alberto Bertoni, Christian Choffrut, Massimiliano Goldwurm Goldwurm, and Violetta Lonati, *On the number of occurrences of a symbol in words of regular languages*, Theoretical Computer Science **302** (2003), no. 1–3, 431–456.
48. A. T. Bharucha-Reid, *Elements of the theory of Markov processes and their applications*, Dover, 1997, A reprint of the original McGraw-Hill edition, 1960.
49. Philippe Biane, *Permutations suivant le type d'excedance et le nombre d'inversions et interprétation combinatoire d'une fraction continue de Heine*, European Journal of Combinatorics **14** (1993), 277–284.
50. Philippe Biane, Jim Pitman, and Marc Yor, *Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions*, Bulletin of the American Mathematical Society (N.S.) **38** (2001), no. 4, 435–465.
51. Norman L. Biggs, *Algebraic graph theory*, Cambridge University Press, 1974.

52. Norman L. Biggs, E. Keith Lloyd, and Robin Wilson, *Graph theory, 1736–1936*, Oxford University Press, 1974.
53. Patrick Billingsley, *Probability and measure*, 2nd ed., John Wiley & Sons, 1986.
54. N. H. Bingham, C. M. Goldie, and J. L. Teugels, *Regular variation*, Encyclopedia of Mathematics and its Applications, vol. 27, Cambridge University Press, Cambridge, 1989.
55. Norman Bleistein and Richard A. Handelsman, *Asymptotic expansions of integrals*, Dover, New York, 1986, A reprint of the second Holt, Rinehart and Winston edition, 1975.
56. Béla Bollobás, *Random graphs*, Academic Press, 1985.
57. Béla Bollobás, Christian Borgs, Jennifer T. Chayes, Jeong Han Kim, and David B. Wilson, *The scaling window of the 2-SAT transition*, Random Structures & Algorithms **18** (2001), no. 3, 201–256.
58. D. Borwein, S. Rankin, and L. Renner, *Enumeration of injective partial transformations*, Discrete Mathematics **73** (1989), 291–296.
59. Jonathan M. Borwein and Peter B. Borwein, *Strange series and high precision fraud*, American Mathematical Monthly **99** (1992), no. 7, 622–640.
60. Jonathan M. Borwein, Peter B. Borwein, and Karl Dilcher, *Pi, Euler numbers and asymptotic expansions*, American Mathematical Monthly **96** (1989), no. 8, 681–687.
61. Jérémie Bourdon and Brigitte Vallée, *Generalized pattern matching statistics*, Mathematics and computer science, II (Versailles, 2002) (B. Chauvin *et al.*, ed.), Trends Math., Birkhäuser, Basel, 2002, pp. 249–265.
62. Mireille Bousquet-Mélou, *A method for the enumeration of various classes of column-convex polygons*, Discrete Math. **154** (1996), no. 1-3, 1–25.
63. ———, *Limit laws for embedded trees. applications to the integrated superbrownian excursion*, ArXiv, 2005, Preprint `math.CO/0501266`. To appear in *Random Structures and Algorithms*.
64. Mireille Bousquet-Mélou and Anthony J. Guttmann, *Enumeration of three-dimensional convex polygons*, Annals of Combinatorics **1** (1997), 27–53.
65. Mireille Bousquet-Mélou and Marko Petkovšek, *Linear recurrences with constant coefficients: the multivariate case*, Discrete Mathematics **225** (2000), no. 1-3, 51–75.
66. Gilles Brassard and Paul Bratley, *Algorithmique: conception et analyse*, Masson, Paris, 1987.
67. Leo Breiman, *Probability*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, Corrected reprint of the 1968 original.
68. W. G. Brown and W. T. Tutte, *On the enumeration of rooted non-separable planar maps*, Canadian Journal of Mathematics **16** (1964), 572–577.
69. W. H. Burge, *An analysis of binary search trees formed from sequences of nondistinct keys*, JACM **23** (1976), no. 3, 451–454.
70. Stanley N. Burris, *Number theoretic density and logical limit laws*, Mathematical Surveys and Monographs, vol. 86, American Mathematical Society, Providence, RI, 2001.
71. E. Rodney Canfield, *Central and local limit theorems for the coefficients of polynomials of binomial type*, Journal of Combinatorial Theory, Series A **23** (1977), 275–290.
72. ———, *Remarks on an asymptotic method in combinatorics*, Journal of Combinatorial Theory, Series A **37** (1984), 348–352.
73. Henri Cartan, *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*, Hermann, 1961.
74. Pierre Cartier and Dominique Foata, *Problèmes combinatoires de commutation et réarrangements*, Lecture Notes in Mathematics, vol. 85, Springer Verlag, 1969.
75. Frédéric Cazals, *Monomer-dimer tilings*, Studies in Automatic Combinatorics **2** (1997), Electronic publication `http://algo.inria.fr/libraries/autocomb/autocomb.html`.
76. Philippe Chassaing and Jean-François Marckert, *Parking functions, empirical processes, and the width of rooted labeled trees*, Electronic Journal of Combinatorics **8** (2001), no. 1, Research Paper 14, 19 pp. (electronic).
77. Philippe Chassaing, Jean-François Marckert, and Marc Yor, *The height and width of simple trees*, Mathematics and computer science (Versailles, 2000), Trends Math., Birkhäuser, Basel, 2000, pp. 17–30.
78. Brigitte Chauvin, Michael Drmota, and Jean Jabbour-Hattab, *The profile of binary search trees*, The Annals of Applied Probability **11** (2001), no. 4, 1042–1062.
79. T. S. Chihara, *An introduction to orthogonal polynomials*, Gordon and Breach, New York, 1978.
80. Noam Chomsky and Marcel Paul Schützenberger, *The algebraic theory of context–free languages*, Computer Programing and Formal Languages (P. Braffort and D. Hirschberg, eds.), North Holland, 1963, pp. 118–161.

