Alcom-FT Technical Report Series
ALCOMFT-TR-03-100

ANALYTIC COMBINATORICS
—
BASIC COMPLEX ASYMPTOTICS

PHILIPPE FLAJOLET & ROBERT SEDGEWICK

Algorithms Project
INRIA Rocquencourt
78153 Le Chesnay
France

Department of Computer Science
Princeton University
Princeton, NJ 08540
USA

First Edition, November 27, 2003
♡ ♡ ♡ ♡ ♡ ♡

ABSTRACT

This booklet develops in about 120 pages the basics of asymptotic enumeration through an approach that revolves around generating functions and complex analysis. Major properties of generating functions that are of interest here are singularities. The text presents the core of the theory with two chapters on complex analytic methods focusing on rational and meromorphic functions. It is largely oriented towards applications of complex analysis to asymptotic enumeration and asymptotic properties of random discrete structures. Many examples are given that relate to words, integer compositions, paths and walks in graphs, lattice paths, trees, and constrained permutations.

Acknowledgements. This work was supported in part by the IST Programme of the EU under contract number IST-1999-14186 (ALCOM-FT). This booklet would be substantially different (and much less informative) without Neil Sloane's *Encyclopedia of Integer Sequences* and 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.

This 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 contains Chapters IV and V; it is a continuation of "Analytic Combinatorics—Symbolic Methods" (by Flajolet & Sedgewick, 2002). Readers are encouraged to check Philippe Flajolet's web pages for regular updates and new developments.

PREFACE

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 a function's 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 Methods* develops systematic "symbolic" relations between some of the major constructions of discrete mathematics and operations on generating functions which exactly encode counting sequences.

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

— *Random Structuress* 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 second 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.

The present booklet specifically exposes *Singular Combinatorics*, which is a unified analytic theory dedicated to the process of extractic asymptotic information from counting generating functions. As it turns out, a collection of general (and simple) theorems provide a systematic translation mechanism between generating functions and asymptotic forms of coefficients. Two chapters compose this booklet. Chapter IV serves as an *introduction to complex-analytic methods* and proceeds with the treatment of *meromorphic functions*, that is, functions whose only singularities are poles, *rational functions* being the simplest case. Chapter V develops *applications of rational and meromorphic asymptotics*, with numerous applications related to words and languages, walks and graphs, as well as permutations. [Future chapters will treat Singularity Analysis (Chapter VI) and its Applications (Chapter VII).]

Contents

CHAPTER IV

Complex Analysis, Rational and Meromorphic Asymptotics

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

— JACQUES HADAMARD

Generating functions are a central concept of combinatorial theory. So far, they have been treated as formal objects, that is, as formal power series. The major theme of Chapters I–III has indeed been to demonstrate how the algebraic structure of generating functions directly reflects the structure of combinatorial classes. From now on, we examine generating functions in the light of *analysis*. This means assigning *values* to the variables that appear in generating functions.

Comparatively little benefit results from assigning only real values to the variable z that figures in a univariate generating function. In contrast assigning *complex* values turns out to have serendipitous consequences. In so doing, a generating function becomes a geometric transformation of the complex plane. This transformation is very regular near the origin—one says that it is analytic or holomorphic. In other words, it only effects initially a smooth distortion of the complex plane.

Farther away from the origin, some "cracks" start appearing in the picture. These cracks—the dignified name is "singularities"—correspond to the disapperance of smoothness. What happens is that knowledge of a function's singularities provide a wealth of information regarding the function's coefficients, and especially their asymptotic rate of growth. Adopting a geometric point of view has a large pay-off.

By focussing on singularities, analytic combinatorics treads in the steps of many respectable older areas of mathematics. For instance, Euler recognized that the fact for the Riemann zeta function $\zeta(s)$ to become infinite at 1 implies the existence of infinitely many prime numbers, while Riemann, Hadamard, and de la Vallée-Poussin uncovered much deeper connections between quantitative properties of the primes and singularities of $1/\zeta(s)$.

In this chapter, we start by recalling the elementary theory of analytic functions and their singularities in a style tuned to the needs of combinatorial theory. Cauchy's integral formula expresses coefficients of analytic functions as contour integrals. Suitable uses of Cauchy's integral formula then make it possible to estimate such coefficients by suitably selecting the contour of integration. For the fairly common case of functions that have singularities at a finite distance, the *exponential growth* formula relates the *location* of the singularities closest to the origin (these are also known as "dominant" singularities) to the exponential order of growth of coefficients. The nature of these *singularities* then dictates the fine structure of the asymptotic of the function's coefficients, especially the *subexponential factors* involved. In this chapter we carry out this programme for rational functions and meromorphic functions, where the latter are defined by the fact their singularities are of the polar type.

Elementary techniques permit us to estimate asymptotically counting sequences, when these are already presented to us in closed form or as simple combinatorial sums. The methods to be exposed require no such explicit forms of counting coefficients to be available. They apply to almost any conceivable combinatorial generating function that has a decent mathematical expression—we already know from Chapters I–III that this covers a very large fragment of elementary combinatorics. In a large number of cases, complex-analytic methods can even be applied to generating functions only accessible *implicitly* from *functional equations*. This paradigm will be extensively explored in this chapter with applications found in denumerants, derangements, surjections, alignments, and several other structures introduced in Chapters I–III.

IV. 1. Generating functions as analytic objects

Generating functions, considered previously as purely *formal* objects subject to algebraic operations, are now going to be interpreted as *analytic* objects. In so doing one gains an easy access to the asymptotic form of their coefficients. This informal section offer a glimpse of themes that form the basis of this chapter and the next one.

In order to introduce the subject softly, let us start with two simple generating functions, one, $f(z)$, being the OGF of the Catalan numbers (starting at index 1), the other, $g(z)$, being the EGF of derangements:

$$f(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right), \qquad g(z) = \frac{\exp(-z)}{1-z}. \tag{1}$$

At this stage, the forms above are merely compact descriptions of formal power series built from the elementary series

$$\begin{array}{lclclcl} (1-u)^{-1} & = & 1 + u + u^2 + \cdots, & \qquad & (1-u)^{1/2} & = & 1 - \frac{1}{2}u - \frac{1}{8}u^2 - \cdots, \\ \exp(u) & = & 1 + \frac{1}{1!}u + \frac{1}{2!}u^2 + \cdots, & & & & \end{array}$$

by standard composition rules. Accordingly, the coefficients of both GFs are known in explicit form

$$f_n := [z^n]f(z) = \frac{1}{n}\binom{2n-2}{n-1}, \qquad g_n := [z^n]g(z) = \left(\frac{1}{0!} - \frac{1}{1!} + \frac{1}{2!} - \cdots + \frac{(-1)^n}{n!}\right).$$

Next, Stirling's formula and comparison with the alternating series giving $\exp(-1)$ provide respectively

$$f_n \underset{n\to\infty}{\sim} \frac{4^n}{\sqrt{\pi n^3}}, \qquad g_n = \underset{n\to\infty}{\sim} e^{-1} \doteq 0.36787. \tag{2}$$

Our purpose is to examine, heuristically for the moment, the relationship between the asymptotic forms (2) and the structure of the corresponding generating functions in (1).

Granted the growth estimates available for f_n and g_n, it is legitimate to substitute in the power series expansions of the GFs $f(z)$ and $g(z)$ any real or complex value of a small enough modulus, the upper bounds on modulus being $\rho_f = \frac{1}{4}$ (for f) and $\rho_g = 1$ (for g). Figure 1 represents the graph of the resulting functions when such real values are assigned to z. The graphs are smooth, representing functions that are differentiable any number of times for z interior to the interval $(-\rho, +\rho)$. However, at the right boundary point, smoothness stops: $g(z)$ become infinite at $z = 1$, and so it even ceases to be finitely

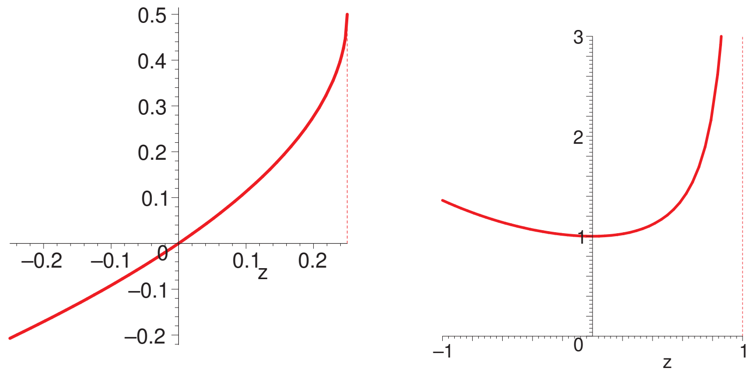


FIGURE 1. Left: the graph of the Catalan OGF, $f(z)$, for $z \in (-\frac{1}{4}, +\frac{1}{4})$; right: the graph of the derangement EGF, $g(z)$, for $z \in (-1, +1)$.

defined; $f(z)$ does tend to the limit $\frac{1}{2}$ as $z \to (\frac{1}{4})^-$, but its derivative becomes infinite there. Such special points at which smoothness stops are called *singularities*, a term that will acquire a precise meaning in the next sections.

Observe also that, by the usual process of analysis, $f(z)$ and $g(z)$ can be *continued* in certain regions, when use is made of the global expressions (1) while $\exp$ and $\sqrt{\ }$ are assigned their usual real-analytic interpretation; for instance:

$$f(-1) = \frac{1}{2}\left(1 - \sqrt{5}\right), \qquad g(-2) = \frac{e^2}{3}.$$

Such "continuation" properties (to the complex realm) will prove essential in developing efficient methods for coefficient asymptotics.

One may proceed similarly with complex numbers, starting with numbers whose modulus is less than the radius of convergence of the series defining the GF. Figure 2 displays the images of regular grids by f and g. This illustrates the fact that a regular grid transforms into an orthogonal network of curves and more precisely that f and g preserve angles—this property corresponds to complex differentiability and is equivalent to analyticity to be introduced shortly. The singularity of f is clearly perceptible on the right of its diagram, since, at $z = \frac{1}{4}$ corresponding to $f(z) = \frac{1}{2}$, the function f folds lines and divides angles by a factor of 2.

Let us now turn to coefficient asymptotics. As is expressed by (2), the coefficients f_n and g_n each belong to a general asymptotic type,

$$A^n \theta(n),$$

corresponding to an exponential growth factor A^n modulated by a tame factor $\theta(n)$, which is subexponential; compare with (2). Here, one has $A = 4$ for f_n and $A = 1$ for g_n; also, $\theta(n) = \frac{1}{4}(\sqrt{\pi n^3})^{-1}$ for f_n and $\theta(n) = e^{-1}$ for g_n. Clearly, A should be related to the radius of convergence of the series. We shall see that, on very general grounds, the exponential rate of growth is given by $A = 1/\rho$, where ρ is the first singularity encountered along the positive real axis. In addition, under general complex-analytic conditions, it will be established that $\theta(n) = O(1)$ is systematically associated to a simple pole of the generating function, while $\theta(n) = O(n^{-3/2})$ systematically arises from a singularity that is of the square-root type. In summary, as this chapter and the next ones will copiously illustrate, one has:

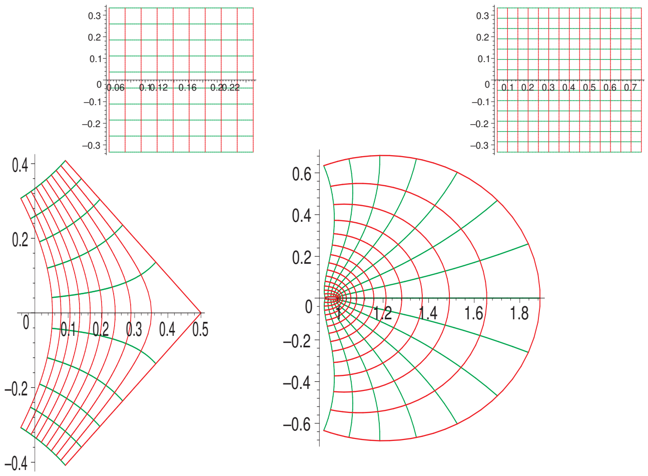


FIGURE 2. The images of regular grids by $f(z)$ (left) and $g(z)$ (right).

> **Fundamental principle of complex coefficient asymptotics.** *The* location *of a function's singularities dictates the* exponential growth *of the function's coefficient, A^n, while the* nature *of the function at its singularities determines the* subexponential factor, $\theta(n)$.

Observe that the rescaling rule,

$$[z^n]f(z) = \rho^{-n}[z^n]f(\rho z),$$

enables one to normalize functions so that they are singular at 1, and so "explains" the fact that the location of a function's singularities should influence the coefficients' approximation by exponential factors. Then various theorems, starting with Theorems IV.6 and IV.7, provide sufficient conditions under which the following central implication is valid,

$$h(z) \sim \sigma(z) \quad \Longrightarrow \quad [z^n]h(z) \sim [z^n]\sigma(z), \tag{3}$$

where $h(z)$ is a function singular at 1 whose Taylor coefficients are to be estimated and $\sigma(z)$ is an approximation near a singularity—usually σ is a much simpler function, typically like $(1-z)^\alpha \log^\beta(1-z)$ whose coefficients are easy to find. Under such conditions, it suffices to estimate a function locally in order to derive its coefficients asymptotically. In other words, the relation (3) provides a mapping between asymptotic scales of functions near singularities and asymptotics scales of coefficients.

$\triangleright$ **1.** *Elementary transfers.* Elementary series manipulation yield the following general result: *Let $h(z)$ be a power series with radius of convergence > 1 and assume that $h(1) \neq 0$; then one has*

$$[z^n]\frac{h(z)}{1-z} \sim h(1), \qquad [z^n]h(z)\sqrt{1-z} \sim -\frac{h(1)}{2\sqrt{\pi n^3}}, \qquad [z^n]h(z)\log\frac{1}{1-z} \sim \frac{h(1)}{n}.$$

See Bender's survey [**9**] for many similar statements. $\triangleleft$

$\triangleright$ **2.** *Asymptotics of generalized derangements.* The EGF of permutations without cycles of length 1 and 2 satisfies

$$j(z) = \frac{e^{-z-z^2/2}}{1-z} \qquad \text{with} \qquad j(z) \underset{z\to 1}{\sim} \frac{e^{-3/2}}{1-z}.$$

Analogy with derangements suggests (Note 1 can justify it) that $[z^n]j(z) \underset{n\to\infty}{\sim} e^{-3/2}$. Here is a table of exact values of $[z^n]j(z)$ (with relative error of the approximation by $e^{-3/2}$ in parentheses):

	$n=5$	$n=10$	$n=20$	$n=50$
j_n :	0.2	0.22317	0.2231301600	0.2231301601484298289332804707640122
error :	(10^{-1})	$(2\cdot 10^{-4})$	$(3\cdot 10^{-10})$	(10^{-33})

The quality of the asymptotic approximation is extremely good. (Such a property is invariably attached to polar singularities.) $\lhd$

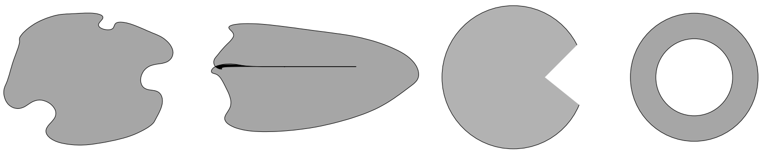
IV. 2. Analytic functions and meromorphic functions

Analytic functions are the primary mathematical concept for complex asymptotics. They can be characterized in two essentially equivalent ways (Subsection IV. 2.1): by means of convergent series expansions (à la Cauchy and Weierstraß) and by differentiability properties (à la Riemann). The first aspect is directly related to the use of generating functions for enumeration; the second one allows for a powerful abstract discussion of closure properties that usually requires little computation. *Meromorphic functions* are nothing but quotients of analytic functions.

Integral calculus with analytic or meromorphic functions (developed in Subsection IV. 2.2) assumes a shape radically different from what it is in the real domain: integrals become quintessentially independent of details of the integration contour, the residue theorem being a prime illustration of this fact. Conceptually, this makes it possible to relate properties of a function at a point (e.g., the coefficients of its expansion at 0) to its properties at another far-away point (e.g., its residue at a pole).

The presentation in this section and the next one is an informal review of basic properties of analytic functions tuned to the needs of asymptotic analysis of counting sequences. For a detailed treatment, we refer the reader to one of the many excellent treatises on the subject, like the books by Dieudonné [**28**], Henrici [**66**], Hille [**67**], Knopp [**72**], Titchmarsh [**109**], or Whittaker and Watson [**114**].

IV. 2.1. Basics. We shall consider functions defined in certain *regions* of the complex domain $\mathbb{C}$. By a region is meant an open subset Ω of the complex plane that is connected. Here are some examples:



simply connected domain slit complex plane indented disc annulus

Classical treatises teach us how to extend to the complex domain the standard functions of real analysis: polynomials are immediately extended as soon as complex addition and multiplication have been defined, while the exponential is definable by means of Euler's formula, and one has for instance

$$z^2 = (x^2 - y^2) + 2ixy, \qquad e^z = e^x \cos y + i e^x \sin y,$$

if $z = x + iy$. Both functions are consequently defined over the whole complex plane $\mathbb{C}$.

The square-root and the logarithm are conveniently described in polar coordinates by

$$\sqrt{z} = \sqrt{\rho}\, e^{i\theta/2}, \qquad \log z = \log \rho + i\theta, \tag{4}$$

if $z = \rho e^{i\theta}$. One can take the domain of validity of (4) to be the complex plane slit along the axis from 0 to $-\infty$, that is, restrict θ to the open interval $(-\pi, +\pi)$, in which case the definitions above specify what is known as the *principal determination*. There is no way for instance to extend by continuity the definition of $\sqrt{z}$ in any domain containing 0 in its interior since, for $a > 0$ and $z \to -a$, one has $\sqrt{z} \to i\sqrt{a}$ as $z \to -a$ from above, while $\sqrt{z} \to -i\sqrt{a}$ as $z \to -a$ from below. (The point $z = 0$ where two determinations "meet" is accordingly known as a *branch point*.)

First comes the main notion of an analytic function that arises from convergent series expansions.

DEFINITION IV.1. *A function $f(z)$ defined over a region Ω is* analytic *at a point $z_0 \in \Omega$ if, for z in some open disc centred at z_0 and contained in Ω, it is representable by a convergent power series expansion*

$$f(z) = \sum_{n\geq 0} c_n (z - z_0)^n. \tag{5}$$

A function is analytic in a region Ω iff it is analytic at every point of Ω.

As derives from an elementary property of power series, given a function f that is analytic at a point z_0, there exists a disc (of possibly infinite radius) with the property that the series representing $f(z)$ is convergent for z inside the disc and divergent for z outside the disc. The disc is called the *disc of convergence* and its radius is the *radius of convergence* of $f(z)$ at $z = z_0$.

The next important notion is a geometric one.

DEFINITION IV.2. *A function $f(z)$ defined over a region Ω is called* complex-differentiable *(also* holomorphic*) at z_0 if the limit, for complex δz,*

$$\lim_{\delta z \to 0} \frac{f(z_0 + \delta z) - f(z_0)}{\delta z}$$

exists. (In particular, the limit is independent of the way δz tends to 0.) This limit is denoted as usual by $f'(z_0)$ or $\frac{d}{dz} f(z)\Big|_{z_0}$ A function is complex-differentiable in Ω iff it is differentiable at every $z_0 \in \Omega$.

Clearly, if $f(z)$ is complex differentiable at z_0, it acts locally as a linear transformation,

$$f(z) - f(z_0) \sim f'(z_0)(z - z_0),$$

whenever $f'(z_0) \neq 0$. Then $f(z)$ locally behaves like a similarity transformation (composed of a translation, a rotation, and a scaling). In particular, it preserves angles[1] and infinitesimal squares get transformed into infinitesimal squares; see Figure 3 for a rendering

It follows from a well known theorem of Riemann (see for instance [**66**, vol. 1, p 143]) that analyticity and complex differentiability are equivalent notions.

First fundamental property of analytic function theory. *A function is analytic in a region Ω if and only if it is complex-differentiable in Ω.*

$\triangleright$ **3.** *Analyticity implies complex-differentiability*. Let $f(z)$ be analytic at 0. Then its derivatives at a point z_0 within the disc of convergence of its expansion at 0 can be obtained by differentiating the series representation of f termwise. Thus: *analytic implies complex-differentiable.* (The converse property requires integration properties and is discussed in Note 10 below.) $\triangleleft$

[1] A mapping that preserves angles is also called a *conformal* map.

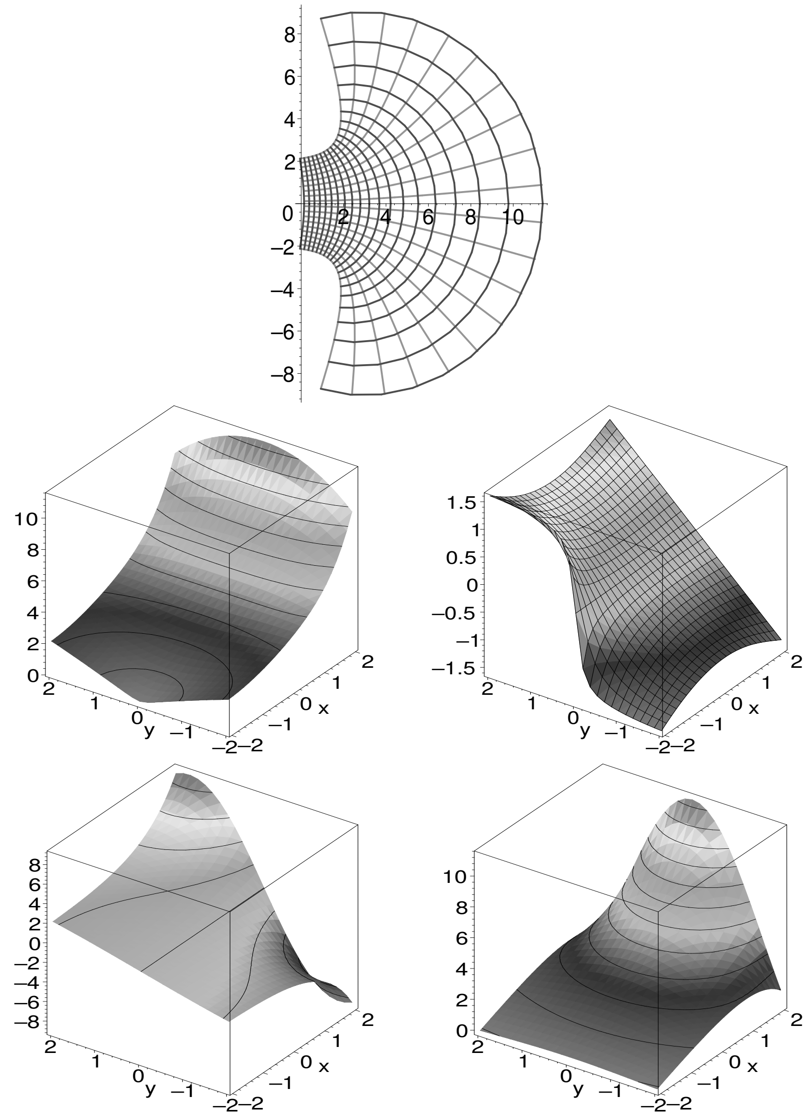


FIGURE 3. Multiple views of an analytic function. The image of the domain $\Omega = \{z \mid |\Re(z)| \leq 2, |\Im(z)| \leq 2\}$ by the function $f(z) = \exp(z) + z + 2$: (top) transformation of a square grid in Ω by f; (middle) the modulus and argument of $f(z)$; (bottom) the real and imaginary parts of $f(z)$.

$\triangleright$ **4.** *Taylor's formula for analytic functions*. With the conventions of Note 3 and as a consequence of simple series rearrangements: *Taylor's formula holds at z_0 and one has*

$$f(z_0 + h) = \sum_{k=0}^{\infty} f^{(k)}(z_0)\frac{h^k}{k!}, \qquad f^{(k)}(z) = \frac{d^k}{dz^k} f(z).$$

for all small enough h. $\triangleleft$

▷ **5.** *Cauchy–Riemann equations.* Let $P(x,y) = \Re f(x+iy)$ and $Q(x,y) = \Im f(x+iy)$. By adopting successively in the definition of complex differentiability $\delta z = h$ and $\delta z = ih$, one finds

$$\frac{\partial P}{\partial x} + i\frac{\partial Q}{\partial x} = \frac{\partial Q}{\partial y} - i\frac{\partial P}{\partial y},$$

implying $P'_x = Q'_y$ and $P'_y = -Q'_x$, known as the Cauchy–Riemann equations. The functions P and Q satisfy the partial differential equations $\Delta f = 0$, where Δ is the 2-dimensional *Laplacian* $\Delta := \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$; such functions are known as *harmonic functions*. ◁

We finally introduce meromorphic functions. The quotient of two analytic functions $f(z)/g(z)$ ceases to be analytic at a point a where $g(a) = 0$. However, a simple structure for quotients of analytic functions prevails.

DEFINITION IV.3. *A function $h(z)$ is* meromorphic *at $z = z_0$ iff in a neighbourhood of $z = z_0$ with $z \neq z_0$ it is representable by an expansion of the form*

$$h(z) = \sum_{n \geq -M} h_n (z - z_0)^n. \tag{6}$$

If $h_{-M} \neq 0$, then $h(z)$ is said to have a pole of order M at $z = a$. The coefficient h_{-1} is called the residue *of $h(z)$ at $z = a$ and is written as*

$$\operatorname{Res}[h(z); z = a].$$

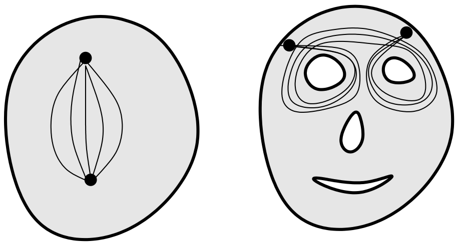
A function is meromorphic in a region iff it is meromorphic at any point of the region.

Equivalently, $h(z)$ is meromorphic at $z = z_0$ iff, in a neighbourhood of z_0, it can be represented as $f(z)/g(z)$, with $f(z)$ and $g(z)$ being analytic at $z = z_0$.

IV. 2.2. Integrals and residues. Integrals along curves in the complex plane are defined in the usual way from curvilinear integrals applied to the real and imaginary parts of the integrand. However integral calculus in the complex plane is of a radically different nature from what it is on the real line—in a way it is much simpler and much more powerful.

A path in a region Ω is described by its parameterization, which is a continuous function γ mapping $[0,1]$ into Ω. Two paths γ, γ' in Ω having the same end points are said to be *homotopic* (in Ω) if one can be continuously deformed into the other while staying within Ω as in the following examples:

homotopic paths:



A closed path is defined by the fact that its end points coincide: $\gamma(0) = \gamma(1)$, and a path is *simple* if the mapping γ is one-to-one. A closed path is said to be a *loop* of Ω if it can be continuously deformed *within* Ω to a single point; in this case one also says that the path is homotopic to 0. In what follows we implicitly restrict attention to paths that are assumed to be rectifiable. Unless otherwise stated, all integration paths will be assumed to be oriented positively.

One has:

Second fundamental property of analytic function theory. *Let f be analytic in Ω and let λ be a loop of Ω. Then $\int_\lambda f = 0$.*

Equivalently, for f analytic in Ω, one has

$$\int_\gamma f = \int_{\gamma'} f, \tag{7}$$

provided γ and γ' are homotopic in γ.

$\triangleright$ **6.** *Proof of the Second Fundamental Principle from analyticity.* Let f be analytic in Ω. It suffices to justify

$$\int_\lambda \left[\sum_{n\ge 0} f_n z^n\right] dz = \sum_{n\ge 0} f_n \left[\int_\lambda z^n \, dz\right] = 0.$$

(The proof does not logically require the First Fundamental Principle.) $\triangleleft$

$\triangleright$ **7.** *Proof of the Second Fundamental Principle from differentiability.* Let f be complex-differentiable in Ω. Then the relation (7) holds. (The proof relies on the Cauchy–Riemann equations guaranteeing that the curvilinear integrals only depend on the endpoints of the contour; it does not logically require the First Fundamental Principle.) $\triangleleft$

The important *Residue Theorem* due to Cauchy relates *global* properties of a meromorphic function, its integral along closed curves, to purely *local* characteristics at designated points, the residues at poles.

THEOREM IV.1 (Cauchy's residue theorem). *Let $h(z)$ be meromorphic in the region Ω and let λ be a simple loop in Ω along which the function is analytic. Then*

$$\frac{1}{2i\pi}\int_\lambda h(z)\, dz = \sum_s \operatorname{Res}[h(z); z = s],$$

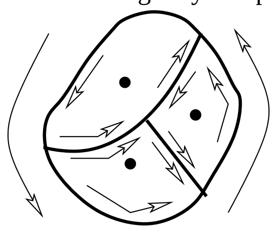
where the sum is extended to all poles s of $h(z)$ enclosed by λ.

PROOF. (Sketch) To see it in the representative case where $h(z)$ has only a pole at $z = 0$, observe by appealing to primitive functions that

$$\int_\lambda h(z)\, dz = \sum_{\substack{n\ge -M \\ n\ne -1}} h_n \left[\frac{z^{n+1}}{n+1}\right]_\lambda + h_{-1}\int_\lambda \frac{dz}{z},$$

where the bracket notation $\big[u(z)\big]_\lambda$ designates the variation of the function $u(z)$ along the contour λ. This expression reduces to its last term, itself equal to $2i\pi h_{-1}$, as is checked by using integration along a circle (set $z = re^{i\theta}$). The computation extends by translation to the case of a unique pole at $z = a$.

In the case of multiple poles, we observe that the simple loop can only enclose finitely many poles (by compactness). The proof then follows from a simple decomposition of the interior domain of λ into cells each containing only one pole. Here is an illustration



in the case of three poles. (Contributions from internal edges cancel.) $\square$

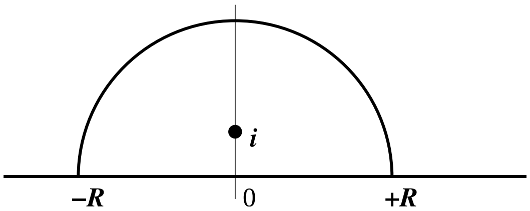
Here is a textbook example of such a reduction from global to local properties. Define the integrals

$$I_m := \int_{-\infty}^{\infty} \frac{dx}{1+x^{2m}},$$

and consider specifically I_1. Elementary calculus teaches us that $I_1 = \pi$ since the anti-derivative of the integrand is an arc tangent:

$$I_1 = \int_{-\infty}^{\infty} \frac{dx}{1+x^2} = [\arctan x]_{-\infty}^{+\infty} = \pi.$$

In the light of the residue theorem, we first consider the integral over the whole line as the limit of integrals over large intervals of the form $[-R, +R]$, then complete the contour of integration by means of a large semi-circle in the upper half-plane, as shown below:



Let γ be the contour comprised of the interval and the semi-circle. Inside γ, the integrand has a pole at $x = i$ ($i = \sqrt{-1}$), where

$$\frac{1}{1+x^2} \equiv \frac{1}{(x+i)(x-i)} = -\frac{i}{2}\frac{1}{x-i} + \frac{1}{4} + \frac{i}{8}(x-i) + \cdots,$$

so that its residue there is $-i/2$. Thus, by the residue theorem, the integral taken over γ is equal to $2\pi i$ times the residue of the integrand at i. As $R \to \infty$, the integral along the semi-circle vanishes (it is $O(R^{-1})$) while the integral along the real segment gives I_1 in the limit. There results the relation giving I_1:

$$I_1 = 2i\pi \operatorname{Res}\left(\frac{1}{1+x^2}, x = i\right) = \pi.$$

Remarkably, the evaluation of the integral in this perspective rests entirely upon the local expansion of the integrand at a special point (the point i).

$\triangleright$ **8.** *The general integral I_m.* Let $\alpha = \exp(\frac{i\pi}{2m})$ so that $\alpha^{2m} = -1$. Contour integration of the type used for I_1 yields

$$I_m = 2i\pi \sum_{j=1}^{m} \operatorname{Res}\left(\frac{1}{1+x^{2m}}; x = \alpha^{2j-1}\right),$$

while, for any $\beta = \alpha^{2j-1}$ with $1 \le j \le m$, one has

$$\frac{1}{1+x^{2m}} \underset{x\to\beta}{\sim} \frac{1}{2m\beta^{2m-1}}\frac{1}{x-\beta} \equiv -\frac{\beta}{2m}\frac{1}{x-\beta}.$$

As a consequence,

$$I_{2m} = -\frac{i\pi}{m}\left(\alpha + \alpha^3 + \cdots + \alpha^{2m-1}\right) = \frac{\pi}{m \sin\frac{\pi}{2m}}.$$

In particular, $I_2 = \pi/\sqrt{2}$, $I_3 = 2\pi/3$, $I_4 = \frac{\pi}{4}\sqrt{2}\sqrt{2+\sqrt{2}}$ as well as $\frac{1}{\pi}I_5$, $\frac{1}{\pi}I_6$ are expressible by radicals, but $\frac{1}{\pi}I_7$, $\frac{1}{\pi}I_9$ are not. The special cases $\frac{1}{\pi}I_{17}$, $\frac{1}{\pi}I_{257}$ are expressible by radicals. $\triangleleft$

$\triangleright$ **9.** *Integrals of rational fractions.* Generally, all integrals of rational functions taken over the whole real line are computable by residues. In particular,

$$J_m = \int_{-\infty}^{+\infty} \frac{dx}{(1+x^2)^m}, \qquad K_m = \int_{-\infty}^{+\infty} \frac{dx}{(1^2+x^2)(2^2+x^2)\cdots(m^2+x^2)}$$

can be explicitly evaluated. $\triangleleft$

Many function-theoretic consequences derive from the residue theorem. For instance, if f is analytic in Ω, $z_0 \in \Omega$ and λ is a simple loop of Ω encircling z_0, one has

$$f(z_0) = \frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{\zeta - z_0}. \tag{8}$$

This follows directly since

$$\operatorname{Res}\left[f(\zeta)(\zeta - z_0)^{-1}; \zeta = z_0\right] = f(z_0).$$

Then, by differentiation with respect to z under the integral sign, one gets similarly

$$\frac{1}{k!} f^{(k)}(z_0) = \frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{(\zeta - z_0)^k}. \tag{9}$$

The values of a function and its derivatives at a point can thus be obtained as values of integrals of the function away from that point.

A very important application of the residue theorem concerns coefficients of analytic functions.

THEOREM IV.2 (Cauchy's Coefficient Formula). *Let $f(z)$ be analytic in a region containing 0 and let λ be a simple loop around 0 that is oriented positively. Then the coefficient $[z^n]f(z)$ admits the integral representation*

$$f_n \equiv [z^n]f(z) = \frac{1}{2i\pi} \int_\lambda f(z) \, \frac{dz}{z^{n+1}} \, .$$

PROOF. This formula follows directly from the equalities

$$\frac{1}{2i\pi} \int_\lambda f(z) \, \frac{dz}{z^{n+1}} = \operatorname{Res}\left[f(z) z^{-n-1}; z = 0\right] = [z^n]f(z),$$

of which the first follows from the residue theorem, and the second from the identification of the residue at 0 as a coefficient. $\square$

$\triangleright$ **10.** *Complex-differentiability implies analyticity.* Formulæ (8) and (9) are by Note 5 consequences of complex-differentiability (without logically relying on the First Fundamental Principle). It is then a simple matter to complete the proof of the First Fundamental Property: one has (for h small enough)

$$\begin{aligned} f(z_0 + h) &= \frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{\zeta - (z_0 + h)} \\ &= \sum_{k \geq 0} \left[\frac{1}{2i\pi} \int_\lambda f(\zeta) \frac{d\zeta}{(\zeta - z_0)^{k+1}}\right] h^k \; = \; \sum_{k\geq 0} f^{(k)}(z_0) \frac{h^k}{k!}, \end{aligned}$$

as results from expanding $(\zeta - z_0 - h)^{-1}$ into powers of h. $\triangleleft$

Analytically, the coefficient formula allows one to deduce information about the coefficients from the values of the function itself, using adequately chosen contours of integration. It thus opens the possibility of estimating the coefficients $[z^n]f(z)$ in the expansion of $f(z)$ near 0 by using information on $f(z)$ *away* from 0. The rest of this chapter will precisely illustrate this process in the case of functions whose singularities are poles, that is, rational and meromorphic functions. Note also that the residue theorem provides the simplest known proof of the Lagrange inversion theorem (see the appendices) whose rôle is *inter alia* central to tree enumerations. The supplements below explore some independent consequences of the residue theorem and the coefficient formula.

$\triangleright$ **11.** *Liouville's Theorem.* If a function $f(z)$ is analytic in the whole of $\mathbb{C}$ and is of modulus bounded by an absolute constant, $|f(z)| \leq B$, then it must be a constant. (By trivial bounds, upon integrating on a large circle, it is found that the Taylor coefficients at the origin of index ≥ 1 are all equal to 0.) Similarly, if $f(z)$ is of at most polynomial growth, $|f(z)| \leq B\,(|z|+1)^r$ over the whole of $\mathbb{C}$, then it must be a polynomial. $\triangleleft$

$\triangleright$ **12.** *Lindelöf integrals.* Let $a(s)$ be analytic in $\Re(s) > \frac{1}{4}$ where it is assumed to satisfy $a(s) = O(\exp((\pi - \epsilon)|s|))$ for some $\epsilon > 0$. Then, one has for $\Re(z) > 0$,

$$\sum_{k=1}^{\infty} a(k)(-z)^k = -\frac{1}{2i\pi}\int_{1/2-i\infty}^{1/2+i\infty} a(s) z^s \, \frac{\pi}{\sin \pi s}\, ds.$$

(Close the integration contour by a large semi-circle on the right.) Such integrals, sometimes called Lindelöf integrals, provide representations for functions determined by an explicit "law" of their Taylor coefficients [**80**].

As a consequence, the generalized *polylogarithm* functions

$$\mathrm{Li}_{\alpha,k}(z) = \sum_{n\geq 1} n^{-\alpha}(\log n)^k z^n$$

are analytic in the complex plane $\mathbb{C}$ slit along $(1+\infty)$. (More properties can be found in [**39, 54**].) For instance, one finds in this way

$$\text{“}\sum_{n=1}^{\infty}(-1)^n \log n\text{”} = -\frac{1}{8\pi}\int_{-\infty}^{+\infty} \frac{\log(\frac{1}{4}+t^2)}{\cosh(\pi t)}\, dt = 0.22579\cdots = \log\sqrt{\frac{\pi}{2}},$$

when the divergent series on the left is interpreted as $\mathrm{Li}_{0,1}(-1) = \lim_{z\to -1^+} \mathrm{Li}_{0,1}(z)$. $\triangleleft$

$\triangleright$ **13.** *Magic duality.* Let ϕ be a function initially defined over the nonnegative integers but admitting a meromorphic extension over the whole of $\mathbb{C}$. Under conditions analogous to those of Note 12, the function

$$F(z) := \sum_{n\geq 1} \phi(n)(-z)^n,$$

which is analytic at the origin, is such that, near positive infinity,

$$F(z) \underset{z\to+\infty}{\sim} E(z) - \sum_{n\geq 1}\phi(-n)(-z)^{-n},$$

for some "elementary" function $E(z)$. (Starting from the representation of Note 12, close the contour of integration by a large semicircle to the left.) In such cases, the function is said to satisfy the principle of *magic duality*—its expansion at 0 and ∞ are given by one and the same "law". Functions

$$\frac{1}{1+z}, \quad \log(1+z), \quad \exp(-x), \quad \mathrm{Li}_2(-z), \quad \mathrm{Li}_3(-z)$$

satisfy magic duality. Ramanujan [**11**] made a great use of this principle, which applies to a wide class of functions including hypergeometric ones; see [**65**, Ch XI] for an insightful discussion. $\triangleleft$

$\triangleright$ **14.** *Euler–Maclaurin and Abel–Plana summations.* Under simple conditions on the analytic function f, one has Plana's (also known as Abel's) complex variables version of the Euler–Maclaurin summation formula:

$$\sum_{n=0}^{\infty} f(n) = \frac{1}{2}f(0) + \int_0^{\infty} f(x)\,dx + \int_0^{\infty}\frac{f(iy)-f(-iy)}{e^{2i\pi y}-1}\,dy.$$

(See [**66**, Vol. 1, p. 274] for a proof and validity conditions.) $\triangleleft$

▷ **15.** *Nörlund-Rice integrals.* Let $a(z)$ be analytic for $\Re(z) > k_0 - \frac{1}{2}$ and of at most polynomial growth in this right half plane. Then, with γ a loop around the interval $[k_0, n]$, one has

$$\sum_{k=k_0}^{n} \binom{n}{k} (-1)^{n-k} a(k) = \frac{1}{2i\pi} \int_\gamma a(s) \frac{n!\, ds}{s(s-1)(s-2)\cdots(s-n)}.$$

If $a(z)$ is meromorphic in a larger region, then the integral can be estimated by residues. For instance, with

$$S_n = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k}, \qquad T_n = \sum_{k=1}^{n} \binom{n}{k} \frac{(-1)^k}{k^2+1},$$

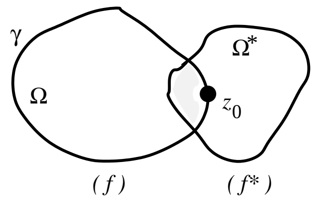
it is found that $S_n = -H_n$ (a harmonic number), while T_n oscillates boundedly as $n \to +\infty$. (This technique is a classical one in the calculus of finite differences, going back to Nörlund [**87**]. In computer science it is known as the method of "Rice's integrals" [**50**] and is used in the analysis of many algorithms and data structures including digital trees and radix sort [**75, 108**].) ◁

IV. 3. Singularities and exponential growth of coefficients

For a given function, a singularity can be informally defined as a point where the function "ceases" to be analytic. Singularities are, as we have stressed repeatedly, essential to coefficient asymptotics. This section presents the bases of a discussion within the framework of analytic function theory.

IV. 3.1. Singularities. Let $f(z)$ be an analytic function defined over the interior region determined by a simple closed curve γ, and let z_0 be a point of the bounding curve γ. If there exists an analytic function $f^*(z)$ defined over some open set Ω^* containing z_0 and such that $f^*(z) = f(z)$ in $\Omega^* \cap \Omega$, one says that f is *analytically continuable* at z_0 and that $f^\star$ is an *immediate analytic continuation* of f.

Analytic continuation:



$f^*(z) = f(z)$ on $\Omega^* \cap \Omega$.

In sharp contrast to real analysis where a function admits of many smooth extensions, analytic continuation is essentially *unique*: for instance, if f^* and f^{**} continue f at z_0, then one must have $f^*(z) = f^{**}(z)$ in the vicinity of z_0. Thus, the notion of immediate analytic continuation is intrinsic. Also the process can be iterated and we say that g is an *analytic continuation*[2] of f, even if their domains of definition do not overlap, provided a finite chain of intermediate function elements connects f and g. This notion is once more intrinsic—this is known as the principle of *unicity of analytic continuation* (along paths). An analytic function is then much like a hologram: as soon as it is specified in any tiny region, it is rigidly determined in any wider region where it can be continued.

DEFINITION IV.4. *Given an f defined in the region interior to γ, a point z_0 on the boundary of the region is a* singular point *or a* singularity[3] *of f if f is* not *analytically continuable at z_0.*

[2]The collection of all function elements continuing a given function gives rise to the notion of Riemann surface, for which manty good books exist, e.g., [**33, 105**]. We shall normally avoid appealing to this theory.

[3]For a detailed discussion, see [**28**, p. 229], [**72**, vol. 1, p. 82], or [**109**].

Granted the intrinsic character of analytic continuation, we can usually dispense with a detailed description of the original domain Ω and the curve γ. In simple terms, a function is singular at z_0 if it cannot be continued as an analytic function beyond z_0. A point at which a function is analytic is also called by contrast a *regular point*.

The two functions $f(z) = 1/(1-z)$ and $g(z) = \sqrt{1-z}$ may be taken as initially defined over the open unit disk by their power series representation. Then, as we already know, they can be analytically continued to larger regions, the punctured plane $\Omega = \mathbb{C}\setminus\{1\}$ for f and the complex plane slit along $(1, +\infty)$ for g. (This is achieved by the usual operations of analysis, upon taking inverses and square roots.) But both are singular at 1: for f, this results from the fact that (say) $f(z) \to \infty$ as $z \to 1$; for g this is due to the branching character of the square-root.

It is easy to check from the definitions that a converging Taylor series is analytic inside its disc of convergence. In other words, it can have no singularity inside this disc. However, it *must* have one on the boundary of the disc, as asserted by the theorem below. In addition, a classical theorem, called Pringsheim's theorem [**109**, Sec. 7.21], provides a refinement of this property in the case of functions with nonnegative coefficients.

THEOREM IV.3 (Boundary singularities). (i) *A function analytic f at the origin whose Taylor expansion at* 0 *has a finite radius of convergence R necessarily has a singularity on the boundary of its disc of convergence, $|z| = R$.*

(ii) [Pringsheim's Theorem] *If in addition f has nonnegative Taylor coefficients, then the point $z = R$ is a singularity of f.*

A figurative way of expressing Theorem IV.4, (i) is as follows:

The radius of convergence of a series equals its "radius of singularity".

(There "radius of singularity" means the first radius at which a singularity appears.) This result together with Pringsheim's is central to asymptotic enumeration as the remainder of this section will demonstrate.

PROOF. (i) Let $f(z)$ be the function and R the radius of convergence of its Taylor series at 0, taken under the form

$$f(z) = \sum_{n\geq 0} f_n z^n. \tag{10}$$

We now that there can be no singularity of f within the disc $|z| < R$. Suppose a contrario that $f(z)$ is analytic in the whole of $|z| < \rho$ for some ρ satisfying $\rho > R$. By Cauchy's coefficient formula (theorem IV.2), upon integrating along the circle λ of radius $r = (R + \rho)/2$, it is seen that the coefficient $[z^n]f(z)$ is $O(r^{-n})$. But then, the series expansion of f would have to converge in the disc of radius $r > R$, a contradiction. (More on this theme below.)

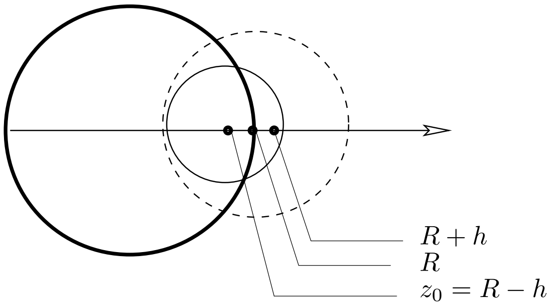
(ii) Suppose a contrario that $f(z)$ is analytic at R, implying that it is analytic in a disc of radius r centred at R. We choose a number h such that $0 < h < \frac{1}{3}r$ and consider the expansion of $f(z)$ around $z_0 = R - h$:

$$f(z) = \sum_{m\geq 0} g_m (z - z_0)^m. \tag{11}$$

By Taylor's formula and the representability of $f(z)$ together with its derivatives at z_0 by means of (10), we have

$$g_m = \sum_{n\geq 0} \binom{n}{m} f_n z_0^m,$$

and in particular, $g_m \geq 0$. By the way h was chosen, the series (11) converges at $z - z_0 = 2h$, as illustrated by the following diagram:



Consequently, one has

$$f(R+h) = \sum_{m \geq 0} \left(\sum_{n \geq 0} \binom{n}{m} f_n z_0^{m-n} \right) (2h)^m.$$

This is a converging double sum of positive terms, so that the sum can be reorganized in any way we like. In particular, one has convergence of all the series involved in

$$\begin{aligned} f(R+h) &= \sum_{m,n \geq 0} \binom{n}{m} f_n (R-h)^{m-n} (2h)^m \\ &= \sum_{n \geq 0} f_n \left[(R-h) + (2h) \right]^n \\ &= \sum_{n \geq 0} f_n (R+h)^n. \end{aligned}$$

This establishes the fact that $f_n = o((R+h)^n)$, thereby reaching a contradiction. Pringsheim's theorem is proved. ☐

Singularities of a function analytic at 0 which are on the boundary of the disc of convergence are called *dominant singularities*. The second part of this theorem appreciably simplifies the search for dominant singularities of combinatorial generating functions since these have nonnegative coefficients.

For instance, the derangement OGF and the surjection EGF,

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

are analytic except for a simple pole at $z = 1$ in the case of $D(z)$, and except for points $z_k = \log 2 + 2ik\pi$ that are simple poles in the case of $S(z)$. Thus the dominant singularities for derangements and surjections are at 1 and $\log 2$ respectively.

It is known that $\sqrt{Z}$ cannot be unambiguously defined as an analytic function in a neighbourhood of $Z = 0$. As a consequence, the function

$$C(z) = (1 - \sqrt{1-4z})/2,$$

which is the generating function of the Catalan numbers, is an analytic function in certain regions that should exclude $1/4$; for instance, one may opt to take the complex plane slit along the ray $(1/4, +\infty)$. Similarly, the function

$$L(z) = \log \frac{1}{1-z}$$

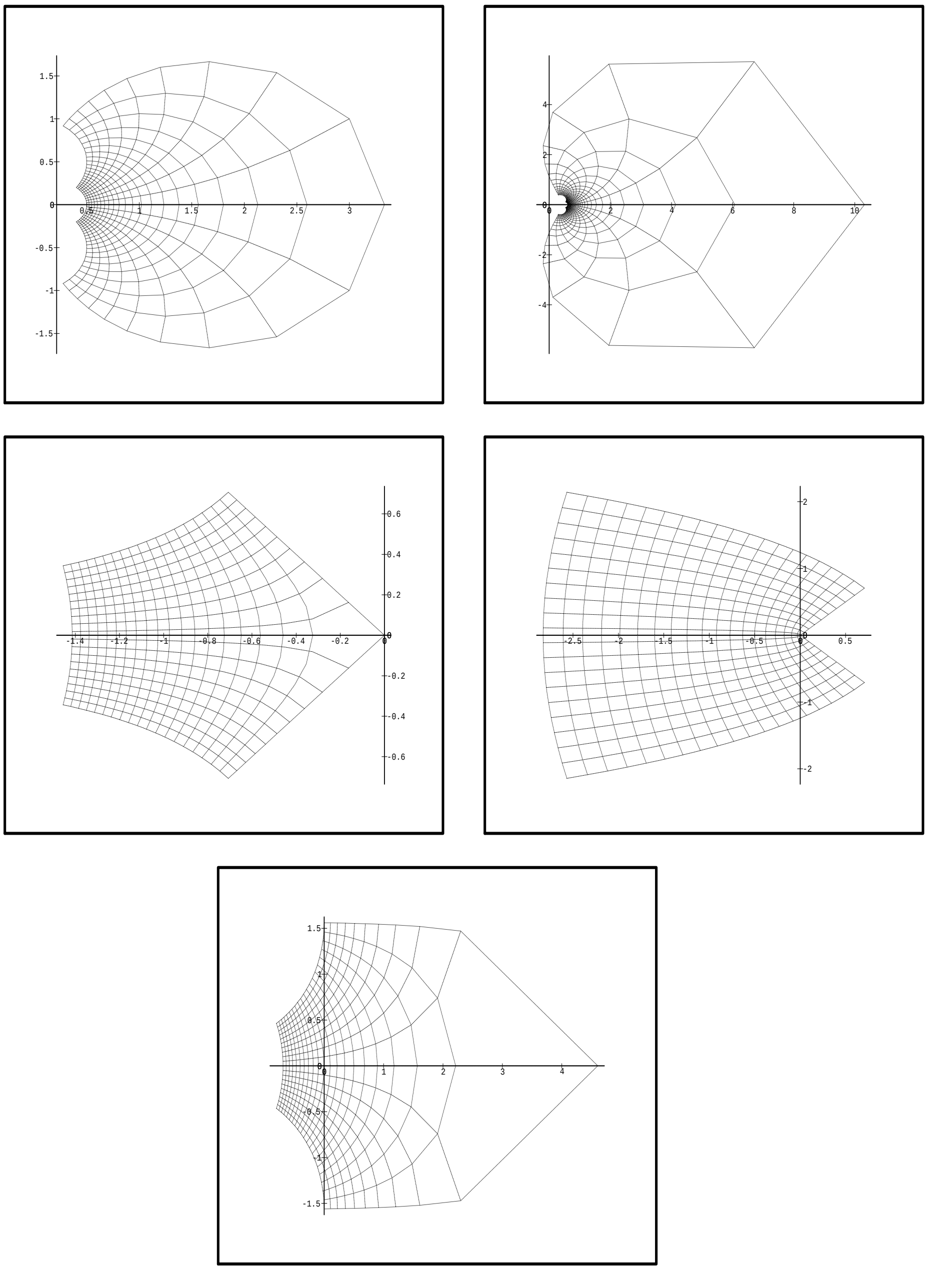


FIGURE 4. The images of a grid on the unit square (with corners $\pm 1 \pm i$) by various functions singular at $z = 1$ reflect the nature of the singularities involved. Here (from top to bottom) $f_0(z) = 1/(1-z)$, $f_1(z) = \exp(z/(1-z))$, $f_2(z) = -(1-z)^{1/2}$, $f_3(z) = -(1-z)^{3/2}$, $f_4(z) = \log(1/(1-z))$. The functions have been normalized to be increasing over the real interval $[-1, 1]$. Singularities are apparent near the right of each diagram where small grid squares get folded or unfolded in various ways. (In the case of functions f_0, f_1, f_4 that become infinite at $z = 1$, the grid has been slightly truncated to the right.)

which is the EGF of cyclic permutations is analytic in the complex plane slit along $(1, +\infty)$. (An alternative way of seeing that $C(z)$ and $L(z)$ are singular at $\frac{1}{4}$ and 1 is to observe that their derivatives become infinite along rays $z \to \frac{1}{4}^-$ and $z \to 1^-$.)

A function having no singularity at a finite distance is called *entire*; its Taylor series then converges everywhere in the complex plane. The EGFs,

$$e^{z+z^2/2},\ e^{e^z-1},$$

associated to involutions and set partitions are entire.

