RANDOM SYSTEMS IN COMBINATORICS

Philippe Flajolet
INRIA Rocquencourt, F-78150 Le Chesnay France.
E-mail: `Philippe.Flajolet@inria.fr`
Web: `http://algo.inria.fr/flajolet`

Summer School on
Statistical Physics and Probabilistic Methods in Computer Science
ICTP, Trieste, August 23 to September 3, 1999

These notes are an informal presentation of Analytic Combinatorics. *They should serve as an introduction to chapters of a forthcoming book that are currently available from the web. See especially:*

I. *Counting and Generating Functions (Chapters 1, 2, 3. 1993, 116p.) INRIA Tech. Rep. #1888.*

II. *Complex Asymptotics and Generating Functions (Chapters 4, 5. 1993, 100p.) INRIA Tech. Rep. #2026.*

III. *Multivariate Asymptotics and Limit Distributions (Chapter 9, May 1997, 123p.) INRIA Tech. Rep. #3162.*

Three basic principles are presented in order to analyse randomness in combinatorics.

Principle I. Enumeration of combinatorial objects that are decomposable is best approached by means of generating functions. Functional equations of GF's then directly reflect the decomposable structure of the objects to be enumerated.

Principle II. The singularities of a function determine precisely the asymptotic behaviour of its Taylor coefficients. In particular, asymptotic counting results derive from singularities of generating functions.

Principle III. For decomposable objects and parameters, algebraic and analytic relations on counting generating functions usually lift to relations between bivariate generating functions. Then *singularity perturbation analysis* shows parameters to admit limit laws, often of the Gaussian type.

RANDOM SYSTEMS IN COMBINATORICS

Philippe Flajolet

Summer School on
Statistical Physics and Probabilistic Methods in Computer Science
ICTP, Trieste, August 23 to September 3, 1999

I. Exact Counting

Combinatorial enumeration. *Combinatorics* is the study of finite arrangements or combinations obeying a finite collection of rules. The domain started largely with concrete problems relative to games of chance, a typical instance being:

> Problem 1. Four players sit at a table and a deck of 52 ($= 13 \times 4$) cards is dealt from. A player wins if she has at least six cards of one suit and she is the only one with this property. What are the chances that there is a winner?

Problems of this kind were studied by Blaise Pascal (1623-1662); they reduce to *enumerative* questions since the probability of a winner is the quotient of the number of favourable cases and of the total number of games (here 52! permutations of 52 cards). A large number of examples often treated by ad hoc methods were then to be studied in the 17th, 18th, and 19th century. Abraham de Moivre (1667-1754) can be considered as a precursor of what may be termed *symbolic methods*.

> Problem 2. Assume we cast a die (with 6 faces) four times. What are the chances of obtaining a sum of scores that equals 17? (Clearly, the number of possible outcomes is $6^4 = 1296$.)

De Moivre's line of reasoning goes as follows. Denote by $d_1, d_2, d_3, d_4, d_5, d_6$ the possible outcomes of casting a die once. Then, all the possible outcomes of 4 casts are described by expanding through distributivity

$$\Omega = (d_1 + d_2 + d_3 + d_4 + d_5 + d_6)^4 = d_1 d_1 d_1 d_1 + \cdots + d_6 d_6 d_6 d_6.$$

In effect, each monomial like $m = d_2 d_5 d_3 d_4$ represents a sequence of draws, for instance $2, 5, 3, 4$ in that order. If we substitute $d_1, \ldots, d_6 \mapsto 1$, we get the number 6^4 as we should. Assume that we now perform the substitution

$$d_1 \mapsto z^1, \ldots, d_6 \mapsto z^6.$$

Then each monomial in Ω with index sum ("score") equal to S transforms into z^S. The number of times the score 17 is attained must then be equal to

$$[\text{coefficient of } z^{17}] \text{ in } \left(z + z^2 + z^3 + z^4 + z^5 + z^6\right)^4.$$

The polynomial in this equation is called nowadays a *generating function* (GF). The term is well–chosen as it is a formal sum

$$f(z) = \sum_n f_n z^n$$

that has the property that the coefficient[1] f_n of z^n in $f(z)$ is precisely the number of configurations with score n that are to be enumerated. In other words, the generating function "enumerates" (i.e., counts) all possibilities.

We can then regard the problem as solved and simple algebra should give us in a matter of minutes the probability of a score equal to 17 as being

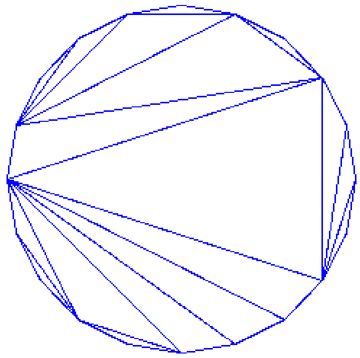
$$\frac{104}{1296} = \frac{13}{162} \doteq 0.08024 \approx 8\%.$$

This is to be contrasted with the couple of hours that it might take to list exhaustively all the possibilities and select the right ones. (For this and many other problems of discrete mathematics, a computer algebra system, like Maple or Mathematica, can prove extremely useful.)