81. Julien Clément, Philippe Flajolet, and Brigitte Vallée, *Dynamical sources in information theory: A general analysis of trie structures*, Algorithmica **29** (2001), no. 1/2, 307–369.
82. Louis Comtet, *Advanced combinatorics*, Reidel, Dordrecht, 1974.
83. Robert M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, *On the Lambert W function*, Advances in Computational Mathematics **5** (1996), 329–359.
84. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, New York, 1990.
85. Thomas M. Cover and Joy A. Thomas, *Elements of information theory*, John Wiley & Sons Inc., New York, 1991, A Wiley-Interscience Publication.
86. David Cox, John Little, and Donal O'Shea, *Ideals, varieties, and algorithms: an introduction to computational algebraic geometry and commutative algebra*, 2nd ed., Springer, 1997.
87. H. E. Daniels, *Saddlepoint approximations in statistics*, Annals of Mathematical Statistics **25** (1954), 631–650.
88. Gaston Darboux, *Mémoire sur l'approximation des fonctions de très grands nombres, et sur une classe étendue de développements en série*, Journal de Mathématiques Pures et Appliquées (1878), 5–56,377–416.
89. H. Davenport, *Multiplicative Number Theory*, revised by H. L. Montgomery, second ed., Springer-Verlag, New York, 1980.
90. F. N. David and D. E. Barton, *Combinatorial chance*, Charles Griffin, London, 1962.
91. N. G. de Bruijn, *A combinatorial problem*, Nederl. Akad. Wetensch., Proc. **49** (1946), 758–764, Also in *Indagationes Math.* **8**, 461–467 (1946).
92. ———, *On Mahler's partition problem*, Indagationes Math. **10** (1948), 210–220, Reprinted from Koninkl. Nederl. Akademie Wetenschappen, Ser. A.
93. ———, *Asymptotic methods in analysis*, Dover, 1981, A reprint of the third North Holland edition, 1970 (first edition, 1958).
94. N. G. de Bruijn and D. A. Klarner, *Multisets of aperiodic cycles*, SIAM Journal on Algebraic and Discrete Methods **3** (1982), 359–368.
95. N. G. de Bruijn, D. E. Knuth, and S. O. Rice, *The average height of planted plane trees*, Graph Theory and Computing (R. C. Read, ed.), Academic Press, 1972, pp. 15–22.
96. Percy Deift, *Integrable systems and combinatorial theory*, Notices Amer. Math. Soc. **47** (2000), no. 6, 631–640.
97. J. M. DeLaurentis and B. G. Pittel, *Random permutations and brownian motion*, Pacific Journal of Mathematics **119** (1985), no. 2, 287–301.
98. Marie-Pierre Delest and Gérard Viennot, *Algebraic languages and polyominoes enumeration*, Theoretical Computer Science **34** (1984), 169–206.
99. Michael Dellnitz, Oliver Schütze, and Qinghua Zheng, *Locating all the zeros of an analytic function in one complex variable*, J. Comput. Appl. Math. **138** (2002), no. 2, 325–333.
100. A. Dembo, A. Vershik, and O. Zeitouni, *Large deviations for integer partitions*, Markov Processes and Related Fields **6** (2000), no. 2, 147–179.
101. Frank den Hollander, *Large deviations*, American Mathematical Society, Providence, RI, 2000.
102. Nachum Dershowitz and Shmuel Zaks, *The cycle lemma and some applications*, European Journal of Combinatorics **11** (1990), 35–40.
103. Robert L. Devaney, *A first course in chaotic dynamical systems*, Addison-Wesley Studies in Nonlinearity, Addison-Wesley Publishing Company Advanced Book Program, Reading, MA, 1992, Theory and experiment, With a separately available computer disk.
104. Luc Devroye, *Limit laws for local counters in random binary search trees*, Random Structures & Algorithms **2** (1991), no. 3, 302–315.
105. P. Dienes, *The taylor series*, Dover, New York, 1958, A reprint of the first Oxford University Press edition, 1931.
106. Jean Dieudonné, *Calcul infinitésimal*, Hermann, Paris, 1968.
107. G. Doetsch, *Handbuch der Laplace-Transformation*, vol. 1–3, Birkhäuser Verlag, Basel, 1955.
108. C. Domb and A.J. Barrett, *Enumeration of ladder graphs*, Discrete Mathematics **9** (1974), 341–358.
109. Peter G. Doyle and J. Laurie Snell, *Random walks and electric networks*, Mathematical Association of America, Washington, DC, 1984.
110. Michael Drmota, *Asymptotic distributions and a multivariate Darboux method in enumeration problems*, Manuscript, November 1990.
111. ———, *Systems of functional equations*, Random Structures & Algorithms **10** (1997), no. 1–2, 103–124.

112. Michael Drmota and Bernhard Gittenberger, *On the profile of random trees*, Random Structures & Algorithms **10** (1997), no. 4, 421–451.
113. Michael Drmota and Michèle Soria, *Marking in combinatorial constructions: Generating functions and limiting distributions*, Theoretical Computer Science **144** (1995), no. 1–2, 67–99.
114. Michael Drmota and Michèle Soria, *Images and preimages in random mappings*, SIAM Journal on Discrete Mathematics **10** (1997), no. 2, 246–269.
115. Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer, *Boltzmann samplers for the random generation of combinatorial structures*, Combinatorics, Probability and Computing **13** (2004), no. 4–5, 577–625, Special issue on Analysis of Algorithms.
116. Thomas Duquesne and Jean-Francçois Le Gall, *Random Trees, Levy Processes and Spatial Branching Processes*, ArXiv, 2005, arXiv:math.PR/0509558.
117. Marianne Durand, *Combinatoire analytique et algorithmique des ensembles de données*, Ph.D. thesis, École Polytechnique, France, 2004.
118. Richard Durrett, *Probability: theory and examples*, second ed., Duxbury Press, Belmont, CA, 1996.
119. Isabelle Dutour and Jean-Marc Fédou, *Object grammars and random generation*, Discrete Mathematics and Theoretical Computer Science **2** (1998), 47–61.
120. A. Dvoretzky and Th. Motzkin, *A problem of arrangements*, Duke Mathematical Journal **14** (1947), 305–313.
121. G. P. Egorychev, *Integral representation and the computation of combinatorial sums*, Translations of Mathematical Monographs, vol. 59, American Mathematical Society, Providence, RI, 1984, Translated from the Russian by H. H. McFadden, Translation edited by Lev J. Leifman.
122. Paul Ehrenfest and Tatiana Ehrenfest, *Über zwei bekannte einwände gegen das Boltzmannsche h-theorem*, Physkalische Zeitschrift **8** (1907), no. 9, 311–314.
123. Samuel Eilenberg, *Automata, languages, and machines*, vol. A, Academic Press, 1974.
124. S. Elizalde and M. Noy, *Consecutive patterns in permutations*, Advances in Applied Mathematics **30** (2003), no. 1-2, 110–125.
125. P. D. T. A. Elliott, *Probabilistic number theory. I*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science], vol. 239, Springer-Verlag, New York, 1979, Mean-value theorems.
126. I. G. Entig, *Generating functions for enumerating self-avoiding rings on the square lattice*, Journal of Physics A: Mathematical and General **18** (1980), 3713–3722.
127. P. Erdős and A. Rényi, *On a classical problem of probability theory*, Magyar Tud. Akad. Mat. Kutató Int. Közl. **6** (1961), 215–220.
128. Paul Erdős and Joseph Lehner, *The distribution of the number of summands in the partitions of a positive integer*, Duke Mathematical Journal **8** (1941), 335–345.
129. H. M. Farkas and I. Kra, *Riemann surfaces*, second ed., Graduate Texts in Mathematics, vol. 71, Springer-Verlag, New York, 1992.
130. Guy Fayolle and Roudolf Iasnogorodski, *Two coupled processors: the reduction to a Riemann-Hilbert problem*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **47** (1979), no. 3, 325–351.
131. Guy Fayolle, Roudolf Iasnogorodski, and Vadim Malyshev, *Random walks in the quarter-plane*, Springer-Verlag, Berlin, 1999.
132. Julien Fayolle, *An average-case analysis of basic parameters of the suffix tree*, Mathematics and Computer Science III: Algorithms, Trees, Combinatorics and Probabilities (M. Drmota *et al.*, ed.), Trends in Mathematics, Birkhäuser Verlag, 2004, pp. 217–227.
133. W. Feller, *An introduction to probability theory and its applications*, third ed., vol. 1, John Wiley, 1968.
134. ______, *An introduction to probability theory and its applications*, vol. 2, John Wiley, 1971.
135. James A. Fill, Philippe Flajolet, and Nevin Kapur, *Singularity analysis, Hadamard products, and tree recurrences*, Journal of Computational and Applied Mathematics **174** (2005), 271–313.
136. James Allen Fill, *On the distribution of binary search trees under the random permutation model*, Random Structures & Algorithms **8** (1996), no. 1, 1–25.
137. Steven Finch, *Mathematical constants*, Cambridge University Press, New-York, 2003.
138. Hans Fischer, *Die verschiedenen formen und funktionen des zentralen grenzwertsatzes in der entwicklung von der klassischen zur modernen wahrscheinlichkeitsrechnung*, Shaker Verlag, Aachen, 2000, 318 p. (ISBN: 3-8265-7767-1).
139. Philippe Flajolet, *Combinatorial aspects of continued fractions*, Discrete Mathematics **32** (1980), 125–161.

