ANALYTIC COMBINATORICS
—
SYMBOLIC COMBINATORICS

PHILIPPE FLAJOLET & ROBERT SEDGEWICK

Algorithms Project
INRIA Rocquencourt
78153 Le Chesnay
France

Department of Computer Science
Princeton University
Princeton, NJ 08540
USA

First Edition, May 25, 2002
♡ ♡ ♡ ♡ ♡ ♡

ABSTRACT

This booklet develops in nearly 200 pages the basics of combinatorial enumeration through an approach that revolves around generating functions. The major objects of interest here are words, trees, graphs, and permutations, which surface recurrently in all areas of discrete mathematics. The text presents the core of the theory with chapters on unlabelled enumeration and ordinary generating functions, labelled enumeration and exponential generating functions, and finally multivariate enumeration and generating functions. It is largely oriented towards applications of combinatorial enumeration to random discrete structures and discrete mathematics models, as they appear in various branches of science, like statistical physics, computational biology, probability theory, and, last not least, computer science and the analysis of algorithms.

Acknowledgements. This work was supported in part by the IST Programme of the EU under contract number IST-1999-14186 (ALCOM-FT). The authors are grateful to Xavier Gourdon who incited us to add a separate chapter on multivariate generating functions and to Brigitte Vallée for many critical suggestions regarding the presentation and global organization of this text. This booklet would be substantially different (and much less informative) without Neil Sloane's *Encyclopedia of Integer Sequences*, Steve Finch's *Mathematical Constants*, both 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 immensely useful.

"*Symbolic Combinatorics*" is a set of lecture notes that are a component of a wider book project titled *Analytic Combinatorics*, which will provide a unified treatment of analytic methods in combinatorics. This text is partly based on an earlier document titled "The Average Case Analysis of Algorithms: Counting and Generating Functions", INRIA Res. Rep. #1888 (1993), 116 pages, which it now subsumes.

FOREWORD

Analytic Combinatorics aims at predicting precisely the asymptotic properties of structured combinatorial configurations, through an approach that bases itself 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. In this context, singularities play a key rôle in extracting the functions' coefficients in asymptotic form and extremely precise estimates result for counting sequences. This chain is applicable to a large number of problems of discrete mathematics relative to words, trees, permutations, graphs, and so on. A suitable adaptation of the theory finally opens the way to the analysis of parameters of large random structures.

Analytic combinatorics can accordingly be organized based on three components:

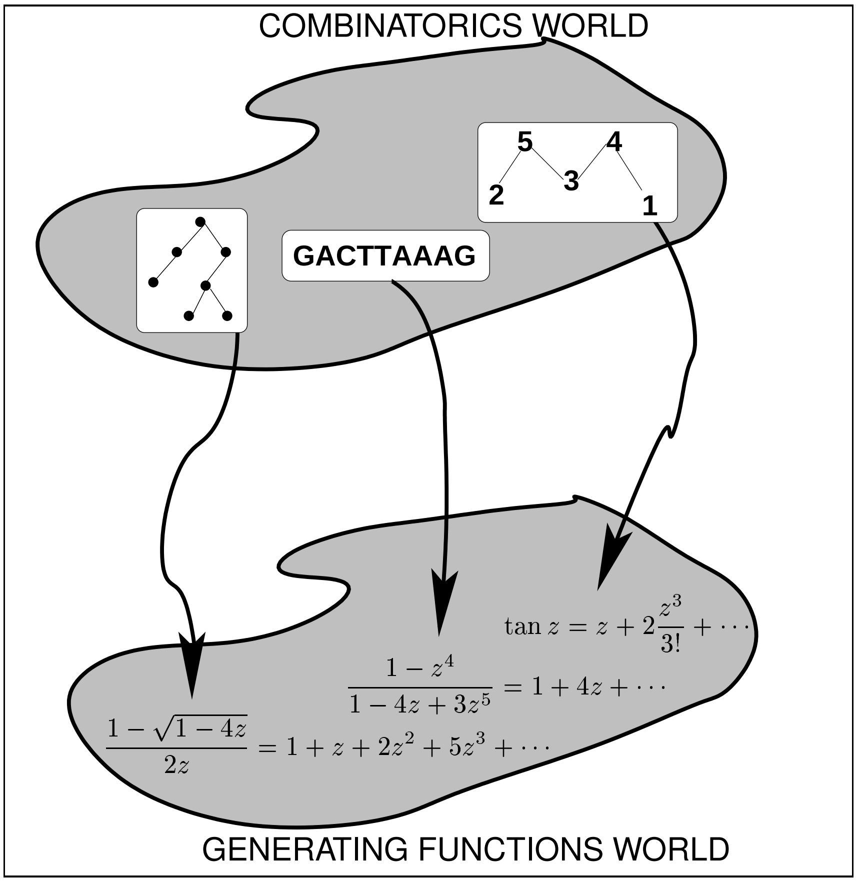
— *Symbolic Combinatorics* develops systematic "symbolic" relations between some of the major constructions of discrete mathematics and operations on generating functions which exactly encode counting sequences.

— *Singular combinatorics* elaborates a collection of methods by which one can extract asymptotic counting informations from generating functions, once these are viewed as analytic (holomorphic) functions over the complex domain. Singularities then appear to be a key determinant of asymptotic behaviour.

— *Random Combinatorics* 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, this corresponds to 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 booklets, of which this is the first installment, expose this view by means of a very large number of examples concerning classical combinatorial structures (like words, trees, permutations, and graphs). What is aimed at eventually is an effective way of quantifying "metric" properties of large random structures. Accordingly, the theory 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.

This booklet specifically exposes *Symbolic Combinatorics*, which is a unified algebraic theory dedicated to the setting up of 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, hence the name of "symbolic combinatorics" that we have adopted to characterize it.) 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 booklet: 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.



Contents

Leonhard Euler (born, 15 April 1707 in Basel, Switzerland; died, 18 Sept 1783 in St Petersburg, Russia) was the first to relate classical analysis and combinatorics in a publication of 1753. Euler showed how to enumerate the triangulations of an n-gon: first, he modelled the combinatorial counting problem by a recurrence, then introduced the corresponding generating function; finally he solved the resulting equation and expanded the generating function using classical analysis, thereby providing a closed-form solution to the original counting problem and discovering the "Catalan numbers".

(Pictures are from *The MacTutor History of Mathematics* archive hosted by the University of St Andrews.)

CHAPTER 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 [**122**]

This chapter and the next are devoted to enumeration, where the question is to determine the number of combinatorial configurations described by finite rules, and do so for all possible sizes. For instance, how many different permutations are there of size 17? of size n, for general n? what if some constraints are imposed, e.g., no four elements of increasing order 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 more 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 is made concrete by means of a "dictionary" based on a core of important constructions, which includes the operations of union, cartesian product, sequence, set, multiset, and cycle. (Supplementary operations like pointing and substitution can be also be similarly treated.)

In this way, a language describing elementary combinatorial classes is set up. The problem of enumerating a class of combinatorial structures then simply reduces to finding a proper *specification*, a sort of formal "grammar", for the class in terms of the basic constructions. The translation into generating functions then becomes a purely mechanical "symbolic" process.

We show here how to describe in such a context integer partitions and compositions, as well as several elementary string and tree enumeration problems. A parallel approach, developed in Chapter II, applies to labelled objects and exponential generating functions, and 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. It also has the great merit of connecting nicely with complex asymptotic methods that exploit analyticity properties and singularities, to the effect that very precise asymptotic estimates are usually available whenever the symbolic method applies—a systematic treatment forms the basis of the next booklet in the series, *Analytic Combinatorics, Singular Combinatorics* (Chapters IV–VI).

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, geometric configurations, permutations, allocations, functions from a finite set into itself, 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: the size of an element is a nonnegative integer; 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 $A_n = \text{card}(\mathcal{A}_n)$ (alternatively, also $a_n = \text{card}(\mathcal{A}_n)$). 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 $A_n = \text{card}(\mathcal{A}_n)$ is the number of objects in class $\mathcal{A}$ that have size n.*

Consider for instance the set $\mathcal{W}$ of binary words, which are words over a binary alphabet,

$$\mathcal{W} := \{\ldots \mathsf{00}, \mathsf{01}, \mathsf{10}, \mathsf{11}, \mathsf{000}, \mathsf{001}, \mathsf{010}, \ldots, \mathsf{1001101}, \ldots\},$$

if the binary alphabet is $\mathcal{A} = \{\mathsf{0},\mathsf{1}\}$. The set $\mathcal{P}$ of permutations is

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

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 $\mathcal{I}_n$; The set $\mathcal{T}$ of triangulations is comprised of triangulations of convex polygonal domains which are decompositions into non-overlapping triangles. (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. 18.) The sets $\mathcal{W}$, $\mathcal{P}$, and $\mathcal{T}$ constitute combinatorial classes, with the convention that the size of a word is its length, the size of a permutation is the number of its elements, the size of a triangulation is the number of triangles it comprises. The corresponding counting sequences are then given by

$$W_n = 2^n, \qquad P_n = n\,!, \qquad T_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!}, \tag{1}$$

where the initial values are

(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

Indeed elementary counting principles, namely, for finite sets $\mathcal{B}$ and $\mathcal{C}$

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

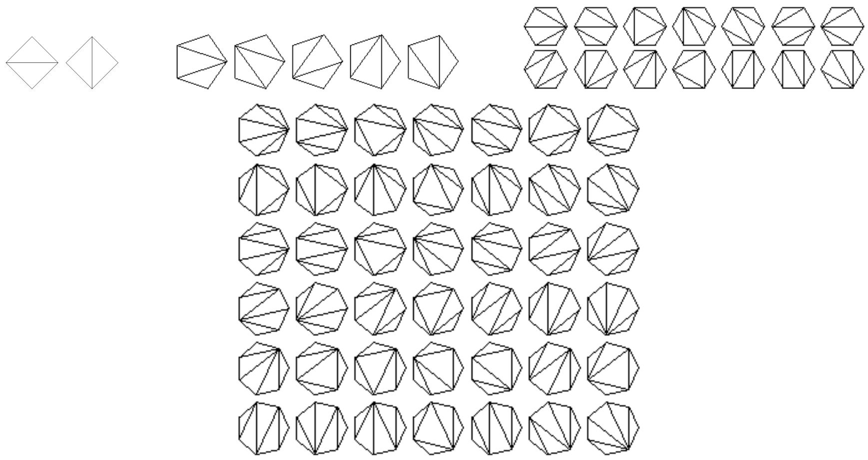


FIGURE 1. The class $\mathcal{T}$ of all triangulations of regular polygons (with size defined as the number of triangles) is a combinatorial class. The counting sequence starts as $T_0 = 1,\ T_1 = 1,\ T_2 = 2,\ T_3 = 5,\ T_4 = 14,\ T_5 = 42,\ T_6 = 132,\ \ldots$. Euler determined the OGF $T(z) = \sum_n T_n z^n$ as

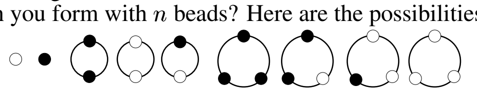
$$T(z) = \frac{1 - \sqrt{1-4z}}{2z},$$

from which there results that $T_n = \frac{1}{n+1}\binom{2n}{n}$. These numbers are known as the *Catalan numbers* (p. 17).

lead directly to expressions for words (W_n) and permutations (P_n). The sequence $W_n = 2^n$ has a well-known legend associated with the invention of the chess game: 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... As to the number of permutations, it has been known for more than 1500 years and Knuth [**86**, p. 23] refers to the Hebrew *Book of Creation* (c. A.D.. 400), and to the *Anuyogadvārasutra* (India, c. A.D. 500) for the explicit formula $n! = 1 \cdot 2 \cdots n$. Following Euler (1707–1783), 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. 17) and are central in combinatorial analysis (Section I. 5.3).

▷ **1.** *Permutations and factorials.* For a permutation in $\mathcal{P}_n$ written as a sequence of distinct numbers, there are n places where one can accommodate n, $n-1$ remaining places for $n-1$, and so on. Therefore, by (3), the number of permutations is $n \cdot (n-1) \cdots = n!$. ◁

▷ **2.** *Necklaces.* You are given tons of beads of two colours, ∘ and •. How many different types of necklace designs can you form with n beads? Here are the possibilities for $n = 1, 2, 3$:



This can be reformulated as the problem of finding the counting sequence of the class of necklaces defined formally as all the possible circular arrangements of two letters. The counting sequence starts as $2, 3, 4, 6, 8, 14, 20, 36, 60, 108, 188, 352$. The solution appears later in this chapter, p. 40. ◁

Two combinatorial classes $\mathcal{A}$ and $\mathcal{B}$ are said to be *isomorphic*, which is written $\mathcal{A} \cong \mathcal{B}$, iff their counting sequences are identical. This is equivalent to saying that there exists a bijection from $\mathcal{A}$ to $\mathcal{B}$ that preserves size, and one also says that $\mathcal{A}$ and $\mathcal{B}$ are *bijectively* *equivalent*. Since we are only interested in counting problems, it proves often convenient to identify isomorphic classes and plainly consider them as identical. We then confine the notation $\mathcal{A} \cong \mathcal{B}$ (instead of $\mathcal{A} = \mathcal{B}$) to the few cases where combinatorial isomorphism rather than plain identity needs to be emphasized.

DEFINITION I.3. *The* ordinary generating function *(OGF) of a sequence* $\{A_n\}$ *is the formal power series*

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

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

$$A(z) = \sum_{n\geq 0} A_n z^n = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|}. \tag{5}$$

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

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. Also, 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 notation reads as "coefficient of z^n in $f(z)$".)

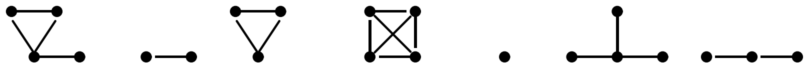
The OGF's corresponding to Eq. (1) are then

$$W(z) = \frac{1}{1-2z}, \qquad P(z) = \sum_{n=0}^{\infty} n!\, z^n, \qquad T(z) = \frac{1-2z-\sqrt{1-4z}}{2z}.$$

The OGF's $W(z)$ and $T(z)$ can be interpreted as standard analytic objects, upon assigning to the formal variable z values in the complex domain $\mathcal{C}$. In effect, the series $W(z)$ and $T(z)$ converge in a neighbourhood of 0 and represent complex functions analytic at the origin, while 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; see APPENDIX: *Formal power series*, p. 169. (Permutation enumeration is most conveniently approached by exponential generating functions developed in Chapter II.)

The second "combinatorial" form 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. This form shows that generating functions are nothing but a reduced representation of the combinatorial class[1], where "internal" structures are destroyed and elements contributing to size ("atoms") are replaced by the variable z. Here is an illustration: start with a (finite) family of graphs $\mathcal{G}$, with size taken as the number of vertices [line 1]. Each vertex in each graph is replaced by the variable z and the graph structure is "forgotten" [line 2]; then the monomials corresponding to each graph are formed [line 3] and the generating function is obtained [line 4] by gathering all the monomials:

[1]This observation of which great use was made by Schützenberger as early as the 1950's and 1960's "explains" why many similarities are to be found between combinatorial structures and generating functions.

$\mathcal{G} =$ 

$zzzz$ zz zzz $zzzz$ z $zzzz$ zzz

$+ z^4$ $+ z^2$ $+ z^3$ $+ z^4$ $+ z$ $+ z^4$ $+ z^3$

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

For instance, there are three graphs of size 4, in agreement with the fact that $[z^4]G(z) = 3$. If size had been instead defined by number of edges, another generating function would have resulted, namely, with y marking 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, $G(z, y) = z + z^2y + z^3y^2 + z^3y^3 + z^4y^3 + z^4y^4 + z^4y^6$; such multivariate generating functions are developed systematically in Chapter III.

A path often taken in the older or more traditional literature consists in decomposing the structures to be enumerated into smaller structures either of the same type or of simpler types, and then in extracting 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, then introduced as a mere technical artefact.

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. In this chapter, the generating functions considered are ordinary generating functions.

DEFINITION I.4. *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 $\mathcal{B}$ and $\mathcal{C}$:

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

In that case, there exists a well defined operator Ψ relating the associated ordinary generating functions

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

As an introductory example, take the construction of cartesian product that forms ordered pairs (equivalently, "records" in classical programming languages):

$$(a) \qquad \mathcal{A} = \mathcal{B} \times \mathcal{C} \quad \text{iff} \quad \mathcal{A} = \{\, \alpha = (\beta, \gamma) \mid \beta \in \mathcal{B},\ \gamma \in \mathcal{C}\,\},$$

the size of a pair $\alpha = (\beta, \gamma)$ being defined by $|\alpha|_{\mathcal{A}} = |\beta|_{\mathcal{B}} + |\gamma|_{\mathcal{C}}$. Then, considering all possibilities, the counting sequences corresponding to $\mathcal{A}, \mathcal{B}, \mathcal{C}$ are related by the convolution relation

$$(b) \qquad A_n = \sum_{k=0}^{n} B_k C_{n-k}.$$

We recognize here the formula for a product of two power series. Therefore, with $A(z) = \sum_{n\geq 0} A_n z^n$ etc, one has

$$(c) \qquad A(z) = B(z) \cdot C(z).$$

Thus in our terminology, the cartesian product is admissible: *A cartesian product translates as a product of OGF's*.

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

$$(d) \qquad \mathcal{A} = \mathcal{B} \cup \mathcal{C}, \qquad \text{with} \quad \mathcal{B} \cap \mathcal{C} = \emptyset,$$

with size defined in a consistent manner. One has

$$(e) \qquad A_n = B_n + C_n,$$

which, at generating function level, means

$$(f) \qquad A(z) = B(z) + C(z).$$

Thus, *a disjoint union translates as a sum of generating functions*.

The correspondences Eq. (a)–(c) and (d)–(f) summarized by the table

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

are clearly very general ones. (Compare with Eq. (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 disjoint union or a cartesian product is recognized, it becomes possible to dispense altogether with the intermediate stages of writing explicitly coefficient relations like (f) or recurrences like (b). 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 (or sums) and on Cartesian products that we have just discussed. We shall introduce the constructions of sequence, cycle, multiset, and powerset. A class is (fully) constructible 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 Figure 2.

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; the atom may 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 may 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

1. The main constructions of union, 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	$\mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$	$A(z) = \dfrac{1}{1 - B(z)}$
Set	$\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$	$A(z) = \exp\left(B(z) - \frac{1}{2}B(z^2) + \cdots\right)$
Multiset	$\mathcal{A} = \mathfrak{M}\{\mathcal{B}\}$	$A(z) = \exp\left(B(z) + \frac{1}{2}B(z^2) + \cdots\right)$
Cycle	$\mathcal{A} = \mathfrak{C}\{\mathcal{B}\}$	$A(z) = \log\dfrac{1}{1 - B(z)} + \dfrac{1}{2}\log\dfrac{1}{1 - B(z^2)} + \cdots$

2. The translation for sets, multisets, and cycles constrained by the number of components (Theorem I.2).

$$
\begin{aligned}
\mathfrak{S}_k\{\mathcal{B}\} &: \quad B(z)^k \\
\mathfrak{P}_2\{\mathcal{B}\} &: \quad \tfrac{B(z)^2}{2} - \tfrac{B(z^2)}{2} \\
\mathfrak{M}_2\{\mathcal{B}\} &: \quad \tfrac{B(z)^2}{2} + \tfrac{B(z^2)}{2} \\
\mathfrak{C}_2\{\mathcal{B}\} &: \quad \tfrac{B(z)^2}{2} + \tfrac{B(z^2)}{2} \\
\mathfrak{P}_3\{\mathcal{B}\} &: \quad \tfrac{B(z)^3}{6} - \tfrac{B(z)\,B(z^2)}{2} + \tfrac{B(z^3)}{3} \\
\mathfrak{M}_3\{\mathcal{B}\} &: \quad \tfrac{B(z)^3}{6} + \tfrac{B(z)\,B(z^2)}{2} + \tfrac{B(z^3)}{3} \\
\mathfrak{C}_3\{\mathcal{B}\} &: \quad \tfrac{B(z)^3}{3} + \tfrac{2B(z^3)}{3} \\
\mathfrak{P}_4\{\mathcal{B}\} &: \quad \tfrac{B(z)^4}{24} - \tfrac{B(z)^2 B(z^2)}{4} + \tfrac{B(z)B(z^3)}{3} + \tfrac{B(z^2)^2}{8} - \tfrac{B(z^4)}{4} \\
\mathfrak{M}_4\{\mathcal{B}\} &: \quad \tfrac{B(z)^4}{24} + \tfrac{B(z)^2 B(z^2)}{4} + \tfrac{B(z)B(z^3)}{3} + \tfrac{B(z^2)^2}{8} + \tfrac{B(z^4)}{4} \\
\mathfrak{C}_4\{\mathcal{B}\} &: \quad \tfrac{B(z)^4}{4} + \tfrac{B(z^2)^2}{2} + \tfrac{B(z^4)}{2}.
\end{aligned}
$$

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 2. 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 2 of Chapter II, p. 67.)

trees. Similarly, we may introduce $\mathcal{E}_\Box, \mathcal{E}_1, \mathcal{E}_2$ to denote a class comprising the neutral objects $\Box, \epsilon_1, \epsilon_2$ 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.

I. 2.1. Basic constructions. Here are described a few powerful constructions that build upon disjoint unions and cartesian products, and form sequences, sets, and cycles.

First consider the *disjoint union* also called the *combinatorial sum* of classes, the intent being to capture the union of disjoint sets, but without the burden of carrying extraneous disjointness conditions. 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 $\mathcal{B}^\Box$ and $\mathcal{C}^\Diamond$, 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 $\Box$-colour and the elements of $\mathcal{C}$ with the $\Diamond$-colour. This is made precise by introducing two distinct "markers" $\Box$ 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,

$$\mathcal{B}+\mathcal{C} := (\{\Box\} \times \mathcal{B}) \cup (\{\Diamond\} \times \mathcal{C}) .$$

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 of origin. One reason behind the definition[2] adopted here of disjoint union is that the combinatorial sum of two classes is always well-defined. Furthermore, we have ($\cong$ represents combinatorial isomorphism)

$$\mathcal{B}+\mathcal{C} \cong \mathcal{B}\cup\mathcal{C} \quad \text{whenever} \quad \mathcal{B}\cap\mathcal{C} = \emptyset.$$

Disjoint union in the above sense is thus equivalent to a standard union whenever it is applied to disjoint sets. Then, 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 admissible since

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

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.

With the convention of identifying isomorphic classes, sum and product acquire pleasant algebraic properties: 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 and the notion of combinatorial isomorphism.)

Next, we turn to the sequence construction. If $\mathcal{C}$ is a class then the *sequence* class $\mathfrak{S}\{\mathcal{C}\}$ is defined as the infinite sum

$$\mathfrak{S}\{\mathcal{C}\} = \{\epsilon\} + \mathcal{C} + (\mathcal{C}\times\mathcal{C}) + (\mathcal{C}\times\mathcal{C}\times\mathcal{C}) + \cdots$$

[2]It would have been inconvenient to have a construction that translates into generating functions under some external condition—disjointness—of a logical nature that would need to be established separately in each particular case.

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: *Regular languages*, p. 171.) It is then readily checked that the construction $\mathcal{A} = \mathfrak{S}\{\mathcal{C}\}$ 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:

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

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

$\triangleright$ **4.** *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 $\mathcal{C} = \mathfrak{S}\{\mathcal{A}\}$ 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 $\mathcal{C} = \mathfrak{S}\{\mathcal{Z} + (\mathcal{Z} \times \mathcal{Z})\}$ are (isomorphic to) the *coverings* of the interval $[0, n]$ by matches of length either 1 or 2. The generating function

$$C(z) = 1 + z + 2\,z^2 + 3\,z^3 + 5\,z^4 + 8\,z^5 + 13\,z^6 + 21\,z^7 + 34\,z^8 + 55\,z^9 + 89\,z^{10} + \cdots,$$

is, as we shall see shortly (p. 24), the OGF of Fibonacci numbers. $\triangleleft$

Cycles are merely sequences defined up to a circular shift of their components, the notation being $\mathfrak{C}\{\mathcal{B}\}$. Thus, $\mathfrak{C}\{\mathcal{B}\} := \mathfrak{S}\{\mathcal{B}\}/\mathbf{S}$ with $\mathbf{S}$ the equivalence relation between sequences defined by $(\alpha_1, \dots, \alpha_r)\,\mathbf{S}\,(\beta_1, \dots, \beta_r)$ iff there exists some circular shift σ of $[1\,..\,n]$ such that for all j, $\beta_j = \alpha_{\sigma(j)}$; 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 $\mathbf{S}$.

$n=3$	$n=4$
aaa	$aaaa$
$aab\ aba\ baa$	$aaab\ aaba\ abaa\ baaa$
$abb\ bba\ bab$	$aabb\ bbba\ bbaa\ baab$
bbb	$abab\ baba$
	$abbb\ bbba\ bbab\ babb$
	$bbbb$

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.

Multisets are like finite sets (that is the order between element does not count) but arbitrary repetitions of elements are allowed. The notation is $\mathcal{A} = \mathfrak{M}\{\mathcal{B}\}$ when $\mathcal{A}$ is obtained by forming all finite multisets of elements from $\mathcal{B}$. The precise way of defining $\mathfrak{M}\{\mathcal{B}\}$ is as a quotient: $\mathfrak{M}\{\mathcal{B}\} := \mathfrak{S}\{\mathcal{B}\}/\mathbf{R}$ with $\mathbf{R}$ the equivalence relation between sequences defined by $(\alpha_1, \dots, \alpha_r)\,\mathbf{R}\,(\beta_1, \dots, \beta_r)$ iff, there exists some arbitrary permutation σ of $[1\,..\,n]$ such that for all j, $\beta_j = \alpha_{\sigma(j)}$. The *powerset* class (or set class) $\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$ is defined as the class consisting of all *finite* subsets of class $\mathcal{B}$, or equivalently, as the class $\mathfrak{P}\{\mathcal{B}\} \subset \mathfrak{M}\{\mathcal{B}\}$ 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.

In what follows, we also want to impose restrictions on the number of components allowed in sequences, sets, multisets, and cycles. Let $\mathfrak{K}$ be any of $\mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$ 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

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

refer to sequences whose number of components are exactly k, larger than k, or in the interval $1\,..\,k$ respectively. For example, one has

$$\mathfrak{S}_k\{\mathcal{B}\} := \overbrace{\mathcal{B}\times\cdots\mathcal{B}}^{k \text{ times}} \equiv \mathcal{B}^k, \qquad \mathfrak{M}_k\{\mathcal{B}\} := \mathfrak{S}_k\{\mathcal{B}\}/\mathbf{R}, \quad \mathfrak{S}_{\geq k}\{\mathcal{B}\} \cong \mathcal{B}^k \times \mathfrak{S}\{\mathcal{B}\}.$$

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

I. 2.2. The admissibility theorem for ordinary generating functions. This section shows 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: *Arithmetical functions*, p. 165).

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

$$\begin{array}{llll}
\textit{Union:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(z) = B(z) + C(z) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \times \mathcal{C} & \Longrightarrow & A(z) = B(z) \cdot C(z) \\
\textit{Sequence:} & \mathcal{A} = \mathfrak{S}\{\mathcal{B}\} & \Longrightarrow & A(z) = \dfrac{1}{1 - B(z)} \\
\textit{Cycle:} & \mathcal{A} = \mathfrak{C}\{\mathcal{B}\} & \Longrightarrow & A(z) = \displaystyle\sum_{k=1}^{\infty} \frac{\varphi(k)}{k} \log \frac{1}{1 - B(z^k)}. \\
\textit{Multiset:} & \mathcal{A} = \mathfrak{M}\{\mathcal{B}\} & \Longrightarrow & A(z) = \begin{cases} \displaystyle\prod_{n\geq 1} (1 - z^n)^{-B_n} \\ \exp\left(\displaystyle\sum_{k=1}^{\infty} \frac{1}{k} B(z^k)\right) \end{cases} \\
\textit{Powerset:} & \mathcal{A} = \mathfrak{P}\{\mathcal{B}\} & \Longrightarrow & A(z) = \begin{cases} \displaystyle\prod_{n\geq 1} (1 + z^n)^{B_n} \\ \exp\left(\displaystyle\sum_{k=1}^{\infty} \frac{(-1)^{k-1}}{k} B(z^k)\right) \end{cases}
\end{array}$$

For the sequence, cycle, and set constructions, it is assumed that $B_0 = 0$.

The class $\mathcal{E} = \{\epsilon\}$ consisting of the neutral structure 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. *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 has discussed earlier. Alternatively, the translation rule follows directly from the combinatorial form of generating functions as

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

Cartesian Product: The admissibility result for $\mathcal{A} = \mathcal{B}\times\mathcal{C}$ has been discussed already. It follows from $A_n = \sum_{k=0}^{n} B_k C_{n-k}$ that

$$A(z) = B(z) \times C(z).$$

Note also the alternative direct derivation based on the combinatorial form of GF's,

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

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

Sequence: Admissibility for $\mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$ (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.

Set (or powerset) construction: Let $\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$ 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,

$$\mathfrak{P}\{\mathcal{B}\} \cong \prod_{\beta\in\mathcal{B}} (\{\epsilon\} + \{\beta\})$$

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. The technique is similar to what is required to establish identities like

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

where all combinations of variables appear. Then, directly from the combinatorial form (4) of OGF's 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

$$\begin{aligned} A(z) &= \exp\Big(\sum_{n=1}^{\infty} B_n \log(1+z^n)\Big) \\ &= \exp\Big(\sum_{n=1}^{\infty} B_n \cdot \sum_{k=1}^{\infty} \frac{z^{nk}}{k}\Big) \\ &= \exp\Big(\frac{B(z)}{1} - \frac{B(z^2)}{2} + \frac{B(z^3)}{3} - \cdots\Big), \end{aligned} \tag{7}$$

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 summations.

The proof 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 $\mathcal{B}^{(\le m)} = \sum_{k=1}^{m} \mathcal{A}_j$ and $\mathcal{A}^{(\le m)} = \mathfrak{P}\{\mathcal{B}^{(\le m)}\}$. Then, with $O(z^{m+1})$ denoting any series that has no term of degree $\le m$, one has

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

On the other hand, $A^{(\le m)}(z)$ and $B^{(\le m)}(z)$ are connected by the fundamental exponential relation (7) , since $\mathcal{A}^{(\le m)}$ is finite. Letting m tend to infinity, there follows in the limit

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

(See APPENDIX: *Formal power series*, p. 169 for definitions of formal convergence.) The necessary condition for validity is that $[z^0]B(z) = 0$, a restriction that also applies to multisets and cycles.

Multiset: First for finite $\mathcal{B}$ (with $\mathcal{B}_0 = \emptyset$), the multiset class $\mathcal{A} = \mathfrak{M}\{\mathcal{B}\}$ is definable by

$$\mathfrak{M}\{\mathcal{B}\} \cong \prod_{\beta\in\mathcal{B}} \mathfrak{S}\{\beta\}.$$

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,

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

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

Cycle: The translation of the cycle relation $\mathcal{A} = \mathfrak{C}\{\mathcal{B}\}$ is

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

where $\varphi(k)$ is the Euler totient function: $\varphi(k)$ equals the number of integers in $[1, k]$ that are relatively prime to k, with $\varphi(1) = 1$. The first terms, with $L_k = \log(1 - B(z^k))^{-1}$ are

$$A(z) = \frac{1}{1}L_1 + \frac{1}{2}L_2 + \frac{2}{3}L_3 + \frac{2}{4}L_4 + \frac{4}{5}L_5 + \frac{2}{6}L_6 + \frac{6}{7}L_7 + \cdots.$$

This translation was first established by Read within the framework of Pólya's theory of counting [**115**]. An elementary combinatorial derivation based on [**58**] is given in APPENDIX: *Cycle construction*, p. 168. □

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 [**113, 115**]. This theory is exposed in many textbooks, see for instance [**28, 76**]. 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 [**13**].

Restricted constructions. An immediate formula for OGF's 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 $\mathcal{A} = \mathfrak{P}_2\{\mathcal{B}\}$. 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,

$$\mathfrak{P}_2\{\mathcal{B}\} + \mathfrak{P}_2\{\mathcal{B}\} + \Delta(B \times B) \cong B \times B,$$

meaning that

$$2A(z) + B(z^2) = B(z)^2.$$

The resulting translation into OGFs is thus

$$\mathcal{A} = \mathfrak{P}_2\{\mathcal{B}\} \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 - \frac{1}{2}B(z^2).$$

Similarly, for multisets, we find

$$\mathcal{A} = \mathfrak{M}_2\{\mathcal{B}\} \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 + \frac{1}{2}B(z^2),$$

while for cycles one has $\mathfrak{C}_2 \cong \mathfrak{M}_2$, and

$$\mathcal{A} = \mathfrak{C}_2\{\mathcal{B}\} \qquad \Longrightarrow \qquad A(z) = \frac{1}{2}B(z)^2 + \frac{1}{2}B(z^2).$$

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.2 (Component-restricted constructions). *The OGF of sequences with k components $\mathcal{A} = \mathfrak{S}_k\{\mathcal{B}\}$ satisfies*

$$A(z) = B(z)^k.$$

The OGF of sets, $\mathcal{A} = \mathfrak{P}_k\{\mathcal{B}\}$, is a polynomial in the quantities $B(z), \ldots, B(z^k)$,

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

The OGF of multisets, $\mathcal{A} = \mathfrak{M}_k\{\mathcal{B}\}$, 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, $\mathcal{A} = \mathfrak{C}_k\{\mathcal{B}\}$, 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 2.

PROOF. The result for sequences is obvious since $\mathfrak{S}_k\{\mathcal{B}\}$ 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 $\mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$, 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

$$\begin{array}{lll} A_{n,k} & := & \text{card}\left\{\alpha \in \mathcal{A} \;\middle|\; |\alpha| = n,\, \chi(\alpha) = k\right\} \\ A(z,u) & := & \displaystyle\sum_{n,k} A_{n,k} u^k z^n = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|} u^{\chi(\alpha)}. \end{array}$$

For instance, a direct calculation shows that, for sequences, there holds

$$\begin{array}{lll} A(z,u) & = & \displaystyle\sum_{k \geq 0} u^k B(z)^k \\ & = & \displaystyle\frac{1}{1 - uB(z)}. \end{array}$$

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 $[u^k]A(z,u)$. □

▷ **5.** *Vallée's identity.* Let $\mathcal{M} = \mathfrak{M}\{\mathcal{C}\}$, $\mathcal{P} = \mathfrak{P}\{\mathcal{C}\}$. 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 on page 28. ◁

▷ **6.** *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 for instance in the analysis of polynomial factorization algorithms [**49**]. ◁

▷ **7.** *Sequences without repeated components.* These have generating function formally given by

$$\int_0^\infty \exp\left(\sum_{k \geq 1} (-1)^{j-1} \frac{u^j}{j} A(z^j) \right) e^{-u}\, du.$$

(This form is based on the Eulerian integral: $k! = \int_0^\infty e^u u^k\, dn$.) ◁

I. 2.3. Constructibility and combinatorial specifications. In the framework just introduced, the class of all binary words is described by

$$\mathcal{W} = \mathfrak{S}\{\mathcal{A}\} \qquad \text{where} \qquad \mathcal{A} = \{a, b\},$$

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 (Ex. 2, p. 3) and the natural integers (Ex. 3, p. 9) respectively defined by

$$\mathcal{N} = \mathfrak{C}\{\mathcal{Z} + \mathcal{Z}\} \qquad \text{and} \qquad \mathcal{I} = \mathfrak{S}_{\geq 1}\{\mathcal{Z}\}.$$

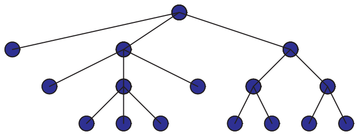
From there, one can construct ever more complicated objects. For instance,

$$\mathcal{P} = \mathfrak{M}\{\mathcal{I}\} \equiv \mathfrak{M}\{\mathfrak{S}_{\geq 1}\{\mathcal{Z}\}\}$$

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 $+, \times, \mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$. An iterative specification can be equivalently listed by naming some of the subterms (for instance partitions in terms of natural integers themselves defined as sets of atoms).

We next turn our attention to trees (cf. also APPENDIX: *Tree concepts*, p. 174 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—the "root". Computer scientists commonly make use of trees called *plane* that are rooted but also embedded in the plane. In other words, 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 [**130**]. (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 this will result in a different tree—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

$$\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\}, \tag{8}$$

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

Although recursive definitions are familiar to computer scientists, the specification (8) 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

$$\mathcal{G}^{[j+1]} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}^{[j]}\}.$$

For instance, $\mathcal{G}^{[1]} = \mathcal{Z} \times \mathfrak{S}\{\emptyset\} = \{(\zeta, \epsilon)\} \cong \{\zeta\}$ describes (the linear representation of) the tree of size 1, and

$$\begin{array}{rcl}
\mathcal{G}^{[2]} & = & \Big\{\zeta,\ \zeta\boxed{\zeta},\ \zeta\boxed{\zeta,\ \zeta},\ \zeta\boxed{\zeta,\ \zeta,\ \zeta},\ \ldots,\Big\} \\
\mathcal{G}^{[3]} & = & \Big\{\zeta,\ \zeta\boxed{\zeta},\ \zeta\boxed{\zeta,\ \zeta},\ \zeta\boxed{\zeta,\ \zeta,\ \zeta},\ \ldots \\
& & \quad \zeta\boxed{\zeta,\boxed{\zeta}},\ \zeta\boxed{\zeta,\boxed{\zeta,\ \zeta}},\ \zeta\boxed{\zeta,\boxed{\zeta,\ \zeta}},\ \zeta\boxed{\zeta,\boxed{\zeta,\ \zeta},\ \boxed{\zeta,\ \zeta}},\ldots\Big\}.
\end{array}$$

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 $\mathcal{G}^{[j]}$: $\mathcal{G} := \bigcup_j \mathcal{G}^{[j]}$. (There, '$\cup$' represents the usual set-theoretic union.)

$\triangleright$ **8.** *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 OGF's satisfy $A^{[\infty]}(z) = \lim_{j\to\infty} A^{[j]}(z)$ in the formal topology (APPENDIX: *Formal power series*, p. 169). $\triangleleft$

In all generality, a *specification* for an r–tuple $\vec{\mathcal{A}} = (\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})$ of classes is a collection of r equations,

$$\left\{\begin{array}{rcl}
\mathcal{A}^{(1)} & = & \Xi_1(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)}) \\
\mathcal{A}^{(2)} & = & \Xi_2(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)}) \\
& \cdots & \\
\mathcal{A}^{(r)} & = & \Xi_r(\mathcal{A}^{(1)}, \ldots, \mathcal{A}^{(r)})
\end{array}\right. \tag{9}$$

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. The system (9) corresponds to an iterative specification if it is stictly 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.5. *A class of combinatorial structures is said to be* constructible *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 [**56**] for the description of such a system.

COROLLARY I.1. *The generating function of a constructible class is a component of a system of generating function equations whose terms are built from*

$$1,\ z,\ +\ ,\ \times\ ,\ \Phi_{\mathfrak{S}},\ \Phi_{\mathfrak{C}},\ \Phi_{\mathfrak{M}},\ \Phi_{\mathfrak{P}},$$

where
$$\begin{cases} \Phi_{\mathfrak{S}}[f] = \dfrac{1}{1-f}, & \Phi_{\mathfrak{C}}[f] = \displaystyle\sum_{k=1}^{\infty} \frac{\varphi(k)}{k}\log\frac{1}{1-f(z^k)}, \\ \Phi_{\mathfrak{M}}[f] = \exp\left(\displaystyle\sum_{k=1}^{\infty}\frac{f(z^k)}{k}\right), & \Phi_{\mathfrak{P}}[f] = \exp\left(\displaystyle\sum_{k=1}^{\infty}(-1)^{k-1}\frac{f(z^k)}{k}\right). \end{cases}$$

Thus, iterative classes have explicit generating functions involving compositions of the basic operators only, while recursive structures have OGF's that are only 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.

For instance, the OGF of binary words corresponding to $\mathcal{W} = \mathfrak{S}(\mathcal{Z}+\mathcal{Z})$ is

$$W(z) = \frac{1}{1-2z},$$

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

For the class $\mathcal{G}$ of general trees, constructibility leads to an equation defining $G(z)$ implicitly,

$$G(z) = \frac{z}{1-G(z)}.$$

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 $\overline{G}(z)$ is to be discarded since it involves a term z^{-1} as well as negative coefficients; the expansion results from Newton's binomial theorem applied to $(1+x)^{1/2}$ at $x = -4z$.)

The numbers

$$\text{(10)} \qquad \mathrm{C}_n = \frac{1}{n+1}\binom{2n}{n} = \frac{(2n)!}{(n+1)!\,n!} \qquad \text{with OGF} \quad \mathrm{C}(z) = \frac{1-\sqrt{1-4z}}{2z}$$

are known as the Catalan numbers (*EIS* **A000108**)[3] in the honour of Eugène Catalan (1814-1894), a French and Belgian mathematician who developed many of their properties. In

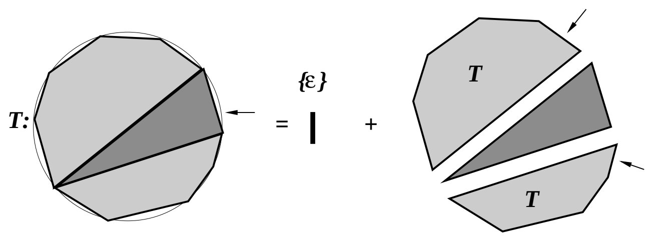
[3] Throughout this book, a reference like *EIS* **A000108** points to Sloane's *Encyclopedia of Integer Sequences* that is available in electronic form [**132**] or as a book by Sloane and Ploufe [**133**].

summary, *general trees are enumerated by Catalan numbers:*

$$G_n = \mathrm{C}_{n-1} \equiv \frac{1}{n}\binom{2n-2}{n-1}, \qquad \textit{where } \mathrm{C}_n \textit{ is a Catalan number.}$$

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

We can now conclude with the enumeration of triangulations, one of our three leading examples at the beginning of this chapter. Fix n points regularly spaced on a circle and conventionally numbered from 0 to $n-1$ (for instance the nth roots of unity). A triangulation is defined as a maximal decomposition of the regular n-gon into $n-2$ triangles; the size of the triangulation is taken as the number of triangles, that is, $n-2$. 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 (where the arrow points to a possible choice of roots):



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

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

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*:

$$T_n = \mathrm{C}_n \equiv \frac{1}{n+1}\binom{2n}{n}, \qquad \textit{where } \mathrm{C}_n \textit{ is a Catalan number.}$$

This particular result goes back to Euler and Segner (1753), a century before Catalan; see Figure 1 for first values and p. 48 for related bijections.

$\triangleright$ **9.** *A variant specification of triangulations.* Consider the class $\mathcal{U}$ of "nonempty" triangulation of the n-gon, that is, we exclude the 2-gon and the coresponding "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 $U = z(1+U)^2$. $\triangleleft$

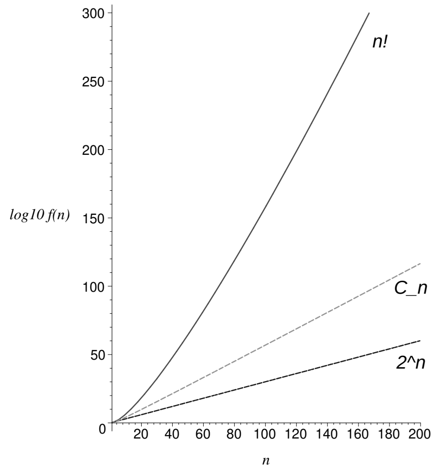


FIGURE 3. The growth regimes of three sequences $f(n) = 2^n,\ \mathrm{C}_n,\ n!$, with a plot of $\log_{10} f(n)$ versus n.