IV. 3.2. The Exponential Growth Formula. We say that a number sequence $\{a_n\}$ is of *exponential order* K^n which we abbreviate as (the symbol $\bowtie$ is a "bowtie")

$$a_n \bowtie K^n \qquad \text{iff} \qquad \limsup |a_n|^{1/n} = K.$$

The relation $X \bowtie Y$ reads as "*X is of exponential order Y*". In other words, for any $\epsilon > 0$:

- $|a_n| >_{i.o} (K-\epsilon)^n$, that is to say, $|a_n|$ exceeds $(K-\epsilon)^n$ infinitely often (for infinitely many values of n);
- $|a_n| <_{a.e.} (K+\epsilon)^n$, that is to say, $|a_n|$ is dominated by $(K+\epsilon)^n$ almost everywhere (except for possibly finitely many values of n).

This relation can be rephrased as $a_n = \vartheta(n)K^n$, where ϑ is a *subexponential factor* satisfying

$$\limsup |\theta(n)|^{1/n} = 1;$$

such a factor is thus bounded from above almost everywhere by any increasing exponential (of the form $(1+\epsilon)^n$) and bounded from below infinitely often by any decaying exponential (of the form $(1-\epsilon)^n$). Typical subexponential factors are

$$1,\ n^3,\ (\log n)^2,\ \sqrt{n},\ \frac{1}{\sqrt[3]{\log n}},\ n^{-3/2},\ \log\log n.$$

(Note that functions like $e^{\sqrt{n}}$ and $\exp(\log^2 n)$ must be treated as subexponential factors for the purpose of this discussion.) In this and the next chapters, we shall see general methods that enable one to extract such subexponential factors from generating functions.

THEOREM IV.4 (Exponential Growth Formula). *If $f(z)$ is analytic at 0 and R is the modulus of a singularity of $f(z)$ nearest to the origin,*

$$R = \min\{\, |z|\, , z \in \operatorname{Sing}(f)\,\},$$

then the coefficient $f_n = [z^n]f(z)$ satisfies

$$f_n \bowtie \left(\frac{1}{R}\right)^n, \qquad \textit{equivalently} \quad f_n = \left(\frac{1}{R}\right)^n \theta(n) \textit{ with } \limsup |\theta(n)|^{1/n} = 1.$$

PROOF. The lower bound follows since otherwise the series would converge (and hence be analytic) in a larger domain. Trivial bounds on Cauchy's coefficient formula upon taking as contour λ a circle of radius $R-\eta$,

$$\begin{aligned} |f_n| &\le \frac{1}{2\pi}\frac{\max\{|f(z)|\,/\,|z| = R-\eta\}}{|R-\eta|^{n+1}} \cdot (2\pi R) \\ &\le \mathcal{O}((R-\eta)^{-n}), \end{aligned}$$

yield the upper bound. □

The exponential growth formula thus directly relates the exponential order of growth of coefficients of a function to the *location* of its singularities nearest to the origin. Several direct applications to combinatorial enumeration are given below.

EXAMPLE 1. *Exponential growth and combinatorial enumeration*. Here are a few immediate applications of of exponential bounds to surjections, derangements, integer partitions, and unary binary trees.

Surjections. The function

$$R(z) = (2 - e^z)^{-1}$$

is the EGF of surjections. The denominator is an entire function, so that singularities may only arise from its zeros, to be found at the points

$$\chi_k = \log 2 + 2ik\pi, \qquad k \in \mathbb{Z}.$$

The dominant singularity of R is then at $\rho = \chi_0 = \log 2$. Thus, with $r_n = [z^n]R(z)$,

$$r_n \bowtie (\frac{1}{\log 2})^n.$$

Similarly, if "double" surjections are considered (each value in the range of the surjection is taken at least twice), the corresponding EGF is

$$R^*(z) = \frac{1}{2 - z - e^z};$$

the dominant singularity is at ρ^* defined as the positive root of equation $e^{\rho^*} - \rho^* = 2$, and the coefficient r_n^* satisfies: $r_n^* \bowtie (\frac{1}{\rho^*})^n$ Numerically, this gives

$$r_n \bowtie 1.44269^n \qquad \text{and} \qquad r_n^* \bowtie 0.87245^n,$$

with the actual figures for the corresponding logarithms being

n	$\frac{1}{n}\log r_n$	$\frac{1}{n}\log r_n^*$
10	0.33385	0.80208
20	0.35018	0.80830
50	0.35998	0.81202
100	0.36325	0.81327
∞	0.36651	0.81451
	$(\log 1/\rho)$	$(\log(1/\rho^*)$

These estimates constitutes a weak form of a more precise result to be established later in this chapter: If random surjections of size n are taken equally likely, the probability of a surjection being a double surjection is exponentially small.

Derangements. There, for $d_{1,n} = [z^n]e^{-z}(1-z)^{-1}$ and $d_{2,n} = [z^n]e^{-z-z^2/2}(1-z)^{-1}$ we have, from the poles at $z = 1$,

$$d_{1,n} \bowtie 1^n \quad \text{and} \quad d_{2,n} \bowtie 1^n.$$

The upper bound is combinatorially trivial. The lower bound expresses that the probability for a random permutation to be a derangement is *not* exponentially small. For $d_{1,n}$, we have already proved by an elementary argument the stronger result $d_{1,n} \to e^{-1}$; in the case of $d_{2,n}$, we shall establish later the precise asymptotic equivalent $d_{2,n} \to e^{-3/2}$, in accordance with what was announced in the introduction.

Unary-Binary trees. The expression

$$U(z) = \frac{1 - z - \sqrt{1 - 2z - 3z^2}}{2z} = z + z^2 + 2\,z^3 + 4\,z^4 + 9\,z^5 + \cdots,$$

represents the OGF of (plane unlabelled) unary-binary trees. From the equivalent form,

$$U(z) = \frac{1 - z - \sqrt{(1-3z)(1+z)}}{2z},$$

it follows that $U(z)$ is analytic in the complex plane slit along $(\frac{1}{3}, +\infty)$ and $(-\infty, -1)$ and is singular at $z = -1$ and $z = 1/3$ where it has branch points. The closest singularity to the origin being at $\frac{1}{3}$, one has

$$U_n \bowtie 3^n.$$

In this case, the stronger upper bound $U_n \leq 3^n$ results directly from the possibility of encoding such trees by words over a ternary alphabet using Łukasiewicz codes (Chapter I). A complete asymptotic expansion will be obtained in the next chapter. □

The exponential growth formula expressed by Theorem IV.4 can be supplemented by effective upper bounds which are very easy to derive and often turn out to be surprisingly accurate. We state:

PROPOSITION IV.1 (Saddle-Point bounds). *Let $f(z)$ be analytic in the disc $|z| < R$ with $0 < R \leq \infty$. Then, one has, for any r in $(0, R)$, the family of* saddle point upper bounds

(12)

$$[z^n]f(z) \quad \leq \quad \frac{\sup_{|z|=r} |f(z)|}{r^n} \quad \textit{(any } r\textit{)}, \quad \textit{and} \quad [z^n]f(z) \quad \leq \quad \inf_{s \in (0,R)} \frac{\sup_{|z|=s} |f(z)|}{s^n}.$$

If in addition $f(z)$ has nonnegative coefficients at 0, then

(13)
$$[z^n]f(z) \quad \leq \quad \frac{f(r)}{r^n} \quad \textit{(any } r\textit{)}, \quad \textit{and} \quad [z^n]f(z) \quad \leq \quad \inf_{s \in (0,R)} \frac{f(s)}{s^n}.$$

PROOF. The first bound in (12) results from trivial bounds applied to the Cauchy coefficient formula, when integration is performed along a circle. It is consequently valid for any r smaller than the radius of convergence of f at 0. The best possible such bound is then given by the second inequality; it can be determined by cancelling a derivative,

$$s \; : \; s\frac{f'(s)}{f(s)} = n.$$

Note that because of the first inequality, any approximate solution of this last equation will in fact provide a valid upper bound.

The bounds (13) can be viewed as a specialization of (12). Alternatively, they can be obtained elementarily since

$$f_n \leq \frac{f_0}{r^n} + \cdots + \frac{f_{n-1}}{r^{n-1}} + f_n + \frac{f_{n+1}}{r^{n+1}} + \cdots,$$

whenever the f_k are nonnegative. □

For reasons well explained by the saddle point method (Chapter VI), these bounds usually capture the actual asymptotic behaviour up to a polynomial factor only. A typical instance is the weak form of Stirling's formula,

$$\frac{1}{n!} \equiv [z^n]e^z \leq \frac{e^n}{n^n},$$

which only overestimates the true asymptotic value by a factor of $\sqrt{2\pi n}$.

EXAMPLE 2. *Combinatorial examples of saddle point bounds.* Here are applications to fragmented permutations, set partitions (Bell numbers), involutions, and integer partitions.

n	$\widetilde{I}_n$	I_n
100	$0.106579 \cdot 10^{85}$	$0.240533 \cdot 10^{83}$
200	$0.231809 \cdot 10^{195}$	$0.367247 \cdot 10^{193}$
300	$0.383502 \cdot 10^{316}$	$0.494575 \cdot 10^{314}$
400	$0.869362 \cdot 10^{444}$	$0.968454 \cdot 10^{442}$
500	$0.425391 \cdot 10^{578}$	$0.423108 \cdot 10^{576}$

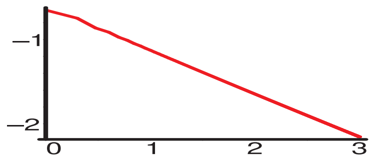


FIGURE 5. The comparison between the exact number of involutions I_n and its approximation $\widetilde{I}_n = n!e^{\sqrt{n}+n/2}n^{-n/2}$: (left) a table; (right) a plot of $\log_{10}(I_n/\widetilde{I}_n)$ against $\log_{10} n$ suggesting that the ratio is $\sim K \cdot n^{-1/2}$.

Fragmented permutations. Consider first the EGF of "fragmented permutations" (Chapter II) defined by $\mathcal{F} = \mathfrak{P}(\mathfrak{S}_{\geq 1}(\ddagger))$ in the labelled universe. We claim that

$$\frac{1}{n!}F_n \equiv [z^n]e^{z/(1-z)} \leq e^{2\sqrt{n}-\frac{1}{2}+O(n^{-1/2})}. \tag{14}$$

Indeed, the minimizing value of r in (13) is r_0 such that

$$0 = \frac{d}{dr}\left(\frac{r}{1-r} - n\log r\right)_{r=r_0} = \frac{1}{(1-r_0)^2} - \frac{n}{r_0}.$$

The equation is solved by $r_0 = (2n+1-\sqrt{4n+1})/(2n)$. One can either use this exact value and perform asymptotic approximation of $f(r_0)/z_0^n$, or adopt the approximate value $r_1 = 1 - 1/\sqrt{n}$, which leads to simpler calculations. The estimate (14) results.

Bell numbers and set partitions. Another immediate applications is an upper bound on Bell numbers enumerating set partitions with EGF e^{e^z-1}. The best saddle point bound is

$$\frac{1}{n!}B_n \leq e^{e^r-1-n\log r}, \qquad r\ :\ re^r = n, \tag{15}$$

with $r \sim \log n - \log\log n$.

Involutions. Regarding involutions, their EGF is $I(z) = \exp(z + \frac{1}{2}z^2)$, and one determines (see Figure 5 for numerical data)

$$\frac{1}{n!}I_n \leq \frac{e^{\sqrt{n}+n/2}}{n^{n/2}}. \tag{16}$$

Similar bounds hold for permutations with all cycle lengths $\leq k$ and permutations σ such that $\sigma^k = Id$.

Integer partitions. The function

$$P(z) = \prod_{k=1}^{\infty}\frac{1}{1-z^k} = \exp\left(\sum_{\ell=1}^{\infty}\frac{1}{\ell}\frac{z^\ell}{1-z^\ell}\right) \tag{17}$$

is the OGF of integer partitions, the unlabelled analogue of set partitions. Its radius of convergence is *a priori* bounded from above by 1, since the set $\mathcal{P}$ is infinite and the second form of $P(z)$ shows that it is exactly equal to 1. Therefore $P_n \bowtie 1^n$. A finer upper bound results from the estimate

$$\Lambda(t) := \log P(e^{-t}) \sim \frac{\pi^2}{6t} + \log\sqrt{\frac{t}{2\pi}} - \frac{1}{24}t + O(t^2), \tag{18}$$

which obtains from Euler–Maclaurin summation or, better, from a Mellin analysis following APPENDIX: *Mellin transform*, p. 120. Indeed, the Mellin transform of Λ is, by the harmonic sum rule,

$$\Lambda^\star(s) = \zeta(s)\zeta(s+1)\Gamma(s), \qquad s \in \langle 1, +\infty \rangle,$$

and the successive leftmost poles at $s = 1$ (simple pole), $s = 0$ (double pole), and $s = -1$ (simple pole) translate into the asymptotic expansion (18). When $z \to 1^-$, this means that

$$P(z) \sim \frac{e^{-\pi^2/12}}{\sqrt{2\pi}} \sqrt{1-z} \exp\left(\frac{\pi^2}{6(1-z)}\right), \tag{19}$$

from which we derive the upper bound,

$$P_n \le C n^{1/4} e^{\pi\sqrt{2n/3}}$$

(for some $C > 0$) in a way analogous to fragmented permutations above. This last bound loses only a polynomial factor, as we shall prove when studying the saddle point method in Chapter VIII. □

▷ **16.** *A natural boundary.* One has $P(re^{i\theta}) \to \infty$ as $r \to 1^-$, for any angle θ that is a rational multiple of 2π. Such points being dense on the unit circle, the function $P(z)$ admits the unit circle as a *natural boundary*, i.e., it cannot be analytically continued beyond this circle. ◁

▷ **17.** *Meinardus' method.* The combination of Mellin transforms and saddle point analysis in the theory of partitions is known as Meinardus' method [**4**, Ch. 6]. Consider the set $\mathcal{R}$ of compositions into rth powers ($r \ge 2$). The OGF satisfies

$$\Lambda(t) := \log R(e^{-t}) = \sum_{\ell \ge 1} \frac{1}{\ell} \frac{e^{-\ell^r t}}{1 - e^{-\ell^r t}},$$

with Mellin transform $\Lambda^\star(s) = \zeta(rs)\zeta(s+1)\Gamma(s)$ defined for $\Re(s) > r^{-1}$. From the pole of $\Lambda^\star$ at $s = 1/r$, one gets

$$R(z) = \exp\left(\frac{\xi}{(1-z)^{1/r}}\right)(1 + o(1)), \qquad \xi := \frac{1}{r}\zeta(1+\frac{1}{r})\Gamma(\frac{1}{r}).$$

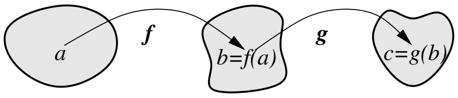
The minimizing value s_0 for saddle point bounds satisfies $1 - s_0\ (rn/\xi)^{-r/(r+1)}$, and

$$\log R_n \le C n^{\frac{1}{r+1}}(1 + o(1))$$

(for some $C > 0$). See Andrews' book [**4**, Ch. 6] for precise asymptotics and a general setting. ◁

IV. 3.3. Closure properties and computable bounds. The functions analytic at a point $z = a$ are closed under sum and product, and hence form a ring. If $f(z)$ and $g(z)$ are analytic at $z = a$, then so is their quotient $f(z)/g(z)$ provided $g(a) \ne 0$. Meromorphic functions are furthermore closed under quotient and hence form a field. Such properties are proved most easily using complex-differentiability and extending the usual relations from real analysis, $(f+g)' = f' + g'$, $(fg)' = fg' = f'g$, and so on.

Analytic functions are also closed under composition: if $f(z)$ is analytic at $z = a$ and $g(w)$ is analytic at $b = f(a)$, then $g \circ f(z)$ is analytic at $z = a$. Graphically:



The proof based on complex-differentiability closely mimicks the real case. Inverse functions exist conditionally: if $f'(a) \neq 0$, then $f(z)$ is locally linear near a, hence invertible, so that there exists a g satisfying $f \circ g = g \circ f = Id$, where Id is the identity function, $Id(z) \equiv z$. The inverse function is itself locally linear, hence complex differentiable, hence analytic. In short, the inverse of an analytic function f at a place where its derivative does not vanish is an analytic function.

One way to establish closure properties, as suggested above, is to deduce analyticity criteria from complex differentiability by way of the "First Fundamental Property". An alternative approach, closer to the original notion of analyticity, can be based on a two-step process: (i) closure properties are shown to hold true for formal power series; (ii) the resulting formal power series are proved to be locally convergent by means of suitable majorizations on their coefficients. This is the basis of the classical method of *majorant series* originating with Cauchy.

▷ **18.** *The majorant series technique.* Given two power series, define $f(z) \preceq g(z)$ if $|[z^n]f(z)| \leq [z^n]g(z)$ for all $n \geq 0$. The following two conditions are equivalent: (i) $f(z)$ is analytic in the disc $|z| < \rho$; (ii) for any $r < \rho$ there exists a c such that

$$f(z) \preceq \frac{c}{1-rz}.$$

If f, g are majorized by $c/(1-rz), d/(1-rz)$ respectively, then $f+g$ and $f \cdot g$ are majorized,

$$f(z)+g(z) \preceq \frac{c+d}{1-rz}, \qquad f(z)\cdot g(z) \preceq \frac{e}{1-sz},$$

for any $s < r$ and some e dependent on s. If f, g are majorized by $c/(1-rz), dz/(1-rz)$ respectively, then $f \circ g$ is majorized:

$$f \circ g(z) \preceq \frac{cz}{1-r(1+d)z}.$$

Constructions for $1/f$ and for the functional inverse of f can be similarly developed. See Cartan's book [**17**] and van der Hoeven's study [**110**] for a systematic treatment. ◁

For functions defined by analytic expressions, singularities can be determined inductively in an intuitively transparent manner. If $\mathrm{Sing}(f)$ and $\mathrm{Zero}(f)$ are the set of singularities and zeros of function f, then, due to closure properties of analytic functions, the following informally stated guidelines apply.

$$\left\{\begin{array}{lll}
\mathrm{Sing}(f \pm g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \\
\mathrm{Sing}(f \times g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \\
\mathrm{Sing}(f/g) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Sing}(g) \cup \mathrm{Zero}(g) \\
\mathrm{Sing}(f \circ g) & \subseteq & \mathrm{Sing}(g) \cup g^{(-1)}(\mathrm{Sing}(f)) \\
\mathrm{Sing}(\sqrt{f}) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Zero}(f) \\
\mathrm{Sing}(\log(f)) & \subseteq & \mathrm{Sing}(f) \cup \mathrm{Zero}(f) \\
\mathrm{Sing}(f^{(-1)}) & \subseteq & f(\mathrm{Sing}(f)) \cup f(\mathrm{Zero}(f')).
\end{array}\right.$$

A mathematically rigorous treatment would require considering multivalued functions and Riemann surfaces, so that we do not state detailed validity conditions and, at this stage, keep for these formulæ the status of useful heuristics. In fact, because of Pringsheim's theorem, the search of dominant singularities of combinatorial generating function can normally avoid considering the multivalued structure of functions, since only some initial segment of the positive real half–line needs to be considered. This in turn implies a powerful and easy way of determining the exponential order of coefficients of a wide variety of generating functions, as we now explain.

As defined in Chapters I and II, a combinatorial class is *constructible* if it can be specified by a finite set of equations involving only basic constructors. A specification is *iterative* if the dependency graph of the specification is acyclic, that is, no recursion is involved and a single functional term (written with sums, products, as well as sequence, set, and cycle constructions) describes the specification. We state:

THEOREM IV.5 (Computability of growth). *Let $\mathcal{C}$ be a* constructible *unlabelled class that admits of an* iterative *specification in terms of* $(1, \mathcal{Z}; \mathfrak{S}, \mathfrak{P}, \mathfrak{M}, \mathfrak{C}; +, \times)$. *Then the radius of convergence ρ_C of the OGF $C(z)$ of $\mathcal{C}$ is a nonzero computable real number.*

Let $\mathcal{D}$ be a constructible *labelled class that admits of an* iterative *specification in terms of* $(1, \mathcal{Z}; \mathfrak{S}, \mathfrak{P}, \mathfrak{C}; +, \star)$. *Then the radius of convergence ρ_D of the EGF $D(z)$ of $\mathcal{D}$ is a nonzero computable real number.*

Accordingly, the exponential rate of growth of the coefficients $[z^n]C(z)$ and $[z^n]D(z)$ are computable real numbers.

A real number α is computable iff there exists a program Π_α which on input m outputs a rational number α_m that is within $\pm 10^{-m}$ of α. The theorem immediately implies that the exponential growth estimates,

$$[z^n]C(z) \equiv C_n \bowtie \left(\frac{1}{\rho_C}\right)^n, \qquad [z^n]D(z) \equiv \frac{1}{n!}D_n \bowtie \left(\frac{1}{\rho_D}\right)^n,$$

for coefficients are automatically computable from the specification itself.

PROOF. In both cases, the proof proceeds by induction on the structural specification of the class. For each class $\mathcal{F}$, with generating function $F(z)$, we associate a *signature*, which is an ordered pair $\langle \rho_F, \tau_F \rangle$, where ρ_F is the radius of convergence of F and τ_F is the value of F at ρ_F, precisely,

$$\tau_F := \lim_{x \to \rho_F^-} F(x).$$

(The value τ_F is well defined as an element of $\mathbb{R} \cup \{+\infty\}$ since F, being a counting generating function, is necessarily increasing on $(0, \rho_F)$.) We prove the assertion of the theorem together with the additional property that $\tau_F = \infty$ and as soon as one of the unary constructors ($\mathfrak{S}, \mathfrak{M}, \mathfrak{P}, \mathfrak{C}$) intervenes in the specification, that is, as soon as the class is infinite. In that case, since the OGF includes infinitely many terms of the form z^n, it must be divergent at 1, so that $\rho_F \leq 1$ holds *a priori* for all infinite classes under consideration.

Consider the unlabelled case first. The signatures of the neutral class 1 and the atomic class $\mathcal{Z}$, with OGF 1 and z, are $\langle +\infty, 1 \rangle$ and $\langle +\infty, +\infty \rangle$. Any nonconstant polynomial which is the OGF of a finite set has the signature $\langle +\infty, +\infty \rangle$. The assertion is thus easily verified in these cases.

Next, let $\mathcal{F} = \mathfrak{S}(\mathcal{G})$. The OGF $G(z)$ must be nonconstant and in fact satisfy $G(0) = 0$ in order for the sequence construction to be properly defined. Thus, by the induction hypothesis, one has $0 < \rho_G \leq +\infty$ and $\tau_G = +\infty$. Now, the function G being increasing and continuous along the positive axis, there must exist a value β such that $0 < \beta < \rho_G$ with $G(\beta) = 1$. For $z \in (0, \beta)$, the quasi-inverse $F(z) = (1 - G(z))^{-1}$ is well defined and analytic; as z approaches β from the left, $F(z)$ increases unboundedly. Thus, the smallest singularity of F along the positive axis is at β, and by Pringsheim's theorem, one has $\rho_F = \beta$. The argument also shows that $\tau_F = +\infty$. There only remains to check that β is computable. The coefficients of G form a computable sequence of integers, so that $G(x)$, which can be well approximated via truncated Taylor series, is an effectively computable number[4] if x is itself a positive computable number less than ρ_G. Then dichotomic search constitutes effectively an algorithm for determining β.

Next, we consider the multiset construction, $\mathcal{F} = \mathfrak{M}(\mathcal{G})$, whose translation into OGFs necessitates the "Pólya exponential":

$$F(z) = \operatorname{Exp}(G(z)) \quad \text{where} \quad \operatorname{Exp}(h(z)) := \exp\left(h(z) + \frac{1}{2}h(z^2) + \frac{1}{3}h(z^3) + \cdots\right).$$

Once more, the induction hypothesis is assumed for G. If G is polynomial, then F is a variant of the OGF of integer partitions, and in fact is expressible as a finite product of terms of the form $P(z), P(z^2), P(z^3), \ldots$ Thus, $\rho_F = 1$ and $\tau_F = \infty$ in that particular case. In the general case of $\mathcal{F} = \mathfrak{M}(\mathcal{G})$ with $\mathcal{G}$ infinite, we start by fixing arbitrarily a number r such that $0 < r < \rho_G \leq 1$ and examine $F(z)$ for $z \in (0, r)$. The expression for F rewrites as

$$\operatorname{Exp}(G(z)) = e^{G(z)} \cdot \exp\left(\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots\right).$$

The first factor is analytic for z on $(0, \rho_G)$ since, the exponential function being entire, e^G has the singularities of G. As to the second factor, one has $G(0) = 0$ (in order for the set construction to be well-defined), while $G(x)$ is convex for $x \in [0, r]$ (since its second derivative is positive). Thus, there exists a positive constant K such that $G(x) \leq Kx$ when $x \in [0, r]$. Then, the series $\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots$ has its terms dominated by those of the convergent series

$$\frac{K}{2}r^2 + \frac{K}{3}r^3 + \cdots = K \log(1-r)^{-1} - Kr.$$

By a well known theorem of analytic function theory, a uniformly convergent sum of analytic functions is itself analytic; consequently, $\frac{1}{2}G(z^2) + \frac{1}{3}G(z^3) + \cdots$ is analytic at all z of $(0, r)$. Analyticity is then preserved by the exponential, so that $F(z)$, being analytic at $z \in (0, r)$ for any $r < \rho_G$ has a radius of convergence that satisfies $\rho_F \geq \rho_G$. On the other hand, since $F(z)$ dominates termwise $G(z)$, one has $\rho_F \leq \rho_G$. Thus finally one has $\rho_F = \rho_G$. Also, $\tau_G = +\infty$ implies $\tau_F = +\infty$.

A completely parallel discussion covers the case of the powerset construction ($\mathfrak{P}$) whose associated functional $\overline{\operatorname{Exp}}$ is a minor modification of the Pólya exponential Exp. The cycle construction can be treated by similar arguments based on consideration of "Pólya's logarithm" as $\mathcal{F} = \mathfrak{C}(\mathcal{G})$ corresponds to

$$F(z) = \operatorname{Log}\frac{1}{1 - G(z)}, \qquad \text{where} \quad \operatorname{Log} h(z) = \log h(z) + \frac{1}{2}\log h(z^2) + \cdots.$$

In order to conclude with the unlabelled case, there only remains to discuss the binary constructors $+$, $\times$, which give rise to $F = G + H$, $F = G \cdot H$. It is easily verified that $\rho_F = \min(\rho_G, \rho_H)$ and $\tau_F = \tau_G \circ \tau_H$ with $\circ$ being $+$ or $\times$. Computability is granted since the minimum of two computable numbers is computable.

The labelled case is covered by the same type of argument as above. The discussion is even simpler, since the ordinary exponential and logarithm replace the Pólya operators Exp and Log. It is still a fact that all the EGFs of infinite families are infinite at their dominant positive singularity, though the radii of convergence can now be of any magnitude (w.r.t. 1). $\square$

$\triangleright$ **19.** *Syntactically decidable properties.* In the unlabelled case, $\rho_F = 1$ iff the specification of $\mathcal{F}$ only involves $(1, \mathcal{Z}; \mathfrak{P}, \mathfrak{M}; +, \times)$ and at least one of $\mathfrak{P}, \mathfrak{M}$. $\triangleleft$

[4]The present argument only establishes non-constructively the *existence* of a program, based on the fact that truncated Taylor series converge geometrically fast at an interior point of their disc of convergence. Making explict this program and the involved parameters however represents a harder problem that is not touched upon here.

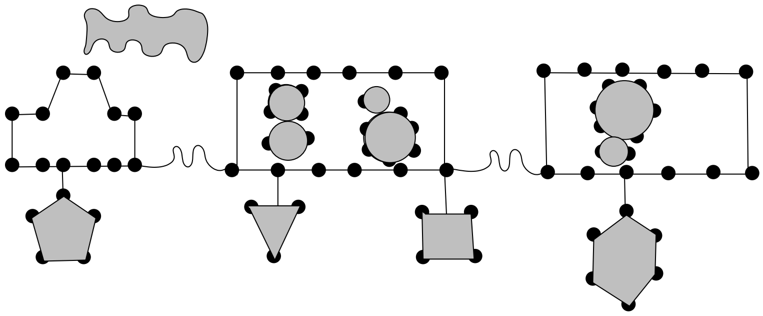


FIGURE 6. A random train.

$\triangleright$ **20.** *Nonconstructibility of permutations and graphs.* The class $\mathcal{P}$ of all permutations cannot be specified as a constructible unlabeled class since the OGF $P(z) = \sum_n n! z^n$ has radius of convergence 0. (It is of course constructible as a labelled class.) Graphs, whether labelled or unlabelled, are too numerous to form a constructible class. $\triangleleft$

Theorem IV.5 establishes a link between analytic combinatorics, computability theory, and symbolic manipulation systems. It is based on an article of Flajolet, Salvy, and Zimmermann [**49**] devoted to such computability issues in exact and asymptotic enumeration. Recursive specifications are not discussed now since they tend to give rise to branch points, themselves amenable to singularity analysis techniques to be developed in the next chapter.

EXAMPLE 3. *Combinatorial trains.* This somewhat artificial example from [**38**] serves to illustrate the scope of Theorem IV.5 and demonstrate its inner mechanisms at work. Define the class of all labelled trains by the following specification,

$$\left\{\begin{array}{lcl} \mathcal{T} & = & \mathcal{W}a \star \mathfrak{S}(\mathcal{W}a \star \mathfrak{P}(\mathcal{P}a)), \\ \mathcal{W}a & = & \mathfrak{S}_{\geq 1}(\mathcal{P}\ell), \\ \mathcal{P}\ell & = & \mathcal{Z} \star \mathcal{Z} \star (\mathbf{1} + \mathfrak{C}(\mathcal{Z})), \\ \mathcal{P}a & = & \mathfrak{C}(\mathcal{Z}) \star \mathfrak{C}(\mathcal{Z}). \end{array}\right. \tag{20}$$

In figurative terms, a train ($\mathcal{T}$) is composed of a first wagon ($\mathcal{W}a$) to which is appended a sequence of passenger wagons, each of the latter capable of containing a set of passengers ($\mathcal{P}a$). A wagon is itself composed of "planks" ($\mathcal{P}\ell$) determined by their end points ($\mathcal{Z} \star \mathcal{Z}$) and to which a circular wheel ($\mathfrak{C}(\mathcal{Z})$) may be attached. A passenger is composed of a head and a belly that are each circular arrangements of atoms (see Figure 6).

The translation into a set of EGF equations is immediate and a symbolic manipulation system readily provides the form of the EGF of trains, $T(z)$, as

$$T(z) = \frac{z^2\left(1+\log((1-z)^{-1})\right)}{\left(1-z^2\left(1+\log((1-z)^{-1})\right)\right)}\left(1-\frac{z^2\left(1+\log((1-z)^{-1})\right)e^{\left(\log((1-z)^{-1})\right)^2}}{1-z^2\left(1+\log((1-z)^{-1})\right)}\right)^{-1},$$

together with the expansion

$$T(z) = 2\,\frac{z^2}{2!} + 6\,\frac{z^3}{3!} + 60\,\frac{z^4}{4!} + 520\,\frac{z^5}{5!} + 6660\,\frac{z^6}{6!} + 93408\,\frac{z^7}{7!} + \cdots.$$

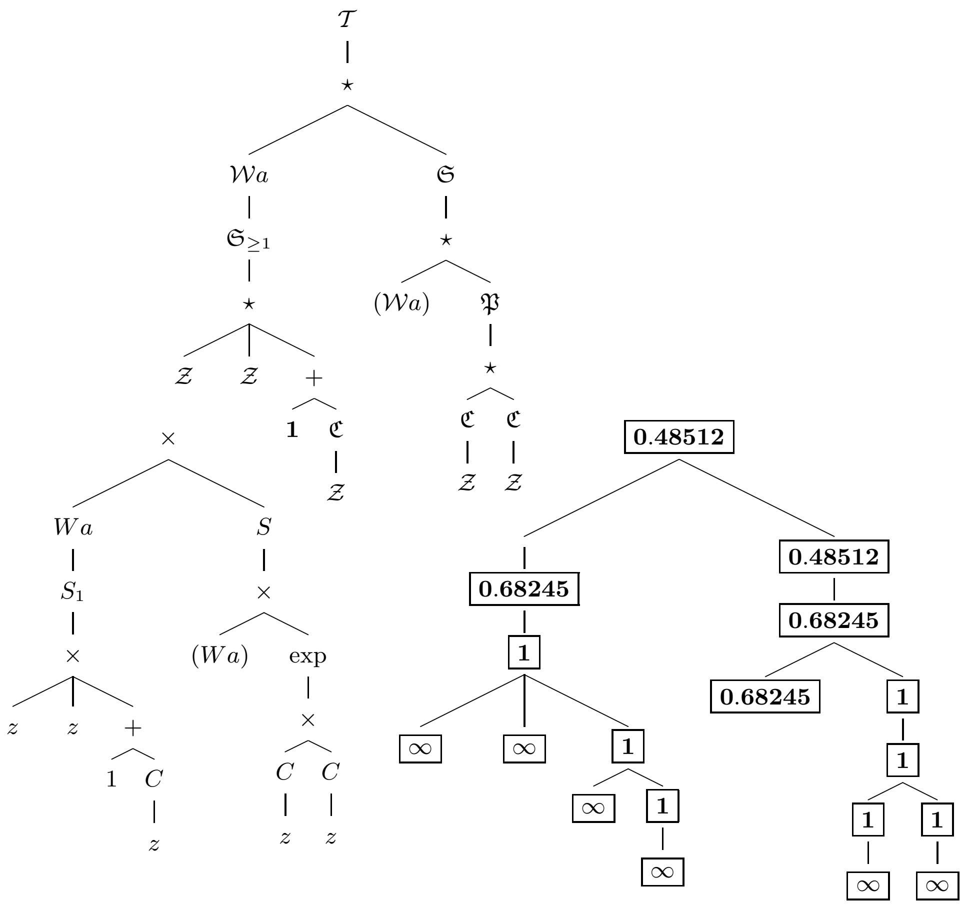


FIGURE 7. The inductive determination of the radius of convergence of the EGF of trains, $T(z)$: (top) a hierarchical view of the specification of $\mathcal{T}$; (bottom left) the corresponding expression tree of the EGF $T(z)$; (bottom right) the value of the radii for each subexpression of $T(z)$. (Notations: $L(y) = \log(1-y)^{-1}$, $S(y) = (1-y)^{-1}$, $S_1(y) = yS(y)$.)

The specification (20) has a hierarchical structure, as suggested by the top representation of Figure 7, and this structure is itself directly reflected by the form of the expression tree of the GF $T(z)$. Then each node in the expression tree of $T(z)$ can be tagged with the corresponding value of the radius of convergence. This is done according to the principles of Theorem IV.5; see the bottom-right part of Figure 7. For instance, the quantity 0.68245 associated to $Wa(z)$ is given by the sequence rule and is determined as smallest positive solution to the equation

$$z^2\left(1 - \log(1-z)^{-1}\right) = 1.$$

The tagging process works upwards till the root of the tree is reached; here the radius of convergence of T is determined to be $\rho \doteq 0.48512\cdots$, a quantity that happens to coincide with the ratio $[z^{49}]T(z)/[z^{50}]T(z)$ to more than 15 decimal places. $\square$

IV. 4. Rational and meromorphic functions

The first principle that we have just discussed in great detail is:

> *The location of singularities of an analytic function determines the exponential order of growth of its Taylor coefficients.*

The second principle which refines the first one is:

> *The nature of the singularities determines the way the dominant exponential term in coefficients is modulated by a subexponential factor.*

We are now going to develop the correspondence between singular expansions and asymptotic behaviours of coefficients in the case of rational and meromorphic functions. Rational functions (fractions) are the simpler ones, and from their basic partial fraction expansion closed forms are derived for their coefficients. Next in order of difficulty comes the class of meromorphic functions; their Taylor coefficients appear to admit very accurate asymptotic expansions with error terms that are exponentially small, as results from an adequate use of the residue theore.

In the case of rational and, more generally, meromorphic functions, the net effect is summarized by the correspondence:

$$\text{Polar singularities} \rightsquigarrow \text{Subexponential factors } \theta(n) \text{ are polynomials.}$$

A distinguishing feature is the extremely good quality of the asymptotic approximations obtained; for instance 15 digits of accuracy is not uncommon in coefficients of index as low as 50.

IV. 4.1. Rational functions. A function $f(z)$ is a *rational function* iff it is of the form $f(z) = \frac{N(z)}{D(z)}$, with $N(z)$ and $D(z)$ being polynomials, which me may always assume to be relatively prime. For rational functions that are generating functions, we have $D(0) \neq 0$.

Sequences $\{f_n\}_{n\geq 0}$ that are coefficients of rational functions coincide with sequences that satisfy linear recurrence relations with constant coefficients. To see it, compute $[z^n]f(z)\cdot D(z)$, with $n > \deg(N(z))$. If $D(z) = d_0 + d_1 z + \cdots + d_m z^m$, then for $n > m$, one has:

$$\sum_{j=0}^{m} d_j f_{n-j} = 0.$$

The main theorem we prove here provides an *exact* finite expression for coefficients of $f(z)$ in terms of the poles of $f(z)$. Individual terms in corresponding expressions are sometimes called exponential polynomials.

THEOREM IV.6 (Expansion of rational functions). *If $f(z)$ is a rational function that is analytic at zero and has poles at points $\alpha_1, \alpha_2, \ldots, \alpha_m$, then there exist m polynomials $\{\Pi_j(x)\}_{j=1}^m$ such that:*

$$f_n \equiv [z^n]f(z) = \sum_{j=1}^{m} \Pi_j(n)\alpha_j^{-n}. \tag{21}$$

Furthermore the degree of Π_j is equal to the order of the pole of f at α_j minus one.

An expression of the form (21) is sometimes called an *exponential polynomial*.

PROOF. Since $f(z)$ is rational it admits a partial fraction expansion. Thus, assuming without loss of generality that $\deg(D) > \deg(N)$, we can decompose f into a finite sum

$$f(z) = \sum_{(\alpha,r)} \frac{c_{\alpha,r}}{(z-\alpha)^r},$$

where α ranges over the poles of $f(z)$ and r is bounded from above by the multiplicity of α as a pole of f. Coefficient extraction in this expression results from Newton's expansion,

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

The binomial coefficient is a polynomial of degree $r-1$ in n, and collecting terms associated with a given α yields the statement of the theorem. $\square$

Notice that the expansion (21) is also an asymptotic expansion in disguise: when grouping terms according to the α's of increasing modulus, each group appears to be *exponentially smaller* than the previous one. A classical instance is the OGF of Fibonacci numbers,

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

with poles at

$$\frac{-1+\sqrt{5}}{2} \doteq 0.61803, \qquad \frac{-1-\sqrt{5}}{2} \doteq -1.61803,$$

so that

$$F_n = \frac{1}{\sqrt{5}}\varphi^n - \frac{1}{\sqrt{5}}\bar{\varphi}^n = \frac{\varphi^n}{\sqrt{5}} + O(\frac{1}{\varphi^n}),$$

with $\varphi = (1+\sqrt{5})/2$ the golden ratio, and $\bar{\varphi}$ its conjugate.

The next example is certainly an artificial one. It is simply designed to demonstrate that all the details of the full decomposition are usually not required. The rational function

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

has a pole of order 5 at $z=1$, poles of order 2 at $z=\omega,\omega^2$ ($\omega = e^{2i\pi/3}$ a cubic root of unity), a pole of order 3 at $z=-1$, and simple poles at $z=\pm\sqrt{2}$. Therefore,

$$f_n = P_1(n) + P_2(n)\omega^{-n} + P_3(n)\omega^{-2n} + P_4(n)(-1)^n + \\ +P_5(n)2^{-n/2} + P_6(n)(-1)^n 2^{-n/2}$$

where the degrees of $P_1,\ldots,P_6$ are respectively $4,1,1,2,0,0$. For an asymptotic equivalent of f_n, only the pole at $z=1$ needs to be considered since it corresponds to the fastest exponential growth; in addition, at $z=1$, only the term of fastest growth needs to be taken into account since it gives the dominant contribution to coefficients. Thus, we have the correspondence

$$f(z) \sim \frac{1}{3^2\cdot 2^3\cdot(\frac{1}{2})}\frac{1}{(1-z)^5} \implies f_n \sim \frac{1}{3^2\cdot 2^3\cdot(\frac{1}{2})}\binom{n+4}{4} \sim \frac{n^4}{864}.$$

EXAMPLE 4. *Asymptotics of denumerants.* Denumerants are synonymous to integer partitions with summands restricted to be from a *fixed* finite set (Chapter I). We let $\mathcal{P}^{\mathcal{T}}$ be the class relative to set $\mathcal{T}$, with the known OGF,

$$P^{\mathcal{T}}(z) = \prod_{\omega\in\mathcal{T}}\frac{1}{1-z^\omega}.$$

A particular case is the one of integer partitions whose summands are in $\{1,2,\ldots,r\}$,

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

The GF has all its poles that are roots of unity. At $z = 1$, the order of the pole is r, and one has

$$P^{\{1,\ldots,r\}}(z) \sim \frac{1}{r!}\frac{1}{(1-z)^r},$$

as $z \to 1$. Other poles have smaller multiplicity: for instance the multiplicity of $z = -1$ is equal to the number of factors $(1-z^{2j})^{-1}$ in $P^{\{1,\ldots,r\}}$, that is $\lfloor r/2 \rfloor$; in general a primitive qth root of unity is found to have multiplicity $\lfloor r/q \rfloor$. There results that $z = 1$ contributes a term of the form n^{r-1} to the coefficient of order n, while each of the other poles contributes a term of order at most $n^{\lfloor r/2 \rfloor}$. We thus find

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

The same argument provides the asymptotic form of $P_n^{\mathcal{T}}$, since, to first order asymptotics, only the pole at $z = 1$ counts. One then has:

PROPOSITION IV.2. *Let $\mathcal{T}$ be a finite set of integers without a common divisor* ($\gcd(\mathcal{T}) = 1$)*. The number of partitions with summands restricted to $\mathcal{T}$ satisfies*

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

For instance, in a country that would have pennies (1 cent), nickels (5 cents), dimes (10 cents) and quarters (25 cents), the number of ways to make change for a total of n cents is

$$[z^n]\frac{1}{(1-z)(1-z^5)(1-z^{10})(1-z^{25})} \sim \frac{1}{1\cdot 5\cdot 10\cdot 25}\frac{n^3}{3!} \equiv \frac{n^3}{7500},$$

asymptotically. □

IV. 4.2. Meromorphic Functions. An expansion very similar to that of Theorem IV.6 given for rational functions holds true for the larger class of coefficients of meromorphic functions.

THEOREM IV.7 (Expansion of meromorphic functions). *Let $f(z)$ be a function meromorphic for $|z| \le R$ with poles at points $\alpha_1, \alpha_2, \ldots, \alpha_m$, and analytic for $|z| = R$ and $z = 0$. Then there exist m polynomials $\{\Pi_j(x)\}_{j=1}^m$ such that:*

$$f_n \equiv [z^n]f(z) = \sum_{j=1}^m \Pi_j(n)\alpha_j^{-n} + \mathcal{O}(R^{-n}). \tag{22}$$

Furthermore the degree of Π_j is equal to the order of the pole of f at α_j minus one.

PROOF. We offer two different proofs, one based on subtracted singularities, the other one based on contour integration.

(i) *Subtracted singularities.* Around any pole α, $f(z)$ can be expanded locally:

$$f(z) = \sum_{k\ge -M} c_{\alpha,k}(z-\alpha)^k \tag{23}$$

$$= S_\alpha(z) + H_\alpha(z) \tag{24}$$

where the "singular part" $S_\alpha(z)$ is obtained by collecting all the terms with index in $[-M\,..-1]$ ($S_\alpha(z) = N_\alpha(z)/(z-\alpha)^M$ with $N_\alpha(z)$ a polynomial of degree less than M) and $H_\alpha(z)$ is analytic at α. Thus setting $R(z) = \sum_j S_{\alpha_j}(z)$, we observe that $f(z) - S(z)$ is analytic for $|z| \le R$. In other words, by collecting the singular parts of the expansions and subtracting them, we have "removed" the singularities of $f(z)$, whence the name of "*method of subtracted singularities*" sometimes given to the method [**66**, vol. 2, p. 448].

Taking coefficients, we get:

$$[z^n]f(z) = [z^n]S(z) + [z^n](f(z) - S(z)).$$

The coefficient of $[z^n]$ in the rational function $S(z)$ is obtained from Theorem 1. It suffices to prove that the coefficient of z^n in $f(z) - S(z)$, a function analytic for $|z| \le R$, is $\mathcal{O}(R^{-n})$. This fact follows from trivial bounds applied to Cauchy's integral formula with the contour of integration being $\lambda = \{z \,/\, |z| = R\}$, as in the proof of Theorem IV.4:

$$|[z^n](f(z) - S(z))| = \frac{1}{2\pi}\left|\int_{|z|=R} (f(z) - S(z))\,\frac{dz}{z^{n+1}}\right| \le \frac{1}{2\pi}\frac{\mathcal{O}(1)}{R^{n+1}}2\pi R.$$

(ii) *Contour integration.* There is another line of proof for Theorem IV.7 which we briefly sketch as it provides an insight which is useful for applications to other types of singularities treated in Chapter V. It consists in using directly Cauchy's coefficient formula and "pushing" the contour of integration past singularities. In other words, one computes directly the integral

$$I_n = \frac{1}{2i\pi}\int_{|z|=R} f(z)\,\frac{dz}{z^{n+1}}$$

by residues. There is a pole at $z = 0$ with residue f_n and poles at the α_j with residues corresponding to the terms in the expansion stated in Theorem IV.7; for instance, if $f(z) \sim c/(z-a)$ as $z \to a$, then

$$\mathrm{Res}(f(z)z^{-n-1}, z = a) = \mathrm{Res}(\frac{c}{(z-a)}z^{-n-1}, z = a) = \frac{c}{a^{n+1}}.$$

Finally, by the same trivial bounds as before, I_n is $\mathcal{O}(R^{-n})$. □

EXAMPLE 5. *Surjections and alignments.* The surjection EGF is $R(z) = (2 - e^z)^{-1}$, and we have already determined its poles: the one of smallest modulus is at $\log 2 \doteq 0.69314$. At the dominant pole, as z tends to $\log 2$, one has $R(z) \sim -\frac{1}{2}(z - \log 2)^{-1}$. This implies an approximation for the number of surjections:

$$R_n \equiv n![z^n]R(z) \sim \xi(n), \qquad \text{with} \quad \xi(n) := \frac{n!}{2}\cdot\big(\frac{1}{\log 2}\big)^{n+1}.$$

Here is, for $n = 2, 4, \ldots, 32$, a table of the values of the surjection numbers (left) compared with the asymptotic approximation rounded[5] to the nearest integer, $\lceil \xi(n) \rfloor$: It is piquant to see that $\lceil \xi(n) \rfloor$ provides the exact value of R_n for all values of $n = 1, \ldots, 15$, and it starts losing one digit for $n = 17$, after which point a few "wrong" digits gradually appear, but in very limited number; see Figure 8 The explanation of such a faithful asymptotic representation owes to the fact that the error terms provided by meromorphic asymptotics are exponentially small. In effect, there is no other pole in $|z| \le 6$, the next ones being at $\log 2 \pm 2i\pi$ with modulus of about 6.32. Thus, for $r_n = [z^n]R(z)$, there holds

$$\frac{R_n}{n!} \sim \frac{1}{2}\cdot\big(\frac{1}{\log 2}\big)^{n+1} + \mathcal{O}(6^{-n}). \tag{25}$$

For the double surjection problem, $R^*(z) = (2 + z - e^z)$, we get similarly

$$[z^n]R^*(z) \sim \frac{1}{e^{\rho^*} - 1}\,(\rho^*)^{-n-1},$$

with $\rho^* = 1.14619$ the smallest positive root of $e^{\rho^*} - \rho^* = 2$.

[5]The notation $\lceil x \rfloor$ represnets x rounded to the nearest integer: $\lceil x \rfloor := \lfloor x + \frac{1}{2} \rfloor$.

3	3
75	75
4683	4683
545835	545835
102247563	102247563
28091567595	28091567595
10641342970443	10641342970443
5315654681981355	5315654681981355
3385534663256845323	338553466325684532 ***6***
2677687796244384203115	2677687796244384203 ***088***
2574844419803190384544203	2574844419803190384544 ***450***
2958279121074145472650648875	295827912107414547265064 ***6597***
4002225759844168492486127539083	40022257598441684924861275 ***55859***
6297562064950066033518373935334635	6297562064950066033518373935 ***416161***
11403568794011880483742464196184901963	1140356879401188048374246419617 ***4527074***
23545154085734896649184490637144855476395	2354515408573489664918449063714 ***5314147690***

FIGURE 8. The surjection numbers pyramid: for $n = 2, 4, \ldots, 32$, the exact values of the numbers R_n (left) compared to the approximation $\lceil \xi(n) \rfloor$ with discrepant digits in boldface (right).

Alignments are sequences of cycles, with EGF

$$f(z) = \frac{1}{1 - \log(1-z)^{-1}}.$$

There is a singularity when $\log(1-z)^{-1} = 1$, which is at $z = 1 - e^{-1}$ and arises before $z = 1$ where the logarithm becomes singular. Thus the computation of the asymptotic form of f_n only needs a local expansion near $(1 - e^{-1})$:

$$f(z) \sim \frac{-e^{-1}}{z - 1 + e^{-1}} \qquad \Longrightarrow \qquad [z^n]f(z) \sim \frac{e^{-1}}{(1-e^{-1})^{n+1}}.$$

$\square$

$\triangleright$ **21.** *Some "supernecklaces".* One estimates

$$[z^n] \log\left(\frac{1}{1 - \log\frac{1}{1-z}}\right) \sim \frac{1}{n}(1-e^{-1})^{-n},$$

where the EGF enumerates (labelled) cycles of cycles. [Hint: Take derivatives.] $\triangleleft$

EXAMPLE 6. *Generalized derangements.* The probability that the shortest cycle in a random permutation of size n has length larger than k is

$$[z^n]\frac{e^{-\frac{z}{1}-\frac{z^2}{2}-\cdots-\frac{z^k}{k}}}{1-z}.$$

For any fixed k, the generating function, call it $f(z)$, is equivalent to $e^{-H_k}/(1-z)$ as $z \to 1$. Accordingly the coefficients $[z^n]f(z)$ tend to e^{-H_k} as $n \to \infty$. Thus, due to meromorphy, we have the characteristic implication

$$f(z) \sim \frac{e^{-H_k}}{1-z} \qquad \Longrightarrow \qquad [z^n]f(z) \sim e^{-H_k}.$$

Since the difference between $f(z)$ and the approximation at 1 is an entire function, the error is exponentially small:

$$[z^n]\frac{e^{-\frac{z}{1}-\frac{z^2}{2}-\cdots-\frac{z^k}{k}}}{1-z} = e^{-H_k} + O(R^{-n}), \tag{26}$$

for *fixed* k and *any* $R > 1$. The cases $k = 1, 2$ in particular justify the estimates mentioned in the introduction on p. 4.

As a side remark, the classical approximation of the harmonic numbers, $H_k \approx \log k + \gamma$ suggests $e^{-\gamma}/k$ as a further approximation to (26) that might be valid for both large n and large k in suitable regions. This can be made precise; in accordance with this heuristic argument, the expected length of the shortest cycle in a random permutation of size n is symptotic to

$$\sum_{k=1}^{n} \frac{e^{-\gamma}}{k} \sim e^{-\gamma} \log n,$$

as first proved by Shepp and Lloyd in [**101**]. $\square$

$\triangleright$ **22.** *Shortest cycles of permutations are not too long.* Let S_n be the random variable denoting the length of the shortest cycle in a random permutation of size n. Using the circle $|z| = 2$ to estimate the error in the approximation e^{-H_k} above, one finds that, for $k \leq \log n$,

$$\left| \mathbb{P}(S_n > k) - e^{-H_k} \right| \leq \frac{1}{2^n} e^{2^k},$$

which is exponentially small in this range of k-values. Thus, the approximation e^{-H_k} remains good when k is allowed to tend sufficiently slowly to ∞ with n. One can also explore the possibility of better bounds and larger regions of validity of the main approximation. (See Panario and Richmond's study [**93**] for a general theory of smallest components in sets.) $\triangleleft$

EXAMPLE 7. *Smirnov words and Carlitz compositions.* This examples illustrates the analysis of a group of rational generating functions (Smirnov words) paralleling nicely the enumeration of a special type of integer composition (Carlitz compositions) resorting to meromorphic asymptotics.

Bernoulli trials have been discussed in Chapter III, in relation to weighted word models. Take the class $\mathcal{W}$ of all words over an r-ary alphabet, where letter j is assigned probability p_j and letters of words are drawn independently. With this weighting, the GF of all words is

$$W(z) = \frac{1}{1 - \sum p_j z} = \frac{1}{1 - z}.$$

Consider the problem of determining the probability that has a random word of length n is of Smirnov type, i.e., all blocks of length 2 are formed with two distinct letters (see also [**60**, p. 69]).

By our discussion of Section III.6, the GF of Smirnov words (again with the probabilistic weighting) is

$$S(z) = \frac{1}{1 - \sum \frac{p_j z}{1 + p_j z}}.$$

This is a rational function with a unique dominant singularity at ρ such that

$$\sum_{j=1}^{r} \frac{p_j \rho}{1 + p_j \rho} = 1. \tag{27}$$

(It is easy to verify by monotonicity that this equation has a unique positive solution.) Thus, ρ is a well characterized algebraic number defined implicitly by an equation of degree r. There results that the probability for a word to be Smirnov is (not too surprisingly) exponentially small, with the precise formula being

$$[z^n] S(z) \sim C \cdot \rho^{-n}, \qquad C = \left(\rho \sum \frac{p_i \rho}{1 + p_i \rho} \right)^{-1}.$$

A similar analysis, but with bivariate generating functions shows that in a random word of length n conditioned to be Smirnov, the letter j appears with frequency asymptotic to

$$q_j = \frac{p_j \rho}{1 + p_j \rho}, \tag{28}$$

in the sense that mean number of occurrences of letter j is asymptotic to $q_j n$. All these results are seen to be consistent with the equiprobable letter case $p_j = 1/r$, for which $\rho = r/(r-1)$.