> PROBLEM 3. A die with f faces is cast t times. Show that the probability of observing a total score of S is
>
> $$[z^S]\left(z\frac{1-z^f}{1-z}\right)^t.$$
>
> Use the binomial theorem twice to come up with a finite sum that expresses the probability of observing score S.

[1] In accordance with standard practice in discrete mathematics, we let $[z^n]f(z)$ denote the coefficient of z^n in the power series expansion of f.

A triangulation



Leonhard Euler (1707–1783) arrived at the notion of generating function from a slightly different angle. A question at the time was the following.

> PROBLEM 4. What is the number of ways of decomposing a regular $(n+2)$-gon into n triangles? (Such a decomposition is called a triangulation.)

Euler's approach starts with a *recurrence*. Let T_n be the sought number of triangulations. Then, $T_1 = 1, T_2 = 2$. For larger n, we have the recurrence

$$T_n = \sum_{j=0}^{n-1} T_j T_{n-1-j},$$

where the recurrence is seen to hold provided we assume $T_0 = 1$ (the "empty" triangulation). (Hint: A triangulation splits canonically into two smaller triangulations and its "root triangle" defined as the triangle built on the edge connecting the first and the second vertex of the polygon.) Introduce the generating function

$$T(z) := \sum_{n \geq 0} T_n z^n = 1 + z + 2z^2 + 5z^3 + 14z^4 + 42z^5 + \cdots .$$

Then, the recurrence translates into a *functional equation* over generating function

$$T(z) = 1 + zT(z)^2,$$

as seen from the fact that coefficients of products of functions are given by convolutions. The functional equation is here of a very simple form: it is quadratic and solving it, we find

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

Now, Newton's binomial expansion

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

and a trite calculation with $\alpha = \frac{1}{2}, x = -4z$ shows that

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

The problem is solved exactly by an exact formula and we find easily the following values.

$$\begin{array}{lcl}
T_1 & = & 1 \\
T_2 & = & 2 \\
T_3 & = & 5 \\
T_4 & = & 14 \\
T_5 & = & 42 \\
T_6 & = & 132 \\
T_7 & = & 429 \\
T_8 & = & 1430 \\
T_9 & = & 4862 \\
T_{10} & = & 16796 \\
T_{11} & = & 58786 \\
T_{12} & = & 208012 \\
T_{13} & = & 742900 \\
T_{14} & = & 2674440 \\
T_{15} & = & 9694845 \\
T_{16} & = & 35357670 \\
T_{17} & = & 129644790 \\
T_{18} & = & 477638700 \\
T_{19} & = & 1767263190 \\
T_{20} & = & 6564120420
\end{array}$$

These numbers were later called the Catalan numbers (Eugène Catalan, 1814–1894). They are now known to arise in literally several hundred combinatorial problems!

Symbolic methods. It may be of interest to come back to the enumeration of triangulations from a "symbolic" perspective. We may say that the class $\mathcal{T}$ of all triangulations is specified by an equation

$$\mathcal{T} = \{\epsilon\} + \mathcal{T} \cdot \Delta \cdot \mathcal{T},$$

where Δ denotes a generic triangle and ϵ represents the "empty" triangulation. Observe next that $T(z)$ has an alternative form,

$$T(z) = \sum_{\tau \in \mathcal{T}} z^{|\tau|},$$

with the size $|\tau|$ of a triangulation being its number of triangles. Then, it becomes obvious that $T(z)$ satisfies the functional equation

$$T(z) = 1 + T(z) \cdot z \cdot T(z).$$

We have seen here at work the philosophy of the symbolic method in combinatorial enumerations:

> **Principle I.** *Enumeration of combinatorial objects that are decomposable is best approached by means of generating functions. Functional equations of GF's then directly reflect the decomposable structure of the objects to be enumerated.*

Symbolic methods remained subliminal in the 19th century and in the early 20th century. Then starting around the 1950's, several schools —at MIT (Rota, Stanley, Gessel), the University of Paris (Schützenberger and Foata), and later the university of Quebec (Joyal and the UQAM group)— came up with general frameworks that make Principle I an active principle of combinatorial enumeration. This revival of early ideas was accompanied and supported by a tremendous flow of problems coming from application domains like computer science where the work of Knuth continues to play a remarkable rôle[2]. As a result, there now exist a number of *dictionaries* relating combinatorial constructions and operations on counting generating function. These dictionaries afford a direct translation from the combinatorial world to the realm of generating functions. Here is a small table that gives a flavour of what goes on.

