# The Average Case Analysis of Algorithms : Multivariate Asymptotics and Limit Distributions

Philippe Flajolet, Robert Sedgewick

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET AUTOMATIQUE

# The Average Case Analysis of Algorithms: Multivariate Asymptotics and Limit Distributions

Philippe FLAJOLET, Robert SEDGEWICK

**N° 3162**

Mai 1997

THÈME 2

Rapport de recherche

ISSN 0249-6399

# THE AVERAGE CASE ANALYSIS OF ALGORITHMS:
## *Multivariate Asymptotics and Limit Distributions*

PHILIPPE FLAJOLET[1] & ROBERT SEDGEWICK[2]

**Abstract.** *This report is part of a series whose aim is to present in a synthetic way the major methods of "analytic combinatorics" needed in the average–case analysis of algorithms. It develops a general approach to the distributional analysis of parameters of elementary combinatorial structures like strings, trees, graphs, permutations, and so on. The methods are essentially analytic and relie on multivariate generating functions, singularity analysis, and continuity theorems. The limit laws that are derived mostly belong to the Gaussian, Poisson, or geometric type.*

---

# L'ANALYSE EN MOYENNE D'ALGORITHMES:
## *Asymptotique multivariée et distributions limites*

**Résumé.** Ce rapport fait partie d'une série dont le but est de présenter de manière unifiée les principales méthodes de "combinatoire analytique" utiles à l'analyse d'algorithmes. On y développe une approche générale à l'analyse en distribution des paramètres de structures combinatoires élémentaires, telles les mots, arbres, graphes, permutations, etc. Les méthodes utilisées reposent sur les séries génératrices multivariées, l'analyse de singularité et les théorèmes de continuité. Les lois limites qui apparaissent sont fréquemment de type gaussien, poissonnien, ou géométrique.

[1]Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay (France)
[2]Department of Computer Science, Princeton University, Princeton, New-Jersey 08544 (USA)

# Foreword

This report is part of a series whose aim is to present in a synthetic way the major methods and models in analytic combinatorics and the analysis of algorithms. The series comprises the following chapters:

1. Symbolic Enumeration and Ordinary Generating Functions;
2. Labelled Structures and Exponential Generating Functions;
3. Parameters and Multivariate Generating Functions;
4. Complex Asymptotic Methods;
5. Singularity Analysis of Generating Functions;
6. Saddle Point Asymptotics;
7. Mellin Transform Asymptotics;
8. Functional Equations and Generating Functions (in preparation);
9. Multivariate Asymptotics and Limit Distributions (this report);
10. Anaytic Combinatorics of Classical Structures (in preparation).

For partly historical reasons, the general heading of this series is "The Average Case Analysis of Algorithms". The whole series, after suitable editing, is destined to be transformed into a book with the titlle

*"Analytic Combinatorics"*

The present report *"Multivariate Asymptotics and Limit Distributions"* constitutes Chapter 9. The following chapters have already appeared as INRIA Research Report under the global heading, "The Average Case Analysis of Algorithms": Chapters 1–3 ("Counting and Generating Functions", RR 1888, 116 pages, 1993); Chapters 4–5 ("Complex Asymptotics and Generating Functions", RR 2026, 100 pages, 1993); Chapter 6 ("Saddle Point Asymptotics", RR 2376, 55 pages, 1994); Chapter 7 ("Mellin Transform Asymptotics", RR 2956, 93 pages, 1996).

**Acknowledgements.** The work of Philippe Flajolet was supported in part by the Long Term Research Project of the European Union ALCOM-IT (# 20244).

The authors are grateful to Hsien-Kuei Hwang and Michèle Soria for permission to include results from the joint paper [21].

# Chapter 9

# Multivariate Asymptotics and Limit Distributions

*Un problème relatif aux jeux du hasard,*
*proposé à un austère janseniste par un homme du monde,*
*a été à l'origine du Calcul des Probabilités.*

— Siméon–Denis Poisson

Analytic combinatorics concerns the enumeration of combinatorial structures in relation to algebraic and analytic properties of generating functions. Previous chapters have been mostly devoted to finding the number of objects of a given size in a combinatorial class, by an analysis of singularities or saddle points. This is a typically "univariate" problem, as it only involves the given size of combinatorial objects to be enumerated and, accordingly, properties of generating functions in a single variable.