I. 2.4. Asymptotic interpretation of counting sequences. Even in simplest cases, counting sequences delivered by the symbolic method may not be too easy to interpret directly. On the other hand, 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 asymptotic analysis, of which a thorough treatment is presented in Chapters III–V. Here, we content ourselves with a few remarks based on elementary real analysis. (The basic notations are described in APPENDIX: *Asymptotic Notation*, p. 166.)

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 what is known as "Stirling's formula", that is, an approximation to the factorial numbers due to the Scottish mathematician James Stirling (1692–1770):

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n}\left(1 + O(\frac{1}{n})\right) \qquad (n \to +\infty). \tag{11}$$

This formula shows that the factorial numbers grow superexponentially fast, and in particular, grow much faster than W_n. 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

shows 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.

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

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

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 52445 89437 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 4. The Catalan numbers C_n, their Stirling approximation $\mathrm{C}_n^\star = 4^n/\sqrt{\pi n^3}$, and the ratio $\mathrm{C}_n^\star \,/\, \mathrm{C}_n$.

which simplifies to

$$\mathrm{C}_n \sim \frac{4^n}{\sqrt{\pi n^3}}. \tag{12}$$

Thus, the growth of Catalan numbers is roughly comparable to an exponential, 4^n, modulated by a "polynomial" factor, here $1/\sqrt{\pi n^3}$. A surprising consequence of this asymptotic estimate to the area of boolean function complexity appears in Example 12 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 1 exemplifies the quality of the approximation with subtler phenomena 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 Ex. 21 for related coding arguments. $\triangleleft$

$\triangleright$ **11.** *Experimental asymptotics.* From the data of Figure 4, 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., [**89**] for related asymptotic expansions and [**22**] 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, lead to similarly simple asymptotic estimates.

I. 3. Integer compositions and partitions

This section and the next one provide first illustrations of the symbolic method and of counting via specifications. In this framework, generating functions are obtained with hardly any computation. 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 they simply combine two levels of constructions of type $\mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$.

I. 3.1. Compositions and partitions. First the definitions:

DEFINITION I.6. *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.

Graphically, compositions may be seen as as "ragged-landscapes" (represent the summands vertically) or equivalently as alignments of balls with dividing lines, the "balls-and-bars" model; in contrast, partitions appear as "staircases" also known as Ferrers diagrams [**28**, p. 100]; see Figure 5. We let $\mathcal{C}$ and $\mathcal{P}$ denote the class of of 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 (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

$$I(z) = \sum_{n \geq 1} z^n = \frac{z}{1-z}, \tag{13}$$

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,

$$\mathcal{I} = \{1,\ 2,\ 3,\ \ldots\} = \{\bullet,\ \bullet\bullet,\ \bullet\bullet\bullet,\ \ldots\} \cong \mathfrak{S}_{\geq 1}\{\bullet\}, \tag{14}$$

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

From their definition, the classes $\mathcal{C}$ and $\mathcal{P}$ can then be specified as

$$\mathcal{C} = \mathfrak{S}\{\mathcal{I}\}, \qquad \mathcal{P} = \mathfrak{M}\{\mathcal{I}\}. \tag{15}$$

In sequences, the order of components is taken into accout, which precisely models compositions. In multisets, order is not taken into account (while repetitions are allowed), so that we do have an adequate specification of partitions. In both cases, size is correctly inherited additively from summands.

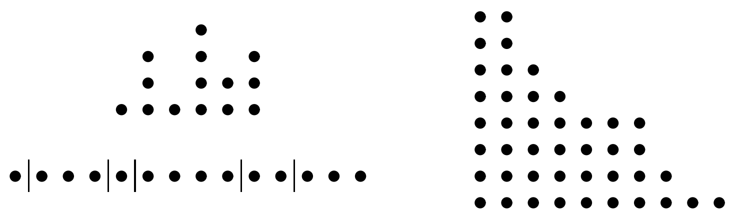


FIGURE 5. Graphical representations of compositions and partitions: (left) the composition $1+3+1+4+2+3 = 14$ with its "ragged-landscape" and "balls-and-bars" models; (right) the partition $8+8+6+5+4+4+4+2+1+1 = 43$ with its staircase (Ferrers diagram) model.

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 6. 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 $P_n = e^{O(\sqrt{n})}$.

First, the specification $\mathcal{C} = \mathfrak{S}\{\mathcal{I}\}$ admits, by Theorem I.1, a direct translation into OGF:

$$C(z) = \frac{z}{1 - I(z)}. \tag{16}$$

The collection of equations (13), (16) 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

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

implying

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

(Naturally, the C_n bear no relation to the Catalan numbers, C_n.) This agrees with basic combinatorics since a composition of n can be viewed as the placement of $n-1$ separation bars between n aligned balls (the "balls and bars" model of Figure 5), of which there are clearly 2^{n-1} possibilities.

Next, the form of the partition generating function derives from Theorem I.1; the general translation mechanism provides the relation

$$P(z) = \exp\left(I(z) + \frac{1}{2} I(z^2) + \frac{1}{3} I(z^3) + \cdots \right) \quad \text{with} \ \ I(z) = \frac{z}{1-z}. \tag{17}$$

In a special case like this, it is just as easy, however, to appeal directly to the product representation and get the more familiar form

(18)

$$\begin{aligned} P(z) &= \prod_{m=1}^{\infty} \frac{1}{1 - z^m} \\ &= 1 + z + 2\,z^2 + 3\,z^3 + 5\,z^4 + 7\,z^5 + 11\,z^6 + 15\,z^7 + 22\,z^8 + 30\,z^9 + \cdots . \end{aligned}$$

	Spec.	OGF	coeff.	asympt.
Composition	$\mathfrak{S}\{\mathfrak{S}_{\geq 1}\{Z\}\}$	$\frac{1-z}{1-2z}$	2^{n-1}	$\frac{1}{2}2^n$
—, sum. $\leq r$	$\mathfrak{S}\{\mathfrak{S}_{1\,..\,r}\{Z\}\}$	$\frac{1-z}{1-2z+z^{r+2}}$	Eq. (19)	$c_r\rho_r^{-n}$
—, k sum.	$\mathfrak{S}_k\{\mathfrak{S}_{\geq 1}\{Z\}\}$	$\frac{z^k}{(1-z)^k}$	$\binom{n-1}{k-1}$	$\frac{n^{k-1}}{(k-1)!}$
Partitions	$\mathfrak{M}\{\mathfrak{S}_{\geq 1}\{Z\}\}$	$\prod_{m=1}^{\infty}(1-z^m)^{-1}$	—	$\frac{1}{4n\sqrt{3}}e^{\pi\sqrt{\frac{2n}{3}}}$
—, sum. $\leq r$	$\mathfrak{M}\{\mathfrak{S}_{1\,..\,r}\{Z\}\}$	$\prod_{m=1}^{r}(1-z^m)^{-1}$	—	$\frac{n^{r-1}}{r!(r-1)!}$
—, $\leq k$ sum.	$\cong \mathfrak{M}\{\mathfrak{S}_{1\,..\,k}\{Z\}\}$	$\prod_{m=1}^{k}(1-z^m)^{-1}$	—	$\frac{n^{k-1}}{k!(k-1)!}$
Cyclic comp.	$\mathfrak{C}\{\mathfrak{S}_{\geq 1}\{Z\}\}$	Eq. (23)	Eq. (24)	$\frac{2^n}{n}$
Part., distinct sum.	$\mathfrak{P}\{\mathfrak{S}_{\geq 1}\{Z\}\}$	$\prod_{m=1}^{\infty}(1+z^m)$	—	$\frac{3^{3/4}}{12n^{3/4}}e^{\pi\sqrt{\frac{n}{3}}}$

FIGURE 7. Partitions and compositions: specifications, generating functions, counting sequences, and asymptotic approximation.

Contrary to compositions that are counted by the explicit formula 2^{n-1}, so simple form exists for p_n. Asymptotic analysis of the OGF (17) based on the saddle point shows that $P_n = e^{O(\sqrt{n})}$. In fact a very famous theorem of Hardy and Ramanujan later improved by Rademacher, see [**4**], 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 much fewer partitions than compositions (Figure 6).

$\triangleright$ **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 also Ex. 33. Ex. 18 further lowers the bound in the case of partitions to $O(N\sqrt{N})$.) $\triangleleft$

When considering variations of the scheme (14), a number of counting results follow rather straightfowardly. We discuss below the case of compositions and partitions with restricted summands, as well as with a fixed number of parts. First, we state:

PROPOSITION I.1. *Let $\mathcal{T} \subseteq \mathcal{I}$ be a subset of the positive integers. The OGF of the classes $\mathcal{C}^{\mathcal{T}} := \mathfrak{S}\{\mathfrak{S}_{\mathcal{T}}\{\mathcal{Z}\}\}$ and $\mathcal{P}^{\mathcal{T}} := \mathfrak{M}\{\mathfrak{S}_{\mathcal{T}}\{\mathcal{Z}\}\}$ 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. $\square$

EXAMPLE 1. *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

$$\mathcal{C}^{\{1,2\}} = \mathfrak{S}\{\mathcal{I}^{\{1,2\}}\} \qquad \text{with } \mathcal{I}^{\{1,2\}} = \{1,2\}.$$

Thus, in terms of generating functions, the relation

$$C^{\{1,2\}}(z) = \frac{1}{1 - I^{\{1,2\}}(z)}$$

holds (see Eq. (16)), with

$$I^{\{1,2\}}(z) = z + z^2.$$

Then,

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

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

$$(19) \qquad C_n^{\{1,\ldots,r\}} = [z^n]\sum_j \left(\frac{z(1-z^r)}{(1-z)}\right)^j = \sum_{j,k}(-1)^k \binom{j}{k}\binom{n-rk-1}{j-1}.$$

Asymptotically, for any fixed r, one checks that there is a unique root ρ_r of the denominator $1 - 2z + z^{r+1}$ in $(\frac{1}{2}, 1)$, that this root dominates all the other roots, and that it is simple. Consequently, one has

$$(20) \qquad C_n^{\{1,\ldots,r\}} \sim c_r \rho_r^{-n} \quad \text{for fixed } r \text{ as } n \to \infty.$$

The quantity ρ_r plays a rôle similar to that of the golden ratio when $r = 2$. Details of the asymptotic analysis are discussed in Chapter 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 4. ◁

EXAMPLE 2. *Partitions with resricted summands and 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 2 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

$$[z^{99}]\,\frac{1}{(1-z)(1-z^5)(1-z^{10})(1-z^{25})} = 242.$$

In the same vein, one proves [**28**, p. 108] that

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

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 [**28**, p. 108].

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

$$P^{\{1,\ldots,r\}}(z) = \prod_{m=1}^{r} \frac{1}{1-z^m}. \tag{21}$$

In other words, we are enumerating partitions according to the value of the largest summand. One then has by looking at the poles

$$P_n^{\{1,\ldots,r\}} \sim c_k n^{k-1} \text{ with } c_k = \frac{1}{k!(k-1)!}.$$

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 is due to Schur and is proved in Chapter IV. □

We next examine the statistic of the number of summands. Let $\mathcal{C}^{(k)}$ denote the class of compositions made of k summands, k a fixed integer ≥ 1. One has

$$\mathcal{C}^{(k)} = \mathcal{I} \times \mathcal{I} \times \cdots \times \mathcal{I},$$

where the number of terms in the cartesian product is k, and $\mathcal{I}$ still represents the summands, i.e., the class of positive integers. 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 directly from the "balls and bars" model of compositions (Figure 5).

Partitions, are naturally represented as collections of points (the staircase model of Figure 5) in the $\mathbb{N} \times \mathbb{N}$ lattice. A geometric symmetry around the main diagonal (also known in the specialized literature as conjugation) exchanges number of summands and value of largest summand, so that the OGF $P^{(\leq k)}(z)$ of partitions with at most k summands coincides with the OGF of partitions with summands all at most k already enumerated in (21)

$$P^{(\leq k)}(z) \equiv P^{\{1,\ldots,k\}} = \prod_{m=1}^{k} \frac{1}{1-z^m}; \tag{22}$$

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

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

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

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

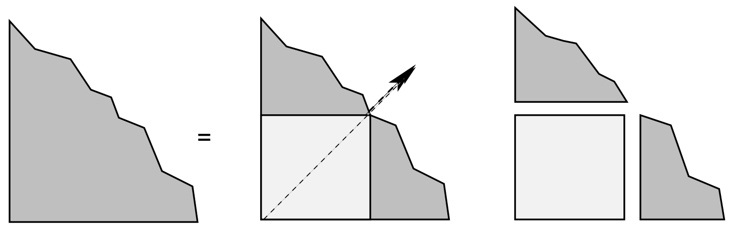
and is expressible as a simple binomial convolution. $\triangleleft$

$\triangleright$ **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 [**4, 28**]. $\triangleleft$

The last problem of this section exemplifies the close interplay betwen combinatorial decompositions and special function identities, which constitutes a recurrent theme of classical combinatorial analysis. The diagram of any partition contains a uniquely determined square (the "Durfee square") that is maximal, as exemplified by the following diagram:



This decomposition gives the identity

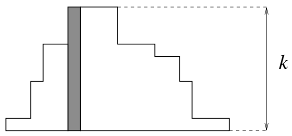
$$\prod_{n=1}^{\infty}\frac{1}{1-z^n} = \sum_{k\geq 0}\frac{z^{k^2}}{((1-z)\cdots(1-z^k))^2},$$

expressing, via (21) and (22), the combinatorial isomorphism (k is the size of the Durfee square)

$$\mathcal{P} \cong \bigcup_{k\geq 0}\left(\mathcal{Z}^{k^2} \times \mathcal{P}^{(\leq k)} \times \mathcal{P}^{\{1,\ldots,k\}}\right),$$

itself nothing but a formal rewriting of the geometric decomposition. As time goes, we shall make greater and greater use of such “direct” translations of object descriptions into generating function equations.

$\triangleright$ **16.** *Stack polyominos.* These are diagrams of compositions such that for some j, one has $1 \leq x_1 \leq x_2 \leq \cdots \leq x_j \geq x_{j+1} \geq \cdots \geq x_k \geq 1$. The diagram representation of stack polyominos,



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

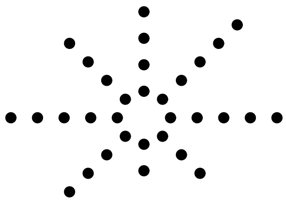
translates immediately into the OGF

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

once use is made of the partition GFs $P^{\{1,\dots,k}(z)$ of (21). The book of van Rensburg [**144**] describes many such constructions and their relation to certain models of statistical physics. $\triangleleft$

I. 3.2. Integer related constructions. Finally, we say a few words about the two constructions of cycle and powerset that haven't been yet applied to $\mathcal{I}$. First, the class $\mathcal{D} = \mathfrak{C}\{I\}$ comprises *cyclic compositions*, that is, compositions defined up to circular shift; 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 segments (taken up to circular symmetry).

A “wheel” (cyclic composition):



By the cycle construction, the OGF is
(23)

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

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

$$D_n = \frac{1}{n} \sum_{k \mid n} \varphi(k)(2^{n/k} - 1) \equiv -1 + \frac{1}{n} \sum_{k \mid n} \varphi(k) 2^{n/k} \sim \frac{2^n}{n}. \tag{24}$$

(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 $D_n \sim C_n/(2n)$.)

More interestingly perhaps, the class $\mathcal{Q} = \mathfrak{P}\{I\}$ is the subclass of $\mathcal{P} = \mathfrak{M}\{I\}$ corresponding to *partitions into distinct summands*: these are determined like in Definition I.6 but with the strict inequalities $x_k > \cdots > x_1$, so that the OGF is

$$Q(z) = \prod_{n\geq 1} (1 + z^n).$$

The coefficients are not amenable to closed from. However the saddle point method (Chapter 6) yields the approximation:

$$Q_n \sim \frac{3^{3/4}}{12n^{3/4}} \exp\left(\pi\sqrt{\frac{n}{3}}\right), \tag{25}$$

which has a shape similar to that of P_n.

▷ **17.** *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). ◁

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

$$\begin{aligned}
P^{\rm pow}(z) &= \prod_{j=0}^{\infty}\frac{1}{1-z^{2^j}} \\
&= 1+z+2\,z^2+2\,z^3+4\,z^4+4\,z^5+6\,z^6+6\,z^7+10\,z^8+10\,z^9+\cdots \\
Q^{\rm pow}(z) &= \prod_{j=0}^{\infty}(1+z^{2^j}) \\
&= 1+z+z^2+z^3+z^4+z^5+\cdots.
\end{aligned}$$

The first sequence $1, 1, 2, 2, \ldots$ is the "binary partition sequence" (*EIS* **A018819**); the difficult asymptotic analysis was performed by de Bruijn [**34**] 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 [**4, 28**] for an introduction to this fascinating subject.

▷ **18.** *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 [**28**, p. 105]. As a consequence, the numbers $\{P_j\}_{j=0}^{N}$ can be determined in $O(N\sqrt{N})$ arithmetic operations. ◁

▷ **19.** *Lattice points.* The number of lattice points with integer coordinates that belong to the closed ball of radius n in d–dimensional 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 OGF's are useful in cryptography [**93**] and estimates may be obtained from the saddle point method.) ◁

I. 4. Words and regular languages

First a finite *alphabet* $\mathcal{A}$ whose elements are called *letters* is fixed. 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 ($\mathcal{A} = \{\mathsf{a},\ldots,\mathsf{z}\}$), sequences written as ygololihq, philology, zgrmblglps are words. The set of all words (often written as $\mathcal{A}^\star$ in formal linguistics) will be consistently denoted by $\mathcal{W}$ here. 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

$$\mathcal{W} \cong \mathfrak{S}\{\mathcal{A}\} \quad \text{implying} \quad W(z) = \frac{1}{1-mz}, \tag{26}$$

where m is the cardinality of the alphabet, *i.e.*, the number of letters. The generating function gives us (in an admittedly devious way) the counting result

$$W_n = m^n.$$

As is usual with symbolic methods, many enumerative consequences usually result from a given construction, and it is precisely the purpose of this section to examine some of them.

We shall introduce 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. It turns out that both frameworks determine the same family of languages, the *regular languages*, though the equivalence is nontrivial, and each particular problem usually admits a preferred representation. The resulting GFs are invariably rational functions.

I. 4.1. Regular specifications. Consider first 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[4] symbols, we have the alternative decomposition:

$$\mathcal{W} \cong \mathfrak{S}\{a\}\,\mathfrak{S}\{b\,\mathfrak{S}\{a\}\}. \tag{27}$$

[4] As is usual when dealing with words, we omit writing explicitly redundant braces '$\{,\}$' and cartesian products '$\times$'. Thus, for instance, $\mathfrak{S}\{a+b\}$ and $\{ab\}$ are shorthand notations for $\mathfrak{S}\{\{a\}+\{b\}\}$ and $\{\{a\}\times\{b\}\}$.

A check is provided by computing the OGF corresponding to this new specification,

$$W(z) = \frac{1}{1-z}\frac{1}{1-z\frac{1}{1-z}}, \tag{28}$$

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

The interest of the decomposition just seen is to take into account various other interesting properties, for example longest runs. Denote by $a^{<k} := \mathfrak{S}_{<k}\{a\}$ the collection of all words formed with the letter a only and whose length is between 0 and $k-1$; the corresponding OGF is $1 + z + z^2 + \cdots + z^{k-1} = (1-z^k)/(1-z)$. The collection $\mathcal{W}^{\langle k\rangle}$ of words which do not have k consecutive a's is described by an amended form of (27), namely

$$\mathcal{W}^{\langle k\rangle} = a^{<k}\,\mathfrak{S}\{ba^{<k}\}. \tag{29}$$

The corresponding OGF obtains immediately from (29)

$$W^{\langle k\rangle}(z) = \frac{1-z^k}{1-z}\cdot\frac{1}{1-z\frac{1-z^k}{1-z}} = \frac{1-z^k}{1-2z+z^{k+1}}.$$

This is therefore the generating functions of words whose longest run of consecutive a's is of length $< k$. From this computation and some asymptotic analysis, it can be deduced that the longest run of a's in a random binary string of length n is about $\log_2 n$. Such asymptotic aspects will be further explored in later chapters.

$\triangleright$ **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 expressive power of nested constructions involving sequences. We set:

DEFINITION I.7. *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 regular specification $\mathcal{R}$ such that $\mathcal{L}$ and $\mathcal{R}$ are combinatorially isomorphic, $\mathcal{L} \cong \mathcal{R}$.*

It is a non-trivial fact that the notion of S-regularity introduced here coincides with the usual notion of regularity in formal language theory. See APPENDIX: *Regular languages*, p. 171 for explanations. From the definition and the basic theorem regarding admissibility (Theorem I.1), one has 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 as well as their relation to finite automata discussed in the next subsection.

EXAMPLE 3. *Combinations and spacings.* The specification $\mathcal{L} = \mathfrak{S}\{a\}\,(b\mathfrak{S}\{a\})^k$ describes unambiguously the set of words that contain exactly k occurrences of the letter b.

The OGF is $L(z) = z^k/(1-z)^{k+1}$, and the number of words in the language satisfies

$$L_n = [z^n]\frac{z^k}{(1-z)^{k+1}} = \binom{n}{k}.$$

Each word of length n is characterized by the positions of its letters b, which means the choices of k positions amongst n possible ones. Formal language theory thus gives 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 another element. The refinement

$$\mathcal{L}^{[d]} = \mathfrak{S}\{a\}\left(b\mathfrak{S}_{<d}\{a\}\right)^{k-1}\left(b\mathfrak{S}\{a\}\right)$$

provides the generating function

$$\sum_{n\geq 0}\binom{n}{k}_{<d} z^n = \frac{z^k(1-z^d)^{k-1}}{(1-z)^{k+1}},$$

which is equivalent to a binomial convolution expression for $\binom{n}{k}_{<d}$. (This problem is clearly analogous to compositions with bounded summands.) $\square$

EXAMPLE 4. *Double run statistics.* By forming maximal groups of equal letters in words, one finds easily that, for a binary alphabet,

$$\mathcal{W} = \mathfrak{S}\{b\}\,\mathfrak{S}\{a\mathfrak{S}\{a\}\,b\mathfrak{S}\{b\}\}\,\mathfrak{S}\{a\}.$$

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 $\mathfrak{S}\{a\}, \mathfrak{S}\{b\}$ by $\mathfrak{S}_{<\alpha}\{a\}, \mathfrak{S}_{<\beta}\{b\}$ internally, and by $\mathfrak{S}_{\leq\alpha}\{a\}, \mathfrak{S}_{\leq\beta}\{b\}$ externally. In particular, the OGF of binary words that never have more than r consecutive equal letters is found to be (set $\alpha = \beta = r$)

$$W^{\langle r,r\rangle} = \frac{(1-z^{r+1})^2}{1-2z+2z^{r+2}-z^{2r+2}} \tag{30}$$

Révész in [**121**] 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 (30). The probabilities are then easily computed using any symbolic algebra 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". Hence, the statistician simply selects the slips that contain runs of length 6 or more. Et voilà! □

▷ **21.** *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 will be called "acceptable".

From the GF, one finds $[z^{11}]W^{\langle 4,4\rangle}(z) = 1546 > 2^{10} = 1024$. Consequently, a code can be built that translates 10 bit blocks into acceptable 11 bit blocks, and only needs a built-in table of size 1024. Such a code has a loss factor of 10%.

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 $1 - 2\alpha + 2\alpha^6 - \alpha^{10} = 0$, which is a pole of $W^{\langle 4,4\rangle}$. One has $\log_2(1/\alpha) = 1.05621$.) Thus, a loss of at least 5% *must* be incurred because of the coding constraint. See Ex. 10 for related coding theory arguments. This limit rate of 1.056 can be approached arbitrarily well, albeit with codes of growing complexity. ◁

EXAMPLE 5. *Patterns in a random 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)

$$\boxed{\text{comb}}\text{at }\boxed{\text{in}}\text{ which our v}\boxed{\text{a}}\text{lian}\boxed{\text{t}}\text{ Hamlet [...] f}\boxed{\text{or}}\text{fe}\boxed{\text{i}}\text{t [...] Whi}\boxed{\text{c}}\text{h he }\boxed{\text{s}}\text{tood ...}$$

A census shows that there are in fact $1.63\,10^{39}$ occurrences hidden somewhere amongst the 120,057 letters that constitute the text. Is this the sign of a secret encouragement passed to us by the author of Hamlet?

Take a fixed finite alphabet $\mathcal{A}$ comprising m letters ($m = 26$ for English). Let $\mathfrak{p} = p_1 p_2 \cdots p_k$ be a word of length k. Consider the regular specification

$$\mathcal{O} = \mathfrak{S}\{\mathcal{A}\}\, p_1\, \mathfrak{S}\{\mathcal{A}\}\, p_2\, \mathfrak{S}\{\mathcal{A}\} \cdots \mathfrak{S}\{\mathcal{A}\}\, p_{k-1}\, \mathfrak{S}\{\mathcal{A}\}\, p_k\, \mathfrak{S}\{\mathcal{A}\}.$$

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

$$\Omega_n = \frac{[z^n]O(z)}{m^n} = m^{-k}\binom{n}{k}, \tag{31}$$

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 $6.96\,10^{38}$. 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.

In the same vein, one can describe all the occurrences of a fixed word $\mathfrak{p} = p_1 p_2 \cdots p_k$ as a *contiguous* block (a "factor") in texts:

$$\widehat{\mathcal{O}} = \mathfrak{S}\{\mathcal{A}\}\,(p_1 p_2 \cdots p_k)\,\mathfrak{S}\{\mathcal{A}\},$$

so that the OGF is

$$\widehat{O}(z) = \frac{z^k}{(1-mz)^2}.$$

Consequently, the expected number of such contiguous occurrences satisfies

$$\widehat{\Omega}_n = m^{-k}(n-k+1) \sim \frac{n}{m^k}. \tag{32}$$

□

For patterns, the estimation of the mean in (31) and (32) can be easily obtained by direct probabilistic reasoning. The example is only meant to demonstrate a symbolic approach to pattern statistics that proves extremely versatile: it can accommodate various notions of patterns (e.g., we may impose maximal spacings between letters) and provide valuable informations on probability distributions as well; see [**50**]. Such methods are of interest in the statistical analysis of texts and in assessing the significance of patterns detected in molecular biology; see [**149**, Ch. 12] for an introduction. From the combinatorial standpoint, these examples illustrate the counting of structures that are richer than words (namely, pattern occurrences) by means of regular specifications.

▷ **22.** *Patterns with gaps.* If less than d symbols of the text must separate the letters of the pattern in order to form a valid occurrence, then the OGF of occurrences is

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

See [**50**] for variations of this theme. ◁

I. 4.2. Finite automata. Let again a finite alphabet $\mathcal{A}$ be fixed. We first define a simple device that is able to "process" words over the alphabet and has wide descriptive power as regards structural properties of words.

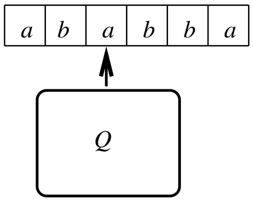
DEFINITION I.8. *A* finite automaton *is a directed multigraph whose edges are labelled by letters of the alphabet. It is customary to call the vertices by the name of* states *and denote by Q the set of states. An initial state $q_0 \in Q$ and a set of final states $Q_f \subseteq Q$ are also designated. A word $w = w_1 \ldots w_n$ 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 $w_1, \ldots, w_n$.*

An 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 language is said to be A–*regular *(automaton regular) if it coincides with the set of words accepted by a deterministic finite automaton.*

The following equivalence theorem is briefly discussed in the Appendix (see APPENDIX: *Regular languages*, p. 171):

> THEOREM (Kleene–Rabin–Scott). *For a language, the following four conditions are equivalent: (i) to be S-regular (i.e., representable by a regular specification); (ii) to be A-regular (i.e., recognizable by a deterministic finite automaton); (iii) to be the set of words accepted by a nondeterministic finite automaton; (iv) to be described by a standard regular* expression.

In the case of a deterministic automaton, it is easy to determine whether a word w is accepted: 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. A deterministic automaton is thus a simple processing device that has a finite instruction set governing its evolution when characters are read. Here is a rendering:



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 I. 4.2. The initial state is q_0, and there is a unique final state q_3.

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 [**26**]. 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.

PROPOSITION I.3. *Let G be a deterministic finite automaton with state set $Q = \{q_0, \ldots, q_s\}$, initial state q_0, and set of final states $\overline{Q} = \{q_{i_1}, \ldots, q_{i_f}\}$. 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

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

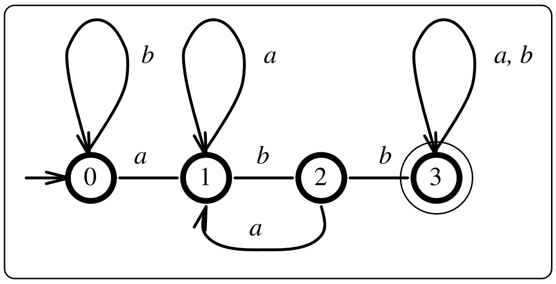


FIGURE 8. Words that contain the pattern abb are recognized by a 4–state automaton with initial state q_0 and final state q_3.

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

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 $j \in \{0, \ldots, s\}$, 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 having read w. The following relation holds for any j:

$$\mathcal{L}_j \cong \Delta_j + \left(\sum_{\alpha \in \mathcal{A}} \{\alpha\} \mathcal{L}_{(q_j \circ \alpha)}\right); \tag{33}$$

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 (33) is then immediate:

$$L_j(z) = [\![q_j \in \overline{Q}]\!] + z \sum_{\alpha \in \mathcal{A}} L_{(q_j \circ \alpha)}(z). \tag{34}$$

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 $L(z) \equiv L_0(z)$. $\square$

For instance, consider the automaton recognizing the pattern abb as given in Figure 8. The languages $\mathcal{L}_j$ (where L_j is the set of accepted words when starting fom 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

$$L_0(z) = \frac{z^3}{(1-z)(1-2z)(1-z-z^2)}. \tag{35}$$

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 the previous subsection, a random word contains a large number, about $\sim n/8$, of occurrences of the pattern abb.)

This example is simple enough that one can also come up with an equivalent regular expression describing $\mathcal{L}_0$: an accepting path in the automaton of Figure 8 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

$$\mathcal{L}_0 = \mathfrak{S}\{b\}\, a\mathfrak{S}\{a\}b\, \mathfrak{S}\{a\mathfrak{S}\{a\}b\}\, b\mathfrak{S}\{a+b\},$$

which gives back a form equivalent to (35), namely,

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

The general construction that reduces systematically finite automata to regular specifications is due to the logician Kleene and is discussed in APPENDIX: *Regular languages*, p. 171.

EXAMPLE 6. *Words containing or excluding a pattern.* Fix an arbitrary pattern $\mathfrak{p} = p_1p_2\cdots p_k$ and let $\mathcal{L}$ be the language of words containing *at least* one occurrence of $\mathfrak{p}$ as a contiguous block. The construction given for the particular pattern $\mathfrak{p} = abb$ generalizes in an easy manner: there exists a deterministic finite automaton with $k+1$ states that recognizes $\mathcal{L}$, the states corresponding to the prefixes of the pattern $\mathfrak{p}$. Thus, the OGF $L(z)$ is *a priori* a rational function of degree at most $k+1$. (The corresponding automaton is in fact known as a Knuth–Morris–Pratt automaton [**88**].)

The automaton construction provides the OGF $L(z)$ in determinantal form but the relation between this rational form and the structure of the pattern is not transparent. An explicit construction due to Guibas and Odlyzko [**74**] 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 $c = (c_0, \ldots, c_{k-1})$ is most conveniently defined in terms of Iverson's bracket as

$$c_i = [\![p_1p_2\cdots p_{k-i} = p_{i+1}p_{i+2}\cdots p_k]\!].$$

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. 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

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

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

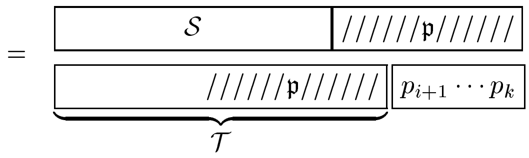
Let $\mathcal{S}$ be the language of words with *no* occurrence of $\mathfrak{p}$ and $\mathcal{T}$ the language of words that end with $\mathfrak{p}$ but have no other occurrence of $\mathfrak{p}$. First, by appending a letter to a word of $\mathcal{S}$, one finds a nonempty word either in $\mathcal{S}$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

$$\mathcal{S} + \mathcal{T} = \{\epsilon\} + \mathcal{S} \times \mathcal{A}. \tag{36}$$

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

$$\mathcal{S} \times \{\mathfrak{p}\} = \mathcal{T} \times \sum_{c_i \neq 0} \{p_{i+1} p_{i+2} \cdots p_k\}, \tag{37}$$

corresponding to the configurations



The translation of the system (36), (37) into OGF's then gives:

The OGF of words not *containing the pattern* $\mathfrak{p}$ *is*

$$S(z) = \frac{c(z)}{z^k + (1 - mz)c(z)}, \tag{38}$$

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

Similarly, the GF's 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. $\square$

$\triangleright$ **23.** *Waiting times in strings.* Let $\mathcal{L} \subset \mathfrak{S}\{a, b\}$ be a language and $S = \{a, b\}^\infty$ be the set of infinite strings with the product probability induced by $\Pr(a) = \mathbb{P}(b) = \frac{1}{2}$. The probability that a random string $\omega \in S$ starts with a word of L is $\widehat{L}(1/2)$, where $\widehat{L}(z)$ is the OGF of the "prefix language" of L, that is, the set of words $w \in L$ 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 $1/2\widehat{L}'(1/2)$. For a regular language, this quantity must be a rational number. $\triangleleft$

$\triangleright$ **24.** *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 [**74**] 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

$\mathfrak{p}$: $3, 4, 5, 5, 6, 6, 7, 8, 8, 8, 8, 9, 9, 10, 11, 14, 15, 15, 16, 21$

$\mathfrak{p}'$: $3, 4, 8, 8, 9, 10, 11, 11, 11, 12, 17, 22, 23, 27, 27, 27, 44, 47, 52, 52.$

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, no contradiction occurs.) $\triangleleft$

$\triangleright$ **25.** *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 contiguous blocks) with probability tending to 1 exponentially fast as $n \to \infty$. (Reason: the rational functions $S(z/2)$ with $S(z)$ as in (38) have no pole in $|z| \leq 1$; see also Chapter 4.)

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 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 that may lead to more explicit expressions.

$\triangleright$ **26.** *Variable length codes.* A finite set $\mathcal{F} \subset \mathcal{W}$, where $\mathcal{W} = \mathfrak{S}\{\mathcal{A}\}$ 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 a construction by automaton.) 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 of Berstel and Perrin [**16**] develops systematically the theory of such "variable-length" codes; see also the construction of the "Aho–Corasick" automaton in [**1**]. $\triangleleft$

$\triangleright$ **27.** *Knight's tours.* For the number of knight's tours on an $n \times w$ chessboard (with fixed w and varying n), the OGF is a rational function. In statistical physics, such automata related methods are commonly used and known as *transfer matrix methods*. $\triangleleft$

I. 4.3. Word related constructions. Words can 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.

EXAMPLE 7. *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 9). Let $\mathcal{S}_n^{(k)}$ denote the collection of all partitions of the set $[1\,..\,n]$ into k non–empty blocks and $S_n^{(k)} = \mathrm{card}(\mathcal{S}_n^{(k)})$ 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 $\mathcal{S}_n^{(k)}$ of an n–set into k blocks by words over a k letter alphabet, $\mathcal{B} = \{b_1, b_2, \ldots, b_k\}$ as follows:

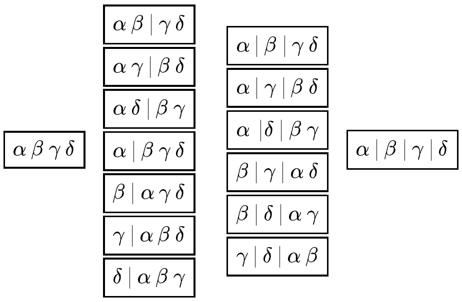


FIGURE 9. The 15 ways to partition a four-element domain into blocks correspond to $S_4^{(1)} = 1, \quad S_4^{(2)} = 7, \quad S_4^{(3)} = 6, \quad S_4^{(4)} = 1$.

Consider a set partition ϖ that is formed of k 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 k 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 r, produce the letter b_r.

For instance to $n = 6$, $k = 3$, 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,

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

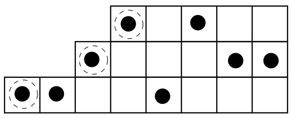
so that the encoding is

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

In this way, a partition is encoded as a word of length n over $\mathcal{B}$ with the additional properties that: (i) all k 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 $\mathcal{S}_n^{(k)}$ is mapped $\mathcal{S}_n^{(k)}$ into words of length n in the language

$$(39) \qquad b_1\,\mathfrak{S}\{b_1\} \cdot b_2\,\mathfrak{S}\{b_1+b_2\} \cdot b_3\,\mathfrak{S}\{b_1+b_2+b_3\} \;\cdots\; b_k\,\mathfrak{S}\{b_1+b_2+\cdots+b_k\}.$$

(The encoding is clearly revertible.) Graphically, this can be rendered by an "irregular staircase" representation, like



where the staircase has length n and height k, each column contains exactly one element, and the columns exposed North-West are systematically filled.

The language specification immediately gives the OGF

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

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

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

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 $S_n^{(k)}$ are nowadays usually denoted by $\left\{ {n \atop k} \right\}$; see APPENDIX: *Stirling numbers*, p. 173. ☐

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.

We conclude this section with a brief mention of "circular words". Let $\mathcal{A}$ be a binary alphabet, viewed as comprised of beads of two distinct colours. The class $\mathcal{N} = \mathfrak{C}\{\mathcal{A}\}$ represents the set of words to taken up to circular shift of their letters. Equivalently, with $\mathcal{A} = \{\bullet, \circ\}$, the class $\mathcal{N}$ describes "necklaces" (p. 3). The OGF of necklaces is given the cyle construction operator:

$$\begin{aligned} N(z) &= \sum_{k=1}^{\infty}\frac{\varphi(k)}{k}\log\frac{1}{1-2z^k} \\ &= 2\,z + 3\,z^2 + 4\,z^3 + 6\,z^4 + 8\,z^5 + 14\,z^6 + 20\,z^7 + 36\,z^8 + 60\,z^9 + \cdots. \end{aligned}$$

Consequently, one has

$$N_n = \frac{1}{n}\sum_{k\ \mid\ n}\varphi(k)2^{n/k}. \tag{40}$$

This is sequence *EIS* **A000031** and one has $N_n = D_n + 1$ where D_n is the wheel count, p. 27. [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 (40).

I. 5. Trees and tree-like structures

This section is concerned with basic tree enumerations. Trees are, as we saw, 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 Chapter 5. a certain type of analytic behaviour appears to be universal in trees, namely a $\sqrt{}$–singularity; as a consequence, most trees families occurring in the combinatorial world have counting sequences obeying the asymptotic form $C\,A^n n^{-3/2}$.

I. 5.1. Plane trees. Plane trees are also sometimes called ordered trees. There, the 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: *Tree concepts*, p. 174 for key concepts associated with trees. They are precisely described in terms of unions, cartesian products, and sequence constructions. Here, we restrict attention to rooted trees.

First, consider the class $\mathcal{G}$ of "general" plane trees where all node degrees are allowed; it satisfies the recursive specification (already discussed on p. 17,

$$\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\}, \tag{41}$$

and, accordingly, $G(z)$ is determined by

$$G(z) = \frac{z}{1 - G(z)}, \qquad \text{hence} \qquad G(z) = \frac{1 - \sqrt{1-4z}}{2}.$$

As a result, the number of general trees of size n is the Catalan number C_{n-1}:

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

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. Thus, for instance $\Omega = \{0, 2\}$ determines binary trees, where each node has either 0 or 2 descendants; $\Omega = \{0, 1, 2\}$ and $\Omega = \{0, 3\}$ determine respectively unary-binary trees and ternary trees; the case of general trees corresponds to $\Omega = \mathbb{Z}_{\geq 0}$. In what follows, an essential rôle is played by the (ordinary) characteristic function of Ω, namely

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

It is in terms of this characteristic function that Ω-restricted trees can be enumerated as shown by the following statement:

PROPOSITION I.4. *The ordinary generating function $T^\Omega(z)$ of the class $\mathcal{T}^\Omega$ of Ω-restricted trees is determined implicitly by the equation*

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

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

$$T_n^\Omega \equiv [z^n]T^\Omega(n) = \frac{1}{n}\,[u^{n-1}]\phi(u)^n. \tag{42}$$

PROOF. The GF equation is a direct consequence of the specification $\mathcal{T}^\Omega = \mathcal{Z}\mathfrak{S}_\Omega\{\mathcal{T}\}$ and of the obvious translation of Ω-restricted sequences:

$$\mathcal{A} = \mathfrak{S}_\Omega\{B\} \quad \Longrightarrow \quad A(z) = \phi(B(z)).$$

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

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

The Lagrange Inversion Theorem precisely states that the expansion of an inverse function (here T) are determined simply by coefficients of powers of the "direct" function (that involves ϕ): see APPENDIX: *Lagrange Inversion*, p. 170. This is precisely what is expressed by (42. □

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.

▷ **28.** *Forests.* Consider ordered k-forests of trees defined by $\mathcal{F} = \mathfrak{S}_k\{\mathcal{T}\}$. 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 8. *"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. 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

$$A_{t\nu+1} = \frac{1}{t\nu+1}\binom{t\nu+1}{\nu} = \frac{1}{(t-1)\nu+1}\binom{t\nu}{\nu}. \tag{43}$$

A particular rôle is played by binary trees. Then a form equivalent to (43) 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 (43) 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 $\widehat{\mathcal{A}}$

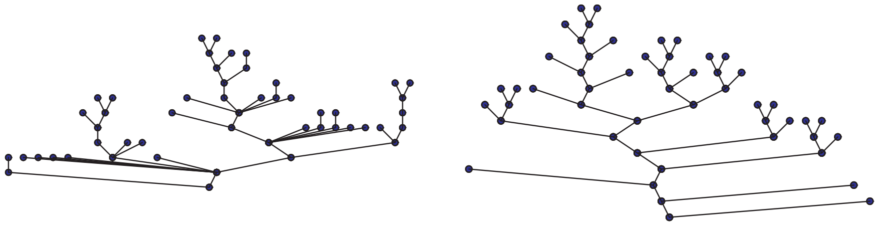


FIGURE 10. A general tree of $\mathcal{G}_{51}$ (left) and a binary tree of $\mathcal{T}_{51}^{\{0,2\}}$ (right) drawn uniformly at random amongst the C_{50} and C_{25} possible trees respectively, with $\mathrm{C}_n = \frac{1}{n+1}\binom{2n}{n}$ the nth Catalan number.