140. ______, *Analyse d'algorithmes de manipulation d'arbres et de fichiers*, Cahiers du Bureau Universitaire de Recherche Opérationnelle, vol. 34–35, Université Pierre et Marie Curie, Paris, 1981, 209 pages.
141. ______, *On congruences and continued fractions for some classical combinatorial quantities*, Discrete Mathematics **41** (1982), 145–153.
142. ______, *On the performance evaluation of extendible hashing and trie searching*, Acta Informatica **20** (1983), 345–369.
143. ______, *Approximate counting: A detailed analysis*, BIT **25** (1985), 113–134.
144. ______, *Elements of a general theory of combinatorial structures*, Fundamentals of Computation Theory (Lothar Budach, ed.), Lecture Notes in Computer Science, vol. 199, Springer Verlag, 1985, Proceedings of FCT'85, Cottbus, GDR, September 1985 (Invited Lecture), pp. 112–127.
145. ______, *Analytic models and ambiguity of context–free languages*, Theoretical Computer Science **49** (1987), 283–309.
146. ______, *Mathematical methods in the analysis of algorithms and data structures*, Trends in Theoretical Computer Science (Egon Börger, ed.), Computer Science Press, Rockville, Maryland, 1988, (Lecture Notes for *A Graduate Course in Computation Theory*, Udine, 1984), pp. 225–304.
147. ______, *Singularity analysis and asymptotics of Bernoulli sums*, Theoretical Computer Science **215** (1999), no. 1-2, 371–381.
148. ______, *Counting by coin tossings*, Proceedings of ASIAN'04 (Ninth Asian Computing Science Conference) (M. Maher, ed.), Lecture Notes in Computer Science, vol. 3321, 2004, (Text of Opening Keynote Address.), pp. 1–12.
149. Philippe Flajolet, Jean Françon, and Jean Vuillemin, *Sequence of operations analysis for dynamic data structures*, Journal of Algorithms **1** (1980), 111–141.
150. Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond, *The distribution of heights of binary trees and other simple trees*, Combinatorics, Probability and Computing **2** (1993), 145–156.
151. Philippe Flajolet, Danièle Gardy, and Loÿs Thimonier, *Birthday paradox, coupon collectors, caching algorithms, and self–organizing search*, Discrete Applied Mathematics **39** (1992), 207–229.
152. Philippe Flajolet, Stefan Gerhold, and Bruno Salvy, *On the non-holonomic character of logarithms, powers, and the nth prime function*, Electronic Journal of Combinatorics **11(2)** (2005), no. A1, 1–16.
153. Philippe Flajolet, Xavier Gourdon, and Philippe Dumas, *Mellin transforms and asymptotics: Harmonic sums*, Theoretical Computer Science **144** (1995), no. 1–2, 3–58.
154. Philippe Flajolet, Xavier Gourdon, and Conrado Martínez, *Patterns in random binary search trees*, Random Structures & Algorithms **11** (1997), no. 3, 223–244.
155. Philippe Flajolet, Xavier Gourdon, and Daniel Panario, *The complete analysis of a polynomial factorization algorithm over finite fields*, Journal of Algorithms **40** (2001), no. 1, 37–81.
156. Philippe Flajolet, Peter Grabner, Peter Kirschenhofer, and Helmut Prodinger, *On Ramanujan's Q–function*, Journal of Computational and Applied Mathematics **58** (1995), no. 1, 103–116.
157. Philippe Flajolet and Fabrice Guillemin, *The formal theory of birth-and-death processes, lattice path combinatorics, and continued fractions*, Advances in Applied Probability **32** (2000), 750–778.
158. Philippe Flajolet, Yves Guivarc'h, Wojtek Szpankowski, and Brigitte Vallée, *Hidden pattern statistics*, Automata, Languages, and Programming (F. Orejas, P. Spirakis, and J. van Leeuwen, eds.), Lecture Notes in Computer Science, no. 2076, Springer Verlag, 2001, Proceedings of the 28th ICALP Conference, Crete, July 2001., pp. 152–165.
159. Philippe Flajolet, Kostas Hatzis, Sotiris Nikoletseas, and Paul Spirakis, *On the robustness of interconnections in random graphs: A symbolic approach*, Theoretical Computer Science **287** (2002), no. 2, 513–534.
160. Philippe Flajolet, Hsien-Kuei Hwang, and Michèle Soria, *The ubiquitous Gaussian law in analytic combinatorics*, 1997, In preparation.
161. Philippe Flajolet, Peter Kirschenhofer, and Robert F. Tichy, *Deviations from uniformity in random strings*, Probability Theory and Related Fields **80** (1988), 139–150.
162. Philippe Flajolet, Donald E. Knuth, and Boris Pittel, *The first cycles in an evolving graph*, Discrete Mathematics **75** (1989), 167–215.
163. Philippe Flajolet and Thomas Lafforgue, *Search costs in quadtrees and singularity perturbation asymptotics*, Discrete and Computational Geometry **12** (1994), no. 4, 151–175.
164. Philippe Flajolet and Marc Noy, *Analytic combinatorics of non-crossing configurations*, Discrete Mathematics **204** (1999), no. 1-3, 203–229, (Selected papers in honor of Henry W. Gould).
165. Philippe Flajolet and Andrew M. Odlyzko, *The average height of binary trees and other simple trees*, Journal of Computer and System Sciences **25** (1982), 171–213.

166. ______, *Random mapping statistics*, Advances in Cryptology (J-J. Quisquater and J. Vandewalle, eds.), Lecture Notes in Computer Science, vol. 434, Springer Verlag, 1990, Proceedings of EUROCRYPT'89, Houtalen, Belgium, April 1989, pp. 329–354.
167. ______, *Singularity analysis of generating functions*, SIAM Journal on Algebraic and Discrete Methods **3** (1990), no. 2, 216–240.
168. Philippe Flajolet, Patricio Poblete, and Alfredo Viola, *On the analysis of linear probing hashing*, Algorithmica **22** (1998), no. 4, 490–515.
169. Philippe Flajolet and Helmut Prodinger, *Level number sequences for trees*, Discrete Mathematics **65** (1987), 149–156.
170. ______, *On Stirling numbers for complex argument and Hankel contours*, SIAM Journal on Discrete Mathematics **12** (1999), no. 2, 155–159.
171. Philippe Flajolet, Mireille Régnier, and Robert Sedgewick, *Some uses of the Mellin integral transform in the analysis of algorithms*, Combinatorial Algorithms on Words (A. Apostolico and Z. Galil, eds.), NATO Advance Science Institute Series. Series F: Computer and Systems Sciences, vol. 12, Springer Verlag, 1985, (Invited Lecture), pp. 241–254.
172. Philippe Flajolet, Bruno Salvy, and Gilles Schaeffer, *Airy phenomena and analytic combinatorics of connected graphs*, Electronic Journal of Combinatorics **11** (2004), no. 2:#R34, 1–30.
173. Philippe Flajolet, Bruno Salvy, and Paul Zimmermann, *Automatic average–case analysis of algorithms*, Theoretical Computer Science **79** (1991), no. 1, 37–109.
174. Philippe Flajolet and Robert Sedgewick, *Mellin transforms and asymptotics: finite differences and Rice's integrals*, Theoretical Computer Science **144** (1995), no. 1–2, 101–124.
175. ______, *The average case analysis of algorithms: Mellin transform asymptotics*, Research Report 2956, Institut National de Recherche en Informatique et en Automatique, 1996, 93 pages.
176. Philippe Flajolet, Paolo Sipala, and Jean-Marc Steyaert, *Analytic variations on the common subexpression problem*, Automata, Languages, and Programming (M. S. Paterson, ed.), Lecture Notes in Computer Science, vol. 443, 1990, Proceedings of the 17th ICALP Conference, Warwick, July 1990, pp. 220–234.
177. Philippe Flajolet and Michèle Soria, *Gaussian limiting distributions for the number of components in combinatorial structures*, Journal of Combinatorial Theory, Series A **53** (1990), 165–182.
178. ______, *The cycle construction*, SIAM Journal on Discrete Mathematics **4** (1991), no. 1, 58–60.
179. ______, *General combinatorial schemas: Gaussian limit distributions and exponential tails*, Discrete Mathematics **114** (1993), 159–180.
180. Philippe Flajolet and Jean-Marc Steyaert, *A complexity calculus for classes of recursive search programs over tree structures*, Proceedings of the 22nd Annual Symposium on Foundations of Computer Science, IEEE Computer Society Press, 1981, pp. 386–393.
181. ______, *A complexity calculus for recursive tree algorithms*, Mathematical Systems Theory **19** (1987), 301–331.
182. Philippe Flajolet, Wojciech Szpankowski, and Brigitte Vallée, *Hidden word statistics*, Preprint, 2002, 34 pages; submitted to *J. of the ACM*.
183. Philippe Flajolet, Paul Zimmerman, and Bernard Van Cutsem, *A calculus for the random generation of labelled combinatorial structures*, Theoretical Computer Science **132** (1994), no. 1-2, 1–35.
184. Dominique Foata, *La série génératrice exponentielle dans les problèmes d'énumération*, S.M.S, Montreal University Press, 1974.
185. Dominique Foata, Bodo Lass, and Guo-Niu Han, *Les nombres hyperharmoniques et la fratrie du collectionneur de vignettes*, Seminaire Lotharingien de Combinatoire **47** (2001), Paper B47a.
186. Dominique Foata and Marcel-P. Schützenberger, *Théorie géométrique des polynômes Euleriens*, Lecture Notes in Mathematics, vol. 138, Springer Verlag, 1970.
187. W. B. Ford, *Studies on divergent series and summability and the asymptotic developments of functions defined by Maclaurin series*, 3rd ed., Chelsea Publishing Company, New York, 1960, (From two books originally published in 1916 and 1936.).
188. Jean Françon and Gérard Viennot, *Permutations selon leurs pics, creux, doubles montées et doubles descentes, nombres d'Euler et de Genocchi*, Discrete Mathematics **28** (1979), 21–35.
189. G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, Sitz.-Ber. Akad. Wiss., Phys-Math Klasse, Berlin (1912), 456–477.
190. William Fulton, *Algebraic curves*, W.A. Benjamin, Inc., New York, Amsterdam, 1969.
191. F. R. Gantmacher, *Matrizentheorie*, Deutscher Verlag der Wissenschaften, Berlin, 1986, A translation of the Russian original Teoria Matriz, Nauka, Moscow, 1966.