Carlitz compositions illustrate a similar situation, in which the alphabet is in a sense infinite, while letters have different sizes. Recall that a Carlitz composition of the integer n is a composition of n such that no two adjacent summands have equal values. Consider first compositions with a bound m on the largest allowable summand. The OGF of such Carlitz compositions is

$$C^{[m]}(z) = \left(1 - \sum_{j=1}^{m} \frac{z^j}{1+z^j}\right)^{-1},$$

and the OGF of all Carlitz compositions is obtained by letting m tend to infinity:

$$C^{[\infty]}(z) = \left(1 - \sum_{j=1}^{\infty} \frac{z^j}{1+z^j}\right)^{-1}. \tag{29}$$

In particular, we get *EIS* **A003242**[6]:

$$C^{[\infty]}(z) = 1 + z + z^2 + 3z^3 + 4z^4 + 7z^5 + 14z^6 + 23z^7 + 39z^8 + 71z^9 + \cdots .$$

The asymptotic form of the number of Carlitz compositions is then easily found by singularity analysis of meromorphic functions. The OGF has a simple pole at ρ which is the smallest positive root of the equation

$$\sum_{j=1}^{\infty} \frac{\rho^j}{1+\rho^j} = 1. \tag{30}$$

(Note the formal analogy with (27) due to commonality of the combinatorial argument.)

$$C_n^{[\infty]} \sim C \cdot \alpha^n, \quad C \doteq 0.45638, \quad \alpha \doteq 1.75024\,.$$

There, $\alpha = \rho$ with ρ as in (30). In a way analogous to Smirnov words, the asymptotic frequency of summand k appears to be $\rho^k/(1+\rho^k)$; see [**71, 83**] for further properties. □

IV. 5. Localization of singularities

There are situations where a function possesses several dominant singularities, that is, several singularities are present on the boundary of its disk of convergence. We examine here the induced effect on the coefficient's coefficients and discuss ways to localize such dominant singularities.

[6]The *EIS* designates Sloane's On-Line Encyclopedia of Integer Sequences [**102**]; see [**103**] for an earlier printed version.

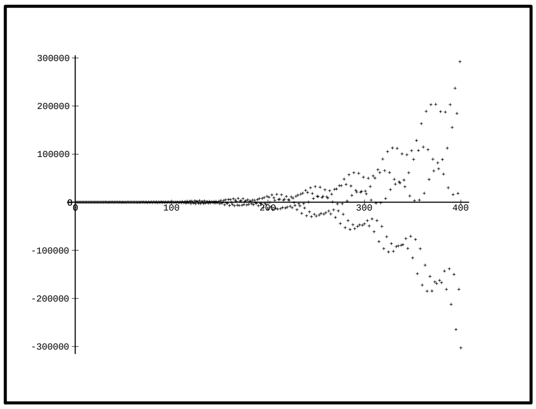


FIGURE 9. The coefficients $[z^n]f(z)$, where $f(z) = \left(1+1.02z^4\right)^{-3}\left(1-1.05z^5\right)^{-1}$ illustrate a periodic superposition of smooth behaviours that depend on the residue class of n modulo 20.

IV. 5.1. Multiple singularities. In the presence of multiple singularities on the circle of convergence of a series, several geometric terms of the form α^n sharing the same modulus must be combined. In simpler cases, such terms induce a periodic behaviour for coefficients that is easy to describe; in the more general case, fluctuations of a somewhat "arithmetic nature" result. Finally, consideration of *all* singularities (whether dominant or not) of a meromorphic functions may lead to explicit summations expressing their coefficients.

Periodicities. When several singularities of $f(z)$ have the same modulus, they may induce complete cancellations, so that different regimes will be present in the coefficients of f. For instance

$$\frac{1}{1+z^2} = 1 - z^2 + z^4 - z^6 + z^8 - \cdots, \quad \frac{1}{1-z^3} = 1 + z^3 + z^6 + z^9 + \cdots,$$

exhibit patterns of periods 4 and 3 respectively, this corresponding to roots of unity or order 4 $(\pm i)$, and 3. Accordingly,

$$\phi(z) = \frac{1}{1+z^2} + \frac{1}{1-z^3} = \frac{2 - z^2 + z^3 + z^4 + z^8 + z^9 - z^{10}}{1 - z^{12}}$$

has a pattern of period 12, and the coefficients ϕ_n such that $n \equiv 1, 5, 6, 7, 11$ modulo 12 are zero. Consequently, if we analyze

$$[z^n]\psi(z) \qquad \text{where} \qquad \psi(z) = \phi(z) + \frac{1}{1 - z/2},$$

we see that a different exponential growth manifests itself when n is taken congruent to $1, 5, 6, 7, 11 \bmod 12$. In many combinatorial applications, generating functions involving periodicities can be decomposed "at sight", and the corresponding asymptotic subproblems generated are then solved separately.

$\triangleright$ **23.** *Decidability of polynomial properties.* Given a polynomial $p(z) \in \mathbb{Q}[z]$, the following properties are decidable: (i) whether one of the zeros of p is a root of unity; (ii) whether one of the zeros of p has an argument that is commensurate with π. [One can use resultants. An algorithmic discussion of this and related issues is given in [**62**].] $\triangleleft$

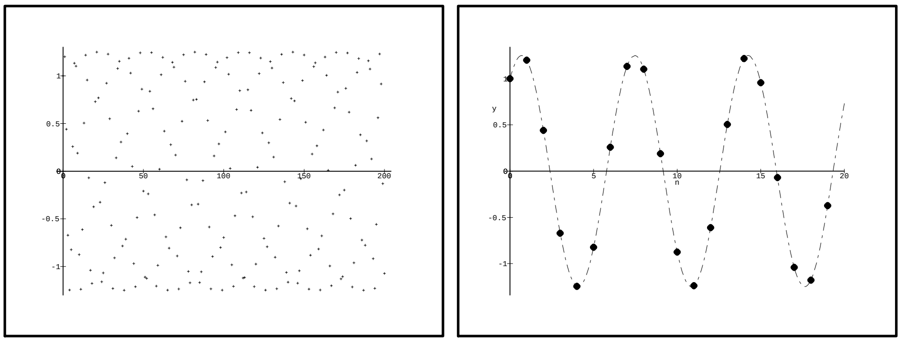


FIGURE 10. The coefficients of $f = 1/(1 - \frac{6}{5}z + z^2)$ exhibit an apparently chaotic behaviour (left) which in fact corresponds to a discrete sampling of a sine function (right), reflecting the presence of two conjugate complex poles.

Nonperiodic fluctuations. Take the polynomial $D(z) = 1 - \frac{6}{5}z + z^2$, whose roots are

$$\alpha = \frac{3}{5} + i\frac{4}{5}, \ \bar{\alpha} = \frac{3}{5} - i\frac{4}{5},$$

both of modulus 1 (the numbers $3, 4, 5$ form a Pythagorean triple), with argument $\pm\theta$ where $\theta = \arctan(\frac{4}{3}) = 0.9279$. The expansion of the function $f(z) = 1/D(z)$ starts as

$$\frac{1}{1 - \frac{6}{5}z + z^2} \quad = \quad 1 + \frac{6}{5}z + \frac{11}{25}z^2 - \frac{84}{125}z^3 - \frac{779}{625}z^4 - \frac{2574}{3125}z^5 + \cdots$$

the sign sequence being

$$+++---++++---+++----+++----+++---,$$

which indicates a mildly irregular oscillating behaviour, where blocks of 3 or 4 pluses follow blocks of 3 or 4 minuses.

The exact form of the coefficients of f results from a partial fraction expansion:

$$f(z) = \frac{a}{1 - z/\alpha} + \frac{b}{1 - z/\bar{\alpha}} \qquad \text{with } a = \frac{1}{2} + \frac{3}{8}i, \ b = \frac{1}{2} - \frac{3}{8}i.$$

Accordingly,

$$\begin{aligned} f_n &= ae^{-in\theta_0} + be^{in\theta_0} \\ &= \frac{\sin((n+1)\theta_0)}{\sin(\theta_0)}. \end{aligned}$$

This explains the sign changes observed. Since the angle θ_0 is not commensurate with π, the coefficients fluctuate but, unlike in our earlier examples, no exact periodicity is present in the sign patterns. See Figure 10 for a rendering and Figure 10 below for a meromorphic case linked to compositions into prime summands.

Complicated problems of an arithmetical nature may occur if several such singularities with non–commensurable arguments combine, and some open problem remain in the analysis of linear recurring sequences. (For instance no decision procedure is known to determine whether such a sequence ever vanishes.) Fortunately, such problems occur infrequently in combinatorial enumerations where zeros of rational functions tend to have a simple geometry.

Exact formulæ. The error terms appearing in the asymptotic expansion of coefficients of meromorphic functions are already exponentially small. By “pealing off” the singularities of a meromorphic function layer by layer, in order of increasing modulus, one is led to extremely precise expansions for the coefficients. Sometimes even, “exact” expressions may result. The latter is the case for the Bernoulli numbers B_n, the surjection numbers R_n, the Secant numbers E_{2n} and the Tangent numbers E_{2n+1} defined by

$$\begin{array}{lcll} \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} & = & \frac{z}{e^z-1} & \text{(Bernoulli numbers)} \\ \sum_{n=0}^{\infty} R_n \frac{z^n}{n!} & = & \frac{1}{2-e^z} & \text{(Surjection numbers)} \\ \sum_{n=0}^{\infty} E_{2n} \frac{z^{2n}}{(2n)!} & = & \frac{1}{\cos(z)} & \text{(Secant numbers)} \\ \sum_{n=0}^{\infty} E_{2n+1} \frac{z^{2n+1}}{(2n+1)!} & = & \tan(z) & \text{(Tangent numbers).} \end{array}$$

Bernoulli numbers have an EGF $z/(e^z-1)$ that has poles at the points $\chi_k = 2ik\pi$, with $k \in \mathbb{Z} \setminus \{0\}$. The residue at χ_k is equal to χ_k,

$$\frac{z}{e^z-1} \sim \frac{\chi_k}{z-\chi_k} \qquad (z \to \chi_k).$$

The expansion theorem for meromorphic functions is applicable here. To see it use the Cauchy integral formula, and proceed as in the proof of Theorem IV.7, using as external contours large circles that pass half way between poles. Along these contours, the integrand tends to 0 because the Cauchy “kernel” z^{-n-1} decreases with the radius of the integration contour while the EGF stays bounded. In the limit, corresponding to an infinitely large contour, the coefficient integral becomes equal to the sum of all residues of the meromorphic function over the whole of the complex plane.

From this argument, we thus get: $\frac{B_n}{n!} = -\sum_{k\in\mathbb{Z}\setminus\{0\}} \chi_k^{-n}$. This proves that $B_n = 0$ if n is odd. If n is even, then grouping terms two by two, we get the exact representation (which also serves as an asymptotic expansion):

$$\frac{B_{2n}}{(2n)!} = (-1)^{n-1} 2^{1-2n} \pi^{-2n} \sum_{k=0}^{\infty} \frac{1}{k^{2n}}. \tag{31}$$

Reverting the equality, we have also established that

$$\zeta(2n) = (-1)^{n-1} 2^{2n-1} \pi^{2n} \frac{B_{2n}}{(2n)!} \qquad \text{with} \quad \zeta(s) = \sum_{k=1}^{\infty} \frac{1}{k^s}, \quad B_n = n![z^n]\frac{z}{e^z-1},$$

a well-known identity that provides values of the Riemann zeta function ($\zeta(s)$) at even integers as rational multiples of powers of π.

In the same vein, the *surjection numbers* have as EGF $R(z) = (2-e^z)^{-1}$ with simple poles at

$$\chi_k = \log 2 + 2ik\pi \qquad \text{where} \qquad R(z) \sim \frac{1}{2}\frac{1}{\chi_k - z}.$$

Since $R(z)$ stays bounded on circles passing half way in between poles, we find the exact formula, $\frac{R_n}{n!} = \frac{1}{2}\sum_{k\in\mathbb{Z}} \chi_k^{-n-1}$. An equivalent real formulation is

$$\frac{R_n}{n!} = \frac{1}{2}\left(\frac{1}{\log 2}\right)^{n+1} + \sum_{k=1}^{\infty} \frac{\cos((n+1)\theta_k)}{(\log^2 2 + 4k^2\pi^2)^{(n+1)/2}} \text{ with } \theta_k = \arctan(\frac{2k\pi}{\log 2}), \tag{32}$$

which shows the hidden occurrence of infinitely many "harmonics" of fast decaying amplitude.

$\triangleright$ **24.** *Alternating permutations, tangent and secant numbers.* The relation (31) also provides a representation of the *tangent numbers* since $E_{2n-1} = (-1)^{n-1}B_{2n}4^n(4^n-1)/(2n)$. The secant numbers E_{2n} satisfy

$$\sum_{k=1}^{\infty} \frac{(-1)^k}{(2k+1)^{2n+1}} = \frac{(\pi/2)^{2n+1}}{2\,(2n)!}E_{2n},$$

which can be read either as providing an asymptotic expansion of E_{2n} or as an evaluation of the sums on the left (the values of a Dirichlet L-function) in terms of π. The asymptotic number of alternating permutations (Chapter II) is consequently known to great accuracy. $\triangleleft$

$\triangleright$ **25.** *Solutions to the equation* $\tan(x) = x$. Let x_n be the nth positive root of the equation $\tan(x) = x$. For any integer $r \geq 1$, the sum $\sum_n x_n^{-2r}$ is a computable rational number. [From folklore and *The American Mathematical Monthly*.] $\triangleleft$

IV. 5.2. Localization of zeros and poles. We gather here a few results that often prove useful in determining the location of zeros of analytic functions, and hence of poles of meromorphic functions. A detailed treatment of this topic may be found in Henrici's book [**66**].

Let $f(z)$ be an analytic function in a region Ω and let γ be a simple closed curve interior to Ω, and on which f is assumed to have no zeros. We claim that the quantity

$$N(f;\gamma) = \frac{1}{2i\pi}\int_\gamma \frac{f'(z)}{f(z)}\,dz$$

exactly equals the number of zeros of f inside γ counted with multiplicity. The reason is that the function f'/f has its poles exactly at the zeros of f, and its residue at each pole is 1, so that the assertion directly results from the residue theorem

Since a primitive function of f'/f is $\log f$, the integral also represents the variation of $\log f$ along γ, which is written $[\log f]_\gamma$. The variation $[\log f]_\gamma$ reduces to i times the variation of the argument of f along γ as $\log(re^{i\theta}) = \log r + i\theta$ and the modulus r has variation equal to 0 along a closed contour, $[\log\rho]_\gamma = 0$. The quantity $[\theta]_\gamma$ is, by its definition, the number of times the transformed contour $f(\gamma)$ winds about the origin. This observation is known as the *Argument Principle*:

> **Argument Principle.** *The number of zeros of $f(z)$ (counted with multiplicities) inside γ equals the winding number of the transformed contour $f(\gamma)$ around the origin.*

By the same argument, if f is meromorphic in $\Omega \ni \gamma$, then $N(f;\gamma)$ equals the difference between the number of zeros and the number of poles of f inside γ, multiplicities being taken into account. Figure 11 exemplifies the use of the argument principle in localizing zeros of a polynomial.

By similar devices, we get Rouché's theorem:

> **Rouché's theorem.** *Let the functions $f(z)$ and $g(z)$ be analytic in a region containing in its interior the closed simple curve γ. Assume that f and g satisfy $|g(z)| < |f(z)|$* on *the curve γ. Then $f(z)$ and*

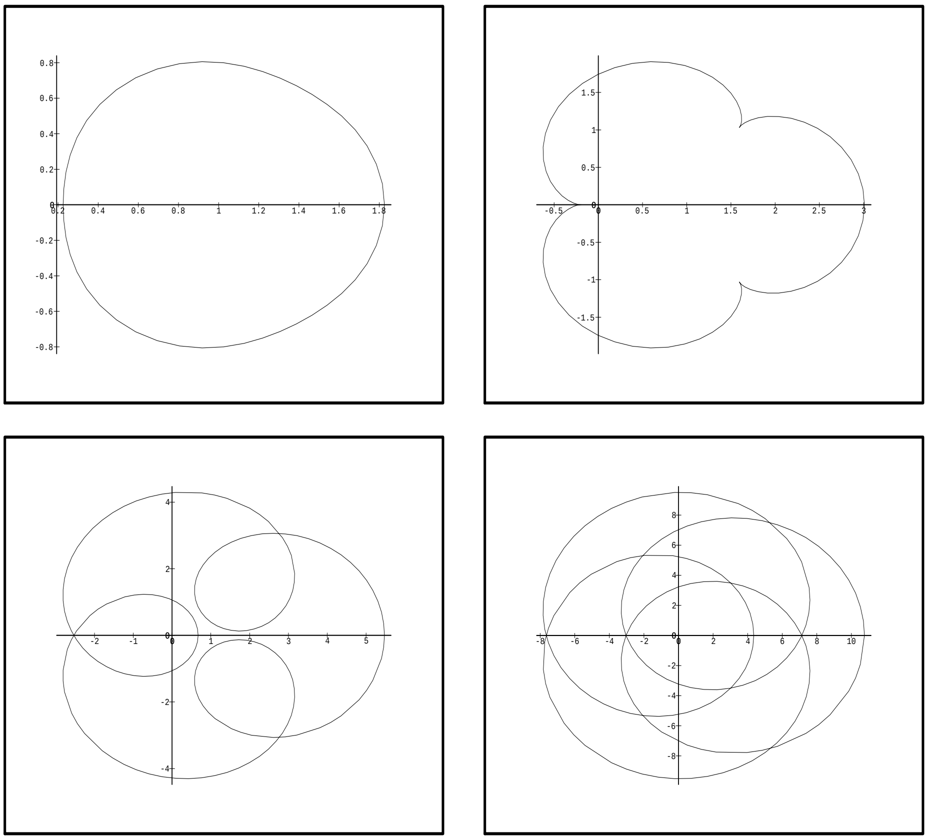


FIGURE 11. The transforms of $\gamma_j = \{|z| = \frac{4j}{10}\}$ by $P_4(z) = 1 - 2z + z^4$, for $j = 1, 2, 3, 4$, demonstrate that $P_4(z)$ has no zero inside $|z| < 0.4$, one zero inside $|z| < 0.8$, two zeros inside $|z| < 1.2$ and four zeros inside $|z| < 1.6$. The actual zeros are at $\rho_4 = 0.54368$, 1 and $1.11514 \pm 0.77184i$.

> $f(z) + g(z)$ *have the same number of zeros* inside *the interior domain delimited by* γ.

The intuition behind Rouché's theorem is that, since $|g| < |f|$, then $f(\gamma)$ and $(f+g)(\gamma)$ must have the same winding number.

$\triangleright$ **26.** *Proof of Rouché's theorem.* Under the hypothesis of Rouché's theorem, for $0 \le t \le 1$ $h(z) = (f(z) + tg(z))$ is such that $N(h;\gamma)$ is both an integer and a continuous function of t in the given range. The conclusion of the theorem follows. $\triangleleft$

$\triangleright$ **27.** *The fundamental theorem of algebra.* Every complex polynomial $p(z)$ of degree n has exactly n roots. A proof follows by Rouché's theorem from the fact that, for large enough $|z| = R$, the polynomial assumed to be monic is a "perturbation" of its leading term, z^n. $\triangleleft$

These principles form the basis of numerical algorithms for locating zeros of analytic functions. For instance, one can start from an initial domain and recursively subdivide it until roots have been isolated with enough precision—the number of roots in a subdomain being at each stage determined by numerical integration; see Figure 11 and refer for instance to [**27**] for a discussion. Such algorithms can even acquire the status of full proofs if one operates with guaranteed precision routines (using, e.g., careful implementations of interval arithmetics). Examples of use of the method will appear in the next sections.

$\triangleright$ **28.** *The analytic Implicit Function Theorem from residues.* The sum of the roots of the equation $g(y) = 0$ interior to γ equals

$$\frac{1}{2i\pi}\int_\gamma \frac{g'(y)}{g(y)}\, y\, dy.$$

Let $F(z, y)$ be an analytic function in both z and y (i.e., it admits a convergent series expansion). If $F'_y(z_0, y_0) \neq 0$, then the function $y(z)$ implicitly defined by $F(z, y) = 0$ and such that $y(z_0) = y_0$ is given by

$$y(z) = \frac{1}{2i\pi}\int_\gamma \frac{F'_y(z, y)}{F(z, y)}\, y\, dy,$$

where γ is a small loop around y_0. Deduce that $y(z)$ is analytic at z_0. (Note: this requires a modicum of analytic functions of two complex variables as is to be found, e.g., in [**17**].) $\triangleleft$

IV. 5.3. The example of patterns in words. All patterns are not born equal. Surprisingly, in a random sequence of coin tossings, the pattern *HTT* is likely to occur much sooner (after 8 tosses on average) than the pattern *HHH* (needing 14 tosses on average); see the preliminary discussion in Chapter I. Questions of this sort are of obvious interest in the statistical analysis of genetic sequences. Say you discover that a sequence of length 100,000 on the four letters A, G, C, T contains the pattern *TACTAC* twice. Can this be assigned to chance or is this is likely to be a meaningful signal of some yet unknown structure? The difficulty here lies in quantifying precisely where the asymptotic regime starts, since, by Borges's Theorem (see the Note in Chapter I), sufficiently long texts will almost certainly contain any fixed pattern. The analysis of rational generating functions supplemented by Rouché's theorem provides definite answers to such questions.

We consider here the class $\mathcal{W}$ of words over an alphabet $\mathcal{A}$ of cardinality $m \geq 2$. A pattern $\mathfrak{p}$ of some length k is given. As seen in Chapters I and III, its autocorrelation polynomial is central to enumeration. This polynomial is defined as $c(z) = \sum_{j=0}^{k-1} c_j z^j$, where c_j is 1 if $\mathfrak{p}$ coincides with its kth shifted version and 0 otherwise. We consider here the enumeration of words containing the pattern $\mathfrak{p}$ at least once, and dually of words excluding the pattern $\mathfrak{p}$. In other words, we look at problems like: What is the probability that a random of words of length n does (or does not) contain your name as a block of consecutive letters?

The OGF of the class of words excluding $\mathfrak{p}$ is, we recall,

$$S(z) = \frac{c(z)}{z^k + (1 - mz)c(z)}. \tag{33}$$

and we shall start with the case $m = 2$ of a binary alphabet. The function $S(z)$ is simply a rational function, but the location and nature of its poles is yet unknown. We only know *a priori* that it should have a pole in the positive interval somewhere between $\frac{1}{2}$ and 1 (by Pringsheim's Theorem and since its coefficients are in the interval $[1, 2^n]$, for n large enough). Here is a small list for patterns of length $k = 3, 4$ of the pole ρ nearest to the origin:

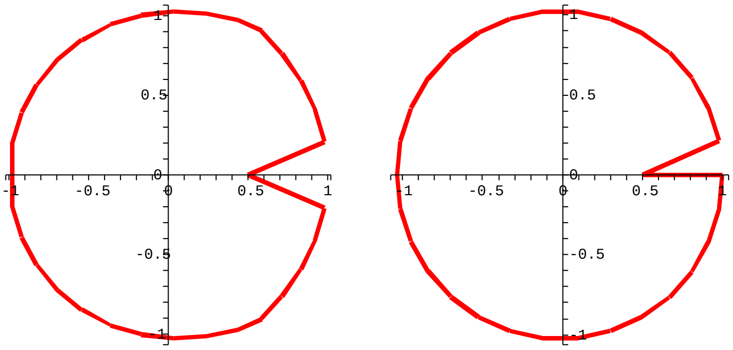


FIGURE 12. Complex zeros of $z^{31}+(1-2z)c(z)$ represented as joined by a polygonal line: (left) correlated pattern $a(ba)^{15}$; (right) uncorrelated pattern $a(ab)^{15}$.

Length (k)	Types	$c(z)$	ρ
$k=3$	aab, abb, ...	1	0.61803
	aba, bab	$1+z^2$	0.56984
	aaa, bbb	$1+z+z^2$	0.54368
$k=4$	aaab, aabb, abbb, ...	1	0.54368
	aaba, abba, abaa, ...	$1+z^3$	0.53568
	abab, baba	$1+z^2$	0.53101
	aaaa, bbbb	$1+z+z^2+z^3$	0.51879

We thus expect ρ to be close to $\frac{1}{2}$ as soon as the pattern is long enough. In order to prove this, we are going to apply Rouché's Theorem to the denominator of (33).

As regards termwise domination of coefficients, the autocorrelation polynomials lies between 1 (for less correlated patterns like aaa...b) and $1+z+\cdots+z^{k-1}$ (for the special case aaa...a). We set aside the special case of $\mathfrak{p}$ having only equal letters, i.e., a "maximal" autocorrelation polynomial—this case is discussed at length in the next chapter. Thus, in this scenario, the autocorrelation polynomial starts as $1+z^{\ell}+\cdots$ for some $\ell \geq 2$. Fix the number $A=0.6$. On $|z|=A$, we have

$$|c(z)| \geq \left|1-(A^2+A^3+\cdots)\right| = \left|1-\frac{A^2}{1-A}\right| = \frac{1}{10}. \tag{34}$$

In addition, the quantity $(1-2z)$ ranges over the circle of diameter $[-0.2, 1.2]$ as z varies along $|z|=A$, so that $|1-2z| \geq 0.2$. All in all, we have found that, for $|z|=A$,

$$|(1-2z)c(z)| \geq 0.02.$$

On the other hand, for $k>7$, we have $|z^k| < 0.017$ on the circle $|z|=A$. Then, amongst the two terms composing the denominator of (33), the first is strictly dominated by the second along $|z|=A$. By virtue of Rouché's Theorem, the number of roots of the denominator inside $|z| \leq A$ is then same as the number of roots of $(1-2z)c(z)$. The latter number is 1 (due to the root $\frac{1}{2}$) since $c(z)$ cannot be 0 by the argument of (34). Figure 12 exemplifies the extremely well-behaved characters of the complex zeros.

In summary, we have found that for all patterns with at least two different letters ($\ell \geq 2$) and length $k \geq 8$, the denominator has a unique root in $|z| \leq A = 0.6$. The property for lengths k satisfying $4 \leq k \leq 7$ is then easily verified directly. The case $\ell = 1$ can be subjected to an entirely similar argument (see Chapter V for details). Therefore, unicity of a simple pole ρ of $S(z)$ in the interval $(0.5, 0.6)$ is granted.

It is then a simple matter to determine the local expansion of $s(z)$ near $z = \rho$,

$$S(z) \underset{z\to\rho}{\sim} \frac{\widetilde{\Lambda}}{\rho - z}, \qquad \widetilde{\Lambda} := \frac{c(\rho)}{2c(\rho) - k\rho^{k-1}},$$

from which a precise estimate for coefficients derives by Theorems IV.6 and IV.7.

The computation finally extends almost verbatim to nonbinary alphabets, with ρ being now close to $\frac{1}{m}$. It suffices to use the disc of radius $A = 1.2/m$. The Rouché part of the argument grants us unicity of the dominant pole in the interval $(1/m, A)$ for $k \geq 5$ when $m = 3$, and for $k \geq 4$ and any $m \geq 4$. (The remaining cases are easily checked individually.)

PROPOSITION IV.3. *Consider an m-ary alphabet. Let $\mathfrak{p}$ be a pattern of length $k \geq 4$ with autocorrelation polynomial $c(z)$. Then the probability that a random word of length n does not contain $\mathfrak{p}$ as a pattern (a block of consecutive leters) satisfies*

$$\mathbb{P}_{\mathcal{W}_n}(\mathfrak{p} \textit{ does not occur}) = \Lambda_{\mathfrak{p}}(m\rho)^{-n-1} + O\left((\frac{5}{6})^n \right), \tag{35}$$

where $\rho \equiv \rho_{\mathfrak{p}}$ is the unique root in $(\frac{1}{m}, \frac{6}{5m})$ of the equation $z^k + (1 - mz)c(z) = 0$ and

$$\Lambda_{\mathfrak{p}} = \frac{mc(\rho)}{mc(\rho) - k\rho^{k-1}}.$$

Despite their austere appearance, these formulæ have indeed an a fairly intuitive content. First, the equation satisfied by ρ can be put under the form $mz = 1 + m^{-k}/c(z)$, and, since ρ is close to $\frac{1}{m}$, we may expect the approximation

$$m\rho \approx 1 + \frac{1}{\gamma m^k},$$

where $\gamma := c(m^{-1})$ satisfies $1 \leq \gamma < m/(m-1)$. By similar principles, the probabilities in (35) should be approximately

$$\mathbb{P}_{\mathcal{W}_n}(\mathfrak{p} \text{ does not occur}) \approx \left(1 + \frac{1}{\gamma m^k}\right)^{-n} \approx e^{-n/(\gamma m^k)}.$$

For a binary alphabet, this tells us that the occurrence of a pattern of length k starts becoming likely when n is of the order of 2^k, that is, when k is of the order of $\log_2 n$. The more precise moment when this happens must depend (via γ) on the autocorrelation of the pattern, with strongly correlated patterns having a tendency to occur a little late. (This vastly generalizes our empirical observations of Chapter I.) However, observe that the mean number of occurrences of a pattern in a text of length n does not depend on the shape of the pattern. This apparent paradox is easily resolved: correlated patterns tend to occur late, but they lend themselves to appearing in clusters. Thus, the late pattern aaa when it occurs still has probability $\frac{1}{2}$ to occur at the next position as well, and cash in another occurrence, whereas no such possibility is available to the early pattern aab whose occurrences must be somewhat spread out.

Such analyses are important as they can be used to develop a precise understanding of the behaviour of data compression algorithms (the Lempel–Ziv scheme); see Julien Fayolle's memoir (Master Thesis, Paris, 2002) for details.

▷ **29.** *Multiple pattern occurrences.* A similar analysis applies to the generating function $S^{\langle s\rangle}(z)$ of words containing a fixed number s of occurrences of a pattern $\mathfrak{p}$. The OGF is obtained by expanding (with respect to u) the BGF $W(z,u)$ obtained in Chapter III by means of an inclusion-exclusion argument. For $s \geq 1$, one finds

$$S^{\langle s\rangle}(z) = z^k \frac{N(z)^{s-1}}{D(z)^{s+1}}, \qquad D(z) = z^k + (1-mz)c(z), \quad N(z) = z^k + (1-mz)(c(z)-1)),$$

which now has a pole of multiplicity $s+1$ at $z=\rho$. ◁

▷ **30.** *Patterns in Bernoulli sequences.* Similar results hold when letters are assigned nonuniform probabilities, $p_j = \mathbb{P}(a_j)$, for $a_j \in \mathcal{A}$. One only needs to define the weighted autocorrelation polynomial by its coefficient c_j being $c_j = \mathbb{P}(\mathfrak{p}_1 \cdots \mathfrak{p}_j)$, when $\mathfrak{p}$ coincides with its jth shifted version. Multiple pattern occurrences can be also analysed. ◁

IV. 6. Singularities and functional equations

In the various combinatorial examples discussed so far in this chapter, we have been dealing with functions that are given by explicit expressions. Such situations essentially cover nonrecursive structures as well as the simplest recursive structures, like Catalan or Motzkin trees, whose generating functions are expressible in terms of radicals. In fact, as will shall see extensively in this book, complex analytic methods are instrumental in analysing coefficients of functions *implicitly* specified by functional equations. In other words: *the very nature of a functional equation can often provide clues regarding the singularities of its solution*. Chapter V will illustrate this philosophy in the case of rational functions defined by systems of positive equations; a very large number of examples will then be given in Chapters VI and VII, where singularities much more general than mere poles are treated. The purpose of this subsection is simply to offer a preliminary discussion of the way dominant singularities can be located in many cases by means means of simple iteration or inversion properties of analytic functions. Three typical functional equations are to be discussed here:

$$f(z) = ze^{f(z)}, \qquad f(z) = z + f(z^2+z^3), \qquad f(z) = \frac{1}{1-zf(z^2)}.$$

Inverse functions. We start with a generic problem: given a function ψ analytic at a point y_0 with $z_0 = \psi(y_0)$ what can be said about its inverse, namely the solution(s) to the equation $\psi(y) = z$ when z is near z_0 and y near y_0? Two cases occur depending on the value of $\psi'(y_0)$.

Regular case. If $\psi'(y_0) \neq 0$, then ψ admits an analytic expansion near y_0:

$$\psi(y) = \psi(y_0) + (y-y_0)\psi'(y_0) + \frac{1}{2}(y-y_0)^2\psi''(y_0) + \cdots .$$

Solving formally for y indicates a *locally linear* dependency,

$$y - y_0 \sim \frac{1}{\psi'(y_0)}(z-z_0). \tag{36}$$

A full formal expansion of $y-y_0$ in powers of $z-z_0$ is obtained by repeated substitution,

$$y - y_0 = c_1(z-z_0) + c_2(z-z_0)^2 + \cdots \tag{37}$$

and the method of majorizing series shows that the series so obtained converges locally in a sufficiently small neighbourhood of z_0. Rouché's theorem (equivalently, the analytic version of the Implicit Function Theorem, see Note 28), implies that the equation $\psi(y) = z$ admits there a unique analytic solution. In summary, *an analytic function locally admits an analytic inverse near any point where its first derivative is nonzero.*

Singular case. If to the contrary one has $\psi'(y_0) = 0$ and $\psi''(y_0) \neq 0$, then the expansion of ψ is of the form

$$\psi(y) = \psi(y_0) + \frac{1}{2}(y - y_0)^2 \psi''(y_0) + \cdots . \tag{38}$$

Solving formally for y now indicates a *locally quadratic* dependency

$$(y - y_0)^2 \sim \frac{2}{\psi''(y_0)}(z - z_0),$$

and the inversion problem admits two solutions satisfying

$$y - y_0 \sim \pm\sqrt{\frac{2}{\psi''(y_0)}}(z - z_0)^{1/2}.$$

The point z_0 is thus a branch point.

A similar reasoning applies whenever the first nonzero derivative of ψ at y_0 is of order $r \geq 2$ (with a local behaviour for y then of the form $(z - z_0)^{1/r}$). Thus, the dependency between y and z cannot be analytic around (y_0, z_0). In other words, *an analytic function is not locally invertible in an analytic manner in the vicinity of any point where its first derivative is zero.*

We can now consider the problem of obtaining information on the coefficients of a function $y(z)$ defined by an implicit equation

$$y(z) = z\phi(y(z)). \tag{39}$$

For simplicity, we shall momentarily assume $\phi(u)$ to be a nonlinear entire function (possibly a polynomial of degree ≥ 2) with nonnegative coefficients. In order for the problem to be (formally) well-posed we assume that $\phi(0) \neq 0$.

The equation (39) occurs in the counting of various types of trees. For instance, $\phi(u) = e^u$ corresponds to labelled Cayley trees, $\phi(u) = (1 + u)^2$ to binary trees, and $\phi(u) = 1 + u + u^2$ to plane unary–binary trees (Motzkin trees). A full analysis of the problem was developed by Meir and Moon [**85**], themselves elaborating on earlier ideas of Pólya [**97, 98**] and Otter [**92**].

Equation (39) may be rephrased as

$$\psi(y(z)) = z \qquad \text{where} \qquad \psi(u) = \frac{u}{\phi(u)}, \tag{40}$$

so that it is a generic instance of the inversion problem for analytic functions: $y = \psi^{(-1)}$. We first observe that (39) and (40) admit unique formal power series solutions by the method of indeterminate coefficients. An application of the technique of majorizing series shows that this formal solution also represents an analytic function near the origin, with $y(0) = 0$. In addition, the coefficients of $y(z)$ are all nonnegative.

Now comes the hunt for singularities. The function $y(z)$ increases along the positive real axis. The equation $\psi'(\tau) = 0$ which is expected to create singularities for $y(z)$ is in terms of ϕ:

$$\phi(\tau) - \tau\phi'(\tau) = 0. \tag{41}$$

The function $\phi(u) = \sum_{k=0}^{\infty} \phi_k u^k$ being by assumption entire, the equation (41) is equivalent to

$$\phi_0 = \phi_2 \tau^2 + 2\phi_3 \tau^3 + \cdots ,$$

which admits a unique positive solution.

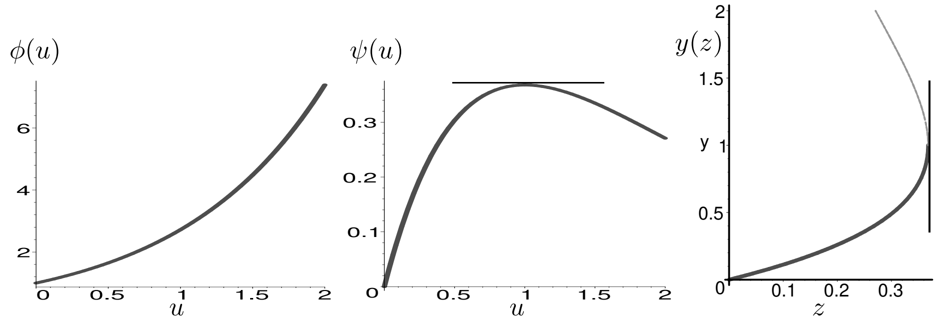


FIGURE 13. Singularities of inverse functions: $\phi(u) = e^u$ (left); $\psi(u) = u/\phi(u)$ (middle); $y = \mathrm{Inv}(\psi)$ (right).

As z increases, starting from 0 along the positive real axis, $y(z)$ increases. Let $\rho \leq \infty$ be the dominant positive singularity of $y(z)$. We are going to prove *a contrario* that $y(\rho) = \tau$ (technically, we should define $y(\rho)$ as the limit of $y(x)$ as $x \to \rho^-$). Assume that $y(\rho) < \tau$; then $y(z)$ could be analytically continued at $z = \rho$, by the discussion above of inverse functions in the regular case, since $\phi'(y(\rho)) > 0$. If on the other hand, we had $y(\rho) > \tau$, then, there would be a value $\rho^* < \rho$ such that $y(\rho^*) = \tau$; but there, we would have $\psi'(y(\rho^*)) = 0$, so that $y(z)$ would be singular at $z = \rho^*$ by the discussion on inverse functions in the singular case. Thus, in both cases, the assumption $y(\rho) \neq \tau$ leads to a contradiction. We thus obtain that $y(\rho) = \tau$, and, since y and ψ are inverse functions, this corresponds to

$$\rho = \psi(\tau) = \tau/\phi(\tau).$$

Equipped with this discussion, we state a result which covers situations more general than the case of ϕ being entire.

PROPOSITION IV.4. *Let ϕ be a nonlinear function that is analytic at 0, with nonnegative Taylor coefficients and radius of convergence $R \leq +\infty$. Assume that there exists $\tau \in (0, R)$ such that*

$$\frac{\tau\phi'(\tau)}{\phi(\tau)} = 1. \tag{42}$$

Let $y(z)$ be the solution analytic at the origin of the equation $y(z) = \phi(y(z))$. Then, one has the exponential growth formula:

$$[z^n]\, y(z) \bowtie \left(\frac{1}{\rho}\right)^n \qquad \textit{where} \quad \rho = \frac{\tau}{\phi(\tau)} = \frac{1}{\phi'(\tau)}.$$

Note that, by Supplement 31 below, there can be *at most* one solution of the characteristic equation (42) in $(0, R)$, a necessary and sufficient condition for the existence of a solution in the open interval $(0, R)$ being $\lim_{x \to R^-} \frac{x\phi'(x)}{\phi(x)} > 1$. This last condition is automatically granted as soon as $\phi(R) = +\infty$.

PROOF. The discussion above applies verbatim. The function $y(z)$ is analytic around 0 (by majorizing series techniques). By the already seen argument, its value $y(\rho)$ cannot be different from τ, so that its radius of convergence must equal ρ. The form of y_n then results from general exponential bounds. $\square$

$\triangleright$ **31.** *Convexity of GFs and the Variance Lemma.* Let $\phi(z)$ be a nonlinear GF with nonnegative coefficients and a nonzero radius of convergence R. For $x \in (0, R)$ a parameter, define the *Boltzmann* *random variable* Ξ (of parameter x) by the property

$$\mathbb{P}(\Xi = n) = \frac{\phi_n x^n}{\phi(x)}, \qquad \text{with} \quad \mathbb{E}(s^{\Xi}) = \frac{\phi(sx)}{\phi(x)}$$

the probability generating function of Ξ. By differentiation, the first two moments of Ξ are

$$\mathbb{E}(\Xi) = \frac{x\phi'(x)}{\phi(x)}, \qquad \mathbb{E}(\Xi^2) = \frac{x^2\phi''(x)}{\phi(x)} + \frac{x\phi'(x)}{\phi(x)}.$$

There results, for any nonlinear GF $\phi(x)$, the general convexity inequality

$$\frac{d}{dx}\left(\frac{x\phi'(x)}{\phi(x)}\right) > 0,$$

since the variance of a nondegenerate random variable is always positive. Equivalently, the function $\log(\phi(e^t))$ is convex for $t \in (-\infty, \log R)$. $\lhd$

Take for instance general Catalan trees corresponding to

$$y = \frac{z}{1 - y(z)}, \qquad \text{so that} \quad \phi(u) = \frac{1}{1-u}.$$

We have $R = 1$ and the characteristic equation reads

$$\frac{\tau}{1-\tau} = 1,$$

implying $\tau = \frac{1}{2}$, so that $\rho = \frac{1}{4}$. We obtain as anticipated $y_n \bowtie 4^n$, a weak asymptotic formula for the Catalan numbers. Similarly, for Cayley trees, we have $\phi(u) = e^u$, the characteristic equation reduces to $(\tau - 1)e^{\tau} = 0$, so that $\tau = 1$ and $\rho = e^{-1}$, giving a weak form of Stirling's formula:

$$[z^n]y(z) = \frac{n^{n-1}}{n!} \bowtie e^n.$$

Here is a table of a few cases of application of the method to structures already encountered in previous chapters.

Type	$\phi(u)$	(R)	τ, ρ	$y_n \bowtie \rho^{-n}$
gen. Catalan tree	$\dfrac{1}{1-u}$	(1)	$\frac{1}{2}, \frac{1}{4}$	$y_n \bowtie 4^n$
binary tree	$(1+u)^2$	(∞)	$1, \frac{1}{4}$	$y_n \bowtie 4^n$
Motzkin tree	$1 + u + u^2$	(∞)	$1, \frac{1}{3}$	$y_n \bowtie 3^n$
Cayley tree	e^u	(∞)	$1, e^{-1}$	$y_n \bowtie e^n$

In fact, for all such problems, the dominant singularity is always of the square-root type as our previous discussion suggests. Accordingly, the asymptotic form of coefficients is invariably of the type

$$[z^n]\, y(z) \sim C \cdot \rho^{-n} n^{-3/2},$$

as we shall prove in Chapter VI by means of the singularity analysis method.

$\rhd$ **32.** *A variant form of the inversion problem.* Consider the equation $y = z + \phi(y)$, where ϕ is assumed to be entire and $\phi(u) = O(u^2)$ at $u = 0$. This corresponds to a simple variety of trees in which trees are counted by the number of their leaves only. For instance, we have already encountered labelled hierarchies (phylogenetic trees) in Section II.6 corresponding to $\phi(u) = e^u - 1 - u$, which is one of "Schröder's problems". Let $\widetilde{\tau}$ be the root of $\phi'(\widetilde{\tau}) = 1$ and set $\widetilde{\rho} = \widetilde{\tau} - \phi(\widetilde{\tau})$. Then $[z^n]y(z) \bowtie \rho^{-n}$. For the EGF L of labelled hierarchies ($L = z + e^L - 1 - L$), this gives $L_n/n! \bowtie (2\log 2 - 1)^{-n}$. (Observe that Lagrange inversion also provides $[z^n]y(z) = \frac{1}{n}[w^{n-1}](1 - y^{-1}\phi(y))^{-n}$.) $\lhd$

Iteration. Consider the class $\mathcal{E}$ of *balanced 2–3 trees* defined as trees whose node degrees are restricted to the set $\{0, 2, 3\}$, with the additional property that all leaves are at the same distance from the root. Such tree trees, which are particular cases of B-trees, are a useful data structure for implementing dynamic dictionaries [**75**]. We adopt as notion of size the number of leaves (also called external nodes). The OGF of $\mathcal{E}$ satisfies the functional equation

$$E(z) = z + E(z^2 + z^3), \tag{43}$$

which reflects an inductive definition involving a substitution: given an existing tree, a new tree is obtained by substituting in all possible ways to each external node ($\Box$) either a pair $(\Box, \Box)$ or a triple $(\Box, \Box, \Box)$. On other words, we have

$$\mathcal{E}[\Box] = \Box + \mathcal{E}\left[\Box \to (\Box\Box + \Box\Box\Box)\right].$$

Equation (43) implies the seemingly innocuous recurrence

$$E_n = \sum_{k=0}^{n} \binom{k}{n-2k} E_k \qquad \text{with} \quad E_0 = 0,\ E_1 = 1,$$

but no closed-form solution is known (nor likely to exist) for E_n or $E(z)$. The expansion starts as (the coefficients are *EIS* **A014535**)

$$E(z) = z + z^2 + z^3 + z^4 + 2\,z^5 + 2\,z^6 + 3\,z^7 + 4\,z^8 + 5\,z^9 + 8\,z^{10} + \cdots .$$

We present here the first stage of an analysis due to Odlyzko [**88**] and corresponding to exponential bounds. Let $\sigma(z) = z^2 + z^3$. Equation (43) can be expanded by iteration in the ring of formal power series,

$$E(z) = z + \sigma(z) + \sigma^{[2]}(z) + \sigma^{[3]}(z) + \cdots , \tag{44}$$

where $\sigma^{[j]}(z)$ denotes the jth iterate of the polynomial σ:

$$\sigma^{[0]}(z) = z, \qquad \sigma^{[h+1]}(z) = \sigma^{[h]}(\sigma(z)) = \sigma(\sigma^{[h]}(z)).$$

Thus, $E(z)$ is nothing but the sum of all iterates of σ. The problem is to determine the radius of convergence of $E(z)$, and by Pringsheim's theorem, the quest for dominant singularities can be limited to the positive real line.

For $z > 0$, the polynomial $\sigma(z)$ has a unique fixed point, $\rho = \sigma(\rho)$, at

$$\rho = \frac{1}{\varphi} \qquad \text{where} \qquad \varphi = \frac{1+\sqrt{5}}{2},$$

the golden ratio. Also, for any positive x satisfying $x < \rho$, the iterates $\sigma^{[j]}(x)$ must converge to 0; see Fig. 14. Furthermore, since $\sigma(z) \sim z^2$ near 0, these iterates converge to 0 doubly exponentially fast. First, for $x \in [0, \frac{1}{2}]$, one has $\sigma(x) \le \frac{3}{2}x^2$ for $x \in [0, \frac{1}{2}]$, so that there

$$\sigma^{[j]}(x) \le \left(\frac{3}{2}\right)^{2^j - 1} x^{2^j}. \tag{45}$$

Second, for $x \in [0, A]$, where A is any number $< \rho$, there is a number k_A such that $\sigma^{[k_A]}(x) < \frac{1}{2}$, so that, by (45), there holds:

$$\sigma^{[k]}(x) \le \frac{3}{2}\left(\frac{3}{4}\right)^{2^{k-k_A}}.$$

Thus, the series of iterates of σ is quadratically convergent for $z \in [0, A]$, any $A < \rho$.

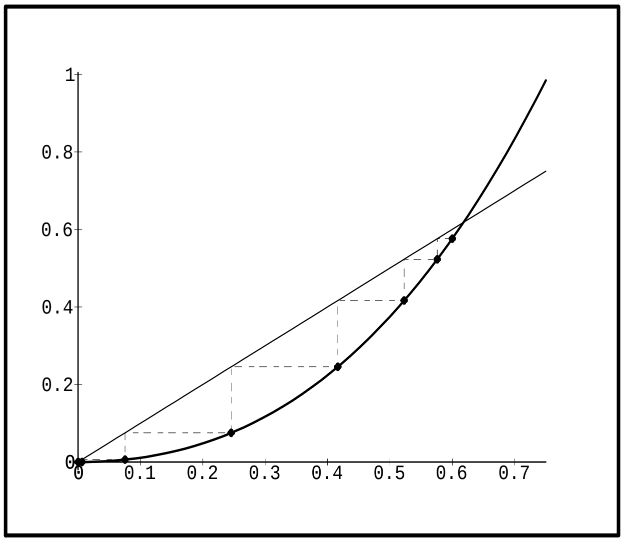


FIGURE 14. The iterates of a point $x_0 \in [0, \frac{1}{\varphi}[$ (here $x_0 = 0.6$) by $\sigma(z) = z^2 + z^3$ converge fast to 0.

By the triangular inequality, $|\sigma(z)| \le (\sigma(|z|)$, the sum in (44) is a normally converging sum of analytic functions, and is thus itself analytic. Consequently $E(z)$ is analytic in the whole of the open disk $|z| < \rho$.

It remains to prove that the radius of convergence of $E(z)$ is exactly equal to ρ. To that purpose it suffices to observe that $E(z)$, as given by (44), satisfies

$$E(x) \to +\infty \qquad \text{as} \qquad x \to \rho^-.$$

Let N be an arbitrarily large but fixed integer. It is possible to select a positive x_N sufficiently close to ρ with $x_N < \rho$, such that the Nth iterate $\sigma^{[N]}(x_N)$ is larger than $\frac{1}{2}$ (the function $\sigma^{[N]}(x)$ admits ρ as a fixed point and it is continuous and increasing at ρ). Given the sum expression (44), this entails the lower bound $E(x_N) > \frac{N}{2}$ for such an $x_N < \rho$ so that $E(x)$ is unbounded as $x \to \rho^-$.

The dominant positive real singularity of $E(z)$ is thus $\rho = \frac{1}{\varphi}$, and application of Cauchy bounds shows that

$$[z^n]\, E(z) \bowtie \left(\frac{1+\sqrt{5}}{2}\right)^n. \tag{46}$$

It is notable that this estimate could be established so simply by a purely qualitative examination of the basic functional equation and of a fixed point of the associated iteration scheme.

The complete asymptotic analysis of the E_n was given by Odlyzko [**88**] in a classic paper. It requires the full power of singularity analysis methods to be developed in Chapter VI. Equation (47) below states the end result, which involves periodic fluctuations; see Figure 15 (right). There is overconvergence of the representation (44), that is, convergence in certain directions beyond the disc of convergence of $E(z)$, as illustrated by Figure 15 (left). The proof techniques involve an investigation of the behaviour of iterates of σ in the complex plane, an area launched by Fatou and Julia in the first half of the past century and nowadays well-studied under the name of "complex dynamics".

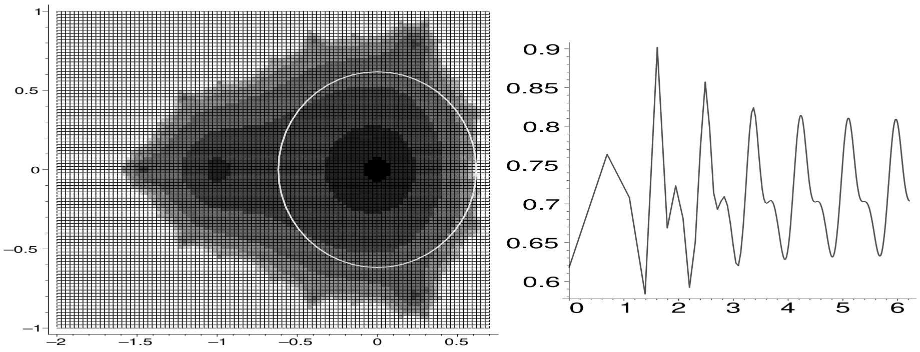


FIGURE 15. Left: the fractal domain of analyticity of $E(z)$ in gray with darker areas representing faster convergence of the sum of iterates of σ. Right: the ratio $E_n/(\varphi^n n^{-1})$ plotted against $\log n$ for $n = 1\,..\,500$ confirms that $E_n \bowtie \varphi^n$ and illustrates the periodic fluctuations expressed by Equation (47).

$\triangleright$ **33.** *The asymptotic number of 2–3 trees.* This analysis is from [**88, 89**]. The number of 2–tree trees satisfies asymptotically

$$E_n = \frac{\varphi^n}{n}\Omega(\log n) + O\left(\frac{\varphi^n}{n^2}\right), \tag{47}$$

where Ω is a periodic function with mean value $\varphi(\log(4-\varphi) \doteq 0.71208$ and period $\log(4-\phi) \doteq$. Thus oscillations are inherent in E_n. A plot of the ratio $E_n/(\phi^n/n)$ is offered in Figure 15. $\triangleleft$

Complete asymptotics of a functional equation. This is Pólya's counting of certain molecules, a case where only a functional equation is known for a generating function, $M(z) = \sum_n M_n z^n$:

$$M(z) = \frac{1}{1 - zM(z^2)}. \tag{48}$$

The M_n represent the number of chemical isomeres of alcohols $C_nH_{2n+1}OH$ without asymmetric carbon atoms, and the series starts as (*EIS* **A000621**)

$$M(z) = 1 + z + z^2 + 2z^3 + 3z^4 + 5z^5 + 8z^6 + 14z^7 + 23z^8 + 39z^9 + \cdots . \tag{49}$$

By iteration of the functional equation, one finds a continued fraction representation:

$$M(z) = \cfrac{1}{1 - \cfrac{z}{1 - \cfrac{z^2}{1 - \cfrac{z^4}{\ddots}}}}.$$

Pólya [**98**] who established this functional equation in the historical paper that introduced "Pólya Theory" developed at the same time a precise asymptotic estimate for M_n.