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 (43), since, by basic combinatorics, $\widehat{A}_\nu = A_{t\nu+1}$. $\square$

$\triangleright$ **29.** *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**). $\triangleleft$

$\triangleright$ **30.** *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

$$\widetilde{\mathcal{A}} = \mathcal{Z} + \mathfrak{S}_k\{\widetilde{\mathcal{A}}\}.$$

The binomial formula for $\widetilde{A}_n$ follows from Lagrange inversion applied to $\widetilde{A} = z/(1 - \widetilde{A}^{t-1})$. $\triangleleft$

EXAMPLE 9. *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

$$\mathcal{S} = \mathcal{Z} + \mathfrak{S}_{\geq 2}\{\mathcal{S}\}, \quad \mathcal{Z} = \{x\}. \tag{44}$$

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

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

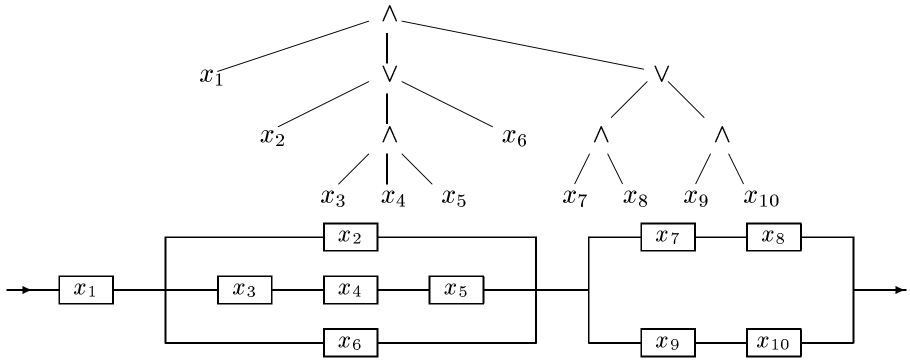


FIGURE 11. An and–or positive proposition of the conjunctive type (top), its associated tree (middle), and an equivalent planar series-paralllel network of the serial type (bottom).

degree at least 2 (while not contributing to size). The functional equation satisfied by the OGF is then

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

This is not *a priori* of the type corresponding to Proposition I.4 because *not* all nodes contribute to size in this particular application. However, the quadratic equation induced by (45) 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 11, bottom), where placement in the plane matters.)

In an instructive paper, Stanley [**136**] 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[5] (*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

[5]This was first observed by David Hough in 1994; see [**136**]. In [**75**], 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.

Tree variety		1 2 3 4 5 6 7 8	n	$+\infty$
Plane gen.	$\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\}$	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.	$\mathcal{T} = \mathcal{Z} + \mathfrak{S}_2\{\mathcal{T}\}$	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.	$\mathcal{H} = \mathcal{Z} \times \mathfrak{M}\{\mathcal{H}\}$	1 1 2 4 9 20 48 115	–	$\sim \lambda \cdot \beta^n / n^{3/2}$
Unord. bin.	$\mathcal{U} = \mathcal{Z} + \mathfrak{M}_2\{\mathcal{U}\}$	1 1 1 2 3 6 11 23	–	$\lambda_2 \cdot \beta_2^n / n^{3/2}$

FIGURE 12. The number of rooted trees of type plane/unordered and general/binary for $n = 1\,..\,8$ and the corresponding asymptotic forms where $\lambda \doteq 0.43992$, $\beta \doteq 2.95576$; $\lambda_2 \doteq 0.79160$, $\beta_2 \doteq 2.48325$. For binary trees, size is by convention the number of external nodes.

and–or connectives in alternation (no "negation"), upon starting from the top in some conventional fashion (e.g, with an and-clause); see Figure 11[6]. Hipparchus was naturally not cognizant of generating functions, but with the technology of the time (and a rather remarkable mind!), he would still be able to discover a recurrence equivalent to (45),

$$S_n = [\![n \geq 2]\!] \left(\sum_{n_1+\cdots+n_k=n} S_{n_1} S_{n_2} \cdots S_n \right) + [\![n = 1]\!], \tag{46}$$

where the sum has only 42 essentially different terms for $n = 10$ (see [**136**] for a discussion), and finally determine S_{10}. ☐

▷ **31.** *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? ◁

▷ **32.** *Faster determination of Schröder numbers.* By forming a differential equation satisfied by $S(z)$ and extracting coefficients, one obtains a recurrence

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

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. ◁

[6]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.

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

$$\mathcal{H} = \mathcal{Z} \times \mathfrak{M}\{\mathcal{H}\},$$

which translates into the *functional equation*

$$\begin{aligned} H(z) &= z(1-z)^{-H_1}(1-z^2)^{-H_2}(1-z^3)^{-H_3}\cdots \\ &= z\exp\left(H(z) + \frac{1}{2}H(z^2) + \frac{1}{3}H(z^3) + \cdots\right). \end{aligned}$$

The first form is due to Cayley in 1857 [**17**, p. 43]; it does not admit a closed form solution, though the equation permits one to determine all the H_n recurrently (*EIS* **A000081**)

$$H(z) = z + z^2 + 2z^3 + 4z^4 + 9z^5 + 20z^6 + 48z^7 + 115z^8 + 286z^9 + 719z^{10} + \cdots .$$

In addition, the local analysis of the singularities of $H(z)$ (Chapter 4) yields a *bona fide* asymptotic expansion for H_n, a fact first discovered by Pólya [**115**] who proved that

$$H_n \sim \lambda \cdot \frac{\beta^n}{n^{3/2}}, \tag{47}$$

for some positive constants $\lambda \doteq 0.43992$ and $\beta \doteq 2.95576$.

$\triangleright$ **33.** *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. 23.) $\triangleleft$

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.5. *Let $\Omega \subset \mathbb{N}$ be a finite set of integers containing 0. Define the "exponential characteristic"*

$$\overline{\phi}(u) = \sum_{\omega\in\Omega} \frac{u^\omega}{\omega!}.$$

The OGF $U(z)$ of nonplane trees with degrees constrained to lie in Ω satisfies the functional equation

$$U(z) = z\overline{\phi}(U(z)) + z\Phi(U(z^2), U(z^3), \ldots),$$

for some polynomial Φ.

PROOF. The class of trees satisfies the combinatorial equation,

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

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.2 implies that the translation of $\mathfrak{M}_k\{\mathcal{A}\}$ is $A(z)^k/k!$ plus a polynomial form in $\{A(z^k)\}_{k\geq 2}$; 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 4, 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 12),

$$U_n^\Omega \sim \lambda_\Omega \frac{(\beta_\Omega)^n}{\sqrt{n^3}}. \tag{48}$$

Many of these questions have their origin in combinatorial chemistry, starting with Cayley in the 19th century [**17**, Ch. 4]. Pólya reexamined these questions, and in his important paper published in 1937 [**113**] he developed at the same time a general theory of combinatorial enumerations under group actions and of asymptotics methods giving rise to estimates like (48). See the book by Harary and Palmer [**76**] for more on this topic or Read's edition of Pólya's paper [**115**].

▷ **34.** *Binary nonplane trees.* Unordered binary trees with size measured by the number of external nodes are described by the equation $\mathcal{U} = \mathcal{Z} + \mathfrak{M}_2\{\mathcal{U}\}$. The functional equation determining $U(z)$ is

$$U(z) = z + \frac{1}{2}U(z)^2 + \frac{1}{2}U(z^2); \quad U(z) = z + z^2 + z^3 + 2z^4 + 3z^5 + \cdots. \tag{49}$$

The asymptotic analysis of the coefficients (*EIS* **A001190**) was carried out by Otter [**111**] who established an estimate of type (48). (The values of the constants are summarized in Figure 12.) The quantity U_n is also the number of structurally distinct products of n elements under a commutative nonassociative binary operation. ◁

▷ **35.** *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 [**17**, 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**)

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

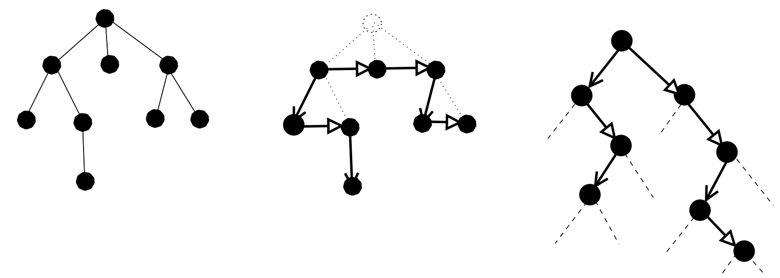
These numbers also enumerate topologically equivalent series-parallel networks (with no plane embedding imposed) as well as hierarchies in statistical classification theory [**142**]. They are the nonplanar analogues of the Hipparchus–Schröder's numbers on p. 43. ◁

I. 5.3. Tree related constructions. Trees underlie recursive structures of all sorts. A first illustration is provided by the fact that the Catalan numbers, $C_n = \frac{1}{n+1}\binom{2n}{n}$ 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 [**137**, 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 the "explain" the occurrence of Catalan numbers in relation to the already encountered classes $\mathcal{G}, \mathcal{B}, \mathcal{T}$.

The combinatorial isomorphism relating $\mathcal{G}$ and $\mathcal{B}$ (albeit with a shift in size) coincides with a classical technique of computer science [**85**, Sec. 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:

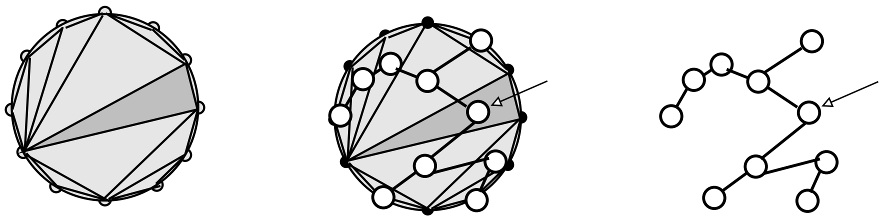
$$\mathcal{G}_n \cong \mathcal{B}_{n-1}.$$

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.

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.

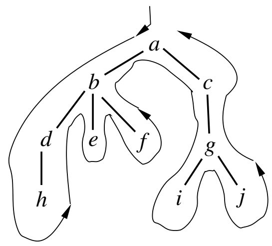


Under this correspondence, tree nodes correspond to triangle faces, while edges connect adjacent triangles. What this correspondence proves is the combinatorial isomorphism

$$\mathcal{T}_n \cong \mathcal{B}_n.$$

We turn next to different types of objects that are in correspondence with trees. These can be interpreted as words encoding tree traversals, and interpretd geometrically as paths in the discrete plane $\mathbb{Z} \times \mathbb{Z}$.

EXAMPLE 10. *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

(50) $\tau =$ 

the first visits to nodes take place in the following order

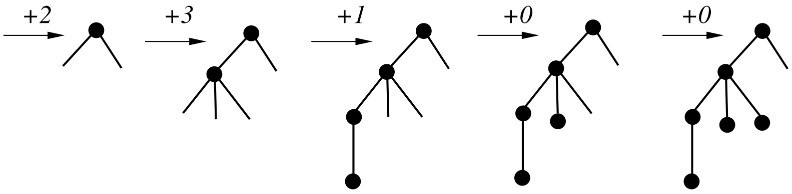
$$a,\quad b,\quad d,\quad h,\quad e,\quad f,\quad c,\quad g,\quad i,\quad j\,.$$

(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 (50), 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,

$$\sigma \rightsquigarrow f_2 f_3 f_1 f_0 f_0 f_0 f_1 f_2 f_0 f_0.$$

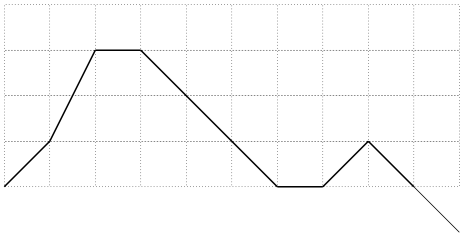
This can be interpreted in logical language a denotation for a functional term built out symbols from $\mathcal{F}$, where f_j represents a "function" of degree j. The correspondence even becomes obvious if superfluous parentheses are added at appropriate place to delimitate scope:

$$\sigma \rightsquigarrow f_2(f_3(f_1(f_0), f_0, f_0), f_1(f_2(f_0, f_0))).$$

Such codes are known as Łukasiewicz codes[7], 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.

[7] A less dignified name is "Polish prefix notation". The "reverse Polish notation" is a variant based on postorder that has been used in calculators since the 1970's.

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:

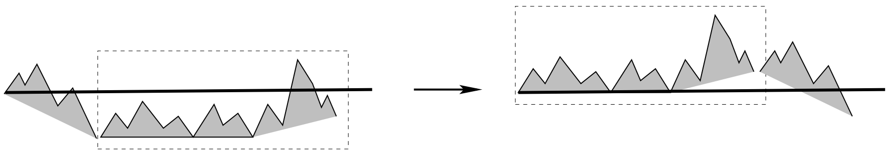


f_2	f_3	f_1	f_0	f_0	f_0	f_1	f_2	f_0	f_0
1	2	0	-1	-1	-1	0	1	-1	-1

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 the its points, except for the very last step, are always in the upper half-plane.

By this correspondence, the number of Łukasiewicz paths with n steps is the shifted Catalan number, $\frac{1}{n}\binom{2n-2}{n-1}$. $\square$

$\triangleright$ **36.** *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 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" [**119**] or the "cycle lemma" [**40**]. Raney has derived from it a purely combinatorial proof of the Lagrange inversion formula [**119**] while Dvoretzky & Motzkin [**40**] have employed this technique to solve a number of counting problems related to circular arrangements. $\triangleleft$

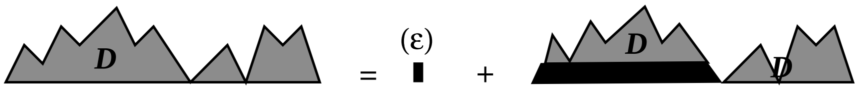
EXAMPLE 11. *Binary tree codes and Dyck paths.* Walks associated with binary trees have a very special form since the vertical displacements can only be $+1$ or -1. The resulting paths of Łukasiewicz type are then equivalently characterized as sequences of numbers $x = (x_0, x_1, \ldots, x_{2n}, x_{2n+1})$ satisfying the conditions

$$(51)\quad x_0 = 0; \qquad x_j \geq 0 \quad \text{for } 1 \leq j \leq 2n; \qquad |x_{j+1} - x_j| = 1; \qquad x_{2n+1} = -1.$$

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 (51) are known in combinatorics as *Dyck paths*[8]. 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

(52)



$$\mathcal{D} \quad = \quad \{\epsilon\} \; + \; (\nearrow \mathcal{D} \searrow) \times \mathcal{D}$$

induces for the OGF of Dyck paths the quadratic equation

$$D(z) = 1 + (zD(z)z)\, D(z),$$

from which the Catalan GF results, and $D_{2n} = \frac{1}{n+1}\binom{2n}{n}$, as expected. The decomposition (52) 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.

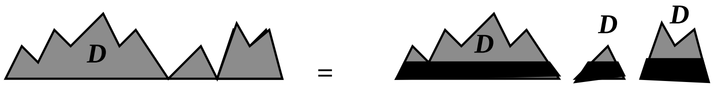
Dyck paths also arise in connection will well-parenthetized expressions. These are recognized by keeping a counter that records at each stage the excess of the number of opening brackets '(' over closing brackets ')'. Finally, one of the origins of Dyck path is the famous "ballot problem", which goes back to the nineteenth century [**99**]: 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 of science is quite remarkable. □

▷ **37.** *Dyck paths and general trees.* The class of Dyck paths admits an alternative sequence decomposition

(53)



$$\mathcal{D} = \mathfrak{S}\{\mathcal{Z} \times \mathcal{D} \times \mathcal{Z}\},$$

[8]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.

which again leads to the Catalan GF. The decomposition (53) is known as the "arch decomposition". 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). ◁

▷ **38.** *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 Ex. 36, it suffices to generate uniformly a sequence of n a's and $n+1$ b's, then reorganize it according to the conjugacy principle. ◁

▷ **39.** *Motzkin paths and unary-binary trees.* Motzkin paths are defined by changing the third condition of (51) defining Dyck paths into $|x_{j+1} - x_j| \le 1$. They appear as codes for unary-binary trees and are enumerated by the Motzkin numbers of Ex. 29. ◁

EXAMPLE 12. *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 ('$\wedge$'); 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

$$D_n = \left(\frac{1}{n+1}\binom{2n}{n}\right) \cdot 2^n \cdot (2m)^{n+1}, \tag{54}$$

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

$$\sum_{j=0}^{\nu} D_j < \|\Omega\|, \tag{55}$$

holds, then there are not enough descriptions of size $\le \nu$ to exhaust Ω. In other terms, there must exist at least one object in Ω whose complexity exceeds ν. If the left side of (55) 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 (12) is available. There results from (54) 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

$$\nu(m) := \frac{2^m}{\log_2 m},$$

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 [**143**], especially the chapter by Li and Vitányi, for a discussion of such counting techniques within the framework of complexity theory. ☐

We finally conclude with a vast generalization of the previous examples.

DEFINITION I.9. *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*

$$\left\{ \begin{array}{lcl} \mathcal{S}_1 & = & \Phi_1(\mathcal{Z}, \mathcal{S}_1, \ldots, \mathcal{S}_r) \\ \vdots & \vdots & \vdots \\ \mathcal{S}_r & = & \Phi_r(\mathcal{Z}, \mathcal{S}_1, \ldots, \mathcal{S}_r), \end{array} \right. \tag{56}$$

where each Φ_j is a constructor that involves only the operations of combinatorial sum ($+$) and cartesian product ($\times$).

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 Schútzenberger [**26**] in the course of their research relating formal languages and formal power series:

PROPOSITION I.6. *A combinatorial class $\mathcal{C}$ that is* context-free *admits an OGF that is an* algebraic function. *In other words, there exists a bivariate polynomial $P(z, y) \in \mathbb{C}[z, y]$ such that*

$$P(z, C(z)) = 0.$$

PROOF. The context-free system (56) 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 performing polynomial elimination are well-known in algebra: one may appeal to a repeated use of resultants [**94**] or to Groebner basis algorithms. See Lang's classic treatise on algebra for resultants [**95**, V.§10] and the excellent introduction to Groebner bases provided by Cox, Little, and O'Shea in [**31**]. □

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. 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 presentation of two types of mechanisms that enrich the framework of constructions: the constructions of pointing and substitution, as well as the use of implicit combinatorial definitions,

I. 6.1. 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.10. *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*

$$\mathcal{B} \circ \mathcal{C} \equiv \mathcal{B}[\mathcal{C}] := \sum_{k \geq 0} \mathcal{B}_k \times \mathfrak{S}_k\{\mathcal{C}\}.$$

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"[9] 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.10, and the intuitive interpretation of pointing and substitution.

THEOREM I.3 (Pointing and substitution). *The constructions of pointing and substitution are admissible*[10]*:*

$$\begin{array}{lll} \mathcal{A} = \Theta\mathcal{B} & \Longrightarrow & A(z) = z\partial_z B(z) \quad \partial_z := \dfrac{d}{dz} \\ \mathcal{A} = \mathcal{B} \circ \mathcal{C} & \Longrightarrow & A(z) = B(C(z)) \end{array}$$

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. $\square$

$\triangleright$ **40.** *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" $\partial(f \circ g) = ((\partial f) \circ g) \cdot \partial g$. $\triangleleft$

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

$$\mathcal{P} = \mathcal{E} + \Theta(\mathcal{Z} \times \mathcal{P}), \qquad \mathcal{E} = \{\epsilon\},$$

---

[9]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., [**103, 109, 152**] for the general theory as well as [**118, 157**] for particular cases like necklaces and trees.

[10]In this book, we borrow from differential algebra the convenient notation $\partial_z := \frac{d}{dz}$ to represent derivatives.

and to the corresponding ordinary differential equation for the OGF,

$$P(z) = 1 + z\frac{d}{dz}(zP(z)),$$

whose formal solution is $P(z) = \sum_{n\geq 0} n!z^n$.

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:

$$\mathcal{M} = \mathcal{B} \circ \mathfrak{S}\{\mathcal{Z}\} \quad \text{and} \quad M(z) = B\left(\frac{z}{1-z}\right).$$

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. 43 (Motzkin numbers).

I. 6.2. 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. 36. 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.4 (Implicit specifications). *The generating functions associated to the implicit equations in* $\mathcal{X}$

$$\mathcal{A} = \mathfrak{S}\{\mathcal{X}\}, \quad \mathcal{A} = \mathfrak{M}\{\mathcal{X}\}$$

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 Moebius 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:

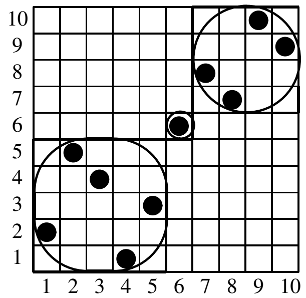
$$\log(A(z)) = \sum_{k=1}^{\infty} \frac{1}{k} X(z^k).$$

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 Moebius inversion; see APPENDIX: *Arithmetical functions*, p. 165. □

EXAMPLE 13. *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 $\sigma = 2\,5\,4\,1\,3\,6\,8\,7\,10\,9$:



$$\sigma = \boxed{2\,5\,4\,1\,3}\ \boxed{6}\ \boxed{8\,7\,10\,9}$$

Thus the class $\mathcal{P}$ of all permutations and the class $\mathcal{I}$ of indecomposable ones are related by

$$\mathcal{P} = \mathfrak{S}\{\mathcal{I}\}.$$

This determines $I(z)$ implicitly, and Theorem I.4 gives:

$$I(z) = 1 - \frac{1}{P(z)} \quad \text{where} \quad P(z) = \sum_{n \geq 1} n!\, z^n .$$

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: *Formal power series*, p. 169). 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 [**28**, p. 262]. □

▷ **41.** *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 [**114**]. See [**24**] for similar multidimensional extensions. ◁

EXAMPLE 14. *Irreducible polynomials over finite fields.* Objects apparently "non–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

$$P \cong \mathfrak{S}\{\mathbb{F}_p\} \quad \text{and} \quad P(z) = \frac{1}{1-pz}, \quad P_n = p^n.$$

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

$$\mathcal{P} \cong \mathfrak{M}\{\mathcal{I}\}$$

expresses precisely the unique factorization property. Thus, the irreducibles are determined implicitly from the class of all polynomials whose OGF is known. Theorem I.4 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 [**32**]), 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 [**83**] for an abstract discussion of statistical properties of arithmetical semigroups.) □

▷ **42.** *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. 14) $Q(z) = P(z)/P(z^2)$, hence

$$Q_n = p^n - p^{n-1} \qquad (n \geq 1).$$

Berlekamp's book [**14**] discusses such facts together with relations to error correcting codes. ◁

▷ **43.** *Balanced trees.* The class $\mathcal{O}$ of balanced 2-3 trees is a familiar data structure [**86**], defined as (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. Balanced trees satisfy an implicit equation based on combinatorial substitution:

$$\mathcal{O} = \mathcal{Z} + \mathcal{O}[(\mathcal{Z} \times \mathcal{Z}) + (\mathcal{Z} \times \mathcal{Z} \times \mathcal{Z})], \qquad O(z) = z + O(z^2 + z^3).$$

Odlyzko [**110**] has determined the growth of O_n (it is like φ^n/n, where $\varphi = (1+\sqrt{5})/2$ is the golden ratio, but involves subtle fluctuations). $\lhd$

I. 7. Notes

There are several lessons to be learnt from the uses that we have surveyed of symbolic combinatorics.

First, for a given class of problems, symbolic methods lead to a unified treatment that reveals a natural class of functions in which generating functions lie. Thus 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, in such a case, based on partial fraction expansion. 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 is necessarily a rational function. Theorems of this sort establish a bridge between combinatorial analysis and special functions. 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.

Second, 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 and V.

Modern presentations of combinatorial analysis appear in the books of Comtet [**28**] (a beautiful book largely example driven), Stanley [**135, 137**] (a rich set with an algebraic orientation), and Wilf [**153**] (generating functions oriented). An elementary but insightful presentation of the basic techniques appears in Graham, Knuth, and Patashnik's classic [**71**], a popular book with a highly original design. An encyclopedic reference is the book of Jackson & Goulden [**68**] 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 [**101**] *Combinatory Analysis* first published in 1917, the introduction of denumerant generating functions by Pólya as exposed in [**116**], or the "domino theory" in [**71**, Sec. 7.1].) One source in recent times is the Chomsky–Schützenberger theory of formal languages and enumerations [**26**]. Rota [**122**] and Stanley [**134, 137**] developed an approach which is largely based on partially ordered sets. Bender and Goldman developed a theory of "prefabs" [**11**] whose purposes are similar to the theory developed here. Joyal [**79**] 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 [**13**]. Parallel (but independent) developments by the "Russian School" are nicely synthetized in the books by Sachkov [**124, 125**].

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 [**87**]. The symbolic ideas exposed here have been applied to the analysis of algorithms in surveys [**46, 147**] and are further exposed in our book [**130**]. Flajolet, Salvy, and Zimmermann [**56**] have shown how to use them in order to automate the analysis of some well characterized classes of combinatorial structures.

CHAPTER II

Labelled Structures and Exponential Generating Functions

Cette approche évacue pratiquement tous les calculs.
— DOMINIQUE FOATA &
MARCEL P. SCHÜTZENBERGER [**64**]

Many objects of classical combinatorics present themselves naturally as labelled structures where "atoms" of an object (typically nodes in a graph or a tree) bear distinctive integer labels. For instance the cycle decomposition of a permutation represents the permutation as an unordered collection of circular graphs whose nodes are labelled by integers.

Commonly encountered classes of labelled objects are permutations, set partitions, labelled graphs and labelled trees, graphs and mappings of a finite set into itself, as well as structures related to occupancy problems.

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.

The labelled constructions translate over exponential generating functions. The translation schemes are analytically simpler than in the unlabelled case considered in the previous chapter. Labelled constructions enable us to take into account structures that are in many ways combinatorially richer, in particular as regards order properties. They therefore constitute a facet with powerful descriptive powers of the symbolic method for combinatorial enumeration.

II. 1. Labelled classes and labelled product

Throughout this chapter, we consider *combinatorial classes* as broadly defined in Chapter I: we deal exclusively with finite objects; a combinatorial class is a set of objects, with a notion of size attached, so that the number of objects of each size is finite. However, the objects are now *labelled* in the sense that each "atom" carries with it an integer label and all the labels occurring in an object are distinct. Precisely, a *weakly labelled* object of size n bears n distinct labels that are integers in $\mathbb{Z}_{\geq 0}$. An object of size n is said to be (strongly or well) *labelled* if it is is weakly labelled and its collection of labels is the consecutive integer interval $[1\,..\,n]$. For a labelled class, the *size* function is systematically defined as the number of labels that the object contains.

As an example, consider the class $\mathcal{G}$ of labelled graphs. An element is by definition an undirected graph such that labels are supported by vertices. A particular labelled graph of size 4 is then

$$g = \quad \begin{array}{ccc} 1 & \text{—} & 3 \\ | & & | \\ 4 & \text{—} & 2 \end{array}\ ,$$

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 abstract graph structure counts, so that this is the same abstract graph as in the alternative visual representations

$$g = \quad \begin{array}{ccc} 1 & \text{—} & 4 \\ | & & | \\ 3 & \text{—} & 2 \end{array}\ ,\quad \begin{array}{ccc} 3 & \text{—} & 2 \\ | & & | \\ 1 & \text{—} & 4 \end{array}\ ,$$

since in all three cases, the lists of edges coincide. However, this graph is different from

$$h = \quad \begin{array}{ccc} 4 & \text{—} & 1 \\ | & & | \\ 3 & \text{—} & 2 \end{array}\ ,$$

since, for instance, 1 and 2 have become adjacent. Altogether, it can be seen that there are 3 different ways to build labelled graphs out of the common unlabelled quadrangle graph

$$\begin{array}{ccc} \star & \text{—} & \star \\ | & & | \\ \star & \text{—} & \star \end{array}\ .$$

See Figure 1 for details.

It is also convenient to introduce the neutral (empty, null) object ϵ that has size 0 and bears no label at all, and consider it as a special case of a labelled object; the *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 bears the integer label ①.

The counting of labelled objects is normally achieved by means of exponential generating functions.

DEFINITION II.1. *The* exponential generating function *(EGF) of a sequence* $\{A_n\}$ *is the formal power series*

$$A(z) = \sum_{n=0}^{\infty} A_n \frac{z^n}{n\,!}. \tag{1}$$

The exponential generating function *(EGF) of a class* $\mathcal{A}_n$ *is the generating function of the numbers* $A_n = \text{card}(\mathcal{A}_n)$. *Equivalently, the EGF of class* $\mathcal{A}$ *is*

$$A(z) = \sum_{n\geq 0} A_n \frac{z^n}{n\,!} = \sum_{\alpha\in\mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|\,!}.$$

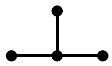
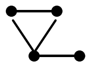
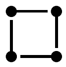
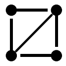
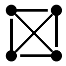
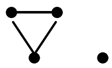
It is also said that the variable z marks *size in the generating function.*

With the standard notation for coefficients of series, the coefficient A_n in an exponential generating function is then recovered by

$$A_n = n\,! \cdot [z^n]\, A(z),$$

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 $a(z)$) will all be

	Unlab.	Lab.
	1	12
	1	4
	1	12
	1	3
	1	6
	1	1
	1	4
	1	12
	1	3
	1	6
	1	1
Total:	**11** Unlab.	**64** Lab.

There are $\widehat{G}_4 = \mathbf{11}$ unlabelled graphs of size 4, i.e., comprising 4 nodes when any number of edges is allowed (left column).
Each unlabelled graph corresponds to a variable number of labelled graphs (indicated in each case by the figure in the right column). For instance, the totally disconnected graph and the complete graph have only 1 labelling. In contrast the line graph has $\frac{1}{2}4! = 12$ possible labellings. For size 4, the number of labellings is seen here to vary between 1 and 12.
The total number of labelled graphs found is $G_4 = \mathbf{64} = 2^6$, in agreement with the general formula

$$G_n = 2^{n(n-1)/2}.$$

See p. 70 for details.

FIGURE 1. Unlabelled versus labelled graphs for size $n = 4$.

denoted by the same group of letters. Clearly, the EGF's of the neutral class and the atomic class are respectively

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

EXAMPLE 1. *Permutations.* The class $\{\mathcal{P}\}$ of all permutations is prototypical of labelled classes. Under the linear representation of permutations, it starts as

$$\mathcal{P} = \left\{ \epsilon\,,\ \textcircled{1}\,,\ \begin{array}{c} \textcircled{1}-\textcircled{2} \\ \textcircled{2}-\textcircled{2} \end{array}\,,\ \begin{array}{c} \textcircled{1}-\textcircled{2}-\textcircled{3} \\ \textcircled{2}-\textcircled{3}-\textcircled{1} \\ \textcircled{3}-\textcircled{1}-\textcircled{2} \\ \textcircled{2}-\textcircled{1}-\textcircled{3} \\ \textcircled{1}-\textcircled{3}-\textcircled{2} \\ \textcircled{3}-\textcircled{1}-\textcircled{2} \end{array}\,,\ \ldots \right\},$$

so that $P_0 = 1$, $P_1 = 1$, $P_2 = 2$, $P_3 = 6$, etc. There, by definition, all the possible orderings between the distinct atoms 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 places at which to place the element 1, then $(n-1)$ possible places for 2, etc). 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. □

EXAMPLE 2. *Urns.* The class $\mathcal{U}$ of totally disconnected graphs starts as

$$\mathcal{U} = \left\{ \epsilon\,,\ \textcircled{1}\,,\ \boxed{\textcircled{1}\ \ \textcircled{2}}\,,\ \boxed{\begin{matrix}\textcircled{1} & \textcircled{2}\\ \multicolumn{2}{c}{\textcircled{3}}\end{matrix}}\,,\ \boxed{\begin{matrix}\textcircled{1} & \textcircled{2}\\ \textcircled{3} & \textcircled{4}\end{matrix}}\,,\ \boxed{\begin{matrix}\textcircled{1} & & \textcircled{2}\\ & \textcircled{5} & \\ \textcircled{3} & & \textcircled{4}\end{matrix}}\,, \dots \right\}.$$

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

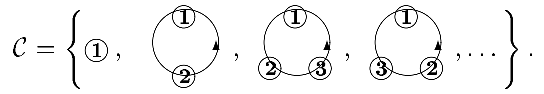
$$U(z) = \sum_{n \geq 0} 1\, \frac{z^n}{n!} = \exp(z) = e^z.$$

(The fact that the EGF of the constant sequence $\{1\}$ is the exponential function explains the term "exponential generating function".) Alternatively, presenting elements of an urn in sorted order leads to a representation of urns as *sorted* linear graphs; for instance,

$$\textcircled{1}-\textcircled{2}-\textcircled{3}-\textcircled{4}-\textcircled{5}$$

is such an equivalent representation of the urn of size 5. Though urns may 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. □

EXAMPLE 3. *Circular graphs.* Finally, the class of circular graphs, where cycles are oriented in some conventional manner (say, positively here) is



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

$$C(z) = \sum_{n \geq 1} (n-1)! \frac{z^n}{n!} = \sum_{n \geq 1} \frac{z^n}{n} = \log \frac{1}{1-z},$$

where, as we shall see shortly, the logarithm is characteristic of circular arrangements of labelled objects. □

II. 2. Admissible labelled constructions

We now describe a toolkit of *constructions* that make it possible to build complex classes from simpler ones. Combinatorial sum or disjoint union is defined exactly as in Chapter I: it is the union of disjoint copies. Novelty here lies in the definition of a product that is adapted to labelled structures. The usual cartesian product is unsuitable since an ordered pair of two labelled objects is not well labelled—for instance the label 1 would invariably appear repeated twice. The labelled product translates naturally into exponential generating functions, and from there simple translation rules follow for labelled sequences, sets, and cycles.

As a preparation to the translation of labelled constructions, we first briefly review the effect of products over EGF's. If $a(z), b(z), c(z)$ are EGF's, with $a(z) = \sum_n a_n z^n/n!$ and so on, we have the *binomial convolution* formula

$$a(z) = b(z) \cdot c(z) \qquad \Longrightarrow \qquad a_n = \sum_{k=0}^{n} \binom{n}{k} b_k c_{n-k}, \tag{2}$$

since, by 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,

$$a(z) = a^{(1)}(z)\, a^{(2)}(z) \cdots a^{(r)}(z) \implies$$

$$a_n = \sum_{n_1+n_2+\cdots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} a_n^{(1)} a_n^{(2)} \cdots a_n^{(r)}. \tag{3}$$

In Eq. (3) there occurs the multinomial coefficient

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}.$$

This multinomial coefficient also counts the number of ways of splitting n elements into r distinguished classes of cardinalities $n_1, \ldots, n_r$. This fact lies at the very heart of most enumerative applications of binomial convolutions and EGF's.

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 between labels.* Then two dual modes of relabellings prove important:

- *Reduction*: For a non–canonically 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 reduction of the structure α.
- *Expansion*: This operation is defined relative to a relabelling function $e \in [1\,..\,n] \mapsto \mathbb{Z}_{\geq 1}$ 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.

We next define a product called the *labelled product*, or simply *product* (originally this was named *partitional product* by Foata who proposed an early formalization in [**62**]). Given two labelled structures $\beta \in \mathcal{B}$ and $\gamma \in \mathcal{C}$, the product $\beta \star \gamma$ comprises the finite collection of objects that are ordered pairs (β', γ') of relabelled copies of (β, γ),

$$\beta \star \gamma := \{\, (\beta', \gamma') \;\big|\; (\beta', \gamma') \text{ is well–labelled},\ \rho(\beta') = \beta,\ \rho(\gamma') = \gamma \,\}, \tag{4}$$

the relabellings preserving the order structure present in β and γ. An equivalent form is via expansion of labels:

$$\beta \star \gamma = \{\, (e(\beta), f(\gamma) \;\big|\; \mathrm{Im}(e) \cap \mathrm{Im}(f) = \emptyset,\ \mathrm{Im}(e) \cup \mathrm{Im}(f) = [1\,..\,|\beta| + |\gamma|] \,\}, \tag{5}$$

where e, f are again assumed to be *increasing* with ranges $\mathrm{Im}(e), \mathrm{Im}(f)$. For instance, one has

$$\left(\begin{array}{ccc} \textcircled{2} \text{---} \textcircled{6} \\ | \qquad\quad | \\ \textcircled{7} \text{---} \textcircled{4} \end{array}, \begin{array}{c} \textcircled{3} \text{---} \textcircled{5} \\ \backslash \qquad / \\ \textcircled{1} \end{array}\right) \in \left(\begin{array}{c} \textcircled{1} \text{---} \textcircled{3} \\ | \qquad\quad | \\ \textcircled{4} \text{---} \textcircled{2} \end{array} \star \begin{array}{c} \textcircled{2} \text{---} \textcircled{3} \\ \backslash \qquad / \\ \textcircled{1} \end{array}\right),$$

as seen by reduction of the left pair or, dually, by expansion of the right pair.

If $\mathcal{B}$ and $\mathcal{C}$ are two 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.2. *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).*

The corresponding counting sequences satisfy the relation,

$$A_n = \sum_{n_1+n_2=n} \binom{n}{n_1, n_2} B_{n_1} C_{n_2}.$$

There the binomial arises since the the number of relabellings involved in forming all the elements of $(\beta \star \gamma)$ is $\binom{n}{n_1, n_2}$, if $|\beta| - n_1$, $|\gamma| = n_2$ and $n_1 + n_2 = n$. The product $B_{n_1} C_{n_2}$ keeps track of all the possibilities for the $\mathcal{B}$ and $\mathcal{C}$ components. By (2), the binomial convolution corresponds to the product relation,

$$A(z) = B(z) \cdot C(z),$$

relating EGFs. Thus, the labelled product simply translates into the product operation on exponential generating functions.

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 $\mathfrak{S}_k\{\mathcal{B}\}$. This corresponds to forming k–sequences and performing all consistent relabellings. The (labelled) *sequence* class of $\mathcal{B}$ is denoted by $\mathfrak{S}\{\mathcal{B}\}$ and is defined by

$$\mathfrak{S}\{\mathcal{B}\} \stackrel{\text{def}}{=} \{\epsilon\} + \mathcal{B} + (\mathcal{B} * \mathcal{B}) + (\mathcal{B} * \mathcal{B} * \mathcal{B}) + \cdots = \bigcup_{k \geq 0} \mathfrak{S}_k\{\mathcal{B}\}.$$

The product relation for EGF's clearly extends to arbitrary products, as seen from the multinomial convolution formula (3), so that

$$\mathcal{A} = \mathfrak{S}_k\{\mathcal{B}\} \implies B(z) = B(z)^k,$$

and (assuming $B_0 \neq 0$)

$$\mathcal{A} = \mathfrak{S}\{\mathcal{B}\} \implies A(z) = \sum_{k=0}^{\infty} B(z)^k = \frac{1}{1 - B(z)}.$$

We denote by $\mathfrak{P}_k\{\mathcal{B}\}$ the class of k–sets formed from $\mathcal{B}$. The powerset class is defined formally, like in the unlabelled case, as the quotient $\mathfrak{P}\{\mathcal{B}\} := \mathfrak{S}_k\{\mathcal{B}\}/\mathbf{R}$ where the equivalence relation $\mathbf{R}$ indentifies two sequences when the components of one are a permutation of the components of the other (p. 9). In simple terms, a "set" is like a

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	$\mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$	$A(z) = \dfrac{1}{1 - B(z)}$
Set	$\mathcal{A} = \mathfrak{P}\{\mathcal{B}\}$	$A(z) = \exp(B(z))$
Cycle	$\mathcal{A} = \mathfrak{C}\{\mathcal{B}\}$	$A(z) = \log \dfrac{1}{1 - B(z)}$

2. The translation for sets, multisets, and cycles of fixed cardinality.

Construction		EGF
Sequence	$\mathcal{A} = \mathfrak{S}_k\{\mathcal{B}\}$	$A(z) = A(z)^k$
Set	$\mathcal{A} = \mathfrak{P}_k\{\mathcal{B}\}$	$A(z) = \dfrac{1}{k!} A(z)^k$
Cycle	$\mathcal{A} = \mathfrak{C}_k\{\mathcal{B}\}$	$A(z) = \dfrac{1}{k} A(z)^k$

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 2. A "dictionary" of *labelled* constructions together with their translation into *exponential* generating functions (EGF's). 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 2 of Chapter I, p. 2.)

sequence, but the order between components is immaterial. The (labelled) *powerset* class of $\mathcal{B}$, denoted $\mathfrak{P}\{\mathcal{B}\}$, is defined by

$$\mathfrak{P}\{\mathcal{B}\} \stackrel{\text{def}}{=} \{\epsilon\} + \mathcal{B} + \mathfrak{P}_2\{\mathcal{B}\} + \cdots = \bigcup_{k\geq 0} \mathfrak{P}_k\{\mathcal{B}\}.$$

A labelled k–set is associated with exactly $k!$ different sequences. (There is here is subtle difference with the the unlabelled case where formulæ are more complex as an unlabelled sequence may contain repeated elements while components of a labelled sequence are all distinguished by their labels.) Thus in terms of EGF's, one has (assuming $\mathcal{B}_0 = \emptyset$)

$$\mathcal{A} = \mathfrak{P}_k\{\mathcal{B}\} \implies A(z) = \frac{B(z)^k}{k!},$$

$$\mathcal{A} = \mathfrak{P}\{\mathcal{B}\} \implies A(z) = \sum_{k=0}^{\infty} \frac{B(z)^k}{k!} = \exp(B(z)).$$

Note that the distinction between multisets and powersets is here immaterial, since by definition components of a set all have distinct labels: in the labelled universe, we have $\mathfrak{M} \equiv \mathfrak{P}$.

We also introduce the class of k–cycles, $\mathfrak{C}_k\{\mathcal{B}\}$ and the cycle class. The cycle class is defined formally, like in the unlabelled case, as the quotient $\mathfrak{C}\{\mathcal{B}\} := \mathfrak{S}_k\{\mathcal{B}\}/\mathbf{S}$ where the equivalence relation $\mathbf{S}$ indentifies two sequences when the components of one are a cyclic permutation of the components of the other (p. 9). In simple terms, a "cycle" is like a sequence, but components can be circularly shifted. In terms of EGF's, we have (assuming $\mathcal{B}_0 = \emptyset$)

$$\mathcal{A} = \mathfrak{C}_k\{\mathcal{B}\} \implies A(z) = \frac{B(z)^k}{k},$$

$$\mathcal{A} = \mathfrak{C}\{\mathcal{B}\} \implies A(z) = \sum_{k=0}^{\infty} \frac{B(z)^k}{k} = \log \frac{1}{1 - B(z)},$$

since each cycle admits exactly k representations as a sequence. In summary:

THEOREM II.1. *The constructions of labelled product, k–th power, and sequence class,*

$$\mathcal{A} = \mathcal{B} \star \mathcal{C}, \qquad \mathcal{A} = \mathfrak{S}_k\{\mathcal{B}\}, \qquad \mathcal{A} = \mathfrak{S}\{\mathcal{B}\}$$

are admissible:

$$A(z) = B(z) \cdot C(z), \qquad A(z) = B(z)^k, \qquad A(z) = \frac{1}{1 - B(z)}.$$

The constructions of k–set and powerset class

$$\mathcal{A} = \mathfrak{P}_k\{\mathcal{B}\}, \qquad \mathcal{A} = \mathfrak{P}\{\mathcal{B}\},$$

are admissible:

$$A(z) = \frac{1}{k!} B(z)^k, \qquad A(z) = \exp(B(z)).$$

The constructions of k–cycle and cycle class,

$$\mathcal{A} = \mathfrak{C}_k\{\mathcal{B}\}, \qquad \mathcal{A} = \mathfrak{C}\{\mathcal{B}\},$$

are admissible:

$$A(z) = \frac{1}{k} B(z)^k, \qquad A(z) = \log \frac{1}{1 - B(z)}.$$

Constructible classes. Like 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 node $\mathcal{N} = \{1\}$ of size 1. Amongst the elementary classes discussed in Section II. 1, one immediately recognizes that

$$\mathcal{P} = \mathfrak{S}\{\mathcal{Z}\}, \quad \mathcal{U} = \mathfrak{P}\{\mathcal{Z}\}, \quad \mathcal{C} = \mathfrak{C}\{\mathcal{Z}\},$$

specify permutations, urns, and circular graphs respectively. These are basic building blocks out of which more complex objects can be constructed. Set partitions ($\mathcal{S}$), surjections ($\mathcal{R}$), permutations ($\mathcal{P}$) under their cycle decomposition, and alignments ($\mathcal{O}$) are then particular constructible classes corresponding to

$$\mathcal{S} \simeq \mathfrak{P}\{\mathfrak{P}_{\geq 1}\{\mathcal{Z}\}\}, \quad \mathcal{R} \simeq \mathfrak{S}\{\mathfrak{P}_{\geq 1}\{\mathcal{Z}\}\}, \quad \mathcal{P} \simeq \mathfrak{P}\{\mathfrak{C}_{\geq 1}\{\mathcal{Z}\}\}, \quad \mathcal{O} \simeq \mathfrak{S}\{\mathfrak{C}_{\geq 1}\{\mathcal{Z}\}\}.$$

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 $\mathfrak{S}_k$), $f^k/k!$ (for $\mathfrak{P}_k$), and f^k/k (for $\mathfrak{C}_k$) are to be added to the list.