[2]There is added demand regarding quantification of randomness in combinatorics because of the importance of "randomized algorithms". See Motwani–Raghavan, *Randomized Algorithms*, Cambridge, 1995.

Combinatorial constructions	Operations on GF's
null object	1
atom of size 1	z
(disjoint) Union ($\cup$)	$+$
Product	$\times$
Sequence	$\dfrac{1}{1-\cdot}$
Cycle	$\log\dfrac{1}{1-\cdot}$
Set	$\exp(\cdot)$

There are in fact two "universes" of combinatorial problems. Think of the objects to be enumerated as some sort of graphs. Then, the nodes may be all of the same kind or they may alternatively be all distinguishable by, say, being labelled by consecutive integers starting from 1. We speak of the *unlabelled universe* in the first case, and of the *labelled universe* in the second case. To illustrate this fundamental dichotomy, consider the class $\mathcal{L}$ of all linear graphs of the type

$$\mathcal{L} = \{\circ \to \circ \to \circ \to \cdots \to \circ\}.$$

Define as usual the *size* as the number of nodes. Then, as regards counts of objects of size n, one has: (i) in the unlabelled case, $L_n^{\text{unl.}} = 1$ as there is only one type of object for each size; (ii) in the labelled case, $L_n^{\text{lab.}} = n!$ as there are as many objects of size n as there are permutations of the labels $\{1, \ldots n\}$.

Unlabelled object are conveniently enumerated by ordinary generating functions (OGF),

$$\text{OGF:} \quad f_n \mapsto f(z) = \sum_n f_n z^n$$

while labelled objects are normally enumerated by exponential generating functions (EGF),

$$\text{EGF:} \quad f_n \mapsto f(z) = \sum_n f_n \frac{z^n}{n!}.$$

There, f_n is the number of objects of size n in some class $\mathcal{F}$ to be enumerated. Chapters 1 and 2 of the lecture notes by Flajolet and Sedgewick[3] develop thoroughly this approach and cover unlabelled objects and OGF's as well

[3] See `http://algo.inria.fr/flajolet/Publications` for these notes and more.

as labelled objects and EGF's. Proper versions of the basic dictionary are given there. Chapter 3 shows that the symbolic method also applies to the description of characteristic parameters of decomposable structures by means of generating functions.

Excellent references for these topics are:

— at an elementary level: the books by Wilf (*Generatingfunctionology*, Academic Press, 1990), by Graham-Knuth-Patashnik (*Concrete Mathematics*, Addison Wesley, 1989), and by Sedgewick-Flajolet (*An Introduction to the Analysis of Algorithms*, Addison-Wesley Publishing Company, 1996);

— at an advanced level: the books by Comtet (*Advanced Combinatorics*, Reidel, 1974), by Goulden-Jackson (*Combinatorial Enumeration*, Wiley, 1983), and by Bergeron-Labelle-Leroux (*Combinatorial species and tree-like structures*, Cambridge Univ. Press, 1998);

— for applications to the analysis of algorithms, the reference books by Knuth (*The Art of Computer Programming*, Volumes 1–3, Addison-Wesley).

II. Asymptotic Counting

The shortest path between two truths in the real domain passes through the complex plane.

— Jacques Hadamard (1865–1963)

Asymptotics and combinatorics. It is only in very simple cases that couting problems admit of closed-form solutions. Even in this case, it may not be immediately obvious how fast the number of objects of a given size grows with n. One then adopts the point of view of *asymptotic analysis* in order to come up with approximate expressions that may be easier to interpret. As a first example, consider the Catalan numbers. Given Stirling's famous formula that reads

$$n! = n^n e^{-n}\sqrt{2\pi n}\left(1 + O\left(\frac{1}{n}\right)\right),$$

we easily find that the Catalan numbers satisfy

$$T_n \equiv \frac{1}{n+1}\binom{2n}{n} \sim \frac{4^n}{\sqrt{\pi n^3}}.$$

(The formula is asymptotic in the sense that the relative error provided by the approximation tends to 0 as $n \to +\infty$.) As seen here, an asymptotic formula is much easier to interpret: the number of triangulations of an $(n+2)$-gon grows roughly exponentially like 4^n, and there is a further modulation by a subexponential divisor of the form $\sqrt{\pi n^3}$; as an illustration, we can instantly predict that the number T_{1000} has exactly 598 digits in decimal notation, etc.

From asymptotic estimates, we can also recognize easily properties of random combinatorial objects. For instance, the root triangle in a random triangulation of an $(n+2)$-gon lies completely on the left (i.e., is formed with vertices $n, 1, 2$) with probability

$$\varpi_n = \frac{T_{n-1}}{T_n} \sim \frac{1}{4}.$$

The chain followed here is then typical of *elementary asymptotic combinatorics*:

Combinatorial problem

$\downarrow$

Generating function enumeration

$\swarrow$

Exact counts (GF expansion)

$\searrow$

Asymptotic counts

Let us turn next to an example that is borrowed from 19th century recreational mathematics.