PROPOSITION IV.5. *Let $M(z)$ be the solution analytic around 0 of the functional equation*

$$M(z) = \frac{1}{1 - zM(z^2)}.$$

Then, there exist constants K and α such that

$$M_n \sim K \cdot \alpha^n, \qquad \alpha \doteq 1.68136\,75244, \quad K \doteq 0.36071\,40971.$$

PROOF. We offer two proofs. The first one is based on direct consideration of the functional equation and is of a high degree of applicability. The second one, following Pólya, makes explicit a linear structure present in the problem and leads to more explicit results.

First proof. The first few coefficients of M are determined by the functional equation and known (49). Then, by positivity of the functional equation, $M(z)$ dominates coefficientwise any GF $(1 - zM_{<m}(z^2)^{-1}$, where $M_{<m}(z)$ is the mth truncation of $M(z)$. In particular, one has the domination relation (use $M_{<2}(z) = 1 + z$)

$$M(z) \preceq \frac{1}{1 - z - z^3}.$$

Since the rational fraction has a dominant poles at $z \doteq 0.68232$, this implies that the radius ρ of convergence of $M(z)$ satisfies $\rho < 0.69 < 1$. In the other direction, since $M(z^2) < M(z)$ for $z \in (0, \rho)$, then, one has the numerical inequality

$$M(z) \leq \frac{1}{1 - zM(z)}, \qquad 0 \leq z < \rho.$$

This can be used to show that the Catalan generating function $C(z) = (1 - \sqrt{1 - 4z})/(2z)$ is a majorant of $M(z)$ on the interval $(0, \frac{1}{4})$ and that $M(z)$ exists for $z \in (0, \frac{1}{4})$. In other words, one has $\frac{1}{4} \leq \rho < 0.69$. At any rate, the radius of convergence of M is strictly between 0 and 1.

▷ **34.** *Alcohols and trees.* Since $M(z)$ starts as $1+z+z^2+\cdots$ while $C(z)$ starts as $1+z+2z^2+\cdots$, there is a small interval $(0, \epsilon)$ such that $M(z) \leq C(z)$. By the functional equation of $M(z)$, one has $M(z) \leq C(z)$ for z in the larger interval $(0, \sqrt{\epsilon})$. One can then bootstrap and show that $M(z) \leq C(z)$ for $z \in (0, \frac{1}{4})$. ◁

Next, as $z \to \rho^-$, one must have $zM(z^2) \to 1$. Indeed, if this was not the case, we would have $zM(z^2) < A < 1$ for some A. But then, since $\rho^2 < \rho$, the quantity $(1 - zM(z^2))^{-1}$ would be analytic at $z = \rho$, a clear contradiction. Thus, ρ is determined implicitly by the equation

$$\rho M(\rho^2) = 1,$$

and by monotonicity, there can be only one such solution. Numerically, one can estimate ρ as the limit of quantities ρ_m satisfying

$$\sum_{n=0}^{m} M_n \rho_m^{2n+1} = 1,$$

together with $\rho \in [\frac{1}{4}, 0.069]$. In each case, only a few of the M_n are needed. One obtains in this way:

$$\rho_{10} \doteq 0.595,\ \rho_{20} \doteq 0.594756,\ \rho_{30} \doteq 0.59475397,\ \rho_{40} \doteq 0.594753964,$$

and it is not hard to verify that this provides a geometrically convergent scheme to the limit $\rho \doteq 0.59475\,39639$. (Note: Pólya determined ρ to five decimals by hand!)

The previous discussion also implies that ρ is a pole, which must be simple. Thus

$$(50) \qquad M(z) \sim z \to \rho K \frac{1}{1 - z/\rho}, \qquad K := \frac{1}{\rho M(\rho^2) + 2\rho^3 M'(\rho^2)}.$$

The argument shows at the same time that $M(z)$ is meromorphic in $|z| < \sqrt{\rho} \doteq 0.77$. That $M(z)$ is a the only pole on $|z| = \rho$ can be seen from the fact that $zM(z^2) = z + z^3 + \cdots$ is unperiodic in the sense of Chapter V. (We don't detail the argument here as the property is also implied by the developments of the second proof.) The translation of the singular expansion (50) yields the statement.

Second proof. First, a sequence of formal approximants follows from (48) starting with

$$1,\ \frac{1}{1-z},\ \frac{1}{1-\dfrac{z}{1-z^2}} = \frac{1-z^2}{1-z-z^2},\ \frac{1}{1-\dfrac{z}{1-\dfrac{z^2}{1-z^4}}} = \frac{1-z^2-z^4}{1-z-z^2-z^4+z^5}.$$

which permits to compute any number of terms of the series $M(z)$. Closer examination of (48) suggests to set

$$M(z) = \frac{\psi(z^2)}{\psi(z)},$$

where

$$\psi(z) = 1 - z - z^2 - z^4 + z^5 - z^8 + z^9 + z^{10} + z^{17} + z^{18} + z^{20} - z^{21} - z^{37} - \cdots$$

Back substitution into (48) yields

$$\frac{\psi(z^2)}{\psi(z)} = \frac{1}{1 - z\dfrac{\psi(z^4)}{\psi(z^2)}} \text{ or } \frac{\psi(z^2)}{\psi(z)} = \frac{\psi(z^2)}{\psi(z^2) - z\psi(z^4)},$$

which shows $\psi(z)$ to be a solution of the functional equation

$$\psi(z) = \psi(z^2) - z\psi(z^4).$$

The coefficients of ψ are all in the set $\{0, -1, +1\}$, as they satisfy the recurrence

$$\psi_{4n} = \psi_{2n},\ \psi_{4n+1} = -\psi_n,\ \psi_{4n+2} = \psi_{2n+1},\ \psi_{4n+3} = 0.$$

Thus, $M(z)$ appears as the quotient of two function, $\psi(z^2)/\psi(z)$; since $\psi(z)$ whose coefficients are bounded by 1 in absolute value, it is analytic in the unit disk, $M(z)$ is itself meromorphic in the unit disc. A numerical plot shows that that $\psi(z)$ has its smallest positive real zero at $\rho \doteq 0.59475$, which is a simple zero of $\psi(z)$ and thus a pole of $M(z)$ as $\psi(\rho^2) \neq 0$. Thus

$$M(z) \sim \frac{\psi(\rho^2)}{(z-\rho)\psi'(\rho)} \implies M_n \sim -\frac{\psi(\rho^2)}{\rho\psi'(\rho)}\left(\frac{1}{\rho}\right)^n.$$

Numerical computations then yield Pólya's estimate. Et voilà! □

The example of Pólya's alcohols is exemplary, both from a historical point of view and from a methodological perspective. It demonstrates that quite a lot of information can be pulled out of a functional equation without solving it. (A very similar situation will be discussed in Chapter V, see the enumeration of coin fountains.) In passing, we have made great use of the fact that if $f(z)$ is analytic in $|z| < r$ and some bounds imply the strict inequalities $0 < r < 1$, then one can regard functions like $f(z^2)$, $f(z^3)$, and so on, as "known" since they are analytic in the disc of convergence of f and even beyond, a situation evocative of our earlier discussion of Pólya operators in Subsection IV. 3.3. Globally, the lesson is that functional equation, even very complicated ones, can often be used to bootstrap the local singular behaviour of solutions and one can do so despite the absence of any explicit solution. Then, the transition from singularities to coefficient asymptotics is a simple jump.

▷ **35.** *An arithmetic exercise* Find a characterization of $\psi_n = [z^n]\psi(z)$ based on the binary representation of n. Tabulate ψ_n for all $n \in (10^{1000}, 10^{1000} + 10^{500})$, possibly using some compressed format. Find the asymptotic proportion of the ψ_n for $n \in [1\,..\,N]$ that are nonzero. ◁

IV. 7. Notes

This chapter has been designed to serve as a refresher of basic complex analysis, with special emphasis on methods relevant for analytic combinatorics. References most useful for the discussion given in this chapter include the books of Titchmarsh [**109**] (oriented towards classical analysis), Whittaker and Watson [**114**] (stressing special functions), Dieudonné [**28**] and Knopp [**72**]. Henrici [**66**] presents complex analysis under the perspective of constructive and numerical methods, a highly valuable point of view for this book. References dealing specifically with asymptotic analysis are discussed at the end of the next chapter.

As demonstrated by the first batch of examples sprinkled over this chapter, singularities provide a royal road to coefficient asymptotics. In this regard, the two main statements of this chapter are are the theorems relative to the expansion of rational and meromorphic functions, Theorems IV.6 and IV.7. They are of course extremely classical (and easy) results. Issai Schur (1875–1941) is to be counted amongst the very first mathematicians who recognized the rôle of analytic methods in combinatorial enumerations (Example 4). This thread was developed by George Pólya in his famous paper of 1937 (see [**97, 98**]), which Read in [**98**, p. 96] describes as a "landmark in the history of combinatorial analysis". There, Pólya founded at the same time combinatorial chemistry, the enumeration of objects under group actions, and the complex-asymptotic theory of graphs and trees.

De Bruijn's classic booklet [**25**] is a wonderfully concrete introduction to effective asymptotic theory, and it contains many examples from discrete mathematics thoroughly worked out. The state of affairs in 1995 regarding analytic methods in combinatorial enumeration is superbly summarized in Odlyzko's scholarly chapter [**89**]. Wilf devotes his Chapter 5 of *Generatingfunctionoloy* [**116**] to this question. The books by Hofri [**68**] and Szpankowski [**108**] contain useful accounts in the perspective of analysis of algorithms. See also our book [**100**] for a light introduction and the chapter by Vitter and Flajolet [**112**] for more on this topic.

Paraphrasing the number theorist Hecke, we may feel confident in stating: *A function's singularities contain a wealth of asymptotic information on the function's coefficients; a generating function contains a wealth of information on the corresponding combinatorial structures*. This philosophy furthermore unites analytic combinatorics and analytic number theory. It is the purpose of the next four chapters to illustrate it thoroughly by means of a great variety of combinatorial examples.

CHAPTER V

Applications of Rational and Meromorphic Asymptotics

Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision.

— ANDREW ODLYZKO [**89**]

The primary goal of this chapter is to provide combinatorial illustrations of the power of complex analytic methods, and specifically of the rational–meromorphic framework. At the same time, we shift gears and envisage counting problems at a new level of generality. Precisely, we consider combinatorial-analytic *schemas*, which, broadly speaking, are wide *families* of combinatorial types amenable to a common analytic framework and associated with a common collection of asymptotic properties.

The first schema comprises regular specifications and languages, which *a priori* leads to rational generating functions and thus systematically resort to Theorem IV.6. This is not the end of the story, however, since in general one is interested not just in a single set of combinatorial objects, but rather in a whole family of classes. The case of patterns in words at the end of the previous chapter has already exemplified this situation. Here, we extend the analysis to the determination of longest runs, corresponding to maximal sequences of good (or bad) luck in games of chance. In so doing, we develop analytical methods that apply in many cases to largest components. We then consider an important class of regular specifications, the ones that are built on nested sequences and combinatorially correspond to lattice paths. Besides providing a precise quantification of height in Dyck paths, this also leads to the determination of height in random (general) Catalan trees. The treatment is to a large extent made possible because nested sequence constructions lead naturally to nested quasi-inverses, that is, continued fractions. And continued fractions enjoy a wealth of algebraic and analytic properties.

Next, we discuss a general schema of analytic combinatorics known as the *supercritical sequence* schema, which provides a prime illustration of the power of meromorphic asymptotics while being of a very wide applicability. For instance, one can predict very precisely (and easily) the number of ways that an integer can be decomposed additively as a sum of primes (or twin primes), this even though many details of the distribution of primes are still surrounded in mystery.

Last we discuss positive linear systems of generating functions: although the resulting generating functions are once more bound to be rational, there is benefit in examining them as defined implicitly (rather than solving explicitly) and work out singularities directly. The crucial technical tool there is the Perron-Frobenus theory of nonnegative matrices, whose importance has been long recognized in the theory of finite Markov chains. A general discussion of singularities can then be conducted, leading to valuable consequences on a variety of models—paths in graphs, finite automata, and transfer matrices.

All these cases illustrate the power of rational and meromorphic asymptotics. The last example discussed treats locally constrained permutations, where rational functions even provide an entry to the world of permutations.

Universality is a term originating with statistical physics that is also nowadays increasingly used in probability theory. By universality is meant a collection of key properties that are shared by a wide family of models and are largely independent of particulars of each models. For instance, in statistical physics, random placements of pieces or random walks on a regular lattice share common properties that do not depend on the particular geometry of the lattice, whether square, triangular, or honeycomb. In probability theory, it is established that sums of random variable converge to a Gaussian limit, so that the Gaussian law is universal for sums of random variables (under suitably mild moment conditions). In this spirit, we can describe the supercritical sequence as universal accross combinatorics as it covers a large family of models simply characterized by the presence of an external sequence construction ($\mathcal{F} = \mathfrak{S}(\mathcal{G})$) accompanied with a natural analytic assumption ("supercriticality"). Alignments, compositions, and surjections for instance find themselves sheltered under a common umbrella and analytic theory tells us that they *must* share many features, like having a linear number of components in the mean and with high probability, an asymptotically predetermined proportion of components of each possible type, and so on. In a similar spirit, one can regard exponential-polynomial behaviour as universal across all problems described by regular expressions (Sections V. 1 and V. 2) or by finite state models (Section V. 4 and V. 5).

V. 1. Regular specification and languages

A combinatorial specification is said to be *regular* if it is nonrecursive ("iterative") and it involves only the constructions of Atom, Union, Product, and Sequence; see Chapter I. For convenience and without loss of analytic generality, we consider here unlabelled structures. Since the operators translating these constructions into generating functions are all of a rational nature, it follows that the corresponding OGFs are invariably rational. Then Theorem IV.6 applies directly:

THEOREM V.1 (Regular specification asymptotics). *Let $\mathcal{C}$ be an unlabelled class that is described by a regular specification. Then the coefficients of the OGF $C(z)$ satisfy an exponential-polynomial formula,*

$$C_n \equiv [z^n]C(z) = \sum_{j=1}^{m} \Pi_j(n)\alpha_j^{-n}, \tag{1}$$

for a family of algebraic numbers α_j and a family of polynomials Π_j.

General trees of bounded height, denumerants, as well as partitions and compositions into summands at most r constitute prime examples of structures admitting regular specifications.

The name "regular specification" has been chosen so as to be in agreement with the notions of regular expression and regular language from formal language theory introduced in Chapter I. We saw there that a language is called S–regular ("specification regular") if it is combinatorially isomorphic with a class $\mathcal{R}$ which admits a regular specification.

The most frequent case is that of a language specified by a *regular expression*, involving letters of the alphabet, union, catenation, and Kleene star. If the regular expression is unambiguous, i.e., every word is uniquely parsable (see APPENDIX: *Regular expressions*, p. 125), it is combinatorially isomorphic to a regular specification. In the general case, one may encounter regular expressions that are ambiguous; then, the systematic application of the translation rules amounts to counting every word with its multiplicity, that is, the number of ways in which it can be parsed.

PROPOSITION V.1 (Regular expression counting). *Given a regular expression R (assumed to be of finite ambiguity), the ordinary generating function $L_R(z)$ of the language $\mathcal{L}(R)$, counting* with multiplicity, *is given by the inductive rules:*

$$\epsilon \mapsto 1, \quad a \mapsto z, \quad \cup \mapsto +, \quad \cdot \mapsto \times, \quad \star \mapsto (1-(.))^{-1}.$$

In particular, if R is unambiguous, then the ordinary generating function satisfies $L_R(z) = L(z)$ and is given directly by the rules above. In both cases, the coefficients $[z^n]L_R(z)$ admit of an exponential-polynomial form.

Note. If R is ambiguous, it is known that one can build an unambigous R' such that $\mathcal{L}(R) = \mathcal{L}(R')$. Consequently, the conclusions of Proposition V.1 extend in principle to counting without multiplicities words in *any* regular language. One then has however to rely on an indirect automaton construction (see the appendices) which computational complexity is in general exponential.

PROOF. Formal rules associate to any proper regular expression R a specification $\mathcal{R}$:

$$\begin{aligned} &\epsilon \mapsto 1 \text{ (the empty object)}, \quad a \mapsto Z_a \ (Z_a \text{ an atom}),\\ &\cup \mapsto +, \cdot \mapsto \times, \quad \star \mapsto \mathfrak{S}\{.\} \end{aligned}$$

It is readily recognized that this mapping is such that $\mathcal{R}$ generates exactly the collection of all parsings of words according to R. The translation rules of Chapter 1 then yield the first part of the statement. The second part follows since $L(z) = L_R(z)$ whenever R is unambiguous. □

EXAMPLE 1. *A potpourri of regular specifications.* We briefly recapitulate here a number of combinatorial problems already encountered in Chapters I–III that are reducible to regular specifications.

Compositions of integers (Section I.3) are specified by $\mathcal{C} = \mathfrak{S}(\mathfrak{S}_{\geq 1}(\mathcal{Z}))$, whence the OGF $(1-z)/(1-2z)$ and the closed form $C_n = 2^{n-1}$, an especially trivial exponential-polynomial form. Polar singularities are also present for compositions into k summands ($\mathfrak{S}_k(\mathfrak{S}_{\geq 1}(\mathcal{Z}))$ and for compositions whose summands are restricted to the interval $[1, r]$ ($\mathfrak{S}(\mathfrak{S}_{1\,..\,r}(\mathcal{Z}))$, with corresponding generating functions

$$\frac{z^k}{(1-z)^k}, \qquad \frac{1-z}{1-2z+z^{r+2}}.$$

In the first case, one has an explicit form for the coefficients, $\binom{n-1}{k-1}$, which is also a particular exponential-polynomial form (with the basis of the exponential being 1). The second case requires a dedicated analysis of the dominant polar singularity, a task that is undertaken in Example 2 below for the closely related problem of determining longest runs in random binary words. We shall also see later (Section V. 3 and Example 9) that a rich class of summand-restricted compositions resorts to the framework of meromorphic asymptotics.

Integer partitions involve the multiset construction. However, when summands are restricted to the interval $[1\,..\,r]$, the specification satisfies the combinatorial identity (Section I.3),

$$\mathfrak{M}(\mathfrak{S}_{1\,..\,r}(\mathcal{Z})) \simeq \mathfrak{S}(\mathcal{Z}) \times \mathfrak{S}(\mathcal{Z}^{\times 2}) \times \cdots \mathfrak{S}(\mathcal{Z}^{\times r}),$$

corresponding to the OGF

$$\prod_{j=1}^{r} \frac{1}{1-z^j}.$$

This case has already served as a leading example in our discussion of denumerants in Example IV.4, where the analysis of the pole at 1 furnishes the dominant asymptotic behaviour ($n^{k-1}/(k!(k-1)!)$) of these special partitions.

Words lead to many problems that are prototypical of the regular specification framework. In Section 1.4, we saw that one could give a regular expression describing the set of words containing the pattern abb, from which the exact and asymptotic forms of counting coefficients derive. The case of long runs of a single letter is similarly amenable to to regular specifications and is detailed below. Note however that, for a general pattern $\mathfrak{p}$, the generating functions of words constrained to include (or dually exclude) $\mathfrak{p}$ are best based on the inclusion-exclusion argument of Section III.6.4. The corresponding asymptotic analysis has already served as a pilot example in Section IV. 5.3 of the previous chapter.

Words can also be analysed under the Bernoulli model, where letter i is selected with probability p_i; cf Section III.5 for a general discussion. We saw there that one can put regular specifications to good use in order to analyse the mean number of records in a random word.

Set partitions are typically labelled objects. However, when suitably constrained, they can sometimes be encoded by words described by regular expressions; see Section I.4.3 for partitions into k classes, where the OGF has been found to be

$$S^{(k)}(z) = \frac{z^k}{(1-z)(1-2z)\cdots(1-kz)} \qquad \text{implying} \qquad S_n^{(k)} \sim \frac{k^n}{k!},$$

where the asymptotic estimate results from the dominant pole at $1/k$.

Trees have generating functions that, in all nondegenerate cases, are beyond rational functions. However, the generating function of general (Catalan) trees of height $\leq h$ *is* rational; see Section III.7 relative to extremal parameters. The corresponding analysis is detailed below, Section V. 2 and Example 6, in relation to the enumeration of Dyck paths in a strip. $\square$

$\triangleright$ **1.** *Partially commutative monoids.* Let $\mathcal{W} = \mathcal{A}^\star$ be the set of all words over a finite alphabet $\mathcal{A}$ whose letters are also considered as formal indeterminates. Consider a set C of commutation rules between pairs of elements of A. For instance, if $\mathcal{A} = \{a, b, c\}$, then $\mathrm{C} = \{ab = ba, ac = ca\}$ means that a commutes with both b and c, but bc is not a commuting pair: $bc \neq cb$. Let $\mathcal{M} = \mathcal{W}/[\mathrm{C}]$ be the set of equivalent classes of words (monomials) under the rules induced by C. $\mathcal{M}$ is called a *partially commutative monoid* or a trace monoid.

If $A = \{a, b\}$, then the two possibilities for C are $\mathrm{C} = \emptyset$ and $\mathrm{C} := \{ab = ba\}$. Normal forms for $\mathcal{M}$ are given by the regular expressions $(a+b)^\star$ and $a^\star b^\star$ corresponding to the OGFs

$$\frac{1}{1-a-b}, \qquad \frac{1}{1-a-b+ab}.$$

If $\mathcal{A} = \{a, b, c\}$, the possibilities for C, the corresponding normal forms, and the OGFs M are as follows. If $\mathrm{C} = \emptyset$, then $\mathcal{M} \simeq (a+b)^\star$ with OGF $(1-a-b-c)^{-1}$; the other cases are

$$\begin{array}{ccc} ab = ba & ab = ba,\ ac = ca & ab = ba,\ ac = ca,\ bc = cb \\ (a^\star b^\star c)^\star a^\star b^\star & a^\star (b+c)^\star & a^\star b^\star c^\star \\ \dfrac{1}{1-a-b-c+ab} & \dfrac{1}{1-a-b-c+ab+ac+bc} & \dfrac{1}{1-a-b-c+ab+ac+bc-abc}. \end{array}$$

Cartier and Foata [**18**] have proved the general result (based on extended Moebius inversion),

$$M = \left(\sum_F (-1)^{|F|} F\right)^{-1},$$

in which the sum is over all monomials F formed with distinct letters that all commute pairwise. Goldwurm and Santini [**59**] have proved that $[z^n]M(z) \sim K \cdot \alpha^n$ for some $K, \alpha > 0$. $\lhd$

EXAMPLE 2. *Longest runs in words* The analysis of longest runs in words provides an illustration of the technique of localizing dominant singularities in rational functions and of the corresponding coefficient extraction process. In Chapter I, we have determined the family of OGFs describing the length L of the longest run of consecutive a's in a binary word over the alphabet $\mathcal{W} = \{a, b\}$. The counting GF associated with the property $(L < k)$ for a fixed k is a rational function. Determining the probability distribution of L over the set of all words of length n is then equivalent to analysing the *whole* family of GFs indexed by k. The probabilistic problem is a famous one, discussed by Feller in [**34**], as it represents a basic question in the analysis of runs of good (or bad) luck in a succession of independent events. Our presentation closely follows an insightful note of Knuth [**73**] whose research was motivated by the related problem of analysing carry propagation in certain binary adders.

PROPOSITION V.2. *The longest run parameter L taken over the set of binary words of length n (endowed with the uniform distribution) satisfies, for h in any bounded set of $\mathbb{Z}$, the uniform estimate*[1]

$$\mathbb{P}_n\left(L < \lfloor \lg n \rfloor + h\right) = e^{-\alpha(n)2^{-h}} + O\left(\frac{\log n}{n}\right), \qquad \alpha(n) := 2^{\{\lg n\}}. \tag{2}$$

In particular, the mean and variance satisfy $\mathbb{E}_n(L) = \lg n + O(1)$ and $\mathbb{V}_n(L) = O(1)$, and the distribution is concentrated around its mean.

The probability distribution in (2) is called a *double exponential distribution* (Figure 1). In fact, the formula is an *asymptotic* one. It does not represent a unique *limit distribution* in the usual sense, but rather a whole family depending on the fractional part of $\lg n$, that is, on the way n is placed with respect to powers of 2. This phenomenon is further reflected by the fact that the second asymptotic term in the mean is subject to fluctuations (albeit of a tiny amplitude), see the discussion of $\Phi(x)$ below.

PROOF. The specification $\mathcal{W}^{\langle k\rangle} = a^{<k}\mathfrak{S}(ba^{<k})$ describes those words for which this length is strictly less than k. The expression of the OGF,

$$W^{\langle k\rangle}(z) = \frac{1-z^k}{1-z} \cdot \frac{1}{1-z\frac{1-z^k}{1-z}} = \frac{1-z^k}{1-2z+z^{k+1}}, \tag{3}$$

results. Quite clearly, one should locate the dominant pole, separate it from the other poles (as this leads to constructive error terms), as well as estimate the contribution to the coefficients arising from this dominant pole.

[1]The symbol $\lg x$ denotes the binary logarithm, $\lg x = \log_2 x$.

(i) *Locating the dominant pole.* The OGF $W^{\langle k \rangle}$ has, by the first form of (3) a dominant pole ρ_k which is a root of the equation $1 = s(\rho_k)$, where $s(z) = z(1-z^k)/(1-z)$. We consider $k \geq 2$. Since $s(z)$ is an increasing polynomial and $s(0) = 0$, $s(\frac{1}{2}) < 1$, $s(1) = 1$, the root ρ_k must lie in the open interval $(\frac{1}{2}, 1)$. In fact, as one easily verifies, the condition $k \geq 2$ guarantees that $s(0.6) > 1$, hence the refined estimate

$$\frac{1}{2} < \rho_k < \frac{3}{5} \qquad (k \geq 2). \tag{4}$$

It now becomes possible to derive very precise estimates by bootstrapping. (This technique is a form of iteration for approaching a fixed point—its use in the context of asymptotic expansions is detailed in De Bruijn's book [**25**].) Writing the defining equation for ρ_k as a fixed point equation,

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

and making use of the rough estimates (4) yields next

$$\frac{1}{2}\left(1 + (\frac{1}{2})^k\right) < \rho_k < \frac{1}{2}\left(1 + (\frac{3}{5})^k\right). \tag{5}$$

Thus, ρ_k is exponentially close to $\frac{1}{2}$, and a further iteration from (5) shows

$$\rho_k = \frac{1}{2} + \frac{1}{2^{k+1}} + O\left(\frac{k}{2^{2k}}\right), \tag{6}$$

which constitutes a very precise estimate.

(ii) *Contribution from the dominant pole.* A straight calculation provides the value of the residue,

$$R_{n,k} := -\operatorname{Res}\left[W^{\langle k \rangle}(z) z^{-n-1}; z = \rho_k\right] = \frac{1 - \rho_k^k}{2 - (k+1)\rho^k} \rho_k^{-n-1}, \tag{7}$$

which is expected to provide the main approximation to the coefficients of $W^{\langle k \rangle}$ as $n \to \infty$. The meaning of (7) is better grasped if one notes that the residue resembles $2^n e^{-n/2^k}$. We shall return to such approximations shortly.

(iii) *Separation of the subdominant poles.* Consider the circle $|z| = \frac{3}{4}$ and take the second form of the denominator of $W^{\langle k \rangle}$, namely

$$1 - 2z + z^{k+1}.$$

In view of Rouché's theorem, we may regard this polynomial as the sum $f(z) + g(z)$, where $f(z) = 1 - 2z$ and $g(z) = z^{k+1}$. The term $f(z)$ has on the circle a modulus that varies between $\frac{1}{2}$ and $\frac{5}{2}$; the term $g(z)$ is at most $\frac{27}{64}$ for any $k \geq 2$. Thus, on the circle $|z| = \frac{3}{4}$, one has $|g(z)| < |f(z)|$, so that $f(z)$ and $f(z) + g(z)$ have the same number of zeros inside the circle. Since $f(z)$ admits $z = \frac{1}{2}$ as only zero there, the denominator must also have a unique root in $|z| \leq \frac{3}{4}$, and that root must coincide with ρ_k.

Similar arguments also give bounds on the error term when the number of words with longest run of length at most k is estimated by the residue (7) at the dominant pole. On the disc $|z| = \frac{3}{4}$, the denominator of $W^{\langle k \rangle}$ stays bounded away from 0 (its modulus is at least $\frac{5}{64}$ when $k \geq 2$, by previous considerations). Thus, the modulus of the remainder integral is $O((4/3)^n)$, and in fact bounded from above by $35(4/3)^n$. In summary, if we let $q_{n,k}$ represent the probability that the longest run in a random word of length n is less than k, one has available the main estimate

$$q_{n,k} := \mathbb{P}_n(L < k) = \frac{1 - \rho_k^k}{1 - (k+1)\rho_k^k/2} \left(\frac{1}{2\rho_k}\right)^{n+1} + O\left((\frac{2}{3})^n\right), \tag{8}$$

uniformly with respect to k. Here is table of the numerical values of the quantities appearing in the approximation of $q_{n,k}$ when written under the form $c_k \cdot (2\rho_k)^{-n}$:

k	$c_k \cdot (2\rho_k)^{-n}$
2	$1.17082 \cdot 0.80901^n$
3	$1.13745 \cdot 0.91964^n$
4	$1.09166 \cdot 0.96378^n$
5	$1.05753 \cdot 0.98297^n$
10	$1.00394 \cdot 0.99950^n$

(iv) There finally remains to transform the main estimate (8) into the limit form asserted in the statement. First, the "tail inequalities"

$$\mathbb{P}_n\left(L < \frac{3}{4}\lg n\right) = O\left(e^{-\frac{1}{2}\sqrt[4]{n}}\right), \qquad \mathbb{P}_n\left(L \geq 2\lg n\right) = O\left(\frac{1}{n}\right), \tag{9}$$

describe the tail of the probability distribution of L_n. They derive from simple bounding techniques applied to the main approximation (8) using (6). Thus, for asymptotic purposes, only a small region around $\lg n$ needs to be considered.

Regarding the central regime, for $k = \lg n + x$ and x in $[-\frac{1}{4}\lg n, \lg n]$, the approximation (6) of ρ_k and related quantities applies, and one finds

$$(2\rho_k)^{-n} = \exp\left(-\frac{n}{2^k} + O(kn2^{-2k})\right) = e^{-n/2^k}\left(1 + O(\frac{\log n}{n})\right).$$

(This results from standard expansions like $(1-a)^n = e^{-na}\exp(O(na^2))$.) At the same time, the coefficient of this quantity in (8) is

$$1 + O(k\rho_k^k) = 1 + O\left(\frac{\log n}{n^{3/4}}\right).$$

Thus a double exponential approximation holds (Figure 1) and for $k = \lg n + x$ with x in $[-\frac{1}{4}\lg n, \lg n]$, one has (uniformly)

$$q_{n,k} = e^{-n/2^k}\left(1 + O\left(\frac{\log n}{n^{3/4}}\right)\right). \tag{10}$$

In particular, upon setting $k = \lfloor \lg n \rfloor + h$, the first part of the statement follows. (The floor function takes into account the fact that k must be an integer.)

The mean and variance estimates derive from the fact that the distribution quickly decays at values away from $\lg n$ while it satisfies (10) in the central region. The mean is given by

$$\mathbb{E}_n(L) := \sum_{h\geq 0}[1 - \mathbb{P}_n(L < h)] = \Phi(n) + O\left(\frac{\log^2 n}{n}\right), \quad \Phi(x) := \sum_{h\geq 0}\left[1 - e^{-x/2^h}\right].$$

Consider the three cases $h < h_0$, $h \in [h_0, h_1]$, and $h > h_1$ with $h_0 = \lg x - \log\log x$ and $h_1 = \lg x + \log\log x$, where the general term is (respectively) close to 1, between 0 and 1, and close to 0. By summing, one finds elementarily $\Phi(x) = \lg x + O(\log\log x)$ as $x \to \infty$, and elementary ways of catching the next $O(1)$ term are discussed for instance in [**100**, p. 403].

The method of choice for precise asymptotics is to treat $\Phi(x)$ as a harmonic sum and apply Mellin transform techniques (APPENDIX: *Mellin Transform*, p. 120). The Mellin transform of $\Phi(x)$ is

$$\Phi^\star(s) := \int_0^\infty \Phi(x)x^{s-1}\,dx = \frac{\Gamma(s)}{1-2^s} \qquad \Re(s) \in (-1, 0).$$

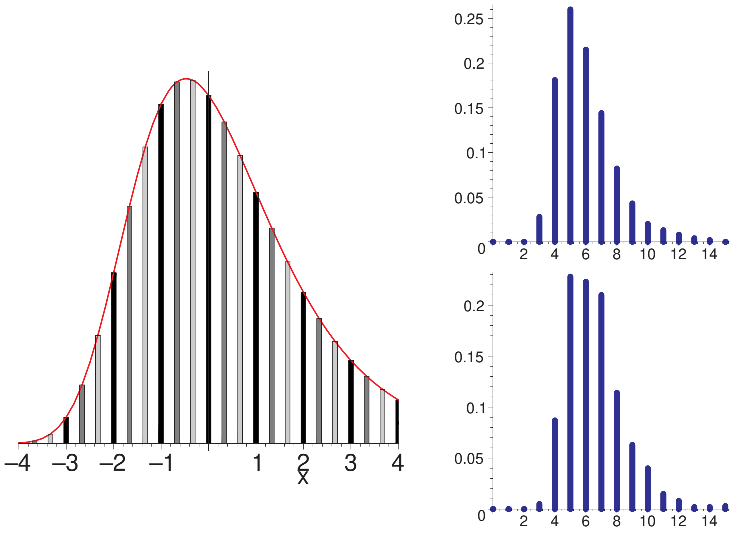


FIGURE 1. The double exponential laws: Left, histograms for n at 2^p (black), $2^{p+1/3}$ (dark gray), and $2^{p+2/3}$ (light gray), where $x = k - \lg n$. Right, empirical histograms for 1000 simulations with $n = 100$ (top) and $n = 140$ (bottom).

The double pole of $\Phi^\star$ at 0 and the simple poles at $s = \frac{2ik\pi}{\log 2}$ are reflected by the asymptotic expansion:

$$\Phi(x) = \lg x + \frac{\gamma}{\log 2} + \frac{1}{2} + P(\lg x) + O(x^{-1}), \quad \text{where} \quad P(w) = \sum_{k \in \mathbb{Z}\setminus\{0\}} \Gamma\left(\frac{2ik\pi}{\log 2}\right) e^{2ik\pi w}.$$

The oscillating function $P(w)$ has amplitude of the order of 10^{-6}. (See [**43, 73, 108**] for more on this topic.) The variance is similarly analysed. □

The analysis is closely related to the case of words excluding a patterns in Chapter IV. There, we conducted a global analysis applicable to any pattern. Here, we have specialized the discussion to patterns $\mathsf{aaa}\cdots\mathsf{a}$ and effectively extracted a whole family of limit distributions. What is striking is the existence of an *infinite family of limit laws*, which depend on the fractional part of $\lg n$. □

▷ **2.** *Longest runs in Bernoulli sequences.* Consider an alphabet $\mathcal{A}$ with letters independently chosen according to the probability distribution $\{p_j\}$. Then, the OGF of words where each letter is repeated at most k times derives from the construction of Smirnov words and is

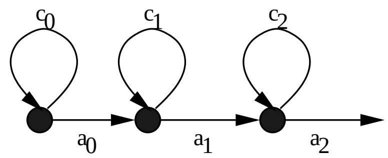
$$W^{[k]}(z) = \left(1 - \sum_i p_i z \frac{1-(p_i z)^k}{1-(p_i z)^{k+1}}\right)^{-1}.$$

Let $p_{\max}$ be the smallest of the p_j. Then the expected length of the longest run of any letter is $\log n / \log p_{\max} + O(1)$, and very precise quantitative information can be derived from the OGFs by methods akin to Example 7 (Smirnov words and Carlitz compositions) in Chapter IV, p. 32. ◁

The next batch of examples in this section develops the analysis of walks in a special type of graphs. These examples serve two purposes: they illustrate further cases of modelling by means of regular specifications, and, at the same time, provide a bridge with the analysis of lattice paths in the next section.

EXAMPLE 3. *Walks of the pure-birth type.* Consider a walk on the nonnegative integers that starts at 0 and is only allowed to either stay at the same place or progress by an increment of $+1$. Our goal is to enumerate the possible configurations that start from 0 and reach point $m-1$ in n steps. A step from j to $j+1$ will be encoded by a letter a_j; a step from state j to state j will be encoded by c_j. A diagram representing these steps is then:

(11)



(Compare with (21).) The language encoding all legal walks from state 0 to state $m-1$ can be described by a regular expression,

$$\mathcal{H}_{0,m-1} = (c_0)^\star a_0 (c_1)^\star a_1 \cdots (c_{m-2})^\star a_{m-2} (c_{m-1})^\star,$$

and the representation is certainly unambiguous. Symbolicly using letters as variables, the corresponding ordinary multivariate generating function is then

$$H_{0,m-1}(\vec{a},\vec{c}) = \frac{a_0 a_1 \cdots a_{m-2}}{(1-c_0)(1-c_1)\cdots(1-c_{m-1})}.$$

Assume that the steps are assigned weights, with α_j corresponding to a_j and γ_j to c_j. Weights of letters are extended multiplicatively to words in the usual way (cf Chapter III). If in addition, one takes $\gamma_j = 1-\alpha_j$, one obtains a probabilistic weighting: the walker starts from position 0, and, if at j, at each clock tick, she either stays at the same place with probability $1-\alpha_j$ or moves to the right with probability α_j. The OGF of such weighted walks then becomes

$$H_{0,m-1}(z) = \frac{\alpha_0\alpha_1\cdots\alpha_{m-2} z^{m-1}}{(1-(1-\alpha_0)z)(1-(1-\alpha_1)z)\cdots(1-(1-\alpha_{m-1})z)}, \tag{12}$$

and $[z^n]H$ is the probability for the walker to be found at position j at (discrete) time n. This walk process can be alternatively interpreted as a (discrete-time) pure *birth process* in the usual sense of probability theory: There is a population of individuals and, at each discrete epoch, a new birth may take place, the probability of a birth being α_j when the population is of size j.

The form (12) readily lends itself to a partial fraction decomposition. The poles of H are at the points $(1-\alpha_j)^{-1}$ and one finds as $z\to(1-\alpha_j)^{-1}$:

$$H_{0,m-1}(z) \sim \frac{r_{j,m-1}(1-\alpha_j)}{1-z(1-\alpha_j)} \quad \text{where} \quad r_{j,m-1} := \frac{\alpha_0\alpha_1\cdots\alpha_{m-2}}{\prod_{k\in[0,m-1],\ k\neq\{j\}}(\alpha_k-\alpha_j)}.$$

Thus, the probability of being in state $m-1$ at time n is

$$[z^n]H_{0,m-1}(z) = \sum_{j=0}^{m-1} r_{j,m-1}(1-\alpha_j)^{n+1}.$$

This has the form of an alternating sum that can be evaluated in each particular instance.

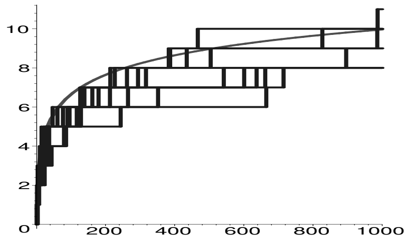


FIGURE 2. A simulation of 10 trajectories of the pure-birth process till $n = 1024$, with geometric probabilities corresponding to $q = 1/2$, compared to the curve $\log_2 x$.

An especially interesting case of the pure-birth walk is when the quantities α_k are geometric: $\alpha_k = q^{k+1}$ for some q with $0 < 1 < k$. In that case, the probability of being in state $m-1$ after n transitions becomes

$$\sum_{j=0}^{m-1} \frac{(-1)^j q^{\binom{j}{2}}}{(q)_j (q)_{m-j-1}} (1 - q^{m-j-1})^{n+1}, \quad (q)_j := (1-q)(1-q^2)\cdots(1-q^j).$$

This corresponds to a stochastic progression in a medium with exponentially increasing hardness or, equivalently, to the growth of a population where the current size of the population adversely affects fertility in an exponential manner. On intuitive grounds, we expect an evolution of the process to stay reasonably close to the curve $y = \log_{1/q} x$; see Figure 2 for a simulation confirming this fact, which can be justified by means of the analytic formulæ just described. This particular analysis is borrowed from [**37**], where it was initially developed in connection with the algorithm called "approximate counting" to be described below. □

Note. The theory of pure birth processes is discussed under a calculational and non measure-theoretic angle in the book by Bharucha-Reid [**14**]. See also the *Course* by Karlin and Taylor [**70**] for a concrete presentation.

EXAMPLE 4. *Approximate Counting.* Assume you need to keep a counter that is able to record the number of certain events (say impulses) and should have the capability of keeping counts till a certain maximal value N. A standard information-theoretic argument (with ℓ bits, one can only keep track of 2^ℓ possibilities) implies that one needs $\lceil \log_2 N+1 \rceil$ bits to perform the task—a standard binary counter will indeed do the job. However, in 1977, Robert Morris has proposed a way to maintain counters that only requires of the order of $\log\log N$ bits. What's the catch?

Morris' elegant idea consists in relaxing the constraint of exactness in the counting process and, by playing with probabilities, tolerate a small error on the counts obtained. Precisely, his solution maintains a random quantity Q which is initialized by $Q = 0$. Upon receiving an impulse, one updates Q according to the following simple procedure (with $q \in (0, 1)$ a design parameter):

procedure Update(Q);
with probability q^{Q+1} do Q := $Q + 1$ (else keep Q unchanged).

When asked the number of impulses (number of times the update procedure was called) at any moment, simply use the following procedure to return an estimate:

procedure Answer(Q);
output $\dfrac{q^{-Q} - 1}{1 - q}$.

Let Q_n be the value of the random quantity Q after n executions of the update procedure and X_n the corresponding estimate output by the algorithm. It is easy to verify (by recurrence or by generating functions, see Note 3 below) that

$$\mathbb{E}(q^{-Q_n}) = n(1-q)+1, \qquad \text{so that} \quad \mathbb{E}(X_n) = n. \tag{13}$$

Thus the answer provided at any instant is an *unbiased estimator* (in a mean value sense) of the actual count n. On the other hand, the analysis of the geometric pure-birth process in the previous example applies. In particular, the exponential approximation $(1-\alpha)^n \approx e^{-n\alpha}$ in conjunction with the basic formulæ show that for large n and m sufficiently near to $\log_{1/q} n$, one has (asymptotically) the *geometric-birth distribution*

$$\mathbb{P}\left(Q_n = \log_{1/q} n + x\right) = \sum_{j=0}^{\infty} \frac{(-1)^j q^{\binom{j}{2}}}{(q)_j (q)_\infty} \exp(-q^{x-j-1}) + o(1). \tag{14}$$

(We refer to [**37**] for details.) Such calculations imply that Q_n is with high probability (w.h.p.) close to $\log_{1/q} n$. Thus, if $n \le N$, the value of Q_n will be w.h.p. bounded from above by $(1+\epsilon)\log_{1/q} N$, with ϵ a small constant. But this means that the integer Q, which can itself be represented in binary, will only require

$$\log_2 \log n + O(1) \tag{15}$$

bits for storage, for fixed q.

A closer examination of the formulæ reveals that the accuracy of the estimate improves considerably when q becomes close to 1. The *standard error* is defined as $\frac{1}{n}\sqrt{\mathbb{V}(X_n)}$ and it measures (in a mean quadratic sense) the relative error to likely to be made. The variance of Q_n is, like the mean, determined by recurrence or generating functions, and one finds

$$\mathbb{V}(q^{-Q_n}) = \binom{n}{2}\frac{(1-q)^3}{q}, \qquad \frac{1}{n}\sqrt{\mathbb{V}(X_n)} \sim \sqrt{\frac{1-q}{q}}. \tag{16}$$

This means that accuracy increases as q approaches 1 and, by suitably dimensioning q, one can make it as small as desired. In summary, (13), (16), and (15) express the following property: *Approximate counting makes it possible to count till N using only about $\log\log N$ bits of storage, while achieving a standard error that is almost a constant and can be set to any prescribed value.* Morris' trick is now fully understood.

For instance, with $q = 2^{-1/16}$, it proves possible to count up to $2^{16} = 65536$ using only 8 bits (instead of 16), with an error likely not to exceed 20%. Naturally, there's not too much reason to appeal to the algorithm when a *single* counter needs to be managed. (Everybody can afford a few bits!) Approximate Counting turns out to be useful when a very large number of counts need to be kept *simultaneously*. It constitutes one of the early examples of a probabilistic algorithm in the management of large volumes of data, also known as *data mining*.

Functions akin to those of (14) also surface in other areas of probability theory. Guillemin, Robert, and Zwart [**64**] have detected them in processes that combine an additive increase and a multiplicative decrease (AIMD processes), in a context motivated by the adaptive transmission of "windows" of varying sizes in large communication networks (the TCP protocol of the internet). Biane, Bertoin, and Yor [**12**] encountered a function identical to (14) in their study of exponential functionals of Poisson processes. □

$\triangleright$ **3.** *Moments of* q^{-Q_n}. It is a perhaps surprising fact that any integral moment of q^{-Q_n} is a polynomial in n and q, like in (13), (16). To see it, define

$$\Phi(w) \equiv \Phi(w, \xi, q) := \sum_{m \geq 0} q^{m(m+1)/2} \frac{\xi^m w^m}{(1+\xi q)(1+\xi q^2)\cdots(1+\xi q^{m+1})}.$$

By (12), one has

$$\sum_{m \geq 0} H_{0,m} w^m = \frac{1}{1-z} \Phi\left(w; \frac{z}{1-z}, q\right).$$

On the other hand, Φ satisfies $\Phi(w) = 1 - q\xi(1-w)\Phi(qw)$, hence the q–identity,

$$\Phi(w) = \sum_{j \geq 0} (-q\xi)^j \left[(1-w)(1-qw)\cdots(1-q^{j-1}w)\right],$$

which resorts to q-calculus[2]. Thus $\Phi(q^{-r}; \xi, q)$ is a polynomial for any $r \in \mathbb{Z}_{\geq 0}$, as the expansion terminates. $\triangleleft$

Our last example makes use of regular expressions in order to estimate moments. Note that ambiguous representations are purposely used to accomplish the task.

EXAMPLE 5. *Occurrences of "hidden" patterns in texts.* Fix an alphabet $\mathcal{A} = \{a_1, \ldots, a_r\}$ of cardinality r and assume a probability distribution on $\mathcal{A}$ to be given, with p_j the probability of letter a_j. We consider the Bernoulli model on $\mathcal{W} = \mathfrak{S}(\mathcal{A})$, where the probability of a word is the product of the probabilities of its letters (cf Section III.5). A word $\mathfrak{p} = y_1 \cdots y_k$ called the pattern is fixed. The problem is to gather information on the random variable X representing the number of occurrences of $\mathfrak{p}$ in the set $\mathcal{W}_n$, where occurrences as a *"hidden pattern"*, i.e., as a *subsequence*, are counted (Section I.4.1). This is a basic example where counting with ambiguity proves useful.

The generating function associated to $\mathcal{W}$ endowed with its probabilistic weighting is

$$W(z) = \frac{1}{1 - \sum p_j z} = \frac{1}{1-z}.$$

The regular expression

$$\mathcal{O} = \mathfrak{S}(\mathcal{A}) y_1 \mathfrak{S}(\mathcal{A}) \cdots \mathfrak{S}(\mathcal{A}) y_{k-1} \mathfrak{S}(\mathcal{A}) y_k \mathfrak{S}(\mathcal{A}) \tag{17}$$

describes all contexts of occurrences of $\mathfrak{p}$ as a subsequence in all words. Graphically, this may be rendered as follows for a pattern of length 3, $\mathfrak{p} = y_1 y_2 y_3$:

$$\text{———}\boxed{y_1}\text{———}\boxed{y_2}\text{———}\boxed{y_3}\text{———} \tag{18}$$

There the boxes indicate distinguished positions where letters of the pattern appear and the horizontal lines represent arbitrary separating words ($\mathfrak{S}(\mathcal{A})$). The corresponding OGF

$$O(z) = \frac{\pi(\mathfrak{p}) z^k}{(1-z)^{k+1}}, \qquad \pi(\mathfrak{p}) := p_{y_1} \cdots p_{y_{k-1}} p_{y_k} \tag{19}$$

counts elements of $\mathcal{W}$ with *ambiguity*, where the ambiguity coefficient of a word $w \in \mathcal{W}$ is precisely equal to the number of occurrences of $\mathfrak{p}$ as a subsequence in w. There results that the expected number of hidden occurrences of $\mathfrak{p}$ in a random word of length n is

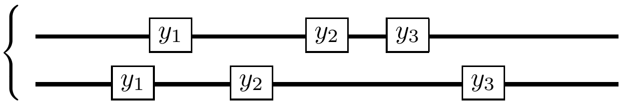
$$[z^n] O(z) = \pi(\mathfrak{p}) \binom{n}{k}, \tag{20}$$

which is consistent with what a direct probabilistic reasoning would give.

[2]By q–calculus is roughly meant the collection of special function identities relating power series of the form $\sum a_n(q) z^n$, where $a_n(q)$ is a rational fraction whose degree is quadratic in n. See [**5**, Ch. 10] for basics and [**57**] for more advanced (q–hypergeometric) material.

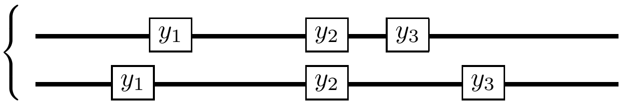
We next proceed to determine the variance of X over $\mathcal{W}_n$. In order to do so, we need contexts in which *pairs* of occurrences appear. Let $\mathcal{Q}$ denote the set of all words in $\mathcal{W}$ with *two* occurrences (i.e., an ordered pair of occurrences) of $\mathfrak{p}$ as a subsequence being distinguished. Then clearly $[z^n]Q(z)$ must represent $\mathbb{E}_{\mathcal{W}_n}[X^2]$. There are several cases to be considered. Graphically, a pair of occurrences may be interleaved and share no common position, like in what follows:

(21)

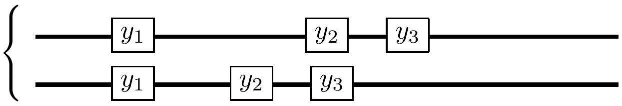


But they may also have one or several overlapping positions, like in

(22)



(23)



(This last situation necessitates $y_2 = y_3$, typical patterns being abb and aaa.)

In the first case corresponding to (21), where there are no overlapping positions, the configurations of interest have OGF

$$Q^{[0]}(z) = \binom{2k}{k} \frac{\pi(\mathfrak{p})^2 z^{2k}}{(1-z)^{2k+1}}. \tag{24}$$

There, the binomial coefficient $\binom{2k}{k}$ counts the total number of ways of freely interleaving two copies of $\mathfrak{p}$; the quantity $\pi(\mathfrak{p})^2 z^{2k}$ takes into account the $2k$ distinct positions where the letters of the two copies appear; the factor $(1-z)^{-2k-1}$ corresponds to all the possible $2k+1$ fillings of the gaps between letters.

In the second case, let us start by considering pairs where exactly one position is overlapping, like in (22). Say this position corresponds to the rth and sth letters of $\mathfrak{p}$ (r and s may not be equal). Obviously, we need $y_r = y_s$ for this to be possible. The OGF of the configurations is now

$$\binom{r+s-2}{r-1}\binom{2m-r-s}{m-r} \frac{\pi(\mathfrak{p})^2 (p_{y_r})^{-1} z^{2k-1}}{(1-z)^{2k}}.$$

There, the first binomial coefficient $\binom{r+s-2}{r-1}$ counts the total number of ways of interleaving $y_1 \cdots y_{r-1}$ and $y_1 \cdots y_{s-1}$; the second binomial $\binom{2m-r-s}{m-r}$ is similarly associated to the interleavings of $y_{r+1} \cdots y_k$ and $y_{s+1} \cdots y_k$; the numerator takes into account the fact that $2k-1$ positions are now occupied by predetermined letters; finally the factor $(1-z)^{-2k}$ corresponds to all the $2k$ fillings of the gaps between letters. Summing over all possibilities for r, s gives the OGF of pairs with one overlapping position as

$$Q^{[1]}(z) = \left(\sum_{1 \le r,s \le k} \binom{r+s-2}{r-1}\binom{2m-r-s}{m-r} \frac{[\![y_r = y_s]\!]}{p_{y_r}} \right) \frac{\pi(\mathfrak{p})^2 z^{2k-1}}{(1-z)^{2k}}. \tag{25}$$

Similar arguments show that the OGF of pairs of occurrences with at least *two* shared positions (see, e.g., 23)) is of the form, with P a polynomial,

$$Q^{[\geq 2]}(z) = \frac{P(z)}{(1-z)^{2k-1}}, \tag{26}$$

for the essential reason that, in the finitely many remaining situations, there are at most $(2k-1)$ possible gaps.

We can now examine (24), (25), (26) in the light of singularities. The coefficient $[z^n]Q^{[0]}(z)$ is seen to cancel to first asymptotic order with the square of the mean as given in (20). The contribution of the coefficient $[z^n]Q^{[\geq 2]}(z)$ appears to be negligible as it is $O(n^{2k-2})$. The coefficient $[z^n]Q^{[1]}(z)$, which is $O(n^{2k-1})$, is seen to contribute to the asymptotic growth of the variance. In summary, after a trite calculation, we obtain:

PROPOSITION V.3. *The number X of occurrences of a hidden pattern $\mathfrak{p}$ in a random text of size n obeying a Bernoulli model satisfies*

$$\mathbb{E}_{\mathcal{W}_n}[X] = \pi(\mathfrak{p})\binom{n}{k} \sim \frac{\pi(\mathfrak{p})}{k!}n^k, \qquad \mathbb{V}_{\mathcal{W}_n}[X] = \frac{\pi(\mathfrak{p})^2\kappa(\mathfrak{p})^2}{(2k-1)!}n^{2k-1}\left(1 + O(\frac{1}{n})\right),$$

where the "correlation coefficient" $\kappa(\mathfrak{p})^2$ is given by

$$\kappa(\mathfrak{p})^2 = \sum_{1 \leq r,s \leq k} \binom{r+s-2}{r-1}\binom{2m-r-s}{m-r}[\![y_r = y_s]\!]\left(\frac{1}{p_{y_r}} - 1\right).$$

In particular, the distribution of X is concentrated around its mean.