II. 2.2. Labelled versus unlabelled? Let $\mathcal{A}$ be a labelled class. If this class is constructible, it automatically has an unlabelled counterpart $\widehat{\mathcal{A}}$ that is obtained by interpreting all the intervening constructions as unlabelled ones, in the sense of Chapter I. Equivalently, one may view objects in $\widehat{\mathcal{A}}$ as obtained from objects of $\mathcal{A}$ by "forgetting the labels". This is formalized by identifying two labelled object if there is an *arbitrary* relabelling (not just order-consistent ones, as have 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. We state:

PROPOSITION II.1. *The counts of a labelled class* $\mathcal{A}$ *and its unlabelled counterpart* $\widehat{\mathcal{A}}$ *are related by*

$$\widehat{A}_n \leq A_n \leq n!\,\widehat{A}_n \quad \textit{or equivalently} \quad 1 \leq \frac{A_n}{\widehat{A}_n} \leq n!. \tag{7}$$

EXAMPLE 4. *Labelled and Unlabelled graphs.* This phenomenon has been already encountered in our discussion of graphs, where 4 labellings can be attached to the unlabelled quadrangle graph of size 4. If one considers instead the totally disconnected graph of size 4, then there exists exactly one labelled version (the "urn" of size 4) and one unlabelled version. 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 $n = 1\,..\,18$

$\widehat{G}_n$ (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
64001015704527557894928	1329227995784915872903807060280344576
245935864153532932683719776	87112285931760246646623899502532662132736
1787577725145611700547878190848	11417981541647679048466287755595961091061972992

The sequence $\{\widehat{G}_n\}$ constitutes *EIS* **A000088**, which can be obtained by an extension of methods of Chapter I; see [**76**, 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 (7) 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
18	$0.6387404239 \cdot 10^{16}$	0.9976618880

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 [**76**] 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.) $\square$

The case of urns and totally disconnect 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, apart from computing separately the two GF's and comparing coefficients.

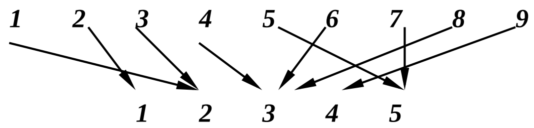
II. 3. Surjections, set partitions, and words

This section and the next are devoted to what could be termed nonrecursive structures of "level 2" defined by the fact that they combine two constructions. Here, we examine classes

$$\mathcal{R} = \mathfrak{S}\{\mathfrak{P}_{\geq 1}\{Z\}\} \qquad \text{and} \qquad \mathcal{S} = \mathfrak{P}\{\mathfrak{P}_{\geq 1}\{Z\}\},$$

corresponding to sequences-of-sets ($\mathcal{R}$) and sets-of-sets ($\mathcal{S}$) respectively. We shall see shortly (Section II. 3.1) that such abstract specifications model classical objects of discrete mathematics, namely surjections ($\mathcal{R}$) and set partitions ($\mathcal{S}$). (These constitute in a way labelled analogues of integer compositions and integer partitions in the unlabelled universe.) The symbolic methodology 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. 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 unto 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 of $\mathcal{R}_9^{(5)}$:



We set $\mathcal{R}^{(r)} = \bigcup_n \mathcal{R}_n^{(r)}$ and proceed to determine 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 preimages, $\big(\phi^{-1}(1), \phi^{-1}(2), \ldots, \phi^{-1}(r)\big)$, themselves disjoint nonempty sets of integers that cover the interval $[1\,..\,n]$. In other words, one has the combinatorial specification

$$\mathcal{R}^{(r)} = \mathfrak{S}_r\{\mathcal{V}\}, \quad \mathcal{V} = \mathcal{U} \setminus \{\epsilon\} = \mathfrak{P}_{\geq 1}\{\mathcal{Z}\},$$

where $\mathcal{V}$ designates the class of urns ($\mathcal{U}$) that are nonempty. Consequently, the EGF satisfies

$$R^{(r)}(z) = (e^z - 1)^r, \tag{8}$$

in view of our earlier discussion of urns ($\mathcal{U}$) with EGF $U(z) = e^z$.

Equation (8) does solve the counting problem for surjections. For small r, one finds

$$R^{(2)}(z) = e^{2z} - 2e^z + 1, \qquad R^{(3)}(z) = e^{3z} - 3e^{2z} + 3e^z - 1,$$

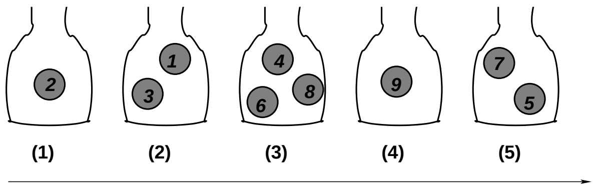
whence, by expanding,

$$R_n^{(2)} = 2^n - 2, \qquad R_n^{(3)} = 3^n - 3 \cdot 2^n + 3\,.$$

A surjection, here the mapping from $[1\,..\,9]$ onto $[1\,..\,5]$ given by the table

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 2 & 1 & 2 & 3 & 5 & 3 & 5 & 3 & 4 \end{pmatrix},$$

may be viewed as an ordered tuple of nonempty urns, or equivalently, linear sorted graphs



$$\begin{aligned} \sigma &= [\,\{2\}, \quad \{3,1\}, \quad \{6,4,8\}, \quad \{9\}, \quad \{5,7\}\,] \\ &= \Big[\,②, \quad ①\!-\!③, \quad ④\!-\!⑥\!-\!⑧, \quad ⑨, \quad ⑤\!-\!⑦\,\Big] \end{aligned}$$

corresponding to the collection of preimages of $1, 2, 3, 4, 5$.

FIGURE 3. The decomposition of surjections as sequences-of-sets.

The general formula follows similarly from expanding the rth power in (8) by the binomial theorem, and then extracting coefficients:

$$\begin{aligned} R_n^{(r)} &= n!\,[z^n] \sum_{j=0}^{r} \binom{r}{j} (-1)^j e^{r-j} z \\ &= \sum_{j=0}^{r} \binom{r}{j} (-1)^j (r-j)^n. \end{aligned} \tag{9}$$

$\triangleright$ **1.** *A direct derivation of the surjection EGF.* One may verify the result provided by the symbolic method by returning to first principles. Since each preimage of a surjection is a nonempty set, the number of r-surjections is expressed by an r-fold convolution,

$$R_n^{(r)} = \sum_{(n_1, n_2, \ldots, n_r)} \binom{n}{n_1, n_2, \ldots, n_r}, \tag{10}$$

the sum being taken over $n_j \geq 1,\ n_1 + n_2 + \cdots + n_r = n$. (In this formula the indices n_j vary over all allowable cardinalities of preimages, and the multinomial coefficient counts the number of ways of distributing the elements of $[1\,..\,n]$ amongst the r preimages.) Introduce the numbers V_n by $V_0 = 0$ and $V_n = 1$ if $n \geq 1$. The formula (10) then assumes the simpler form

$$R_n^{(r)} \equiv \sum_{n_1, n_2, \ldots, n_r} \binom{n}{n_1, n_2, \ldots, n_r} V_{n_1} V_{n_2} \cdots V_{n_r}, \tag{11}$$

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 (11) leads to (8). $\triangleleft$

Let $\mathcal{S}_n^{(r)}$ denote the number of ways of partitioning the set $[1\,..\,n]$ into r disjoint and nonempty equivalence classes. 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, each of which is a non-empty urn. Thus, one has

$$\mathcal{S}^{(r)} = \mathfrak{P}_r\{\mathcal{V}\}, \qquad \mathcal{V} = \mathfrak{P}_{\geq 1}\{\mathcal{Z}\}.$$

The basic formula connecting the two counting sequences results from there (or from direct reasoning):

(12) $$S_n^{(r)} = \frac{1}{r!} R_n^{(r)} \qquad \text{and} \qquad S^{(r)}(z) = \frac{1}{r!}\left(e^z - 1\right)^r.$$

The rationale for (12) is that 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 (Example 7 and Figure 9 of Chapter I). Knuth, following Karamata, advocated for the $S_n^{(r)}$ the notation $\left\{ {n \atop r} \right\}$. From (9), an explicit form also exists:

(13) $$S_n^{(r)} \equiv \left\{ {n \atop r} \right\} = \frac{1}{r!} \sum_{j=0}^{r} \binom{r}{j} (-1)^j (r-j)^n.$$

The books by Graham, Knuth, and Patashnik [**71**] and Comtet [**28**] contain a thorough discussion of these numbers; see also APPENDIX: *Stirling numbers*, p. 173.

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

$$\mathcal{R} = \mathfrak{S}\{\mathcal{V}\}, \qquad \mathcal{S} = \mathfrak{P}\{\mathcal{V}\}, \qquad \text{with} \quad \mathcal{V} = \mathfrak{P}_{\geq 1}\{\mathcal{Z}\}.$$

From there one finds

(14) $$R(z) = \frac{1}{2 - e^z}, \qquad S(z) = e^{e^z - 1},$$

since $V(z) = e^z - 1$ and $R(z) = (1 - V(z))^{-1}$, $S(z) = e^{V(z)}$. The numbers $R_n = n!\,[z^n]R(z)$ and $S_n = n![z^n]S(z)$ are called the *surjection numbers* (also, "preferential arrangements" numbers, *EIS* **A000670**) and the *Bell numbers* (*EIS* **A000110**) respectively. These numbers are well determined by expanding the EGFs:

$$\begin{aligned} R(z) &= 1 + z + 3\frac{z^2}{2!} + 13\frac{z^3}{3!} + 75\frac{z^4}{4!} + 541\frac{z^5}{5!} + 4683\frac{z^6}{6!} + 47293\frac{z^7}{7!} + \cdots \\ S(z) &= 1 + z + 2\frac{z^2}{2!} + 5\frac{z^3}{3!} + 15\frac{z^4}{4!} + 52\frac{z^5}{5!} + 203\frac{z^6}{6!} + 877\frac{z^7}{7!} + \cdots . \end{aligned}$$

Explicit expressions as finite double sums result from summing Stirling numbers,

$$R_n = \sum_{k \geq 0} k! \left\{ {n \atop k} \right\}, \qquad R_n = \sum_{k \geq 0} k! \left\{ {n \atop k} \right\},$$

where each Stirling number is itself a sum given by (13).

Alternatively, single (though infinite) sums result from the expansions

$$\left\{ \begin{aligned} R(z) &= \frac{1}{2}\frac{1}{1 - \frac{1}{2}e^z} \\ &= \sum_{k=0}^{\infty} \frac{1}{2^{k+1}} e^{kz} \end{aligned} \right. \qquad \text{and} \qquad \left\{ \begin{aligned} S(z) &= e^{e^z - 1} = \frac{1}{e} e^{e^z} \\ &= \frac{1}{e} \sum_{k=0}^{\infty} e^{kz}, \end{aligned} \right.$$

$n = 1$, $S_1 = 1$:

$\boxed{\{1\}}$

$n = 2$, $S_2 = 2$:

$\boxed{\{2\},\{1\}}, \boxed{\{1,2\}}$

$n = 3$, $S_3 = 5$:

$\boxed{\{1\},\{2,3\}}, \boxed{\{1,2\},\{3\}}, \boxed{\{1,2,3\}}, \boxed{\{2\},\{1\},\{3\}}, \boxed{\{1,3\},\{2\}}$

$n = 4$, $S_4 = 15$:

$\boxed{\{1,2,4\},\{3\}}, \boxed{\{1,3,4\},\{2\}}, \boxed{\{1\},\{2,4\},\{3\}}, \boxed{\{1,3\},\{2\},\{4\}}, \boxed{\{1,2,3\},\{4\}}, \boxed{\{4\},\{1,2\},\{3\}}, \boxed{\{3,4\},\{1,2\}},$
$\boxed{\{1,2,3,4\}}, \boxed{\{1\},\{2,3,4\}}, \boxed{\{2,3\},\{1,4\}}, \boxed{\{2\},\{1,4\},\{3\}}, \boxed{\{3,4\},\{2\},\{1\}}, \boxed{\{2\},\{1\},\{4\},\{3\}}, \boxed{\{1,3\},\{2,4\}},$
$\boxed{\{1\},\{4\},\{2,3\}}$

$n = 5$, $S_5 = 52$:

$\boxed{\{2\},\{1\},\{4\},\{3,5\}}, \boxed{\{1,2,3\},\{5\},\{4\}}, \boxed{\{1,2,4,5\},\{3\}}, \boxed{\{1\},\{2,4\},\{3,5\}}, \boxed{\{5\},\{1\},\{2,4\},\{3\}},$
$\boxed{\{1,5\},\{2,4\},\{3\}}, \boxed{\{1\},\{2,3\},\{4,5\}}, \boxed{\{1,3,5\},\{2,4\}}, \boxed{\{1\},\{4\},\{2,5\},\{3\}}, \boxed{\{2,4,5\},\{1,3\}}, \boxed{\{5\},\{1,2,3,4\}},$
$\boxed{\{1,5\},\{2,3,4\}}, \boxed{\{2\},\{1\},\{3\},\{4,5\}}, \boxed{\{5\},\{2\},\{1,4\},\{3\}}, \boxed{\{2,3,4,5\},\{1\}}, \boxed{\{1,3,5\},\{2\},\{4\}}, \boxed{\{2,4,5\},\{1\},\{3\}},$
$\boxed{\{1,2,3,5\},\{4\}}, \boxed{\{2,3,5\},\{1,4\}}, \boxed{\{1,3\},\{5\},\{2\},\{4\}}, \boxed{\{4\},\{1,2\},\{3,5\}}, \boxed{\{1,2,3\},\{4,5\}}, \boxed{\{1,2\},\{3\},\{4,5\}},$
$\boxed{\{1,5\},\{2\},\{4\},\{3\}}, \boxed{\{1,3,4\},\{2,5\}}, \boxed{\{3,4\},\{5\},\{2\},\{1\}}, \boxed{\{1,3\},\{5\},\{2,4\}}, \boxed{\{2\},\{1,3,4,5\}}, \boxed{\{1,3\},\{2\},\{4,5\}},$
$\boxed{\{5\},\{1\},\{4\},\{2,3\}}, \boxed{\{5\},\{4\},\{1,2\},\{3\}}, \boxed{\{3,4,5\},\{1,2\}}, \boxed{\{5\},\{2,3\},\{1,4\}}, \boxed{\{1,2,3,4,5\}}, \boxed{\{1,3,4\},\{5\},\{2\}},$
$\boxed{\{2\},\{1,4,5\},\{3\}}, \boxed{\{3,4\},\{1\},\{2,5\}}, \boxed{\{5\},\{1\},\{2,3,4\}}, \boxed{\{4\},\{1,2,5\},\{3\}}, \boxed{\{1,2,4\},\{3,5\}}, \boxed{\{3,4\},\{1,2,5\}},$
$\boxed{\{1,4\},\{2,5\},\{3\}}, \boxed{\{3,4\},\{5\},\{1,2\}}, \boxed{\{5\},\{1,2,4\},\{3\}}, \boxed{\{1,3\},\{4\},\{2,5\}}, \boxed{\{2,3\},\{1,4,5\}}, \boxed{\{2\},\{1\},\{3,4,5\}},$
$\boxed{\{2,3,5\},\{1\},\{4\}}, \boxed{\{5\},\{2\},\{1\},\{4\},\{3\}}, \boxed{\{2\},\{1,4\},\{3,5\}}, \boxed{\{1,5\},\{3,4\},\{2\}}, \boxed{\{1,5\},\{4\},\{2,3\}}$

FIGURE 4. A listing of all set partitions for sizes $n = 1, 2, 3, 4, 5$.

from which coefficient extraction yields

$$R_n \quad = \quad \frac{1}{2} \sum_{k=0}^{\infty} \frac{k^n}{2^k} \qquad \text{and} \qquad S_n \quad = \quad \frac{1}{e} \sum_{k=0}^{\infty} \frac{k^n}{k!}.$$

The formula for the Bell numbers was found by Dobinski in 1877.

The asymptotic analysis of the surjection numbers (R_n) will be performed in a later chapter by means of singularity analysis of the meromorphic function $R(z)$; that of Bell's partition numbers (S_n) is best done by means of the saddle point method. The asymptotic forms found are

$$(15) \qquad R_n \sim \frac{n!}{2} \frac{1}{(\log 2)^{n+1}} \qquad \text{and} \qquad S_n \sim n! \, \frac{e^{e^{r(n)}-1}}{r(n)^{n+1}\sqrt{2\pi \exp(r(n))}},$$

where $r(n)$ is the positive root of the equation $re^r = n$. One has $r(n) \approx \log n - \log\log n$, so that

$$\log S_n = n \left(\log n - \log\log n - 1 + o(1) \right).$$

Elementary derivations of these asymptotic forms are explored in the notes that follow.

▷ **2.** *Laplace's method for sums.* By examining ratios between successive terms in the sum expressing S_n, one determines the index k_0 near which the terms in Dobinski's formula are maximal. The general term of index $k = k_0 \pm h$, after scaling, is then found to be well approximated by the Gaussian function e^{-x^2}. A comparison with the Riemann sum of the Gaussian functions leads to the asymptotic form stated for S_n. This is an instance of the Laplace method for sums that is detailed in

De Bruijn's book [**35**]; see also [**130**]. The asymptotic estimation of R_n can be subjected to a similar treatment (Comtet). $\lhd$

$\rhd$ **3.** *Cauchy's method for generating functions.* An approach different from the one in Ex. 2 bases itself on the fact that $R(z)$ has a singularity at a finite distance. Indeed, the function

$$R(z) - \frac{1}{2}\frac{1}{\log 2 - z}$$

is analytic for $|z| \leq 6$. (The singularity of $R(z)$ at log 2 has been removed and the next poles are at $\log 2 \pm 2i\pi$.) Thus, one has

$$\frac{R_n}{n!} = \frac{1}{2}\left(\frac{1}{(\log 2)^{n+1}} + O(\frac{1}{6^n})\right)$$

by virtue of Cauchy's bounds for coefficients of analytic functions; see Chapter IV for details. $\lhd$

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

$$S^{(A,B)}(z) = \beta(\alpha(z)) \qquad \textit{where} \qquad \alpha(z) = \sum_{a\in A} \frac{z^a}{a!}, \quad \beta(z) = \sum_{b\in B} \frac{z^b}{b!}.$$

PROOF. One has

$$\mathcal{R}^{(A,B)} = \mathfrak{S}_A\{\mathfrak{P}_B\{Z\}\} \qquad \text{and} \qquad \mathcal{S}^{(A,B)} = \mathfrak{P}_A\{\mathfrak{P}_B\{Z\}\},$$

where $\mathfrak{K}_X$ represents a construction with a number of components restricted to the integer set X. $\square$

EXAMPLE 5. *Set partitions with bounded block sizes.* 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

$$S^{\langle \leq b\rangle}(z) = \exp(e_b(z) - 1), \quad S^{\langle > b\rangle}(z) = \exp(e^z - e_b(z)),$$

correspond to partitions with all blocks of size $\leq b$ and all blocks of size $> b$, respectively. $\square$

$\rhd$ **4.** 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 $(2 - z - e^z)^{-1}$. $\lhd$

EXAMPLE 6. *Comtet's square.* An exercise in Comtet's book [**28**, Ex. 13, p. 225] serves beautifully to illustrate the power of symbolic methods. The question is to enumerate set partitions such that a parity constrained is satisfied by the number of blocks and/or the number of elements in each block. Then, the EGF's are tabulated as follows:

Set partitions	Any number of blocks	Odd number of blocks	Even number 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

$$\left\{\begin{array}{c}\exp\\ \sinh\\ \cosh\end{array}\right\} \circ \left\{\begin{array}{c}\exp -1\\ \sinh\\ \cosh -1\end{array}\right\},$$

in accordance with general principles. ☐

II. 3.2. Applications to words and random allocations. The examples discussed now deal with enumerative problems that present themselves when analysing statistics on letters in words. They find applications in random allocations and the so-called "hashing algorithms" of computer science [**130**]. Fix an alphabet

$$\mathcal{X} = \{a_1, a_2, \ldots, a_r\}$$

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, $w \in \mathcal{W}_n$, can be viewed as 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 $a_1 = a, \ldots, a_5 = r$; for the word $w =$ '*abracadbra*', 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:

$$\overbrace{\{1,4,6,8,11\}}^{a=a_1}, \quad \overbrace{\{2,9\}}^{b=a_2}, \quad \overbrace{\{5\}}^{c=a_3}, \quad \overbrace{\{7\}}^{d=a_4}, \quad \overbrace{\{3,10\}}^{r=a_5}.$$

(Here, all preimages are nonempty, but this need not always the case.) The decomposition based on preimages then gives

$$\mathcal{W} \simeq \mathcal{U}^r \equiv \mathfrak{S}_r\{\mathcal{U}\}, \tag{16}$$

where $\mathcal{U}$ represents a possibly empty urn. As the EGF of $\mathcal{U}$ is $U(z) = e^z$, this construction implies that the EGF of all words is

$$W(z) = (e^z)^r = e^{rz}, \tag{17}$$

which yields back $W_n = r^n$, as was to be expected. For the situation where restrictions are imposed on the number of occurrences of letters, the decomposition (16) 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*

$$W^{(A)}(z) = (\alpha(z))^r \qquad \text{where} \qquad \alpha(z) = \sum_{a \in A} \frac{z^a}{a!}. \tag{18}$$

Though this result is technically a shallow consequence of the symbolic method, it has several important applications in discrete probability; see [**130**, Ch. 8] for a discussion along the lines of the symbolic method.

EXAMPLE 7. *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

$$\mathcal{W}^{\langle \le b \rangle}(z) = (e_b(z))^r, \qquad \mathcal{W}^{\langle > b \rangle}(z) = (e^z - e_b(z))^r, \tag{19}$$

respectively. Taking $b = 1$ in the first formula gives the number of r–arrangements of n elements (i.e., ordered combinations of r elements amongst n) as

$$n!\,[z^n](1+z)^r = r!\binom{n}{r} = n(n-1)\cdots(n-r+1), \tag{20}$$

as anticipated; taking $b = 1$, but now in the second formula, gives back the number of r-surjections.

For general b, the generating functions of (19) contain valuable information on the least frequent and most frequent letter in random words. Some consequences are explored below. □

▷ **5.** *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.) ◁

▷ **6.** *Arrangements.* Define an *arrangement* of size n as an ordered combination of (some) elements of $[1\,..\,n]$, and 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

$$\mathcal{A} \simeq \mathcal{U} \star \mathcal{P},\ \mathcal{U} = \mathfrak{P}\{\mathcal{Z}\},\ \mathcal{P} = \mathfrak{S}\{\mathcal{Z}\} \qquad \Longrightarrow \qquad A(z) = \frac{e^z}{1-z}.$$

The resulting counting sequence

$$A_n = \sum_{k=0}^{n} \frac{n!}{k!}$$

starts as $1, 2, 5, 16, 65, 326, 1957, 13700$ (*EIS* **A000522**); see also [**28**, p. 75]. ◁

▷ **7.** *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 EGF of the number of moves of duration $2n$ 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: this the EGF enumerates mappings where each preimage has an even cardinality.] From there, one can generalize to the case where A contains k balls initially and ℓ balls finally. (This is Ehrenfest's simplified model of heat transfer that is analysed thoroughly in [**69**] by combinatorial methods.) ◁

EXAMPLE 8. *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, the probabilistic model[1] has

$$\begin{aligned} \mathbb{P}\{\text{Max} \le b\} &= n!\,[z^n] e_b\left(\tfrac{z}{m}\right)^m \\ \mathbb{P}\{\text{Max} > b\} &= n!\,[z^n]\left(e^{z/m} - e_b\left(\tfrac{z}{m}\right)\right)^m . \end{aligned} \tag{21}$$

The justification of this formula relies on the easy identity

$$\frac{1}{m^n}\,[z^n]f(z) \equiv [z^n]f(\frac{z}{m}), \tag{22}$$

and on the fact that a probability is determined as the ratio between the number of favorable cases (given by (19) and the total number of cases (m^n).

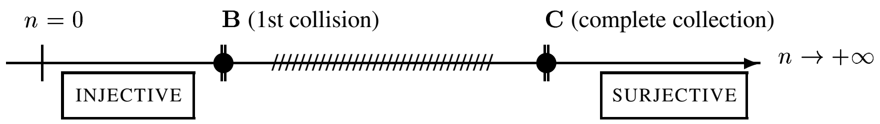
An especially interesting case is when m and n are asymptotically proportional, that is, $n/m = \alpha$ and α lies in a compact subinterval of $(0, +\infty)$. In that case, with probability tending to 1 as n tends to infinity, one has

$$\text{Min} = 0, \qquad \text{Max} \sim \frac{\log n}{\log\log n}.$$

In other words, there are almost surely empty urns (in fact many of them, see Ex. 8 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 VI) based on exact generating representations like (19) and (21). They form the core of the reference book [**92**] by Kolchin, Sevastyanov, and Chistyakov. The resulting estimates are in turn invaluable in the analysis of hashing algorithms [**67, 86, 130**] to which the balls-in-bins model has been recognized to apply with great accuracy [**100**]. □

The next two examples illustrate applications of EGF's to two classical problems of probability theory, the "birthday paradox" and the "coupon collector problem". Assume 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:

[1]We let $\mathbb{P}(E)$ represent the probability of an event E and $\mathbb{E}(X)$ the expectation of the random variable X. Whenever necessary, subscripts may be used to indicate the probabilistic model of use.



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 9. *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 $\beta_1, \ldots, \beta_n$ has no repetition. In other words, the function β from $[1\,..\,n]$ to $\mathcal{X}$ must be injective; equivalently, $\beta_1, \ldots, \beta_n$ is an n-arrangement of r objects. Thus, we have the fundamental relation

$$\begin{aligned}
\mathbb{P}\{B > n\} &= \frac{r(r-1)\cdots(r-n+1)}{r^n} \\
&= \frac{n!}{r^n}[z^n](1+z)^r \\
&= n!\,[z^n]\left(1+\frac{z}{r}\right)^r,
\end{aligned} \tag{23}$$

where the second line repeats (20) and the third results from the series transformation (22).

The expectation of the random variable B is

$$\mathbb{E}(B) = \sum_{n=0}^{\infty} \mathbb{P}\{B > n\}, \tag{24}$$

by virtue of a general formula valid for all discrete random variables. From (23), 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 (23), line 3. Let f be an entire function with nonnegative coefficients. Then the formula

$$f(z) = \sum_{n\geq 0} f_n z^n \quad\Longrightarrow\quad S := \sum_{n=0}^{\infty} f_n n! = \int_0^{\infty} e^{-t} f(t)\,dt, \tag{25}$$

is valid provided either the sum or the integral on the right converges. The reason is the usual Eulerian representation of factorials,

$$n! = \int_0^{\infty} e^{-t} t^n\,dt.$$

Applying this principle to (24) with the probabilities given by (23) (third line), one finds

$$\mathbb{E}(B) = \int_0^\infty e^{-t}\left(1+\frac{t}{r}\right)^r dt. \tag{26}$$

This last form is easily amenable to asymptotic analysis and the Laplace method[2] delivers the estimation

$$\mathbb{E}(B) = \sqrt{\frac{\pi r}{2}} + \frac{2}{3} + O(r^{-1/2}), \tag{27}$$

as r tends to infinity. For instance, the asymptotic approximation provided by the first two terms of (27) is $24.61\mathit{119}$, 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

$$I(r,b) := \int_0^\infty e^{-t} e_b\left(\frac{t}{r}\right)^r dt. \tag{28}$$

(The basic birthday paradox corresponds to $b=2$.) The formula (28) was first derived by Klamkin and Newman in 1967; their paper [**80**] 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. □

▷ **8.** *Birthday paradox with leap years.* Assume that the 29th of February exists precisely once every fourth year. What is the amplitude of the effect on the expectation of the first birthday collision? (Hint: one may wish to treat the general case of nonuniform date distributions;see Ex. 10 below.) ◁

EXAMPLE 10. *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". (The name "coupon collector" is due to the fact that in former times, chocolate bars would contain different coupons or images and collectors would be awarded some gift in exchange for a full collection.) In other words, the event $\{C\le n\}$ means the equality between sets, $\{\beta_1,\ldots,\beta_n\}=\mathcal{X}$. Thus, the probabilities satisfy

$$\begin{aligned}\mathbb{P}\{C\le n\} &= \frac{R_n^{(r)}}{r^n} = \frac{n!\left\{ {r \atop n} \right\}}{r^n}\\ &= \frac{n!}{r^n}[z^n](e^z-1)^r\\ &= n![z^n]\left(e^{z/r}-1\right)^r,\end{aligned} \tag{29}$$

by our earlier enumeration of surjections. The complementary probabilities are then

$$\mathbb{P}\{C>n\} = 1-\mathbb{P}\{C\le n\} = n![z^n]\left(e^z - \left(e^{z/r}-1\right)^r\right).$$

[2] Knuth [**85**, 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 $\mathbb{E}(B)=1+Q(r)$.

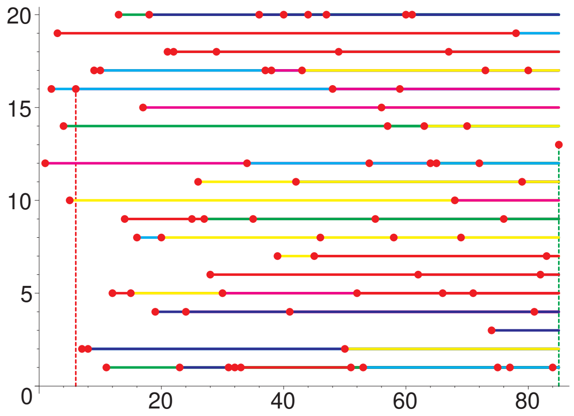


FIGURE 5. A sample realization of the "birthday paradox" and "coupon collection" with an alphabet of $r = 20$ letters. The first collision occurs at time $B = 6$ while the collection becomes complete at time $C = 87$.

An application of the Eulerian integral trick of (26) then provides a representation of the expectation of the time needed for a full collection as

$$\mathbb{E}(C) = \int_0^\infty \left(1 - (1 - e^{-t/r})^r\right) dt. \tag{30}$$

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 (30), perform the change of variables $v = 1 - e^{-t/r}$, then expand and integrate termwise; this process provides the more tractable form

$$\mathbb{E}(C) = r\,\mathrm{H}_r, \tag{31}$$

where H_r is the harmonic number:

$$\mathrm{H}_r = 1 + \frac{1}{2} + \frac{1}{3} + \cdots + \frac{1}{r}.$$

(Formula (31) is by the way easy to interpret directly: one needs on average $1 = r/r$ trials to get the first day, then $r/(r-1)$ to get a different day, etc.)

Regarding (31), one has available the well-known formula (by comparing sums with integrals or by Euler-Maclaurin summation),

$$\mathrm{H}_r = \log r + \gamma + \frac{1}{2r} + O(r^{-2}), \quad \gamma \doteq 0.57721\,56649,$$

where γ is known as Euler's constant. Thus, the expected time for a full collection satisfies

$$\mathbb{E}(C) = r \log r + \gamma r + \frac{1}{2} + O(r^{-1}). \tag{32}$$

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 (32) yields 2364.64625, representing a relative error of only one in ten millions.

Like before, 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

$$J(r,b) = \int_0^\infty \left(1 - \left(1 - e_b(t/r)e^{-t/r}\right)^r\right)\, dt,$$

an expression that vastly generalizes (32). From there, one finds [**108**]

$$J(r,b) = n\left(\log n + (b-1)\log\log n + \gamma - \log(b-1)! + o(1)\right),$$

so that only a few more trials are needed in order to obtain additional collections. □

▷ **9.** *The little sister.* The coupon collector has a little sister to whom he gives his duplicates. Foata, Lass, and Han [**63**] show that the little sister misses on average H_r coupons when her big brother first obtains a complete collection. ◁

▷ **10.** *The original coupon collector problem.* A company issues coupons of r different types, type j being issued with probability p_j. Let C be the random variable representing the number of coupons that one needs to gather until a full collection with r different coupons is obtained. By the multivariate techniques of Chapter III, one has

$$\mathbb{P}\{C \le n\} = n!\,[z^n]\prod_{j=1}^{r}(e^{p_j z} - 1).$$

The Eulerian integral gives for the expectation:

$$\mathbb{E}(C) = \int_0^\infty \left(1 - \prod_{j=1}^{r}(1 - e^{-p_i x})\right)\, dx.$$

See [**47**] for several variations on this theme and p. 141 for related context. ◁

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

$$\widehat{W}(z) \equiv \sum_n W_n z^n = \frac{1}{1-rz} \qquad \text{and} \qquad W(z) \equiv \sum_n W_n \frac{z^n}{n!} = e^{rz},$$

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,

$$\widehat{S}^{(r)}(z) = \frac{z^r}{(1-z)(1-2z)\cdots(1-rz)} \qquad \text{and} \qquad S^{(r)}(z) = \frac{(e^z-1)^r}{r!},$$

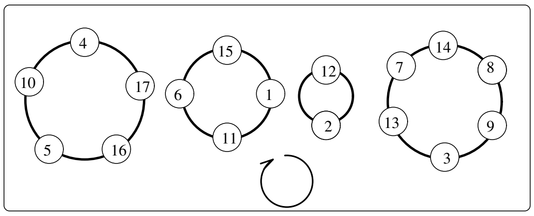
by viewing these either as unlabelled structures (an encoding via words of a regular language, see Section I.4.3) or directly as labelled structures.

II. 4. Alignments, permutations, and related structures

In this section, we start by considering the specifications,

$$\mathcal{O} = \mathfrak{S}\{\mathfrak{C}\{\mathcal{Z}\}\}, \qquad \text{and} \qquad \mathcal{P} = \mathfrak{P}\{\{\mathcal{Z}\}\}, \tag{33}$$

built by piling up two constructions, sequences-of-cycles and sets-of-cycles respectively. They define a new class of objects, called alignments ($\mathcal{O}$), while serving to specify permutations ($\mathcal{P}$) in a novel way as detailed below. (These specifications otherwise parallel surjections and set partitions.) Permutations are in this context examined under their cycle



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 clockwise and i is connected to σ_i in the graph.)

FIGURE 6. The cycle decomposition of permutations.

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. Define first an alignment as a well-labelled sequence of cycles and let $\mathcal{O}$ be the class of all alignments. Let $\mathcal{P}$ be defined momentarily as the class of all sets of cycles. The corresponding specifications are then clearly the ones of (33).

By the symbolic method, alignments have EGF

$$\begin{aligned} O(z) &= \frac{1}{1-\log(1-z)^{-1}} \\ &= 1 + z + 3\frac{z^2}{21} + 14\frac{z^3}{3!} + 88\frac{z^4}{4!} + 694\frac{z^5}{5!} + \cdots, \end{aligned}$$

which does not simplify. The coefficients form *EIS* **A007840** ("ordered factorizations of permutations into cycles").

From elementary mathematics, it is known that a permutation admits a unique decomposition into cycles. Let $\sigma = \sigma_1 \ldots \sigma_n$ be as 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 (33)) is isomorphic to the class of permutations as defined in Section II. 1:

$$\mathcal{P} = \mathfrak{P}\{\mathfrak{C}\{\mathcal{Z}\}\} \cong \mathfrak{S}\{\mathcal{Z}\}.$$

This combinatorial isomorphism is reflected by the obvious series identity

$$P(z) = \exp\left(\log\frac{1}{1-z}\right) = \frac{1}{1-z}.$$

In a sense, 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 11. *Stirling cycle numbers.* The number of permutations of size n comprised of r cycles is determined by the explicit generating function

$$P_n^{(r)} = \frac{n!}{r!} [z^n] \left(\log \frac{1}{1-z} \right)^r. \tag{34}$$

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 **[71]** where they are denoted by $\left[{n \atop r} \right]$. See APPENDIX: *Stirling numbers*, p. 173. (Note that the number of alignments formed with r cycles is $r! \left[{n \atop r} \right]$.) As we shall see shortly (p. 99) 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 put on 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 favorables 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 "reasonable" values of n from (34), preferably 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

so that, for this value of n, 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: *On average, a random permutation of size 100 has a little more than 5 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

$$\sum_{k=0}^{n} \begin{bmatrix} n \\ k \end{bmatrix} u^k = u(u+1)(u+2)\cdots(u+n-1) \tag{35}$$

describes precisely the distribution of Stirling cycle numbers for any fixed value of n. From there, the expected number of cycles, $\mu_n := \sum_k k p_{n,k}$ is easily found (use logarithmic differentiation of (35)),

$$\mu_n = \mathrm{H}_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}.$$

In particular, one has $\mu_{100} \equiv \mathrm{H}_{100} \doteq 5.18738$. In general: *The mean number of cycles in a random permutation of size n grows logarithmically with n, $\mu_n \sim \log n$.* □

EXAMPLE 12. *Involutions and permutations without long cycles.* A permutation σ is an *involution* if $\sigma^2 = Id$ with Id the identity permutation. Quite clearly, an involution can have only cycles of sizes 1 and 2. The class $\mathcal{I}$ of all involutions thus satisfies $\mathcal{I} = \mathfrak{P}\{\mathfrak{C}_{1,2}\{\mathcal{Z}\}\}$. The translation is immediate:

$$I(z) \equiv \sum_n I_n \frac{z^n}{n!} = \exp\left(z + \frac{z^2}{2}\right). \tag{36}$$

This last equation then provides the formula

$$I_n = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{(n-2k)!2^k k!},$$

which solves the counting problem explicitly. A *pairing* is an involution without fixed point; in other words, only cycles of length 2 are allowed. The EGF and the number of all pairings are given by

$$J(z) = e^{z^2/2}, \qquad J_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1)$$

as was to be anticipated from 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. □

All perms	Derangements	Involutions	Pairings
$\frac{1}{1-z}$	$\frac{e^{-z}}{1-z}$	$e^{z+z^2/2}$	$e^{z^2/2}$
Shortest cycle $> r$		Longest cycle $\leq r$	
$\frac{e^{-\ell_r(z)}}{1-z}$		$e^{\ell_r(z)}$	

FIGURE 7. A summary of major EGFs related to permutations. There, $\ell_r(z) := \sum_{j=1}^{r} \frac{z^j}{j}$ is the "truncated logarithm".

$\triangleright$ **11.** *Permutations such that $\sigma^e = Id$.* Such permutations are "roots of unity" in the symmetric group. Their EGF is

$$\exp\left(\sum_{d \mid e} \frac{z^d}{d}\right),$$

where the sum extends to all divisors d of e. $\triangleleft$

EXAMPLE 13. *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

$$\mathcal{D}^{(r)} = \mathfrak{P}\{\mathfrak{C}_{>r}\{\mathcal{Z}\}\}, \tag{37}$$

the corresponding EGF being then

$$D^{(r)}(z) = \exp\left(\sum_{j>r} \frac{z^j}{j}\right) = \frac{\exp(-\sum_{j=1}^{r} \frac{z^j}{j})}{1-z}. \tag{38}$$

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 [**28**]: "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 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 [**107**, 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

$$\frac{D_n^{(r)}}{n!} \sim e^{-\mathrm{H}_r}, \tag{39}$$

(for any fixed r), as can be proved easily by complex asymptotic methods (Chapter IV). $\square$

Like several other structures that we have been considering previously, permutation allow for transparent connections between structural constraints and the shapes of generating functions. The major counting results encountered in this section are summarized in Figure 7.

$\triangleright$ **12.** *Parity constraints in permutations.* The EGF's 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 EGF's of permutations having an even number of cycles ($E^*(z)$) and an odd number of cycles ($O^*(z)$) are

$$E^*(z) = \cosh(\log\frac{1}{1-z}) = \frac{1}{2}\,\frac{2-z^2}{1-z}, \qquad O^*(z) = \sinh(\log\frac{1}{1-z}) = \frac{1}{2}\,\frac{z^2}{1-z}.$$

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.) $\triangleleft$

II. 4.2. Second level structures. Consider the three basic constructors of labelled sequence ($\mathfrak{S}$), set ($\mathfrak{P}$), and cycle ($\mathfrak{C}$). 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 9 possibilities summarized by the following table:

ext.\int.	$\mathfrak{S}_{\geq 1}$	$\mathfrak{P}_{\geq 1}$	$\mathfrak{C}$
$\mathfrak{S}$	*"Labelled compositions" ($\mathcal{L}$)* $\mathfrak{S}\circ\mathfrak{S}$ $\dfrac{1-z}{1-2z}$	Surjections ($\mathcal{R}$) $\mathfrak{S}\circ\mathfrak{P}$ $\dfrac{1}{2-e^z}$	Alignments ($\mathcal{O}$) $\mathfrak{S}\circ\mathfrak{C}$ $\dfrac{1}{1-\log(1-z)^{-1}}$
$\mathfrak{P}$	*"Fragmented permutations" ($\mathcal{F}$)* $\mathfrak{P}\circ\mathfrak{S}$ $e^{z/(1-z)}$	Set partitions ($\mathcal{S}$) $\mathfrak{P}\circ\mathfrak{P}$ e^{e^z-1}	Permutations ($\mathcal{P}$) $\mathfrak{P}\circ\mathfrak{C}$ $\dfrac{1}{1-z}$
$\mathfrak{C}$	*"Supernecklaces[1]"* $\mathfrak{C}\circ\mathfrak{S}$ $\log\dfrac{1-z}{1-2z}$	*"Supernecklaces[2]"* $\mathfrak{C}\circ\mathfrak{P}$ $\log(1-e^z)^{-1}$	*"Supernecklaces[3]"* $\mathfrak{C}\circ\mathfrak{C}$ $\log\dfrac{1}{1-\log(1-z)^{-1}}$

The classes of surjections, alignments, set partitions, and permutations appear naturally as $\mathfrak{S}\circ\mathfrak{P}$, $\mathfrak{S}\circ\mathfrak{C}$, $\mathfrak{P}\circ\mathfrak{P}$, and $\mathfrak{P}\circ\mathfrak{C}$. The other ones represent essentially nonclassical objects. The case of $\mathcal{L}$ corresponding to $\mathfrak{S}\circ\mathfrak{S}$ describes a class whose elements are (ordered) sequences of linear graphs. This can be interpreted as permutations with separators inserted, e.g, $53|264|5$, or alternatively as integer compositions with a labelling superimposed. Finally, the class $\mathcal{F} = \mathfrak{P}\{\mathfrak{S}_{\geq 1}\{\mathcal{Z}\}\}$ 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

$$F(z) = e^{z/(1-z)} = 1 + z + 3\frac{z^2}{2!} + 13\frac{z^3}{3} + 73\frac{z^4}{4!} + \cdots,$$

whose coefficients constitute *EIS* **A000262** ("sets of lists"). What we termed "supernecklaces" in the last column 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. . .