192. Zhicheng Gao and L. Bruce Richmond, *Central and local limit theorems applied to asymptotic enumerations IV: Multivariate generating functions*, Journal of Computational and Applied Mathematics **41** (1992), 177–186.
193. Danièle Gardy, *Méthode de col et lois limites en analyse combinatoire*, Theoretical Computer Science **92** (1992), no. 2, 261–280.
194. ______, *Normal limiting distributions for projection and semijoin sizes*, SIAM Journal on Discrete Mathematics **5** (1992), no. 2, 219–248.
195. ______, *Some results on the asymptotic behaviour of coefficients of large powers of functions*, Discrete Mathematics **139** (1995), no. 1-3, 189–217.
196. George Gasper and Mizan Rahman, *Basic hypergeometric series*, Encyclopedia of Mathematics and its Applications, vol. 35, Cambridge University Press, 1990.
197. Ira M. Gessel, *A factorization for formal Laurent series and lattice path enumeration*, J. Combin. Theory Ser. A **28** (1980), no. 3, 321–337.
198. ______, *A noncommutative generalization and q–analog of the Lagrange inversion formula*, Transactions of the American Mathematical Society **257** (1980), no. 2, 455–482.
199. Ira M. Gessel, *Symmetric functions and P–recursiveness*, Journal of Combinatorial Theory, Series A **53** (1990), 257–285.
200. M. L. Glasser, *A Watson sum for a cubic lattice*, Journal of Mathematical Physics **13** (1972), 1145–1146.
201. B. V. Gnedenko and A. N. Kolmogorov, *Limit distributions for sums of independent random variables*, Addison-Wesley, 1968, Translated from the Russian original (1949).
202. C. D. Godsil, *Algebraic combinatorics*, Chapman and Hall, 1993.
203. Massimiliano Goldwurm and Massimo Santini, *Clique polynomials have a unique root of smallest modulus*, Information Processing Letters **75** (2000), no. 3, 127–132.
204. V. Goncharov, *On the field of combinatory analysis*, Soviet Math. Izv., Ser. Math. **8** (1944), 3–48, In Russian.
205. Gaston H. Gonnet, *Notes on the derivation of asymptotic expressions from summations*, Information Processing Letters **7** (1978), no. 4, 165–169.
206. Gaston H. Gonnet, *Expected length of the longest probe sequence in hash code searching*, Journal of the ACM **28** (1981), no. 2, 289–304.
207. I. J. Good, *Random motion and analytic continued fractions*, Proceedings of the Cambridge Philosophical Society **54** (1958), 43–47.
208. Ian P. Goulden and David M. Jackson, *Combinatorial enumeration*, John Wiley, New York, 1983.
209. ______, *Distributions, continued fractions, and the Ehrenfest urn model*, Journal of Combinatorial Theory. Series A **41** (1986), no. 1, 21–31.
210. Xavier Gourdon, *Largest component in random combinatorial structures*, Discrete Mathematics **180** (1998), no. 1-3, 185–209.
211. Xavier Gourdon and Bruno Salvy, *Asymptotics of linear recurrences with rational coefficients*, Tech. Report 1887, INRIA, March 1993, To appear in Proceedings FPACS'93.
212. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, Addison Wesley, 1989.
213. D. H. Greene and D. E. Knuth, *Mathematics for the analysis of algorithms*, Birkhäuser, Boston, 1981.
214. ______, *Mathematics for the analysis of algorithms*, second ed., Birkhauser, Boston, 1982.
215. Daniel Hill Greene, *Labelled formal languages and their uses*, Ph.D. thesis, Stanford University, June 1983, Available as Report STAN-CS-83-982.
216. L. J. Guibas and A. M. Odlyzko, *Long repetitive patterns in random sequences*, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete **53** (1980), no. 3, 241–262.
217. ______, *String overlaps, pattern matching, and nontransitive games*, Journal of Combinatorial Theory. Series A **30** (1981), no. 2, 183–208.
218. Leo J. Guibas and Andrew M. Odlyzko, *Periods in strings*, Journal of Combinatorial Theory, Series A **30** (1981), 19–42.
219. Fabrice Guillemin, Philippe Robert, and Bert Zwart, *AIMD algorithms and exponential functionals*, Annals of Applied Probability **14** (2004), no. 1, 90–117.
220. Laurent Habsieger, Maxime Kazarian, and Sergei Lando, *On the second number of Plutarch*, American Mathematical Monthly **105** (1998), 446–447.
221. Jennie C. Hansen, *A functional central limit theorem for random mappings*, Annals of Probability **17** (1989), no. 1.