> Problem 5. There are n persons that go to the opera and leave their hats at the cloakroom. After the show, each person picks up a hat at random. What is the probability that nobody has his hat on his head.

In more dignified terms, we ask for the fraction of permutations σ of $\{1, \ldots, n\}$ that have no fixed point: $\sigma(x) \neq x$ for each x. Permutations without fixed points are also known as "derangements"; let D_n be their number corresponding to size n. We are obviously in a labelled universe and EGF's must be used. Symbolic methods then grant us that the EGF

$$D(z) := \sum D_n \frac{z^n}{n!}$$

satisfies the (explicit!) functional equation

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

This reads symbolically as follows: a derangement decomposes as a set (the exponential) of cycles (the logarithm) where the derangement property prohibits the singleton cycle (the term z is to be subtracted). Then, by expanding, we get

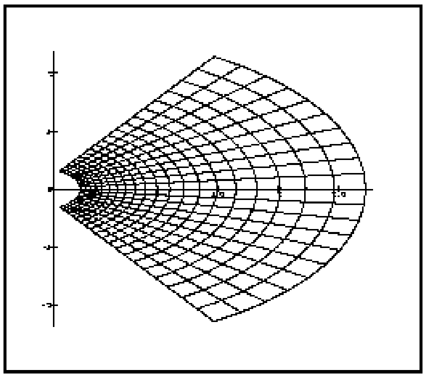
$$\frac{D_n}{n!} = 1 - \frac{1}{1!} + \frac{1}{2!} - \frac{1}{3!} + \frac{(-1)^n}{n!}.$$

The alert reader will recognize the beginning of the expansion of $\exp(-1)$. The series converges fast and we have thus

$$\frac{D_n}{n!} \to e^{-1} \doteq 0.36787,$$

with at least 5 digits of accuracy as soon as $n \geq 10$. Thus the asymptotic answer is: "37% almost exactly for all n greater than 10." This answer is both simple (contrast with the exact formula) and accurate. What we witness here is a great phenomenon of *asymptotic simplification.*

An analytic function ($z \mapsto e^{-z}$) transforms the "unit grid" and preserves angles.



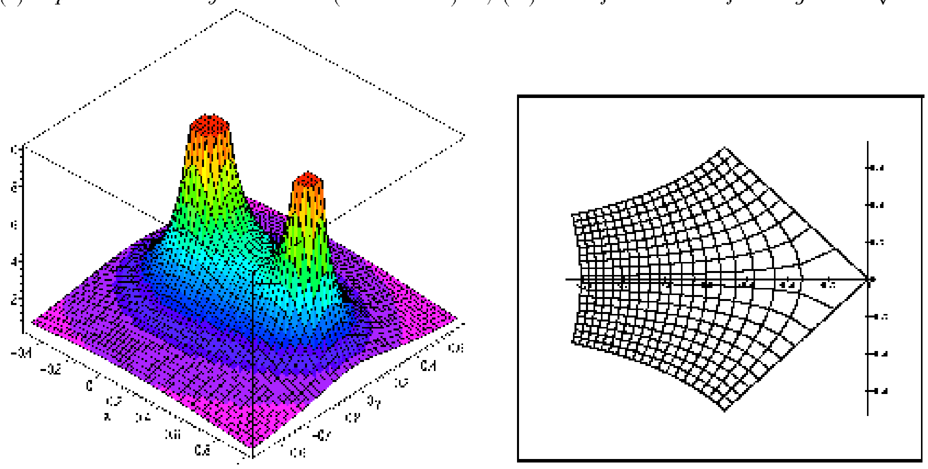
Complex asymptotics and singularity analysis. Real life problems are usually much more complex than these two toys examples. Fortunately, the road that goes through exact expressions (a bottleneck of methods seen so far) can be avoided. It turns out that following early ideas of Cauchy and then Riemann a jump into the realm of *analytic functions* does the job very nicely!.

An analytic function (also known as holomorphic function) is nothing but a function from the complex plane to the complex plane that is complex-differentiable, that is,

$$f'(z) = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

exists independently of the way (= the direction along which) Δz tends to 0. Most classical functions given by explicit expressions are known to be analytic. A point at which a function ceases to be analytic is called a *singularity.*

Two functions with singularities:
(i) representation by modulus: $(1 - z - z^2)^{-1}$; *(ii) transformation of unit grid:* $-\sqrt{1-z}$



The principle[4] that governs complex asymptotics is then the following.

> **Principle II.** *The singularities of a function determine precisely the asymptotic behaviour of its Taylor coefficients. In particular, asymptotic counting results derive from singularities of generating functions.*

[4] A similar principle was recognized early by Bernhard Riemann (1826–1866) in the context of analytic number theory where it gives access to the Prime Number Theorem ("the density of primes near n is $\sim 1/\log n$").