$\triangleright$ **13.** *Counting specifications of level n.* The algebra of constructions satisfies the combinatorial isomorphism $\mathfrak{P}\{\mathfrak{C}\{\mathcal{X}\}\} \cong \mathfrak{S}\{X\}$ for all $\mathcal{X}$. How many different terms of length n can be built from three symbols $\mathfrak{C}, \mathfrak{P}, \mathfrak{S}$ satisfying a semi-group law ('$\circ$') together with the relation $\mathfrak{P} \circ \mathfrak{C} = \mathfrak{S}$? This determines the number of specifications of level n. (Hint: the OGF is rational as it corresponds to words with an excluded pattern.) $\triangleleft$

II. 5. Labelled trees, mappings, and graphs

In this section, we consider labelled trees and certain labelled objects 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.

II. 5.1. Trees. The trees to be studied here are rooted and labelled, meaning as usual that a node is distinguished as the root and that nodes bear distinct integer labels. 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 $\Omega \ni 0$.

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

$$\mathcal{A} = \mathcal{Z} \star \mathfrak{S}_{\Omega}\{\mathcal{A}\},$$

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 subtrees

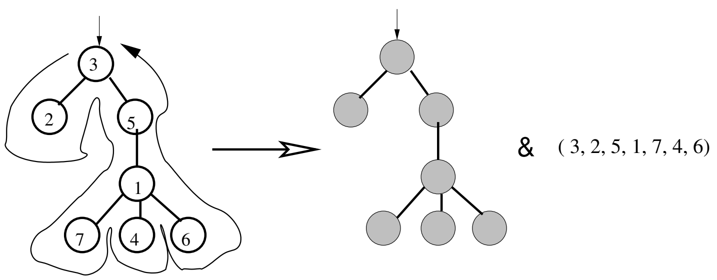


FIGURE 8. A labelled plane tree is determined by an unlabelled tree (the "shape") and a permutation of the labels $1, \ldots, n$.

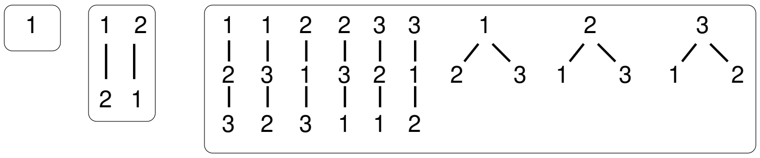


FIGURE 9. There are $T_1 = 1, T_2 = 2, T_3 = 9$, and in general $T_n = n^{n-1}$ Cayley trees of size n.

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.4). 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 unrooted 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**.

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 the symbolic equation

$$\mathcal{T} = \mathcal{Z} \star \mathfrak{P}\{\mathcal{T}\}, \tag{40}$$

where the set construction translates the fact that subtees stemming from the root are not ordered between themselves. From the specification (40), the EGF $T(z)$ is defined implicitly by the "functional equation"

$$T(z) = ze^{T(z)}. \tag{41}$$

The first few values are easily found:

$$T(z) = z + 2^1\frac{z^2}{2!} + 3^2\frac{z^3}{3!} + 4^3\frac{z^4}{4!} + 5^4\frac{z^5}{5!} + \cdots.$$

This leads to expect that

$$T_n = n^{n-1} \tag{42}$$

a fact proved (once more) by the Lagrange Inversion Theorem (see APPENDIX: *Lagrange Inversion*, p. 170):

$$\frac{T_n}{n!} = [z^n]T(z) = \frac{1}{n}[u^{n-1}](e^z)^n = \frac{n^{n-1}}{n!}.$$

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 (42) given by Cayley in 1889 is often referred to as "Cayley's formula" and unrestricted non-plane labelled tees are often called "Cayley trees". See [**17**, p. 51] for a historical discussion. The simplicity of Cayley's formula calls for a combinatorial explanation. The most famous one due to Prüfer (in 1918); see [**17**, p. 53] or [**105**, p. 5] for a description of the Prüfer encoding of trees by sequences. The function $T(z)$ is also known as the (Cayley) "tree function"; it is a close relative of the W–function [**29**] 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 trees where all (out)degrees of nodes are restricted to lie in a set Ω. This corresponds to the specification

$$\mathcal{T}^{(\Omega)} = \mathcal{Z} \star \mathfrak{P}_\Omega\{\mathcal{T}^{(\Omega)}\},$$

which translates directly into an EGF equation,

$$T^{(\Omega)}(z) = z\overline{\phi}(T^{(\Omega)}(z)) \qquad \text{where} \quad \overline{\phi}(u) = \sum_{\omega\in\Omega} \frac{u^\omega}{\omega!},$$

and is amenable to Lagrange inversion. What this last formula involves is the "exponential characteristic" of the degree sequence (as opposed to the ordinary characteristic, in the planar case). In summary:

PROPOSITION II.5. *The number of trees, where all nodes have their outdegree in Ω, is*

$$T_n^{(\Omega)} = (n-1)![u^{n-1}](\overline{\phi}(u))^n \qquad \textit{where} \quad \overline{\phi}(u) = \sum_{\omega\in\Omega} \frac{u^\omega}{\omega!}.$$

▷ **14.** *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. ◁

▷ **15.** *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

$$\mathcal{L} = \mathcal{Z} + \mathfrak{P}_{\geq 2}\{\mathcal{L}\},$$

so that $L(z)$ satisfies $L = z + e^L - 1 - L$, and

$$L(z) = T(\frac{1}{2}e^{z/2-1/2}) + \frac{z}{2} - \frac{1}{2} = z + \frac{z^2}{2!} + 4\frac{z^3}{3!} + 26\frac{z^4}{4!} + 236\frac{z^5}{5!} + \cdots$$

(*EIS* **A000311**), with T being the Cayley tree function. The numbers count "phylogenetic trees" (used to describe the evolution of a genetically related group of organisms) and correspond to Schröder's "fourth problem"; see [**28**, 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 corresponds to $\mathcal{M} = \mathcal{Z} + \mathfrak{P}_2\{\mathcal{M}\}$, so that

$$M(z) = 1 - \sqrt{1-2z} \qquad \text{and} \qquad M_n = 1\cdot 3 \cdots (2n-3),$$

where the counting numbers are 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.

Given a mapping (or function) f, upon starting from any point x_0, the succession of (directed) edges in the graph traverses the iterates of the mapping, $x_0, f(x_0), f(f(x_0)), \ldots$, and since the domain is finite, each such sequence must eventually loop on itself. When the operation is repeated, the elements 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:

$$\left\{\begin{array}{lcl} \mathcal{F} & = & \mathfrak{P}\{\mathcal{K}\} \\ \mathcal{K} & = & \mathfrak{C}\{\mathcal{T}\} \\ \mathcal{T} & = & \mathcal{Z} \star \mathfrak{P}\{\mathcal{T}\}. \end{array}\right. \tag{43}$$

This translates at sight into a set of equations for EGF's

$$\left\{\begin{array}{lcl} F(z) & = & e^{K(z)} \\ K(z) & = & \log \dfrac{1}{1-T(z)} \\ T(z) & = & ze^{T(z)}. \end{array}\right. \tag{44}$$

Eventually, the EGF $F(z)$ is found to satisfy $F = (1-T)^{-1}$. It can be checked from there, by Lagrange inversion once more, 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 $\log(1-$

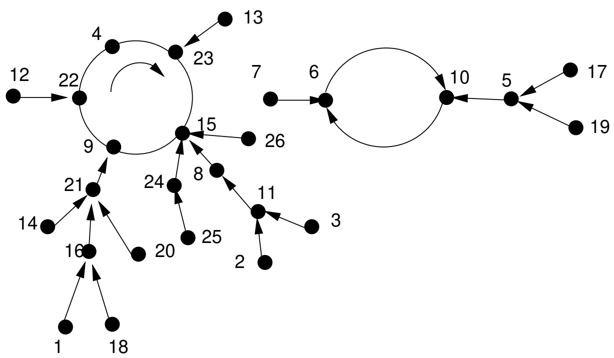


FIGURE 10. A functional graph of size $n = 26$ associated to the mapping φ such that $\varphi(1) = 16$, $\varphi(2) = \varphi(3) = 11$, $\varphi(4) = 23$, and so on.

$T)^{-1}$ and recover coefficients by Bürmann's form):

$$K_n = n^{n-1}Q(n) \quad \text{where} \quad Q(n) := 1 + \frac{n-1}{n} + \frac{(n-1)(n-2)}{n^2} + \cdots. \tag{45}$$

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 (26)). 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 Riemman sum by an integral: this is an instance of the Laplace method for sums (see [**85**, Sec. 1.2.11.3], [**130**, Sec. 4.7] as well as Ex. 2). 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 IV and V. 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.

As is customary with the symbolic method, the constructions (43) also lead to a large number of related counting results. For instance, 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. (39)). 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 Chapter IV.

Next, idempotent mappings satisfying $f(f(x)) = f(x)$ correspond to $\mathcal{I} \cong \mathfrak{P}\{\mathcal{Z} \star \mathfrak{P}\{\mathcal{Z}\}\}$, 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 V.

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 [**86**, p. 371] and [**130**, Sec 8.8]. The algorithm once suitably optimised first led to the factorization of the Fermat number $F_8 = 2^{2^8} + 1$ obtained by Brent in 1980.

$\triangleright$ **16.** *Binary mappings.* The class $\mathcal{B}F$ of binary mappings, where each point has either 0 or 2 preimages, is specified by

$$\mathcal{BF} = \mathfrak{P}\{\mathcal{K}\},\ \mathcal{K} = \mathfrak{C}\{\mathcal{P}\},\ \mathcal{P} = \mathcal{Z} \star \mathcal{B},\ \mathcal{B} = \mathcal{Z} \star \mathfrak{P}_{0,2}\{\mathcal{B}\}$$

All mappings $\dfrac{1}{1-T}$	Partial $\dfrac{e^T}{1-T}$	Injective partial $\dfrac{1}{1-z}e^{z/(1-z)}$	Surjection $\dfrac{1}{2-e^z}$	Bijection $\dfrac{1}{1-z}$
Connected ($\mathcal{K}$) $\log\dfrac{1}{1-T}$	No fixed point $\dfrac{e^{-T}}{1-T}$	Involution $e^{z+z^2/2}$	Idempotent e^{ze^z}	Binary $\dfrac{1}{\sqrt{1-2z^2}}$

FIGURE 11. A summary of various counting EGFs relative to mappings.

(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 [**6, 53**] for a general enumerative theory of random mappings including degree-restricted ones. $\lhd$

$\rhd$ **17.** *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 $\mathcal{PF} = \mathfrak{P}\{\mathcal{T}\} \star \mathcal{F}$, so that

$$PF(z) = \frac{e^{T(z)}}{1-T(z)} \qquad \text{and} \qquad PF_n = (n+1)^n.$$

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, $PFI = \mathfrak{P}\{\mathfrak{C}\{\mathcal{Z}\} + \mathfrak{S}_{\geq 1}\{\mathcal{Z}\}\}$, 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 [**20**].) $\lhd$

The results of this section and the previous ones offer a wide number of counting results relative to maps satisfying various constraints. These are summarized in Figure 11.

II. 5.3. Labelled graphs. Random graphs form a major chapter of the theory of random discrete structures [**19, 78**]. 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 [**51, 77**].)

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 IV and V imply immediately

$$A_n \sim e^{1/2}\, n^{n-2}.$$

Surprisingly, perhaps, there are barely more acyclic graphs than unrooted trees.

The *excess* of a graph is defined as the difference between the number of vertices and the number of nodes. For a connected graph, this is always -1 or more with 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 $\mathfrak{UC}$ representing the (new) undirected cycle construction, one thus has

$$\mathcal{W}_0 \cong \mathfrak{UC}_{\geq 3}\{\mathcal{T}\}.$$

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

$$\mathcal{W}_0 + \mathcal{W}_0 \cong \mathfrak{C}_{\geq 3}\{\mathcal{T}\},$$

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 IV):

$$n![z^n]W_0 \sim \frac{1}{4}\sqrt{2\pi}n^{n-1/2} - \frac{5}{3}n^{n-1} + \frac{1}{48}\sqrt{2\pi}n^{n-3/2} + \cdots . \tag{47}$$

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,

$$n![z^n]e^{W_{-1}+W_0} = \Gamma(3/4)2^{-1/4}e^{-1/2}\pi^{-1/2}n^{n-1/4}\left(1+O(n^{-1/2})\right).$$

Such graphs stand just next to acyclic graphs in order of structural complexity.

▷ **18.** *2-Regular graphs.* This is based on Comtet's account [**28**, 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

$$R(z) = \frac{e^{-z/2-z^2/4}}{\sqrt{1-z}}.$$

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 general enumeration of r–regular graphs becomes somewhat more difficult when $r > 2$. Algebraic aspects are discussed in [**65, 68**] while Bender and Canfield [**9**] 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. ◁

The previous discussion suggests considering more generally the enumeration of connected graphs according to excess. E. M. Wright made important contributions in this area [**154, 155, 156**] that are revisited in the famous "giant paper on the giant component" by Janson, Knuth, Łuczak, and Pittel [**77**]. 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 (not given here, see Wright's original papers or [**77**]) 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:

$$W_1 = \frac{1}{24}\frac{T^4(6-T)}{(1-T)^3}, \qquad W_2 = \frac{1}{2}\frac{T^4(2+28T-23T^2+9T^3-T^4)}{(1-T)^6}.$$

As explained in the giant paper [**77**], 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)[3], 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 $m = \frac{1}{2}n + O(n^{1/3})$, 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.

[3] 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 continuous measures.

In each case, refined estimates follow from a detailed analysis of corresponding generating functions, which is a main theme of [**51**] and especially [**77**]. 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 [**19, 78**] for a probabilistic context and the paper [**10**] 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 [**76**].

II. 6. Additional constructions

Like in the unlabelled case, pointing and substitution are available in the world of labelled structures (Section II. 6.1). Implicit definitions enlarge the scope of the symbolic method (Section II. 6.2) The inversion process needed to enumerate implicit structures are even simpler, since in the labelled universe sets and cycles have 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* 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

$$A_n = n \cdot B_n \implies A(z) = z\frac{d}{dz}A(z).$$

The *composition* or *substitution* can be defined so that it corresponds *a priori* to composition of generating functions. It is formally defined as

$$\mathcal{B} \circ \mathcal{C} = \sum_{k=0}^{\infty} \mathcal{B}_k \times \mathfrak{P}_k\{\mathcal{C}\},$$

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 $\mathcal{B} \circ \mathcal{C}$, is as follows: select some element of $\mathcal{B}$ of some size k, then 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 labelled node of value r in $\mathcal{B}$.

THEOREM II.3. *The combinatorial constructions of pointing and substitution are admissible.*

$$\begin{aligned} \mathcal{A} = \Theta\mathcal{B} \quad &\implies \quad A(z) = z\partial_z A(z), \qquad \partial_z \equiv \frac{d}{dz} \\ \mathcal{A} = \mathcal{B} \circ \mathcal{C} \quad &\implies \quad A(z) = B(C(z)). \end{aligned}$$

For instance, the EGF of (relabelled) pairings of elements drawn from $\mathcal{A}$ is

$$e^{A(z)+A(z)^2/2},$$

since the EGF of involutions is $e^{z+z^2/2}$.

▷ **19.** *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 powerset class of $\mathcal{A}$ may be defined as $\mathcal{U}\circ\mathcal{A}$ where $\mathcal{U}$ is the class of all urns. Thus,

$$\mathfrak{S}\{\mathcal{A}\}\cong\mathcal{P}\circ\mathcal{A},\ \mathfrak{P}\{\mathcal{A}\}\cong\mathcal{U}\circ\mathcal{A}.$$

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" [**79**] and the book by Bergeron, Labelle, and Leroux [**13**] make a great use of such ideas and show that an extensive theory of combinatorial enumeration can be based on such ideas. ◁

▷ **20.** *Distinct component sizes.* The EGF's 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 [**72**, Sec. 4.1.6] for a Tauberian argument. ◁

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 $\mathfrak{S},\mathfrak{P},\mathfrak{C}$, the algebra is equally easy.

THEOREM II.4 (Implicit specifications). *The generating functions associated to the implicit equations in $\mathcal{X}$*

$$\mathcal{A}=\mathfrak{S}\{\mathcal{X}\},\qquad \mathcal{A}=\mathfrak{P}\{\mathcal{X}\},\qquad \mathcal{A}=\mathfrak{C}\{\mathcal{X}\},$$

are respectively

$$X(z)=1-\frac{1}{A(z)},\qquad X(z)=\log A(z),\qquad X(z)=1-e^{-A(z)}.$$

EXAMPLE 14. *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 $\mathcal{K}\subset\mathcal{G}$:

$$\mathcal{G}=\mathfrak{P}\{\mathcal{K}\}\implies G(z)=e^{K(z)}.$$

This basic formula is known in graph theory [**76**] 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,

$$\begin{aligned} K(z) &= \log\left(1+\sum_{n\ge1}2^{\binom{n}{2}}\frac{z^n}{n!}\right)\\ &= z+\frac{z^2}{2!}+4\frac{z^3}{3!}+38\frac{z^4}{4!}+728\frac{z^5}{5!},\end{aligned}$$

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. 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 second term in the asymptotic formula (50). Notice that here good use could be made of a purely divergent generating function for asymptotic enumeration purposes. □

▷ **21.** *Bipartite graphs.* A plane bipartite graph is a pair (G, ω) where G is 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 [**153**, 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. (This assertion can be established rigorously by complex analysis since EGF's of constructible classes must have a nonzero radius of convergence.) In contrast, the special graphs encountered in this chapter, including graphs of fixed excess, are all constructible.

II. 6.3. Order constraints. A construction well suited to taking into account many order properties of combinatorial structures the modified labelled product,

$$\mathcal{A} = (\mathcal{B}^{\Box} \star \mathcal{C}).$$

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.*

$$\mathcal{A} = (\mathcal{B}^{\square} \star \mathcal{C}) \quad \Longrightarrow \quad A(z) = \int_0^z (\partial_t B(t)) \cdot C(t)\, dt, \qquad \partial_t \equiv \frac{d}{dt}. \tag{51}$$

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

$$\frac{A_n}{n!} = \frac{1}{n} \sum_{k=0}^{n} \binom{n}{k} (kB_k)\, C_{n-k},$$

the result follows by taking EGF's. □

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 label values. 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.

▷ **22.** *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."* ◁

EXAMPLE 15. *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 12 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" [**33**, 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

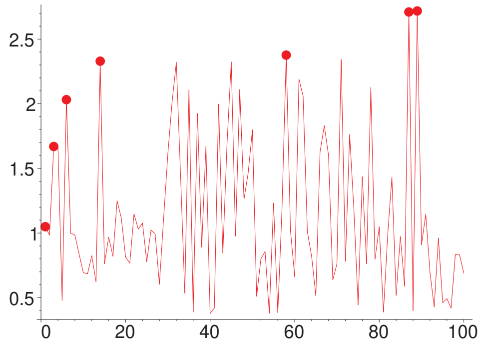


FIGURE 12. A numerical sequence of size 100 with records marked by circles: there are 7 records that occur at times $1, 3, 5, 11, 60, 86, 88$.

of records obeys the same laws as in a random permutation of $[1\,..\,n]$. 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

$$\mathcal{P}^{(k)} = \mathfrak{P}_k\{\mathcal{Q}\}.$$

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[4] is easily revertible. Here is an illustration, with leaders underlined:

$$\begin{aligned}\{(\underline{\mathbf{7}},2,6,1),\ (\underline{\mathbf{4}},3),\ (\underline{\mathbf{9}},8,5)\} &\cong [(\underline{\mathbf{4}},3),\ (\underline{\mathbf{7}},2,6,1),\ (\underline{\mathbf{9}},8,5))] \\ &\cong \underline{\mathbf{4}},3,\underline{\mathbf{7}},2,6,1,\underline{\mathbf{9}},8,5.\end{aligned}$$

Thus, the number of permutations with k records is determined by

$$P^{(k)}(z) = \frac{1}{k!}\left(\log\frac{1}{1-z}\right)^k, \qquad P_n^{(k)} = \left[\begin{matrix}n\\k\end{matrix}\right],$$

where we recognize Stirling cycle numbers from Example 11. In other words:

> *The number of permutations of size n having k records is counted by the Stirling "cycle" number $\left[\begin{smallmatrix}n\\k\end{smallmatrix}\right]$.*

[4]This correspondence is easily extended to a transformation on permutations that maps the number of records to the number of cycles.In this case, it is known as Foata's fundamental correspondence [**98**, Sec. 10.2].

Returning to our statistical problem, the treatment of Example 84 (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 12 to have been generated independently at random (and indeed they were). $\square$

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

$$\begin{array}{llll} \mathcal{F} = \mathfrak{S}\{\mathcal{G}\} & \Longrightarrow & f(z) = \dfrac{1}{1-g(z)}, & f = 1 + gf \\ \mathcal{F} = \mathfrak{P}\{\mathcal{G}\} & \Longrightarrow & f(z) = e^{g(z)}, & f = \displaystyle\int g' f \\ \mathcal{F} = \mathfrak{C}\{\mathcal{G}\} & \Longrightarrow & f(z) = \log \dfrac{1}{1-g(z)}, & f = \displaystyle\int g' \dfrac{1}{1-g} \end{array}$$

The last column is easily checked to provide an alternative form of the standard operator corresponding to sequences, powersets, and cycles. Each case is then itself deduced directly from Theorem II.5 and the labelled product rule:

Sequences: they obey the recursive definition

$$\mathcal{F} = \mathfrak{S}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong \{\epsilon\} + (\mathcal{G} \star \mathcal{F}).$$

Sets: we have

$$\mathcal{F} = \mathfrak{P}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong \{\epsilon\} + (\mathcal{G}^{\blacksquare} \star \mathcal{F}),$$

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

$$\mathcal{F} = \mathfrak{C}\{\mathcal{G}\} \qquad \Longrightarrow \qquad \mathcal{F} \cong (\mathcal{G}^{\blacksquare} \times \mathfrak{S}\{\mathcal{G}\}).$$

Greene [**73**] 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.

$\triangleright$ **23.** *Higher order constraints.* Let the symbols $\square$, $\boxdot$, $\blacksquare$ represent smallest, second smallest, and largest labels respectively. One has the correspondences (with $\partial_z = \frac{d}{dz}$)

$$\begin{array}{ll} \mathcal{A} = \left(\mathcal{B}^{\square} \star \mathcal{C}^{\blacksquare}\right) & \partial_z^2 A(z) = (\partial_z B(z)) \cdot (\partial_z C(z)) \\ \mathcal{A} = \left(\mathcal{B}^{\square\,\blacksquare} \star \mathcal{C}\right) & \partial_z^2 A(z) = \left(\partial_z^2 B(z)\right) \cdot C(z) \\ \mathcal{A} = \left(\mathcal{B}^{\square} \star \mathcal{C}^{\boxdot} \star \mathcal{D}^{\blacksquare}\right) & \partial_z^3 A(z) = (\partial_z B(z)) \cdot (\partial_z C(z)) \cdot (\partial_z D(z)), \end{array}$$

and so on. These can be transformed into (iterated) integral representations. [See [**73**] for more.] $\triangleleft$

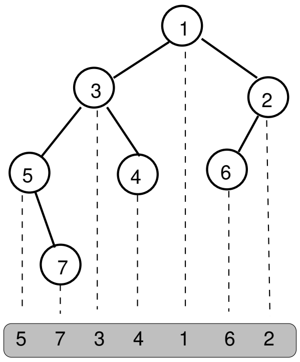


FIGURE 13. A permutation of size 7 and its increasing binary tree lifting.

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 turn gives access to new characteristics of permutations.

EXAMPLE 16. *Increasing binary trees and alternating permutations.* To each permutation, one can associate bijectively a binary tree of a special type[5] 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.

The correspondence (Figure 13) 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 [**135**, p. 23–25] who says that "it has been primarily developed by the French", pointing at [**64**].)

Thus, the family $\mathcal{I}$ of binary increasing trees satisfies the recursive definition

$$\mathcal{I} = \{\epsilon\} + \left(\mathcal{Z}^{\Box} \star \mathcal{I} \star \mathcal{I}\right),$$

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.

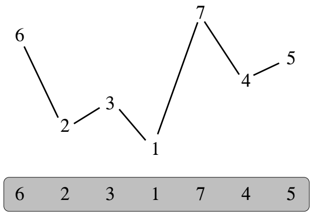
[5]Such trees are closely related to classical data structures of computer science, like heaps and binomial queues [**30, 129**].

The construction of increasing trees associated with permutation is instrumental in deriving EGF's relative to various local order patterns in permutations, like the number of ascents and descents, rises, falls, peaks and troughs, etc. We illustrate its use here by counting the number of *up-and-down* (or *zig-zag*) permutations, also known as *alternating* permutations. The result 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

$$\mathcal{J} = \mathcal{Z} + \left(\mathcal{Z}^{\Box} \star \mathcal{J} \star \mathcal{J}\right),$$

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

$$\overline{\mathcal{J}} = \{\epsilon\} + \left(\mathcal{Z}^{\Box} \star \mathcal{J} \star \overline{\mathcal{J}}\right),$$

since now all internal nodes of the tree representation are binary, except for the rightmost one that only branches on the left. Thus, $J'(z) = \tan(z)J(z)$, 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**). □

Use will be made later in this book (Chapter III, p. 17) 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 [**102, 130, 147, 148**].

▷ **24.** *Combinatorics of trigonometrics.* Interpret $\tan \frac{z}{1-z},\ \tan\tan z,\ \tan(e^z - 1)$ as EGFs. ◁

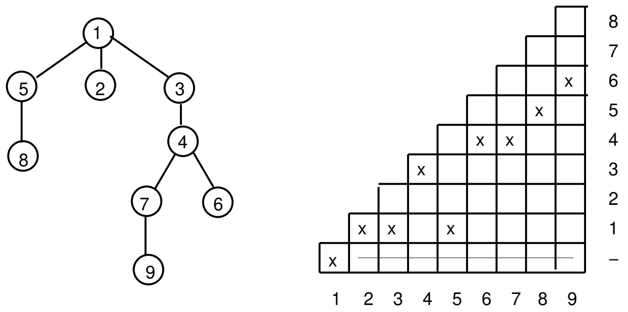


FIGURE 14. An increasing Cayley tree (left) and its associated regressive mapping (right).

EXAMPLE 17. *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

$$\mathcal{K} = \left(\mathcal{Z}^{\Box} \star \mathfrak{P}\{\mathcal{K}\}\right).$$

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 [**104**] 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 14). 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 fact that

$$K_n = 1 \cdot 2 \cdot 3 \cdots (n-1)$$

receives a natural interpretation. $\Box$

Regressive mappings can be also related directly to permutations. The construction that associates a regressive mapping to a permutation is called the "inversion table" construction; see [**86, 130**]. In short, given a permutation $\sigma = \sigma_1, \ldots, \sigma_n$, one can associate to it a function $\psi = \psi_\sigma$ from $[1 \mathinner{.\,.} n]$ to $[0 \mathinner{.\,.} n-1]$, 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.) Summarizing, we have a double combinatorial connection,

$$\text{Increasing Cayley tree} \cong \text{Regressive mappings} \cong \text{Permutations},$$

that opens the way to yet more permutation enumerations.

$\triangleright$ **25.** *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. 48) then gives rise to a binary increasing tree. Hence, increasing Cayley trees and increasing binary trees are also directly related. $\triangleleft$

II. 7. Notes

Labelled constructions are a frequently used paradigm of combinatorial analysis with applications to order statistics and graphical enumerations for instance. See the books by Comtet [**28**], Wilf [**153**], Stanley [**135**], or Goulden and Jackson [**68**] for many examples.

The labelled set construction and the exponential formula were recognized early by researchers working in the area of graphical enumerations [**76**]. Foata [**62**] 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 [**134**]. This is parallel to the concept of "prefab" due to Bender and Goldman [**11**].

Greene developed a general framework of "labelled grammars" largely based on the boxed product with implications for the random generation of combinatorial structures. Joyal's theory of species [**79**], already mentioned in the previous chapter, 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 [**56**] how to compile automatically specifications into generating functions; this is complemented by a calculus that produces fast random generation algorithms [**61**].

CHAPTER III

Combinatorial Parameters and Multivariate Generating Functions

Generating functions find averages, etc.
— HERBERT WILF [**153**]

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.
— JEAN-JACQUES ROUSSEAU, *Les Confessions*, Livre VI

Many scientific endeavours, in probability theory and statistics, computer science and analysis of algorithms, statistical physics and computational biology demand precise quantitative informations on probabilistic properties of *parameters* of combinatorial objects. For the purpose of designing, analysing, and optimizing a sorting algorithm, it is for instance 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. The "exact" problem is then a refined counting problem with two parameters, size and additional characteristic; 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. As demonstrated in this chapter, 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.

Multivariate generating functions—ordinary or exponential—can keep track of the number of components in a composite construction, like a sequence, a (multi)set, or a cycle. Generally, multivariate generating functions give access to "inherited" parameters defined inductively over combinatorial objects. This includes the number of occurrences of designated "patterns" to be found in an object of a given size. From such generating functions, there result either explicit probability distributions or, at least, mean and variance evaluations. Essentially all the combinatorial classes discussed in the first two chapters are amenable to such a treatment. Typical applications 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. Technically, the translation schemes that relate combinatorial constructions and multivariate generating functions present no major difficulty, since they appear to be natural (notational, even) refinements of the paradigm developed in Chapters I and II for the univariate case.

Beyond its technical aspects anchored in "symbolic combinatorics", this chapter also serves as a first encounter with the general area of "random combinatorics". The question is: *What does a random object of large size look like?* Multivariate generating functions when combined with probabilistic inequalities often offer definitive answers. 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 even further in Chapters IV–VII, after complex-asymptotic methods have been grafted on exact modelling by generating functions.

This chapter is organized as follows. Section III. 1 first introduces the basic notions of multivariate enumeration and multivariate generating function. There, we shall also discuss the relations with discrete probabilistic models, as the language of elementary probability theory does provide an intuitively appealing way to conceive of multivariate counting data. The symbolic method *per se* declined in its multivariate version is centrally developed in Sections III. 2 and III. 3: with suitable multi-index notations, the extension to the multivariate case is almost immediate. Recursive parameters that often arise from tree statistics form the subject of Section III. 4, while "universal" generating functions and combinatorial models are discussed in Section III. 5. 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. 6. The chapter concludes with Section III. 7 that presents a brief abstract discussion of extremal parameters like height in trees or smallest and largest components in composite structures, which leads to families of univariate generating functions.

III. 1. Parameters, generating functions, and distributions

Our purpose here is to analyse various characteristics of combinatorial structures. Most of the time, we shall be interested in enumeration according to size *and* a single auxiliary parameter. However, the theory is best developed in full generality for the joint analysis of a *finite collection* of parameters.

DEFINITION III.1. *Consider a combinatorial class $\mathcal{A}$. A* parameter $\chi = (\chi_1, \dots, \chi_d)$ *on the class is a function from $\mathcal{A}$ to the set $\mathbb{N}^d$ of d-tuples of natural numbers. The* counting sequence *of $\mathcal{A}$ with respect to size and the parameter χ is then defined by*

$$A_{n,k_1,\dots,k_d} = \text{card}\left\{\alpha \;\middle|\; |\alpha| = n,\ \chi_1(\alpha) = k_1, \dots, \chi_d(\alpha) = k_d\right\}.$$

We sometimes refer to such a parameter as a "multiparameter" (in particular when $d > 1$), as a "simple" or "scalar" parameter otherwise. One may take for $\mathcal{A}$ the class $\mathcal{P}$ of all permutations, and for $\chi \equiv \chi_1$ the parameter that associates to a permutation the number of its cycles. Natural questions are then: How many permutations of size n have k cycles? What is the expected number of cycles in a random permutation? Does this parameter have a distribution that can be made explicit? What are the features of this distribution in terms of "shape", "concentration", or limiting asymptotic behaviour. See Figure 1 for a first example related to binary words and Figure 2 for histograms relative to words and to cycles in permutations.

III. 1.1. Multivariate generating functions. Not too unexpectedly, the treatment of parameters in this book will be in terms of generating functions. The multi-index convention employed in various branches of mathematics greatly simplifies notations and is as follows: let $\mathbf{u} = (u_1, \dots, u_d)$ be a vector of d formal variables and $\mathbf{k} = (k_1, \dots, k_d)$

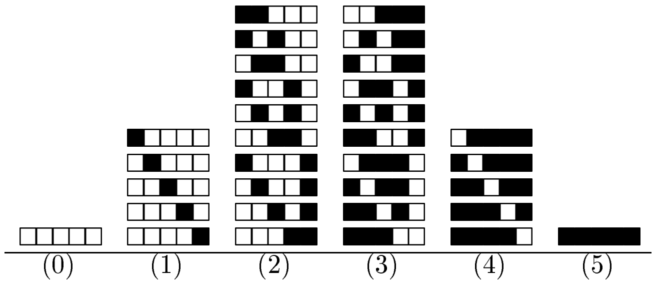


FIGURE 1. The set $\mathcal{W}_5$ of the 32 binary words over the alphabet $\{\square, \blacksquare\}$ enumerated according to the number of occurrences of the letter '$\blacksquare$' gives rise to the bivariate counting sequence $\{W_{5,j}\} = 1, 5, 10, 10, 5, 1$.

be a vector of integers of the same dimension; then, the multi-power $\mathbf{u}^{\mathbf{k}}$ is defined as the monomial

$$\mathbf{u}^{\mathbf{k}} := u_1^{k_1} u_2^{k_2} \cdots u_d^{k_d}. \tag{1}$$

With this notation, we have:

DEFINITION III.2. *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 by*

$$\begin{aligned} A(z, \mathbf{u}) &= \sum_{n,k} A_{n,\mathbf{k}} \mathbf{u}^{\mathbf{k}} z^n && \textit{(ordinary MGF)} \\ A(z, \mathbf{u}) &= \sum_{n,k} A_{n,\mathbf{k}} \mathbf{u}^{\mathbf{k}} \frac{z^n}{n!} && \textit{(exponential MGF)}, \end{aligned} \tag{2}$$

where the multi-index convention is in force.

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

$$\begin{aligned} A(z, \mathbf{u}) &= \sum_{\alpha \in \mathcal{A}} \mathbf{u}^{\chi(\alpha)} z^{|\alpha|} && \textit{(ordinary MGF; unlabelled case)} \\ A(z, \mathbf{u}) &= \sum_{\alpha \in \mathcal{A}} \mathbf{u}^{\chi(\alpha)} \frac{z^{|\alpha|}}{|\alpha|!} && \textit{(exponential MGF; labelled case).} \end{aligned} \tag{3}$$

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 parameter u, with the property for the multivariate GF to reduce to the univariate counting GF at $u = 1$.

In the case of a single parameter, these formulæ give rise to a *bivariate generating function*, also abbreviated as BGF. As already pointed out, this is the most frequently encountered situation in this book. In the many cases where the univariate versus multivariate distinction does not need to be stressed, we shall allow ourselves to use common (italic) letters to represent both scalars and vectors (so that $\mathbf{u} \mapsto u$ and $\mathbf{k} \mapsto k$): in such cases, the multi-index convention is automatically understood as soon as $d > 1$. (In this way, generating functions can be written with the less intrusive notation $A(z, u)$.)

The counting of $\mathcal{A}$–structures according to size and values of the scalar parameter χ is entirely encoded into a bivariate generating function. In order to put the OGF and EGF cases under a common umbrella, set

$$\omega_n = 1 \text{ (OGF, unlabelled case)}, \qquad \omega_n = n! \text{ (EGF, labelled case)}.$$

One may then arrange the BGF either in powers of z or in powers of u:

$$\begin{aligned} A(z,u) &= \sum_n A_n(u)\frac{z^n}{\omega_n} \\ &= \sum_k A^{\langle k\rangle}(z)u^k. \end{aligned}$$

If one views the table of coefficients as a 2-dimensional table, the $A_n(u)$ describe the behaviour of χ over all objects of some fixed size n—these are sometimes called the "horizontal" GF's associated to the BGF; the $A^{\langle k\rangle}(z)$, also called "vertical" generating functions, count the objects in $\mathcal{A}$ associated to fixed values of the parameter χ. Here is a diagram that displays the GFs stemming from a single BGF $A(z, u)$ and justifies this "horizontal–vertical" terminology.

$$
\begin{array}{|c|c|c|c||c}
\multicolumn{4}{c}{\overbrace{\hphantom{A^{\langle 0\rangle}(z)\quad A^{\langle 1\rangle}(z)\quad A^{\langle 2\rangle}(z)}}^{\text{vertical GF's}}} & \\
A^{\langle 0\rangle}(z) & A^{\langle 1\rangle}(z) & A^{\langle 2\rangle}(z) & & \\
\downarrow & \downarrow & \downarrow & & \\
\hline
+A_{0,0}u^0z^0 & +A_{0,1}u^1z^0 & +A_{0,2}u^2z^0 & \cdots\cdots & \longleftarrow A_0(u) \\
+A_{1,0}u^0z^1 & +A_{1,1}u^1z^1 & +A_{1,2}u^2z^1 & \cdots\cdots & \longleftarrow A_1(u) \\
+A_{2,0}u^0z^2 & +A_{2,1}u^1z^2 & +A_{2,2}u^2z^2 & \cdots\cdots & \longleftarrow A_2(u) \\
+A_{3,0}u^0z^3 & +A_{3,1}u^1z^3 & +A_{3,2}u^2z^3 & \cdots\cdots & \longleftarrow A_3(u) \\
\vdots & \vdots & \vdots & & \\
\hline
\end{array}
\left.\vphantom{\begin{array}{c}1\\1\\1\\1\end{array}}\right\}\ \text{horizontal GFs;}
$$

see also [**130**]. (Technically, we are taking advantage of the isomorphism between formal power series: $\mathbb{C}[\![z,u]\!] \cong \mathbb{C}[\![u]\!][\![z]\!] \cong \mathbb{C}[\![z]\!][\![u]\!]$.) Accordingly, the coefficients $A_{n,k}$ are recovered by applying the coefficient operator repeatedly in any convenient order. For instance, for a simple parameter

$$A_{n,k} = \omega_n \cdot [u^k z^n]A(z,u) \equiv \omega_n \cdot [z^n]\Big([u^k]A(z,u)\Big) \equiv \omega_n \cdot [u^k]\Big([z^n]A(z,u)\Big),$$

As a first illustration, consider the binomial coefficient $\binom{n}{k}$, already discussed from a univariate point of view in Chapter I, as it counts the binary words of length n having k occurrences of a designated letter; see Figure 1. In order to compose the bivariate GF, start from the simplest case of Newton's binomial theorem and form directly the horizontal GFs:

$$W_n(u) := \sum_{k=0}^{n}\binom{n}{k}u^k = (1+u)^n,$$

Then a summation over all values of n gives the ordinary BGF

$$W(z,u) = \sum_{k,n\ge 0}\binom{n}{k}u^k z^n = \sum_{n\ge 0}(1+u)^n z^n = \frac{1}{1-z(1+u)}. \tag{4}$$

(There, the second equality results from a computation in $\mathbb{C}[[u]][[z]]$.) The vertical OGFs of the binomial coefficients are

$$W^{\langle k\rangle}(z) = \sum_{n\geq 0} \binom{n}{k} z^n = \frac{z^k}{(1-z)^{k+1}},$$

as results from a direct calculation based on Newton's binomial theorem with negative exponents, or via an expansion of the BGF with respect to u:

$$W(z,u) = \frac{1}{1-z}\frac{1}{1-u\frac{z}{1-z}} = \sum_{k\geq 0} u^k \frac{z^k}{(1-z)^{k+1}}.$$

Such calculations are typical of MGF manipulations. Observe that (4) reduces to the OGF $(1-2z)^{-1}$ of binary words, as it should, upon setting $u=1$.

▷ **1.** *Exponential GFs of binomial coefficients.* The exponential BGF of binomial coefficients is

$$\widetilde{W}(z,u) = \sum_{k,n} \binom{n}{k} u^k \frac{z^n}{n!} = \sum (1+u)^n \frac{z^n}{n!} = e^{z(1+u)}. \tag{5}$$

The vertical EGFs are $e^z z^k/k!$. The horizontal GFs are $(1+u)^n$, like in the ordinary case. ◁

As a second illustration, we saw in Chapter II (Example 11) that the number of permutations of size n having k cycles is the Stirling cycle number $\left[{n \atop k}\right]$. The EGF is, for fixed k, given by

$$P^{\langle k\rangle}(z) := \sum_n \begin{bmatrix} n \\ k \end{bmatrix} \frac{z^n}{n!} = \frac{L(z)^k}{k!}, \qquad L(z) := \log\frac{1}{1-z}.$$

The starting point is thus a collection of vertical EGFs. From there, the exponential BGF is easily formed as follows:

$$\begin{aligned} P(z,u) &:= \sum_k P^{\langle k\rangle}(z) u^k \\ &= \sum_k \frac{u^k}{k!} L(z)^k \quad = \quad e^{uL(z)} \\ &= (1-z)^{-u}. \end{aligned} \tag{6}$$

The simplification is quite remarkable but altogether quite typical, as we shall see shortly, in the context of a labelled set construction.

An expansion of the BGF according to the variable z further gives by virtue of Newton's binomial theorem:

$$\begin{aligned} P(z,u) &= \sum_{n\geq 0} \binom{n+u-1}{n} z^n \\ P_n(u) &= u(u+1)\cdots(u+n-1) \equiv \sum_k \begin{bmatrix} n \\ k \end{bmatrix} u^k. \end{aligned}$$

This last polynomial is a horizontal GF 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 passing, note that the relation

$$P_n(u) = P_{n-1}(u)(u+(n-1)),$$

is equivalent to a recurrence

$$\begin{bmatrix} n \\ k \end{bmatrix} = (n-1)\begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n-1 \\ k-1 \end{bmatrix},$$

by which Stirling numbers are often defined and easily evaluated numerically; see also APPENDIX: *Stirling numbers*, p. 173. (The recurrence is otherwise susceptible to a direct combinatorial interpretation—add n either to an existing cycle or as a "new" singleton.)

▷ **2.** *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

$$P(z, u_1, u_2) = \frac{\exp\left((u_1-1)z + (u_2-1)\frac{z^2}{2}\right)}{1-z}. \tag{7}$$

(This is to be proved later in this chapter, p. 137.) 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 off *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 the number of singleton cycles. This last BGF itself interpolates between the EGF of derangements $(u=0)$ and the EGF of all permutations $(u=1)$. ◁

Concise expressions for BGFs like (4), (5), (6), or (7) are precious for deriving moments, variance, and even finer characteristics of distributions, as we see next.