This example is based on an article by Flajolet, Szpankowski, and Vallée [**53**]. There the authors show further that the asymptotic behaviour of moments of higher order can be worked out. By the moment convergence theorem described in Chapter VII, this calculation entails that *the distribution of X over $\mathcal{W}_n$ is asymptotically normal*. The method also extends to a much more general notion of "hidden" pattern, e.g., distances between letters of $\mathfrak{p}$ can be constrained in various ways so as to determine a valid occurrence in the text [**53**]. It also extends to the very general framework of dynamical sources [**16**], which include Markov models as a special case. The two references [**16, 53**] thus provide a set of analyses that interpolate between the two extreme notions of pattern occurrence—as a block of consecutive symbols or as a subsequence ("hidden pattern"). Such studies demonstrate that hidden patterns are with high probability bound to occur an extremely large number of times in a long enough text—this might cast some doubts on numerological interpretations encountered in various cultures. □

▷ **4.** *Hidden patterns and shuffle relations.* To each pairs u, v of words over $\mathcal{A}$ associate the weighted-shuffle polynomial in the indeterminates $\mathcal{A}$ denoted by $\left(\!\binom{u}{v}\!\right)_t$ and defined by the properties

$$\begin{cases} \left(\!\!\binom{xu}{yv}\!\!\right)_t = x\left(\!\!\binom{u}{yv}\!\!\right)_t + y\left(\!\!\binom{xu}{v}\!\!\right)_t + t[\![x=y]\!]x\left(\!\!\binom{u}{v}\!\!\right)_t \\ \left(\!\!\binom{\mathbf{1}}{u}\!\!\right)_t = \left(\!\!\binom{u}{\mathbf{1}}\!\!\right)_t = u \end{cases}$$

where t is a parameter, x, y are elements of $\mathcal{A}$, and $\mathbf{1}$ is the empty word. Then the OGF of $Q(z)$ above is

$$Q(z) = \sigma\left[\left(\!\!\binom{\mathfrak{p}}{\mathfrak{p}}\!\!\right)_{(1-z)}\right]\frac{1}{(1-z)^{2k+1}},$$

where σ is the substitution $a_j \mapsto p_j z$. ◁

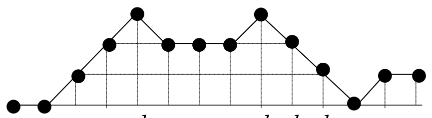
V. 2. Lattice paths and walks on the line.

In this section, we consider *lattice paths* that are fundamental objects of combinatorics. Indeed, they relate to trees, permutations, and set partitions, to name a few. They also correspond to walks on the integer half-line and as such they relate to classical 1-dimensional random walks and to birth-and-death processes of probability theory. The lattice paths discussed here have steps that correspond to movements either immediately to the left or to the right. Combinatorially, such paths are the limit of paths of bounded height, themselves definable as nested sequences. As a consequence, the OGF's obtained involve a cascade of quasi-inverses, $1/(1-f)$, so that they are of the continued fraction type.

DEFINITION V.1 (Lattice path). *A (lattice)* path $\upsilon = (U_0, U_1, \ldots, U_n)$ *is a sequence of points in the lattice* $\mathbb{N}\times\mathbb{N}$ *such that if* $U_j = (x_j, y_j)$*, then* $x_j = j$ *and* $|y_{j+1} - y_j| \leq 1$*. An edge* $\langle U_j, U_{j+1}\rangle$ *is called an* ascent *(*$\underline{a}$*) if* $y_{j+1} - y_j = +1$*, a* descent *(*$\underline{b}$*) if* $y_{j+1} - y_j = -1$*, and a* level step *(*$\underline{c}$*) if* $y_{j+1} - y_j = 0$.

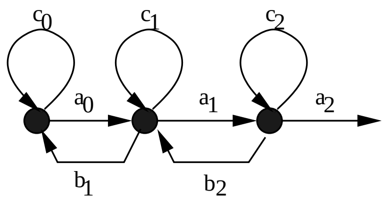
The quantity n *is the* length *of the path,* $o(\upsilon) := y_0$ *is the* initial altitude*,* $h(\upsilon) := y_n$ *is the* final altitude*. A path is called an* excursion *if both its initial and final altitudes are zero. The extremal quantities* $\sup\{\upsilon\} := \max_j y_j$ *and* $\inf\{\upsilon\} := \min_j y_j$ *are called the* height *and* depth *of the path.*

It is assumed that paths are normalized by the condition $x_0 = 0$. With this normalization, a path of length n is encoded by a word with a, b, c representing ascents, descents, and level steps, respectively. What we call the *standard encoding* is such a word in which each step a, b, c is (redundantly) subscripted by the value of the y-coordinate of its associated point. For instance,



$w = c_0\ a_0\ a_1\ a_2\ b_3\ c_2\ c_2\ a_2\ b_3\ b_2\ b_1\ a_0\ c_1$

encodes a path that connects the initial point $(0, 0)$ to the point $(13, 1)$. Such a path can also be regarded as a rendering of the evolution in discrete time of a walk over the integer line:



Equivalently, lattice paths cane be read as trajectories of birth-and-death processes. (Compare with the pure-birth case in (11) above.)

Let $\mathcal{H}$ be the set of all lattice paths. Given a geometric condition (Q), it is then possible to associate to it a "language" $\mathcal{H}[Q]$ that comprises the collection of all path encodings satisfying the condition Q. This language can be viewed either as a set or as a formal sum,

$$H[Q] = \sum_{\{w \mid Q\}} w,$$

in which case it becomes the generating function in infinitely many indeterminates of the corresponding condition.

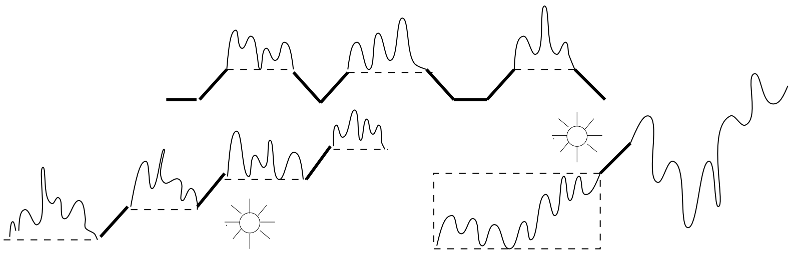


FIGURE 3. The three major decompositions of lattice paths: the arch decomposition (top), the last passages decomposition (bottom left), and the first passage decomposition (bottom right).

The general subclass of paths of interest in this subsection is defined by arbitrary combinations of flooring (m), ceiling (h), as well as fixing initial (k) and final (l) altitudes:

$$H_{k,l}^{[\geq m,<h]} \quad = \quad \{w \in H : o(w) = k,\ h(w) = l,\ \inf\{w\} \geq m,\ \sup\{w\} < h\}.$$

We also need the specializations, $H_{k,l}^{[<h]} = H_{k,l}^{[\geq 0,<h]}$, $H_{k,l}^{[\geq m]} = H_{k,l}^{[\geq m,<\infty]}$, $H_{k,l} = H_{k,l}^{[\geq 0,<\infty]}$. Three simple combinatorial decompositions of paths then suffice to derive all the basic formulæ.

Arch decomposition: An excursion from and to level 0 consists of a sequence of "arches", each made of either a c_0 or a $a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1$, so that

$$\mathcal{H}_{0,0} = \left(c_0 \cup a_0\mathcal{H}_{1,1}^{[\geq 1]}b_1\right)^\star, \tag{27}$$

which relativizes to height $< h$.

Last passages decomposition. Recording the times at which each level $0, \ldots, k$ is last traversed gives

$$\mathcal{H}_{0,k} = \mathcal{H}_{0,0}^{[\geq 0]}a_0\mathcal{H}_{1,1}^{[\geq 1]}a_1 \cdots a_{k-1}\mathcal{H}_{k,k}^{[\geq k]} \tag{28}$$

First passage decomposition. The quantities $H_{k,l}$ with $k \leq l$ are implicitly determined by the first passage through k in a path connecting level 0 to l, so that

$$\mathcal{H}_{0,l} \quad = \quad \mathcal{H}_{0,k-1}^{[<k]}a_{k-1}\mathcal{H}_{k,l} \quad (k \leq l), \tag{29}$$

(A dual decomposition holds when $k \geq l$.)

The basic results express the generating functions in terms of a fundamental continued fraction and its associated convergent polynomials. They involve the "numerator" and "denominator" polynomials, denoted by P_h and Q_h that are defined as solutions to the second order (or "three-term") recurrence equation

$$Y_{h+1} = (1 - c_h)Y_h - a_{h-1}b_hY_{h-1},\ h \geq 1, \tag{30}$$

together with the initial conditions $(P_{-1}, Q_{-1}) = (1, 0)$, $(P_0, Q_0) = (0, 1)$, and with the convention $a_{-1}b_0 = 1$. In other words, setting $C_j = 1 - c_j$ and $A_j = a_{j-1}b_j$, we have:

$$\begin{array}{llll} P_0 = 0, & P_1 = 1, & P_2 = C_2, & P_3 = C_1C_2 - A_2 \\ Q_0 = 0, & Q_1 = C_0, & Q_2 = C_0C_1 - A_1, & Q_3 = C_0C_1C_2 - C_2A_1 - C_0A_2 \end{array}$$

These polynomials are known as continuant polynomials [**76, 113**].

THEOREM V.2 (Path continued fractions [**36**]). (i) *The generating function* $H_{0,0}$ *of all excursions is represented by the fundamental continued fraction:*

$$H_{0,0} = \cfrac{1}{1 - c_0 - \cfrac{a_0 b_1}{1 - c_1 - \cfrac{a_1 b_2}{1 - c_2 - \cfrac{a_2 b_3}{\ddots}}}}. \tag{31}$$

(ii) *The generating function of ceiled excursion* $H_{0,0}^{[<h]}$ *is given by a convergent of the fundamental fraction (with* P_h, Q_h *given by (30):*

$$H_{0,0}^{[<h]} = \cfrac{1}{1 - c_0 - \cfrac{a_0 b_1}{1 - c_1 - \cfrac{a_1 b_2}{\cfrac{\ddots}{1 - c_{h-1}}}}} \tag{32}$$

$$= \frac{P_h}{Q_h}. \tag{33}$$

(iii) *The generating function of floored excursions is given by the truncation of the fundamental fraction:*

$$H_{h,h}^{[\geq h]} = \cfrac{1}{1 - c_h - \cfrac{a_h b_{h+1}}{1 - c_{h+1} - \cfrac{a_{h+1} b_{h+2}}{\ddots}}} \tag{34}$$

$$= \frac{1}{a_{h-1} b_h} \frac{Q_h H_{0,0} - P_h}{Q_{h-1} H_{0,0} - P_{h-1}}, \tag{35}$$

PROOF. Repeated use of the arch decomposition (27) provides a form of $H_{0,0}^{[<h]}$ with nested quasi-inverses $(1-f)^{-1}$ that is the finite fraction representation (32), for instance,

$$\mathcal{H}_{00}^{[<1]} \cong \mathfrak{S}\{c_0\}, \quad \mathcal{H}_{00}^{[<2]} \cong \mathfrak{S}\{c_0 + a_0 \mathfrak{S}\{c_1\} b_1\},$$
$$\mathcal{H}_{00}^{[<3]} \cong \mathfrak{S}\{c_0 + a_0 \mathfrak{S}\{c_1 + a_0 \mathfrak{S}\{c_2\} b_2\} b_1\}.$$

The continued fraction representation for basic paths (namely $H_{0,0}$) is then obtained by letting $h \to \infty$ in (32). Finally, the continued fraction form (34) for ceiled excursions is nothing but the fundamental form (31), when the indices are shifted. The three continued fraction expressions (31), (32), (34) are hence established.

Finding explicit expressions for the fractions $H_{0,0}^{[<h]}$ and $H_{h,h}^{[\geq h]}$ next requires determining the polynomials that appear in the convergents of the basic fraction (31). By definition, the convergent polynomials P_h and Q_h are the numerator and denominator of the fraction $H_{0,0}^{[<h]}$. For the computation of $H_{0,0}^{[<h]}$ and P_h, Q_h, one classically introduces the linear fractional transformations

$$g_j(y) = \frac{1}{1 - c_j - a_j b_{j+1} y},$$

so that

$$H_{0,0}^{[<h]} = g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(0) \text{ and } H_{0,0} = g_0 \circ g_1 \circ g_2 \circ \cdots \;,. \tag{36}$$

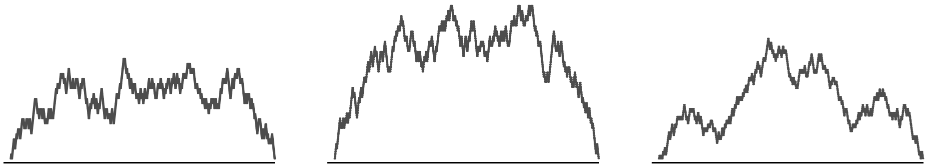


FIGURE 4. Three random Dyck paths of length $2n = 500$ have altitudes resp. 20, 31, 24: the distribution is spread, see Proposition V.4.

Now, linear fractional transformations are representable by 2×2-matrices

$$\frac{ay+b}{cy+d} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \tag{37}$$

in such a way that composition corresponds to matrix product. By induction on the compositions that build up $H_{0,0}^{[<h]}$, there follows the equality

$$g_0 \circ g_1 \circ g_2 \circ \cdots \circ g_{h-1}(y) = \frac{P_h - P_{h-1}a_{h-1}b_h y}{Q_h - Q_{h-1}a_{h-1}b_h y}, \tag{38}$$

where P_h and Q_h are seen to satisfy the recurrence (30). Setting $y = 0$ in (38) proves (33).

Finally, $H_{h,h}^{[\geq h]}$ is determined implicitly as the root y of the equation $g_0 \circ \cdots \circ g_{h-1}(y) = H_{0,0}$, an equation that, when solved using (38), yields the form (35). □

A large number of generating functions can be derived by similar techniques. We refer to the article [**36**], where this theory was first systematically developed and to the exposition given in [**60**, Chapter 5]. Our presentation here draws upon [**44**] where the theory was put to further use in order to develop a formal algebraic theory of the general birth-and-death process in continuous time.

▷ **5.** *Transitions and crossings.* The lattice paths $\mathcal{H}_{0,l}$ corresponding to the transitions from altitude 0 to l and $\mathcal{H}_{k,0}$ (from k to 0) have OGFs

$$H_{0,l} = \frac{1}{\beta_l}\left(Q_l H_{0,0} - P_l\right), \qquad H_{k,0} = \frac{1}{\alpha_k}(Q_k H_{0,0} - P_k).$$

The crossings $\mathcal{H}_{0,h-1}^{[<h]}$ and $\mathcal{H}_{h-1,0}^{[<h]}$ have OGFs,

$$H_{0,h-1}^{[<h]} = \frac{\alpha_{h-1}}{Q_h}, \qquad H_{h-1,0}^{[<h]} = \frac{\beta_{h-1}}{Q_h},$$

obtained from the last passages decomposition. (Abbreviations used are: $\alpha_m = a_0 \cdots a_{m-1}$, $\beta_m = b_1 \cdots b_m$.) This gives combinatorial interpretations for fractions of the form $1/Q$ and results from the basic decompositions combined with Theorem V.2; see [**36, 44**] for details. ◁

We examine next a few specializations of the general formulæ provided by Theorem V.2.

EXAMPLE 6. *Height of standard lattice paths.* In order to count lattice paths, it suffices to effect one of the substitutions,

$$\sigma_M : a_j \mapsto z,\ b_j \mapsto z, c_j \mapsto z; \qquad \sigma_D : a_j \mapsto z,\ b_j \mapsto z, c_j \mapsto 0.$$

In the former case, all three step types are taken into account, giving rise to so-called "Motzkin paths"; in the latter case level steps are disallowed, and one obtains so-called "Dyck paths".

We henceforth restrict attention to the case of Dyck paths. See Figure 4 for three simulations suggesting that the distribution of height is somewhat spread. The continued fraction expressing $H_{0,0}$ is in this case purely periodic, and it represents a quadratic function:

$$H_{0,0}(z) = \cfrac{1}{1 - \cfrac{z^2}{1 - \cfrac{z^2}{1 - \ddots}}} = \frac{1}{2z^2}\left(1 - \sqrt{1-4z^2}\right),$$

since $H_{0,0}$ satisfies $y = (1 - z^2 y)^{-1}$. The families of polynomials P_h, Q_h are in this case determined by a recurrence with constant coefficients and they coincide, up to a shift of indices. Define classically the Fibonacci polynomials by the recurrence

$$F_{h+2}(z) = F_{h+1}(z) - zF_h(z), \quad F_0(z) = 0, \quad F_1(z) = 1. \tag{39}$$

One finds $Q_h = F_{h+1}(z^2)$ and $P_h = F_h(z^2)$. (The Fibonacci polynomials are essentially reciprocals of Chebyshev polynomials.) By Theorem V.2, the GF of paths of height $< h$ is then

$$H_{00}^{[<h]}(z) = \frac{F_h(z^2)}{F_{h+1}(z^2)}.$$

(We get more and, for instance, the number of ways of crossing a strip of width $h-1$ is $H_{0,h-1}^{[<h]}(z) = z^{h-1}/F_{h+1}(z^2)$.) Note that the polynomials have an explicit form,

$$F_h(z) = \sum_{k=0}^{\lfloor (h-1)/2 \rfloor} \binom{h-1-k}{k}(-z)^k,$$

as follows from the generating function expression: $\sum_h F_h(z)y^h = y/(1-y+zy^2)$.

The equivalence between Dyck paths and (general) plane tree traversals discussed in Chapter I implies that trees of height at most h and size $n+1$ are equinumerous with Dyck paths of length $2n$ and height at most h. Set for convenience

$$G^{[h]}(z) = zH_{00}^{[<h+1]}(z^{1/2}) = z\frac{F_{h+1}(z)}{F_{h+2}(z)},$$

which is precisely the OGF of general plane trees having height $\leq h$. (This is otherwise in agreement with the continued fraction form obtained directly in Chapter III). It is possible to go much further as first shown by De Bruijn, Knuth, and Rice in a beautiful paper [**26**], which also constitutes the historic application of Mellin transforms in analytic combinatorics. (We refer to this paper for and historical context and references.)

First, solving the linear recurrence (39) with z treated as a parameter yields the alternative closed form expression

$$F_h(z) = \frac{G^h - \overline{G}^h}{G - \overline{G}}, \qquad G = \frac{1-\sqrt{1-4z}}{2}, \quad \overline{G} = \frac{1+\sqrt{1-4z}}{2}. \tag{40}$$

There, $G(z)$ is the OGF of all trees, and an equivalent form of $G^{[h]}$ is provided by

$$G - G^{[h-2]} = \sqrt{1-4z}\frac{u^h(z)}{1-u^h}, \quad \text{where} \quad u = \frac{1-\sqrt{1-4z}}{1+\sqrt{1-4z}} = \frac{G^2}{z}, \tag{41}$$

as is easily verified. Thus $G^{[h]}$ can be expressed in terms of $G(z)$ and z:

$$G - G^{[h-2]} = \sqrt{1-4z}\sum_{j\geq 1} z^{-jh}G(z)^{2jh}.$$

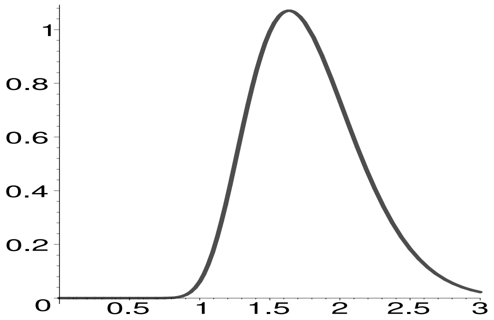


FIGURE 5. The limit density of the distribution of height $-\Theta'(x)$.

The Lagrange-Bürmann inversion theorem then gives after a simple calculation

$$G_{n+1} - G_{n+1}^{[h-2]} = \sum_{j\geq 1} \Delta^2 \binom{2n}{n-jh}, \tag{42}$$

where

$$\Delta^2 \binom{2n}{n-m} := \binom{2n}{n+1-m} - 2\binom{2n}{n-m} + \binom{2n}{n-1-m}.$$

Consequently, the number of trees of height $\geq h-1$ admits of closed form: it is a "sampled" sum by steps of h of the $2n$th line of Pascal's triangle (upon taking second order differences).

The relation (42) leads easily to the asymptotic distribution of height in random trees of size n. Stirling's formula yields the Gaussian approximation of binomial numbers: for $k = o(n^{3/4})$ and with $w = k/\sqrt{n}$, one finds

$$\frac{\binom{2n}{n-k}}{\binom{2n}{n}} \sim e^{-w^2}\left(1 - \frac{w^4+3w^2}{6n} + \frac{5w^8+6w^6-45w^4-60}{360n^2} + \cdots\right). \tag{43}$$

The use of the Gaussian approximation (43) inside the exact formula (42) then implies: *The probability that a tree of size $n+1$ has height at least $h-1$ satisfies uniformly for $h \in [\alpha\sqrt{n}, \beta\sqrt{n}]$ (with $0 < \alpha < \beta < \infty$) the estimate*

$$\frac{G_{n+1} - G_{n+1}^{[h-2]}}{G_{n+1}} = \Theta\left(\frac{h}{\sqrt{n}}\right) + O\left(\frac{1}{n}\right), \qquad \Theta(x) := \sum_{j\geq 1} e^{-j^2x^2}(4j^2x^2-2). \tag{44}$$

The function $\Theta(x)$ is a "theta function" which classically arises in the theory of elliptic functions [**114**]. Since binomial coefficients decay fast away from the center, simple bounds also show that the probability of height to be at least $n^{1/2+\epsilon}$ decays like $\exp(-n^{2\epsilon})$, hence is exponentially small. Note also that the probability distribution of height H itself admits of an exact expression obtained by differencing (42), which is reflected asymptotically by differentiation of the estimate of (44):

(45)
$$\mathbb{P}_{\mathcal{G}_{n+1}}\left[H = \lfloor x\sqrt{n}\rfloor\right] = -\frac{1}{\sqrt{n}}\Theta'(x) + O\left(\frac{1}{n}\right), \qquad \Theta'(x) := \sum_{j\geq 1} e^{-j^2x^2}(12j^2x - 8j^4x^3).$$

The forms (44) and (45) also give access to moments of the distribution of height. We find

$$\mathbb{E}_{\mathcal{G}_{n+1}}\big[H^r\big] \sim \frac{1}{\sqrt{n}} S_r\left(\frac{1}{\sqrt{n}}\right), \qquad \text{where} \quad S_r(y) := -\sum_{h\geq 1} h^r \Theta'(hy).$$

The quantity $y^{r+1}S_r(y)$ is a Riemann sum relative to the function $-x^r\Theta'(x)$, and the step $y = n^{-1/2}$ decreases to 0 as $n \to \infty$. Approximating the sum by the integral, one gets:

$$\mathbb{E}_{\mathcal{G}_{n+1}}\big[H^r\big] \sim n^{r/2}\mu_r \qquad \text{where} \quad \mu_r := -\int_0^\infty x^r \Theta'(x)\,dx.$$

The integral giving μ_r is a Mellin transform in disguise (set $s = r+1$) to which the treatment of harmonic sums applies. We then get upon replacing $n+1$ to n:

PROPOSITION V.4. *The expected height of a random plane rooted tree comprising n nodes is*

$$\sqrt{\pi n} - \frac{1}{2} + o(1).$$

More generally, the moment of order r of height is asymptotic to

$$\mu_r n^{r/2} \qquad \textit{where} \quad \mu_r = r(r-1)\Gamma(r/2)\zeta(r).$$

The random variable $H/\sqrt{n}$ obeys asymptotically a Theta distribution, in the sense of both the "central" estimate (44) and the "local" estimate (45). The same asymptotic estimates hold for height of Dyck paths having length $2n$.

The improved estimate of the mean is from [**26**]. The general moment forms are in fact valid for any real r (not just integers). An alternative formula for the Theta function appears in the Note below. Figure 5 plots the limit density $-\Theta'(x)$. $\square$

$\triangleright$ **6.** *Height, Fibonacci and Chebyshev polynomials.* The reciprocal polynomials $U_h(z) = z^h F_h(1/z)$ satisfy $U_h(\cos(\theta)) = \sin((h+1)\theta)/\sin(\theta)$ as is readily verified from the recurrence (39) and elementary trigonometry. Thus, the roots of $F_h(z)$ are $(4\cos^2 j\pi/h)^{-1}$ and the partial fraction expansion of $G^{[h]}(z)$ can be worked out explicitly [**26**]. There results

$$G_{n+1}^{[h-2]} = \frac{4^{n+1}}{h} \sum_{1\leq j\leq h/2} \sin^2\frac{j\pi}{h}\cos^{2n}\frac{j\pi}{h}, \tag{46}$$

which provides in particular an asymptotic form for any fixed h. (This formula can also be found directly from the sampled sum (42) by multisection of series.) Asymptotic analysis of this last expression when $h = x\sqrt{n}$ yields the alternative expression

$$\lim_{n\to\infty} \mathbb{P}_{\mathcal{G}_{n+1}}\big[H \leq x\sqrt{n}\big] = 4\pi^{5/2}x^{-3}\sum_{j\geq 0} j^2 e^{-j^2\pi^2/x^2} \quad (\equiv 1-\Theta(x)),$$

which reflects a classical transformation formula of theta functions. See the study by Biane, Pitman, and Yor [**15**] for fascinating connections between this formula, Brownian motion, and the functional equation of the Riemann zeta function. $\triangleleft$

$\triangleright$ **7.** *Motzkin paths.* The OGF of Motzkin paths of height $< h$ is

$$\frac{1}{1-z}\cdot {}^{\mathrm{D}}H_{0,0}^{[h]}\left(\frac{z}{1-z}\right),$$

where ${}^{\mathrm{D}}H_{0,0}^{[h]}$ above refers to Dyck paths. Therefore, such paths of length n can be enumerated exactly by formulæ derived from (42) and 46). In particular, the expected height is $\sim \sqrt{\pi n/3}$. $\triangleleft$

▷ **8.** *Height in simple varieties of trees.* Consider a simple variety of trees corresponding to the GF equation $Y(z) = z\phi(Y(z))$ (see Chapter III) and values of n such that there exists a tree of size n. Assume that there exists a positive τ strictly within the disc of convergence of ϕ such that $\tau\phi'(\tau) - \phi(\tau) = 0$. Then, the rth moment of height ($\overline{H}$) is asymptotically $\xi^{r/2} r(r-1)\Gamma(r/2)\zeta(r)n^{r/2}$. The normalized quantity $H = H/\xi$ obeys asymptotically a Theta distribution in the sense of both the central estimate (44) and the local estimate (45). [This is from [**47**] and [**41**] respectively.] For instance, $\xi = 2$ for plane binary trees and $\xi = \sqrt{2}$ for Cayley trees. ◁

EXAMPLE 7. *Area under Dyck path and coin fountains.* Consider the case of Dyck path and the parameter equal to the area below the path. Area under a lattice path can be defined as the sum of the indices (i.e., the starting altitudes) of all the variables that enter the standard encoding of the path. Thus, the BGF $D(z, q)$ of Dyck path with z marking half-length and q marking area is obtained by the substitution

$$a_j \mapsto q^j z, \quad b_j \mapsto q^j, \quad c_j \mapsto 0$$

inside the fundamental continued fraction (31). It proves convenient to operate with the continued fraction

$$F(z, q) = \cfrac{1}{1 - \cfrac{zq}{1 - \cfrac{zq^2}{\ddots}}}, \tag{47}$$

so that $D(z, q) = F(q^{-1}z, q^2)$. Since F and D satisfy difference equations, for instance,

$$F(z, q) = \frac{1}{1 - zqF(qz, q)}, \tag{48}$$

moments of area can be determined by differentiating and setting $q = 1$ (see Chapter III for such a direct approach).

A general trick from q–calculus is effective to derive an alternative expression of F. Attempt to express the continued fraction F of (47) as a quotient $F(z, q) = A(z)/B(z)$. Then, the relation (48) implies

$$\frac{A(z)}{B(z)} = \frac{1}{1 - qz\frac{A(qz)}{B(qz)}}, \quad \text{hence } A(z) = B(qz),\ B(z) = B(qz) - qzB(q^2 z),$$

where q is treated as a parameter. The difference equation satisfied by $B(z)$ is readily solved by indeterminate coefficients: this classical technique was introduced in the theory of integer partitions by Euler. With $B(z) = \sum b_n z^n$, the coefficients satisfy the recurrence

$$b_0 = 1, \quad b_n = q^n b_n - q^{2n-1} b_{n-1}.$$

This is a first order recurrence on b_n that unwinds to give

$$b_n = (-1)^n \frac{q^{n^2}}{(1-q)(1-q^2)\cdots(1-q^n)}.$$

In other words, introducing the "q-exponential function",

$$E(z, q) = \sum_{n=0}^{\infty} \frac{(-z)^n q^{n^2}}{(q)_n}, \quad \text{where} \quad (q)_n = (1-q)(1-q^2)\cdots(1-q^n), \tag{49}$$

one finds

$$F(z,q) = \frac{E(qz,q)}{E(z,q)}. \tag{50}$$

Given the importance of the functions under discussion in various branches of mathematics, we cannot resist a quick digression. The name of the q-exponential comes form the obvious property that $E(z(q-1),q)$ reduces to e^{-z} as $q \to 1^-$. The explicit form (49) constitutes in fact the "easy half" of the proof of the celebrated Rogers-Ramanujan identities, namely,

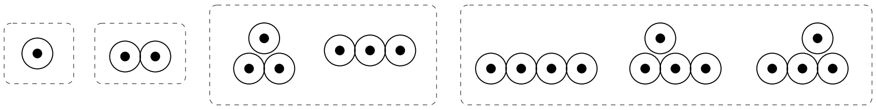
$$\begin{aligned} E(-1,q) = \sum_{n=0}^{\infty} \frac{q^{n^2}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+1})^{-1}(1-q^{5n+4})^{-1} \\ E(-q,q) = \sum_{n=0}^{\infty} \frac{q^{n(n+1)}}{(q)_n} &= \prod_{n=0}^{\infty} (1-q^{5n+2})^{-1}(1-q^{5n+3})^{-1}, \end{aligned} \tag{51}$$

that relate the q-exponential to modular forms. See Andrews' book [**4**, Ch. 7] for context.

Here is finally a cute application of these ideas to asymptotic enumeration. Odlyzko and Wilf define in [**91, 89**] an (n,m) coin fountain as an arrangement of n coins in rows in such a way that there are m coins in the bottom row, and that each coin in a higher row touches exactly two coins in the next lower row. Let $C_{n,m}$ be the number of (n,m) fountains and $C(q,z)$ be the corresponding BGF with q marking n and z marking m. Set $C(q) = C(q,1)$. The question is to determine the total number of coin fountains of area n, $[q^n]C(q)$. The series starts as (this is *EIS* **A005169**)

$$C(q) = 1 + q + q^2 + 2q^3 + 3q^4 + 5q^5 + 9q^6 + 15q^7 + 26q^8 + \cdots,$$

as results from inspection of the first few cases.



The function $C(q)$ is *a priori* meromorphic in $|q| < 1$. From the bijection with Dyck paths and area, one finds

$$C(q) = \cfrac{1}{1 - \cfrac{q}{1 - \cfrac{q^2}{1 - \cfrac{q^3}{\ddots}}}}.$$

The identity (50) implies

$$C(q) = \frac{E(q,q)}{E(1,q)}.$$

An exponential lower bound of the form 1.6^n holds on $[q^n]C(q)$, since $(1-q)/(1-q-q^2)$ is dominated by $C(q)$ for $q > 0$. At the same time, the number $[q^n]C(q)$ is majorized by the number of compositions, which is 2^{n-1}. Thus, the radius of convergence of $C(q)$ has to lie somewhere between 0.5 and $0.61803\ldots$. It is then easy to check by numerical analysis the existence of a simple zero of the denominator, $E(-1,q)$, near $\rho \doteq 0.57614$. Routine computations based on Rouché's theorem then makes it possible to verify formally that ρ is the only simple pole in $|q| < 3/5$ (the process is detailed in [**89**]). Thus, singularity analysis of meromorphic functions applies:

Objects	*Weights* $(\alpha_j, \beta_j \gamma_j)$	*Counting*	*Orth. pol.*
Simple paths	$1, 1, 0$	Catalan #	Chebyshev
Permutations	$j+1, j, 2j+1$	Factorial #	Laguerre
Alternating perm.	$j+1, j, 0$	Secant #	Meixner
Involutions	$1, j, 0$	Odd factorial #	Hermite
Set partition	$1, j, j+1$	Bell #	Poisson-Charlier
Nonoverlap. set part.	$1, 1, j+1$	Bessel #	Lommel

FIGURE 6. Some special families of combinatorial objects together with corresponding weights, moments, and orthogonal polynomials.

PROPOSITION V.5. *The number of coin fountains made of n coins satisfies asymptotically*

$$[q^n]C(q) = cA^n + O((5/3)^n), \quad c \doteq 0.31236, \quad A = \rho^{-1} \doteq 1.73566.$$

This example illustrates the power of modelling by continued fractions as well as the smooth articulation with meromorphic function asymptotics. □

The systematic theory of lattice path enumerations and continued fractions was developed initially because of the need to count weighted lattice paths, notably in the context of the analysis of dynamic data structures in computer science [**40**]. In this framework, a system of multiplicative weights $\alpha_j, \beta_j, \gamma_j$ is associated with the steps a_j, b_j, c_j, each weight being an integer that represents a number of "possibilities" for the corresponding step type. A system of weighted lattice paths has counting generating functions given by an easy specialization of the corresponding multivariate expressions we have just developed, namely,

$$a_j \mapsto \alpha_j z, \quad b_j \mapsto \beta_j z, \quad c_j \mapsto \gamma_j z, \tag{52}$$

where z marks the length of paths. One can then sometimes solve an enumeration problem expressible in this way by reverse-engineering the known collection of continued fractions as found in a reference book like Wall's treatise [**113**]. Next, for general reasons, the polynomials P, Q are always elementary variants of a family of orthogonal polynomials that is determined by the weights [**22, 36, 107**]. When the multiplicities have enough structural regularity, the weighted lattice paths are likely to correspond to classical combinatorial objects and to classical families of orthogonal polynomials; see [**36, 40, 58, 60**] and Figure 6 for an outline. We illustrate this by a simple example due to Lagarias, Odlyzko, and Zagier [**79**].

EXAMPLE 8. *Interconnection networks and involutions.* The problem considered here was introduced by Lagarias, Odlyzko, and Zagier in [**79**]: *There are $2n$ points on a line, with n point-to-point connections between pairs of points. What is the probable behaviour of the* width *of such an interconnection network?* Imagine the points to be $1, \ldots, 2n$, the connections as circular arcs between points, and let a vertical line sweep from left to right; width is defined as the maximum number of edges encountered by such a line. One may freely imagine a tunnel of fixed capacity (this corresponds to the width) inside which wires can be placed to connect points pairwise. See Figure 7.

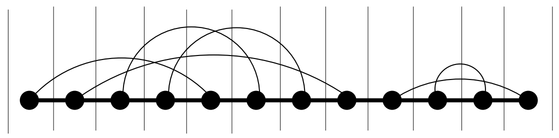


FIGURE 7. An interconnection network on $2n = 12$ points.

Let $\mathcal{I}_{2n}$ be the class of all interconnection networks on $2n$ points, which is precisely the collection of ways of grouping $2n$ elements into n pairs, or, equivalently, the class of all involutions (i.e., permutations with cycles of length 2 only). The number I_{2n} equals the "odd factorial",

$$I_{2n} = 1 \cdot 3 \cdot 5 \cdots (2n-1),$$

whose EGF is $e^{z^2/2}$ (see Chapter 2). The problem calls for determining the quantity $I_{2n}^{[h]}$ that is the number of networks corresponding to a width $\leq h$.

The relation to lattice paths is as follows. First, when sweeping a vertical line across a network, define an active arc at an abscissa as one that straddles that abscissa. Then build the sequence of active arcs counts at half-integer positions $\frac{1}{2}, \frac{3}{2}, \ldots, 2n - \frac{1}{2}, 2n + \frac{1}{2}$. This constitutes a sequence of integers where each member is ± 1 the previous one, that is, a lattice path without level steps. In other words, there is an ascent in the lattice path for each element that is smaller in its cycle and a descent otherwise. One may view ascents as associated to situations where a node "opens" a new cycle, while descents correspond to "closing" a cycle.

Involutions are much more numerous than lattice paths, so that the correspondence from involutions to lattice paths is many-to-one. However, one can easily enrich lattice paths, so that the enriched objects are in one-to-one correspondence with involutions. Consider again a scanning position at a half-integer where the vertical line crosses ℓ (active) arcs. If the next node is of the closing type, there are ℓ possibilities to choose from. If the next node is of the opening type, then there is only one possibility, namely, to start a new cycle. A complete encoding of a network is obtained by recording additionally the sequence of the n possible choices corresponding to descents in the lattice path (some canonical order is fixed, for instance, oldest first). If we write these choices as superscripts, this means that the set of all enriched encodings of networks is obtained from the set of standard lattice path encodings by effecting the substitutions

$$b_j \mapsto \sum_{k=1}^{j} b_j^{(k)}.$$

The OGF of all involutions is obtained from the generic continued fraction of Theorem V.2 by the substitution

$$a_j \mapsto z, \quad b_j \mapsto j\, z,$$

where z records the number of steps in the enriched lattice path, or equivalently, the number of nodes in the network. In other words, we have obtained combinatorially a formal

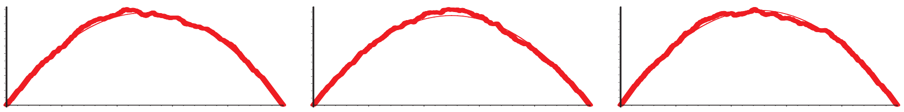


FIGURE 8. Three simulations of random networks with $2n = 1000$ illustrate the tendency of the profile to conform to a parabola with height close to $n/2 = 250$.

continued fraction representation,

$$\sum_{n=0}^{\infty}(1\cdot 3\cdots(2n-1))z^{2n} = \cfrac{1}{1-\cfrac{1\cdot z^2}{1-\cfrac{2\cdot z^2}{1-\cfrac{3\cdot z^2}{\ddots}}}},$$

which was originally discovered by Gauß [**113**]. Theorem V.2 then gives immediately the OGF of involutions of width at most h as a quotient of polynomials. Define

$$I^{[h]}(z) := \sum_{n\geq 0} I^{[h]}_{2n} z^{2n}.$$

One has

$$I^{[h]}(z) = \cfrac{1}{1-\cfrac{1\cdot z^2}{1-\cfrac{2\cdot z^2}{\cfrac{\ddots}{1-h\cdot z^2}}}} = \frac{P_h(z)}{Q_h(z)}$$

where P_h and Q_h satisfy the recurrence

$$Y_{h+1} = Y_h - hz^2 Y_{h-1}.$$

The polynomials are readily determined by their generating functions that satisfies a first-order linear differential equation reflecting the recurrence. In this way, the denominator polynomials are identified to be reciprocals of the Hermite polynomials,

$$Q_h(z) = (z/2)^h H_h(\frac{1}{2z}),$$

themselves defined classically [**1**, Ch. 22] as orthogonal with respect to the measure $e^{-x^2}\,dx$ on $(-\infty,\infty)$ and expressible via

$$H_m(x) = \sum_{m=0}^{\lfloor m/2\rfloor} \frac{(-1)^j m!}{j!(m-2j)!}(2x)^{m-2j}, \qquad \sum_{m\geq 0} H_m(x)\frac{t^m}{m!} = e^{xt-t^2}.$$

In particular, one finds

$$I^{[0]} = 1, \quad I^{[1]} = \frac{1}{1-z^2}, \quad I^{[2]} = \frac{1-2z^2}{1-3z^2}, \quad I^{[3]} = \frac{1-5z^2}{1-6z^2+3z^4}, \quad \&c.$$

The interesting analysis of the dominant poles of the rational GF's, for any fixed h, is discussed in the paper [**79**]. Furthermore, simulations strongly suggest that the width of a random interconnection network on $2n$ nodes is tightly concentrated around $n/2$; see Figure 8. Louchard [**82**] succeeded in proving this fact and a good deal more: With high probability, the profile (the profile is defined here as the number of active arcs as time evolves) of a random network conforms asymptotically to a deterministic parabola $2nx(1-x/(2*n))$ to which are superimposed random fluctuations of amplitude only $O(\sqrt{n})$ well-characterized by a Gaussian process. In particular, *the width of a random network of* $2n$ *nodes converges in probability to* $\frac{n}{2}$. $\square$

V. 3. The supercritical sequence and its applications

We have seen earlier in this section that surjections and alignments with EGFs

$$\frac{1}{2-\exp(z)}, \qquad \frac{1}{1-\log(1-z)^{-1}}$$

have coefficients that satisfy simple asymptotic estimates of the form $C \cdot A^n$. A similar property holds for integer compositions, where there is even an exact counting formula, namely, 2^{n-1}. The common feature of these examples is that that they all involve a sequence construction in their specification and correspond to the schema $\mathcal{F} = \mathfrak{S}(\mathcal{G})$, in either the labelled or the unlabelled case.

We thus consider a sequence construction $\mathcal{F} = \mathfrak{S}(\mathcal{G})$, with the associated GFs (either ordinary or exponential) satisfying the usual relation

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

and $G(0)=0$ for well-foundedness. We shall write $f_n = [z^n]F(z)$ and $g_n = [z^n]G(z)$. We also restrict attention to the case where the radius of convergence of G is nonzero, in which case, the radius of convergence of F is also nonzero by virtue of closure properties of analytic functions. We set:

DEFINITION V.2. *Let F, G be GFs with nonnegative coefficients that are analytic at 0, with $G(0)=0$. The schema $F(z) = (1-G(z))^{-1}$ is said to be* supercritical *if $G(\rho) > 1$, where $\rho = \rho_G$ is the radius of convergence of G.*

Note that $G(\rho)$ is well defined as the limit $\lim_{x\to\rho^-} G(x)$ since $G(x)$ increases along the positive real axis. (The value $G(\rho)$ corresponds to what has been denoted earlier by τ_G when discussing "signatures" in Section IV. 3.3.) We assume that $G(z)$ is *unperiodic* in the sense that there does not exist an integer $d \geq 2$ such that $G(z) = h(z^d)$ for some h analytic at 0. (This normalization is merely a convenience that entails no loss of generality.) One has

THEOREM V.3 (Supercritical sequence asymptotics). *Let the schema $F = (1-G)^{-1}$ be supercritical and assume that G is unperiodic. Then, one has*

$$[z^n]F(z) = \frac{1}{\sigma G'(\sigma)} \cdot \sigma^{-n}\left(1 + O(A^n)\right),$$

where σ is the root in $(0, \rho_G)$ of $G(\sigma) = 1$, and A is a number less than 1. The number X of $\mathcal{G}$–components in a random $\mathcal{C}$–structure of size n has mean and variance satisfying

$$\begin{aligned}
\mathbb{E}_n(X) &= \frac{1}{\rho G'(\rho)} \cdot (n+1) - 1 + \frac{G''(\rho)}{G'(\rho)^2} + O(A^n) \\
\mathbb{V}_n(X) &= \frac{\rho G''(\rho) + G'(\rho) - \rho G'(\rho)^2}{\rho^2 G'(\rho)^3} \cdot n + O(1).
\end{aligned}$$

In particular, the distribution is concentrated.

PROOF [**52, 104**]. The basic observation is that G increases continuously from $G(0) = 0$ to $G(\rho_G) = \tau_G$ (with $\tau_G > 1$ by assumption) when x increases from 0 to ρ_G. Therefore, the positive number σ, which satisfies $G(\sigma) = 1$ is well defined. Then, F is analytic at all points of the interval $(0, \sigma)$. The function G being analytic at σ, satisfies, in a neighbourhood of σ

$$G(z) = 1 + G'(\sigma)(z - \sigma) + \frac{1}{2!}G''(\sigma)(z - \sigma)^2 + \cdots .$$

so that $F(z)$ has a pole at $z = \sigma$; also, this pole is simple since $G'(\rho) > 0$. Pringsheim's theorem then implies that the radius of convergence of F must coincide with σ.

There remains to show that $F(z)$ is meromorphic in a disc of some radius $R > \sigma$ with the point σ as the only singularity inside the disc. This results from the assumption that G is unperiodic. In effect, one has $G(\sigma e^{i\theta}) \le 1$ for all θ by the triangular inequality. It suffices to verify that $G(\sigma e^{i\theta}) \ne 1$ for $\theta \in [-\pi, \pi] \setminus \{0\}$ to ensure that F is analytic at points of the circle $|z| = \sigma$, with the sole exception of σ. A contrario, $G(\sigma e^{i\theta}) = 1$ would imply, by the converse of the triangle inequality that

$$g_n \sigma^n e^{in\theta} = g_n \sigma^n,$$

for all values of n such that $g_n \ne 0$. This in turn is only possible if there is a root of unity, $\omega = e^{2i\pi/d}$, such that $\omega^n = 1$ whenever $g_n \ne 1$. This last fact is itself incompatible with the assumption that $G(z)$ is unperiodic.

In summary, $F(z)$ has a simple pole at $z = \sigma$ and is otherwise analytic at all points of $|z| = \sigma$. Thus, by compactness, there exists a disc of radius $R > \sigma$ in which F is analytic except for a unique pole at σ. Take r such that $\sigma < r < R$ and apply the main theorem of meromorphic function asymptotics to deduce the stated formula with $A = \sigma/r$.

Consider next the number of $\mathcal{G}$-components in a random $\mathcal{F}$ structure of size n. Bivariate generating functions give access to the expectation of this random variable:

$$\begin{aligned} \mathbb{E}_n(X) &= \frac{1}{f_n}[z^n] \left. \frac{\partial}{\partial u} \frac{1}{1 - uG(z)} \right|_{u=1} \\ &= \frac{1}{f_n}[z^n] \frac{G(z)}{(1 - G(z))^2}. \end{aligned}$$

The problem is now reduced to extracting coefficients in a univariate generating function with a double pole at $z = \rho$, and it suffices to expand the GF locally at ρ. The variance calculation is similar though it involves a triple pole. □

When a sequence construction is supercritical, the number of components is in the mean $\asymp n$ while its standard deviation is $\asymp \sqrt{n}$. Thus, the distribution is concentrated (see Chapter III). In fact, there results from a general theorem of Bender [**8**] that the distribution of the number of components is asymptotically Gaussian; see later chapters for details.

Direct cases of application to combinatorial generating functions are

$$a_1(z) = \frac{z}{1 - z},\ a_2(z) = e^z - 1,\ a_3(z) = \log(1 - z)^{-1},$$

corresponding respectively to integer compositions (OGF), surjections (EGF), and alignments. Thus:

- The expected number of summands in a random composition of the integer n is $\sim \frac{n+1}{2}$, with variance $\sim \frac{n}{4}$.
- The expected cardinality of the range of a random surjection whose domain has cardinality n is asymptotic to βn with $\beta = 1/(2\log 2)$;
- The expected number of components in a random alignment of size n is asymptotic to $n/(e - 1)$.

10	16	1***5***
20	732	73***4***
30	36039	360***57***
40	1772207	17722***61***
50	87109263	871092***48***
60	4281550047	42815***49331***
70	210444532770	21044453***0095***
80	10343662267187	1034366226***5182***
90	508406414757253	5084064147***81706***
100	24988932929490838	24988932929***612479***

FIGURE 9. The pyramid relative to compositions into prime summands for $n = 10\,..\,100$: (left: exact values; right: asymptotic formula rounded).

EXAMPLE 9. *Compositions with restricted summands, compositions into primes.* Unrestricted integer compositions are well understood as regards enumeration: their number is exactly $C_n = 2^{n-1}$, their OGF is $C(z) = (1-z)/(1-z)$, and compositions with k summands are enumerated by binomial coefficients. Such simple exact formulæ disappear when restricted compositions are considered, but, as we now show, asymptotics is much more robust to changes in specifications.

Let S be a subset of the integers $\mathbb{Z}_{\geq 1}$ such that $\gcd(S) = 1$, i.e., not all members of S are multiples of a common divisor $d \geq 2$. In order to avoid trivialities, we also assume that $S \neq \{1\}$. The class $\mathcal{C}^S$ of compositions with summands constrained to the set S then satisfies:

$$\textit{Specification}: \quad \mathcal{C}^S = \mathfrak{S}(\mathfrak{S}_S(\mathcal{Z}));$$
$$\textit{OGF}: \quad D(z) = \frac{1}{1-S(z)}, \qquad S(z) = \sum_{s \in S} z^s.$$

By assumption, $S(z)$ is unperiodic, so that Theorem V.3 applies directly. There is a well-defined number σ such that

$$S(\sigma) = 1, \qquad 0 < \sigma < 1,$$

and the number of S–restricted compositions satisfies

$$C_n^S := [z^n]C^S(z) = \frac{1}{\sigma S'(\sigma)} \cdot \sigma^{-n}\left(1 + O(A^n)\right). \tag{53}$$

Amongst the already discussed cases, $S = \{1, 2\}$ gives rise to Fibonacci numbers and, more generally, $S = \{1, \ldots, r\}$ corresponds to partitions with summands at most r. In this case, the OGF,

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

is a simple variant of the OGF associated to longest runs in strings. The treatment of the latter can be copied almost verbatim to the effect that the largest component in a random composition of n is found to be $\lg n + O(1)$, both on average and with high probability.

Here is a surprising application of the general theory. Consider the case where S is taken to be the set of prime numbers, $\mathrm{Prime} = \{2, 3, 5, 7, 11, \ldots\}$, thereby defining the class of *compositions into prime summands*. The sequence starts as

$$1, 0, 1, 1, 1, 3, 2, 6, 6, 10, 16, 20, 35, 46, 72, 105,$$

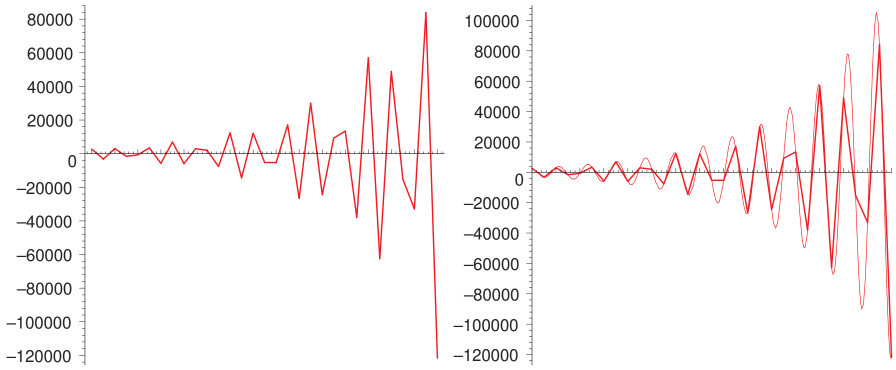


FIGURE 10. Errors in the approximation of the number of compositions into primes for $n = 70\,.\,.\,100$: left, the values of $C_n^{\rm Prime} - g(n)$; right, the correction $g_2(n)$ arising from the next two poles, which are complex conjugate and the continuous extrapolation of this approximation.

corresponding to the OGF $1+z^2+\cdots$, and is *EIS* **A023360** in Sloane's encyclopedia. The formula (53) applies to provide the asymptotic form of the number of such compositions. It is also well worth noting that the constants appearing in (53) are easily determined to great numerical precision, as we now explain.

By (53) and the preceding equation, the dominant singularity of the OGF of compositions into prime is the positive root $\sigma < 1$ of the characteristic equation

$$S(z) \equiv \sum_{p \ \rm Prime} z^p = 1.$$

Fix a threshold value m_0 (for instance $m_0 = 10$ or 100) and introduce the two series

$$S^-(z) := \sum_{s\in S,\ s<m_0} z^s, \quad S^+(z) := \left(\sum_{s\in S,\ s<m_0} z^s\right) + \frac{z^{m_0}}{1-z}.$$

Clearly, for $x \in (0,1)$, one has $S^-(x) < S(x) < S^+(x)$. Define two constants σ^-, σ^+ by the conditions

$$S^-(\sigma^-) = 1, \quad S^+(\sigma^+) = 1, \qquad 0 < \sigma^-, \sigma^+ < 1.$$

These constants are algebraic numbers that are accessible to computation. At the same time, they satisfy $\sigma^+ < \sigma < \sigma^-$. As the order of truncation, m_0, increases, the values of σ^+, σ^- are expected to provide better and better approximations to σ together with an interval in which σ provably lies. For instance, $m_0 = 10$ is enough to determine that $0.66 < \sigma < 0.69$, and the choice $m_0 = 100$ gives σ to 15 guaranteed digits of accuracy, namely, $\sigma \doteq 0.67740\,17761\,30660$. Then, the asymptotic formula (53) instantiates as

$$C_n^{\rm Prime} \sim g(n), \qquad g(n) := 0.30365\,52633 \cdot 1.47622\,87836^n. \tag{54}$$

The constant $\sigma^{-1} \doteq 1.47622$ is akin to the family of Backhouse constants described in [**35**].

Once more, the asymptotic approximation is very good as shown by the pyramid of Figure 9. The difference between C_n^{Prime} and its approximation $g(n)$ from Eq. (54) is plotted on the left of Figure 10. The seemingly haphazard oscillations that manifest themselves are well explained by the principles discussed in the previous section. It appears that the next poles of the OGF are complex conjugate and lie near $-0.76 \pm 0.44i$, having modulus about 0.88. The corresponding residues then jointly contribute a quantity of the form

$$g_2(n) = c \cdot A^n \sin(\omega n + \omega_0), \qquad A \doteq 1.13290,$$

for some constants c, ω, ω_0. Comparing the left and right parts of Figure 10 shows this next layer of poles to explain quite well the residual error $C_n^{\mathrm{Prime}} - g(n)$. (The diagram on the right in Figure 10 also displays the values of the continuous interpolation to $g_2(n)$.)