One can then set up a second dictionary. Here is small sample of easy functions with a singularity at 1, together with the exact and asymptotic forms of their Taylor coefficients.

Function	Coeff.	Asymptotics
$-\sqrt{1-z}$	$2\,4^{-n+1}\frac{1}{n}\binom{2n-2}{n-1}$	$\sim \frac{2}{\sqrt{\pi n^3}}$
$\frac{1}{\sqrt{1-z}}$	$4^{-n}\binom{2n}{n}$	$\sim \frac{1}{\sqrt{\pi n}}$
$\frac{1}{1-z}$	1	~ 1
$\frac{1}{1-z}\log\frac{1}{1-z}$	$1+\frac{1}{2}+\cdots+\frac{1}{n}$	$\sim \log n + \gamma\ (\gamma = 0.57721)$
$\frac{1}{(1-z)^2}$	$n+1$	$\sim n$

We see on this examples that larger functions at a singularity lead to larger coefficients. This fact is well explained by an underlying theory that is known as "singularity analysis"[5].

This dictionary applies for functions singular at 1 where it allows us to *transfer* informations on the *local* behaviour of the function at its singularity, 1, directly into an asymptotic estimate of coefficients.

Let's come back to the derangement problem in this perspective. Equipped with this dictionary and the right theorems, we may legitimately reason as follows.

1. The derangement function (EGF) is singular at $z = 1$ (only) where it satisfies

$$D(z) \equiv \frac{e^{-z}}{1-z} \sim \frac{e^{-1}}{1-z}.$$

2. The coefficient of the local singular approximation transfers to the function yielding the valid estimate

$$[z^n]D(z) \sim [z^n]\frac{e^{-z}}{1-z} \equiv e^{-1}.$$

In more pompous terms, we may say that we have a commutative diagram

[5]This theory derives from Cauvhy's integral formula $[z^n]f(z) = \frac{1}{2i\pi}\int_{0+} f(z)\,\frac{dz}{z^{n+1}}$ together with contour deformation techniques.

$$\begin{array}{ccc} \text{Function: } f(z) & \overset{\text{asympt}}{\Longrightarrow} & \sigma(z) \\ \downarrow & & \downarrow \\ \text{Coefficient: } f_n & \overset{\text{asympt}}{\Longrightarrow} & \sigma_n \end{array}$$

We may similarly revisit the Catalan generating function. Scale $T(z)$ by considering $T(z/4)$ which is now singular at 1, that is, scale T_n by considering $T_n/4^n$. The exponent[6] at the singularity is $\frac{1}{2}$, as results from Euler's calculation discussed previously. This square-root singularity gives rise to an $n^{-3/2}$; the scaling that brings the singularity to 1 gives a factor of 4^n, which is consistent with our previous estimates.

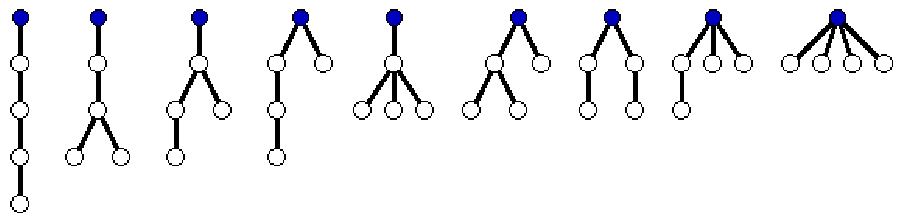
The path taken here is of course of considerably greater generality as it suffices to *estimate* the generating function near its singularity. What we've found is the general case of singularity analysis:

$$\begin{array}{lcl} \text{Location of singularity } (\rho) & \rightsquigarrow & \text{Exponential factor } (\dfrac{1}{\rho})^n \\ \text{Nature of singularity} & \rightsquigarrow & \text{Subexponential factor } (n, \log n, 1, n^{-1.2}) \end{array}$$

From a fundamental perspective, this seemingly abstract detour via complex variables has many advantages. We may well cope with explicit generating functions whose expansions would be horrendous if at all possible, and we may additionally cope with many generating functions that are not even explicit. A precursor of this method in combinatorics was George Pólya (1887–1985) who developed a very general theory for the enumeration of objects under symmetry and then investigated the corresponding asymptotic solutions. A good example treated by Pólya is described below.

> PROBLEM 6. Enumerate structurally different graphs with n nodes that are trees together with a designated root node.

The 9 rooted trees of size $n = 5$
(after `http://sue.csc.uvic.ca/ cos/inf/tree/RootedTree.html`*)*



[6]Singular exponents are reminiscent of critical exponents in phase transition phenomena of statistical physics.