III. 1.2. Distributions, moments, and generating functions. As indicated in the preamble to this chapter, the eventual goal of multivariate enumeration is the quantification of properties present with high regularity in large random structures. With this subsection and the next one, we momentarily digress from our primary objective in order to introduce the basic concepts of discrete probability needed to interpret multivariate counting sequences.

Consider a pair $\langle \mathcal{A}, \chi \rangle$, where $\mathcal{A}$ is a class and χ a parameter. The *uniform probability distribution* over $\mathcal{A}_n$ is defined as follows: the probability of any $\alpha \in \mathcal{A}_n$ is equal to $1/A_n$ and the probability of any set (or "event") $\mathcal{E} \subseteq \mathcal{A}_n$ is

$$\mathbb{P}\{\mathcal{E}\} = \frac{\text{card}(\mathcal{E})}{A_n}$$

("the number of favorable cases over the total number of cases"). 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, take for simplicity the parameter χ to be scalar (i.e., $d=1$). We regard χ as defining over each $\mathcal{A}_n$ a (discrete) *random variable* defined over the (discrete) probability space $\mathcal{A}_n$:

$$\mathbb{P}_{\mathcal{A}_n}\{\chi(\alpha) = k\} = \frac{A_{n,k}}{A_n} = \frac{A_{n,k}}{\sum_k A_{n,k}}.$$

This way of thinking enables us to make use of whichever probabilistic intuition might be available in any particular case, while allowing for a natural interpretation of data. Indeed, instead of noting that there are 381922055502195 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 above k indicates the value of $\mathbb{P}\{X = k\}$. Figure 2 displays in this way two classical combinatorial distributions. Given

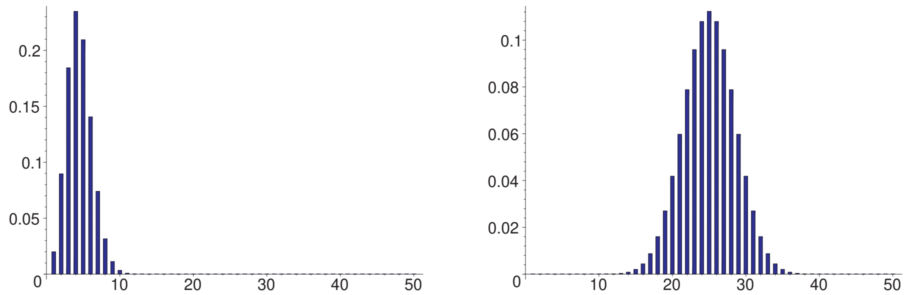


FIGURE 2. Histograms of two distributions. Left: the number of cycles in a random permutation of size 50 (Stirling cycle distribution). Right: the number of occurrences of a designated letter in a random binary word of length 50 (binomial distribution).

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 1.

An important information is provided by *moments*. Given a discrete random variable (RV) X, 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

$$\mathbb{E}(X^r) = \sum_k \mathbb{P}\{X = k\} \cdot k^r.$$

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 $\mathbb{E}(X)$.

Bivariate generating functions can be put to use in order to determine probability generating function and moments of parameters. Consider a BGF $A(z, u)$, where z marks size and u marks the parameter χ. Coefficient extraction then yields a polynomial

$$A_n(u) := \omega_n \cdot [z^n]A(z, u),$$

whose coefficients enumerate the configurations $\alpha \in \mathcal{A}_n$ according to the value of the χ parameter. Also, we have $A_n = A_n(1)$ the total number of objects in $\mathcal{A}_n$ having size n.

Consequently, the normalized polynomial

$$p_n(u) := \frac{A_n(u)}{A_n(1)} = \frac{[z^n]A(z,u)}{[z^n]A(z,1)}$$

is the *probability generating function* (PGF) of χ on $\mathcal{A}_n$ in the sense that

$$[u^k]p_n(u) = \mathbb{P}_{\mathcal{A}_n}\{\chi = k\}, \qquad \text{equivalently,} \qquad p_n(u) = \sum_k \mathbb{P}_{\mathcal{A}_n}\{\chi = k\}u^k.$$

Successive differentiations then give access to the moments of χ on A_n. In particular, one has

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}_n}(\chi) &= \left(\partial_u p_n(u)\right)_{u=1} && \partial_u := \frac{\partial}{\partial u} \\
\mathbb{E}_{\mathcal{A}_n}(\chi^2) &= \left(\partial_u^2 p_n(u) + \partial_u p_n(u)\right)_{u=1} && \partial_u^2 := \frac{\partial^2}{\partial u^2}, \text{ etc.}
\end{aligned}$$

We thus get:

PROPOSITION III.1 (Moments from BGFs). *The moments of order 1 (mean) and of order 2 of a parameter χ are determined from the BGF $A(z,u)$ by differentiation and specialization at 1 as follows:*

$$\begin{aligned}
\mathbb{E}_{\mathcal{A}_n}(\chi) &= \frac{[z^n]\partial_u A(z,1)}{[z^n]A(z,1)} \\
\mathbb{E}_{\mathcal{A}_n}(\chi^2) &= \frac{[z^n]\partial_u^2 A(z,1)}{[z^n]A(z,1)} + \frac{[z^n]\partial_u A(z,1)}{[z^n]A(z,1)}.
\end{aligned}$$

In particular, the *standard deviation* is recovered from there by the usual formula,

$$\sigma(\chi)^2 = \mathbb{E}(\chi^2) - \mathbb{E}(\chi)^2.$$

As seen from basic definitions, the quantities

$$\Omega_n^{(k)} := \omega_n \cdot \left([z^n]\,\partial_u^k A(z,u)\big|_{u=1}\right)$$

give, up to normalization, the so-called *factorial moments*

$$\mathbb{E}\left(\chi(\chi-1)\cdots(\chi-k+1)\right) = \frac{1}{A_n}\Omega_n^{(k)}.$$

(Factorial moments and power moments are clearly connected by linear relations; as a matter of fact, the connection coefficients are Stirling numbers.) Most notably, $\Omega_n^{(1)}$ is the *cumulated value* of χ over all objects of $\mathcal{A}_n$:

$$\Omega_n^{(1)} \equiv \omega_n \cdot [z^n]\,\partial_u A(z,u)\big|_{u=1} = \sum_{\alpha\in\mathcal{A}_n} \chi(\alpha) \equiv A_n \cdot \mathbb{E}_{\mathcal{A}_n}(\chi).$$

EXAMPLE 1. *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 [**84**].

We are dealing with labelled objects, hence exponential generating functions. As seen earlier on p. 111, 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

$$\mathbb{E}_n(\chi) = [z^n]\frac{1}{1-z}\log\frac{1}{1-z} = 1 + \frac{1}{2} + \cdots + \frac{1}{n}, \tag{8}$$

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.

For the variance, a further differentiation of the bivariate EGF gives

$$\sum_{n\geq 0} \mathbb{E}_n(\chi(\chi-1))z^n = \frac{1}{1-z}\left(\log\frac{1}{1-z}\right)^2. \tag{9}$$

From this expression (or from the Stirling polynomials), a calculation shows that

$$\sigma_n^2 = \left(\sum_{k=1}^{n}\frac{1}{k}\right) - \left(\sum_{k=1}^{n}\frac{1}{k^2}\right). \tag{10}$$

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 [**66**] and Chapter VII. □

▷ **3.** *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 (8) and (10). 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

$$(1-z)^{-\alpha}\log\frac{1}{1-z} = \sum_{n\geq 0}\left(\frac{1}{\alpha} + \frac{1}{\alpha+1} + \cdots + \frac{1}{n-1+\alpha}\right)\binom{n+\alpha-1}{n}z^n.$$

while the next differentiation gives access to (10). ◁

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. 4 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 VII). In all cases, the BGF's 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 BGF's themselves.

III. 1.3. Moment inequalities. We conclude this section by a few remarks that make precise our earlier informal discussion of concentration.

Qualitatively speaking, families of distributions can be classified in two categories: the ones that are "concentrated" (i.e., the standard deviation is much smaller than the mean) and the ones that are "spread" (i.e., the standard deviation is at least as large as the mean). Figure 2 illustrates the phenomena at stake and suggests that both the Stirling cycle distributions and the binomial distributions are somehow concentrated. In contrast, the uniform distributions over $[0, n]$, which have totally flat histograms, are spread. Such informal observations are indeed supported by the Markov-Chebyshev inequalities:

PROPOSITION III.2 (Markov-Chebyshev inequalities). *Let X be a nonnegative random variable and Y an arbitrary real variable. One has*

$$\begin{aligned} \mathbb{P}\{X \geq t\mathbb{E}(X)\} &\leq \frac{1}{t} && \textit{(Markov inequality)} \\ \mathbb{P}\{|Y - \mathbb{E}(Y)| \geq t\sigma(X)\} &\leq \frac{1}{t^2} && \textit{(Chebyshev inequality).} \end{aligned}$$

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 $X = |Y - \mathbb{E}(Y)|^2$. □

Proposition III.2 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). These bounds are *universal* in the sense that they hold for all random variables. In fact, in most cases of combinatorial interest, it is the case that far stronger decay rates—of an exponential nature—hold: see Chapter VII on multivariate asymptotics and limit distributions.

The next proposition formalizes a notion of concentration for distributions. It applies to a *family* of distributions indexed by the integers, typically the values of a scalar parameter χ on the subclasses $\{\mathcal{A}_n\}_{n\geq 0}$ indexed by size.

PROPOSITION III.3 (Concentration of distribution). *Consider a family of random variables X_n, e.g., values of 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*

$$\lim_{n\to+\infty} \mathbb{P}\left\{1 - \epsilon \leq \frac{X_n}{\mu_n} \leq 1 + \epsilon\right\} = 1. \tag{11}$$

PROOF. It is a direct consequence of Chebyshev's inequality. □

In probability theory, the concentration property (11) is called *convergence in probability* and is then written more concisely as

$$\frac{X_n}{\mu_n} \xrightarrow{P} 1 \qquad \text{or} \qquad X_n \xrightarrow{P} \mu_n.$$

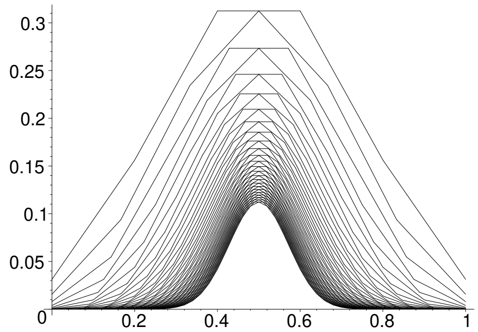


FIGURE 3. Plots of the binomial distributions for $n = 5, \ldots, 50$. The horizontal axis is normalized and rescaled to 1, so that the curves display $\left\{\mathbb{P}(\frac{X_n}{n} = x)\right\}$, for $x = 0, \frac{1}{n}, \frac{2}{n}, \ldots$.

It 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 with high probability (w.h.p.)*". When this property is satisfied, the expected value is in a strong sense a typical value.

For instance, 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 3 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 variance computation (10) implies that the number of cycles in a random permutation of large size is concentrated. (At the opposite end of the spectrum, the uniform distributions over $[1\,..\,n]$ are *not* concentrated.)

Moment inequalities are discussed for instance in Billingsley's reference treatise [**18**, 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 [**3**] 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 [**19, 78**]. Finally, moment inequalities serve to design, analyse, and optimize randomized algorithms, a theme excellently covered in the book by Motwani and Raghavan [**107**].

Finer estimates on distributions form the subject of our Chapter VII dedicated to limit laws. The reader may get a feeling of some of the phenomena at stake when re-examining Figure 3: the visible emergence of a continuous curve (the bell curve) corresponds to a common asymptotic shape for the whole family of distributions (the Gaussian law).

III. 2. Inherited parameters and ordinary multivariate generating functions

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 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 here 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 (2) assumes an extremely simple and symmetrical form:

$$A(\mathbf{z}) = \sum_{\mathbf{k}} A_{\mathbf{k}} \mathbf{z}^{\mathbf{k}} = \sum_{\alpha \in \mathcal{A}} \mathbf{z}^{\overline{\chi}(\alpha)}. \tag{12}$$

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 (1) is in force,

$$\mathbf{z}^{\mathbf{k}} := z_0^{k_0} z_1^{k_1} \cdots {z_d}^{k_d}, \tag{13}$$

but it is now applied to $(d+1)$-dimensional vectors.

Next, we define inherited parameters.

DEFINITION III.3. *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 any of $\mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$, 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

$$\langle \mathcal{A}, \chi \rangle = \langle \mathcal{B}, \xi \rangle + \langle \mathcal{C}, \zeta \rangle, \quad \langle \mathcal{A}, \chi \rangle = \langle \mathcal{B}, \xi \rangle \times \langle \mathcal{C}, \zeta \rangle, \quad \langle \mathcal{A}, \chi \rangle = \mathfrak{K}\left\{ \langle \mathcal{B}, \xi \rangle \right\}.$$

For instance, the class $\mathcal{I}$ of natural numbers, $\mathcal{I} = \mathfrak{S}_{\geq 1}\{\mathcal{Z}\}$ has OGF $I(z) = z/(1-z)$. Let ξ be the parameter that takes the constant value 1 on all elements of $\mathcal{I}$. The ordinary MGF of $\langle \mathcal{I}, \xi \rangle$ is simply

$$I(z,u) = zu + z^2u + z^3u + \cdots = \frac{zu}{1-z}.$$

The class $\mathcal{C}$ of integer compositions is, as seen in Chapter I, specified as the class of all sequences of natural integers: $\mathcal{C} = \mathfrak{S}\{\mathcal{I}\}$, with OGF

$$C(z) = \frac{1}{1 - \frac{z}{1-z}} = \frac{1-z}{1-2z}, \qquad \text{so that} \quad C_n = 2^{n-1}.$$

The constant parameter ξ is unimportant *per se*; however, the parameter χ on $\mathcal{C}$ inherited from $\langle \mathcal{I}, \xi \rangle$ carries some useful information as it represents the number of summands (or parts) that enters a composition. Its ordinary BGF is written $C(z,u)$, or $C(z_0, z_1)$ under the multi-index convention. It turns out (see below, p. 120) that the schemes translating admissible constructions in the univariate case (Chapter I) transport almost verbatim to the multivariate case, so that

$$C(z,u) = \frac{1}{1 - I(z,u)} = \frac{1}{1 - u\frac{z}{1-z}} = \frac{1-z}{1-z(u+1)}. \tag{14}$$

We have an altogether nontrivial result obtained without any computation, which directly derives from the basic specification $\mathcal{C} = \mathfrak{S}\{\mathcal{I}\}$ relating compositions to integers. This is precisely the spirit of the symbolic method applied to parameters.

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 is used. The associated operators on ordinary MGFs are then:*

$$\begin{array}{llcl}
\textit{Union:} & \mathcal{A} = \mathcal{B} + \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) + C(\mathbf{z}) \\
\textit{Product:} & \mathcal{A} = \mathcal{B} \times \mathcal{C} & \Longrightarrow & A(\mathbf{z}) = B(\mathbf{z}) \cdot C(\mathbf{z}) \\
\textit{Sequence:} & \mathcal{A} = \mathfrak{S}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z}) = \dfrac{1}{1 - B(\mathbf{z})} \\
\textit{Cycle:} & \mathcal{A} = \mathfrak{C}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z}) = \displaystyle\sum_{\ell=1}^{\infty} \frac{\varphi(\ell)}{\ell} \log \frac{1}{1 - B(\mathbf{z}^\ell)}. \\
\textit{Multiset:} & \mathcal{A} = \mathfrak{M}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z}) = \exp\left(\displaystyle\sum_{\ell=1}^{\infty} \frac{1}{\ell} B(\mathbf{z}^\ell) \right) \\
\textit{Powerset:} & \mathcal{A} = \mathfrak{P}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z}) = \exp\left(\displaystyle\sum_{\ell=1}^{\infty} \frac{(-1)^{\ell-1}}{\ell} B(\mathbf{z}^\ell) \right)
\end{array}$$

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 are then built up from the union and product schemes, in exactly the same manner as in the proof of Theorem I.1. □

This theorem is shallow. However, its importance devolves from its extremely wide range of combinatorial consequences as well as the ease with which it can be applied. The reader is especially encouraged to study carefully the example that follows as it illustrates in its bare bones version the power of the symbolic method for taking into account combinatorial parameters.

EXAMPLE 2. *Summands in integer compositions.* Let us return to integer compositions, $\mathcal{C}$. The BGF of compositions with χ the scalar parameter equal to the number of summands is

$$C(z_0, z_1) = \frac{1}{1 - I(z_0, z_1)} = \frac{1}{1 - z_1 I(z)} = \frac{1}{1 - z_0 z_1 (1 - z_0)^{-1}}, \tag{15}$$

which, up to notations, is exactly Equation (14) that is now justified. Consider next the double parameter χ where χ_1 is the number of parts equal to 1 and χ_2 the number of parts equal to 2. This is inherited from the corresponding parameter on the class $\mathcal{I}$ of natural numbers, with MGF

$$I(z_0, z_1, z_2) = z_1 z_0 + z_2 z_2 + \frac{z_0^3}{1 - z_0} = \frac{z_0}{1 - z_0} + (z_1 - 1) z_0 + (z_2 - 1) z_0^2. \tag{16}$$

Consequently, the trivariate MGF of $\langle \mathcal{C}, \chi \rangle$ is

$$C(z_0, z_1, z_2) = \frac{1}{1 - I(z_0, z_1, z_2)}. \tag{17}$$

Observe that the marking variables betray their origin. For instance, in (16) and (17), one enumerates compositions through a marking by means of dedicated variables of the configurations to be recorded, while at the same time, the usual rules translating constructions are applied. Much use of this way of envisioning the technique will be made in the remainder of this chapter.

MGFs like (14) or (15) can then be exploited in the usual way through formal power series expansions. For instance, the number of compositions of n with k parts is, by (14),

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

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,

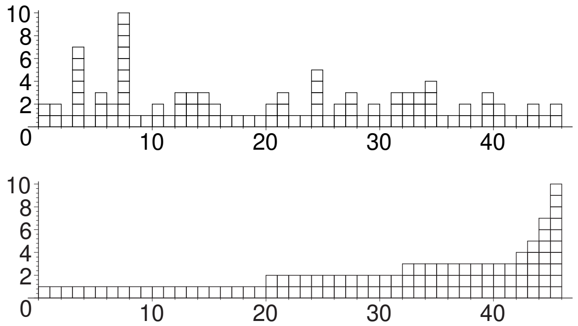


FIGURE 4. A random composition of $n = 100$ represented as a ragged landscape (top); its associated profile $1^{20}2^{12}3^{10}4^{1}5^{1}7^{1}10^{1}$, defined as the partition obtained by sorting the summands (bottom).

upon setting $z_0 = z$, $z_1 = u$ and $z_2 = 1$,

$$[z^n u^k]\frac{1-z}{1-uz-\frac{z^2}{(1-z)}} = [z^n]\frac{(1-z)^{k+1}}{(1-z-z^2)^k},$$

where the OGF closely resembles a power of the OGF of Fibonacci numbers.

Following the discussion of Section III. 1, such MGFs also carry complete information on moments. For instance, the cumulated value of the number of parts in all compositions of n has OGF

$$\partial_u C(z,u)|_{u=1} = \frac{1-z}{(1-2z)^2},$$

as seen from Section III. 1.2, since cumulated values are obtained via differentiation of a BGF. Therefore, the expected number of parts in a random composition of n is

$$\frac{1}{2^{n-1}}[z^n]\frac{z(1-z)}{(1-2z)^2} = \frac{1}{2}(n+1).$$

What we have shown is a property of random compositions: *On average, a random composition of the integer n has about $n/2$ summands.* 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. *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 found to have BGF

$$\widehat{C}(z,u) = \left(1 - \left(\frac{z}{1-z} + (u-1)z^r\right)\right)^{-1}.$$

Though expanding this expression explicitly would be cumbersome, one can still pull out the number of r-summands in a random composition of size n. The differentiated form

$$\partial_u \widehat{C}(z,u)\Big|_{u=1} = \frac{z^r(1-z)^2}{(1-2z)^2}.$$

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. Another differentiation gives access to the second moment. Consequently, one has (take the nth coefficient and divide by 2^{n-1}): *The number of r summands in a composition of size n has mean*

$$\frac{n}{2^{r+1}} + O(1);$$

the standard deviation is of order $\sqrt{n}$, which ensures concentration of distribution. □

From the point of view of random combinatorics, 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 and V.

▷ **4.** *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 second moment methods.) ◁

EXAMPLE 3. *Number of components in abstract schemas I.* Consider now a relation $\mathcal{A} = \mathfrak{K}\{\mathcal{B}\}$, where $\mathfrak{K}$ is any *unlabelled* constructor amongst $\mathfrak{S}, \mathfrak{C}, \mathfrak{M}, \mathfrak{P}$. The parameter "number of components", χ, defined on $\mathcal{A}$ is inherited from the constant parameter ξ equal to 1 on $\mathcal{B}$. The BGF of $\langle \mathcal{B}, \xi \rangle$ is simply

$$B(z, u) = uB(z),$$

with $B(z)$ the OGF of $\mathcal{B}$. The BGF of $\langle \mathcal{A}, \chi \rangle$ is then given by Theorem III.1. Finally, the cumulated quantities of the number of components,

$$\Omega_n := \sum_{\alpha \in \mathcal{A}} \chi(\alpha), \qquad \Omega(z) := \sum_n \Omega_n z^n,$$

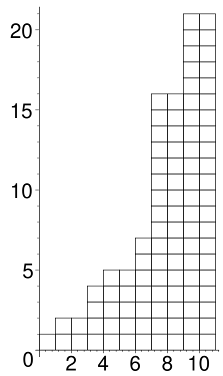


FIGURE 5. A random partition of size $n = 100$ has an aspect rather different from the profile of a random composition of the same size (Figure 4).

are given by the usual differentiation process $\partial_u(\cdot)|_{u=1}$. The easy computations are summarized by the following table:

(18)

$\mathfrak{K}$	MGF $(A(z,u))$	Cumul. OGF $(\Omega(z))$
Sequence:	$\dfrac{1}{1-uB(z)}$	$A(z)\cdot B(z) = \dfrac{B(z)}{(1-B(z))^2}$
Set:	$\exp\left(\sum_{k=1}^{\infty}(-1)^{k-1}\dfrac{u^k}{k}B(z^k)\right)$	$A(z)\cdot\sum_{k=1}^{\infty}(-1)^{k-1}B(z^k)$
Multiset:	$\exp\left(\sum_{k=1}^{\infty}\dfrac{u^k}{k}B(z^k)\right)$	$A(z)\cdot\sum_{k=1}^{\infty}B(z^k)$
Cycle:	$\sum_{k=1}^{\infty}\dfrac{\varphi(k)}{k}\log\dfrac{1}{1-u^kB(z^k)}$	$\sum_{k=1}^{\infty}\varphi(k)\dfrac{B(z^k)}{1-B(z^k)}.$

Mean values are then recovered as

$$\mathbb{E}_n(\chi) = \frac{\Omega_n}{A_n},$$

in accordance with the usual formula. $\square$

$\triangleright$ **5.** *r-Components in abstract schemas I.* 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 ($\mathfrak{K} = \mathfrak{S}$) and multisets ($\mathfrak{K} = \mathfrak{M}$), respectively. $\triangleleft$

As a next illustration, we discuss the profile of random partitions (Figure 5).

EXAMPLE 4. *The profile of partitions.* Let $\mathcal{P} = \mathfrak{M}\{\mathcal{I}\}$ be the class of all integer partitions. The BGF of $\mathcal{P}$ with u marking the number χ of parts (or summands) is

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

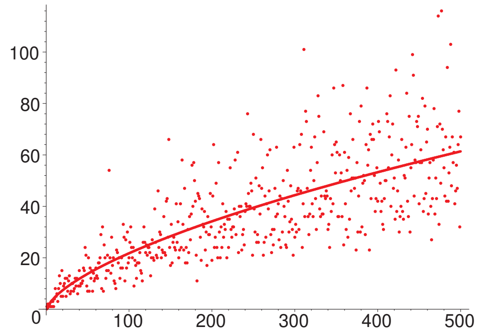


FIGURE 6. The number of parts in random partitions of size $1, \dots, 500$: exact values of the mean and simulations (circles, one for each value of n).

as results from first principles (see also (18)). The OGF of cumulated values,

$$\Omega(z) = P(z) \cdot \sum_{k=1}^{\infty} \frac{z^k}{1-z^k}, \tag{19}$$

is obtained by logarithmic differentiation. Now, the factor on the right in (19) can be expanded: one has

$$\sum_{k=1}^{\infty} \frac{z^k}{1-z^k} = \sum_{n=1}^{\infty} d(n) z^n,$$

with $d(n)$ the number of divisors of n. Thus, the mean value of χ is

$$\mathbb{E}_n(\chi) = \frac{1}{P_n} \sum_{j=1}^{n} d(j) P_{n-j}. \tag{20}$$

The same technique applies to the number of parts equal to r. The form of BGF

$$\widetilde{P}(z,u) = \frac{1-z^r}{1-uz^r} \cdot P(z),$$

implies that the mean number of r-parts is (apply ∂_u, the set $u=1$)

$$\mathbb{E}_n(\widetilde{\chi}) = \frac{1}{P_n}[z^n] \left(P(z) \cdot \frac{z^r}{1-z^r} \right) = \frac{1}{P_n} \left(P_{n-r} + P_{n-2r} + P_{n-3r} + \cdots \right).$$

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. □

The comparison between Figures 4 and 5 together with the supporting analysis shows that different combinatorial models may well lead to rather different types of probabilistic behaviours. Figure 6 displays the exact value of the mean number of parts in random partitions of size $n = 1, \dots, 500$, (as calculated from (20)) accompanied with the observed values of one random sample for each value of n in the range. The mean number of parts is 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 [**42**].

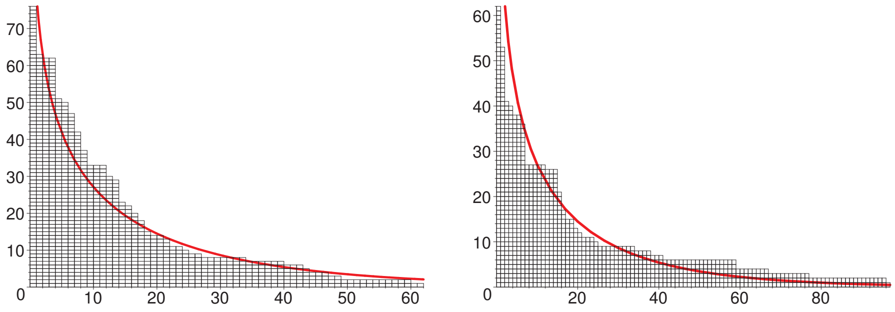


FIGURE 7. Two partitions of $\mathcal{P}_{1000}$ drawn at random, compared to the limiting shape $\Psi(x)$ defined by (21).

In recent years, Vershik and his collaborators [**38, 145**] have shown that most integer partitions tend to conform to a definite profile given (after normalization by $\sqrt{n}$) by the continous plane curve $y = \Psi(x)$ defined implicitly by

$$y = \Psi(x) \qquad \text{iff} \qquad e^{\alpha x} + e^{\alpha y} = 1, \qquad \alpha = \frac{\pi}{\sqrt{6}}. \tag{21}$$

This is illustrated in Figure 7 by two randomly drawn elements of $\mathcal{P}_{1000}$ drawn against the "most likely" limit shape. The theoretical result explains the huge 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. More "global" tree parameters (e.g., number of leaves and path length) that need a recursive definition will be discussed in Section III. 4 below.

EXAMPLE 5. *Root degree in general Catalan trees.* Consider the parameter χ equal to the degree of the root in a tree. Take the class $\mathcal{G}$ of all plane unlabelled trees, aka Catalan trees. A plane tree is a root to which is appended a sequence of trees,

$$\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\},$$

where the atomic class $\mathcal{Z}$ is the formed of a single node, so that

$$G(z) = \frac{z}{1 - G(z)}.$$

The bivariate GF with u marking χ is then

$$G(z, u) = \frac{z}{1 - uG(z)}.$$

(To see it from first principles, simply rewrite trees as roots appended to forests

$$\mathcal{G} = \mathcal{Z} \times \mathcal{F}, \qquad \mathcal{F} = \mathfrak{S}\{\mathcal{G}\}.$$

and define ξ on $\mathcal{F}$ as the number of components in the forest: χ on $\mathcal{G}$ is inherited from ζ on $\mathcal{F}$ and the constant weight 0 on the factor Z corresponding to the root. The parameter ζ on $\mathcal{F}$ is given by the usual rules for the number of components in sequences.)

From there, the cumulative GF is found,

$$\Omega(z) = \frac{zG(z)}{(1-G(z))^2}.$$

The recursive relation satisfied by G entails a further simplification,

$$\Omega(z) = \frac{1}{z}G(z)^3 = \left(\frac{1}{z}-1\right)G(z) - 1.$$

A closed form for the coefficient results, and the mean root degree is found to be

$$\mathbb{E}_n(\chi) = \frac{1}{G_n}\left(G_{n+1} - G_n\right) = 3\frac{n-1}{n+1},$$

which is clearly asymptotic to 3.

A closer analysis reveals that the probability that the root degree equals r is

$$\mathbb{P}_n\{\chi = r\} = \frac{1}{G_n}[z^n]z\,G(z)^r \sim r2^{-r-1}.$$

A random plane tree is thus usually composed of a small number of root subtrees, at least one of which should be accordingly fairly large. $\square$

III. 3. Inherited parameters and exponential multivariate generating functions

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. With a suitable use of multi-index conventions, the translation mechanisms developed in the univariate case (Chapter II) remain in vigour.

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 (13) defining $\mathbf{z}^{\mathbf{k}}$ has been brought into the game, the exponential MGF of $\langle\mathcal{A}, \chi\rangle$ (see Definition III.2) can be rephrased as

$$A(\mathbf{z}) = \sum_{\alpha\in\mathcal{A}} \frac{\mathbf{z}^{\mathbf{k}}}{k_0!} = \sum_{\alpha\in\mathcal{A}} \frac{\mathbf{z}^{\overline{\chi}(\alpha)}}{|\alpha|!}. \tag{22}$$

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 only consider 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 *acceptable* 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 acceptable and it is defined by cases on disjoint unions and determined additively on labelled products—this is Definition III.3 with labelled products replacing cartesian products. In particular, 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 provided the multi-index convention* (22) *is used. The associated operators on exponential MGFs are then:*

$$\begin{array}{llll}
\textit{Union:} & \mathcal{A}=\mathcal{B}+\mathcal{C} & \Longrightarrow & A(\mathbf{z})=B(\mathbf{z})+C(\mathbf{z}) \\
\textit{Product:} & \mathcal{A}=\mathcal{B}\star\mathcal{C} & \Longrightarrow & A(\mathbf{z})=B(\mathbf{z})\cdot C(\mathbf{z}) \\
\textit{Sequence:} & \mathcal{A}=\mathfrak{S}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z})=\dfrac{1}{1-B(\mathbf{z})} \\
\textit{Cycle:} & \mathcal{A}=\mathfrak{C}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z})=\log\dfrac{1}{1-B(\mathbf{z})}. \\
\textit{Set:} & \mathcal{A}=\mathfrak{P}\{\mathcal{B}\} & \Longrightarrow & A(\mathbf{z})=\exp\big(B(\mathbf{z})\big).
\end{array}$$

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. ☐

This theorem can be exploited to determine moments, in a way that entirely parallels its unlabelled counterpart.

EXAMPLE 6. *The profile of permutations.* Let $\mathcal{P}$ be the class of all permutations and χ the number of components. The parameter χ is inherited from the parameter having constant value 1 on all cyclic permutations. Therefore, the exponential BGF is

$$P(z,u)=\exp\left(u\log\frac{1}{1-z}\right)=(1-z)^{-u},$$

as was already obtained by an *ad hoc* calculation in (6). We also know (page 115) 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.

Let $\widetilde{\chi}$ be the number of cycles of length r. The exponential BGF is

$$\widetilde{P}(z,u)=\exp\left(\log\frac{1}{1-z}+(u-1)\frac{z^r}{r}\right).$$

The EGF of cumulated values is then obtained by differentiating and with respect to u and setting $u=1$:

$$\widetilde{\Omega}(z)=\frac{z^r}{r}\,\frac{1}{1-z}.$$

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* $r\le n$*.*

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. This formula sheds a new light on the harmonic number formula for the mean number of cycles. In particular, 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 by Shepp and Lloyd [**131**] for an original discussion of largest and smallest cycles).

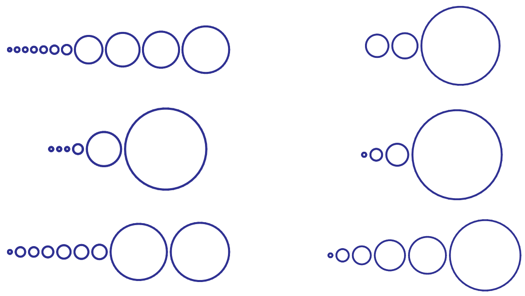


FIGURE 8. The profile of permutations: a rendering of the cycle structure of six random permutations of size 500, where circle areas are drawn in proportion to cyle lengths. Permutations tend to have a few small cycles (of size $O(1)$), a few large ones (of size $\Theta(n)$), and altogether have $\mathrm{H}_n \sim \log n$ cycles on average.

Since formuæ for labelled objects are so simple, one can get more. The BGF of the number of r-cycles is

$$\widetilde{P}(z,u) = \frac{e^{-z^r/r}}{1-z}\, e^{uz^r/r},$$

so that

$$\mathbb{P}\{\overline{\chi} = k\} = \frac{1}{k!\, r^k}[z^{n-kr}]\frac{e^{-z^r/r}}{1-z},$$

where one recognizes in the last factor the EGF of permutations without cycles of length r. From this (and the asymptotics of generalized derangement numbers in Chapter IV),one proves easily that the asymptotic law of the number of r-cycles is Poisson[1] of rate $\frac{1}{r}$. (This interesting property to be established in later chapters constitutes the starting point of [**131**].) □

EXAMPLE 7. *Number of components in abstract schemas II.* 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 $\mathfrak{S}, \mathfrak{C}, \mathfrak{P}$. The exponential BGF $A(z,u)$ and the exponential GF $\Omega(z)$ of cumulated values are given by the following table:

(23)

$\mathfrak{K}$	exp. MGF $(A(z,u))$	Cumul. EGF $(\Omega(z))$
Sequence:	$\dfrac{1}{1-uB(z)}$	$A(z)\cdot B(z) = \dfrac{B(z)}{(1-B(z))^2}$
Set:	$\exp(uB(z))$	$A(z)\cdot B(z) = B(z)e^{B(z)}$
Cycle:	$\log\dfrac{1}{1-uB(z)}$	$\dfrac{B(z)}{1-B(z)}$.

Mean values are then easily recovered, and one finds

$$\mathbb{E}_n(\chi) = \frac{\Omega_n}{A_n} = \frac{[z^n]\Omega(z)}{[z^n]A(z)},$$

[1] The Poisson distribution of rate $\lambda > 0$ is supported by the nonnegative integers and determined by

$$\mathbb{P}\{X\} = e^{-\lambda}\frac{\lambda^k}{k!}.$$

by the same formula as in the unlabelled case. □

EXAMPLE 8. *Set partitions.* Set partitions $\mathcal{S}$ are built of blocks, $\mathcal{S} = \mathfrak{P}\{\mathfrak{P}_{\geq 1}\{\mathcal{Z}\}\}$, and the construction is reflected by the EGF equation

$$S(z) = e^{V(z)} \qquad \text{with} \qquad V(z) = e^z - 1.$$

The bivariate EGF with u marking the number of blocks is then

$$S(z,u) = e^{uV(z)} = e^{u(e^z-1)}.$$

Since set partitions are otherwise known to be enumerated by the Stirling partition numbers, one has

$$\sum_{n,k} \left\{ {n \atop k} \right\} u^k \frac{z^n}{n!} = e^{u(e^z-1)}, \qquad \sum_n \left\{ {n \atop k} \right\} \frac{z^n}{n!} = \frac{1}{k!}(e^z-1)^k,$$

which is consistent with earlier calculations of Chapter II.

The EGF of mean values, $\Omega(z)$ is then

$$\Omega(z) = V(z)e^{V(z)} = (e^z-1)e^{e^z-1}.$$

Due to the simple shape of $V(z)$, this 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 [**127**] based on the asymptotic expansion of the Bell numbers reveals the expected value and the standard deviation to be respectively asymptotic to

$$\frac{n}{\log n}, \qquad \frac{\sqrt{n}}{\log n}.$$

Similarly the exponential BGF of the number of blocks of size k is

$$e^{e^z + (u-1)\frac{z^k}{k!}},$$

out of which mean and variance can be derived once the asymptotic form of Bell numbers is known. □

EXAMPLE 9. *Root degree in Cayley trees.* For the class $\mathcal{T}$ of non–plane labelled trees (Cayley trees) the basic EGF equation is

$$T(z) = z\, e^{T(z)},$$

since non–planarity is taken into account by a set construction. In that case, the bivariate EGF satisfies $T(z,u) = ze^{uT(z)}$, and we find

$$\Omega(z) = zT(z)e^{T(z)} = (T(z))^2,$$

so that the mean root degree is, by Lagrange inversion,

$$2(1 - \frac{1}{n}) \sim 2.$$

A similar calculation shows that the fraction of trees with root degree k is asymptotically

$$\frac{e^{-1}}{(k-1)!}, \qquad k \geq 1,$$

which is a shifted Poisson law of rate 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 present in plane trees. □

▷ **6.** *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 IV imply that the number of components in a random alignment has expectation $\sim n/(e-1)$ and standard deviation $\Theta(\sqrt{n})$. ◁

▷ **7.** *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 Chapter IV imply that the image cardinality of a random surjection has expectation $n/(2\log 2)$ and standard deviation $\Theta(\sqrt{n})$. ◁

Postscript: Towards a theory of schemas. Let us look back and recapitulate some of the information gathered in pages 120—130 regarding the number of components in composite structures. The classes considered in the table below are compositions of two constructions, either in the unlabelled (**U**) or the labelled (**L**) 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.

Integer partitions, $\mathfrak{M} \circ \mathfrak{S}$ (**U**) $\exp\left(u\frac{z}{1-z}+\frac{u^2}{2}\frac{z^2}{1-z^2}+\cdots\right)$ $\sim \frac{\sqrt{n}\log n}{\pi\sqrt{2/3}}, \quad \Theta(\sqrt{n})$	Integer compositions, $\mathfrak{S} \circ \mathfrak{S}$ (**U**) $\left(1-u\frac{z}{1-z}\right)^{-1}$ $\frac{n}{2}, \quad \Theta(\sqrt{n})$
Set partitions, $\mathfrak{P} \circ \mathfrak{P}$ (**L**) $\exp\left(u\left(e^z-1\right)\right)$ $\sim \frac{n}{\log n} \quad \sim \frac{\sqrt{n}}{\log n}$	Surjections, $\mathfrak{S} \circ \mathfrak{P}$ (**L**) $\left(1-u\left(e^z-1\right)\right)^{-1}$ $\sim \frac{n}{2\log 2}, \quad \Theta(\sqrt{n})$
Permutations, $\mathfrak{P} \circ \mathfrak{C}$ (**L**) $\exp\left(u\log(1-z)^{-1}\right)$ $\sim \log n, \quad \sim \sqrt{\log n}$	Alignments, $\mathfrak{S} \circ \mathfrak{C}$ (**L**) $\left(1-u\log(1-z)^{-1}\right)^{-1}$ $\sim \frac{n}{e-1}, \quad \Theta(\sqrt{n})$

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, set partitions, and permutations) are associated with a greater variety of asymptotic regimes. 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.) Eventually, such facts can be organized into broad analytic schemas, as will be seen in Chapters IV–VII.

$\triangleright$ **8.** *Balls in bins: occupancy.* There are n balls thrown into m bins in all possible ways (m fixed). The bivariate EGF with z marking the number of balls and u marking the number of bins that contain k balls is

$$\left(e^z + (u-1)\frac{z^k}{k!}\right)^m.$$

Let m and n tend to infinity in such a way that $\frac{n}{m} = \alpha$, a fixed constant. The proportion of bins containing k elements tends (on average and in probability) to the limit

$$e^{-\alpha}\frac{\alpha^k}{k!}.$$

Thus a Poisson law of rate α describes the occupancy of bins in a random allocation. $\triangleleft$

$\triangleright$ **9.** *Distinct component sizes in sets.* Take the number of *distinct* block sizes and cycle sizes in set partitions and permutations. The bivariate EGF's are

$$\prod_{n=1}^{\infty}\left(1-u+ue^{z^n/n!}\right), \quad \prod_{n=1}^{\infty}\left(1-u+ue^{z^n/n}\right).$$

Find a comparable OGF for the number of distinct summands in an integer partition. $\triangleleft$

III. 4. Recursive parameters

In this section, we adapt the general methodology of previous sections in oder to treat parameters that are defined by recursive rules over structures that are themselves recursively specified. Typical applications concern trees and tree-like structures.

Consider a combinatorial class specified recursively

$$\mathcal{Y} = \mathfrak{K}\{\mathcal{Y}\}, \tag{24}$$

where $\mathfrak{K}$ is any composition of basic constructors and atoms. By distinguishing a finite set of configurations $\mathcal{X} \subset \mathcal{Y}$ considered to be "small" size, one can rephrase the specification (24) in the form

$$\mathcal{Y} = \mathcal{X} + \mathcal{V}, \qquad \mathcal{V} = \mathfrak{K}_+\{\mathcal{Y}\}. \tag{25}$$

A certain functional equation will then result for the counting GFs:

$$Y(z) = X(z) + V(z), \qquad V(z) = \Upsilon[Y(z)]. \tag{26}$$

For instance, general plane trees ($\mathcal{G}$) and Cayley trees ($\mathcal{T}$) admit the equivalent specifications

$$\begin{array}{rclrcl} \mathcal{G} &=& \mathcal{Z}\times\mathfrak{S}\{\mathcal{G}\}, & \quad \mathcal{G} &=& \mathcal{Z}+\mathcal{Z}\times\mathfrak{S}_{\geq 1}\{\mathcal{G}\} \\ \mathcal{T} &=& \mathcal{Z}\star\mathfrak{P}\{\mathcal{T}\}, & \quad \mathcal{T} &=& \mathcal{Z}+\mathcal{Z}\star\mathfrak{P}_{\geq 1}\{\mathcal{T}\}. \end{array}$$

In other words, the "small" objects of size 1 have been moved out of the original construction. In the case at hand, we individualize *leaves*[2] of trees.

[2] A leaf in a rooted tree is a node without descendents.

First, consider a parameter χ on $\mathcal{Y}$ that is inherited from a parameter β defined on $\mathcal{V}$ and another parameter ξ (the "initial conditions") especially defined on the "small" structures of $\mathcal{X}$, with $\mathcal{X}, \mathcal{V}$ as on the right of (25). Then, with the auxiliary variable u marking χ on $\mathcal{A}$ as well as ξ and β on $\mathcal{X}$ and $\mathcal{V}$, general principles lead to a functional relation,

$$Y_\chi(z,u) = X_\xi(z,u) + \Upsilon[Y_\beta(z,u)]. \tag{27}$$

Here, we have indicated the involved parameters by subscripts for clarity. For instance the parameter χ equal to "root–degree" of a tree is of this structural type, being inherited from $\xi \equiv 0$ on a leaf $\mathcal{Z}$ and $\beta \equiv 1$ on components of $\mathfrak{K}_+$.