Here is a final example that demonstrates in a striking way the scope of the method. Define the set Prime_2 of "twinned primes" as the set of primes that belong to a twin prime pair, that is, $p \in \mathrm{Prime}_2$ if one of $p-2, p+2$ is prime. The set Prime_2 starts as $3, 5, 7, 11, 13, 17, 19, 29, 31, \ldots$ (numbers like 23 or 37 are thus excluded). The asymptotic formula for the number of compositions of the integer n into summands that are twinned primes, is

$$C_n^{\mathrm{Prime}_2} \sim 0.18937 \cdot 1.29799^n.$$

It is quite remarkable that the constants involved are still computable real numbers (and of low complexity, even), this despite the fact that it is not known whether the set of twinned primes is finite or infinite. Incidentally, a sequence that starts like $C_n^{\mathrm{Prime}_2}$,

$$1, 0, 0, 1, 0, 1, 1, 1, 2, 1, 3, 4, 3, 7, 7, 8, 14, 15, 21, 28, 33, 47, 58, \ldots$$

and coincides till index 22 included (!), but not beyond, was encountered by P. A. MacMahon[3], as the authors discovered, much to their astonishment, from scanning Sloane's Encyclopedia, where it appears as *EIS* **A002124**. □

Profiles of supercritical sequences. We have seen in Chapter III that integer compositions and integer partitions, when sampled at random, tend to assume rather different aspects. Given a sequence construction, $\mathcal{F} = \mathfrak{S}(\mathcal{G})$, the profile of an element $\alpha \in \mathcal{F}$ is the vector $(X^{\langle 1 \rangle}, X^{\langle 2 \rangle}, \ldots)$ where $X^{\langle 1 \rangle}(\alpha)$ is the number of $\mathcal{G}$–components in α that have size j. In the case of (unrestricted) integer compositions, it could be proved elementarily that, on average and for size n, the number of 1 summands is $\sim n/2$, the number of 2 summands is $\sim n/4$, and so on. Now that meromorphic asymptotic is available, such a property can be placed in a much wider perspective.

PROPOSITION V.6. *Consider a supercritical sequence construction, $\mathcal{F} = \mathfrak{S}(\mathcal{G})$, with the "unperiodic" condition. The number of $\mathcal{G}$–components of any fixed size k in a random $\mathcal{F}$–object of size n satisfies*

$$\mathbb{E}_n(X^{\langle k \rangle}) = \frac{g_k \sigma^k}{\sigma G'(\sigma)} n + O(1), \quad \mathbb{V}_n(X^{\langle k \rangle}) = O(n). \tag{55}$$

There, σ is the root in $(0, \rho_G)$ of $G(\sigma) = 1$, and $g_k = [z^k]G(z)$.

PROOF. The bivariate GF with u marking the number of $\mathcal{G}$–components of size k is

$$F(z, u) = \frac{1}{1 - (G(z) + (u-1) g_k z^k)},$$

[3]See "Properties of prime numbers deduced from the calculus of symmetric functions", *Proc. London Math. Soc.*, 23 (1923), 290-316). MacMahon's sequence corresponds to the compositions into arbitrary *odd* primes.

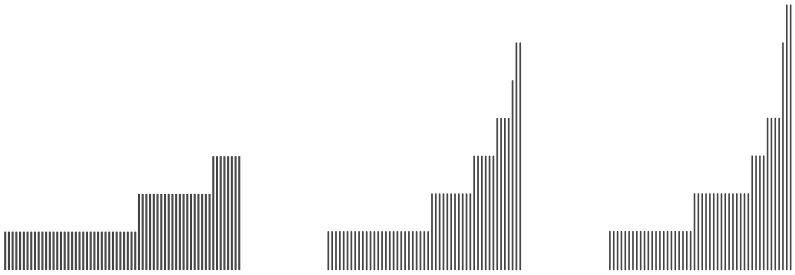


FIGURE 11. Profile of structures drawn at random represented by the sizes of their components in sorted order: (from left to right) a random surjection, alignment, and composition of size $n = 100$.

as results from the theory developed in Chapter III. The mean value is then given by a quotient,

$$\mathbb{E}_n(X^{\langle k\rangle}) = \frac{1}{f_n}\,[z^n]\,\frac{\partial}{\partial u}F(z,u)\bigg|_{u=1} = \frac{1}{f_n}[z^n]\frac{g_k z^k}{(1-G(z))^2}.$$

The GF of cumulated values has a double pole at $z = \sigma$, and the estimate of the mean value follows. The variance is estimated similarly, after two successive differentiations and the analysis of a triple polar singularity. □

The total number of components X satisfies $X = \sum X^{\langle k\rangle}$, and, by Theorem V.3, its mean is asymptotic to $n/(\sigma G'(\sigma))$. Thus, Equation (55) indicates that, at least in some average-value sense, the "proportion" of components of size k amongst all components is given by $g_k\sigma^k$. Also, since $G(\sigma) = 1$, the coefficients $g_k\sigma^k$ add up to 1.

EXAMPLE 10. *The profiles of compositions, surjections, and alignments.* Proposition V.6 immediately applies to compositions (that are sequences of sequences), surjections (sequences of sets), and alignments (sequences of cycles). The following table summarizes the conclusions:

Structures	*Specif.*	*Law* $(g_k\sigma^k)$	*Type*	σ
Compositions	$\mathfrak{S}(\mathfrak{S}_{\geq 1}(\mathcal{Z}))$	$\frac{1}{2^k}$	Geometric	$\frac{1}{2}$
Surjections	$\mathfrak{S}(\mathfrak{P}_{\geq 1}(\mathcal{Z}))$	$\frac{1}{k!}(\log 2)^k$	Poisson	$\log 2$
Alignments	$\mathfrak{S}(\mathfrak{C}(\mathcal{Z}))$	$\frac{1}{k}(1-e^{-1})^k$	Logarithmic	$1-e^{-1}$

The geometric and Poisson law are classical; the *logarithmic distribution* (also called "logarithmic-series distribution") of parameter λ is by definition the law of a random variable Y such that

$$\mathbb{P}(Y = k) = \frac{1}{\log(1-\lambda)^{-1}}\frac{\lambda^k}{k}.$$

The way the internal construction induces the law of component sizes,

$$\text{Sequence} \mapsto \text{Geometric}; \quad \text{Set} \mapsto \text{Poisson}; \quad \text{Cycle} \mapsto \text{Logarithmic},$$

stands out. Figure 11 exemplifies the phenomenon by displaying components sorted by size and represented by vertical segments of corresponding lengths for three randomly drawn objects of size $n = 100$. □

$\triangleright$ **9.** *Proportion of k–components and convergence in probability.* For any fixed k, the random variable $X_n^{\langle k \rangle}/X_n$ converges in probability (the notion is defined in Chapter III) to the value $g_k\sigma^k$,

$$\frac{X_n^{\langle k \rangle}}{X_n} \xrightarrow{P} g_k\sigma^k, \quad \text{i.e.,} \quad \lim_{n\to\infty} \mathbb{P}\left\{ g_k\sigma^k(1-\epsilon) \le \frac{X_n^{\langle k \rangle}}{X_n} \le g_k\sigma^k(1+\epsilon) \right\} = 1,$$

for any $\epsilon > 0$. The proof is an easy consequence of the Chebyshev inequalities (the distributions of X_n and $X_n^{\langle k \rangle}$ are both concentrated). $\triangleleft$

$\triangleright$ **10.** *Random generation of supercritical sequences.* Let $\mathcal{F} = \mathfrak{S}(\mathcal{G})$ be a supercritical sequence scheme. Consider a sequence of i.i.d. (independently identically distributed) random variables $Y_1, Y_2, \ldots$ each of them obeying the discrete law

$$\mathbb{P}(Y = k) = g_k\sigma^k, \qquad k \ge 1.$$

A sequence is said to be hitting n if $Y_1 + \cdots + Y_r = n$ for some $r \ge 1$. The vector $(Y_1, \ldots, Y_r)$ for a sequence conditioned to hit n has the same distribution as the sequence of the lengths of components in a random $\mathcal{F}$–object of size n.

For probabilists, this explains the shape of the formulæ in Theorem V.3, which resemble renewal relations [**34**, Sec. XIII.10]. It also implies that, given a uniform random generator for $\mathcal{G}$–objects, one can generate a random $\mathcal{F}$–object of size n in $O(n)$ steps on average [**31**]. This applies to surjections, alignments, and compositions in particular. $\triangleleft$

$\triangleright$ **11.** *Largest components in supercritical sequences.* Let $\mathcal{F} = \mathfrak{S}(\mathcal{G})$ be a supercritical sequence. Assume that $g_k = [z^k]G(z)$ satisfies the asymptotic "smoothness" condition

$$g_k \underset{k\to\infty}{\sim} c\rho^{-k}k^{\beta}, \qquad c, \rho \in \mathbb{R}_{>0},\ \beta \in \mathbb{R}.$$

Then the size L of the largest $\mathcal{G}$ component in a random $\mathcal{F}$ object satisfies, for size n,

$$\mathbb{E}_{\mathcal{F}_n}(X) = \frac{1}{\log(\rho/\sigma)}\left(\log n + \beta \log\log n\right) + o(\log\log n).$$

This covers integer compositions ($\rho = 1$, $\beta = 0$) and alignments ($\rho = 1$, $\beta = -1$). [The analysis generalizes the case of longest runs in Example 2 and is based on similar principles. The GF of $\mathcal{F}$ objects with $L \le m$ is $F^{\langle m \rangle}(z) = \left(1 - \sum_{k\le m} g_k z^k\right)^{-1}$, according to Section III.7. For m large enough, this has a dominant singularity which is a simple pole at σ_m such that $\sigma_m - \sigma \sim c_1(\sigma/\rho)^m m^{\beta}$. There follows a double-exponential approximation

$$\mathbb{P}_{\mathcal{F}_n}(L \le m) \approx \exp\left(-c_2 n m^{\beta}(\sigma/\rho)^m\right)$$

in the "central" region. See Gourdon's study [**61**] for details.] $\triangleleft$

V. 4. Functional equations: positive rational systems

For rational functions, positivity coupled with some simple ancillary conditions entails a host of important properties, like unicity of the dominant singularity. Such facts result from the classical Perron-Frobenius theory of nonnegative matrices that we summarize in this section. They in turn imply strong properties of large random structures.

The basic case is that of a d-dimensional column vector$\mathrm{y}(z)$ of generating functions satisfying a linear system of the form

$$\mathrm{y}(z) = \mathrm{a} + zT\,\mathrm{y}(z),$$

for some $(d \times d)$ matrix T and vector a. If T satisfies suitable positivity conditions and a is nonnegative, then any component $\mathrm{y}_j(z)$ closely resembles the extremely simple rational function,

$$\frac{1}{1 - \lambda_1 z},$$

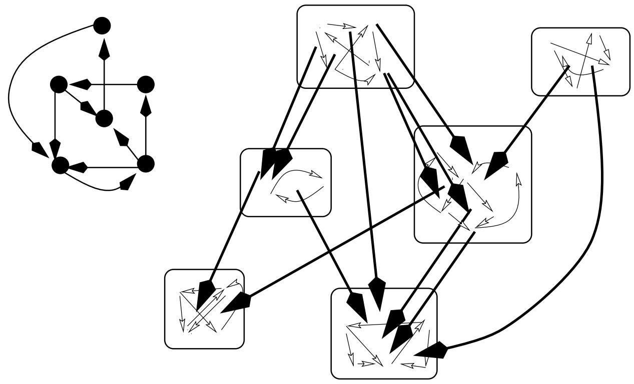


FIGURE 12. The irreducibility conditions of Perron-Frobenius theory. Left: a strongly connected digraph. Right: a weakly connected digraph that is not strongly connected is a collection of strongly connected components related by a directed acyclic graph.

where λ_1 is a well-characterized eigenvalue of T. Accordingly, the asymptotic phenomena associated with such systems are highly predictable. We propose to expose here the general theory and treat in the next section classical applications to paths in graphs and to languages recognized by finite automata.

V. 4.1. Perron-Frobenius theory of nonnegative matrices. For an arbitrary square matrix $A \in \mathbb{R}^{m\times m}$, the *spectrum* is the set of its *eigenvalues*, that is, the set of λ such that $\lambda I - A$ is not invertible (i.e., not of full rank), where I is the unit matrix with the appropriate dimension. A *dominant eigenvalue* is one of largest modulus. Finally, the spectral radius of an arbitrary matrix A is defined as

$$\sigma(A) = \max_j \{|\lambda_j|\}, \tag{56}$$

where the set $\{\lambda_j\}$ is the set of eigenvalues of A (also called spectrum). The spectral radius $\sigma(A)$ describes growth properties associated to the powers of A. Indeed, given the Jordan normal form of matrices, it is easy to see that all entries of A^n are bounded from above by a multiple of $\sigma(A)^n \cdot n^{r-1}$, where r is the maximum multiplicity of any dominant eigenvalue. When analysing a family of combinatorial models that admit a matrix formulation, it is then of obvious interest to determine the value of the spectral radius and the multiplicities attached to dominant eigenvalues.

The properties of positive and of nonnegative matrices have been superbly elicited by Perron [**96**] in 1907 and by Frobenius [**55**] in 1908–1912. The corresponding theory has far-reaching implications: it lies at the basis of the theory of finite Markov chains and it extends to positive operators in infinite-dimensional spaces [**77**].

For A a scalar matrix of dimension $m \times m$ with nonnegative entries, a crucial rôle is played by the *dependency graph*; this is the (directed) graph with vertex set $V = \{1\,..\,m\}$ and edge set containing the directed edge $(a \to b)$ iff $A_{a,b} \neq 0$. The reason for this terminology is the following: Let A represent the linear transformation $\left\{y_i^\star = \sum_j A_{i,j} y_j\right\}_i$; then, the fact that an entry $A_{i,j}$ is nonzero means that $y_i^\star$ depends effectively on y_j and is translated by the directed edge $(i \to j)$ in the dependency graph.

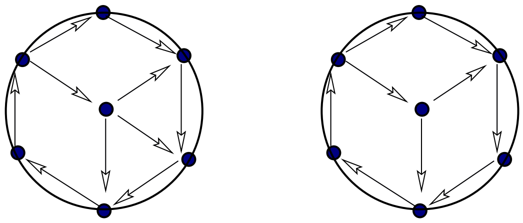


FIGURE 13. The aperiodicity conditions of Perron-Frobenius theory: an aperiodic digraph (left) and a periodic digraph (right).

From this point on, we consider matrices with nonnegative entries. Two notions are essential, irreducibility and aperiodicity (the terms are borrowed from Markov chain theory and matrix theory).

DEFINITION V.3. *The matrix A is called* irreducible *if its dependency graph is strongly connected (i.e., any two vertices are connected by a directed path). A strongly connected digraph G is* periodic *with parameter d iff all its cycles have a length that is a multiple of d. In that case, the graph decomposes into cyclically arranged layers: the vertex set V can be partitioned into d classes, $V = V_0 \cup \cdots \cup V_{d-1}$, in such a way that the edge set E satisfies*

$$E \subseteq \bigcup_{i=0}^{d-1} (V_i \times V_{(i+1) \bmod d}). \tag{57}$$

The maximal possible d is called the period*. If no decomposition exists with $d \geq 2$, so that the period has the trivial value 1, then the graph and all the matrices that admit it as their dependency graph are called* aperiodic*.*

By considering only simple paths, it is then seen that irreducibility is equivalent to the condition that $(I + A)^m$ has all its entries that are strictly positive. See Figure 12 for a graphical rendering of irreducibility and for the general structure of a (weakly connected) digraph. As an illustration of periodicity, a directed 10-cycle is periodic with parameter $d = 1, 2, 5, 10$ and the period is 10. See Figure 13 for representations of a periodic and an aperiodic digraph.

Periodicity also means that the existence of paths of length n between any given pair of nodes $\langle i, j\rangle$ is constrained by the congruence class $n \bmod d$. A contrario, aperiodicity entails the existence, for all n sufficiently large, of paths of length n connecting $\langle i, j\rangle$. From the definition, a matrix A with period d has, up to simultaneous permutation of its rows and columns, a cyclic block structure

$$\begin{pmatrix} 0 & \boxed{A_{0,1}} & 0 & \cdots & 0 \\ 0 & 0 & \boxed{A_{1,2}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \boxed{A_{d-2,d-1}} \\ \boxed{A_{d-1,0}} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

where the blocks $A_{i,i+1}$ are reflexes of the connectivity between V_i and V_{i+1} in (57). ???

THEOREM V.4 (Perron-Frobenius theorem). *Let A be a matrix that is assumed to be irreducible in the sense that its dependency graph is strongly connected.*

(i) *If A has (strictly)* positive elements, *then its eigenvalues can be ordered in such a way that*

$$\lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \cdots,$$

i.e., A has a unique dominant eigenvalue which is positive and simple.

(ii) *If A has* nonnegative elements, *then its eigenvalues can be ordered in such a way that*

$$\lambda_1 = |\lambda_2| = \cdots = |\lambda_d| > |\lambda_{d+1}| \geq |\lambda_{d+2}| \geq \cdots,$$

and each of the dominant eigenvalues is simple, with λ_1 being positive. Furthermore, the quantity d is precisely equal to the period of the dependency graph. If $d = 1$, in particular, then there is unicity of the dominant eigenvalue. If $d \geq 2$, the whole spectrum is invariant under the set of transformations

$$\lambda \mapsto \lambda e^{2ij\pi/d}, \qquad j = 0, 1, \ldots, d-1.$$

For proof techniques including a full proof of Part (i) of the theorem, see APPENDIX: *Perron-Frobenius theory of nonnegative matrices*, p. 125.

For short, one says that a matrix is positive (resp. nonnegative) if all its elements are positive (resp. nonnegative). Here are two useful turnkey results, Corollaries V.1 and V.2.

COROLLARY V.1. *Any one of the following conditions suffices to guarantee the existence of a unique dominant eigenvalue of a nonnegative matrix T:*

(i) *T has (strictly) positive entries;*
(ii) *T is such that, some power T^s is (strictly) positive;*
(iii) *T is irreducible and at least one diagonal element of T is nonzero;*
(iv) *T is irreducible and the dependency graph of T is such that there exist at least two paths from the same source to the same destination that are of relatively prime lengths.*

PROOF. The proof makes use of the well-known correspondence between terms in coefficients of matrix products and paths in graphs (see below Section V. 5 for more). Sufficiency of condition (i) results directly from Case (i) of Theorem V.4. Condition (ii) immediately implies irreducibility. Unicity of the dominant eigenvalue (hence aperiodicity) results from Perron-Frobenius properties of A^s, by which $\lambda_1^s > |\lambda_2|^s$. (Also, by elementary graph combinatorics, one can always take the exponent s to be at most the dimension m.) By basic combinatorics of paths in graphs, Conditions (iii) and (iv) each imply Condition (ii). □

V. 4.2. Positive rational functions. The importance of Perron-Frobenius theory and of its immediate consequence, Corollary V.1, stems from the fact that uniqueness of the dominant eigenvalue is usually related to a host of analytic properties of generating functions as well as probabilistic properties of structures. In particular, as we shall see in the next section, several combinatorial problems (like automata or paths in graphs) can be reduced to the following case.

COROLLARY V.2. *Consider the matrix*

$$F(z) = (I - zT)^{-1},$$

where T, called the "transition matrix", is a scalar nonnegative matrix. It is assumed that T is irreducible. Then each entry $F_{i,j}(z)$ of $F(z)$ has a radius of convergence ρ that coincides with the smallest positive root of the determinantal equation

$$\Delta(z) := \det(I - zT) = 0.$$

Furthermore, the point ρ is a simple pole of any $F_{i,j}(z)$.

In addition, if T is aperiodic or if it satisfies any of the conditions of Corollary V.1, then all singularities other than ρ are strictly dominated in modulus by ρ.

The statement obviously applies to any positive linear combinations of entries of F and thus to solutions of any system of the form $\mathrm{y}(z) = \mathrm{a} + zT\,\mathrm{y}(z)$.

PROOF. Define first (as in the statement) $\rho = 1/\lambda_1$, where λ_1 is the eigenvalue of T of largest modulus that is guaranteed to be simple by assumption of irreducibility and by Perron-Frobenius properties. Next, the relations induced by $F = I + zTF$, namely,

$$F_{i,j}(z) = \delta_{i,j} + z\sum_k T_{i,k}F_{k,j}(z),$$

together with positivity and irreducibility entail that the $F_{i,j}(z)$ must all have the same radius of convergence r. Indeed, each F_{ij} depends positively on all the other ones (by irreducibility) so that any infinite value of an entry in the system must propagate to all the other ones.

The characteristic polynomial

$$\Delta(z) = \det(I - zT),$$

has roots that are inverses of the eigenvalues of T and $\rho = 1/\lambda_1$ is smallest in modulus. Thus, since Δ is the common denominator to all the $F_{i,j}(z)$, poles of any $F_{i,j}(z)$ can only be included in the set of zeros of this determinant, so that the inequality $r \geq \rho$ holds.

It remains to exclude the possibility $r > \rho$, which means that no "cancellations" with the numerator can occur at $z = \rho$. The argument relies on finding a positive combination of some of the $F_{i,j}$ that *must* be singular at ρ. We offer two proofs, each of interest in its own right: one (a) is conveniently based on the Jacobi trace formula, the other (b) is based on supplementary Perron–Frobenius properties.

(a) Jacobi's trace formula for matrices [**60**, p. 11],

$$\det\circ\exp = \exp\circ\,\mathrm{Tr} \qquad \text{or} \qquad \log\circ\det = \mathrm{Tr}\,\circ\log \tag{58}$$

generalizes the scalar identities[4] $e^a e^b = e^{a+b}$ and $\log ab = \log a + \log b$. Here we have (for z small enough)

$$\begin{aligned}\mathrm{Tr}\log(I - zT)^{-1} &= \sum_i \sum_{n\geq 1} T_{i,i,n}\frac{z^n}{n} \\ &= \log\det(I - zT)^{-1},\end{aligned}$$

where the first line results from expansion of the logarithm and the second line is an instance of the trace formula. Thus, by differentiation, the sum $\sum_i M_{i,i}(z)$ is seen to be singular at $\rho = 1/\lambda_1$ and we have established that $r = \rho$.

(b) Alternatively, let v_1 be the eigenvector of T corresponding to λ_1. Perron-Frobenius theory also teaches us that, under the irreducibility and aperiodicity conditions, the vector v_1 has all its coordinates that are nonzero. Then the quantity

$$(1 - zT)^{-1}v_1 = \frac{1}{1 - z\lambda_1}v_1$$

is certainly singular at $1/\lambda_1$. But it is also a linear combination of the $F_{i,j}$'s. Thus at least one of the entries of F (hence all of them by the discussion above) must be singular at $\rho = 1/\lambda_1$. Therefore, we have again $r = \rho$.

Finally, under the additional assumption that T is aperiodic, Perron-Frobenius theory grants us that $\rho = 1/\lambda_1$ is well-separated in modulus from all other singularities F. $\square$

[4] The Jacobi trace formula is readily verified when the matrix is diagonalizable, and from there, it can be extended to all matrices by an algebraic "density" argument.

Several of these arguments will be recycled when we discuss the harder problem of analysing coefficients of positive algebraic functions in Chapter V.

We next proceed to show that properties of the Perron-Frobenius type even extend to a large class of linear systems of equations that have nonnegative polynomial coefficients. Such a case is important because of its applicability to transfer matrices; see Section V. 5 below.

Some definitions extending the ones of scalar matrices must first be set. A polynomial

$$p(z) = \sum_j c_j z^{e_j}, \qquad \text{every } c_j \neq 0,$$

is said to be primitive if the quantity $\delta = \gcd(\{e_j\})$ is equal to 1; it is imprimitive otherwise. Equivalently, $p(z)$ is imprimitive iff $p(z) = q(z^\delta)$ for some *bona fide* polynomial q and some $\delta > 1$. Thus, $z, 1+z, z^2+z^3, z+z^4+2z^8$ are primitive while $1, 1+z^2, z^3+z^6, 1+2z^8+5z^{12}$ are not.

DEFINITION V.4. *A linear system with polynomial entries,*

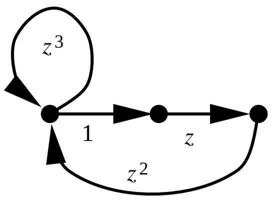
$$f(z) = v(z) + T(z)f(z) \tag{59}$$

where $T \in \mathbb{R}[z]^{r\times r}$, $v \in \mathbb{R}[z]^r$, and $f \in \mathbb{R}[z]^r$ the vector of unknowns is said to be:

- (a) rationally proper (r–proper) *if $T(0)$ is nilpotent, meaning that $T(0)^r$ is the null matrix;*
- (b) rationally nonnegative (r–nonnegative) *if each component $v_j(z)$ and each matrix entry $T_{i,j}(z)$ lies in $\mathbb{R}_{\geq 0}[z]$;*
- (c) rationally irreducible (r–irreducible) *if $(I + T(z))^r$ has all its entries that are nonzero polynomials.*
- (d) rationally aperiodic (r–periodic) *if at least one diagonal entry of some power $T(z)^e$ is a primitive polynomial.*

It is again possible to visualize these properties of matrices by drawing a directed graph whose vertices are labelled $1, 2, \ldots, r$, with the edge connecting i to j that is weighted by the entry $T_{i,j}(z)$ of matrix $T(z)$. Properness means that all sufficiently long paths (and all cycles) must involve some positive power of z— it is a condition satisfied in well-founded combinatorial problems; irreducibility means that the dependency graph is strongly connected by paths whose edges are associated with nonzero polynomials; periodicity means that all closed paths involve weights that are polynomials in some z^e for some $e > 1$.

For instance, if W is a matrix with positive entries, then zW is r–irreducible and r–aperiodic, while z^3W is r–periodic. The matrix $T = \begin{pmatrix} z & z^3 \\ 1 & 0 \end{pmatrix}$ is r–proper, r–irreducible, and r–aperiodic, since $T^2 = \begin{pmatrix} z^2+z^3 & z^4 \\ z & z^3 \end{pmatrix}$. The matrix $T = \begin{pmatrix} z^3 & 1 & 0 \\ 0 & 0 & z \\ z^2 & 0 & 0 \end{pmatrix}$ is r–proper, but it fails to be r–aperiodic since all cycles only involve powers of z^3, as is visible on the associated graph:



By abuse of language, we say that $f(z)$ is a solution of a linear system if it coincides with the first component of a solution vector, $f \equiv f_1$. The following theorem generalizes Corollary V.2.

THEOREM V.5 (Positive rational systems). (i) *Assume that a rational function $f(z)$ is a solution of a system (59) that is r–positive, r–proper, r–irreducible, and r–aperiodic. Then, $f(z)$ has a unique dominant singularity ρ that is positive, and is a simple pole; ρ is the smallest positive solution of*

$$\det(I - T(z)) = 0. \tag{60}$$

(ii) *Assume that $f(z)$ is a solution of a system that is r–positive, r–proper, and r–irreducible (but not necessarily r–aperiodic). Then, the set of dominant singularities of $f(z)$ is of the form $\{\rho_j\}_{j=0}^{d-1}$, where $\rho_0 \in \mathbb{R}_{\geq 0}$, $\rho_j/\rho_0 = \eta$ is a root of unity, and $\rho_j\eta^\ell$ is a dominant singularity for all $\ell = 0, 1, 2, \ldots$. In addition, each ρ_j is a simple pole.*

PROOF. Consider first Case (i). For any fixed $x > 0$, the matrix $T(x)$ satisfies the Perron Frobenius conditions, so that it has a maximal positive eigenvalue $\lambda_1(x)$ that is simple. More information derives from the introduction of matrix norms[5]. Spectral radius and matrix norms are intimately related since

$$\sigma(A) = \lim_{n\to+\infty} (||A^n||)^{1/n}.$$

In particular, this relation entails that the spectral radius is an increasing function of matrix entries: for nonnegative matrices, if $A \leq B$ (in the sense that $A_{i,j} \leq B_{i,j}$ for all i, j), then one has $\sigma(A) \leq \sigma(B)$; if $A < B$ (in the sense that $A_{i,j} < B_{i,j}$ for all i, j), then one has $\sigma(A) < \sigma(B)$. (To see the last inequality, note the existence of $\epsilon > 0$ such that $A \leq (1-\epsilon)B$.)

Returning to the case at hand, equation (56) and the surrounding remarks imply that the spectral radius $\sigma(T(x))$, which also equals $\lambda_1(x)$ for positive x, satisfies

$$\lambda_1(0) = 0, \qquad \lambda_1(x) \text{ strictly increasing}, \qquad \lambda_1(+\infty) = +\infty.$$

(The first condition reflects properness, the second one is a consequence of irreducibility, and the last one derives from simple majorizations.) In particular, the equation $\lambda_1(x) = 1$ admits a unique root ρ on $(0, +\infty)$. (Notice that $\lambda_1(x)$ is a real branch of the algebraic curve $\det(\lambda I - T(x)) = 0$ that dominates all other branches in absolute value for $x > 0$. There results from the general theory of algebraic functions that $\lambda_1(x)$ is analytic at every point $x > 0$.)

[5]A matrix norm $||.||$ satisfies: $||A|| = 0$ implies $A = 0$; $||cA|| = |c| \cdot ||A||$; $||A + B|| \leq ||A| + ||B||$; $||A \times B|| \leq ||A|| \cdot ||B||$.

There remains to prove that: (a) ρ is at most a simple pole of $f(z)$; (b) ρ is actually a pole; (c) there are no other singularities of modulus equal to ρ.

Fact (a) amounts to the property that ρ is a simple root of the equation $\lambda(\rho) = 1$, that is, $\lambda'(\rho) \neq 0$. (To prove $\lambda'(\rho) \neq 0$, we can argue a contrario. First derivatives $\lambda'(\rho), \lambda''(\rho)$, etc, cannot be zero till some *odd* order inclusively since this would contradict the increasing character of $\lambda(x)$ around ρ along the real line. Next, if derivatives till some *even* order ≥ 2 inclusively were zero, then we would have by the local analytic geometry of $\lambda(z)$ near ρ some complex value z_1 satisfying: $|\lambda(z_1)| = 1$ and $|z_1| < \rho$; but for such a value z_1, by irreducibility and aperiodicity, for some exponent e, the entries of $T(z_1)^e$ would be all strictly dominated in absolute value by those of $T(\rho)^e$, hence a contradiction.) Then, $\lambda'(\rho) \neq 0$ holds and by virtue of

$$\det(I - T(z)) = (1 - \lambda_1(z)) \prod_{j\neq 1}(1 - \lambda_j(z)) = (1 - \lambda_1(z)) \frac{\det(I - T(z))}{1 - \lambda_1(z)},$$

the quantity ρ is only a simple root of $\det(I - T(z))$.

Fact (b) means that no "cancellation" may occur at $z = \rho$ between the numerator and the denominator given by Cramer's rule. It derives from an argument similar to the one employed for Corollary V.2. Fact (c) derives from aperiodicity and the Perron-Frobenius properties. □

V. 5. Paths in graphs, automata, and transfer matrices.

A cluster of applications of rational functions is to problems that are naturally described as paths in digraphs, or equivalently as finite automata. In physics, the corresponding treatment is also the basis of what is called the "transfer matrix method". We start our exposition with the enumeration of paths in graphs that constitutes the most direct introduction to the subject.

V. 5.1. Paths in graphs. Let G be a directed graph with vertex set $\{1, \ldots, m\}$, where self-loops are allowed and label each edge (a, b) by the *formal* variable $g_{a,b}$. We introduce the matrix $\mathbf{G}$ such that

$$\mathbf{G}_{a,b} = g_{a,b} \text{ if the edge } (a,b) \in G, \qquad \mathbf{G}_{a,b} = 0 \text{ otherwise,} \tag{61}$$

which is called the *formal adjacency matrix* of G. Then, from the standard definition of matrix products, the powers $\mathbf{G}^r$ have elements that are path polynomials. More precisely, one has the simple but essential relation,

$$(\mathbf{G})^r_{i,j} = \sum_{w \in \mathcal{P}(i,j;r)} w, \tag{62}$$

where $\mathcal{P}(i, j, r)$ is the set of paths in G that connect i to j and have length r, and a path w is assimilated to the monomial in indeterminates $\{g_{i,j}\}$ that represents multiplicatively the succession of its edges; for instance:

$$(\mathbf{G})^3_{i,j} = \sum_{m_1=i, m_2, m_3, m_4=j} g_{m_1,m_2} g_{m_2,m_3} g_{m_3,m_4},$$

In other words, powers of the matrix associated to a graph "generate" all paths in a graph. One may then treat simultaneously all lengths of paths (and all powers of matrices) by introducing the variable z to record length.

PROPOSITION V.7. (i) *Let G be a digraph and let $\mathbf{G}$ be the matrix associated to G by rules* (61) *The OGF $F^{\langle i,j\rangle}(z)$ of the set of all paths from i to j in a digraph G with z marking length and $g_{a,b}$ marking the occurrence of edge (a,b) is the entry i,j of the matrix $(I-z\mathbf{G})^{-1}$, namely*

$$F^{\langle i,j\rangle}(z) = (I - z\mathbf{G})^{-1}\Big|_{i,j} = \frac{\Delta^{\langle i,j\rangle}(z)}{\Delta(z)},$$

where $\Delta(z) = \det(I - z\mathbf{G})$ and $\Delta^{\langle i,j\rangle}(z)$ is the determinant of the minor of index i,j of $I - z\mathbf{G}$.

(ii) *The generating function of nonempty closed paths is given by*

$$\sum_i (F^{\langle i,i\rangle}(z) - 1) = -z\frac{\Delta'(z)}{\Delta(z)}.$$

The quantity $\det(I - z\mathbf{G})$ is obviously the reciprocal polynomial of the characteristic polynomial of $\mathbf{G}$.

PROOF. Part (i) results from the discussion above which implies

$$F^{\langle i,j\rangle}(z) = \sum_{n=0}^{\infty} z^n \left(\mathbf{G}^n\right)_{i,j} = \Big((I - z\mathbf{G})^{-1}\Big)_{i,j},$$

and from the cofactor formula of matrix inversion. Part (ii) results from Jacobi's trace formula (58). Introduce the quantity known as the *zeta function*,

$$\begin{array}{rcl} \zeta(z) & := & \exp\left(\sum_i \sum_{n=1}^{\infty} F_n^{\langle i,i\rangle}\frac{z^n}{n}\right) \quad = \quad \exp\left(\sum_{n=1}^{\infty} \frac{z^n}{n}\mathrm{Tr}\,\mathbf{G}^n\right) \\ & = & \exp\left(\mathrm{Tr}\,\log(I - z\mathbf{G})^{-1}\right) \quad = \quad \det(I - z\mathbf{G})^{-1}, \end{array}$$

where the last line results from the Jacobi trace formula. Thus, $\zeta(z) = \Delta(z)^{-1}$. On the other hand, differentiation combined with the definition of $\zeta(z)$ yields

$$\begin{array}{rcl} z\dfrac{\zeta'(z)}{\zeta(z)} & = & -z\dfrac{\Delta'(z)}{\Delta(z)} \\ & = & \displaystyle\sum_i \sum_{n=1}^{\infty} F_n^{\langle i,i\rangle} z^n, \end{array}$$

and Part (ii) follows. □

The numeric substitution $\sigma: \ g_{a,b} \mapsto 1$ transforms the matrix $\mathbf{G}$ into the usual *adjacency matrix*. In particular, the number of paths of length n is obtained, under this substitution, as $[z^n](1 - zG])^{-1}$. In a similar vein, it is possible to consider weighted graphs, where the $g_{a,b}$ are assigned *real-valued weights*; with the weight of a path being defined by the product of its edges weights, one finds that $[z^n](I - z\mathbf{G})^{-1}$ equals the total weight of all paths of length n. If furthermore the assignment is made in such a way that $\sum_b g_{a,b} = 1$, then the matrix $\mathbf{G}$, which is called a stochastic matrix, can be interpreted as the transition matrix of a Markov chain.

▷ **12.** *Fast computation of the characteristic polynomial.* Observe that

$$z\frac{\zeta'(z)}{\zeta(z)} = \sum_{n\geq 1} z^n \mathrm{Tr}\,\mathbf{G}^n = \sum_{\lambda} \frac{\lambda z}{1 - \lambda z},$$

(the sum is over eigenvalues). From this, one deduces an algorithm that determines the characteristic polynomial of a matrix of dimension m in $\mathcal{O}(m^4)$ arithmetic operations. [Hint: computing the quantities $\mathrm{Tr}\,G^j$ for $j = 1, \ldots, m$ requires precisely m matrix multiplications.] ◁

$\triangleright$ **13.** *The matrix tree theorem.* Let G be a directed graph without loops and associated matrix $\mathbf{G}$, with $g_{a,b}$ marking edge (a,b). The Laplacian matrix $\mathbf{L}[G]$ is defined by

$$\mathbf{L}[G]_{i,j} = -g_{i,j} + [\![i=j]\!] \sum_k g_{i,k}.$$

Let $\mathbf{L}_1[G]$ be the matrix obtained by deleting the first row and first column of $\mathbf{L}[G]$. Then, the "tree polynomial"

$$T_1[G] := \det \mathbf{L}_1[G]$$

enumerates all (oriented) spanning trees of G rooted at node 1. [This classic result belongs to a circle of ideas initiated by Kirchhoff, Sylvester, Borchardt and others in the 19th century. See, e.g., the discussions by Knuth [**74**, p. 582–583] and Moon [**86**].] $\triangleleft$

Let us now assume that positive weights are assigned to the edges of G. In other words, the quantities $g_{a,b}$ in (61) have positive values. If the resulting matrix is irreducible and aperiodic, then Perron-Frobenius theory applies. There exists $\rho = 1/\lambda_1$, with $\lambda_1 > 0$ the dominant eigenvalue of $\mathbf{G}$, and the OGF of weighted paths from i to j has a simple pole at ρ. A host of probabilistic properties of paths result from there, after a certain "residue matrix" has been calculated:

LEMMA V.1 (Iteration of Perron-Frobenius matrices). *Set $M(z) = (I - z\mathbf{G})^{-1}$ where $\mathbf{G}$ has nonnegative entries, is irreducible, and is aperiodic. Let λ_1 be the dominant eigenvalue of $\mathbf{G}$. Then the "residue" matrix R such that*

$$(I - z\mathbf{G})^{-1} = \frac{R}{1 - z\lambda_1} + O(1) \qquad (z \to \lambda_1^{-1}) \tag{63}$$

has entries given by ($\langle x, y\rangle$ represents a scalar product)

$$R_{ij} = \frac{r_i \ell_j}{\langle r, \ell\rangle},$$

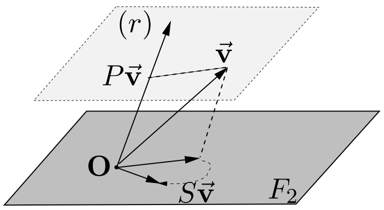
where r and ℓ are right and left eigenvectors of $\mathbf{G}$ corresponding to the eigenvalue λ_1.

PROOF. Let $\mathcal{E}$ be the ambient space. There exists a direct sum decomposition $\mathcal{E} = \mathcal{F}_1 + \mathcal{F}_2$ where $\mathcal{F}_1$ is the 1-dimensional eigenspace generated by the eigenvector (r) corresponding to eigenvalue λ_1 and $\mathcal{F}_2$ is the supplementary space which is the direct sum of characteristic spaces corresponding to the other eigenvalues $\lambda_2, \ldots$. (For the purposes of the present discussion, one may freely think of the matrix as diagonalizable, with $\mathcal{F}_2$ the union of eigenspaces associated to $\lambda_2, \ldots$.) Then $\mathbf{G}$ as a linear operator acting on $\mathcal{F}$ admits the decomposition

$$\mathbf{G} = \lambda_1 P + S,$$

where P is the projector on $\mathcal{F}_1$ and S acts on $\mathcal{F}_2$ with spectral radius $|\lambda_2|$, as illustrated by the diagram:

(64)



By standard properties of projections, $P^2 = P$ and $PS = SP = 0$ so that

$$\mathbf{G}^n = \lambda_1^n P + S.$$

Consequently, there holds,

$$(65)\qquad \begin{aligned}(I - z\mathbf{G})^{-1} &= \sum_{n\geq 0} z^n \lambda_1^n P + z^n S \\ &= \frac{P}{1-\lambda_1 z} + (I - zS)^{-1}.\end{aligned}$$

Thus, the residue matrix R coincides with the projector P.

Now, for any vector w, by general properties of projections, one has ($R \equiv P$):

$$Rw = c(w)r,$$

for some coefficient $c(w)$. Application of this to each of the base vectors e_j (i.e., $e_j = (\delta_{j1}, \ldots, \delta_{jd})^t$) shows that the matrix R has each of its *columns* proportional to the eigenvector r. A similar reasoning with the transpose $\mathbf{G}^t$ of $\mathbf{G}$ and the associated residue matrix R^t shows that the matrix R has each of its *rows* proportional to the eigenvector ℓ. In other words, for some constant γ, one must have

$$R_{i,j} = \gamma \ell_j r_i.$$

The normalization constant γ is itself finally determined by $\ell R r = \langle \ell, r\rangle$.

We finally observe that a full expansion can be obtained:

$$(66)\qquad (I - z\mathbf{G})^{-1}\frac{P}{1-\lambda_1 z} + \sum_{k\geq 0} R_k \left(z - \lambda_1^{-1}\right)^k, \qquad R_k := S^k (I - \lambda_1^{-1} S)^{-k-1}.$$

The proof also reveals that one needs to solve one polynomial equation for determining λ_1, and then the other quantities in (66) are all obtained by inverting matrices in the field of constants extended by the algebraic quantity λ_1. (Numerical procedures are likely to be used instead for large matrices.) □

Equipped with this lemma, we can now state:

THEOREM V.6 (Random paths in digraphs). *Let* $\mathbf{G}$ *be a nonnegative matrix associated to a weighted digraph* G*, assumed to be irreducible and aperiodic. Consider the collection* $\mathcal{P}_{a,b}$ *of (weighted) paths with fixed origin* a *and final destination* b*. Then, the number of traversals of edge* (s,t) *in a random element of* $\mathcal{P}_{a,b}$ *has mean*

$$(67)\qquad \tau_{s,t} n + O(1) \qquad \textit{where} \quad \tau_{s,t} := \frac{\ell_s g_{s,t} r_t}{\lambda_1 \langle \ell, r\rangle}.$$

In other words, a long random path tends to spend asymptotically a fixed (nonzero) fraction of its time traversing any given edge. Accordingly, the number of visits of vertex s is also proportional to n and obtained by summing the expression of (67) according to all the possible values of t.

PROOF. First, the total weight ("number") of paths in $\mathcal{P}_{a,b}$ satisfies

$$(68)\qquad [z^n]\left[(I - z\mathbf{G})^{-1}\right]_{a,b} \sim \lambda_1 \frac{r_a \ell_b}{\langle \ell, r\rangle},$$

as follows from Lemma V.1. Next, introduce the modified matrix $\mathbf{H} = (h_{i,j})$ defined by

$$h_{i,j} = g_{i,j}\, u^{[\![i=s \wedge j=t]\!]}.$$

In other words, we mark each traversal of edge i,j by the variable u. Then, the quantity

$$(69)\qquad [z^n]\left[\frac{\partial}{\partial u}(I - z\mathbf{H})^{-1}\Big|_{u=1}\right]_{a,b}$$

represents the total number of traversals of edge (s,t), with weights taken into account. Simple algebra[6] shows that

$$\frac{\partial}{\partial u}(I - z\mathbf{H})^{-1}\bigg|_{u=1} = (I - z\mathbf{G})^{-1}\,(zH')\,(I - z\mathbf{G}), \tag{70}$$

where $H' := (\partial_u H)_{u=1}$ has all its entries equal to 0, except for the s,t entry whose value is $g_{s,t}$. By the calculation of the residue matrix in Lemma V.1, the coefficient of (69) is then asymptotic to

$$[z^n]\frac{R_{a,s}}{1-\lambda_1 z} g_{s,t} z \frac{R_{t,b}}{1-\lambda_1 z} \sim \upsilon n \lambda_1^{n-1}, \qquad \upsilon := \frac{r_a \ell_s g_{s,t} r_t \ell_b}{\langle \ell, r\rangle^2}. \tag{71}$$

Comparison of (71) and (68) finally yields the result since the relative error terms are $O(n^{-1})$ in each case. □

Another consequence of this last proof and Equation (68) is that the numbers of paths starting at a and ending at either b or c satisfy

$$\lim_{n\to\infty} \frac{P_{a,b,n}}{P_{a,c,n}} = \frac{\ell_b}{\ell_c}. \tag{72}$$

In other words, the quantity

$$\frac{\ell_b}{\sum_j \ell_j}$$

is the probability that a random path with origin fixed at some point a but otherwise unconstrained will end up at point b after n steps. Such properties are strongly evocative of Markov chain theory discussed below in Example 12.

▷ **14.** *Concentration of distribution for the number of passages.* Under the conditions of the theorem, the standard deviation of the number of traversals of a designated node or edge is $O(\sqrt{n})$. Thus in a random long path, the distribution of the number of such traversals is concentrated. [Compared to (70), the calculation of the second moment requires taking a further derivative, which leads to a triple pole. The second moment and the square of the mean, which are each $O(n^2)$, are then found to cancel to main asymptotic order.] ◁

EXAMPLE 11. *Walks on the interval revisited.* As a direct illustration, consider the walks associated to the graph $G(5)$ with vertex set $1,\ldots,5$ and edges being formed of all pairs (i,j) such that $|i-j| \le 1$. The matrix is

$$\mathbf{G}(5) = \begin{pmatrix} 1&1&0&0&0\\ 1&1&1&0&0\\ 0&1&1&1&0\\ 0&0&1&1&1\\ 0&0&0&1&1 \end{pmatrix}.$$

The characteristic polynomial factorizes as

$$\chi_{\mathbf{G}(5)}(z) = z(z-1)(z-2)(z^2-2z-2),$$

and its dominant root is $\lambda_1 = 1+\sqrt{3}$. From there, one finds a left eigenvector (which is also a right eigenvector since the matrix is symmetric):

$$r = \ell^t = (1, \sqrt{3}, 2, \sqrt{3}, 1).$$

[6] If A depends on u, one has $\partial_u A^{-1} = A^{-1}(\partial_u A)A^{-1}$, which is a noncommutative generalization of the usual differentiation rule for inverses.

Thus a random path (with the uniform distribution over all paths corresponding to the weights being equal to 1) visits nodes $1, \dots, 5$ with frequencies proportional to

$$1, \quad 1.732, \quad 2, \quad 1.732, \quad 1,$$

implying that the central nodes are visited more often—such nodes have higher degrees of freedom, hence there tends to be more paths that traverse them.

In fact, this example has structure. For instance, the corresponding problem on an interval of length 11, leads to a matrix with a highly factorable characteristic polynomial

$$\chi_{\mathbf{G}(11)} = z\,(z-1)\,(z-2)\left(z^2-2\,z-2\right)\left(z^2-2\,z-1\right)\left(z^4-4\,z^3+2\,z^2+4\,z-2\right).$$

The reader may have recognized a particular case of lattice paths which resort to to the theory exposed in Section V. 2. For instance, according to Theorem V.2, the OGF of paths from vertex 1 to vertex 1 in the graph with k vertices is given by the continued fraction

$$\cfrac{1}{1-z-\cfrac{z^2}{1-\cfrac{z^2}{\ddots\,\cfrac{}{1-\cfrac{z^2}{1-z}}}}}.$$

(The number of fraction bars is k; the first and last quotients are $1-z$, the others being equal to 1.) From this it can be shown that the characteristic polynomial of G is an elementary variant of the Fibonacci–Chebyshev polynomial of Example 6. The analysis based on Theorem V.6 is simpler, albeit more rudimentary, as it only provides a first-order asymptotic solution to the problem. $\square$

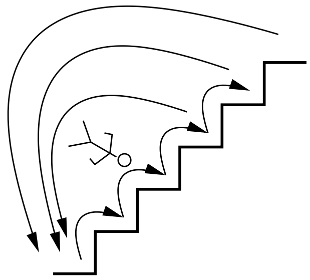
EXAMPLE 12. *Elementary theory of Markov chains.* Consider the case where the row sums of matrix $\mathbf{G}$ are all equal to 1, that is, $\sum_j g_{i,j} = 1$. Such a matrix is called a *stochastic matrix*. The quantity $g_{i,j}$ can then be interpreted as the probability of leaving state i for state j, assuming one is in state i. Assume that the matrix $\mathbf{G}$ is irreducible and aperiodic. Clearly, the matrix $\mathbf{G}$ admits the column vector $r = (1, 1, \dots, 1)^t$ as an eigenvector corresponding to the dominant eigenvalue $\lambda_1 = 1$. The left eigenvector ℓ normalized so that its elements sum to 1 is called the (row) vector of stationary probabilities. It must be determined by linear algebra and it involves finding an element of the kernel of matrix $I - \mathbf{G}$, which can be done in a standard way.

Application of Theorem V.6 and Equation (68) shows immediately the following:

PROPOSITION V.8 (Stationary probabilities of Markov chains). *Consider a weighted graph corresponding to a stochastic matrix $\mathbf{G}$ which is irreducible and aperiodic. Let ℓ be the normalized left eigenvector corresponding to the eigenvalue 1. A random (weighted) path of length n with fixed origin and destination visits node s a mean number of times asymptotic to $\ell_s n$ and traverses edge (s,t) a mean number of times asymptotic to $\ell_s g_{s,t} n$. A random path of length n with fixed origin ends at vertex s with probability asymptotic to ℓ_s.*

This first-order asymptotic property certainly constitutes the most fundamental result in the theory of finite Markov chains. $\square$

The next example illustrates an elementary technique often employed in calculations of eigenvalues and eigenvectors. It presupposes that the matrix to be analysed can be reduced to a sparse form and has a regular enough structure.



$$\mathbf{G} = \begin{pmatrix} 1&1&0&0&0&0\\ 1&0&1&0&0&0\\ 1&0&0&1&0&0\\ 1&0&0&0&1&0\\ 1&0&0&0&0&1\\ 1&0&0&0&0&0 \end{pmatrix} \qquad \widetilde{\mathbf{G}} = \begin{pmatrix} \frac12&\frac12&0&0&0&0\\ \frac12&0&\frac12&0&0&0\\ \frac12&0&0&\frac12&0&0\\ \frac12&0&0&0&\frac12&0\\ \frac12&0&0&0&0&\frac12\\ 1&0&0&0&0&0 \end{pmatrix}$$

FIGURE 14. The devil's staircase ($m = 6$) and the two matrices that can model it.

EXAMPLE 13. *The devil's staircase.* You live in a house that has a staircase with m steps. You come back home a bit loaded and at each second, you can either succeed in climbing a step or fall back all the way down. On the last step, you always stumble and fall back down (Figure 14). Where are you likely to be found at time n?

Precisely, two slightly different models correspond to this informally stated problem. The probabilistic model views it as a Markov chain with equally likely possibilities at each step and is reflected my matrix $\widetilde{\mathbf{G}}$ in Figure 14. The combinatorial model just assumes all possible evolutions ("histories") of the system as equally likely and it corresponds to matrix $\mathbf{G}$. We opt here for the latter, keeping in mind that the same method basically applies to both cases.

We first write down the constraints expressing the joint properties of an eigenvalue λ and its right eigenvector $x = (x_1, \ldots, x_m)^t$. The equations corresponding to $(\lambda I - \mathbf{G})x = 0$ are formed of a first batch of $m - 1$ relations,

$$(\lambda - 1)x_1 - x_2 = 0, \quad -x_1 + \lambda x_2 - x_3 = 0, \quad \cdots, -x_1 + \lambda x_{m_1} - x_m = 0, \tag{73}$$

together with the additional relation (one cannot go higher than the last step):

$$-x_1 + \lambda x_m = 0. \tag{74}$$

The solution to (73) is readily found by pulling out successively $x_2, \ldots, x_m$ as functions of x_1:

$$x_2 = (\lambda - 1)x_1, \quad x_3 = (\lambda^2 - \lambda - 1)x_1, \quad \cdots, \quad x_m = (\lambda^m - \lambda^{m-1} - \cdots - 1)x_1. \tag{75}$$

Combined with the special relation (74), this last relation shows that λ must satisfy the equation

$$1 - 2\lambda^m + \lambda^{m+1}. \tag{76}$$

Let λ_1 be the largest positive root of this equation, existence and dominance beeing guaranteed by Perron-Frobenius properties. Note that the quantity $\rho := 1/\lambda_1$ satisfies the characteristic equation

$$1 - 2\rho + \rho^{m+1} = 0,$$

already encountered when discussing longest runs in words; the discussion of Example 2 then grants us the existence of an isolated ρ near $\frac{1}{2}$, hence the fact that λ_1 is slightly less than 2.

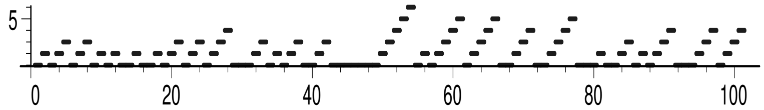
Similar devices yield the left eigenvector $y = (y_1, \ldots, y_m)$. It is found easily that y_j must be proportional to λ_1^{-j}. We thus obtain from Theorem V.6 and Equation (72): *The* *probability of being in state j (i.e., being on step j of the stair) at time n tends to the limit*

$$\varpi_j = \gamma \lambda_1^{-j}$$

where λ_1 is the root near 2 of (76) *and the normalization constant γ is determined by* $\sum_j \varpi_j = 1$. In other words, the distribution of the altitude at time n is a truncated geometric distribution with parameter $1/\lambda_1$. For instance, $m = 6$ leads to $\lambda_1 = 1.98358$, and the asymptotic probabilities of being in states $1, \ldots, 6$ are

$$(77) \qquad 0.50413, \quad 0.25415, \quad 0.12812, \quad 0.06459, \quad 0.03256, \quad 0.01641,$$

exhibiting clearly a geometric decay. Here is the simulation of a random history for $n = 100$:



In this case, the frequencies observed are $0.44,\ 0.26,\ 0.17,\ 0.08,\ 0.04,\ 0.01$, pretty much in agreement with what is expected.