In the unlabelled case (only graphical shapes are considered), Pólya starts from the infinite (!) functional equation

$$U(z) = z \exp\left(\frac{U(z)}{1} + \frac{U(z^2)}{2} + \frac{U(z^3)}{3} + \cdots\right)$$

that represents pretty much the only information available on the problem. Then, one can see that the terms $U(z^2), U(z^3)$, etc, don't affect the nature of the dominant singularity (at $\rho = 0.33832$) that is itself determined by a standard inversion of functions and hence comes out to be of the square–root type—just like in the case of triangulations. From there, one determines easily by singularity analysis that

$$U_n \sim 0.43992 \cdot (2.95576)^n \, n^{-3/2}..$$

This last problem illustrates an advanced chain of considerable generality:

Combinatorial class
↓
Functional equation over GF
↓
Singularity analysis (location & nature)
↓
Asymptotics of coefficients (exponential & polynomial factors)

Chapters 4 and 5 of the lecture notes by Flajolet-Sedgewick provide a thorough discussion of this methodology.

III. Analytic Combinatorics

Parameters and random behaviour. Properties of random combinatorial structures can be approached either by probabilistic methods or by analytic methods. The latter is the path trodden so far and we shall continue our discussion along similar lines. First a motivating example in the form of a problem.

> Problem 7. In the hat-at-the-opera context, what is the probability that no-one has its hat on his head and that no two persons can recover their hats by a pairwise exchange between them? Generalize.

In a more dignified language: how many permutations are such that they have no cycles of length 1 or 2? By symbolic methods, the EGF of such "double derangements" is

$$D^{(2)}(z) = \exp\left(\log\frac{1}{1-z} - z - \frac{z^2}{2}\right) = \frac{e^{-z-z^2/2}}{1-z},$$

as we now need to subtract cycles of lengths 1 and 2. By singularity analysis, the sought probability satisfies asymptotically

$$[z^n]D^{(2)}(z) \sim [z^n]\frac{e^{-3/2}}{1-z} \equiv e^{-3/2}.$$

The answer follows by inspection once the general scheme has been mastered. More generally, the probability that a permutation has no cycle of length $1, \ldots, k$ is instantly recognized to be

$$[z^n]D^{(k)}(z) \sim \exp\left(-\frac{1}{1} - \frac{1}{2} - \cdots - \frac{1}{n}\right) = e^{H_k} \text{ with } H_k = 1 + \cdots + \frac{1}{k},$$

a harmonic number. It may be proven from there that the mean of the size of the smallest cycle in a permutation is asymptotic to $e^{-\gamma}\log n$, where $\gamma \doteq 0.57721$ is Euler's constant. (A plausible reason is that $e^{-H_k} \approx e^{-\gamma}/k$.) We see here that precise results on the behaviour of random objects are attained by very qualitative reasoning on the shape (singularities) of the generating functions, and not on the exact expressions of coefficients that are here already rather formidable multiple alternating sums.

Bivariate generating functions. Many parameters are nicely accessible to the symbolic method. A common framework is that of *bivariate generating functions* (BGF's). Let $\mathcal{F}$ be a combinatorial class, let $|\cdot|$ be size, and let χ be an integer valued parameter; the BGF (of the ordinary type) is defined as

$$F(z,u) = \sum_{\omega\in\mathcal{F}} u^{\chi[\omega]} z^{|\omega|},$$

with a similar definition for exponential BGF's. There are now two parameters[7] according to which enumeration is to be conducted and we introduce accordingly two formal variables. The quantity

$$f_{n,k} = [z^n u^k]F(z,u)$$

[7]The extension to more than two parameters causes no special hardship.

then represents the number of objects of size n with parameter value equal to k. In particular, the sequence with varying k,

$$\varpi_{n,k} = \frac{f_{n,k}}{f_n} = \frac{f_{n,k}}{\sum_k f_{n,k}}$$

describes the probability distribution of χ on objects of size n.

> **Problem 8.** The line interval $[1, n]$ is covered by monomers `[ ]` and dimers `[ | ]`. at random. What is the mean number of monomers in such a random covering?

By symbolic methods, we know that the quasi-inverse $(1 - \cdot)^{-1}$ generates sequences in all possible ways. Indeed, expand symbolically

$$\frac{1}{1 - (\texttt{[ ]} + \texttt{[|]})}$$

into

$$1 + (\texttt{[ ]} + \texttt{[|]}) + (\texttt{[ ][ ]} + \texttt{[ ][|]} + \texttt{[|][ ]} + \texttt{[|][|]}) + \cdots,$$

to realize that all coverings are generated. Now, substitute

$$\texttt{[ ]} \mapsto zu, \qquad \texttt{[|]} \mapsto z^2,$$

giving rise to the OGF and to the (ordinary) BGF

$$F(z) = \frac{1}{1 - z - z^2}, \qquad F(z, u) = \frac{1}{1 - zu - z^2}.$$

From there (take a partial derivative with respect to u and set $u = 1$), it is not hard to find that the mean is asymptotic to

$$\frac{1}{\phi^2} n \doteq 0.38196\, n, \qquad \phi = \frac{1 + \sqrt{5}}{2},$$

the golden ratio. Details depend on the Fibonacci numbers (Leonardo Fibonacci: 1170–1250),