What we have done when passing from ξ to χ is to examine the effect of *one* level of recursion. Assume next that χ and β are one and the same. In other words, there is a *unique* parameter χ defined *through recursion* on objects of the recursive class $\mathcal{Y}$, with β that singles out "initial conditions". An instance is now the total number χ of leaves in a tree: it is either defined to be 1, by a special case or else it is inherited additively as the sum of the values obtained from the root subtrees, cf (25). Indeed, if $\tau = \langle \rho, \tau_1, \ldots, \tau_r \rangle$ is a tree with root ρ and $r \geq 1$, one has

$$\chi(\tau) = \chi(\tau_1) + \cdots + \chi(\tau_r),$$

with χ coinciding with $\xi \equiv 1$ on atoms. With this identification of χ and β, the bivariate generating function $Y(z,u)$ becomes *implicitly defined* by a *functional equation* of the form

$$Y_\chi(z,u) = X_\xi(z,u) + \Upsilon[Y_\chi(z,u)].$$

Once the mechanism is clear, we may as well drop subscripts indicative of parameters and write

$$Y(z,u) = X(z,u) + \Upsilon[Y(z,u)]. \tag{28}$$

This stands out as a "deformation" of the usual univariate functional equation for the GF of $\mathcal{Y}$, to which it reduces when $u = 1$. With a natural extension of notations, we may even write symbolically a recursive specification for class–parameter pairs,

$$\langle \mathcal{Y}, \chi \rangle = \langle \mathcal{X}, \xi \rangle + \Upsilon[\langle \mathcal{Y}, \chi \rangle],$$

and simply apply the common translation mechanisms to get back (28). Naturally, similar considerations apply to vectorial parameters and/or to collections of mutually recursive combinatorial classes.

EXAMPLE 10. *Leaves in special varieties of 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 [**85**, Sec. 2.3] for a motivation to such questions.

Consider once more the class $\mathcal{G}$ of plane unlabelled trees, $\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\}$, enumerated by the Catalan numbers: $G_n = \frac{1}{n}\binom{2n-2}{n-1}$. The number $G_{n,k}$ of trees with n nodes and k leaves is to be determined. Let χ be the parameter "number of leaves" and $G(z,u)$ the associated bivariate OGF. In order to individuate leaves, rewrite the original specification of plane trees as

$$\mathcal{G} = \mathcal{Z} + (\mathcal{Z} \times \mathfrak{S}_{\geq 1}\{\mathcal{G}\}).$$

The parameter χ is additive; hence, to the defining relation, there corresponds termwise

$$G(z,u) = zu + \frac{zG(z,u)}{1 - G(z,u)}.$$

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 ballots problems.) The mean number of leaves then derives from the cumulative GF,

$$\Omega(z) = \partial_u G(z,u)|_{u=1} = \frac{1}{2}z + \frac{1}{2}\frac{z}{\sqrt{1-4z}},$$

so that the mean is $n/2$ exactly for $n \geq 2$. Also, the distribution is concentrated since the standard deviation is easily calculated to be $O(\sqrt{n})$.

In a similar vein, define binary plane trees by the equation,

$$\mathcal{B} = \mathcal{Z} + (\mathcal{B} \times \mathcal{Z}) + (\mathcal{Z} \times B) + (\mathcal{B} \times \mathcal{Z} \times \mathcal{B}), \tag{29}$$

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 (29) gives for the MGF $B = B(z, u_0, u_1, u_2)$ the functional equation

$$B = zu_0 + 2zu_1 + zu_2B^2,$$

by which 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 [**117**]. In particular, the mean number of nodes of each type is asymptotically:

$$\text{leaves: } \sim \frac{n}{4}, \qquad \text{1-nodes: } \sim \frac{n}{2}, \qquad \text{2-nodes: } \sim \frac{n}{2}.$$

Finally, 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 found to be asymptotic to ne^{-1}. □

▷ **10.** *Leaves and node-degree profile in simple varieties of trees.* 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.

More generally, 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 moments can be determined. $\triangleleft$

$\triangleright$ **11.** *Marking in functional graphs.* Consider the class $\mathcal{F}$ of finite mappings discussed in Chapter II:

$$\mathcal{F} = \mathfrak{P}\{\mathcal{K}\},\ \mathcal{K} = \mathfrak{C}\{\mathcal{T}\},\ \mathcal{T} = \{1\} \star \mathfrak{P}\{\mathcal{T}\}.$$

The translation on EGF's is

$$F(z) = e^{K(z)},\ K(z) = \log\frac{1}{1 - T(z)},\ T(z) = e^{T(z)}.$$

Here are bivariate EGF's for (i) the number of components, (ii) the number of maximal trees, (iii) the number of leaves:

$$(i)\ e^{uK(z)}, \qquad (ii)\ \frac{1}{1 - uT(z)}, \qquad (iii)\ \frac{1}{1 - T(z,u)} \quad \text{with} \quad T(z,u) = (u-1)z + ze^{T(z,u)}.$$

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 of the trivariate F involves the Stirling cycle numbers. $\triangleleft$

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.

We next turn to finer characteristics of trees, like 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

$$\sum_{\alpha\in\mathcal{A}} z^{|\alpha|} u^{\chi(\alpha)} = \sum_{\alpha\in\mathcal{A}} z^{|\alpha|} u^{|\alpha|+\chi(\alpha)},$$

that is,

$$A_\chi(z,u) = A_\xi(zu, u). \tag{30}$$

This is clearly a general mechanism: *a linear transformation on parameters induces a monomial substitution on the corresponding marking variables in MGFs.* We now put it to use in the analysis of path length in trees.

EXAMPLE 11. *Path length in trees.* Path length is an important "global" characteristic of trees classically defined as the sum of distances of all nodes to the root of the tree. (Distances are measured by the number of edges on the minimal connecting path.) For instance, when a tree is used as a data structure with nodes containing additional informations, 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 tree-sort, quicksort, and so on [**85, 130**].

From the definition of path length,

$$\lambda(\tau) := \sum_{\nu\in\tau} \mathrm{dist}(\nu, \mathrm{root}(\tau)),$$

there immediately results that

$$\lambda(\tau) = \sum_{\upsilon \text{ root subtree of } \tau} (\lambda(\upsilon) + |\upsilon|)\,. \tag{31}$$

(Distribute nodes in their corresponding subtrees: distances to the subtree roots must be corrected by 1; regroup terms.)

From this point on, we specialize the discussion to general plane trees (see Ex. 12 for more): $\mathcal{G} = \mathcal{Z}\mathfrak{S}\{\mathcal{G}\}$. Introduce momentarily the parameter $\mu(\tau) = |\tau| + \lambda(\tau)$. Then, one has from the inductive definition (31) and the general transformation rule (30):

$$G_\lambda(z,u) = \frac{z}{1 - G_\mu(z,u)} \qquad \text{and} \qquad G_\mu(z,u) = G_\lambda(zu,u).$$

In other words, $G(z,u) \equiv G_\lambda(z,u)$ satisfies a nonlinear functional equation of the difference type:

$$G(z,u) = \frac{z}{1 - G(uz,u)}. \tag{32}$$

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 $\Omega(z) := \partial_u G(z,1)$ satisfies

$$\Omega(z) = \frac{z}{(1 - G(z))^2} \left(zG'(z) + \Omega(z) \right),$$

which is a linear equation that solves to

$$\Omega(z) = z^2 \frac{G'(z)}{(1 - G(z))^2 - z} = \frac{z}{2(1-4z)} - \frac{z}{2\sqrt{1-4z}}$$

where $\delta = 1 - 4z$. Consequently, one has

$$\Omega_n = 2^{2n-1} - \binom{2n-2}{n-1},$$

where the sequence starting 1, 5, 22, 93, 386 for $n \geq 2$ constitutes *EIS* **A000346**. We thus have:

> *The mean path length of a random Catalan tree of size n is asymptotic to $2\sqrt{\pi n^3}$; in short: a branch in a random Catalan tree of size n has expected length of the order of $\sqrt{n}$.*

Under the uniform combinatorial model, trees thus tend to be somewhat imbalanced. □

The imbalance property found for random Catalan trees is a general phenomenon—it applies to binary Catalan and more generally to all simple varieties of trees. Ex. 12 below and Chapter V 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 [**120**], de Bruijn, Knuth and Rice [**37**], Kolchin [**90**], as well as Flajolet, and Odlyzko [**52**]. Figure 9 borrowed from [**130**] 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.)

▷ **12.** *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

$$\Omega(z) \equiv \partial_u \left(T(z,u)\right)_{u=1} = \frac{\phi'(T(z))}{\phi(T(z))}(zT'(z))^2,$$

from which coefficients can be extracted. ◁

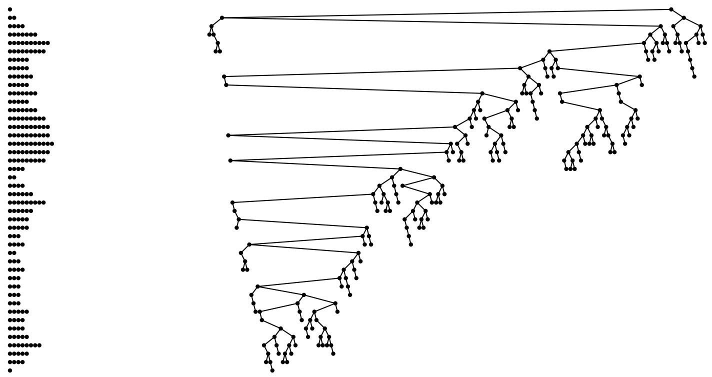


FIGURE 9. A random pruned binary tree of size 256 and its associated level profile: the histogram on the left displays the number of nodes at each level in the tree.

III. 5. "Universal" generating functions and combinatorial models

By a *universal* generating function, we mean a generating function in a number (possibly infinite) of variables that mark a homogeneous collection of characteristics of a combinatorial class. 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 universal MGF naturally entails very detailed knowledge on the enumerative properties of structures to which it is relative. Universal generating functions, given their expressive power, also make weighted models accessible to calculation, a situation that covers in particular Bernoulli trials and branching processes from classical probability theory.

As a basic example, consider the class of all words $\mathcal{W} = \mathfrak{S}\{\mathcal{A}\}$ 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

$$A(z, \mathbf{u}) = zu_1 + zu_2 + \cdots + zu_r,$$

and χ on $\mathcal{W}$ is clearly inherited from χ on $\mathcal{A}$. Thus, by the sequence rule, one has

$$W(z, \mathbf{u}) = \frac{1}{1 - z(u_1 + u_2 + \cdots + u_r)}, \tag{33}$$

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

$$[u_1^{n_1} u_2^{n_2} \cdots u_r^{n_r}]\,(u_1 + u_2 + \cdots + u^r)^n = \binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1! n_2! \cdots n_r}.$$

We are back to the usual multinomial coefficients.

▷ **13.** *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

$$4587555960000615321908476928639\underline{99999999999999}541244403999938467809152307136000000.$$

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 [**85**, p. 23] for a brief historical note. ◁

Next, consider permutations and the various lengths of their cycles. The MGF where u_1, u_2 mark 1-cycles and 2-cycles respectively is

$$\exp\left(u_1\frac{z}{1}+u_2\frac{z^2}{2}+\frac{z^3}{3}+\cdots\right).$$

By analogy, one is led to considering an MGF in infinitely many variables

$$U(z,\mathbf{u})=\exp\left(u_1\frac{z}{1}+u_2\frac{z^2}{2}+u_3\frac{z^3}{3}+\cdots\right). \tag{34}$$

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. Mathematically, an object like U in (34) is perfectly well defined: it suffices to consider $\mathbb{K}=\mathbb{C}(u)$ the field of fractions in infinitely many variables—any element of $\mathbb{K}$ involves only finitely many indeterminates; then calculate normally with formal power series of $\mathbb{K}[\![z]\!]$, assuming $\mathbb{K}$ as the coefficient field. Indeed, with the notion of formal convergence[3] defined in the appendix, one can take limits in $\mathbb{K}[\![z]\!]$ and write legitimately

$$\lim_{m\to\infty}\exp\left(\sum_{j=1}^{m}u_j\frac{z^j}{j}\right)=\exp\left(\lim_{m\to\infty}\sum_{j=1}^{m}u_j\frac{z^j}{j}\right)=U.$$

Henceforth, we shall keep in mind that verifications of formal correctness are always possible by returning to basic definitions.

Universal generating functions are often surprisingly simple to expand. For instance, the equivalent form of (34)

$$U(z,\mathbf{u})=e^{u_1z/1}\cdot e^{u_2z^2/2}\cdot e^{u_3z^3/3}\cdots$$

implies immediately that the number of permutations with n_1 cycles of size 1, n_2 of size 2, etc, is

$$\frac{n!}{c_1!\,c_2!\,\cdots c_n!\,1^{c_1}\,2^{c_2}\cdots n^{c_n}}, \tag{35}$$

provided $\sum jc_j=n$. 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

$$\exp\left(u_1\frac{z}{1!}+u_2\frac{z^2}{2!}+u_3\frac{z^3}{3!}+\cdots\right).$$

A formula analogous to (35), with j^{c_j} being replaced by $j!^{c_j}$ follows. Several examples of such "universal" generating functions are presented in Comtet's book; see [**28**], pages 225 and 233.

[3] In contrast, the quantity evocative of a generating function of words over an infinite alphabet

$$S\overset{!}{=}\left(1-z\sum_{j=1}^{\infty}u_j\right)^{-1}$$

cannot receive a sound definition as a element of the formal domain $\mathbb{K}[\![z]\!]$; for instance, the coefficient of z in the sequence of approximants would not even converge to an element of $\mathbb{K}$ equipped with the discrete topology.

▷ **14.** *Universal GFs for compositions and surjections.* The universal GF's 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 $n=\sum_j jn_j$ are given by

$$\binom{n_1+n_2+\cdots}{n_1,n_2,\ldots}, \qquad \frac{n!}{1!^{n_1}2!^{n_2}\cdots}\binom{n_1+n_2+\cdots}{n_1,n_2,\ldots}.$$

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 [**28**, Sec. 3.3] for a fair treatment of such polynomials. ◁

▷ **15.** *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 (assume $f(0)=g(0)=0$):

$$h_1 = f_1 g_1, \quad h_2 = f_2 g_1^2 + 2f_1 g_2, \ldots .$$

The general form, a mere avatar of the multinomial expansion, is known as Fàa di Bruno's formula [**28**, p. 137]. (Faà di Bruno (1825–1888) was canonized by the Catholic Church in 1988, albeit not for reasons related to his formula.) ◁

▷ **16.** *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:

$$\begin{array}{rclcrcl} B(z) &=& \dfrac{1}{A(-z)}, & & A(z) &=& \dfrac{1}{B(-z)}, \\ C(z) &=& z\dfrac{d}{dz}\log B(z), & & B(z) &=& \exp\displaystyle\int_0^z C(t)\,\frac{dt}{t}. \end{array}$$

Consequently, each of a_n, b_n, c_n is polynomially expressible in terms of any of the other quantities. (The connection coefficients again involve multinomials.) ◁

▷ **17.** *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 [**65**] has shown how to extract explicit expressions from such huge symmetric functions.] ◁

III. 5.1. Word models. The enumeration of words, or "sequences" as they are sometimes also called, constitutes a rich chapter of combinatorial analysis. Applications are to be found in classical probability theory and statistics [**33**] as well as in computer science [**139**] and mathematical models of biology [**149**]. We focus our attention here to problems that involve universal generating functions.

EXAMPLE 12. *Words and records.* Fix an alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$ and let $\mathcal{W} = \mathfrak{S}\{\mathcal{A}\}$ 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 12 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 $a_{i_1}, \ldots, a_{i_k}$ 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

$$\left(a_{i_1}(a_1 + \cdots + a_{i_1})^\star\right) \quad \cdots \quad \left(a_{i_k}(a_1 + \cdots + a_{i_k})^\star\right). \tag{36}$$

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 (36) 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

$$W(z, v, \mathbf{u}) = \prod_{s=1}^{r} \left(1 + zvu_s\,(1 - z(u_1 + \cdots + u_s))^{-1}\right), \tag{37}$$

the rationale being that, for arbitrary quantities y_s, one has

$$\sum_{k=0}^{r} \sum_{1 \le i_1 < \cdots < i_k \le r} y_{i_1} y_{i_2} \cdots y_{i_k} = \prod_{s=1}^{r} (1 + y_s).$$

We shall encounter more applications of (37) 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,

$$\Omega(z) \equiv \frac{\partial}{\partial v} W(z, v)\bigg|_{v=1} = \frac{z}{1 - rz} \sum_{j=1}^{r} \frac{1}{1 - (j-1)z}.$$

Thus, by partial fraction expansion, the mean number of records in $\mathcal{W}_n$ (whose cardinality is r^n) has value

$$\mathbb{E}_{\mathcal{W}_n}(\#\text{ records}) = \mathrm{H}_r - \sum_{j=1}^{r-1} \frac{(j/r)^n}{r - j}. \tag{38}$$

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). □

EXAMPLE 13. *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 $\mathcal{S} \subseteq \mathcal{W} = \mathfrak{S}\{\mathcal{A}\}$ have "universal" GF

$$S(z, u_1, \dots, u_r) = \sum_{w \in S} z^{|w|} u_1^{\chi_1(w)} \cdots u_r^{\chi_r(w)},$$

where $\chi_j(w)$ is the number of occurrences of letter a_j in w. Then one has

$$S(z, \lambda_1, \dots, \lambda_r) = \sum_{w \in S} z^{|w|}\pi(w),$$

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 universal 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 dice. 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 $0 \le \lambda_j \le 1$. 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 $[z^n]S(z, \lambda_1, \dots, \lambda_r)$ 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 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 language describing the events of interest (with varying n) is, as seen in Chapter I,

$$\mathcal{S} = \mathfrak{S}_{<\ell}\{H\}\mathfrak{S}\{T\mathfrak{S}_{<\ell}\{H\}\}.$$

Its universal 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. (Fellers' book [**43**, p. 322–326] offers for instance a classical discussion of the problem.)

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 (38): one finds by logarithmic differentiation of (37)

(39)
$$\mathbb{E}_{\mathcal{W}_n}(\text{\# records}) = [z^n]\Omega(z) \qquad \text{where} \quad \Omega(z) = \frac{z}{1-z}\sum_{j=1}^{r}\frac{p_j}{1-z(p_1+\cdots+p_{j-1})}.$$

The cumulative GF $\Omega(z)$ in (39) 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 (8):

The mean number of records in a random word of length n with non uniform letter probabilities p_j satisfies asymptotically

$$\mathbb{E}_{\mathcal{W}_n}(\text{\# records}) \sim \sum_{j=1}^{r}\frac{p_j}{p_j+p_{j+1}+\cdots+p_r}.$$

This relation and similar ones were obtained by Burge [**25**]; analogous ideas may serve to analyse the sorting algorithm *Quicksort* under equal keys [**128**] as well as the hybrid data structures of Bentley and Sedgewick; see [**12, 27**]. □

Similar considerations apply to weighted EGFs of words. For instance, the probability of having attained a complete coupon collection in case a company issues coupon j with probability p_j (with $1 \le j \le r$), is

$$n![z^n]\prod_{j=1}^{r}(e^{p_j z}-1).$$

The probability that all coupons are different at time n is

$$n![z^n]\prod_{j=1}^{r}(1+p_j z),$$

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.

III. 5.2. Tree models. We examine here two important universal GFs associated with tree models; these provide valuable informations concerning the degree profile and the level profile of trees, while being tightly coupled with an important class of stochastic processes, 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

$$Y(z) = z\phi(Y(z)), \tag{40}$$

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

$$\phi(w) = \sum_{\omega\in\Omega} u^\omega, \qquad \phi(w) = \sum_{\omega\in\Omega} \frac{u^\omega}{\omega!}, \tag{41}$$

where $\Omega \subseteq \mathbb{N}$ is the set of allowed node degrees. Meir and Moon in an important paper [**104**] have described some common properties of tree families that are determined by the Axiom (40). (For instance mean path length is of order $n\sqrt{n}$ and height is $O(\sqrt{n})$.) Following these authors, we shall call *simple variety of trees* any class whose counting GF is defined by an equation of type (40). For each of the two cases of (41), we shall write

$$\phi(w) = \sum_{j=0}^{\infty} \phi_j w^j. \tag{42}$$

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.4 (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"* (42) *is*

$$Y_{n;n_0,n_1,n_2,\ldots} = \omega_n \cdot \frac{1}{n}\binom{n}{n_0, n_1, n_2, \ldots} \phi_0^{n_0}\phi_1^{n_1}\phi_2^{n_2}\cdots . \tag{43}$$

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 the standard multinomial expansion applies, yielding (43).

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 shows the logical equivalence between the general tree counting result of Proposition III.4 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 (43) provides a combinatorial proof of the Lagrange inversion theorem. Such direct derivations have been proposed by Raney [**119**] 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 [**40**]).

The next example demonstrates the usefulness of universal generating functions for investigating the profile of trees.

EXAMPLE 14. *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 universal 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:

$$Y(z, \mathbf{u}) = zu_0\phi\left(zu_1\phi\left(zu_2\phi\left(zu_3\phi(\cdots)\right)\right)\right). \tag{44}$$

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 universal GF is the continued fraction:

$$Y(z, \mathbf{u}) = \cfrac{u_0 z}{1 - \cfrac{u_1 z}{1 - \cfrac{u_2 z}{1 - \cfrac{u_3 z}{\ddots}}}}.$$

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.5 (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* (42) *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 $\sum_j n_j = n$.

(Note that one must always have $n_0 = 1$ for a single tree; the general formula with $n_0 \neq 1$ gives the level profile of forests.)

For instance, the counts for general Catalan trees and for Cayley trees are respectively

$$\binom{n_0+n_1-1}{n_1}\binom{n_1+n_2-1}{n_2}\binom{n_2+n_3-1}{n_3}\cdots, \qquad \frac{(n-1)!}{n_0!n_1!n_2!\cdots}n_0^{n_1}n_1^{n_2}n_2^{n_3}\cdots.$$

The first of these enumerative results is due to Flajolet [**44**] and it places itself within a general combinatorial theory of continued fractions; the second one is due to Rényi and Szekeres [**120**] who developed such a formula in the context of a deep study of the distribution of height in random Cayley trees. □

▷ **18.** *"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

$$G(z,q) = \cfrac{z}{1-\cfrac{zq}{1-\cfrac{zq^2}{\ddots}}}, \qquad T(z,q) = ze^{zqe^{zq^2e^{\cdot^{\cdot^{\cdot}}}}},$$

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.) ◁

It is interesting to compare the counting results provided by universal generating functions. In a way, they contain "all" the information regarding a random object, but in a form that is not necessarily synthetic enough. Thus universal formulæ appear as offering a perspective that complements the analysis of single parameters presented in earlier sections. As we show next, they can also be used to reduce branching processes to combinatorial models.

EXAMPLE 15. *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:

$$\pi(\tau) = \prod_{j=0}^{\infty} \lambda_j^{\chi_j(\tau)}, \tag{45}$$

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

$$\mathbb{P}_{\mathcal{G}_n,\Lambda}(\tau) = \frac{\pi(\tau)}{\sum_{|\tau|=n}\pi(\tau)}. \tag{46}$$

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 (45) and (46) covers any simple variety family of trees: just replace each λ_j by the quantity ϕ_j given by the "generator' (42) 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 (46).

(ii) Let Λ^{**} be defined by $\lambda_j^{**} = \theta^j\lambda_j$ for some nonzero constant θ. Then the weight corresponding to Λ^{**} satisfies $\pi^{**}(\tau) = c^{|\tau|-1}\pi(w)$, 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:

$$\phi(w) \mapsto \phi^*(w) = c\phi(w) \qquad \text{and} \qquad \phi(w) \mapsto \phi^{**}(w) = \phi(\theta w). \tag{47}$$

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.

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 [**7**]. 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 (45) 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 (47), there result 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 universal 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.* $\square$

The reduction of combinatorial tree models to branching processes has been pursued most notably by the "Russian School": see especially the books by Kolchin [**90, 91**] and references therein. Conversely, symbolic-combinatorial methods may be viewed as a systematic way of obtaining equations relative to characteristics of branching processes. We do not proceed further along these lines as this would take us outside of the scope of the present book.

$\triangleright$ **19.** *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 with offspring probabilities that 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 $p_j = e^{-\lambda}\lambda^j/j!$. $\triangleleft$

III. 6. Additional constructions

We discuss here additional constructions already examined in earlier chapters, namely pointing and substitution (Section III. 6.1) as well as order constraints (Section III. 6.2) on the one hand, implicit structures (Section III. 6.3) 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. 6.4, 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. 6.1. 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 (12) and (22). In particular $z_0 = z$ marks size, and z_k marks the component j of the multiparameter k. 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 θ_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.

▷ **20.** *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. ◁

Similarly, the substitution $x \mapsto S(\mathbf{z})$ in a GF $F(z)$, 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 z 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 normally easily obtained by returning to the combinatorial definition of generating functions as "reduced images" of combinatorial classes.

EXAMPLE 16. *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

$$\mathcal{R}_1 = \{(x,y) \mid 1 \leq x \leq y\}, \quad \mathcal{R}_2 = \{(x,y) \mid 1 \leq y \leq 2x\},$$

will 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 an equation of the form

$$F(z,u) = f(zu) + \left(\mathcal{L}\left[F(z,u)\right]\right)_{u\mapsto zu}, \tag{48}$$

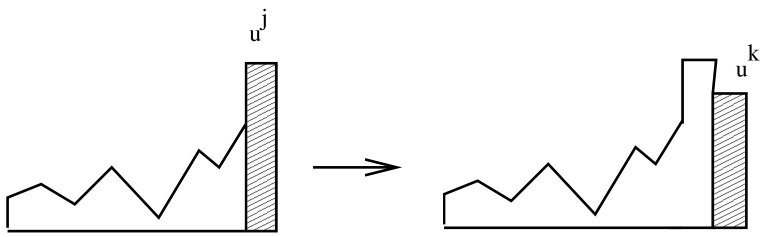


FIGURE 10. The technique of "adding a slice" for enumerating constrained compositions.

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

$$\mathcal{L}[u^j] := \sum_{(j,k)\in\mathcal{R}} u^k. \tag{49}$$

In effect, Equation (48) 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 (49) expresses. In particular, $F(z,1)$ gives back the enumeration of $\mathcal{F}$–objects irrespective of the size of the first column.

For a rule $\mathcal{R}$ that is "simple enough", the basic equation (48) 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 then the same as partitions. Since

$$L[u^j] = u^j + u^{j+1} + u^{j+2} + \cdots = \frac{u^j}{1-u},$$

the function $F(z,u)$ satisfies a functional equation involving a substitution,

$$F(z,u) = \frac{zu}{1-zu} + \frac{1}{1-zu}F(z,zu). \tag{50}$$

This relation iterates: *any linear functional equation of the substitution type*

$$\phi(u) = \alpha(u) + \beta(u)\phi(\sigma(u))$$

is solved formally by

$$\phi(u) = \alpha(u) + \beta(u)\alpha(\sigma(u)) + \beta(u)\beta(\sigma(u))\alpha(\sigma^{\langle 2\rangle}(u)) + \cdots, \tag{51}$$

where $\sigma^{\langle j\rangle}(u)$ designates the jth iterate of u.

Returning to compositions into increasing summands, that is, partitions, the turnkey solution (51) gives, upon iterating on the second argument and with the first argument being treated as a parameter:

$$F(z,u) = \frac{zu}{1-zu} + \frac{z^2u}{(1-zu)(1-z^2u)} + \frac{z^3u}{(1-zu)(1-z^2u)(1-z^3u)} + \cdots. \tag{52}$$

Equivalence with the alternative form

$$F(z,u) = zu + \frac{z^2u^2}{1-z} + \frac{z^3u^3}{(1-z)(1-z^2)} + \frac{z^4u^4}{(1-z)(1-z^2)(1-z^3)} \cdots \tag{53}$$

is then easily verified from (50) 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 (53) 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 partitions whose smallest summand is 1 and whose largest summand is k. (In passing, the equality between (52) and (53) is a shallow but curious identity that is quite typical of the area.)

This same method has been applied in [**54**] 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 (51), 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 $\mathcal{R}_2$:

$$\begin{aligned} F(z,1) \quad &= \quad \frac{\sum_{j\ge1}(-1)^{j-1}Q_{j-1}(z)z^{2^{j+1}-j-2}}{\sum_{j\ge0}(-1)^{j}Q_{j}(z)z^{2^{j+1}-j-2}} \\ &\quad\ \text{where} \quad Q_j(z) = (1-z)(1-z^3)(1-z^7)\cdots(1-z^{2^j-1}). \end{aligned}$$

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 [**54**] for details including very precise asymptotic estimates and Tangora's paper for relations to algebraic topology. □

The reason for presenting this method in some detail is that it is very general. It has been in particular employed to derive a number of original enumerations of polyominos by area, a topic of interest in some branches of statistical mechanics: for instance, the book by Janse van Regsburg [**144**] discusses many applications of such lattice models to polymers and vesicles. See Bousquet-Mélou's review paper [**23**] for a methodological perspective. Some of the origins of the method point to Pólya in the 1930's, see [**114**], and independently to Temperley [**141**, pp. 65–67].

▷ **21.** *Carlitz compositions.* Let $\mathcal{K}$ be the class of compositions such that pairs of adjacent summands are always distinct. These can be generated by the operator $\mathcal{L}[u^j] = u(1-u)^{-1} - u^j$, from which the OGF follows. Alternatively, one may start from Smirnov words (p. 152 below) and effect the substitution $v_j \mapsto z^j$, so that

$$K(z) = \left(1 - \sum_{j=1}^{\infty}\frac{z^j}{1+z^j}\right)^{-1}.$$

For maximal summand $\le r$, replace ∞ by r in the formula above. (Such compositions have been introduced by Carlitz in 1976; see the paper by Knopfmacher and Prodinger [**82**] for early references and asymptotic properties.) ◁

III. 6.2. 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

$$\mathcal{A} = (\mathcal{B}^{\Box} \star \mathcal{C})$$

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

$$A(z,\mathbf{u}) = \int_0^z (\partial_t B(t,\mathbf{u}))\cdot C(t,\mathbf{u})\,dt.$$

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 11. Local order patters in a permutation and the four types of nodes in the corresponding increasing binary tree.

To illustrate this multivariate extension, we shall consider a quadrivariate statistic on permutations.

EXAMPLE 17. *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 summarized in the first two columns of Figure 11. The correspondence with binary increasing trees described in Example 16 of Chapter II then shows the following: peaks and valleys correspond to binary nods 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 markers for the number of nodes of each type, as summarized in Figure 11. 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

$$I(z, \mathbf{u}) = \frac{\delta}{u_2} \frac{v_1 + \delta \tan(z\delta)}{\delta - v_1 \tan(z\delta)} - \frac{v_1}{u_2}, \tag{54}$$

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

$$I = u_0 z + u_0(u_1 + u_1') \frac{z^2}{2!} + u_0((u_1 + u_1')^2 + 2u_0 u_2) \frac{z^3}{3!},$$

which agrees with the small cases. This calculation is consistent with what has been found in Chapter II regarding the EGF of all permutations and of alternating permutations,

$$\frac{1}{1-z}, \quad \tan(z),$$

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 the BGF of Eulerian numbers (61) derived below by other means.

By specialization of the tetravariate 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. □

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% worse than in a perfect binary tee. This

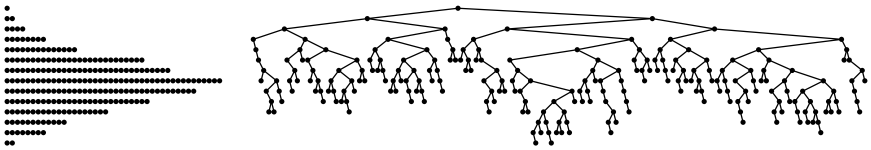


FIGURE 12. The level profile of a random increasing binary tree of size 256. (Compare with Figure 9 for binary trees under the uniform Catalan statistic.)

fact applies to binary search trees and it justifies the performance of such trees to be quite good when applied to random data [**86, 102, 130**] or subjected to randomization [**123**].

III. 6.3. 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,

$$\mathcal{G} = \mathfrak{P}\{\mathcal{K}\},$$

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

$$K(z,u) = \log\left(1 + \sum_{n=1}^{\infty} (1+u)^{n(n-1)/2} \frac{z^n}{n!}\right). \tag{55}$$

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 (55) 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 $\mathbb{C}(u)[\![z]\!]$):

$$\begin{aligned}\log\left(1+\sum_{n=1}^{\infty}(1+u)^{n(n-1)/2}\frac{z^n u^{-n}}{n!}\right) &= \frac{1}{u}W_{-1}(z)+W_0(z)+\cdots \\ &= \frac{1}{u}\left(T-\frac{1}{2}T^2\right)+\left(\frac{1}{2}\log\frac{1}{1-T}-\frac{1}{2}T-\frac{1}{4}T^2\right)+\cdots,\end{aligned} \tag{56}$$

with $T \equiv T(z)$. See Temperley's early works [**140, 141**] as well as the "giant paper on the giant component" [**77**] and the paper [**55**] for direct derivations that eventually constitute analytic alternatives to Wright's combinatorial approach.

EXAMPLE 18. *Smirnov words.* Following the terminology of Jackson and Goulden [**68**], a Smirnov word is a word that has no consecutive equal letters. Let $\mathcal{W} = \mathfrak{S}\{\mathcal{A}\}$ be the set of words over the alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$ 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 that we represent figuratively as

$$\mathcal{W} = \mathcal{S}\big[a_1 \mapsto \mathfrak{S}_{\geq 1}\{a_1\},\ \ldots\ , a_r \mapsto \mathfrak{S}_{\geq 1}\{a_r\}\big].$$

There results the relation

$$W(v_1, \ldots, v_r) = S\left(\frac{v_1}{1-v_1},\ \ldots\ , \frac{v_r}{1-v_r}\right). \tag{57}$$

This relation determines the MGF $S(v_1, \ldots, v_r)$ *implicitly*. Indeed, since the inverse function of $v/(1-v)$ is $v/(1+v)$, one finds

$$S(v_1, \ldots, v_r) = W\left(\frac{v_1}{1+v_1},\ \ldots\ , \frac{v_r}{1+v_r}\right). \tag{58}$$

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 (58) 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 (58) 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

$$\frac{1}{1 - r\dfrac{z\frac{1-z^m}{1-z}}{1+z\frac{1-z^m}{1-z}}} = \frac{1-z^{m+1}}{1-rz+(r-1)z^m}.$$

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 different run length upperbounds depending on each particular letter. For instance, this topic is pursued in several works of Karlin and coauthors (see, e.g., [**106**]), themselves motivated by biological applications. □

III. 6.4. Inclusion-Exclusion. Inclusion-exclusion is a familiar type of reasoning rooted in elementary mathematics. We re-examine it here in the perspective of multivariate generating functions, where it essentially reduces to a combined use of substitution and implicit definitions.

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 or a discrete probability measure on $\mathcal{E}$. The more general formula

$$|A \cup B| = |A| + |B| - |AB| \qquad \text{where} \quad AB := A \cap B,$$

follows immediately. What is called the *inclusion-exclusion principle* or *sieve formula* is the following multivariate generalization, for an arbitrary family $A_1, \ldots, A_r \subset \mathcal{E}$:

(59)

$$\begin{aligned} |A_1 \cup \cdots \cup A_r| &\equiv \left|\mathcal{E} \setminus (\overline{A}_1\overline{A}_2\cdots\overline{A}_r)\right| \qquad \text{where} \quad \overline{A} := \mathcal{E} \setminus A \\ &= \sum_{1\le i\le r} |A_i| - \sum_{1\le i_1<i_2\le r} |A_{i_1}A_{i_2}| + \cdots + (-1)^{r-1}|A_1A_2\cdots A_r|. \end{aligned}$$

The easy proof by induction results from elementary properties of the boolean algebra formed by the subsets of $\mathcal{E}$; see, e.g., [**28**, Ch. IV].) An alternative formulation results from setting $B_j = \overline{A}_j$, $\overline{B}_j = A_j$:

(60) $$|B_1B_2\cdots B_r| = |\mathcal{E}| - \sum_{1\le i\le r} |\overline{B}_i| + \sum_{1\le i_1<i_2\le r} |\overline{B}_{i_1}\overline{B}_{i_2}| - \cdots + (-1)^r|\overline{B}_1\overline{B}_2\cdots\overline{B}_r|.$$

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 members of the $\overline{B}_j$).

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 (60) 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 couting 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.

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 aribitrary 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 then 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 (60), but with the signs "wrong", i.e., all positive.

Introduce now the variable v to mark the distinguished fixed points in objects of $\mathcal{Q}$. The exponential BGF is then

$$Q(z, v) = e^{vz} \frac{1}{1-z}.$$

Let $P(z, u)$ 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 (60) 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 objets 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.

The book by Goulden and Jackson [**68**, 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.

EXAMPLE 19. *Rises and ascending runs in permutations.* A *rise* in a permutation $\sigma = \sigma_1 \cdots \sigma_n$ is a pair of consecutive elements σ_i, σ_{i+1} satisfying $\sigma_i < \sigma_{i+1}$. The problem is to determine the number $A_{n,k}$ of permutations of size having exactly k rises together with the BGF $A(z,u)$.

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 constuction applied to atoms ($\mathcal{Z}$) and clusters ($\mathcal{C}$) as

$$\mathcal{B} = \mathfrak{S}\{\mathcal{Z} + \mathcal{C}\}, \qquad \text{where} \quad \mathcal{C} = \mathfrak{P}_{\geq 2}\{\mathcal{Z}\},$$

since a cluster is an ordered sequence, or equivalently a set, having furthermore 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

$$A(z,u) = \frac{u-1}{u - e^{z(u-1)}}. \tag{61}$$

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 [**28**] or [**71**]. (From Eq. (61), permutations without rises are enumerated by $B(z,-1) = e^z$, an altogether obvious result.)

Moments derive easily from an expansion of (61) at $u = 1$, which gives

$$A(z, u) = \frac{1}{1-z} + \frac{1}{2}\frac{1}{(1-z)^2}(u-1) + \frac{1}{12}\frac{z^3(2+z)}{(1-z)^3}(u-1)^2 + \cdots .$$

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 order ℓ) is a sequence of consecutive elements $\sigma_i\sigma_{i+1}\cdots\sigma_{i+\ell-1}$ such that $\sigma_i < \sigma_{i+1} < \cdots < \sigma_{i+\ell-1}$. An inclusion-exclusion similar to the one seen above shows that the BGF of the number of ascending runs of order ℓ in permutations is

$$(62)\quad B^{\langle \ell \rangle}(z, u) = \left(\frac{1}{1-(z+(e^{zv}-e_{\ell-1}(zv))/v^{\ell-1})}\right)_{v=u-1}, \qquad e_r(z) := \sum_{j=0}^{r}\frac{z^j}{j!}.$$

(Rises correspond to $\ell = 2$.)

The BGF (62) can be exploited to determine quantitative information on long runs in permutations. First, an expansion at $u = 1$ shows that the mean number of ascending runs is $(n - \ell)/\ell!$ exactly, as soon as $n \ge \ell$. This entails that, if $n = o(\ell!)$, the probability of finding an ascending run of order ℓ 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)

$$\mathbb{P}(X \ge 1) = \mathbb{E}([\![X \ge 1]\!]) = \mathbb{E}(\min(X, 1)) \le \mathbb{E}(X).$$

An inequality in the converse direction results from the second moment method. In effect, the variance of the number of ascending runs 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. Thus, by Chebyshev's inequalities, concentration of distribution holds as long as ℓ is such that $\ell! = o(n)$. In this case, with high probability (i.e., with probability tending to 1 as n tends to ∞), there are many ascending runs of order ℓ.

What has been found here is a fairly sharp threshold phenomenon:

> *In a random permutation of size n, with high probability, an ascending run of order ℓ is present if $\ell! = o(n)$ but not present if $n = o(\ell!)$. Let $\ell_0(n)$ be the largest integer such that $\ell_0! \le n$, so that $\ell_0(n) \sim (\log n)/\log\log n$. Then, with high probability, there is no ascending run of length $\ell_0 + 1$ but at least one ascending run (and in fact many) of length $\ell_0 - 1$.*

□

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 [**68**, 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 [**48**] (by combinatorial methods) and Devroye [**39**] (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 [**97, 146**] Solving a problem open for over 20 years, Baik, Deift, and Johansson [**8**] have eventually determined its limiting distribution. (The undemanding survey by Aldous and Diaconis [**2**] discusses some of the background of this problem.)

▷ **22.** *Increasing subsequences in permutations.* This exercise is based on Lifschitz and Pittel's work [**96**] who were amongst the first to provide nontrivial lower bounds by elementary arguments. Let $\lambda(\sigma)$ be the length of the longest increasing subsequence in permutation σ and $\kappa(\sigma)$ the number of increasing subsequences (of all length). Then, the EGF $K(z)$ of cumulated values of κ satisfies

$$K(z) := \sum_{\sigma\in\mathcal{P}} \kappa(\sigma)\frac{z^{|\sigma|}}{|\sigma|!} = \frac{1}{1-z}e^{z/(1-z)}, \qquad [z^n]K(z) = \sum_{k=0}^{n}\binom{n}{k}\frac{1}{k!} \sim \frac{1}{2\sqrt{\pi e}}n^{-1/4}e^{2\sqrt{n}}.$$

Since $2^\lambda \le \kappa$ and the exponential function is convex, the expected length of the longest increasing subsequence has upper bound $c\sqrt{n} + o(n^{1/2})$ with $c = 2/\log 2 \doteq 2.88539$. ◁

EXAMPLE 20. *Patterns in words.* Take the set of all words $W = \mathfrak{S}\{\mathcal{A}\}$ over a finite alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$. A pattern $\mathfrak{p} = p_1p_2\cdots p_k$, 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* 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 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. 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:

$$\begin{array}{l|l}
\overbrace{\mathtt{a\ a\ a\ a\ a}}^{z^5} & z^5 \\
\quad\ \mathtt{a\ a\ a}\overbrace{\mathtt{a\ a}}^{z^2} & \times\,(z + \underline{\mathbf{z^2}} + z^3 + z^4) \\
\qquad\quad \mathtt{a}\overbrace{\mathtt{a\ a\ a\ a}}^{z^4} & \times\,(z + z^2 + z^3 + \underline{\mathbf{z^4}}).
\end{array}$$

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 $(1-\widehat{c}(z))^{-1}$). 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. □

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. We refer globally to Szpankowski's book [**139**], where such questions are treated systematically and in great detail.

▷ **23.** *Moments of number of occurrences.* Observe that the derivatives of $X(z,v)$ at $v = 0$ give access to the factorial moments of the number of occurrences of a pattern. Evaluate the mean and variance of the number of occurrences. (Hint: set $z \mapsto z/r$ and expand the rational fractions involved near $z = 1$.) ◁

▷ **24.** *Words with fixed repetions.* Let $W^{\langle k\rangle}(z) = [u^k]W(z,u)$ be the OGF of words containing a pattern exactly k times. There exist two functions $\lambda(z), \mu(z)$ such that $W^{\langle k\rangle}(z) = \lambda(z)\mu(z)^k$ for any $k \geq 1$. ◁

▷ **25.** *Patterns in Bernoulli sequences.* Work out the BGF of the number of occurrences of a pattern in a random string with nonuniform letter probabilities $p_j = \mathbb{P}(a_j)$. (Hint: one needs to define a weighted correlation polynomial $c(z)$.) ◁

▷ **26.** *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 $uB(zu)$ 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 $u \mapsto u-1$ 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 [**57**]. ◁

III. 7. 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 GF's can then be constructed by the symbolic method in its univariate version.