222. F. Harary, R. W. Robinson, and A. J. Schwenk, *Twenty-step algorithm for determining the asymptotic number of trees of various species*, Journal of the Australian Mathematical Society (Series A) **20** (1975), 483–503.
223. Frank Harary and Edgar M. Palmer, *Graphical enumeration*, Academic Press, 1973.
224. G. H. Hardy, *Ramanujan: Twelve lectures on subjects suggested by his life and work*, third ed., Chelsea Publishing Company, New-York, 1978, Reprinted and Corrected from the First Edition, Cambridge, 1940.
225. Bernard Harris and Lowell Schoenfeld, *Asymptotic expansions for the coefficients of analytic functions*, Illinois Journal of Mathematics **12** (1968), 264–277.
226. Theodore E. Harris, *The theory of branching processes*, Dover Publications, 1989, A reprint of the 1963 edition.
227. W. K. Hayman, *A generalization of Stirling's formula*, Journal für die reine und angewandte Mathematik **196** (1956), 67–95.
228. Erich Hecke, *Vorlesungen über die Theorie der algebraischen Zahlen*, Akademische Verlagsgesellschaft, Leipzig, 1923.
229. Peter Henrici, *Applied and computational complex analysis*, vol. 2, John Wiley, New York, 1974.
230. ———, *Applied and computational complex analysis*, vol. 1, John Wiley, New York, 1974.
231. Dean Hickerson, *Counting horizontally convex polyominoes*, Journal of Integer Sequences **2** (1999), Electronic.
232. E. Hille, *Analytic function theory*, Blaisdell Publishing Company, Waltham, 1962, 2 Volumes.
233. Micha Hofri, *Analysis of algorithms: Computational methods and mathematical tools*, Oxford University Press, 1995.
234. J. A. Howell, T. F. Smith, and M. S. Waterman, *Computation of generating functions for biological molecules*, SIAM Journal on Applied Mathematics **39** (1980), no. 1, 119–133.
235. Hsien-Kuei Hwang, *Théorèmes limites pour les structures combinatoires et les fonctions arithmetiques*, Ph.D. thesis, École Polytechnique, December 1994.
236. ———, *Large deviations for combinatorial distributions. I. Central limit theorems*, The Annals of Applied Probability **6** (1996), no. 1, 297–319.
237. ———, *Large deviations of combinatorial distributions. II. Local limit theorems*, The Annals of Applied Probability **8** (1998), no. 1, 163–181.
238. ———, *On convergence rates in the central limit theorems for combinatorial structures*, European Journal of Combinatorics **19** (1998), no. 3, 329–343.
239. ———, *Asymptotics of Poisson approximation to random discrete distributions: an analytic approach*, Advances in Applied Probability **31** (1999), no. 2, 448–491.
240. Mourad E. H. Ismail, *Classical and quantum orthogonal polynomials in one variable*, Encyclopedia of Mathematics and its Applications, no. 98, Cambridge University Press, 2005.
241. Philippe Jacquet and Mireille Régnier, *Trie partitioning process: Limiting distributions*, CAAP'86 (P. Franchi-Zanetacchi, ed.), Lecture Notes in Computer Science, vol. 214, 1986, Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice France, March 1986., pp. 196–210.
242. Philippe Jacquet and Wojciech Szpankowski, *Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees*, Theoretical Computer Science **144** (1995), no. 1–2, 161–197.
243. Svante Janson, *Random cutting and records in deterministic and random trees*, Technical Report, 2004, *Random Structures & Algorithms*, 42 pages, to appear.
244. Svante Janson, Donald E. Knuth, Tomasz Łuczak, and Boris Pittel, *The birth of the giant component*, Random Structures & Algorithms **4** (1993), no. 3, 233–358.
245. Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience, New York, 2000.
246. Iwan Jensen, *A parallel algorithm for the enumeration of self-avoiding polygons on the square lattice*, Journal of Physics A: Mathematical and General **36** (2003), 5731–5745.
247. William B. Jones and Arne Magnus, *Application of Stieltjes fractions to birth-death processes*, Padé and rational approximation (New York) (E. B. Saff and Richard S. Varga, eds.), Academic Press Inc., 1977, Proceedings of an International Symposium held at the University of South Florida, Tampa, Fla., December 15-17, 1976, pp. 173–179.
248. André Joyal, *Une théorie combinatoire des séries formelles*, Advances in Mathematics **42** (1981), no. 1, 1–82.
249. R. Jungen, *Sur les séries de Taylor n'ayant que des singularités algébrico-logarithmiques sur leur cercle de convergence*, Commentarii Mathematici Helvetici **3** (1931), 266–306.
250. Mark Kac, *Random walk and the theory of Brownian motion*, American Mathematical Monthly **54** (1947), 369–391.

251. Samuel Karlin and James McGregor, *The classification of birth and death processes*, Trans. Amer. Math. Soc. **86** (1957), 366–400.
252. Samuel Karlin and Howard Taylor, *A first course in stochastic processes*, second ed., Academic Press, 1975.
253. Rainer Kemp, *Random multidimensional binary trees*, Journal of Information Processing and Cybernetics (EIK) **29** (1993), 9–36.
254. Frances Kirwan, *Complex algebraic curves*, London Mathematical Society Student Texts, no. 23, Cambridge University Press, 1992.
255. M. S. Klamkin and D. J. Newman, *Extensions of the birthday surprise*, Journal of Combinatorial Theory **3** (1967), 279–282.
256. S. C. Kleene, *Representation of events in nerve nets and finite automata*, Automata studies, Princeton University Press, Princeton, N. J., 1956, pp. 3–41.
257. A. Knopfmacher, A. M. Odlyzko, B. Pittel, L. B. Richmond, D. Stark, G. Szekeres, and N. C. Wormald, *The asymptotic number of set partitions with unequal block sizes*, Electronic Journal of Combinatorics **6** (1999), no. 1, R2:1–37.
258. Arnold Knopfmacher and Helmut Prodinger, *On Carlitz compositions*, European Journal of Combinatorics **19** (1998), no. 5, 579–589.
259. John Knopfmacher, *Abstract analytic number theory*, Dover, 1990.
260. John Knopfmacher and Arnold Knopfmacher, *Counting irreducible factors of polynomials over a finite field*, Discrete Mathematics **112** (1993), 103–118.
261. K. Knopp, *Theory of functions*, Dover Publications, New York, 1945.
262. Donald E. Knuth, *The art of computer programming*, vol. 1: Fundamental Algorithms, Addison-Wesley, 1968, Second edition, 1973.
263. ———, *Mathematical analysis of algorithms*, Information Processing 71, North Holland Publishing Company, 1972, Proceedings of IFIP Congress, Ljubljana, 1971, pp. 19–27.
264. ———, *The art of computer programming*, vol. 3: Sorting and Searching, Addison-Wesley, 1973.
265. ———, *The average time for carry propagation*, Indagationes Mathematicae **40** (1978), 238–242.
266. ———, *The art of computer programming*, 2nd ed., vol. 2: Seminumerical Algorithms, Addison-Wesley, 1981.
267. Donald E. Knuth, *Bracket notation for the 'coefficient of' operator*, E-print `arXiv:math/9402216`, February 1994.
268. Donald E. Knuth, *The art of computer programming*, 3rd ed., vol. 1: Fundamental Algorithms, Addison-Wesley, 1997.
269. ———, *The art of computer programming*, 2nd ed., vol. 3: Sorting and Searching, Addison-Wesley, 1998.
270. ———, *The art of computer programming*, 3rd ed., vol. 2: Seminumerical Algorithms, Addison-Wesley, 1998.
271. ———, *Selected papers on analysis of algorithms*, CSLI Publications, Stanford, CA, 2000.
272. Donald E. Knuth, James H. Morris, Jr., and Vaughan R. Pratt, *Fast pattern matching in strings*, SIAM Journal on Computing **6** (1977), no. 2, 323–350.
273. Donald E. Knuth and Boris Pittel, *A recurrence related to trees*, Proceedings of the American Mathematical Society **105** (1989), no. 2, 335–349.
274. Donald E. Knuth and Arnold Schönhage, *The expected linearity of a simple equivalence algorithm*, Theoretical Computer Science **6** (1978), 281–315.
275. Donald E. Knuth and Ilan Vardi, *Problem 6581 (the asymptotic expansion of $2n$ choose n)*, American Mathematical Monthly **95** (1988), 774.
276. Valentin F. Kolchin, *Random mappings*, Optimization Software Inc., New York, 1986, Translated from *Slučajnye Otobraženija*, Nauka, Moscow, 1984.
277. ———, *Random graphs*, Encyclopedia of Mathematics and its Applications, vol. 53, Cambridge University Press, Cambridge, U.K., 1999.
278. Valentin F. Kolchin, Boris A. Sevastyanov, and Vladimir P. Chistyakov, *Random allocations*, John Wiley and Sons, New York, 1978, Translated from the Russian original *Slučajnye Razmeščeniya*.
279. Jacob Korevaar, *Tauberian theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 329, Springer-Verlag, Berlin, 2004, A century of developments.
280. M. Krasnoselskii, *Positive solutions of operator equations*, P. Noordhoff, Groningen, 1964.
281. Christian Krattenthaler, *Advanced determinant calculus*, Seminaire Lotharingien de Combinatoire **42** (1999), Paper B42q, 66 pp.
282. Jacques Labelle and Yeong Nan Yeh, *Generalized Dyck paths*, Discrete Mathematics **82** (1990), 1–6.