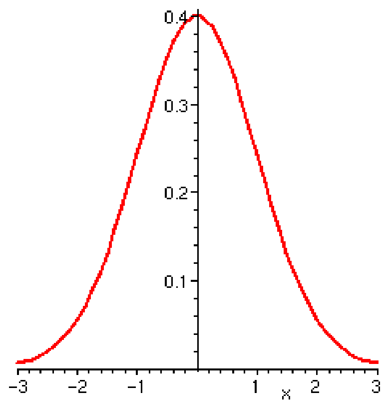
$$0, 1, 1, 2, 3, 5, 8, 13, 21, 35, \ldots, \qquad f_{n+2} = f_{n+1} + f_n.$$

Again, looking at the problem more abstractly in terms of singularities — especially the one nearest to the origin, called "dominant"— makes it possible to solve a vast generalization with very little computation being involved.

Problem 9. The line interval $[1, n]$ is covered by j-mers of size $j = 1, \ldots, m$ at random (m a fixed parameter). What is asymptotically (as $n \to +\infty$) the mean number of j-mers in such a random covering?

A plot of the Gaussian (or normal) density function

$$\Pr(X \in (x, x + dx)) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx.$$



Singularity perturbation. Now the royal path to this and many more combinatorial problems of a similar flavour is once more via complex asymptotic methods. This was pioneered by Goncharov (Moscow, 1943) and Bender (La Jolla, 1970+) following work started by Laplace (Pierre-Simon Laplace, 1749–1827) on the central limit theorem (CLT) and sums of independent random variables[8]. In general, one has available a functional equation on a BGF of the form

$$\Phi(z, u, F(z, u)) = 0,$$

where Φ is an analytic operator that reflects the combinatorics of the problem at hand. For $u = 1$, we're back to the basic counting problem and we may assume the location $\rho(1)$ and the nature (e.g., the corresponding singular exponent $\alpha(1)$) to be known. Analytic relations tend to be robust, so that, in most cases, we'll have $F(z, u)$ that has a singularity characterized by a pair $\langle \rho(u), \alpha(u) \rangle$ (singularity location–singular exponent) when u is near 1. This indicates that the *probability generating function* (PGF) of

[8]The central limit theorems asserts that a large number of random fluctuations have a resultant (i.e., sum) that is asymptotically Gaussian.

the parameter of interest, namely

$$g_\chi(u) = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}$$

must satisfy

$$g_\chi(u) \sim \left(\frac{\rho(1)}{\rho(u)}\right)^n \cdot n^{\alpha(u)-\alpha(1)} \approx \left(\frac{\rho(1)}{\rho(u)}\right)^n.$$

Thus, the PGF of χ is approximately a large power. From this, by differentiation, we'll find that the mean μ_n and the variance σ_n^2 of χ (taken on random objects of size n) both grow linearly in n.

The standardized variable

$$X_n^* = \frac{\chi - \mu_n}{\sigma_n}$$

is then such that its Fourier transform ($E(.)$ denotes expectation) is related to $g_\chi(u)$ by

$$E\left(e^{itX_n^*}\right) = g_\chi\left(e^{itX_n^*}\right) e^{-it\mu_n/\sigma_n}.$$

A simple calculation then shows that, as $n \to +\infty$,

$$E\left(e^{itX_n^*}\right) \to e^{-t^2/2} = \int_{-\infty}^{+\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-x^2/2}\right) e^{ixt}\, dx.$$

By the continuity of Fourier transforms (Lévy's theorem), this implies that the distribution of χ on objects of size n tends to become Gaussian[9] as n gets large. This scheme is again of considerable generality across decomposable combinatorial structures.

> **Principle III.** *For decomposable objects and parameters, algebraic and analytic relations on counting generating functions usually lift to relations between bivariate generating functions. Then* singularity perturbation analysis *shows parameters to admit limit laws, often of the Gaussian type.*

[9] What is meant precisely is that the distribution function $\Pr(X_n^* \leq x)$ converges pointise to $\Pr(X_\infty \leq x)$, with X_∞ a standard Gaussian variable (alternative names: convergence in law or in distribution, weak convergence).