Finally, the similarity with the longest run problem is easily explained. Let u and d be letters representing steps upwards and downwards respectively. The set of paths from state 1 to state 1 is described by the regular expression

$$\mathcal{P}_{1,1} = \left(d + ud + \cdots + u^{m-1}d\right)^{\star},$$

corresponding to the generating function

$$P_{1,1}(z) = \frac{1}{1 - z - z^2 - \cdots - z^m},$$

of variant of the OGF of words without m-runs of the letter u, which also corresponds to the enumeration of compositions with summands $\leq m$. The case of the probabilistic transition matrix $\widetilde{\mathbf{G}}$ is left as an exercise to the reader. □

This last example is typical: in many cases combinatorial problems have some amount of regularity. In such situations, all the resources of linear algebra are available, including the vast body of knowledge gathered over years on calculations of structured determinants; see for instance Krattenthaler's survey [**78**] and the book [**111**].

V. 5.2. Finite automata. Word problems corresponding to regular languages can be treated by the theory of regular specifications whenever they have enough structure and an unambiguous regular expression description is of tractable form. This was the main theme of Sections V. 1 and V. 2. The dual point of view of automata theory proves useful whenever no such direct description is in sight. Finite automata resorting essentially to the theory of paths in graphs, the results from the previous sections apply with only minor adaptation. For convenience, we start by recalling definitions already given in Chapter I.

DEFINITION V.5. *A* finite automaton *A over a finite alphabet $\mathcal{A}$ is a directed multigraph whose vertex set Q is called the set of states and whose edges are labelled by letters of the alphabet. This graph is equipped with a designated initial state $q_0 \in Q$ and a designated set of final states $Q_f \subseteq Q$.*

A word w is said to be accepted by the automaton if there exists a path π in the graph connecting the initial state q_0 to one of the final states $q \in Q_f$, so that the succession of labels of the path π corresponds to the sequence of letters composing w. The path π is then called an accepting path for w. (We can regard the finitely many states as keeping a patial memory of the past, an interpretation that proves useful in design issues.) The set of accepted words is denoted by $\mathcal{L}(A)$.

In all generality, a finite automaton may be a *nondeterministic* device: given a word w, one might not "know" *a priori* which choices to effect at vertices in order to accept it. A finite automaton is said to be *deterministic* if given any state $q \in Q$ and any letter $x \in \mathcal{A}$, there is at most one edge from vertex q that bears label x. In that case, one decides easily (in linear time) whether a word is accepted by just following edges dictated by the sequence of letters in w. All automata to be used in the examples below are deterministic.

PROPOSITION V.9 (Finite state automata counting). *Any language accepted by a deterministic finite automaton has a rational generating function obtained as follows. If the language is specified by the deterministic automaton $A = \langle Q, Q_f, q_0 \rangle$, then the corresponding ordinary generating function $L_0(z)$ is the component $L_0(z)$ of the linear system of equations*

$$\left\{ L_j(z) = \phi_j + z \sum_{a \in \mathcal{A}} L_{\tau(q_j, a)}(z) \right\},$$

where ϕ_j equals 1 if $q_j \in Q_f$ and 0 otherwise, and where $\tau(q_j, a)$ is the index of the state reachable from state q_j when the letter a is read.

As a consequence, the number of words in the language accepted by a finite-state automaton always admits of an exponential-polynomial form.

Note. The most fundamental result of the theory of regular languages is that there is complete equivalence between three descriptive models: regular expressions, deterministic finite automata, and nondeterministic finite automata. The corresponding theorems are due to Kleene (the equivalence between regular expression and nondeterministic finite automata) and to Rabin and Scott (the equivalence between nondeterministic and deterministic automata). Thus, finite automata whether deterministic or not accept ("recognize") the family of all regular languages.

PROOF. By the fundamental equivalence of models, one may freely assume the automaton to be deterministic. The quantity L_j is nothing but the OGF of the language obtained by changing the initial state of the automaton to q_j. Each equation expresses the fact that a word accepted starting from q_j may be the empty word (if q_j is final) or, else, it must consist of a letter a followed by a continuation that is itself accepted when the automaton is started from the "next" state, that is, the state of index $\tau(q_j, a)$. (Equivalently, one may reduce the proof to the enumeration of paths in graphs as detailed above.)

Existence of the exponential-polynomial form immediately results from rationality of the OGF. □

As implied by the statement of the proposition, the OGF of the language defined by a deterministic finite automaton involves a quasi-inverse $(1 - zT)^{-1}$ where the *transition matrix* T is a direct encoding of the automaton's transitions. Corollary V.2 and Lemma V.1 have been precisely custom-tailored for this situation. As is by now usual, we shall allow weights on letters of the alphabet, corresponding to a Bernoulli model on words. We say that an automaton is irreducible (resp. aperiodic) if the underlying graph and the associated matrix are irreducible (resp. aperiodic).

PROPOSITION V.10 (Random words and automata). *Let $\mathcal{L}$ be a language recognized by a deterministic finite automaton A that is irreducible and aperiodic. The number of words of $\mathcal{L}$ satisfies*

$$L_n \sim c\lambda_1^n \left(1 + O(d^{-n})\right),$$

where λ_1 is the dominant (Perron-Frobenius) eigenvalue of the transition matrix of A and c, d are positive constants with $d > 1$.

$$\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

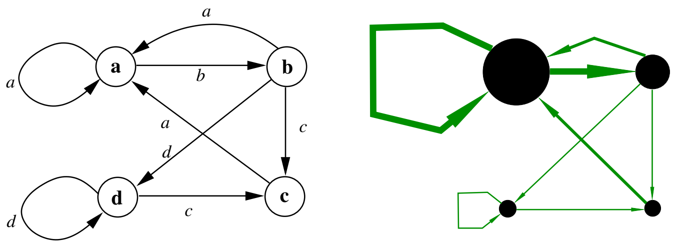


FIGURE 15. Locally constrained words: The transition matrix (T) associated to the forbidden pairs $F = \{ac, ad, bb, cb, cc, cd, da, db\}$, the corresponding automaton, and the graph with widths of vertices and edges drawn in proportion to their asymptotic frequencies.

In a random word of $\mathcal{L}_n$, the number of traversals of a designated vertex or edge has a mean that is asymptotically linear in n and is given by Theorem V.6.

▷ **15.** *Unambiguous automata.* A nondeterministic finite state automaton is said to be unambiguous if the set of accepting paths for any given words comprises at most one element. The translation into generating function as described above also applies to such automata, even though they are nondeterministic. ◁

EXAMPLE 14. *Locally constrained words.* Consider a fixed alphabet $\mathcal{A} = \{a_1, \ldots, a_m\}$ and a set $\mathcal{F} \subseteq \mathcal{A}^2$ of forbidden transitions between consecutive letters. The set of words over $\mathcal{A}$ with no such forbidden transition is denoted by $\mathcal{L}$ and is called a locally constrained language. (The particular case where exactly all pairs of equal letters are forbidden corresponds to Smirnov words and has been discussed on p. 32.)

Clearly, the words of $\mathcal{L}$ are recognized by an automaton whose state space is isomorphic to $\mathcal{A}$: state q simply memorizes the fact that the last letter read was a q. The graph of the automaton is then obtained by the collection of allowed transitions $(q, r) \mapsto a$, with $(q, r) \notin \mathcal{F}$. (In other word, the graph of the automaton is the complete graph in which all edges that correspond to forbidden transitions are deleted.) Consequently, the OGF of any locally constrained language is a rational function. Its OGF is given by

$$(1, 1, \ldots, 1)(I - zT)^{-1}(1, 1, \ldots, 1)^t,$$

where T_{ij} is 0 if $(a_i, a_j) \in \mathcal{F}$ and 1 otherwise. If each letter can follow any other letter in an accepted word, the automaton is irreducible. The graph is aperiodic except in a few degenerate cases (e.g., in the case where the allowed transitions would be $a \to b, c$, $b \to d$, $c \to d$, $d \to a$). Under irreducibility and aperiodicity, the number of words will be $\sim c\lambda_1^{-n}$ and each letter will have on average an asymptotic constant frequency. (See (27) and (28) of Chapter IV for the case of Smirnov words.)

For the example of Figure 15, the alphabet is $\mathcal{A} = \{a, b, c, d\}$. There are eight forbidden transitions and the characteristic polynomial is found to be $\lambda^3(\lambda - 2)$. Thus, one has $\lambda_1 = 2$. The right and left eigenvectors are found to be

$$r = (2, 2, 1, 1)^t, \qquad \ell = (2, 1, 1, 1).$$

Then, the matrix $(\tau_{s,t})$, where $\tau_{s,t}$ represents the asymptotic frequency of transitions from letter s to letter t is found in accordance with Theorem V.6:

$$\Gamma = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{1}{16} & \frac{1}{16} \\ \frac{1}{8} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{16} & \frac{1}{16} \end{pmatrix}.$$

This means that a random path spends a proportion equal to $\frac{1}{4}$ of its time on a transition between an a and a b, but much less ($\frac{1}{16}$) on transitions between pairs of letters bc, bd, cc, ca. The letter frequencies in a random word of $\mathcal{L}$ are $(\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8})$, so that an a is four times more frequent than a c or a d, and so on. See Figure 15 (right) for a rendering.

Various specializations, including multivariate GF's and nonuniform letter models are readily treated by this method. Bertoni *et al.* develop in [**13**] related variance and distribution calculations in the case of the number of occurrences of a symbol in an arbitrary regular language. □

EXAMPLE 15. *De Bruijn graphs.* Two thieves want to break into a house whose entrance is protected by digital lock with an unknown four-digit code. As soon as the four digits of the code are typed consecutively, the gate opens. The first thief proposes to try in order all the four-digit sequences , resulting in as much as 40,000 key strokes in the worst-case. The second thief, who is a mathematician, says he can try *all* four-digit combinations with only 10,003 key strokes. What is the mathematician's trade secret?

Clearly certain optimizations are possible: for instance, for an alphabet of cardinality 2 and codes of 2 letters, the sequence 00110 is better than the naïve one, 00 01 10 11, which is redundant; a few more attempts will lead to an optimal solution for 3–digit codes that has length 11 (rather than 24), for instance,

$$\mathsf{0001110100}.$$

The general question is then: How far can one go and how to construct such sequences?

Fix an alphabet of cardinality m. A sequence that contains as factors (contiguous blocks) *all* the k letter words is called a *de Bruijn sequence*. Clearly, its length must be at least $\delta(m,k) = m^k + k - 1$, as it must have at least m^k positions at distance at least k from the end. Such a sequence of smallest possible length $\delta(m,k)$ is called a *minimal* de Bruijn sequence. Such sequences were discovered by N. G. de Bruijn [**24**] in 1946, in response to a question coming from electrical engineering, where all possible reactions of a device presented as a black box must be tested at minimal cost. We shall expose here the case of a binary alphabet, $m = 2$, the generalization to $m > 2$ being obvious.

Let $\ell = k - 1$ and consider the automaton $\mathcal{B}_\ell$ that memorizes the last block of length ℓ read when scanning the input text from left to right. A state is thus assimilated to a string of length ℓ and the total number of states is 2^ℓ. The transitions are easily calculated: let $q \in \{0,1\}^\ell$ be a state and let $\sigma(w)$ be the function that shifts all letters of a word w one position to the left, dropping the first letter of w in the process (thus σ maps $\{0,1\}^\ell$ to $\{0,1\}^{\ell-1}$); the transitions are

$$q \overset{\mathsf{0}}{\mapsto} \sigma(q)\mathsf{0}, \qquad q \overset{\mathsf{1}}{\mapsto} \sigma(q)\mathsf{1}.$$

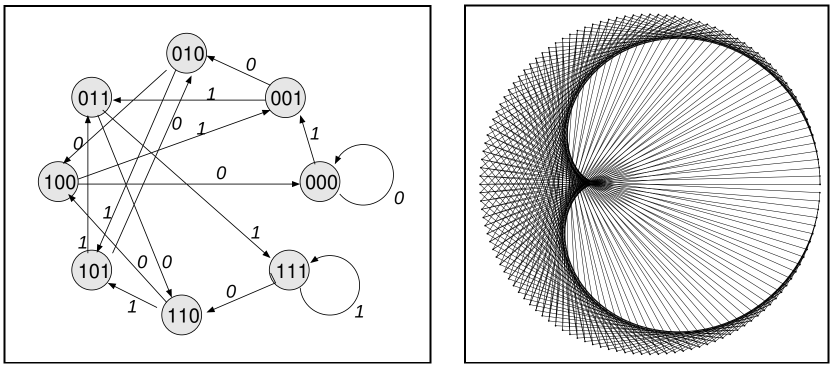


FIGURE 16. The de Bruijn graph: (left) $\ell = 3$; (right) $\ell = 7$.

If one further interprets a state q as the integer in the interval $[0\,..\,2^\ell - 1$ that it represents, then the transition matrix assumes a remarkably simple form:

$$T_{i,j} = \frac{1}{2}[\![(j \equiv 2i \bmod 2^\ell) \text{ or } (j \equiv 2i + 1 \bmod 2^\ell)]\!].$$

See Figure 16 for a rendering borrowed from [**53**].

Combinatorially, the de Bruijn graph is such that each node has indegree 2 and outdegree 2. By a well known theorem going back to Euler: *A necessary and sufficient condition for an undirected connected graph to have an Eulerian circuit (that is, a closed path that traverses each vertex exactly once) is that every node has even degree. For strongly connected digraphs, the condition is that each node should have an outdegree equal to its indegree.* This last condition is obviously satisfied here. Take an Eulerian circuit starting and ending at node 0^ℓ; its length is $2^{\ell+1} = 2^k$. Then, clearly, the sequence of edge labels encountered when prefixed with the word $0^{k-1} = 0^\ell$ constitutes a minimal de Bruijn sequence. In general, the argument gives a de Brujin sequence with minimal length $m^k + k - 1$. Et voilà! The trade secret of the thief-mathematician is exposed.

Back now to enumeration. The de Bruijn matrix is irreducible since a path labelled by sufficiently many zeros always leads any state to the state 0^ℓ, while a path ending with the letters of $w \in \{0,1\}^\ell$ leads to state w. The matrix is aperiodic since it has a loop on states 0^ℓ and 1^ℓ. Thus, by Perron Frobenius properties, it has a unique dominant eigenvalue, and it is not hard to check that its value is $\lambda_1 = 2$, corresponding to the right eigenvector $(1, 1, \ldots, 1)^t$. If one fixes a *pattern* $w \in \{0,1\}^\ell$, Theorem V.6 yields the fact that a random word contains on average $\sim \frac{n}{2^\ell}$ occurrences of pattern w. Note 14 also implies that the distribution of the number of occurrences is concentrated around the mean as the variance is $O(n)$. This gives us in a simple manner a version of what was nicknamed "Borges's Theorem" in Chapter I: *Almost every sufficiently long text contains* all *patterns of some predetermined length* ℓ. As a matter of fact, the de Bruijn graph may be used to quantify many properties of occurrences of patterns in random words, and it has been used for this purpose in several works including [**10, 46, 53**]. □

EXAMPLE 16. *Words with excluded patterns.* Fix a finite set of patterns $\Omega = \{w_1, \ldots, w_r\}$, where each w_j is a word of $\mathcal{A}^\star$. The language $\mathcal{E} \equiv \mathcal{E}^\Omega$ of words that contain no factor in Ω

is described by the extended regular expression

$$\mathcal{E} = \mathcal{A}^\star \setminus \bigcup_{j=1}^{r} (A^\star w_j A^\star),$$

which constitutes a concise but highly ambiguous description. By closure properties of regular languages, $\mathcal{E}$ is itself regular and there must exist a deterministic automaton that recognizes it.

An automaton recognizing $\mathcal{E}$ can be constructed starting from the de Bruijn automaton of index $k = -1 + \max |w_j|$ and deleting all the vertices and edges that correspond to a word of Ω. Precisely, vertex q is deleted whenever q contains a factor in Ω; the transition (edge) from q associated with letter α gets deleted whenever the word $q\alpha$ contains a factor in Ω. The pruned de Bruijn automaton, call it $\mathcal{B}_k^\circ$, accepts all words of $0^k\mathcal{E}$, when it is equiped with the initial state 0^k and all states are final. Thus, the OGF $E(z)$ is in all cases a rational function.

The matrix of $\mathcal{B}_k^\circ$ is the matrix of $\mathcal{B}_k$ with some nonzero entries replaced by 0. Assume that $\mathcal{B}_k^\circ$ is irreducible. This assumption only eliminates a few pathological cases (e.g., $\Omega = \{01\}$ on the alphabet $\{0,1\}$). Then, the matrix of $\mathcal{B}_k^\circ$ admits a simple Perron-Frobenius eigenvalue λ_1. By domination properties ($\Omega \neq \emptyset$), we must have $\lambda_1 < m$, where m is the cardinality of the alphabet. Aperiodicity is automatically granted. We then get by a purely qualitative argument: *The number of words of length n excluding patterns from the finite set Ω is, under the assumption of irreducibility, asymptotic to $c\lambda_1^n$, for some $c > 0$ and $\lambda_1 < \|\mathcal{A}\|$.* This last result is a strong metric form of Borges' Theorem.

The construction of a pruned automaton is clearly a generalization of the case of words obeying local constraints in Example 14 above. $\square$

$\triangleright$ **16.** *Words with excluded patterns and digital trees.* Let S be a finite set of words. An automaton recognizing S, considered as a finite language, can be constructed as a tree. The tree obtained is akin to the classical *digital tree* or *trie* that serves as a data structure for maintaining dictionaries [**75**].

A modification of the construction yields an automaton of size linear in the total number of characters that appear in words of S. [Hint. The construction can be based on the Aho–Corasick automaton [**2**]). $\triangleleft$

$\triangleright$ **17.** *Words excluding a subsequence.* The language formed of all words that do not contain $w_1 \cdots w_k$ as a subsequence (or "hidden pattern"), except at the very end, is described by the unambiguous regular expression

$$(\mathcal{A} \setminus w_1)^\star \, w_1 \, (\mathcal{A} \setminus w_2)^\star \, w_2 \cdots w_{r-1} \, (\mathcal{A} \setminus w_k)^\star \, w_k \, .$$

Assume the alphabet is endowed with a family of weights, with p_j the weight of letter $a_j \in \mathcal{A}$. The OGF $F(z)$ of words *not* containing w as a subsequence satisfies, with $q_j := 1 - p_j$,

$$F(z) = \sum_{j=1}^{k-1} \frac{(p_1 \cdots p_j) z^j}{(1 - q_1 z) \cdots (1 - q_{j+1} z)},$$

from which an asymptotic formula for $[z^n]F(z)$ derives. E.g., in the equiprobable case ($p_i = 1/m$)

$$[z^n]F(z) \sim \frac{1}{m^n} (m-1)^{n-k+1} \frac{n^{k-1}}{(k-1)!}.$$

(This problem is closely related to the discussion of pure-birth processes on page 61.) $\triangleleft$

V. 5.3. Transfer matrix methods. The transfer matrix method constitutes a variant of the modelling by deterministic automata and by paths in graphs. The very general statement of Theorem V.5 applies here with full strength. Here, we shall illustrate the situation by the width of trees following an early article by Odlyzko and Wilf [**90**] and continue with an example that draws its inspiration from the insightful exposition of domino tilings and generating functions in the book of Graham, Knuth, and Patashnik [**63**].

EXAMPLE 17. *Width of trees.* The width of a tree is defined as the maximal number of nodes that can appear on any layer at a fixed distance from the root. If a tree is drawn in the plane, then width and height can be seen as the horizontal and vertical diomensions of the bounding rectangle. Also, width is an indicator of the complexity of traversing the tree in a breadth-first search (by a queue), while height is associated to depth-first search (by a stack).

Transfer matrices are ideally suited to the problem of analysing the number of trees of fixed width. Consider a simple variety of trees corresponding to the equation $Y(z) = z\phi(Y(z))$, where the "generator" ϕ describes the formation of trees and let $Y^{[w]}(z)$ be the GF of trees of width at most w. Such trees are easily built layer by layer. Say there are k nodes at a certain level in the tree (with $1 \le k \le w$); the number of possibilities for attaching ℓ levels at the next level is the number of k-forests of depth 1 having ℓ leaves, that is, the quantity

$$t_{k,\ell} = [y^\ell]\phi(y)^k.$$

Let T be the $w \times w$ matrix with entry $T_{k,\ell} = z^\ell t_{k,\ell}$. Then, clearly, the quantity $z^i(T^h)_{i,j}$ (with $1 \le i, j \le w$) is the number of i-forests of height h, width at most w, with j nodes on level h. Thus, the GF of $\mathcal{Y}$-trees having width at most w is

$$Y^{[w]}(z) = (z, 0, 0, \ldots)(I - T)^{-1}(1, 1, 1, \ldots)^t.$$

For instance, in the case of general Catalan trees, the matrix T has the shape,

$$T = \begin{pmatrix} z\binom{1}{0} & z^2\binom{2}{0} & z^3\binom{3}{0} & z^4\binom{4}{0} \\ z\binom{2}{1} & z^2\binom{3}{1} & z^3\binom{4}{1} & z^4\binom{5}{1} \\ z\binom{3}{2} & z^2\binom{4}{2} & z^3\binom{5}{2} & z^4\binom{6}{2} \\ z\binom{4}{3} & z^2\binom{5}{3} & z^3\binom{6}{3} & z^4\binom{7}{3} \end{pmatrix},$$

for width 4. The analysis of dominant poles provides asymptotic formulae for $[z^n]Y^{[w]}(z)$:

$w=2$	$w=3$	$w=4$	$w=5$	$w=6$
$0.0085 \cdot 2.1701^n$	$0.0026 \cdot 2.8050^n$	$0.0012 \cdot 3.1638^n$	$0.0006 \cdot 3.3829^n$	$0.0004 \cdot 3.5259^n$

Additionally, the exact distribution of height in trees of size n becomes computable in polynomial time (though with a somewhat high degree polynomial).

The character of these generating functions has not been investigated in detail since the original work [**90**], so that, at the moment, analysis stops there. Fortunately, probability theory can take over the problem. Chassaing and Marckert [**20**] have shown, for Cayley trees, that the width satisfies

$$\mathbb{E}_n(W) = \sqrt{\frac{\pi}{2}} + O\left(n^{1/4}\sqrt{\log n}\right), \qquad \mathbb{P}_n(\sqrt{2}W \le x) \to 1 - \Theta(x),$$

where $\Theta(x)$ is the Theta function defined in (44). This answers very precisely an open question of Odlyzko and Wilf [**90**]. The distributional results of [**20**] extend to trees in any simple variety (under mild and natural analytic assumptions on the generator ϕ): see the paper by Chassaing, Marckert, and Yor [**21**], which builds upon earlier results of Drmota and Gittenberger [**30**]. In essence, the conclusion of these works is that the breadth first search traversal of a large tree in a simple variety gives rise to a queue whose size fluctuates asymptotically like a Brownian excursion, and is thus, in a strong sense, of a complexity comparable to depth-first search: trees don't have a preference as to the way they may be traversed. □

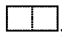
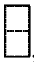
▷ **18.** *A question on width polynomials.* It is unknown whether the following assertion is true. The smallest positive root ρ_k of the denominator of $Y^{[k]}(z)$ satisfies

$$\rho_k = \rho + \frac{c}{k^2} + o(k^{-2}),$$

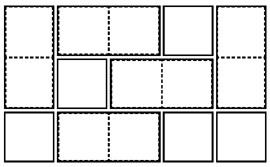
for some $c > 0$. If such an estimate holds together with suitable companion bounds, it would yield a purely analytic proof of the fact that expected width of n–trees is $\Theta(\sqrt{n})$, as well as detailed probability estimates. (The classical theory of Fredholm equations may be useful here.) ◁

EXAMPLE 18. *Monomer-dimer tilings of a rectangle.* Suppose one is given pieces that may be one of the three forms: monomers (m) that are 1×1 squares, and dimers that are dominoes, either vertically (v) oriented 1×2, or horizontally (h) oriented 2×1. In how many ways can an $n \times 3$ rectangle be covered completely and without overlap ('tiled') by such pieces?

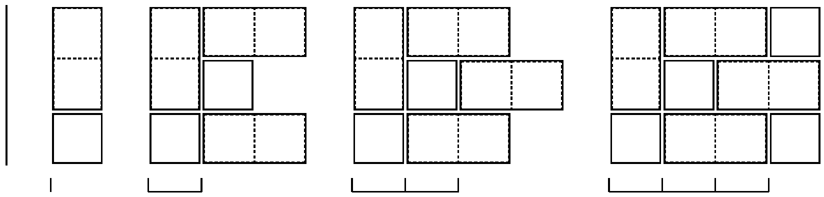
The pieces are thus of the following types,

$m =$ , $h =$ , $v =$ ,

and here is a particular tiling of a 5×3 rectangle:



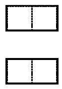
In order to approach this counting problem, one defines a class $\mathcal{C}$ of combinatorial objects called configurations. A configuration relative to an $n \times k$ rectangle is a partial tiling, such that all the first $n - 1$ columns are entirely covered by dominoes while between zero and three unit cells of the last column are covered. Here are for instance, configurations corresponding to the example above.



These diagrams suggest the way configurations can be built by successive addition of dominoes. Starting with the empty rectangle 0×3, one adds at each stage a collection of at most three dominoes in such a way that there is no overlap. This creates a configuration where, like in the example above, the dominoes may not be aligned in a flush-right manner. Continue to add successively dominoes whose left border is at abscissa $1, 2, 3$, etc, in a way that creates no internal "holes".

Depending on the state of filling of their last column, configuration can thus be classified into 8 classes that we may index in binary as $\mathcal{C}_{000}, \ldots, \mathcal{C}_{111}$. For instance $\mathcal{C}_{001}$

represent configurations such that the first two cells (from top to bottom, by convention) are free, while the third one is occupied. Then, a set of rules describes the new type of configuration obtained, when the sweep line is moved one position to the right and dominoes are added. For instance, we have

$\mathcal{C}_{010} \quad \odot$  $\Longrightarrow \quad \mathcal{C}_{101}.$

In this way, one can set up a grammar (resembling a deterministic finite automaton) that expresses all the possible constructions of longer rectangles from shorter ones according to the last layer added. The grammar comprises productions like

$$\begin{aligned} C_{000} \quad = \quad & \epsilon + \underline{mmm}C_{000} + \underline{mv}C_{000} + \underline{vm}C_{000} \\ & + \underline{\cdot mm}C_{100} + \underline{m\cdot m}C_{010} + \underline{mm\cdot}C_{001} + \underline{v\cdot}C_{001} + \underline{\cdot v}C_{100} \\ & + \underline{m\cdot\cdot}C_{011} + \underline{\cdot m\cdot}C_{101} + \underline{\cdot\cdot m}C_{110} + \underline{\cdot\cdot\cdot}C_{111}\,. \end{aligned}$$

In this grammar, a "letter" like $\underline{mv}$ represent the addition of dominoes, in top to bottom order, of types m, v respectively; the letter $\underline{m\cdot m}$ means adding two m-dominoes on the top and on the bottom, etc.

The grammar transforms into a linear system of equations with polynomial coefficients. The substitution $m \mapsto z$, $h, v \mapsto z^2$ then gives the generating functions of configurations with z marking the area covered:

$$C_{000}(z) = \frac{(1-2z^3-z^6)(1+z^3-z^6)}{(1+z^3)(1-5z^3-9z^6+9z^9+z^{12}-z^{15})}.$$

In particular, the coefficient $[z^{3n}]C_{000}(z)$ is the number of tilings of an $n \times 3$ rectangle:

$$C_{000}(z) = 1 + 3z^3 + 22z^6 + 131z^9 + 823z^{12} + 5096z^{15} + \cdots\,.$$

The sequence grows like $c\,\alpha^n$ (for $n \equiv 0 \pmod 3$) where $\alpha \doteq 1.83828$ (α is the cube root of an algebraic number of degree 5). (See [**19**] for a computer algebra session.) On average, for large n, there is a fixed proportion of monomers and the distribution of monomers in a random tiling of a large rectangle is asymptotically normally distributed, as results from the developments of Chapter IX. □

As is typical of the tiling example, one seeks to enumerate a "special" set of configurations $\mathcal{C}_f$. (In the example above, this is $\mathcal{C}_{000}$ representing complete rectangle coverings.) One determines an extended set of configurations $\mathcal{C}$ (the partial coverings, in the example) such that: (i) $\mathcal{C}$ is partitioned into finitely many classes; (ii) there is a finite set of "actions" that operate on the classes; (iii) size is affected in a well-defined additive way by the actions. The similarity with finite automata is apparent: classes play the rôle of states and actions the rôle of letters.

Often, the method of transfer matrices is used to approximate a hard combinatorial problem that is not known to decompose, the approximation being by means of a family of models of increasing "widths". For instance, the enumeration of the number T_n of tilings of an $n \times n$ square by monomers and dimers remains a famous unsolved problem of statistical physics. Here, transfer matrix methods may be used to solve the $n \times w$ version of the monomer–dimer coverings, in principle at least, for any fixed width w: the result will always be a rational function, though its degree, dicated by the dimension of the transfer matrix, will grow exponentially with w. (The "diagonal" sequence of the $n \times w$

rectangular models corresponds to the square model.) It has been at least determined by computer search that the diagonal sequence T_n starts as (this is *EIS* **A028420**):

$$1,\ 7,\ 131,\ 10012,\ 2810694,\ 2989126727,\ 11945257052321,\dots.$$

From this and other numerical data, one estimates numerically that $(T_n)^{1/n^2} \to 1.94021\dots$, but no expression for the constant is known to exist. The difficulty of coping with the finite-width models is that their complexity (as measured , e.g., by the number of states) blows up exponentially with w—such models are best treated by computer algebra; see [**118**]—and no law allowing to take a diagonal is visible. However, the finite width models have the merit of providing at least provable upper and lower bounds on the exponential growth rate of the hard "diagonal problem".

In contrast, for coverings by dimers only, a strong algebraic structure is available and the number of covers of an $n \times n$ square by horizontal and vertical dimers satisfies (n even) beautiful formula originally discovered by Kasteleyn:

$$U_{2n} = 2^{n^2/2} \prod_{j=1}^{n}\prod_{k=1}^{n}\left(\cos^2\frac{j\pi}{n+1} + \cos^2\frac{k\pi}{n+1}\right). \tag{78}$$

This sequence is *EIS* **A004003**,

$$1,\ 2,\ 36,\ 6728,\ 12988816,\ 258584046368,\ 53060477521960000,\ \dots.$$

It is elementary to prove from (78) that

$$\lim_{n\to+\infty}\left(\widehat{U}_{2n}\right)^{1/(2n)^2} = \exp\left(\frac{1}{\pi}\sum_{n=0}^{\infty}\frac{(-1)^n}{(2n+1)^2}\right) = e^{G/\pi} \doteq 1.33851\dots,$$

where G is Catalan's constant. This means in substance that each cell has a number of degrees of freedoms equivalent to 1.33851. See Percus' monograph [**95**] for proofs of this famous result and Finch's book [**35**, Sec. 5.23] for context and references.

▷ **19.** *Powers of Fibonacci numbers.* Consider the OGFs

$$G(z) := \frac{1}{1-z-z^2} = \sum_{n\ge 0} F_{n+1}z^n, \qquad G^{[k]}(z) := \sum_{n\ge 0}(F_{n+1})^k z^n,$$

where F_n is a Fibonacci number. The OGF of monomer–dimer placements on a $k \times n$ board when only monomers (m) and horizontal dimers (h) are allowed is obviously $G^{[k]}(z)$. On the other hand, it is possible to set up a transfer matrix model with state i ($0 \le i \le k$) corresponding to i positions of the current column occupied by a previous domino. Consequently,

$$G^{[k]}(z) = \text{coeff}_{k,k}(I - zT)^{-1}, \qquad \text{where} \quad T_{i,j} = \binom{i}{i+j-k},$$

for $0 \le i, j \le k$. [The denominator of $G^{[k]}(z)$ is otherwise known exactly [**74**, Ex. 1.2.8.30].] ◁

▷ **20.** *Tours on chessboards.* The OGF of Hamiltonian tours on an $n \times w$ rectangle is rational (one is allowed to move from any cell to any other vertically or horizontally adjacent cell). The same holds for king's tours and knight's tours. ◁

▷ **21.** *Cover time of graphs.* Given a fixed digraph G assumed to be strongly connected, and a designated start vertex, one travels at random, moving at each time to any neighbour of the current vertex, making choices with equal likelihood. The expectation of the time to visit all the vertices is a rational number that is effectively (though perhaps not efficiently!) computable. [Hint: set up a transfer matrix, a state of which is a subset of vertices representing those vertices that have been already visited. For an interval $[0,\,..\,m]$, this can be treated by the dedicated theory of walks on the integer interval, as in Section V. 2; for the complete graph, this is equivalent to the coupon collector

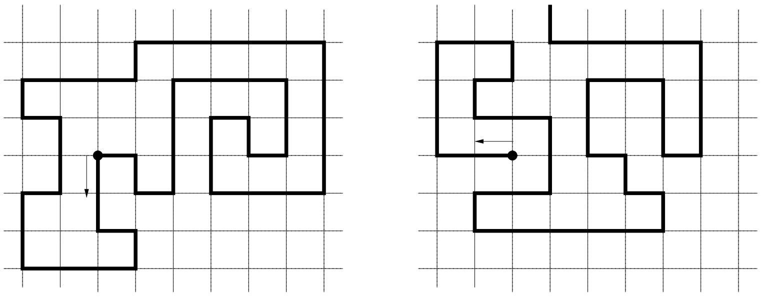
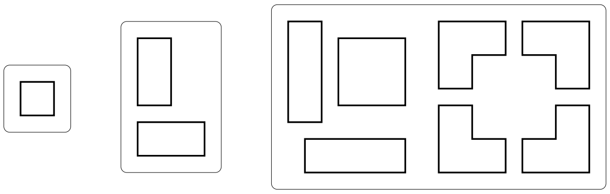


FIGURE 17. A self-avoiding polygon or SAP (left) and a self-avoiding walk or SAW (right).

problem. Most other cases are "hard" to solve analytically and one has to resort to probabilistic approximations; see Aldous and Fill's forthcoming book [**3**].] $\lhd$

EXAMPLE 19. *Self-avoiding walks and polygons.* A long standing open problem shared by statistical physics, combinatorics, and probability theory alike is that of quantifying properties of self-avoiding configurations on the square lattice (Figure 17). Here we consider objects that, starting from the origin (the "root") follow a path, and are solely composed of horizontal and vertical steps of amplitude ± 1. The *self-avoiding walk* or *SAW* can wander but is subject to the condition that it never crosses nor touches itself. The *self-avoiding polygons* or *SAPs*, whose class is denoted by $\mathcal{P}$, are self-avoiding walks, with only an exception at the end, where the end-point must coincide with the origin. We shall focus here on polygons. It proves convenient also to consider *unrooted polygons* (also called simply-connected *polyominoes*), which are polygons where the origin is discarded, so that they plainly represent the possible shapes of SAPs up to translation. For length $2n$, the number p_n of unrooted polygons satisfies $p_n = P_n/(4n)$ since the origin ($2n$ possibilities) and the starting vertex (2 possibilities) of the corresponding SAPs are disregarded in that case. Here is a table, for small values of n, listing polyominoes and the corresponding counting sequences p_n, P_n.



n:	2	3	4	5	6	7	8	9	10
p_n (*EIS* **A002931**):	1	2	7	28	124	588	2938	15268	81826
P_n (*EIS* **A010566**):	8	24	112	560	2976	16464	94016	549648	3273040

Take the (widely open) problem of determining the number P_n of SAPs of perimeter $2n$. This (intractable) problem can be approached as a limit of the (tractable) problem[7] that consists in enumerating the collection $\mathcal{P}^{[w]}$ of SAPs of width w, for increasing values of w. The latter problem is amenable to the transfer matrix method, as first discovered by Entig in 1980; see [**32**]. Indeed, take a polygon and consider a sweepline that moves from

[7]In this version of the text, we limit ourselves to a succinct description and refer to the original papers [**32, 69**] for details.

its left to its right. Once width is fixed, there are at most 2^{2w+2} possibilities for the ways a vertical sweepline may intersect the polygon's edges at half integer abscissæ. (There are $w+1$ edges and for each of these, one should "remember" whether they connect with the upper or lower boundary.) The transitions are then themselves finitely described. In this way, it becomes possible to set up a transfer matrix for any fixed width w. For fixed n, by computing values of $P_n^{[w]}$ with increasing w, one finally determines (in principle) the exact value of any P_n.

The program suggested above has been carried out to record values by the "Melbourne School" under the impulse of Tony Guttmann. For instance, Jensen [**69**] found in 2003 that the number of unrooted polygons of perimeter 100 is

$$p_{50} = 7545649677448506970646886033356862162.$$

Attaining such record values necessitates algorithms that are much more sophisticated than the naïve approach we have just described, as well as a number of highly ingenious programming optimizations.

It is an equally open problem to estimate asymptotically the number of SAPs of perimeter n. Given the exact values till perimeter 100 or more, a battery of fitting tests for asymptotic formula can be applied, leading to highly *convincing* (though still heuristic) formulæ. Thanks to several workers in this area, we can regard the final answer as "known". From the works of Jensen and his predecessors, it results that a reliable empirical estimate is of the form

$$\left\{ \begin{array}{l} p_n = B\mu^{2n}(2n)^{-\beta}(1+o(1)), \\ \mu \doteq 2.63815\,85303, \qquad \beta = -\dfrac{5}{2} \pm 3\cdot 10^{-7}, \qquad B \doteq 0.5623013. \end{array} \right.$$

Thus, the answer is almost certainly of the form $p_n \asymp \mu^{2n} n^{-5/2}$ for unrooted polygons and $P_n \asymp \mu^{2n} n^{-3/2}$ for rooted polygons. It is believed that the same connective constant μ dictates the exponential growth rate of self-avoiding walks. See Finch's book [**35**, Sec. 5.10] for a perspective and numerous references.

There is also great interest in the number $p_{m,n}$ of polyominos with perimeter $2n$ and area m, with area defined as the number of square cells composing the polyomino. Studies conducted by the Melbourne school yield numerical data that are consistent to an amazing degree (e.g., moments till order ten and small–n corrections are considered) with the following assumption: *The distribution of area in a fixed-perimeter polyomino obeys in the asymptotic limit an "Airy area distribution"*. This distribution is defined as the limit distribution of the area under Dyck paths, a problem that was briefly discussed on p. 74 and to which we propose to return in Chapter VII. See [**69, 99**] and references therein for a discussion of polyomino area. It is finally of great interest to note that the interpretation of data was strongly guided by what is already known for exactly solvable models of the type we are repeatedly considering in this book. □

V. 6. Additional constructions

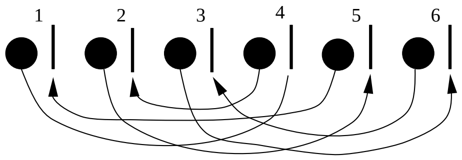
We conclude this chapter with a discussion[8] of a construction that builds on top of rational functions by means of certain transformations. Specifically, it is possible to enumerate constrained permutations by making use of the transfer matrix (or finite automaton) framework.

[8]Contents of this section are supplementary material that can be omitted on first reading.

We examine here problems whose origin lies in nineteenth century recreational mathematics. For instance, the *ménage* problem solved and popularized by Édouard Lucas in 1891, see [**23**], has the following quaint formulation: *What is the number of possible ways one can arrange n married couples (*‘ménages’*) around a table in such a way that men and women alternate, but no woman sits next to her husband?*

The ménage problem is equivalent to a permutation enumeration problem. Sit first conventionally the men at places numbered $0, \dots, n-1$, and let σ_i be the position at the right of which the ith wife is placed. Then, a ménage placement imposes the condition $\sigma_i \neq i$ and $\sigma_i \neq i+1$ for each i. We consider here a linearly arranged table (see remarks at the end for the other classical formulation that considers a round table), so that the condition $\sigma_i \neq i+1$ becomes vacuous when $i = n$. Here is a ménage placement for $n = 6$ corresponding to the permutation

$$\sigma = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 2 & 1 & 3 \end{bmatrix}$$



Clearly, this is a generalization of the derangement problem (for which the weaker condition $\sigma_i \neq i$ is imposed), where the cycle decomposition of permutations suffices to provide a direct solution (see Chapter 2).

Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, any quantity $\sigma_i - i$ is called an *exceedance* of σ. Let Ω be a finite set of integers that we assume to be nonnegative. Then a permutation is said to be Ω-avoiding if none of its exceedances lies in Ω. The counting problem, as we now demonstrate, provides an interesting case of application of the transfer matrix method.

The set Ω being fixed, consider first for all j the class of augmented permutations $\mathcal{P}_{n,j}$ that are permutations of size n such that j of the positions are distinguished and the corresponding exceedances lie in Ω, the remaining positions having arbitrary values (but with the permutation property being satisfied!). Loosely speaking, the objects in $\mathcal{P}_{n,j}$ can be regarded as permutations with “at least” j exceedances in Ω. For instance, with $\Omega = \{1\}$ and

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

there are 5 exceedances that lie in Ω (at positions $1, 2, 3, 5, 6$) and with 3 of these distinguished (say by enclosing them in a box), one obtains an element counted by $\mathcal{P}_{9,3}$ like

$$2\,\boxed{3}\,\boxed{4}\,8\,6\,\boxed{7}\,1\,5\,9.$$

Let $P_{n,j}$ be the cardinality of $\mathcal{P}_{n,j}$. We claim that the number $Q_n = Q_n^{\Omega}$ of Ω-avoiding permutations of size n satisfies

$$Q_n = \sum_{j=0}^{n} (-1)^j P_{n,j}. \tag{79}$$

Equation (79) is typically an *inclusion-exclusion* relation. To prove it formally, define the number $R_{n,k}$ of permutations that have exactly k exceedances in Ω and the generating

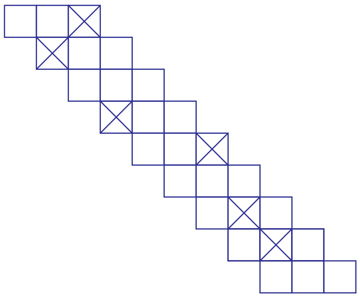


FIGURE 18. A graphical rendering of the legal template **20?02?11?** relative to $\Omega = \{0,1,2\}$.

polynomials

$$P_n(w) = \sum_j P_{n,j} w^j, \qquad R_n(w) = \sum_k R_{n,k} w^k.$$

The GF's are related by

$$P_n(w) = R_n(w+1) \qquad \text{or} \qquad R_n(w) = P_n(w-1)..$$

(The relation $P_n(w) = R_n(w+1)$ simply expresses symbolically the fact that each Ω-exceedance in $\mathcal{R}$ may or may not be taken in when composing an element of $\mathcal{P}$.) In particular, we have $P_n(-1) = R_n(0) = R_{n,0} = Q_n$ as was to be proved.

The preceding discussion shows that everything relies on the enumeration $P_{n,j}$ of permutations with distinguished exceedances in Ω. Introduce the alphabet $\mathcal{A} = \Omega \cup \{\text{'?'}\}$, where the symbol '?' is called the 'don't-care symbol'. A word on $\mathcal{A}$, an instance with $\Omega = \{0,1,2\}$ being 20?02?11?, is called a *template*. To an augmented permutation, one associates a template as follows: each exceedance that is not distinguished is represented by a don't care symbol; each distinguished exceedance (thereby an exceedance with value in Ω) is represented by its value. A template is said to be legal if it arises from an augmented permutation. For instance a template $2\,1\cdots$ cannot be legal since the corresponding constraints, namely $\sigma_1 - 1 = 2$, $\sigma_2 - 2 = 1$, are incompatible with the permutation structure (one should have $\sigma_1 = \sigma_2 = 3$). In contrast, the template 20?02?11? is seen to be legal. Figure 18 is a graphical rendering; there, letters of templates are represented by dominoes, with a cross at the position of a numeric value in Ω, and with the domino being blank in the case of a don't-care symbol.

Let $T_{n,j}$ be the set of legal templates relative to Ω that have length n and comprise j don't care symbols. Any such legal template is associated to exactly $j!$ permutations, since $n-j$ position-value pairs are fixed in the permutation, while the j remaining positions and values can be taken arbitrarily. There results that

$$P_{n,n-j} = j!\,T_{n,j} \qquad \text{and} \qquad Q_n = \sum_{j=0}^{n} (-1)^{n-j} j!\,T_{n,j}, \tag{80}$$

by (79). Thus, the enumeration of avoiding permutations rests entirely on the enumeration of legal templates.

The enumeration of legal templates is finally effected by means of a transfer matrix method, or equivalently, by a finite automaton. If a template $\tau = \tau_1 \cdots \tau_n$ is legal, then the following condition is met,

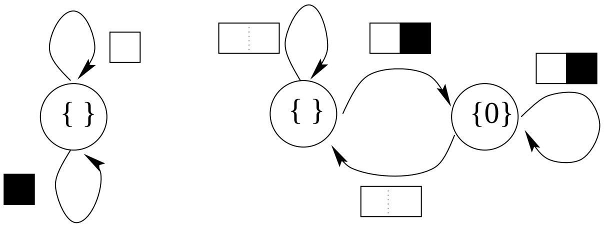
(81) $$\tau_j + j \neq \tau_i + i,$$

for all pairs (i, j) such that $i < j$ and neither of τ_i, τ_j is the don't-care symbol. (There are additional conditions to characterize templates fully, but these only concern a few letters at the end of templates and we may ignore them in this discussion.) In other words, a τ_i with a numerical value preempts the value $\tau_i + i$. Figure 18 exemplifies the situation in the case $\Omega = \{0, 1, 2\}$. The dominoes are shifted one position each time (since it is the value of $\sigma - i$ that is represented) and the compatibility constraint (81) is that no two crosses should be vertically aligned. More precisely the constraints (81) are recognized by a deterministic finite automaton whose states are indexed by subsets of $\{0, \ldots, b-1\}$ where the "span" b is defined as $b = \max_{\omega\in\Omega} \omega$. The initial state is the one associated with the empty set (no constraint is present initially), the transitions are of the form

$$\begin{cases} (q_S, j) \ \mapsto \ q_{S'} & \text{where } S' = ((S-1) \cup \{j-1\}) \cap \{0, \ldots, b-1\},\ j \neq \text{‘?’} \\ (q_S, ?) \ \mapsto \ q_{S'} & \text{where } S' = (S-1) \cap \{0, \ldots, b-1\}; \end{cases}$$

the final state is equal to the initial state (this translates the fact that no domino can protrude from the right, and is implied by the linear character of the ménage problem under consideration). In essence, the automaton only needs a finite memory since the dominoes slide along the diagonal and, accordingly, constraints older than the span can be forgotten. Notice that the complexity of the automaton, as measured by its number of states, is 2^b.

Here are the automata corresponding to $\Omega = \{0\}$ (derangements) and to $\Omega = \{0, 1\}$ (ménages).



For the ménage problem, there are two states depending on whether or not the currently examined value has been preempted at the preceding step.

From the automaton construction, the bivariate GF $T^{\Omega}(z, u)$ of legal templates, with u marking the position of don't care symbols, is a rational function that can be determined in an automatic fashion from Ω. For the derangement and ménage problems, one finds

$$T^{\{0\}}(z, u) = \frac{1}{1 - z(1+u)}, \qquad T^{\{0,1\}}(z, u) = \frac{1-z}{1 - z(2+u) + z^2}.$$

In general, this gives access to the OGF of the corresponding permutations. Consider the partial expansion of $T^{\Omega}(z, u)$ with respect to u, taken under the form

(82) $$T^{\Omega}(z, u) = \sum_r \frac{c_r(z)}{1 - u u_r(z)},$$

assuming for convenience only simple poles. There the sum is finite and it involves algebraic functions c_j and u_j of the variable z. Finally, the OGF of Ω-avoiding permutations is obtained from T^{Ω} by the transformation

$$z^n u^k \quad \mapsto \quad (-z)^n k!,$$

which is the transcription of (80). Define the (divergent) OGF of all permutations,

$$F(y) = \sum_{n=0}^{\infty} n!\, y^n = {}_2F_0[1,1;y],$$

in the terminology of hypergeometric functions. Then, by the remarks above and (82), we find

$$Q^{\Omega}(z) = \sum_r c_r(-z) F(-u_j(-z)).$$

In other words, *the OGF of Ω-avoiding permutations is a combination of compositions of the OGF of the factorial series with algebraic functions.*

The expressions simplify much in the case of ménages and derangements where the denominators of T are of degree 1 in u. One has

$$Q^{\{0\}}(z) = \frac{1}{1+z} F(\frac{z}{1+z}) = 1 + z^2 + 2z^3 + 9z^4 + 44z^5 + 265z^6 + 1854z^7 + \cdots,$$

for derangements, whence a new derivation of the known formula,

$$Q_n^{\{0\}} = \sum_{k=0}^{n} (-1)^k \binom{n}{k} (n-k)!.$$

Similarly, for (linear) ménage placements, one finds

$$Q^{\{0,1\}}(z) = \frac{1}{1+z} F(\frac{z}{(1+z)^2}) = 1 + z^3 + 3z^4 + 16z^5 + 96z^6 + 675z^7 + \cdots,$$

which is *EIS* **A00027** and corresponds to the formula

$$\mathbb{Q}_n^{\{0,1\}} = \sum_{k=0}^{n} (-1)^k \binom{2n-k}{k} (n-k)!.$$

Finally, the same techniques adapts to constraints that "wrap around", that is, constraints taken modulo n. (This corresponds to a round table in the ménage problem.) In that case, what should be considered is the loops in the automaton recognizing templates (see also the previous discussion of the zeta function of graphs). One finds in this way the OGF of the circular (i.e., classical) ménage problem to be *EIS* **A000179**,

$$\widehat{Q}^{\{0,1\}}(z) = \frac{1-z}{1+z} F(\frac{z}{(1+z)^2}) + 2z = 1 + z + z^3 + 2z^4 + 13z^5 + 80z^6 + 579z^7 + \cdots,$$

which yields the classical solution of the (circular) ménage problem,

$$\widehat{Q}_n^{\{0,1\}} = \sum_{k=0}^{n} (-1)^k \frac{2n}{2n-k} \binom{2n-k}{k} (n-k)!,$$

a formula that is due to Touchard; see [**23**, p. 185] for pointers to the vast classical literature on the subject. The algebraic part of the treatment above is close to the inspiring discussion offered in Stanley's book [**106**]. An application to robustness of interconnections in random graphs is presented in [**45**].

For asymptotic analysis purposes, the following general property proves useful: *Let F be the OGF of factorial numbers and assume that $y(z)$ is analytic at the origin where it satisfies $y(z) = z - \lambda z^2 + O(z^3)$; then it is true that*

$$[z^n]F(y(z)) \sim [z^n]F(z(1-\lambda z)) \sim n! e^{-\lambda}. \tag{83}$$

(The proof results from simple manipulations of divergent series in the style of [**9**].) This gives at sight the estimates

$$Q_n^{\{0\}} \sim ne^{-1}, \qquad Q_n^{\{0,1\}} \sim ne^{-2}.$$

More generally, for any set Ω containing λ elements, one has

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}.$$

Furthermore, the number $R^{\Omega}_{n,k}$ of permutations having exactly k occurrences (k fixed) of an exceedance in Ω is asymptotic to

$$Q_n^{\{\Omega\}} \sim ne^{-\lambda}\frac{\lambda^k}{k!}.$$

In other words, the rare event that an exceedance belongs to Ω obeys of Poisson distribution with $\lambda = |\Omega|$. These last two results are established by means of probabilistic techniques in the book [**6**, Sec. 4.3]. The relation (83) points to a way of arriving at such estimates by purely analytic-combinatorial techniques.

$\triangleright$ **22.** *Other constrained permutations.* Given a permutation $\sigma = \sigma_1 \cdots \sigma_n$, a *succession gap* is defined as any difference $\sigma_{i+1} - \sigma_i$. Discuss the counting of permutations whose succession gaps are constrained to lie outside of a finite set Ω. In how many ways can a kangaroo pass through all points of the integer interval $[1, n]$ starting at 1 and ending at n while making hops that belong to $\{-2, -1, 1, 2\}$? $\triangleleft$

$\triangleright$ **23.** *Shuffle products.* Let $\mathcal{L}, \mathcal{M}$ be two languages over two disjoint alphabets. Then, the shuffle product $\mathcal{S}$ of $\mathcal{L}$ and $\mathcal{M}$ is such that $\widehat{S}(z) = \widehat{L}(z) \cdot \widehat{M}(z)$, where $\widehat{S}, \widehat{L}, \widehat{M}$ are the exponential generating functions of $\mathcal{S}, \mathcal{L}, \mathcal{M}$. Accordingly, if the OGF $L(z)$ and $M(z)$ are rational then the OGF $S(z)$ is also rational. [This technique may be used to analyse generalized birthday paradox and coupon collector problems; see [**42**].] $\triangleleft$

V. 7. Notes

Applications of rational functions in discrete and continuosu mathematics are in abundance. Many examples are to be found in Goulden and Jackson's book [**60**]. Stanley [**106**] even devotes a full chapter of his book *Enumerative Combinatorics*, vol. I, to rational generating functions. These two books push the theory further than we can do here, but the corresponding asymptotic aspects which we expose lie outside of their scope. The analytic theory of positive rational functions starts with the works of Perron and Frobenius at the beginning of the twentieth century and is explained in books on matrix theory likes those of Bellman [**7**] and Gantmacher [**56**]. Its importance has been long recognized in the theory of finite Markov chains, so that the basic theory of positive matrices is well developed in many elementary treatises on probability theory. For such aspects, we refer for instance to the classic presentations by Feller [**34**] or Karlin and Taylor [**70**].

The supercritical sequence schema is the first in a list of abstract schemas that neatly exemplify the interplay between combinatorial, analytic, and probabilistic properties of large random structures. The origins of this approach are to be traced to early works of Bender [**8, 9**] followed by Soria and Flajolet [**51, 52, 104**].