III. 7.1. 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, safe 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 $\leq b$,

$$P^{\langle b\rangle}(z) = \exp\left(\frac{z}{1} + \frac{z^2}{2} + \cdots + \frac{z^b}{b}\right),$$

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

$$\mathcal{W} = \mathfrak{S}\{a\} \cdot \mathfrak{S}\{b\,\mathfrak{S}\{a\}\},$$

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 $\mathcal{Y} = \mathfrak{S}\{a\}$. 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 [**70**].

▷ **27.** *Smallest components.* The EGF of permutations with smallest cycle of size $> b$ is

$$\frac{\exp(-\frac{z}{1} - \frac{z^2}{2} - \frac{z^b}{b})}{1 - z}.$$

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 [**112**]. ◁

III. 7.2. 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 GF's 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. 16)

Consider for instance general plane trees defined by

$$\mathcal{G} = \mathcal{N} \times \mathfrak{S}\{\mathcal{G}\} \qquad \text{so that} \qquad G(z) = \frac{z}{1 - G(z)}.$$

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

$$\mathcal{G}^{[0]} = \mathcal{N},\ \mathcal{G}^{[h+1]} = \mathcal{N} \times \mathfrak{S}\{\mathcal{G}^{[h]}\}.$$

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

$$G^{[h]}(z) = \cfrac{z}{1 - \cfrac{z}{1 - \cfrac{z}{\ddots \over 1 - z}}},$$

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 [**37**] have determined the average height of a general plane tree to be $\sim \sqrt{\pi n}$.

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[4]. However, using complex asymptotic methods and singularity analysis, Flajolet and Odlyzko [**52**] 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

$$\mathcal{T} = \{1\} \star \mathfrak{P}\{\mathcal{T}\} \qquad \text{so that} \qquad T(z) = ze^{T(z)}.$$

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.

[4]These polynomials are exactly the much studied Mandelbrot polynomials whose behaviour in the complex plane gives rise to extraordinary graphics.

$\triangleright$ **28.** *Height in general Catalan trees.* The model of height in general plane trees can be solved algebraically. The OGF $G^{[h]}(z)$ of trees of height $\leq h$ is of the form

$$z\frac{F_{h+1}(z)}{F_{h+2}(z)},$$

where the F's are the Fibonacci polynomials

$$F_0(z) = 0,\ F_1(z) = 1,\ F_{h+2}(z) = F_{h+1}(z) - zF_h(z).$$

Express the F_h in terms of $G(z)$ itself. Find explicit forms for the distribution of height in trees of $\mathcal{G}_n$ by means of Lagrange inversion. [This treatment is due to De Bruijn, Knuth, and Rice [**37**].] $\triangleleft$

III. 7.3. 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 $f^{[h]}(z)$ 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.

$\triangleright$ **29.** *Hierarchical partitions.* Let $\varepsilon(z) = e^z - 1$. Find a combinatorial interpretation for

$$\varepsilon(\varepsilon(\cdots(\varepsilon(z)))) \qquad (h \text{ times}).$$

(Such structures show up in statistical classification theory.) $\triangleleft$

$\triangleright$ **30.** *Balanced trees.* Balanced structures lead to counting GF's 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

$$Z^{[h+1]}(z) = Z^{[h]}(z^2 + z^3) = (Z^{[h]}(z))^2 + (Z^{[h]}(z))^3.$$

Express it in terms of the iterates of $\sigma(z) = z^2 + z^3$.

Find the OGF of mean values of the number of internal nodes in such trees. $\triangleleft$

$\triangleright$ **31.** *Extremal statistics in random mappings.* Find the EGF's relative to the largest cycle, longest branch, and diameter of functional graphs. Do the same for the largest tree, largest component. [Hint: see [**53**] for details.] $\triangleleft$

$\triangleright$ **32.** *Deep nodes in trees.* Find the GF of mean values of the number of nodes at maximal depth in a general plane tree and in a Cayley tree. $\triangleleft$

III. 8. Notes

Multivariate generating functions are a common tool from classical combinatorial analysis. Comtet's book [**28**] is once more an excellent source of examples. A systematization of multivariate generating functions for inherited parameters is given in the book by Jackson and Goulden [**68**].

In contrast generating functions for averages seemed to have received relatively little attention before the advent of digital computers and the analysis of algorithms. Many important techniques are implicit in Knuth's books, especially [**85, 86**]. Wilf discusses related issues in his book [**153**] and the paper [**151**]. Early systems specialized to tree algorithms have been proposed by Flajolet and Steyaert in the beginning 1980's [**45, 59, 60, 138**]; see also Berstel and Reutenauer's work [**15**]. Some of the ideas developed there (viewing generating functions of averages as images of combinatorial structures with multiplicities attached) took their inspiration from the well established treatment of formal power series in noncommutative indeterminates (that can be seen as words with multiplicities attached), see Eilenberg's book [**41**] or the proceedings edited by Berstel [**126**].

The global framework of constructible structures affords a neat structural categorization of parameters of combinatorial objects—additively inherited parameters, recursive parameters, largest components, and height. This approach becomes especially powerful when examined in the light of asymptotic properties of structures. In addition, the principles developed here render the analysis of a large class of combinatorial parameters entirely systematic. This includes complexity measures for a closed class of programmes and data structures. Several computations in this area can then even be automated with the help of computer algebra systems [**56, 158**].

APPENDIX A

Auxiliary Results & Notions

1. Arithmetical functions. A general reference for this section is Apostol's book [**5**].

The function $\varphi(k)$ is the *Euler totient function* and it 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 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 *Moebius 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 *Dirichlet generating functions* (DGF). Let $(a_n)_{n\geq 1}$ be a sequence; its Dirichlet series is 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!)

Equation (1) implies elementarily that

$$M(s) := \sum_{n\geq 1} \frac{\mu(n)}{n^s} = \prod_{p\in\mathcal{P}} \left(1 - \frac{1}{p^s}\right) = \frac{1}{\zeta(s)}. \tag{2}$$

The coefficients $\mu(n)$ are known as the Moebius coefficient (or Moebius function). They satisfy

$$\mu(n) = (-1)^r \qquad \text{if } n = p_1 p_2 \cdots p_r \text{ for distinct primes } p_j,$$

and $\mu(n) = 0$ whenever n is divisible by a square.

Finally, if $(a_n), (b_n), (c_n)$ have DGF $\alpha(s), \beta(s), \gamma(s)$, then one has the equivalence

$$\alpha(s) = \beta(s)\gamma(s) \quad \iff \quad a_n = \sum_{d \mid n} b_d c_{n/d}.$$

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 *Moebius inversion*. This relation is used in the enumeration of irreducible polynomials (Section I. 6.2).

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 $\mathbb{S} = \mathbb{Z}_{>0} \cup \{+\infty\}$ and $s_0 = +\infty$, $\mathbb{S} = \mathbb{R}$ and s_0 any point in $\mathbb{R}$, $\mathbb{S} = \mathbb{C}$ or a subset of $\mathbb{C}$, and so on. Two functions ϕ and g from $\mathbb{S} \setminus \{s_0\}$ to $\mathbb{C}$ are given.

— $\mathcal{O}$*–notation*: write

$$\phi(s) \underset{s \to s_0}{=} \mathcal{O}(g(s))$$

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) $c > 0$, there exists a neighbourhood $\mathcal{V}_c$ of s_0 (depending on c), such that

$$|\phi(s)| \le c\,|g(s)|\,, \qquad s \in \mathcal{V}_c, \quad s \ne s_0.$$

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 [**87**, 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 $\phi(s) = \mathcal{O}(s)$ and $\phi(s) = \Omega(s)$. One then says that ϕ is of order exactly g.

For instance, one has as $n \to +\infty$ in $\mathbb{Z}_{>0}$:

$$\sin n = o(\log n); \quad \log n = O(\sqrt{n}); \quad \log n = o(\sqrt{n});$$
$$\binom{n}{2} = \Omega(n\sqrt{n}); \quad \pi n + \sqrt{n} = \Theta(n).$$

As $x \to 1$ in $\mathbb{R}_{\leq 1}$, 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., [**130**, Ch. 4] for a smooth introduction close to the needs of analytic combinatorics and de Bruijn's classic [**35**] for a beautiful presentation.). We shall retain here the fact that Taylor expansions imply asymptotic expansions; for instance, the convergent expansions for $|u| < 1$,

$$\log(1+u) = \sum_{k=1}^{\infty} \frac{(-1)^k}{k} u^k, \quad \exp(u) = \sum_{k\geq 0} \frac{1}{k!} u^k, \quad (1-u)^\alpha = \sum_{k\geq 0} \binom{k+\alpha-1}{k} u^k,$$

imply (as $u \to 0$)

$$\log(1+u) = u + \mathcal{O}(u^2), \quad \exp(u) = 1+u+\frac{u^2}{2}+\mathcal{O}(u^3), \quad (1-u)^{1/2} = 1-\frac{u}{2}+\mathcal{O}(u^2),$$

and, in turn, (as $n \to +\infty$)

$$\log\left(1+\frac{1}{n}\right) = \frac{1}{n} + \mathcal{O}\left(\frac{1}{n^2}\right), \quad \left(1-\frac{1}{\log n}\right)^{1/2} = 1 - \frac{1}{2\log n} + o\left(\frac{1}{\log n}\right).$$

Two important special expansions are Stirling's formula for factorials and the harmonic number approximation,

$$\text{(3)} \qquad \begin{array}{rcll} n! & = & n^n e^{-n}\sqrt{2\pi n}\,(1+\epsilon_n), & 0 < \epsilon_n < \frac{1}{12n} \\ \mathrm{H}_n & = & \log n + \gamma + \dfrac{1}{2n} - \dfrac{1}{12n^2} + \eta_n & \eta_n = \mathcal{O}\left(n^{-4}\right), \quad \gamma \doteq 0.57721, \end{array}$$

that are best established as consequences of the Euler–Maclaurin summation formula [**35, 130**].

$\triangleright$ **1.** *Simplification rules for the asymptotic calculus.* Some of them are

$$\begin{array}{lcll} \mathcal{O}(\lambda f) & \longrightarrow & \mathcal{O}(f) & (\lambda \neq 0) \\ \mathcal{O}(f) \pm \mathcal{O}(g) & \longrightarrow & \mathcal{O}(|f|+|g|) & \\ & \longrightarrow & \mathcal{O}(f) & \text{if } g = \mathcal{O}(f) \\ \mathcal{O}(f\cdot g) & \longrightarrow & O(f)O(g). & \end{array}$$

Similar rules apply for $o(\cdot)$. $\triangleleft$

$\triangleright$ **2.** *Harmonics of harmonics.* The harmonic numbers are readily extended to non-integral index by

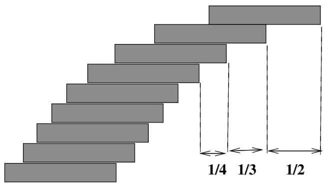
$$\mathrm{H}_x := \sum_{k=1}^{\infty} \left(\frac{1}{k} - \frac{1}{k+x}\right).$$

For instance, $\mathrm{H}_{1/2} = 2 - 2\log 2$. This extension is related to the Gamma function [**150**], 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

$$\mathrm{H}_{\mathrm{H}_n} = \log\log n + \gamma + \frac{\gamma+\frac{1}{2}}{\log n} + \mathcal{O}\left(\frac{1}{\log^2 n}\right).$$

What is the shape of an asymptotic expansion of $\mathrm{H}_{\mathrm{H}_{\mathrm{H}_n}}$? $\triangleleft$

▷ **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! ◁

▷ **4.** *High precision fraud.* Why is it that, to forty decimal places, one finds

$$\begin{aligned} 4\sum_{k=1}^{500{,}000}\frac{(-1)^{k-1}}{2k-1} &\doteq 3.14159\underline{0}6535897932\underline{40}4626433832\underline{6}9502884197 \\ \pi &\doteq 3.1415926535897932384626433832795028841 97, \end{aligned}$$

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 to be found in works by Jon and Peter Borwein [**21, 22**]. ◁

3. Cycle construction. The unlabelled cycle construction is introduced in Chapter 1 and is classically obtained within the framework of Pólya theory [**28, 113, 115**]. The derivation given here is based on an elementary use of symbolic methods tha follows [**58**]. 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 $\mathcal{S} = \mathfrak{S}_{\geq 1}\{\mathcal{A}\}$. 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 $\alpha\beta\alpha\beta$ 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. Moebius inversion then gives

$$PS(z,u) = \sum_{k\geq 1}\mu(k)S(z^k,u^k) = \sum_{k\geq 1}\mu(k)\frac{u^kA(z^k)}{1-u^kA(z^k)}.$$

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,

$$PC(z,u) = \sum_{k\geq 1}\frac{\mu(k)}{k}\log\frac{1}{1-u^kA(z^k)}.$$

Finally, cycles can be composed from arbitrary repetitions of primitive cycles, which yields

$$C(z,u) = \sum_{k\geq 1} PC(z^k, u^k).$$

The arithmetical identity $\sum_{d \mid k} \mu(d)/d = \varphi(k)/k$ gives eventually

$$C(z,u) = \sum_{k\geq 1} \frac{\varphi(k)}{k} \log \frac{1}{1-u^k A(z^k)}, \tag{4}$$

as was to be proved.

Formula (4) is exactly the one that appears in the translation of the cycle construction in the unlabelled case (Theorem III.1). Upon setting $u = 1$, it gives the univariate version (Theorem I.1).

▷ **5.** *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 [**36**]. (The latter fact corresponds to the property that any word can be written as a decreasing product of Lyndon words; it serves to construct bases of free Lie algebras [**98**, Ch. 5].) ◁

▷ **6.** *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

$$PW_n^{(m)} = \sum_{d \mid n} \mu(d) m^{m/d}.$$

For $m = 2$, the sequence starts as $2, 2, 6, 12, 30, 54, 126, 240, 504, 990$ (*EIS* **A027375**). ◁

4. Formal power series. Formal power series extend the usual operations on polynomials to infinite series of the form

$$f = \sum_{n\geq 0} f_n z^n, \tag{5}$$

where z is a formal indeterminate. The notation $f(z)$ is also employed. Let $\mathbb{K}$ be a field of coefficients (usually $\mathbb{Q}, \mathbb{R}, \mathbb{C}$); 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 (5) and endowed with the operations of sum and product,

$$\begin{aligned} \left(\sum_n f_n z^n\right) + \left(\sum_n g_n z^n\right) &:= \sum_n (f_n + g_n)\, z^n \\ \left(\sum_n f_n z^n\right) \times \left(\sum_n g_n z^n\right) &:= \sum_n \left(\sum_{k=0}^{n} f_k g_{n-k}\right) z^n. \end{aligned}$$

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 $\mathrm{val}(f)$. (One sets $\mathrm{val}(0) = +\infty$.) Given two power series f and g, their distance $d(f,g)$ is then defined as $2^{-\mathrm{val}(f-g)}$. With this distance (in fact an ultrametric distance), the space of all formal power series is 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 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 $(1-f)^{-1}$, 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).

$\triangleright$ **7.** *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. 57). 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 of an integral. $\triangleleft$

It can be proved that the usual functional properties of analysis extend to formal power series provided they make sense formally.

5. 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.1. *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)*

$$y(z) = \sum_{n=1}^{\infty} y_n z^n, \qquad \textit{where} \quad y_n = \frac{1}{n}\,[u^{n-1}]\,(\phi(u))^n\,. \tag{6}$$

Furthermore, one has for $k > 0$ (Bürmann form)

$$y(z)^k = \sum_{n=1}^{\infty} y_n^{(k)} z^n, \qquad \textit{where} \quad y_n = \frac{k}{n}\,[u^{n-k}]\,(\phi(u))^n\,. \tag{7}$$

By linearity, a form equivalent to Burmann's (7), with H an arbitrary function, is

$$[z^n]H(y(z)) = [u^{n-1}]\,(H'(u)\phi(u)^n)\,.$$

PROOF. The method of indeterminates coefficients provides a system of polynomial equations for $\{y_n\}$ that is seen to have a unique (polynomial) solution:

$$y_1 = \phi_0, \quad y_2 = \phi_0\phi_1, \quad y_3 = \phi_0\phi_1^2 + \phi_0\phi_2, \ \ldots .$$

Since y_n depends only on the coefficients of $\phi(u)$ till order n, one may assume without loss of generality that ϕ is a polynomial. Then, by general properties of analytic functions, $y(z)$ is analytic at 0 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:

$$\begin{aligned} ny_n &= \frac{1}{2i\pi}\int_{0+} y'(z)\,\frac{dz}{z^n} && \text{(Direct coefficient formula for } y'(z)\text{)} \\ &= \frac{1}{2i\pi}\int_{0+} \frac{dy}{(y/\phi(y))^n} && \text{(Change of variable } z \mapsto y\text{)} \\ &= [y^{n-1}]\,\phi(y)^n && \text{(Reverse coefficient formula for } \phi(y)^n\text{).} \end{aligned} \tag{8}$$

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 entirely similar. □

There exist instructive (but longer) combinatorial proofs based on what is known as the "cyclic lemma" or "conjugacy principle" [**119**] for Łukasiewicz words. (See also Ex. 36 in Chapter I.) Another classical proof due to Henrici relies on properties of iteration matrices [**28**, p. 144-153]; see also Comtet's book for related formulations [**28**].

Lagrange inversion serves most notably to develop explicit formulæ for simple families of trees (either labelled or not), random mappings, and more generally for problems involving coefficients of powers of some fixed function.

▷ **8.** *Lagrange–BÚrmann 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 $\log(y(z)/z)$. ◁

▷ **9.** *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 derive the *Abel's identity*

$$(\alpha+\beta)(n+\alpha+\beta)^{n-1} = \alpha\beta\sum_{k=1}^{n-1}\binom{n}{k}(k+\alpha)^{k-1}(n-k+\beta)^{n-k-1}.$$

◁

6. Regular languages. Two notions of *regularity* for languages are described in the text (Section I. 4): R-regularity, which means definability by regular specifications, and A-regularity, which corresponds to acceptability by a deterministic finite automaton. We indicate briefly here the reasons why the two notions are equivalent. The arguments are minor adaptations of well known facts in the theory of formal languages, and we refer the reader to one of the many good books on the subject for details.

A-regularity implies S-regularity. This construction is due to Kleene [**81**] whose interest had its origin in the formal expressive power of nerve nets. 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

$$\mathcal{L}_{i,j}^{(r)} = \mathcal{L}_{i,j}^{(r-1)} + \mathcal{L}_{i,r}^{(r-1)} \mathfrak{S}\{\mathcal{L}_{r,r}^{(r-1)}\} \mathcal{L}_{r,j}^{(r-1)},$$

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.

S-regularity implies A-regularity. An object described by a regular specification $\mathfrak{r}$ can be viewed as a word decorated with separators that indicate the way it should be parsed. For instance, an element of $\mathfrak{S}\{a + aa\}$ may be viewed as the word

$$\langle a \mid aa \mid aa \mid a \mid aa \rangle,$$

over the enriched alphabet $\mathcal{A} \cup \{`|', `\langle', `\rangle'\}$. The extended representations are then recognizable by automata 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

$$\begin{array}{ccl}
\to\bullet\boxed{\mathfrak{r}+\mathfrak{s}}\bullet\to & = & \begin{array}{l} \nearrow\to\bullet\boxed{\mathfrak{r}}\bullet\to \\ \searrow\to\bullet\boxed{\mathfrak{s}}\bullet\to \end{array} \\
\to\bullet\boxed{\mathfrak{r}\times\mathfrak{s}}\bullet\to & = & \to\bullet\boxed{\mathfrak{r}}\bullet\to\to\bullet\boxed{\mathfrak{s}}\bullet\to \\
\to\bullet\boxed{\mathfrak{S}\{\mathfrak{r}\}}\bullet\to & = & \overline{\downarrow\to\bullet\boxed{\mathfrak{r}}\bullet\to\uparrow}
\end{array}$$

The classical theory of formal languages defines the family of *regular languages* as the smallest family containing the finite languages that is closed under set-theoretic union ($\cup$), catenation product ($\cdot$), and the Kleene star operation $\mathcal{L}^\star = \{\epsilon\} \cup \cup(\mathcal{L} \cdot \mathcal{L}) \cup \cdots$. Any regular language is then denoted by a regular expression. The operations are taken in the set-theoretic sense, so that for instance one has the identity[1] $(a \cup aa)^\star = (a)^\star$. It is a standard result of the theory that any regular language is recognizable by a deterministic finite automaton, so that this notion of regularity is indeed equivalent to the two combinatorial notions of A-regularity and S-regularity. (Note: the reduction of regular languages to deterministic automata goes via the construction of nondeterministic automata followed by a reduction, the Rabin-Scott theorem, that usually involves an exponential blow-up in the number of states.)

---

[1]Union, catenation, and Kleene star resemble sum, cartesian product product, and sequence constructions, respectively. However, there is no systematic correspondence since the set-theoretic operations may be applied ambiguously, in contrast to combinatorial constructions that preserve structure. For instance, in the combinatorial world, one has $\mathfrak{S}\{a\} \neq \mathfrak{S}\{a + aa\}$ and the expression $\mathfrak{S}\{a + aa\}$ denotes structures richer than just words over the letter a (see *coverings* on p. 9).

7. 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 [**71**].

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 the 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 [**28**, p. 216].)

A more important relation is that of the generating polynomials of the $\left[{n \atop r}\right]$ for fixed n,

$$P_n(u) \equiv \sum_{r=1}^{n} P_n^{(r)} u^r = u\cdot(u+1)\cdot(u+2)\cdots(u+n-1).$$

This nicely parallels the OGF for the $\left\{{n \atop r}\right\}$ for fixed r

$$\sum_{n=0}^{\infty}\left\{{n \atop r}\right\} z^n = \frac{z^r}{(1-z)(1-2z)\cdots(1-kz)}.$$

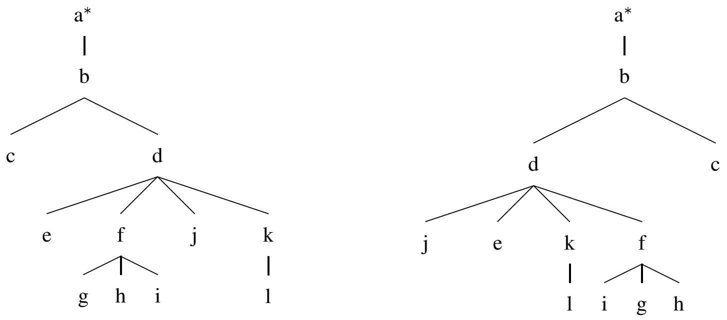
▷ **10.** Schlömilch's formula is established starting from

$$\frac{k!}{n!}\begin{bmatrix} n \\ k \end{bmatrix} = \frac{1}{2i\pi}\oint \log^k \frac{1}{1-z}\,\frac{dz}{z^{n+1}},$$

and performing the change of variable *a la* Lagrange: $z = 1 - e^{-t}$. [**28**, p.216]. ◁

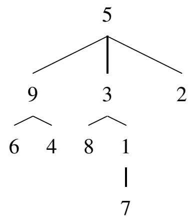
8. Tree concepts. In the abstract graph-theoretic sense, a *forest* is an acyclic (undirected) graph and a *tree* is a forest that consists of just one connected component. A *rooted tree* is a tree in which a specific node is distinguished, the *root*. Rooted trees are drawn with the root either below (the mathematician's and genealogist's convention) or on top (the computer scientist's convention), and in this book, we employ both conventions indifferently. Here are then two planar representations of the same rooted tree

(9)



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:

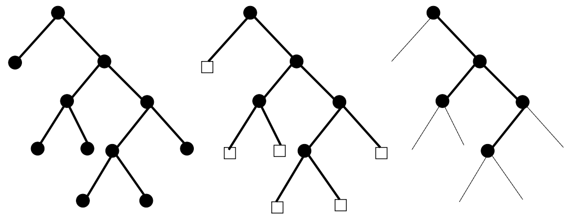
(10)



In a rooted tree, the *outdegree* of a node is the number of its descendants; 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 (10) 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* or *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 (9) 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)	$\mathcal{G} = \mathcal{Z} \times \mathfrak{S}\{\mathcal{G}\}$	(unlabelled)
Binary trees	$\mathcal{A} = \mathcal{Z} + (\mathcal{Z} \times \mathcal{A} \times \mathcal{A})$	(unlabelled)
Pruned binary trees	$\mathcal{B} = \mathbf{1} + (\mathcal{Z} \times \mathcal{B} \times \mathcal{B})$	(unlabelled)
General nonplane trees (Cayley trees)	$\mathcal{T} = \mathcal{Z} \times \mathfrak{P}\{\mathcal{T}\}$	(labelled)

The corresponding GFs are respectively

$$G(z) = \frac{1-\sqrt{1-4z}}{2}, \quad B(z) = \frac{1-\sqrt{1-4z^2}}{2z}, \quad C(z) = \frac{1-\sqrt{1-4z}}{2z}, \quad T(z) = ze^{T(z)},$$

being respectively of type OGF for the first three and EGF for the last one. The corresponding counts are

$$G_n = \frac{1}{n}\binom{2n-2}{n-1}, \quad A_{2\nu+1} = \frac{1}{\nu+1}\binom{2\nu}{\nu}, \quad B_n = \frac{1}{n+1}\binom{2n}{n}, \quad T_n = n^{n-1}.$$

The common occurrence of the Catalan numbers, C_n ($A_{2\nu+1} = B_\nu = G_{\nu+1} = \mathrm{C}_\nu$)is explained by pruning and by the rotation correspondence described on p. 48.

[2] The term "general" refers to the fact that no degree constraints are imposed.

Bibliography

1. Alfred V. Aho and Margaret J. Corasick, *Efficient string matching: an aid to bibliographic search*, Communications of the ACM **18** (1975), 333–340.
2. David Aldous and Persi Diaconis, *Longest increasing subsequences: from patience sorting to the Baik-Deift-Johansson theorem*, Bull. Amer. Math. Soc. (N.S.) **36** (1999), no. 4, 413–432.
3. Noga Alon and Joel H. Spencer, *The probabilistic method*, John Wiley & Sons Inc., New York, 1992.
4. George E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and its Applications, vol. 2, Addison–Wesley, 1976.
5. Tom M. Apostol, *Introduction to analytic number theory*, Springer-Verlag, 1976.
6. J. Arney and E. D. Bender, *Random mappings with constraints on coalescence and number of origins*, Pacific Journal of Mathematics **103** (1982), 269–294.
7. Krishna B. Athreya and Peter E. Ney, *Branching processes*, Springer-Verlag, New York, 1972, Die Grundlehren der mathematischen Wissenschaften, Band 196.
8. 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.
9. 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.
10. 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.
11. Edward A. Bender and Jay R. Goldman, *Enumerative uses of generating functions*, Indiana University Mathematical Journal (1971), 753–765.
12. Jon Bentley and Robert Sedgewick, *Fast algorithms for sorting and searching strings*, Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SIAM Press, 1997.
13. F. Bergeron, G. Labelle, and P. Leroux, *Combinatorial species and tree-like structures*, Cambridge University Press, Cambridge, 1998, Translated from the 1994 French original by Margaret Readdy, With a foreword by Gian-Carlo Rota.
14. Elwyn R. Berlekamp, *Algebraic coding theory*, Mc Graw-Hill, 1968, Revised edition, 1984.
15. J. Berstel and C. Reutenauer, *Recognizable formal power series on trees*, Theoretical Computer Science **18** (1982), 115–148.
16. Jean Berstel and Dominique Perrin, *Theory of codes*, Academic Press Inc., Orlando, Fla., 1985.
17. Norman Biggs, E. Keith Lloyd, and Robin Wilson, *Graph theory, 1736–1936*, Oxford University Press, 1974.
18. Patrick Billingsley, *Probability and measure*, 2nd ed., John Wiley & Sons, 1986.
19. Béla Bollobás, *Random graphs*, Academic Press, 1985.
20. D. Borwein, S. Rankin, and L. Renner, *Enumeration of injective partial transformations*, Discrete Mathematics **73** (1989), 291–296.
21. Jonathan M. Borwein and Peter B. Borwein, *Strange series and high precision fraud*, American Mathematical Monthly **99** (1992), no. 7, 622–640.
22. Jonathan M. Borwein, Peter B. Borwein, and Karl Dilcher, *Pi, Euler numbers and asymptotic expansions*, American Mathematical Monthly **96** (1989), no. 8, 681–687.
23. Mireille Bousquet-Mélou, *A method for the enumeration of various classes of column-convex polygons*, Discrete Math. **154** (1996), no. 1-3, 1–25.
24. Mireille Bousquet-Mélou and Anthony J. Guttmann, *Enumeration of three-dimensional convex polygons*, Annals of Combinatorics **1** (1997), 27–53.
25. W. H. Burge, *An analysis of binary search trees formed from sequences of nondistinct keys*, JACM **23** (1976), no. 3, 451–454.
26. 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.
27. 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.
28. Louis Comtet, *Advanced combinatorics*, Reidel, Dordrecht, 1974.

29. 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.
30. T. H. Cormen, C. E. Leiserson, and R. L. Rivest, *Introduction to Algorithms*, MIT Press, New York, 1990.
31. 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.
32. H. Davenport, *Multiplicative Number Theory*, revised by H. L. Montgomery, second ed., Springer-Verlag, New York, 1980.
33. F. N. David and D. E. Barton, *Combinatorial chance*, Charles Griffin, London, 1962.
34. N. G. De Bruijn, *On Mahler's partition problem*, Indagationes Math. **10** (1948), 210–220, Reprinted from Koninkl. Nederl. Akademie Wetenschappen, Ser. A.
35. N. G. de Bruijn, *Asymptotic methods in analysis*, Dover, 1981, A reprint of the third North Holland edition, 1970 (first edition, 1958).
36. N. G. De Bruijn and D. A. Klarner, *Multisets of aperiodic cycles*, SIAM Journal on Algebraic and Discrete Methods **3** (1982), 359–368.
37. 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.
38. A. Dembo, A. Vershik, and O. Zeitouni, *Large deviations for integer partitions*, Markov Processes and Related Fields **6** (2000), no. 2, 147–179.
39. Luc Devroye, *Limit laws for local counters in random binary search trees*, Random Structures & Algorithms **2** (1991), no. 3, 302–315.
40. A. Dvoretzky and Th. Motzkin, *A problem of arrangements*, Duke Mathematical Journal **14** (1947), 305–313.
41. Samuel Eilenberg, *Automata, languages, and machines*, vol. A, Academic Press, 1974.
42. Paul Erdös and Joseph Lehner, *The distribution of the number of summands in the partitions of a positive integer*, Duke Mathematical Journal **8** (1941), 335–345.
43. W. Feller, *An introduction to probability theory and its applications*, vol. 2, John Wiley, 1971.
44. Philippe Flajolet, *Combinatorial aspects of continued fractions*, Discrete Mathematics **32** (1980), 125–161.
45. ———, *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.
46. ———, *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.
47. Philippe Flajolet, Danièle Gardy, and Loÿs Thimonier, *Birthday paradox, coupon collectors, caching algorithms, and self–organizing search*, Discrete Applied Mathematics **39** (1992), 207–229.
48. Philippe Flajolet, Xavier Gourdon, and Conrado Martínez, *Patterns in random binary search trees*, Random Structures & Algorithms **11** (1997), no. 3, 223–244.
49. 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.
50. 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.
51. Philippe Flajolet, Donald E. Knuth, and Boris Pittel, *The first cycles in an evolving graph*, Discrete Mathematics **75** (1989), 167–215.
52. 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.
53. ———, *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.
54. Philippe Flajolet and Helmut Prodinger, *Level number sequences for trees*, Discrete Mathematics **65** (1987), 149–156.
55. Philippe Flajolet, Bruno Salvy, and Gilles Schaeffer, *Airy phenomena and analytic combinatorics of connected graphs*, Preprint, 2002.
56. Philippe Flajolet, Bruno Salvy, and Paul Zimmermann, *Automatic average–case analysis of algorithms*, Theoretical Computer Science **79** (1991), no. 1, 37–109.
57. 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.
58. Philippe Flajolet and Michèle Soria, *The cycle construction*, SIAM Journal on Discrete Mathematics **4** (1991), no. 1, 58–60.
59. 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.

60. ———, *A complexity calculus for recursive tree algorithms*, Mathematical Systems Theory **19** (1987), 301–331.
61. 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.
62. Dominique Foata, *La série génératrice exponentielle dans les problèmes d'énumération*, S.M.S, Montreal University Press, 1974.
63. 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.
64. Dominique Foata and Marcel-P. Schützenberger, *Théorie géométrique des polynômes Euleriens*, Lecture Notes in Mathematics, vol. 138, Springer Verlag, 1970.
65. Ira M. Gessel, *Symmetric functions and P–recursiveness*, Journal of Combinatorial Theory, Series A **53** (1990), 257–285.
66. V. Goncharov, *On the field of combinatory analysis*, Soviet Math. Izv., Ser. Math. **8**, 3–48, In Russian.
67. Gaston H. Gonnet, *Expected length of the longest probe sequence in hash code searching*, Journal of the ACM **28** (1981), no. 2, 289–304.
68. Ian P. Goulden and David M. Jackson, *Combinatorial enumeration*, John Wiley, New York, 1983.
69. ———, *Distributions, continued fractions, and the Ehrenfest urn model*, Journal of Combinatorial Theory. Series A **41** (1986), no. 1, 21–31.
70. Xavier Gourdon, *Largest component in random combinatorial structures*, Discrete Mathematics **180** (1998), no. 1-3, 185–209.
71. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, Addison Wesley, 1989.
72. D. H. Greene and D. E. Knuth, *Mathematics for the analysis of algorithms*, Birkhäuser, Boston, 1981.
73. Daniel Hill Greene, *Labelled formal languages and their uses*, Ph.D. thesis, Stanford University, June 1983, Available as Report STAN-CS-83-982.
74. L. J. Guibas and A. M. Odlyzko, *String overlaps, pattern matching, and nontransitive games*, Journal of Combinatorial Theory. Series A **30** (1981), no. 2, 183–208.
75. Laurent Habsieger, Maxime Kazarian, and Sergei Lando, *On the second number of Plutarch*, American Mathematical Monthly **105** (1998), 446–447.
76. Frank Harary and Edgar M. Palmer, *Graphical enumeration*, Academic Press, 1973.
77. 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.
78. Svante Janson, Tomasz Łuczak, and Andrzej Rucinski, *Random graphs*, Wiley-Interscience, New York, 2000.
79. André Joyal, *Une théorie combinatoire des séries formelles*, Advances in Mathematics **42** (1981), no. 1, 1–82.
80. M. S. Klamkin and D. J. Newman, *Extensions of the birthday surprise*, Journal of Combinatorial Theory **3** (1967), 279–282.
81. S. C. Kleene, *Representation of events in nerve nets and finite automata*, Automata studies, Princeton University Press, Princeton, N. J., 1956, pp. 3–41.
82. Arnold Knopfmacher and Helmut Prodinger, *On Carlitz compositions*, European Journal of Combinatorics **19** (1998), no. 5, 579–589.
83. John Knopfmacher, *Abstract analytic number theory*, Dover, 1990.
84. Donald E. Knuth, *Mathematical analysis of algorithms*, Information Processing 71, North Holland Publishing Company, 1972, Proceedings of IFIP Congress, Ljubljana, 1971, pp. 19–27.
85. ———, *The art of computer programming*, 3rd ed., vol. 1: Fundamental Algorithms, Addison-Wesley, 1997.
86. ———, *The art of computer programming*, 2nd ed., vol. 3: Sorting and Searching, Addison-Wesley, 1998.
87. ———, *Selected papers on analysis of algorithms*, CSLI Publications, Stanford, CA, 2000.
88. 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.
89. Donald E. Knuth and Ilan Vardi, *Problem 6581 (the asymptotic expansion of $2n$ choose n)*, American Mathematical Monthly **95** (1988), 774.
90. Valentin F. Kolchin, *Random mappings*, Optimization Software Inc., New York, 1986, Translated from *Slučajnye Otobraženija*, Nauka, Moscow, 1984.
91. ———, *Random graphs*, Encyclopedia of Mathematics and its Applications, vol. 53, Cambridge University Press, Cambridge, U.K., 1999.
92. 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*.
93. J. C. Lagarias and A. M. Odlyzko, *Solving low-density subset sum problems*, JACM **32** (1985), no. 1, 229–246.
94. Serge Lang, *Algebra*, Addison-Wesley, Reading, Mass., 1965.
95. ———, *Linear algebra*, Addison-Wesley, Reading, Mass., 1966.
96. V. Lifschitz and B. Pittel, *The number of increasing subsequences of the random permutation*, Journal of Combinatorial Theory, Series A **31** (1981), 1–20.

97. B. F. Logan and L. A. Shepp, *A variational problem for random Young tableaux*, Advances in Mathematics **26** (1977), 206–222.
98. M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 17, Addison–Wesley, 1983.
99. E. Lucas, *Théorie des Nombres*, Gauthier-Villard, Paris, 1891, Reprinted by A. Blanchard, Paris 1961.
100. 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.
101. P. A. MacMahon, *Introduction to combinatory analysis*, Chelsea Publishing Co., New York, 1955, A reprint of the first edition, Cambridge, 1920.
102. Hosam M. Mahmoud, *Evolution of random search trees*, John Wiley, New York, 1992.
103. 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).
104. A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canadian Journal of Mathematics **30** (1978), 997–1015.
105. J. W. Moon, *Counting labelled trees*, Canadian Mathematical Monographs N.1, William Clowes and Sons, 1970.
106. 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.
107. Rajeev Motwani and Prabhakar Raghavan, *Randomized algorithms*, Cambridge University Press, 1995.
108. Donald J. Newman and Lawrence Shepp, *The double dixie cup problem*, American Mathematical Monthly **67** (1960), 58–61.
109. Albert Nijenhuis and Herbert S. Wilf, *Combinatorial algorithms*, second ed., Academic Press, 1978.
110. A. M. Odlyzko, *Periodic oscillations of coefficients of power series that satisfy functional equations*, Advances in Mathematics **44** (1982), 180–205.
111. Richard Otter, *The number of trees*, Annals of Mathematics **49** (1948), no. 3, 583–599.
112. D. Panario and B. Richmond, *Exact largest and smallest size of components*, Algorithmica **31** (2001), no. 3, 413–432.
113. G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Mathematica **68** (1937), 145–254.
114. ———, *On the number of certain lattice polygons*, Journal of Combinatorial Theory, Series A **6** (1969), 102–105.
115. G. Pólya and R. C. Read, *Combinatorial enumeration of groups, graphs and chemical compounds*, Springer Verlag, New York, 1987.
116. George Pólya, Robert E. Tarjan, and Donald R. Woods, *Notes on introductory combinatorics*, Progress in Computer Science, Birkhäuser, 1983.
117. Helmut Prodinger, *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.
118. 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).
119. G. N. Raney, *Functional composition patterns and power series reversion*, Transactions of the American Mathematical Society **94** (1960), 441–451.
120. A. Rényi and G. Szekeres, *On the height of trees*, Australian Journal of Mathematics **7** (1967), 497–507.
121. P. Révész, *Strong theorems on coin tossing*, Proceedings of the International Congress of Mathematicians (Helsinki, 1978) (Helsinki), Acad. Sci. Fennica, 1980, pp. 749–754.
122. Gian-Carlo Rota, *Finite operator calculus*, Academic Press, 1975.
123. 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.
124. Vladimir N. Sachkov, *Combinatorial methods in discrete mathematics*, Encyclopedia of Mathematics and its Applications, vol. 55, Cambridge University Press, 1996.
125. ———, *Probabilistic methods in combinatorial analysis*, Cambridge University Press, Cambridge, 1997, Translated and adapted from the Russian original edition, Nauka, Moscow, 1978.
126. Arto Salomaa and Matti Soittola, *Automata-theoretic aspects of formal power series*, Springer, Berlin, 1978.
127. Bruno Salvy and John Shackell, *Symbolic asymptotics: multiseries of inverse functions*, Journal of Symbolic Computation **27** (1999), no. 6, 543–563.
128. Robert Sedgewick, *Quicksort with equal keys*, SIAM Journal on Computing **6** (1977), no. 2, 240–267.
129. ———, *Algorithms*, second ed., Addison–Wesley, Reading, Mass., 1988.
130. Robert Sedgewick and Philippe Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley Publishing Company, 1996.
131. L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Transactions of the American Mathematical Society **121** (1966), 340–357.
132. N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, 2000, Published electronically at `http://www.research.att.com/~njas/sequences/`.

133. N. J. A. Sloane and Simon Plouffe, *The encyclopedia of integer sequences*, Academic Press, 1995.
134. 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.
135. ______, *Enumerative combinatorics*, vol. I, Wadsworth & Brooks/Cole, 1986.
136. ______, *Hipparchus, Plutarch, Schröder and Hough*, American Mathematical Monthly **104** (1997), 344–350.
137. ______, *Enumerative combinatorics*, vol. II, Cambridge University Press, 1998.
138. Jean-Marc Steyaert, *Structure et complexité des algorithmes*, Doctorat d'état, Université Paris VII, April 1984.
139. Wojciech Szpankowski, *Average-case analysis on algorithms on sequences*, John Wiley, New York, 2001.
140. H. N. V. Temperley, *On the enumeration of the Mayer cluster integrals*, Proc. Phys. Soc. Sect. B. **72** (1959), 1141–1144.
141. ______, *Graph theory and applications*, Ellis Horwood Ltd., Chichester, 1981.
142. Bernard Van Cutsem, *Combinatorial structures and structures for classification*, Comput. Statist. Data Anal. **23** (1996), no. 1, 169–188.
143. J. van Leeuwen (ed.), *Handbook of theoretical computer science*, vol. A: Algorithms and Complexity, North Holland, 1990.
144. E. J. Janse van Rensburg, *The statistical mechanics of interacting walks, polygons, animals and vesicles*, Oxford University Press, Oxford, 2000.
145. A. M. Vershik, *Statistical mechanics of combinatorial partitions, and their limit configurations*, Funktsional'nyĭ Analiz i ego Prilozheniya **30** (1996), no. 2, 19–39.
146. 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.
147. 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.
148. J. Vuillemin, *A unifying look at data structures*, Communications of the ACM **23** (1980), no. 4, 229–239.
149. Michael S. Waterman, *Introduction to computational biology*, Chapman & Hall, 1995.
150. E. T. Whittaker and G. N. Watson, *A course of modern analysis*, fourth ed., Cambridge University Press, 1927, Reprinted 1973.
151. Herbert S. Wilf, *Some examples of combinatorial averaging*, American Mathematical Monthly **92** (1985), 250–261.
152. ______, *Combinatorial algorithms: An update*, CBMS–NSF Regional Conference Series, no. 55, Society for Industrial and Applied Mathematics, Philadelphia, 1989.
153. ______, *Generatingfunctionology*, Academic Press, 1990.
154. E. Maitland Wright, *The number of connected sparsely edged graphs*, Journal of Graph Theory **1** (1977), 317–330.
155. ______, *The number of connected sparsely edged graphs. II. Smooth graphs*, Journal of Graph Theory **2** (1978), 299–305.
156. ______, *The number of connected sparsely edged graphs. III. Asymptotic results*, Journal of Graph Theory **4** (1980), 393–407.
157. 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.
158. Paul Zimmermann, *Séries génératrices et analyse automatique d'algorithmes*, Ph. d. thesis, École Polytechnique, 1991.

Index