283. J. C. Lagarias and A. M. Odlyzko, *Solving low-density subset sum problems*, JACM **32** (1985), no. 1, 229–246.
284. J. C. Lagarias, A. M. Odlyzko, and D. B. Zagier, *On the capacity of disjointly shared networks*, Computer Netwoks **10** (1985), no. 5, 275–285.
285. Steven P. Lalley, *Finite range random walk on free groups and homogeneous trees*, Ann. Probab. **21** (1993), no. 4, 2087–2130.
286. Serge Lang, *Linear algebra*, Addison-Wesley, Reading, Mass., 1966.
287. Pierre-Simon Laplace, *Théorie analytique des probabilités. Vol. I, II*, Éditions Jacques Gabay, Paris, 1995, Reprint of the 1819 and 1820 editions.
288. Gregory F. Lawler, *Intersections of random walks*, Birkhäuser Boston Inc., Boston, MA, 1991.
289. Léonce Lesieur and Jean-Louis Nicolas, *On the Eulerian numbers $M_n = \max A_{n,k}$*, European Journal of Combinatorics **13** (1992), 379–399.
290. L. Lewin (ed.), *Structural properties of polylogarithms*, American Mathematical Society, 1991.
291. V. Lifschitz and B. Pittel, *The number of increasing subsequences of the random permutation*, Journal of Combinatorial Theory, Series A **31** (1981), 1–20.
292. Ernst Lindelöf, *Le calcul des résidus et ses applications à la théorie des fonctions*, Collection de monographies sur la théorie des fonctions, publiée sous la direction de M. Émile Borel, Gauthier-Villars, Paris, 1905, Reprinted by Gabay, Paris, 1989.
293. B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Mathematics **26** (1977), 206–222.
294. M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 17, Addison–Wesley, 1983.
295. M. Lothaire, *Applied combinatorics on words*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, Cambridge, 2005, (A collective work edited by Jean Berstel and Dominique Perrin).
296. G. Louchard, *Kac's formula, Lévy's local time and Brownian excursion*, Journal of Applied Probability **21** (1984), 479–499.
297. Guy Louchard, *The Brownian excursion: a numerical analysis*, Computers and Mathematics with Applications **10** (1984), no. 6, 413–417.
298. ———, *Random walks, Gaussian processes and list structures*, Theoretical Computer Science **53** (1987), no. 1, 99–124.
299. ———, *Probabilistic analysis of some (un)directed animals*, Theoretical Computer Science **159** (1996), no. 1, 65–79.
300. ———, *Probabilistic analysis of column-convex and directed diagonally-convex animals*, Random Structures & Algorithms **11** (1997), no. 2, 151–178.
301. Guy Louchard and Helmut Prodinger, *Probabilistic analysis of Carlitz compositions*, Discrete Mathematics & Theoretical Computer Science **5** (2002), no. 1, 71–96.
302. Guy Louchard, R. Schott, M. Tolley, and P. Zimmermann, *Random walks, heat equations and distributed algorithms*, Journal of Computational and Applied Mathematics **53** (1994), 243–274.
303. E. Lucas, *Théorie des Nombres*, Gauthier-Villard, Paris, 1891, Reprinted by A. Blanchard, Paris 1961.
304. V. Y. Lum, P. S. T. Yuen, and M. Dodd, *Key to address transformations: A fundamental study based on large existing format files*, Communications of the ACM **14** (1971), 228–239.
305. David J. C. MacKay, *Information theory, inference and learning algorithms*, Cambridge University Press, New York, 2003.
306. P. A. MacMahon, *Introduction to combinatory analysis*, Chelsea Publishing Co., New York, 1955, A reprint of the first edition, Cambridge, 1920.
307. Hosam M. Mahmoud, *Evolution of random search trees*, John Wiley, New York, 1992.
308. Conrado Martínez and Xavier Molinero, *A generic approach for the unranking of labeled combinatorial classes*, Random Structures & Algorithms **19** (2001), no. 3-4, 472–497, Analysis of algorithms (Krynica Morska, 2000).
309. J. E. Mazo and A. M. Odlyzko, *Lattice points in high-dimensional spheres*, Monatshefte für Mathematik **110** (1990), no. 1, 47–61.
310. B. D. McKay, D. Bar-Natan, M. Bar-Hillel, and G. Kalai, *Solving the bible code puzzle*, Statistical Science **14** (1999), 150–173.
311. Günter Meinardus, *Asymptotische Aussagen über Partitionen*, Mathematische Zeitschrift **59** (1954), 388–398.
312. A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canadian Journal of Mathematics **30** (1978), 997–1015.
313. ———, *On random mapping patterns*, Combinatorica **4** (1984), no. 1, 61–70.

314. ———, *Erratum: "On an asymptotic method in enumeration"*, J. Combin. Theory Ser. A **52** (1989), no. 1, 163.
315. ———, *On an asymptotic method in enumeration*, J. Combin. Theory Ser. A **51** (1989), no. 1, 77–89.
316. ———, *The asymptotic behaviour of coefficients of powers of certain generating functions*, European Journal of Combinatorics **11** (1990), 581–587.
317. John Milnor, *Dynamics in one complex variable*, Friedr. Vieweg & Sohn, Braunschweig, 1999.
318. J. W. Moon, *Counting labelled trees*, Canadian Mathematical Monographs N.1, William Clowes and Sons, 1970.
319. Macdonald Morris, Gabriel Schachtel, and Samuel Karlin, *Exact formulas for multitype run statistics in a random ordering*, SIAM Journal on Discrete Mathematics **6** (1993), no. 1, 70–86.
320. Leo Moser and Max Wyman, *On the solution of $x^d = 1$ in symmetric groups*, Canadian Journal of Mathematics **7** (1955), 159–168.
321. ———, *Asymptotic expansions*, Canadian Journal of Mathematics **8** (1956), 225–233.
322. ———, *Asymptotic expansions II*, Canadian Journal of Mathematics (1957), 194–209.
323. Rajeev Motwani and Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, 1995.
324. G. Myerson and A. J. van der Poorten, *Some problems concerning recurrence sequences*, The American Mathematical Monthly **102** (1995), no. 8, 698–705.
325. Donald J. Newman and Lawrence Shepp, *The double dixie cup problem*, American Mathematical Monthly **67** (1960), 58–61.
326. Albert Nijenhuis and Herbert S. Wilf, *Combinatorial algorithms*, second ed., Academic Press, 1978.
327. Niels Erik Nörlund, *Vorlesungen über Differenzenrechnung*, Chelsea Publishing Company, New York, 1954.
328. A. M. Odlyzko, *Periodic oscillations of coefficients of power series that satisfy functional equations*, Advances in Mathematics **44** (1982), 180–205.
329. ———, *Explicit Tauberian estimates for functions with positive coefficients*, Journal of Computational and Applied Mathematics **41** (1992), 187–197.
330. ———, *Asymptotic enumeration methods*, Handbook of Combinatorics (R. Graham, M. Grötschel, and L. Lovász, eds.), vol. II, Elsevier, Amsterdam, 1995, pp. 1063–1229.
331. A. M. Odlyzko and L. B. Richmond, *Asymptotic expansions for the coefficients of analytic generating functions*, Aequationes Mathematicae **28** (1985), 50–63.
332. A. M. Odlyzko and H. S. Wilf, *Bandwidths and profiles of trees*, Journal of Combinatorial Theory, Series B **42** (1987), 348–370.
333. ———, *The editor's corner: n coins in a fountain*, American Matematical Monthly **95** (1988), 840–843.
334. F. W. J. Olver, *Asymptotics and special functions*, Academic Press, 1974.
335. Richard Otter, *The number of trees*, Annals of Mathematics **49** (1948), no. 3, 583–599.
336. Daniel Panario, Boris Pittel, Bruce Richmond, and Alfredo Viola, *Analysis of Rabin's irreducibility test for polynomials over finite fields*, Random Structures Algorithms **19** (2001), no. 3-4, 525–551, Analysis of algorithms (Krynica Morska, 2000).
337. Daniel Panario and Bruce Richmond, *Analysis of Ben-Or's polynomial irreducibility test*, Proceedings of the Eighth International Conference "Random Structures and Algorithms" (Poznan, 1997), vol. 13, 1998, pp. 439–456.
338. ———, *Exact largest and smallest size of components*, Algorithmica **31** (2001), no. 3, 413–432.
339. R. B. Paris and D. Kaminski, *Asymptotics and Mellin-Barnes integrals*, Encyclopedia of Mathematics and its Applications, vol. 85, Cambridge University Press, Cambridge, 2001.
340. H.-O. Peitgen and P. H. Richter, *The beauty of fractals*, Springer-Verlag, Berlin, 1986, Images of complex dynamical systems.
341. Robin Pemantle, *Generating functions with high-order poles are nearly polynomial*, Mathematics and computer science (Versailles, 2000), Birkhäuser, Basel, 2000, pp. 305–321.
342. Robin Pemantle and Mark C. Wilson, *Asymptotics of multivariate sequences. I. Smooth points of the singular variety*, Journal of Combinatorial Theory. Series A **97** (2002), no. 1, 129–161.
343. ———, *Asymptotics of multivariate sequences, Part II: Multiple points of the singular variety*, Combinatorics, Probability and Computing **13** (2004), no. 4–5, 735–761.
344. J. K. Percus, *Combinatorial Methods*, Applied Mathematical Sciences, vol. 4, Springer-Verlag, 1971.
345. Oskar Perron, *Über Matrizen*, Mathematische Annalen **64** (1907), 248–263.
346. Valentin V. Petrov, *Limit theorems of probability theory*, Oxford Studies in Probability, vol. 4, The Clarendon Press Oxford University Press, New York, 1995, Sequences of independent random variables, Oxford Science Publications.