Turning to more specific topics, we mention the first global attempt at a combinatorial theory of continued fractions by Flajolet in [**36**] together with related works of Jackson of which an exposition is to be found in [**60**, Ch. 5]. Walks on graphs are well discussed in Godsil's book [**58**]. The discussion of local constraints in permutations based on [**45**] combines the combinatorial elements bound in Stanley's book [**106**] with the general philosophy of analytic combinatorics. Our treatment of words and languages largely draws its inspiration from the line of research started by Schützenberger in the early 1960's and on the subsequent account to be found in Lothaire's book [**81**].

There are many topics that would naturally fit into this chapter but weren't ready for the present edition. Amongst the ones that may be treated (briefly) in future editions, we mention: exactly solvable models of convex polygons, the Ehrenfest urn model, random walks on undirected graphs, shuffles and Laplace transforms, variations on cycles in graphs, digital trees and the Aho-Corasick construction, the Goulden-Jackson cluster method. Future editions will be available from Philippe Flajolet's web page.

APPENDIX B

Basic Complex Analysis

1. The Gamma function. The formulæ of singularity analysis involve the *Gamma function* in an essential manner. The Gamma function extends to nonintegral arguments the factorial function and we collect in this appendix a few classical facts regarding it. Proofs may be found in classic treatises like [**66**] or [**114**]. We first list the basic function–theoretic properties. Next we prove the Hankel contour representation that illustrates a technique fundamental to singularity analysis. Last, we conclude with a few classical expansions that are of use in Chapter V.

Euler introduced the Gamma function as

$$\Gamma(s) = \int_0^\infty e^{-t} t^{s-1}\, dt, \tag{1}$$

where the integral converges provided $\Re(s) > 0$. Through integration by parts, one immediately derives the basic functional equation of the Gamma function,

$$\Gamma(s+1) = s\Gamma(s). \tag{2}$$

Since $\Gamma(1) = 1$, one has $\Gamma(n+1) = n!$, so that the Gamma function serves to extend the factorial function for nonintegral arguments.

From (2) that the Gamma function can be analytically continued to the whole of $\mathbb{C}$ with the exception of poles at $0, -1, -2, \ldots$. As $s \to -m$, with m an integer, the functional equation used backwards yields

$$\Gamma(s) \sim \frac{(-1)^m}{m!} \frac{1}{s+m}\ ,$$

so that the residue of $\Gamma(s)$ at $s = -m$ is $(-1)^m/m!$. Figure 1 depicts the graph of $\Gamma(s)$ for real values of s.

Hankel contour representation. Euler's integral representation of $\Gamma(s)$ used in conjunction with the functional equation permits us to continue $\Gamma(s)$ to the whole of the complex plane. A direct approach due to Hankel provides an alternative integral representation valid for all values of s.

THEOREM B.1 (Hankel's contour integral). *Let $\int_{+\infty}^{(0)}$ denote an integral taken along a contour starting at $+\infty$ in the upper plane, winding counterclockwise around the origin, and proceeding towards $+\infty$ in the lower half plane. Then, for all $s \in \mathbb{C}$,*

$$\frac{1}{\Gamma(s)} = -\frac{1}{2i\pi} \int_{+\infty}^{(0)} (-t)^{-s} e^{-t}\, dt. \tag{3}$$

In (3), $(-t)^{-s}$ is assumed to have its principal determination when t is negative real, and this determination is then extended uniquely by continuity throughout the contour.

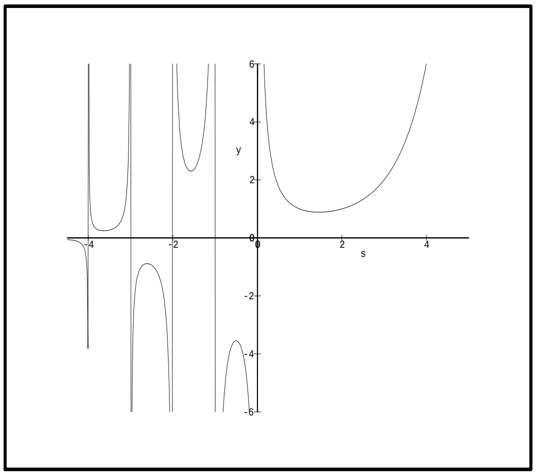


FIGURE 1. A plot of $\Gamma(s)$ for real s.

PROOF. We refer to volume 2 of Henrici's book [**66**, p. 35] or Whittaker and Watson's treatise [**114**, p. 245] for a detailed proof.

A contour of integration that fulfills the conditions of the theorem is typically the contour $\mathcal{H}$ that is at distance 1 of the positive real axis comprising three parts: a line parallel to the positive real axis in the upper half–plane; a connecting semi–circle centered at the origin; a line parallel to the positive real axis in the lower half–plane. More precisely, $\mathcal{H} = \mathcal{H}^- \cup \mathcal{H}^+ \cup \mathcal{H}^\circ$, where

$$\left\{\begin{array}{lcl} \mathcal{H}^- & = & \{z = w - i,\ w \geq 0\} \\ \mathcal{H}^+ & = & \{z = w + i,\ w \geq 0\} \\ \mathcal{H}^\circ & = & \{z = -e^{i\phi},\ \phi \in [-\frac{\pi}{2}, \frac{\pi}{2}]\}. \end{array}\right. \tag{4}$$

Let ϵ be a small positive real number, and denote by $\epsilon \cdot \mathcal{H}$ the image of $\mathcal{H}$ by the transformation $z \mapsto \epsilon z$. By analyticity, for the integral representation, we can equally well adopt as integration path the contour $\epsilon \cdot \mathcal{H}$, for any $\epsilon > 0$. The main idea is then to let ϵ tend to 0.

Assume momentarily that $s < 0$. (The extension to arbitrary s then follows by analytic continuation.) The integral along $\epsilon \cdot \mathcal{H}$ decomposes into three parts:

The integral along the semi–circle is 0 if we take the circle of a vanishing small radius, since $-s > 0$.

The contributions from the upper and lower lines give, as $\epsilon \to 0$

$$\int_{+\infty}^{(0)} (-t)^{-s} e^{-t}\, dt = (-U + L) \int_0^\infty t^{-s} e^{-t}\, dt$$

where U and L denote the determinations of $(-1)^{-s}$ on the half-lines lying in the upper and lower half planes respectively.

By continuity of determinations, $U = (e^{-i\pi})^{-s}$ and $L = (e^{+i\pi})^{-s}$. Therefore, the right hand side of (3) is equal to

$$-\frac{(-e^{i\pi s} + e^{-i\pi s})}{2i\pi}\Gamma(1-s) = -\frac{\sin(\pi s)}{\pi}\Gamma(1-s)$$

which reduces to $1/\Gamma(s)$ by the complement formula for the Gamma function. $\square$

Expansions. The Gamma function has poles at the nonpositive integers but has no zeros. Accordingly, $1/\Gamma(s)$ is an entire function with zeros at $0, -1, \ldots$, and the position of the zeros is reflected by the product decomposition,

(5) $$\frac{1}{\Gamma(s)} = se^{\gamma s} \prod_{m=1}^{\infty} \left[(1 + \frac{s}{n}) e^{-s/n} \right]$$

(of the so–called Weierstraß type). There $\gamma = 0.57721$ denotes Euler's constant

$$\gamma = \lim_{n\to\infty} (H_n - \log n) \equiv \sum_{n=1}^{\infty} \left[\frac{1}{n} - \log(1 + \frac{1}{n}) \right].$$

The logarithmic derivative of the Gamma function is classically known as the psi function and is denoted by $\psi(s)$:

$$\psi(s) = \frac{d}{ds} \log \Gamma(s) = \frac{\Gamma'(s)}{\Gamma(s)}.$$

In accordance with (5), $\psi(s)$ admits a partial fraction decomposition

(6) $$\psi(s+1) = -\gamma - \sum_{n=1}^{\infty} \left[\frac{1}{n+s} - \frac{1}{n} \right].$$

From (6), there results that the Taylor expansion of $\psi(s+1)$, hence of $\Gamma(s+1)$, involves values of the Riemann zeta function,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s},$$

at the positive integers: for $|s| < 1$,

$$\psi(s+1) = -\gamma + \sum_{n=2}^{\infty} (-1)^n \zeta(n) s^{n-1}.$$

so that the coefficients in the expansion of $\Gamma(s)$ around any integer are polynomially expressible in terms of Euler's constant γ and values of the zeta function at the integers. For instance,

$$\Gamma(s+1) = 1 - \gamma\, s + \left(\frac{\pi^2}{12} + \frac{\gamma^2}{2} \right) s^2 + \left(-\frac{\zeta(3)}{3} - \frac{\pi^2\gamma}{12} - \frac{\gamma^3}{6} \right) s^3 + \mathcal{O}(s^4).$$

Another direct consequence of the infinite product formulæ for $\Gamma(s)$ and $\sin \pi s$ is the complement formula for the Gamma function,

(7) $$\Gamma(s)\Gamma(-s) = -\frac{\pi}{s \sin \pi s},$$

which directly results from the factorization of the sine function (due to Euler),

$$\sin s = s \prod_{n=1}^{\infty} \left(1 - \frac{s^2}{n^2\pi^2} \right).$$

In particular, the complement formula entails the special value

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}.$$

▷ **1.** *The duplication formula.* This is

$$2^{2s-1}\Gamma(s)\Gamma(s + \frac{1}{2}) = \pi^{1/2}\Gamma(2s),$$

which provides the expansion of Γ near $1/2$:

$$\Gamma(s+\frac{1}{2}) = \pi^{1/2} - (\gamma + 2\,\log 2)\,\pi^{1/2}s + \left(\frac{\pi^{5/2}}{4} + \frac{(\gamma+2\,\log 2)^2\,\pi^{1/2}}{2}\right)s^2 + \mathcal{O}(s^3).$$

$\lhd$

Finally, a famous asymptotic formula is Stirling's approximation, familiarly known as "Stirling's formula":

$$\Gamma(s+1) = s\Gamma(s) \sim s^s e^{-s}\sqrt{2\pi s}\left[1 + \frac{1}{12s} + \frac{1}{288s^2} - \frac{134}{51840s^4} + \cdots\right].$$

It is valid for (large) real $s > 0$, and more generally for all $s \to \infty$ in $|\operatorname{Arg}(s)| < \pi$. For the purpose of obtaining effective bounds, the following quantitative relation [**114**, p. 253] often proves useful

$$\Gamma(s+1) = s^s e^{-s}(2\pi s)^{1/2}e^{\theta/(12s)} \text{ where } 0 < \theta < 1,$$

an equality that holds now for all $s \geq 1$.

$\rhd$ **2.** *Stirling's formula via the method of Laplace.* Stirling's formula for large s can be derived by applying Laplace's method to the integral

$$\int_0^\infty e^{-t}t^s\,dt \equiv \int_0^\infty e^{-t+s\log t}\,dt,$$

and by expanding near the maximum of the integrand, namely, $t = s$. [See [**23**, p. 267] for an explicit form of the full expansion related to derangement numbers.] $\lhd$

$\rhd$ **3.** *Stirling's formula via Euler–Maclaurin summation.* Stirling's formula can be derived from Euler–Maclaurin summation applied to $\log\Gamma(s)$. $\lhd$

2. Mellin transform. The Mellin transform of a function f defined over $\mathbb{R}_{>0}$ is the complex variable function $f^\star(s)$ defined by the integral

$$f^\star(s) := \int_0^\infty f(x)x^{s-1}\,dx. \tag{8}$$

This transform is also occasionally denoted by $\mathcal{M}[f]$ or $\mathcal{M}[f(x);s]$. Its importance devolves from two properties: (i) it *maps* asymptotic expansions of a function at 0 and $+\infty$ to singularities of the transform; (ii) it *factorizes* harmonic sums (defined below). The conjunction of the mapping property and the harmonic sum property makes it possible to analyse asymptotically rather complicated sums arising from a linear superposition of models taken at different scales. Major properties are sumamrized in Figure 2

It is assumed that f is locally integrable. Then, the two conditions,

$$f(x) \underset{x\to 0^+}{=} O(x^u), \qquad f(x) \underset{x\to+\infty}{=} O(x^v),$$

guarantee that f^* exists for s in a *strip*,

$$s \in \langle -u, -v\rangle, \qquad \text{i.e.,} \qquad -u < \Re(s) < -v.$$

Thus existence of the transform is granted provided $v < u$. The prototypical Mellin transform is the Gamma function discussed elsewhere in this appendix:

$$\Gamma(s) := \int_0^\infty e^{-x}x^{s-1}\,dx = \mathcal{M}[e^{-x};s], \qquad 0 < \Re(s) < \infty.$$

Similarly $f(x) = (1+x)^{-1}$ is $O(x^0)$ at 0 and $O(x^{-1})$ at infinity, and hence it exists in the strip $\langle -1, 0\rangle$; its transform is in fact $\pi/\sin\pi s$. The Heaviside function $H(x) := [\![0 \leq x < 1]\!]$ exists in $\langle 0 < +\infty\rangle$ and has transform $1/s$.

Harmonic sum propery. The Mellin transform is a linear transform. In addition, it satisfies the important rescaling rule

$$f(x) \overset{\mathcal{M}}{\mapsto} f^\star(s) \qquad \text{implies} \qquad f(\mu x) \overset{\mathcal{M}}{\mapsto} \mu^{-s} f^\star(s),$$

for any $\mu > 0$. Linearity then implies the derived rule

$$\sum_k \lambda_k f(\mu_k x) \overset{\mathcal{M}}{\mapsto} \left(\lambda_k \mu_k^{-s}\right) \cdot f^\star(s), \tag{9}$$

valid a priori for any finite set of pairs (λ_k, μ_k) and extending to infinite sums whenever the interchange of $\int$ and $\sum$ is permissible. A sum of the form (9) is called a *harmonic sum*, the function f is the "base function", the λ's the "amplitudes" and the μ's the "frequencies". Equation (9) then yields the "harmonic sum rule": *The Mellin transform of a harmonic sum* factorizes *as the product of the transform of the base function and a generalized Dirichlet series associated to amplitudes and frequencies*. Harmonic sums surface recurrently in the context of analytic combinatorics and Mellin transforms are a method of choice for coping with them.

Here are a few examples of application of the rule (9):

$$\begin{array}{lll} \displaystyle\sum_{k\geq 1} e^{-k^2x^2} & \mapsto \quad \frac{1}{2}\Gamma(s/2)\zeta(s) & \Re(s) > 1 \\ \displaystyle\sum_{k\geq 0} e^{-x2^k} & \mapsto \quad \dfrac{\Gamma(s)}{1-2^{-s}} & \Re(s) > 0 \\ \displaystyle\sum_{k\geq 0} (\log k) e^{-\sqrt{k}x} & \mapsto \quad -\frac{1}{2}\zeta'(s/2)\Gamma(s) & \Re(s) > 2 \\ \displaystyle\sum_{k\geq 1} \dfrac{1}{k(k+x)} & \mapsto \quad \zeta(2-s)\dfrac{\pi}{\sin \pi s} & 0 < \Re(s) < 1 \end{array}$$

$\triangleright$ **4.** *Connection between power series and Dirichlet series.* Let (f_n) be a sequence of numbers with at most polynomial growth, $f_n = O(n^r)$, and with OGF $f(z)$. Then, one has

$$\sum_{n\geq 1} \frac{f_n}{n^s} = \frac{1}{\Gamma(s)} \int_0^\infty f\left(e^{-x}\right) x^{s-1}\, dx, \qquad \Re(s) > k+1.$$

For instance, one obtains the Mellin pairs

$$\frac{e^{-x}}{1-e^{-x}} \overset{\mathcal{M}}{\mapsto} \zeta(s)\Gamma(s) \quad (\Re(s) > 1), \qquad \log \frac{1}{1-e^{x}} \overset{\mathcal{M}}{\mapsto} \zeta(s+1)\Gamma(s) \quad (\Re(s) > 0).$$

These may be used to analyse sums or, conversely, deduce analytic properties of Dirichlet series. $\triangleleft$

Mapping properties. Mellin transfoms map asymptotic terms in the expansions of a function f at 0 and $+\infty$ onto singular terms of the transform $f^\star$. This property stems from the basic identities

$$H(x)x^\alpha \overset{\mathcal{M}}{\mapsto} \frac{1}{s+\alpha}\ (s \in \langle -\alpha, +\infty\rangle), \qquad (1-H(x))x^\beta \overset{\mathcal{M}}{\mapsto} \frac{1}{s+\beta}\ (s \in \langle -\infty, -\beta\rangle),$$

as well as what one obtains by differentiation with respect to α, β.

The converse mapping property also holds. Like for other integral transforms, there is an inversion formula: if f is continuous in an interval containing x, then

$$f(x) = \frac{1}{2i\pi} \int_{c-i\infty}^{c+i\infty} f^\star(s) x^{-s}\, ds, \tag{10}$$

Function ($f(x)$)	Mellin transform ($f^\star(s)$)	
$f(x)$	$\displaystyle\int_0^\infty f(x)x^{s-1}\,dx$	**definition**, $s \in \langle -u, -v\rangle$
$\displaystyle\frac{1}{2i\pi}\int_{c-i\infty}^{c+i\infty} f^\star(s)x^{-s}\,ds$	$f^\star(s)$	**inversion th.**, $-u < c < -v$
$\displaystyle\sum_i \lambda_i f_i(x)$	$\displaystyle\sum_i \lambda_i f_i^\star(s)$	linearity
$f(\mu x)$	$\mu^{-s} f^\star(s)$	scaling rule ($\mu > 0$)
$x^\rho f(x^\theta)$	$\displaystyle\frac{1}{\theta} f^\star\left(\frac{s+\rho}{\theta}\right)$	power rule
$\displaystyle\sum_i \lambda_i f(\mu_i x)$	$\displaystyle\left(\sum_i \lambda_i \mu_i^{-s}\right)\cdot f^\star(s)$	**harmonic sum rule** ($\mu_i > 0$)
$\displaystyle\int_0^\infty \lambda(t) f(tx)\,dt$	$\displaystyle\int_0^\infty \lambda(t)t^{-s}\,dt \cdot f^\star(s)$	harmonic integral rule
$f(x)\log^k x$	$\partial_s^k f^\star(s)$	diff. I, $k \in \mathbb{Z}_{\geq 0}$, $\partial_s := \frac{d}{ds}$
$\partial_x^k f(x)$	$\displaystyle\frac{(-1)^k\Gamma(s)}{\Gamma(s-k)} f^\star(s-k)$	diff. II, $k \in \mathbb{Z}_{\geq 0}$, $\partial_x := \frac{d}{dx}$
$\displaystyle\underset{x\to 0}{\sim} x^\alpha(\log x)^k$	$\displaystyle\underset{s\to -\alpha}{\sim} \frac{(-1)^k k!}{(s+\alpha)^{k+1}}$	**mapping**: $x \to 0$, left poles
$\displaystyle\underset{x\to +\infty}{\sim} x^\beta(\log x)^k$	$\displaystyle\underset{s\to -\beta}{\sim} \frac{(-1)^{k-1} k!}{(s+\beta)^{k+1}}$	**mapping**: $x \to \infty$, right poles

FIGURE 2. Major properties of Mellin transforms.

where the abscissa c should be chosen in the "fundamental strip" of f; for instance any c satisfying $-u < c < -v$ with u, v as above is suitable.

In many cases of practical interest, $f^\star$ is continuable into meromorphic function to the whole of $\mathbb{C}$. If the continuation of $f^\star$ does not grow too fast along vertical lines, then one can estimate the inverse Mellin integral of (10) by residues. This corresponds to shifting the line of integration to some $d \neq c$ and taking poles into account by the residue theorem. Since the residue at a pole s_0 of $f^\star$ involves a factor of x^{-s_0}, the contribution of s_0 will give useful information on $f(x)$ as $x \to \infty$ if s_0 lies to the right of c, and on $f(x)$ as $x \to 0$ if s_0 lies to the left. Higher order poles introduce additional logarithmic factors. The "dictionary" is simply

$$\frac{1}{(s-s_0)^k} \overset{\mathcal{M}^{-1}}{\mapsto} \pm\frac{(-1)^k}{k!}x^{-s_0}(\log x)^k, \tag{11}$$

where the sign is '$+$' for a pole on the left of the fundamental strip and '$-$' for a pole on the right.

Mellin asymptotic summation. The combination of mapping properties and the harmonic sum property constitutes a powerful tool of asymptotic analysis. As an example, let us first investigate the pair

$$F(x) := \sum_{k\geq 1}\frac{1}{1+k^2x^2}, \qquad F^\star(s) = \frac{1}{2}\frac{\pi}{\sin\frac{1}{2}\pi s}\zeta(s),$$

where $F^\star$ results from the harmonic sum rule and is is originally defined in the strip $\langle 1, 2\rangle$. The function is meromorphically continuable to the whole of $\mathbb{C}$ with poles at the points

$0, 1, 2$ and $4, 6, 8, \ldots$. The transform $F^\star$ is small towards infinity, so that application of the dictionary (11) is justified. One then finds mechanically:

$$F(x) \underset{x\to 0^+}{\sim} \frac{\pi}{2x} - \frac{1}{2} + O(x^M), \qquad F(x) \underset{x\to+\infty}{\sim} \frac{\pi^2}{6x^2} - \frac{\pi^4}{90x^4} + \cdots,$$

for any $M > 0$.

A particularly important quantity in analytic combinatorics is the harmonic sum

$$\Phi(x) := \sum_{k=0}^{\infty} \left(1 - e^{-x/2^k}\right).$$

It occurs for instance in the analysis of longest runs on page 59. By the harmonic sum rule, one finds

$$\Phi^\star(s) = -\frac{\Gamma(s)}{1-2^s}, \qquad s \in \langle -1, 0 \rangle$$

(The transform of $e^{-x} - 1$ is also $\Gamma(s)$, but in the shifted strip $\langle -1, 0\rangle$.) The singularities of $\Phi^\star$ are at $s = 0$, where there is a double pole, at $s = -1, -2, \ldots$ which are simple poles, but also at the complex points

$$\chi_k = \frac{2ik\pi}{\log 2}.$$

The Mellin dictionary (11) can still be applied provided one integrates along a long rectangular contour that passes in-between poles. The salient feature is here the presence of fluctuations induced by the imaginary poles, since

$$x^{-\chi_k} = \exp\left(2ik\pi \log_2 x\right),$$

and each pole induces a Fourier element. All in all, one finds (any $M > 0$):

$$\text{(12)} \qquad \begin{cases} \Phi(x) & \underset{x\to+\infty}{\sim} & \log_2 x + \dfrac{\gamma}{\log 2} + \dfrac{1}{2} + P(x) + O(x^M) \\ & & P(x) \ := \ \dfrac{1}{\log 2} \displaystyle\sum_{k\in\mathbb{Z}\setminus\{0\}} \Gamma\left(\frac{2ik\pi}{\log 2}\right) e^{2i\pi \log_2 x}. \end{cases}$$

The analysis for $x \to 0$ is also possible: in this particular case, it yields

$$\Phi(x) \underset{x\to 0}{\sim} \sum_{n\geq 1} \frac{(-1)^{n-1}}{1-2^{-n}} \frac{x^n}{n!},$$

which is what would result from expanding the exponential in $\Phi(x)$ and reorganizing the terms, and consequently constitutes an exact formula.

$\triangleright$ **5.** *Mellin-type derivation of Stirling's formula.* One has the Mellin pair

$$L(x) = \sum_{k\geq 1} \log\left(1 + \frac{x}{k}\right) - \frac{x}{k}, \qquad L^\star(s) = -\frac{\pi}{s \sin \pi s} \zeta(1-s), \quad s \in \langle -2, -1 \rangle$$

Note that $L(x) = \log(e^{-\gamma x}/\Gamma(1+x))$. Mellin asymptotics provides

$$L(x) \underset{x\to+\infty}{\sim} -x\log x - (\gamma - 1)x - \frac{1}{2}\log x - \log\sqrt{2\pi} - \frac{1}{12x} + \frac{1}{360}x^3 - \frac{1}{1260x^5} + \cdots,$$

where one recognizes Stirling's expansion of $x!$,

$$\log x! \underset{x\to+\infty}{\sim} \log\left(x^x e^{-x}\sqrt{2\pi x}\right) + \sum_{n\geq 1} \frac{B_{2n}}{2n(2n-1)} x^{1-2n},$$

with B_n the Bernoulli numbers. $\triangleleft$

▷ **6.** *Mellin-type analysis of the harmonic numbers.* For $\alpha > 0$ a parameter, one has the Mellin pair:

$$K_\alpha(x) = \sum_{k\geq 1}\left(\frac{1}{k} - \frac{1}{(k+x)^\alpha}\right), \qquad K_\alpha^\star(s) = -\zeta(\alpha - s)\frac{\Gamma(s)\Gamma(\alpha-s)}{\Gamma(\alpha)}.$$

This serves to estimate harmonic numbers and their generalisations, for instance

$$H_n \underset{n\to\infty}{\sim} \log n + \gamma + \frac{1}{2n} + \sum_{k\geq 2}\frac{B_k}{k}n^{-k} \sim \log n + \gamma + \frac{1}{2n} - \frac{1}{12n^2} + \frac{1}{120n^4} - \cdots,$$

since $K_1(n) = H_n$. ◁

EXAMPLE 1. *Euler-Maclaurin summation via Mellin analysis.* Let f be continuous on $(0, +\infty)$ and satisfy $f(x) =_{x\to+\infty} O(x^{-1-\delta})$, for some $\delta > 0$, and

$$f(x) \underset{x\to 0^+}{\sim} \sum_{k=0}^{\infty} f_k x^k.$$

The summatory function $F(x)$ satisfies

$$F(x) := \sum_{n\geq 1} f(nx), \qquad F^\star(s) = \zeta(s) f^\star(s),$$

by the harmonic sum rule. By the mapping property, the collection of singular expansions of $f^\star$ at $s = 0, -1, -2, \ldots$ is summarized by the formal sum

$$f^\star(s) \asymp \left(\frac{f_0}{s}\right)_{s=0} + \left(\frac{f_1}{s+1}\right)_{s=1} + \left(\frac{f_2}{s+2}\right)_{s=1} + \cdots.$$

Thus, provided $F^\star(s)$ is small towards $\pm i\infty$ in finite strips, one has

$$F(x) \underset{x\to 0}{\sim} \frac{1}{x}\int_0^\infty f(t)\,dt + \sum_{j=0}^{\infty} f_j \zeta(-j) x^j,$$

where the main term is associated to the singularity of $F^\star$ at 1 and arises from the pole of $\zeta(s)$, with $F^\star(1)$ giving the integral of f. The interest of this approach is that it is very versatile and allows for various forms of asymptotic expansions of f at 0 as well as multipliers like $(-1)^k$, $\log k$, and so on; see [**43**] for details. □

A reference on Mellin transforms are the books by Doetsch [**29**] and Widder [**115**]. The term "harmonic sum" and some of the corresponding technology originates with the abstract [**48**]. This brief presentation is based on the survey article [**43**] to which we refer for a detailed treatment. Mellin analysis of "harmonic integrals" is a classical topic of applied mathematics for which we refer to the books by Wong [**117**] and Paris–Kaminski [**94**]. Good treatments of properties of use in discrete mathematics and analysis of algorithms now appear in the books by Hofri [**68**], Mahmoud [**84**], and Szpankowski [**108**].

3. Perron-Frobenius theory of nonnegative matrices. Perron-Frobenius theory gives access to growth properties associated to nonnegative matrices and hence to the dominant singularities of generating functions that satisfy linear systems of equations with nonnegative coefficients. Applications to rational asymptotics, paths, graphs, and automata are detailed in Chapter IV. The purpose here is only to sketch the main techniques from elementary matrix analysis that intervene in this theory. Excellent treatments are to be found in the books of Bellman [**7**, Ch. 16], Gantmacher [**56**, Ch. 13], as well as Karlin and Taylor [**70**, p. 536–551].

THEOREM B.2. *Let A be a matrix whose entries are all positive. Then, A has a unique eigenvalue $\lambda(A)$ which has greatest modulus. This eigenvalue is positive and simple.*

PROOF. The main idea consists in investigating the set of possible "expansion factors"

$$S := \left\{ \lambda \;\middle|\; \exists \mathrm{v} \geq 0, \quad A\,\mathrm{v} \geq \lambda\,\mathrm{v} \right\}. \tag{13}$$

(There $\mathrm{v} \geq 0$ means that all components of v are nonnegative and $\mathrm{v} \geq \mathrm{w}$ means that the entries of v are at least as large as the corresponding entries of w.) The largest of the expansion factors,

$$\mu := \sup(S),$$

plays a vital rôle in the argument. The proof relies on establishing that it coincides with the dominant eigenvalue $\lambda(A)$. We set $d = \dim(A)$.

Simple inequalities show that S contains at least the interval $[0, d \min_{i,j} a_{i,j}]$. Inequalities relative to the norm $\|\cdot\|_1$ show that $S \subseteq [0, \sum_{i,j} a_{i,j}]$. Thus, μ is finite and nonzero. That the supremum value μ is actually attained (i.e., $\mu \in S$) results from a simple topological argument detailed in [**7**]: take a bounded family $\mathrm{v}^{(j)}$ corresponding to a sequence $\lambda^{(j)}$ tending to μ and extract a convergent subsequence tending to a vector $\mathrm{v}^{(\infty)}$ which must then satisfy $A\,\mathrm{v}^{\infty)} \geq \mu\,\mathrm{v}^{(\infty)}$. We let w be such a vector of $\mathbb{R}^d_{\geq 0}$ satisfying $A\,\mathrm{w} \geq \mu\,\mathrm{w}$.

Next, one has $A\,\mathrm{w} = \mu\,\mathrm{w}$. Indeed, suppose a contrario that this is not the case and that, without loss of generality,

$$\sum_j A_{1,j} w_j - \mu\,\mathrm{w}_1 = \eta, \qquad \sum_j A_{i,j} w_j - \mu\,\mathrm{w}_1 \geq 0 \quad (i = 2, \ldots, d), \tag{14}$$

for $\eta > 0$. Then, given the slack afforded by η, one could construct a small perturbation $\mathrm{w}^\star$ of w (by $\mathrm{w}_j^\star = \mathrm{w}_j$ for $j = 2, \ldots, d$ and $\mathrm{w}_1^\star = w_1 + e_1/(2\mu)$) as well as a value $\mu^\star$ such that $A\,\mathrm{w}^\star \geq \mu^\star\,\mathrm{w}$ with $\mu^\star > \mu$, a contradiction. Thus, μ is an eigenvalue of A and w is an eigenvector corresponding to this eigenvalue.

Furthermore, all eigenvalues are dominated in modulus by μ. Let indeed ν and x be such that $A\,\mathrm{x} = \nu\,\mathrm{x}$. One has $A|\,\mathrm{x}\,| \geq |\nu|\,|\,\mathrm{x}\,|$, where $|\,\mathrm{x}\,|$ designates the vector whose entries are the absolute values of the corresponding entries of x. Thus, by the maximality property defining μ, one must have $|\nu| \leq \mu$. If $|\nu| = \mu$ and x is a corresponding eigenvector, then $A|\,\mathrm{x}\,| \geq \mu|\,\mathrm{x}\,|$, and by the same argument as in (14), one must have $A|\,\mathrm{x}\,| = \mu|x|$. Thus $|x|$ is also an eigenvector corresponding to μ. Then, by the triangular inequality, one has $|A\,\mathrm{x}\,| \geq A|\,\mathrm{x}\,|$, so that in fact $A|\,\mathrm{x}\,| = |A\,\mathrm{x}\,|$, which by the converse triangular inequality implies that $\mathrm{x} = \omega\,\mathrm{y}$, where $\omega \in \mathbb{C}$ and y has nonnegative entries. From this observation and the fact that $A\,\mathrm{y} = \nu\,\mathrm{y}$, it results that ν is positive real, so that $\nu = \mu$. Unicity of the dominant eigenvalue is therefore established.

Finally, simplicity of the eigenvalue μ results from a specific argument based on submatrices. If B_k is obtained from A by deleting the kth row and the kth column, then, on general grounds, one has $\lambda(A) > \lambda(B_k)$. From there, through the equality

$$-\frac{d}{d\lambda}|A - \lambda I| = |B_1 - \lambda I| + \cdots + |B_d - \lambda I|$$

(here $|A| = \det(A)$), it can be verified that the derivative of the characteristic polynomial of A at μ is strictly negative, and in particular nonzero; hence simplicity of the eigenvalue μ. See [**7**] for details. □

4. Regular expressions[1]. A *language* is a set of words over some fixed alphabet $\mathcal{A}$. The structurally simplest (yet nontrivial) languages are the *regular languages* that can be defined in a variety of ways: by regular expressions and by finite automata, either deterministic or nondeterministic.

[1]This entry supplements Appendix A: *Regular languages*.

DEFINITION B.1. *The category* RegExp *of* regular expressions *is defined by the property that it contains all the letters of the alphabet* $(a \in \mathcal{A})$ *as well as the empty symbol* ϵ*, and is such that, if* $R_1, R_2 \in \mathsf{RegExp}$*, then the formal expressions* $R_1 \cup R_2$*,* $R_1 \cdot R_2$ *and* $R_1^\star$ *are regular expressions.*

Regular expressions are meant to specify *languages*. The language $\mathcal{L}(R)$ denoted by a regular expression R is defined inductively by the rules: (i) $\mathcal{L}(R) = \{a\}$ if R is the letter $a \in \mathcal{A}$ and $\mathcal{L}(R) = \{\epsilon\}$ (with ϵ the empty word) if R is the symbol ϵ; (ii) $\mathcal{L}(R_1 \cup R_2) = \mathcal{L}(R_1) \cup \mathcal{L}(R_2)$ (with $\cup$ the set-theoretic union); (iii) $\mathcal{L}(R_1 \cdot R_2) = \mathcal{L}(R_1) \cdot \mathcal{L}(R_2)$ (with $\cdot$ the concatenation of words extended to sets); (iv) $\mathcal{L}(R_1^\star) = \{\epsilon\} + \mathcal{L}(R_1) + \mathcal{L}(R_1) \cdot \mathcal{L}(R_1) + \cdots$. A language is said to be a *regular language* if it is specified by a regular expression.

A language is a *set* of words, but a word $w \in \mathcal{L}(R)$ may be parsable in several ways according to R. More precisely, one defines the *ambiguity coefficient* (or *multiplicity*) of w with respect to the regular expression R as the number of parsings, written $\kappa(w) = \kappa_R(w)$. In symbols, we have

$$\kappa_{R_1 \cup R_2}(w) = \kappa_{R_1}(w) + \kappa_{R_2}(w), \quad \kappa_{R_1 \cdot R_2}(w) = \sum_{u \cdot v = w} \kappa_{R_1}(u) \kappa_{R_2}(v),$$

with natural initial conditions ($\kappa_a(b) = \delta_{a,b}$, $\kappa_\epsilon(w) = \delta_{\epsilon,w}$), and with the definition of $\kappa_{R^\star(w)}$ taken as induced by the definition of $R^\star$ via unions and products, namely,

$$\kappa_{R^\star}(w) = \delta_{\epsilon,w} + \sum_{j=1}^{\infty} \kappa_{R^j}(w).$$

As such, $\kappa(w)$ lies in the completed set $\mathbb{N} \cup \{+\infty\}$. We shall only consider here regular expressions R that are *proper*, in the sense that $\kappa_R(w) < +\infty$. It can be checked that this condition is equivalent to requiring that no $S^\star$ with $\epsilon \in \mathcal{L}(S)$ enters in the inductive definition of the regular expression R. (This condition is substantially equivalent to the notion of well-founded specification in Chapter 1.) A regular expression R is said to be *unambiguous* iff for all w, we have $\kappa_R(w) \in \{0, 1\}$; it is said to be *ambiguous*, otherwise.

Given a language $L = \mathcal{L}(R)$, we are interested in two enumerating sequences

$$L_{R,n} = \sum_{|w|=n} \kappa_R(w), \qquad L_n = \sum_{|w|=n} \mathbf{1}_{w \in L},$$

corresponding to the counting of words in the language, respectively, with and without multiplicities.

Bibliography

1. Milton Abramowitz and Irene A. Stegun, *Handbook of mathematical functions*, Dover, 1973, A reprint of the tenth National Bureau of Standards edition, 1964.
2. Alfred V. Aho and Margaret J. Corasick, *Efficient string matching: an aid to bibliographic search*, Communications of the ACM **18** (1975), 333–340.
3. David Aldous and James A. Fill, *Reversible Markov chains and random walks on graphs*, 2003, Book in preparation; mansucript available electronically.
4. George E. Andrews, *The theory of partitions*, Encyclopedia of Mathematics and its Applications, vol. 2, Addison–Wesley, 1976.
5. George E. Andrews, Richard Askey, and Ranjan Roy, *Special functions*, Encyclopedia of Mathematics and its Applications, vol. 71, Cambridge University Press, 1999.
6. A. D. Barbour, Lars Holst, and Svante Janson, *Poisson approximation*, The Clarendon Press Oxford University Press, New York, 1992, Oxford Science Publications.
7. Richard Bellman, *Matrix analysis*, S.I.A.M. Press, Philadelphia, Pa, 1997, A reprint of the second edition, first published by McGraw-Hill, New York, 1970.
8. Edward A. Bender, *Central and local limit theorems applied to asymptotic enumeration*, Journal of Combinatorial Theory **15** (1973), 91–111.
9. ______, *Asymptotic methods in enumeration*, SIAM Review **16** (1974), no. 4, 485–515.
10. Edward A. Bender and Fred Kochman, *The distribution of subword counts is usually normal*, European Journal of Combinatorics **14** (1993), 265–275.
11. Bruce C. Berndt, *Ramanujan's notebooks, part I*, Springer Verlag, 1985.
12. Jean Bertoin, Philippe Biane, and Marc Yor, *Poissonian exponential functionals, q-series, q-integrals, and the moment problem for log-normal distributions*, Tech. Report PMA-705, Laboratoire de Probabilitś et Modèles Aléatoires, Université Paris VI, 2002.
13. Alberto Bertoni, Christian Choffrut, Massimiliano Goldwurm Goldwurm, and Violetta Lonati, *On the number of occurrences of a symbol in words of regular languages*, Theoretical Computer Science **302** (2003), no. 1–3, 431–456.
14. A. T. Bharucha-Reid, *Elements of the theory of Markov processes and their applications*, Dover, 1997, A reprint of the original McGraw-Hill edition, 1960.
15. Philippe Biane, Jim Pitman, and Marc Yor, *Probability laws related to the Jacobi theta and Riemann zeta functions, and Brownian excursions*, Bulletin of the American Mathematical Society (N.S.) **38** (2001), no. 4, 435–465.
16. Jérémie Bourdon and Brigitte Vallée, *Generalized pattern matching statistics*, Mathematics and computer science, II (Versailles, 2002) (B. Chauvin *et al.*, ed.), Trends Math., Birkhäuser, Basel, 2002, pp. 249–265.
17. Henri Cartan, *Théorie élémentaire des fonctions analytiques d'une ou plusieurs variables complexes*, Hermann, 1961.
18. Pierre Cartier and Dominique Foata, *Problèmes combinatoires de commutation et réarrangements*, Lecture Notes in Mathematics, vol. 85, Springer Verlag, 1969.
19. Frédéric Cazals, *Monomer-dimer tilings*, Studies in Automatic Combinatorics **2** (1997), Electronic publication `http://algo.inria.fr/libraries/autocomb/autocomb.html`.
20. Philippe Chassaing and Jean-François Marckert, *Parking functions, empirical processes, and the width of rooted labeled trees*, Electronic Journal of Combinatorics **8** (2001), no. 1, Research Paper 14, 19 pp. (electronic). MR **2002b:**05037
21. Philippe Chassaing, Jean-François Marckert, and Marc Yor, *The height and width of simple trees*, Mathematics and computer science (Versailles, 2000), Trends Math., Birkhäuser, Basel, 2000, pp. 17–30.
22. T. S. Chihara, *An introduction to orthogonal polynomials*, Gordon and Breach, New York, 1978.
23. Louis Comtet, *Advanced combinatorics*, Reidel, Dordrecht, 1974.
24. N. G. de Bruijn, *A combinatorial problem*, Nederl. Akad. Wetensch., Proc. **49** (1946), 758–764, Also in *Indagationes Math.* **8**, 461–467 (1946).
25. ______, *Asymptotic methods in analysis*, Dover, 1981, A reprint of the third North Holland edition, 1970 (first edition, 1958).
26. 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.

27. Michael Dellnitz, Oliver Schütze, and Qinghua Zheng, *Locating all the zeros of an analytic function in one complex variable*, J. Comput. Appl. Math. **138** (2002), no. 2, 325–333.
28. Jean Dieudonné, *Calcul infinitésimal*, Hermann, Paris, 1968.
29. G. Doetsch, *Handbuch der Laplace Transformation,* vol. 1–3, Birkhäuser Verlag, Basel, 1955.
30. Michael Drmota and Bernhard Gittenberger, *On the profile of random trees*, Random Structures & Algorithms **10** (1997), no. 4, 421–451.
31. Philippe Duchon, Philippe Flajolet, Guy Louchard, and Gilles Schaeffer, *Boltzmann samplers for the random generation of combinatorial structures*, Preprint, 2003, 44 pages. To appear in *Combinatorics, Probability and Computing*.
32. I. G. Entig, *Generating functions for enumerating self-avoiding rings on the square lattice*, Journal of Physics A: Mathematical and General **18** (1980), 3713–3722.
33. H. M. Farkas and I. Kra, *Riemann surfaces*, second ed., Graduate Texts in Mathematics, vol. 71, Springer-Verlag, New York, 1992.
34. W. Feller, *An introduction to probability theory and its applications*, third ed., vol. 1, John Wiley, 1968.
35. Steven Finch, *Mathematical constants*, Cambridge University Press, New-York, 2003.
36. Philippe Flajolet, *Combinatorial aspects of continued fractions*, Discrete Mathematics **32** (1980), 125–161.
37. ______, *Approximate counting: A detailed analysis*, BIT **25** (1985), 113–134.
38. ______, *Elements of a general theory of combinatorial structures*, Fundamentals of Computation Theory (Lothar Budach, ed.), Lecture Notes in Computer Science, vol. 199, Springer Verlag, 1985, Proceedings of FCT'85, Cottbus, GDR, September 1985 (Invited Lecture), pp. 112–127.
39. ______, *Singularity analysis and asymptotics of Bernoulli sums*, Theoretical Computer Science **215** (1999), no. 1-2, 371–381.
40. Philippe Flajolet, Jean Françon, and Jean Vuillemin, *Sequence of operations analysis for dynamic data structures*, Journal of Algorithms **1** (1980), 111–141.
41. Philippe Flajolet, Zhicheng Gao, Andrew Odlyzko, and Bruce Richmond, *The distribution of heights of binary trees and other simple trees*, Combinatorics, Probability and Computing **2** (1993), 145–156.
42. Philippe Flajolet, Danièle Gardy, and Loÿs Thimonier, *Birthday paradox, coupon collectors, caching algorithms, and self–organizing search*, Discrete Applied Mathematics **39** (1992), 207–229.
43. Philippe Flajolet, Xavier Gourdon, and Philippe Dumas, *Mellin transforms and asymptotics: Harmonic sums*, Theoretical Computer Science **144** (1995), no. 1–2, 3–58.
44. Philippe Flajolet and Fabrice Guillemin, *The formal theory of birth-and-death processes, lattice path combinatorics, and continued fractions*, Advances in Applied Probability **32** (2000), 750–778.
45. Philippe Flajolet, Kostas Hatzis, Sotiris Nikoletseas, and Paul Spirakis, *On the robustness of interconnections in random graphs: A symbolic approach*, Theoretical Computer Science **287** (2002), no. 2, 513–534.
46. Philippe Flajolet, Peter Kirschenhofer, and Robert F. Tichy, *Deviations from uniformity in random strings*, Probability Theory and Related Fields **80** (1988), 139–150.
47. 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.
48. Philippe Flajolet, Mireille Régnier, and Robert Sedgewick, *Some uses of the Mellin integral transform in the analysis of algorithms*, Combinatorial Algorithms on Words (A. Apostolico and Z. Galil, eds.), NATO Advance Science Institute Series. Series F: Computer and Systems Sciences, vol. 12, Springer Verlag, 1985, (Invited Lecture), pp. 241–254.
49. Philippe Flajolet, Bruno Salvy, and Paul Zimmermann, *Automatic average–case analysis of algorithms*, Theoretical Computer Science **79** (1991), no. 1, 37–109.
50. Philippe Flajolet and Robert Sedgewick, *Mellin transforms and asymptotics: finite differences and Rice's integrals*, Theoretical Computer Science **144** (1995), no. 1–2, 101–124.
51. Philippe Flajolet and Michèle Soria, *Gaussian limiting distributions for the number of components in combinatorial structures*, Journal of Combinatorial Theory, Series A **53** (1990), 165–182.
52. ______, *General combinatorial schemas: Gaussian limit distributions and exponential tails*, Discrete Mathematics **114** (1993), 159–180.
53. Philippe Flajolet, Wojciech Szpankowski, and Brigitte Vallée, *Hidden word statistics*, Preprint, 2002, 34 pages; submitted to *J. of the ACM*.
54. W. B. Ford, *Studies on divergent series and summability and the asymptotic developments of functions defined by Maclaurin series*, 3rd ed., Chelsea Publishing Company, New York, 1960, (From two books originally published in 1916 and 1936.).
55. G. Frobenius, *Über Matrizen aus nicht negativen Elementen*, Sitz.-Ber. Akad. Wiss., Phys-Math Klasse, Berlin (1912), 456–477.
56. F. R. Gantmacher, *Matrizentheorie*, Deutscher Verlag der Wissenschaften, Berlin, 1986, A translation of the Russian original Teoria Matriz, Nauka, Moscow, 1966.
57. George Gasper and Mizan Rahman, *Basic hypergeometric series*, Encyclopedia of Mathematics and its Applications, vol. 35, Cambridge University Press, 1990.
58. C. D. Godsil, *Algebraic combinatorics*, Chapman and Hall, 1993.
59. Massimiliano Goldwurm and Massimo Santini, *Clique polynomials have a unique root of smallest modulus*, Information Processing Letters **75** (2000), no. 3, 127–132.

60. Ian P. Goulden and David M. Jackson, *Combinatorial enumeration*, John Wiley, New York, 1983.
61. Xavier Gourdon, *Largest component in random combinatorial structures*, Discrete Mathematics **180** (1998), no. 1-3, 185–209.
62. Xavier Gourdon and Bruno Salvy, *Asymptotics of linear recurrences with rational coefficients*, Tech. Report 1887, INRIA, March 1993, To appear in Proceedings FPACS'93.
63. Ronald L. Graham, Donald E. Knuth, and Oren Patashnik, *Concrete mathematics*, Addison Wesley, 1989.
64. Fabrice Guillemin, Philippe Robert, and Bert Zwart, *AIMD algorithms and exponential functionals*, In press, 2003, To appear in *Annals of Applied Probability*. 20 pages.
65. G. H. Hardy, *Ramanujan: Twelve lectures on subjects suggested by his life and work*, third ed., Chelsea Publishing Company, New-York, 1978, Reprinted and Corrected from the First Edition, Cambridge, 1940.
66. Peter Henrici, *Applied and computational complex analysis*, vol. 2, John Wiley, New York, 1974.
67. E. Hille, *Analytic function theory*, Blaisdell Publishing Company, Waltham, 1962, 2 Volumes.
68. Micha Hofri, *Analysis of algorithms: Computational methods and mathematical tools*, Oxford University Press, 1995.
69. Iwan jensen, *A parallel algorithm for the enumeration of self-avoiding polygons on the square lattice*, Journal of Physics A: Mathematical and General **36** (2003), 5731–5745.
70. Samuel Karlin and Howard Taylor, *A first course in stochastic processes*, second ed., Academic Press, 1975.
71. Arnold Knopfmacher and Helmut Prodinger, *On Carlitz compositions*, European Journal of Combinatorics **19** (1998), no. 5, 579–589.
72. K. Knopp, *Theory of functions*, Dover Publications, New York, 1945.
73. Donald E. Knuth, *The average time for carry propagation*, Indagationes Mathematicae **40** (1978), 238–242.
74. ______, *The art of computer programming*, 3rd ed., vol. 1: Fundamental Algorithms, Addison-Wesley, 1997.
75. ______, *The art of computer programming*, 2nd ed., vol. 3: Sorting and Searching, Addison-Wesley, 1998.
76. ______, *The art of computer programming*, 3rd ed., vol. 2: Seminumerical Algorithms, Addison-Wesley, 1998.
77. M. Krasnoselskii, *Positive solutions of operator equations*, P. Noordhoff, Groningen, 1964.
78. Christian Krattenthaler, *Advanced determinant calculus*, Seminaire Lotharingien de Combinatoire **42** (1999), Paper B42q, 66 pp.
79. J. C. Lagarias, A. M. Odlyzko, and D. B. Zagier, *On the capacity of disjointly shared networks*, Computer Netwoks **10** (1985), no. 5, 275–285.
80. Ernst Lindelöf, *Le calcul des résidus et ses applications à la théorie des fonctions*, Collection de monographies sur la théorie des fonctions, publiée sous la direction de M. Émile Borel, Gauthier-Villars, Paris, 1905, Reprinted by Gabay, Paris, 1989.
81. M. Lothaire, *Combinatorics on words*, Encyclopedia of Mathematics and its Applications, vol. 17, Addison–Wesley, 1983.
82. G. Louchard, *Random walks, Gaussian processes and list structures*, Theoretical Computer Science **53** (1987), no. 1, 99–124.
83. Guy Louchard and Helmut Prodinger, *Probabilistic analysis of Carlitz compositions*, Discrete Mathematics & Theoretical Computer Science **5** (2002), no. 1, 71–96.
84. Hosam M. Mahmoud, *Evolution of random search trees*, John Wiley, New York, 1992.
85. A. Meir and J. W. Moon, *On the altitude of nodes in random trees*, Canadian Journal of Mathematics **30** (1978), 997–1015.
86. J. W. Moon, *Counting labelled trees*, Canadian Mathematical Monographs N.1, William Clowes and Sons, 1970.
87. Niels Erik Nörlund, *Vorlesungen über Differenzenrechnung*, Chelsea Publishing Company, New York, 1954.
88. A. M. Odlyzko, *Periodic oscillations of coefficients of power series that satisfy functional equations*, Advances in Mathematics **44** (1982), 180–205.
89. ______, *Asymptotic enumeration methods*, Handbook of Combinatorics (R. Graham, M. Grötschel, and L. Lovász, eds.), vol. II, Elsevier, Amsterdam, 1995, pp. 1063–1229.
90. A. M. Odlyzko and H. S. Wilf, *Bandwidths and profiles of trees*, Journal of Combinatorial Theory, Series B **42** (1987), 348–370.
91. ______, *The editor's corner: n coins in a fountain*, American Matematical Monthly **95** (1988), 840–843.
92. Richard Otter, *The number of trees*, Annals of Mathematics **49** (1948), no. 3, 583–599.
93. D. Panario and B. Richmond, *Exact largest and smallest size of components*, Algorithmica **31** (2001), no. 3, 413–432. MR 1 855 258
94. R. B. Paris and D. Kaminski, *Asymptotics and Mellin-Barnes integrals*, Encyclopedia of Mathematics and its Applications, vol. 85, Cambridge University Press, Cambridge, 2001.
95. J. K. Percus, *Combinatorial Methods*, Applied Mathematical Sciences, vol. 4, Springer-Verlag, 1971.
96. Oskar Perron, *Über Matrizen*, Mathematische Annalen **64** (1907), 248–263.
97. G. Pólya, *Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen*, Acta Mathematica **68** (1937), 145–254.
98. G. Pólya and R. C. Read, *Combinatorial enumeration of groups, graphs and chemical compounds*, Springer Verlag, New York, 1987.

99. Christoph Richard, *Scaling behaviour of two-dimensional polygon models*, Journal of Statistical Physics **108** (2002), no. 3/4, 459–493.
100. Robert Sedgewick and Philippe Flajolet, *An introduction to the analysis of algorithms*, Addison-Wesley Publishing Company, 1996.
101. L. A. Shepp and S. P. Lloyd, *Ordered cycle lengths in a random permutation*, Transactions of the American Mathematical Society **121** (1966), 340–357.
102. N. J. A. Sloane, *The on-line encyclopedia of integer sequences*, 2000, Published electronically at `http://www.research.att.com/~njas/sequences/`.
103. N. J. A. Sloane and Simon Plouffe, *The encyclopedia of integer sequences*, Academic Press, 1995.
104. Michèle Soria-Cousineau, *Méthodes d'analyse pour les constructions combinatoires et les algorithmes*, Doctorat ès sciences, Université de Paris–Sud, Orsay, July 1990.
105. George Springer, *Introduction to Riemann surfaces*, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1957, Reprinted by Chelsea, New York.
106. Richard P. Stanley, *Enumerative combinatorics*, vol. I, Wadsworth & Brooks/Cole, 1986.
107. Gabor Szegő, *Orthogonal polynomials*, Americam Mathematical Society Colloquium Publications, vol. XXIII, A.M.S, Providence, 1989.
108. Wojciech Szpankowski, *Average-case analysis of algorithms on sequences*, John Wiley, New York, 2001.
109. E. C. Titchmarsh, *The theory of functions*, second ed., Oxford University Press, 1939.
110. Joris van der Hoeven, *Majorants for formal power series*, Preprint, 2003, 29 pages. Available from author's webpage.
111. Robert Vein and Paul Dale, *Determinants and their applications in mathematical physics*, Applied Mathematical Sciences, vol. 134, Springer-Verlag, 1998.
112. 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.
113. H. S. Wall, *Analytic theory of continued fractions*, Chelsea Publishing Company, 1948.
114. E. T. Whittaker and G. N. Watson, *A course of modern analysis*, fourth ed., Cambridge University Press, 1927, Reprinted 1973.
115. David Vernon Widder, *The Laplace transform*, Princeton University Press, 1941.
116. Herbert S. Wilf, *Generatingfunctionology*, Academic Press, 1990.
117. Roderick Wong, *Asymptotic approximations of integrals*, Academic Press, 1989.
118. Doron Zeilberger, *Symbol-crunching with the transfer-matrix method in order to count skinny physical creatures*, Integers **0** (2000), Paper A9, Published electronically at `http://www.integers-ejcnt.org/vol0.html`.

Index