From there, research in the field proceeds with a classification of the major classes of combinatorial objects and the associated probabilistic phenomena. Examples covered include *words and languages*, *trees* of various kinds, *graphs*, *permutations*, as well as *allocations* satisfying various constraints, etc. Most can be grouped into general *schemas* that it is of some interest to recognize and organize Sedgewick and I propose to call this field "*analytic combinatorics*". This approach in turns ties well with the analysis of algorithms, especially the ones of *polynomial time complexity* (P) that are often based on decomposability (data structures, the "divide-and-conquer" paradigm, etc).

The accompanying notes present the basic methods and theorems of analytic combinatorics. The domain nicely complements some of the other approaches used to investigate randomness in combinatorics, most notably:

— statistical physics methods are occasionally used in combinatorial optimization problems of the NP type and turn out to be very fruitful in the case of configurations on a regular geometric structure (e.g., lattice configurations);

— probabilistic methods are instrumental in random graph problems[10] and more generally in combinatorial optimization where one has to operate with objects that do not decompose nicely; then, tail inequalities, correlated random variables, martingales, functional limit processes, etc, play an important rôle.

[10]See Bollobás, *Random Graphs*, Academic Press, 1985.

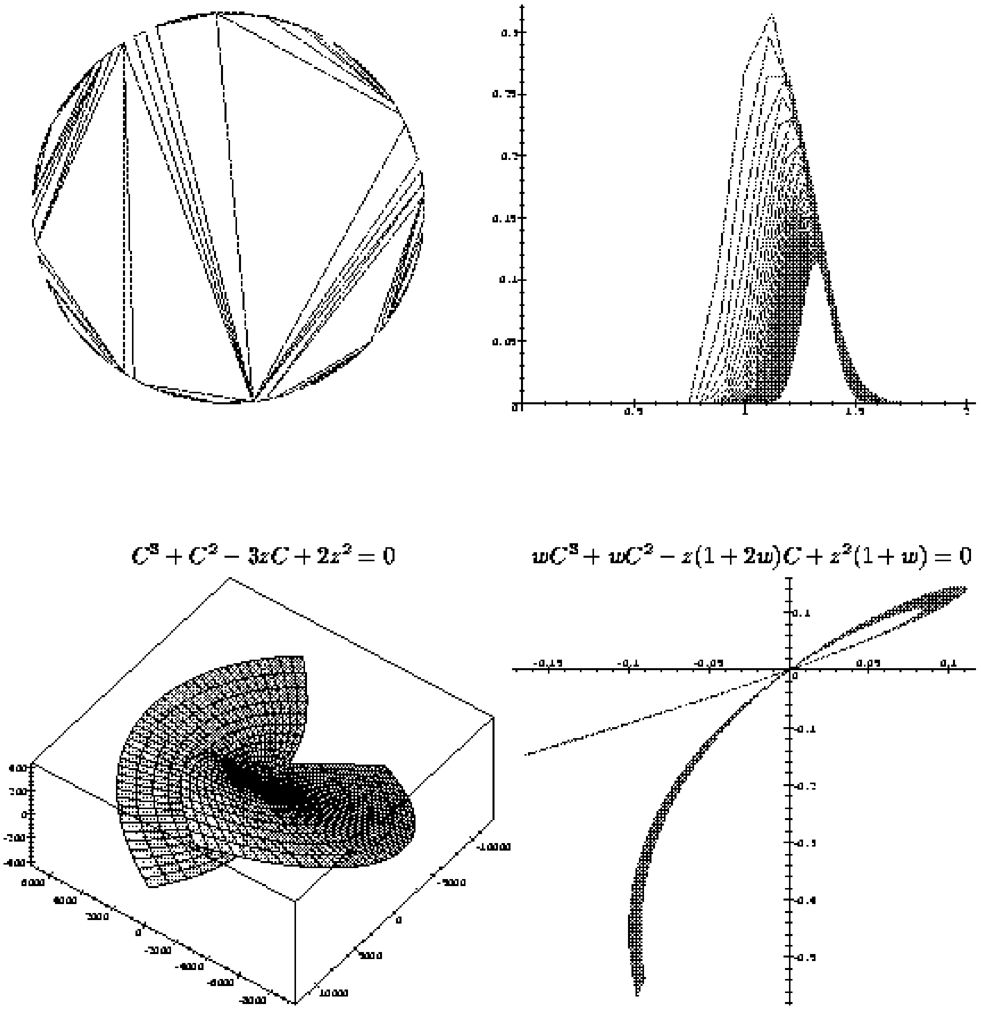


Figure 1: "Analytic combinatorics" of non-crossing graphs: (top-left) a non-crossing connected graph of size 100 drawn at random; (bottom-left) the generating function equation and a rendering of its imaginary part; (bottom-right) singularity perturbation analysis showing a consistently smooth displacement of a square-root singularity for the BGF of number of edges; (top-right) the resulting Gaussian law for the number of edges. (After P. Flajolet, M. Noy. In *Discrete Mathematics* **204**(1-3), 1999, 203-229.)