347. G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Mathematica **68** (1937), 145–254.
348. ———, *On the number of certain lattice polygons*, Journal of Combinatorial Theory, Series A **6** (1969), 102–105.
349. G. Pólya and R. C. Read, *Combinatorial enumeration of groups, graphs and chemical compounds*, Springer Verlag, New York, 1987.
350. George Pólya, Robert E. Tarjan, and Donald R. Woods, *Notes on introductory combinatorics*, Progress in Computer Science, Birkhäuser, 1983.
351. A. G. Postnikov, *Tauberian theory and its applications*, Proceedings of the Steklov Institute of Mathematics, vol. 144, American Mathematical Society, 1980.
352. Nicolas Pouyanne, *On the number of permutations admitting an mth root*, Electronic Journal of Combinatorics **9** (2002), no. 1:R3, 1–12.
353. Thomas Prellberg, *Uniform q-series asymptotics for staircase polygons*, Journal of Physics A: Math. Gen. **28** (1995), 1289–1304.
354. H. A. Priestley, *Introduction to complex analysis*, Oxford University Press, Oxford, 1985.
355. Helmut Prodinger, *Approximate counting via Euler transform*, Mathematica Slovaka **44** (1994), 569–574.
356. ———, *A note on the distribution of the three types of nodes in uniform binary trees*, Séminaire Lotharingien de Combinatoire **38** (1996), Paper B38b, 5 pages.
357. Andrzej Proskurowski, Frank Ruskey, and Malcolm Smith, *Analysis of algorithms for listing equivalence classes of k-ary strings*, SIAM Journal on Discrete Mathematics **11** (1998), no. 1, 94–109 (electronic).
358. E. M. Rains and N. J. A. Sloane, *On Cayley's enumeration of alkanes (or 4-valent trees)*, Journal of Integer Sequences **2** (1999), Article 99.1.1; avilable electronically.
359. G. N. Raney, *Functional composition patterns and power series reversion*, Transactions of the American Mathematical Society **94** (1960), 441–451.
360. Mireille Régnier, *Analysis of grid file algorithms*, BIT **25** (1985), 335–357.
361. A. Rényi and G. Szekeres, *On the height of trees*, Australian Journal of Mathematics **7** (1967), 497–507.
362. P. Révész, *Strong theorems on coin tossing*, Proceedings of the International Congress of Mathematicians (Helsinki, 1978) (Helsinki), Acad. Sci. Fennica, 1980, pp. 749–754.
363. Christoph Richard, *Scaling behaviour of two-dimensional polygon models*, Journal of Statistical Physics **108** (2002), no. 3/4, 459–493.
364. Philippe Robert, *Réseaux et files d'attente: méthodes probabilistes*, Mathématiques & Applications, vol. 35, Springer, Paris, 2000.
365. Gian-Carlo Rota, *Finite operator calculus*, Academic Press, 1975.
366. K. F. Roth and G. Szekeres, *Some asymptotic formulae in the theory of partitions*, Quart. J. Math., Oxford Ser. (2) **5** (1954), 241–259.
367. Salvador Roura and Conrado Martínez, *Randomization of search trees by subtree size*, Algorithms—ESA'96 (Josep Diaz and Maria Serna, eds.), Lecture Notes in Computer Science, no. 1136, 1996, Proceedings of the Fourth European Symposium on Algorithms, Barcelona, September 1996., pp. 91–106.
368. Walter Rudin, *Real and complex analysis*, third ed., McGraw-Hill Book Co., New York, 1987.
369. Vladimir N. Sachkov, *Combinatorial methods in discrete mathematics*, Encyclopedia of Mathematics and its Applications, vol. 55, Cambridge University Press, 1996.
370. ———, *Probabilistic methods in combinatorial analysis*, Cambridge University Press, Cambridge, 1997, Translated and adapted from the Russian original edition, Nauka, Moscow, 1978.
371. V.N. Sachkov, *Verojatnostnye metody v kombinatornom analize*, Nauka, Moscow, 1978.
372. Arto Salomaa and Matti Soittola, *Automata-theoretic aspects of formal power series*, Springer, Berlin, 1978.
373. Bruno Salvy, *Asymptotique automatique et fonctions génératrices*, Ph. D. thesis, École Polytechnique, 1991.
374. Bruno Salvy, *Asymptotics of the stirling numbers of the second kind*, Studies in Automatic Combinatorics **II** (1997), Published electronically.
375. Bruno Salvy and John Shackell, *Symbolic asymptotics: multiseries of inverse functions*, Journal of Symbolic Computation **27** (1999), no. 6, 543–563.
376. Bruno Salvy and Paul Zimmermann, *GFUN: a Maple package for the manipulation of generating and holonomic functions in one variable*, ACM Transactions on Mathematical Software **20** (1994), no. 2, 163–167.

377. Gilles Schaeffer, *Conjugaison d'arbres et cartes combinatoires aléatoires*, Ph.d. thesis, Université de Bordeaux I, December 1998.
378. William R. Schmitt and Michael S. Waterman, *Linear trees and RNA secondary structure*, Discrete Applied Mathematics. Combinatorial Algorithms, Optimization and Computer Science **51** (1994), no. 3, 317–323.
379. Erwin Schrödinger, *Statistical thermodynamics*, A course of seminar lectures delivered in January-March 1944, at the School of Theoretical Physics, Dublin Institute for Advanced Studies. Second edition, reprinted, Cambridge University Press, New York, 1962.
380. Robert Sedgewick, *Quicksort with equal keys*, SIAM Journal on Computing **6** (1977), no. 2, 240–267.
381. ———, *Algorithms*, second ed., Addison–Wesley, Reading, Mass., 1988.
382. Robert Sedgewick and Philippe Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley Publishing Company, 1996.
383. L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Transactions of the American Mathematical Society **121** (1966), 340–357.
384. N. J. A. Sloane, *The On-Line Encyclopedia of Integer Sequences*, 2000, Published electronically at `http://www.research.att.com/~njas/sequences/`.
385. N. J. A. Sloane and Simon Plouffe, *The encyclopedia of integer sequences*, Academic Press, 1995.
386. Neil J. A. Sloane and Thomas Wider, *The number of hierarchical orderings*, Order **??** (2003), ??–??, In print.
387. M. Soittola, *Positive rational sequences*, Theoretical Computer Science **2** (1976), 317–322.
388. Michèle Soria-Cousineau, *Méthodes d'analyse pour les constructions combinatoires et les algorithmes*, Doctorat ès sciences, Université de Paris–Sud, Orsay, July 1990.
389. George Springer, *Introduction to Riemann surfaces*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1957, Reprinted by Chelsea, New York.
390. Richard P. Stanley, *Generating functions*, Studies in Combinatorics, M.A.A. Studies in Mathematics, Vol. 17. (G-C. Rota, ed.), The Mathematical Association of America, 1978, pp. 100–141.
391. ———, *Enumerative combinatorics*, vol. I, Wadsworth & Brooks/Cole, 1986.
392. ———, *Hipparchus, Plutarch, Schröder and Hough*, American Mathematical Monthly **104** (1997), 344–350.
393. ———, *Enumerative combinatorics*, vol. II, Cambridge University Press, 1998.
394. J. Michael Steele, *Probability theory and combinatorial optimization*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.
395. André Stef and Gérald Tenenbaum, *Inversion de Laplace effective*, Ann. Probab. **29** (2001), no. 1, 558–575.
396. P. R. Stein and M. S. Waterman, *On some new sequences generalizing the Catalan and Motzkin numbers*, Discrete Mathematics **26** (1979), no. 3, 261–272.
397. V. E. Stepanov, *On the distribution of the number of vertices in strata of a random tree*, Theory of Probability and Applications **14** (1969), no. 1, 65–78.
398. Jean-Marc Steyaert, *Structure et complexité des algorithmes*, Doctorat d'état, Université Paris VII, April 1984.
399. Jean-Marc Steyaert and Philippe Flajolet, *Patterns and pattern-matching in trees: an analysis*, Information and Control **58** (1983), no. 1–3, 19–58.
400. Gabor Szegő, *Orthogonal polynomials*, American Mathematical Society Colloquium Publications, vol. XXIII, A.M.S, Providence, 1989.
401. Wojciech Szpankowski, *Average-case analysis of algorithms on sequences*, John Wiley, New York, 2001.
402. Lajos Takács, *A Bernoulli excursion and its various applications*, Advances in Applied Probability **23** (1991), 557–585.
403. ———, *Conditional limit theorems for branching processes*, Journal of Applied Mathematics and Stochastic Analysis **4** (1991), no. 4, 263–292.
404. ———, *On a probability problem connected with railway traffic*, Journal of Applied Mathematics and Stochastic Analysis **4** (1991), no. 1, 1–27.
405. ———, *The asymptotic distribution of the total heights of random rooted trees*, Acta Scientifica Mathematica (Szeged) **57** (1993), 613–625.
406. Martin C. Tangora, *Level number sequences of trees and the lambda algebra*, European Journal of Combinatorics **12** (1991), 433–443.
407. H. N. V. Temperley, *On the enumeration of the Mayer cluster integrals*, Proc. Phys. Soc. Sect. B. **72** (1959), 1141–1144.
408. ———, *Graph theory and applications*, Ellis Horwood Ltd., Chichester, 1981.

409. Gérald Tenenbaum, *La méthode du col en théorie analytique des nombres*, Séminaire de Théorie des Nombres, Paris 1985–1986 (C. Goldstein, ed.), Birkhauser, 1988, pp. 411–441.
410. ______, *Introduction to analytic and probabilistic number theory*, Cambridge University Press, Cambridge, 1995, Translated from the second French edition (1995) by C. B. Thomas.
411. E. C. Titchmarsh, *The theory of functions*, second ed., Oxford University Press, 1939.
412. E. C. Titchmarsh and D. R. Heath-Brown, *The theory of the Riemann zeta-function*, second ed., Oxford Science Publications, 1986.
413. W. T. Tutte, *A census of planar maps*, Canadian Journal of Mathematics **15** (1963), 249–271.
414. ______, *On the enumeration of planar maps*, Bull. Amer. Math. Soc. **74** (1968), 64–74.
415. ______, *On the enumeration of four-colored maps*, SIAM Journal on Applied Mathematics **17** (1969), 454–460.
416. ______, *Planar enumeration*, Graph theory and combinatorics (Cambridge, 1983), Academic Press, London, 1984, pp. 315–319.
417. Bernard Van Cutsem, *Combinatorial structures and structures for classification*, Computational Statistics & Data Analysis **23** (1996), no. 1, 169–188.
418. Bernard Van Cutsem and Bernard Ycart, *Indexed dendrograms on random dissimilarities*, Journal of Classification **15** (1998), no. 1, 93–127.
419. Joris van der Hoeven, *Majorants for formal power series*, Preprint, 2003, 29 pages. Available from author's webpage.
420. B. L. Van der Waerden, *On the method of saddle points*, Applied Scientific Research **2** (1951), 33–45.
421. B. L. van der Waerden, *Algebra. Vol. I*, Springer-Verlag, New York, 1991, Based in part on lectures by E. Artin and E. Noether, Translated from the seventh German edition by Fred Blum and John R. Schulenberger. MR **MR1080172 (91h:**00009a)
422. J. van Leeuwen (ed.), *Handbook of theoretical computer science*, vol. A: Algorithms and Complexity, North Holland, 1990.
423. E. J. Janse van Rensburg, *The statistical mechanics of interacting walks, polygons, animals and vesicles*, Oxford University Press, Oxford, 2000.
424. Robert Vein and Paul Dale, *Determinants and their applications in mathematical physics*, Applied Mathematical Sciences, vol. 134, Springer-Verlag, 1998.
425. A. M. Vershik, *Statistical mechanics of combinatorial partitions, and their limit configurations*, Funktsional'nyĭ Analiz i ego Prilozheniya **30** (1996), no. 2, 19–39.
426. A. M. Vershik and S. V. Kerov, *Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables*, Soviet Mathematical Doklady **18** (1977), 527–531.
427. Jeffrey Scott Vitter and Philippe Flajolet, *Analysis of algorithms and data structures*, Handbook of Theoretical Computer Science (J. van Leeuwen, ed.), vol. A: Algorithms and Complexity, North Holland, 1990, pp. 431–524.
428. Joachim von zur Gathen and Jürgen Gerhard, *Modern computer algebra*, Cambridge University Press, New York, 1999.
429. J. Vuillemin, *A unifying look at data structures*, Communications of the ACM **23** (1980), no. 4, 229–239.
430. H. S. Wall, *Analytic theory of continued fractions*, Chelsea Publishing Company, 1948.
431. W. Wasow, *Asymptotic expansions for ordinary differential equations*, Dover, 1987, A reprint of the John Wiley edition, 1965.
432. Michael S. Waterman, *Introduction to computational biology*, Chapman & Hall, 1995.
433. E. T. Whittaker and G. N. Watson, *A course of modern analysis*, fourth ed., Cambridge University Press, 1927, Reprinted 1973.
434. David Vernon Widder, *The Laplace transform*, Princeton University Press, 1941.
435. Herbert S. Wilf, *Some examples of combinatorial averaging*, American Mathematical Monthly **92** (1985), 250–261.
436. ______, *Combinatorial algorithms: An update*, CBMS–NSF Regional Conference Series, no. 55, Society for Industrial and Applied Mathematics, Philadelphia, 1989.
437. ______, *Generatingfunctionology*, Academic Press, 1990.
438. David Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.
439. Roderick Wong, *Asymptotic approximations of integrals*, Academic Press, 1989.
440. Alan R. Woods, *Coloring rules for finite trees, and probabilities of monadic second order sentences*, Random Structures Algorithms **10** (1997), no. 4, 453–485.
441. E. Maitland Wright, *Asymptotic partition formulae: I plane partitions*, Quarterly Journal of Mathematics, Oxford Series **II** (1931), 177–189.

442. ______, *The coefficients of a certain power series*, Journal of the London Mathematical Society **7** (1932), 256–262.
443. ______, *The coefficients of a certain power series*, Journal of the London Mathematical Society **7** (1932), 304–309.
444. ______, *The number of connected sparsely edged graphs*, Journal of Graph Theory **1** (1977), 317–330.
445. ______, *The number of connected sparsely edged graphs. II. Smooth graphs*, Journal of Graph Theory **2** (1978), 299–305.
446. ______, *The number of connected sparsely edged graphs. III. Asymptotic results*, Journal of Graph Theory **4** (1980), 393–407.
447. Robert Alan Wright, Bruce Richmond, Andrew Odlyzko, and Brendan McKay, *Constant time generation of free trees*, SIAM Journal on Computing **15** (1985), no. 2, 540–548.
448. Max Wyman, *The asymptotic behavior of the Laurent coefficients*, Canadian Journal of Mathematics **11** (1959), 534–555.
449. Yifan Yang, *Partitions into primes*, Transactions of the American Mathematical Society **352** (2000), no. 6, 2581–2600.
450. Doron Zeilberger, *Symbol-crunching with the transfer-matrix method in order to count skinny physical creatures*, Integers **0** (2000), Paper A9, Published electronically at `http://www.integers-ejcnt.org/vol0.html`.
451. Paul Zimmermann, *Séries génératrices et analyse automatique d'algorithmes*, Ph. d. thesis, École Polytechnique, 1991.
452. V. M. Zolotarev, *One-dimensional stable distributions*, American Mathematical Society, Providence, RI, 1986, Translated from the Russian by H. H. McFaden, Translation edited by Ben Silver.

Index