Many applications, in discrete mathematics, statistical physics, or analysis of algorithms, require investigating *parameters* of combinatorial structures. It is for instance useful to know that a random permutation of size $n$ has a mean number of runs equal to $(n+1)/2$, still a univariate problem. But it may be equally important to know that such an average is *highly representative* of what occurs in simulations or on actual data that obey the randomness model. For instance, for runs and a permutation of size $n = 1,000$, the probability is less than $10^{-6}$ to observe a case that deviates by more than 10% from the mean value, and the probability decays to less than $10^{-65}$ for $n = 10,000$, and to less than $10^{-653}$ (!!) for $n = 100,000$. Such problems are taking us to the realm of bivariate asymptotics.

A first task consists in locating regions to which the variations of a parameter are most likely to be confined. The analysis of the mean value, standard deviation, and possibly higher moments, provides a simple framework and moment inequalities are often sufficient to prove that a parameter tends to be rather narrowly concentrated around its mean. Such inequalities go back to Chebyshev in the 19th century. They are an important component of the "probabilistic method" that has been so successful in the study of random graphs and hard combinatorial optimization problems. We shall see here that moment methods merge nicely with bivariate generating functions and yield concentration of distribution for many combinatorial parameters like cycles and runs in permutations, leaves in trees of various sorts, patterns in strings, etc.

We already know from Chapter 3 that bivariate generating functions are obtained as soon as problems are sufficiently decomposable. This chapter can be viewed as a collection of analytic techniques for extracting coefficients of bivariate analytic functions, by adapting methods of classical analysis and probability theory.

The histograms of the distribution of a combinatorial parameter (for varying size values) often exhibit a common characteristic "shape". In this case, there is a *limit law* that governs the distribution in the asymptotic limit. In simpler cases, limit laws are *discrete* and, when this happens, they often belong to the geometric or Poisson type. We establish this fact under rather general analytic conditions.

In many cases, limit laws are *continuous*. A fair fraction of this chapter will be devoted to the Gaussian law —described by the famous bell-shaped curve— since it appears so frequently in elementary combinatorial structures. Roughly, we regard here a bivariate generating function as a collection of univariate generating functions, each treated individually by singularity analysis or saddle point techniques. Large powers then tend to occur for coefficients of these GFs —think of quantities of the form $\approx \rho^{-n}$ that arise from radius of convergence bounds. From there, a Gaussian laws is derived along lines that are reminiscent of the central limit theorem of probability theory.

In both the discrete and continuous case, the auxiliary variable marking the combinatorial parameter under study is essentially viewed as a singularity "modifier". Hence the name of singularity perturbation analysis given to this approach.

In fact, almost any classical law of probability theory and statistics is likely to occur somewhere in analytic combinatorics. Conversely, almost any simple combinatorial parameter is likely to be governed by an asymptotic law. The goal of this chapter is to provide fundamental analytic techniques for these facts. It is based on a complete chain

$$\begin{array}{c}\text{Combinatorial structure (constructions)}\\ \downarrow \\ \text{Analytic structure (bivariate generating functions)}\\ \downarrow \\ \text{Singular structure (singularity perturbation analysis)}\\ \downarrow \\ \text{Limit laws.}\end{array}$$

The direct relation that can be established between combinatorial specifications and asymptotic properties, in the form of limit laws, is especially striking here, and it is a characteristic feature of analytic combinatorics.

## 9.1 Moments and combinatorial distributions

A preamble to almost any fine investigation of a combinatorial distribution is the analysis of its moments, especially the first two moments, or equivalently the mean and the standard deviation. These quantities constrain the probability distribution and, for instance, the standard deviation is an indication of plausible dispersions from the mean. Also, asymptotic analysis of the first two moments often provides useful indications regarding the existence of limit laws of either the discrete or continuous type. Here, we start by recalling basic definitions relative to discrete random variables, and specialize the discussion to combinatorial parameters. We then discuss briefly moment inequalities that are an important component of the so-called "probabilistic method". Finally, we examine moment analyses based on bivariate generating functions and discuss several combinatorial applications that are typical.

### 9.1.1 Discrete combinatorial distributions

A random variable $Y$ that is supported by the nonnegative integers is specified by its probability distribution,

$$p_k = \Pr\{Y = k\},$$

or equivalently by its distribution function, a step function,

$$F(x) = \Pr\{Y \le x\} = \sum_{k \le x} p_k.$$

To the distribution of $Y$ is associated the probability generating functions (PGF),

$$p(u) = \sum_{k\geq 0} p_k u^k,$$

a function that is *a priori* defined in $|u| \leq 1$ and analytic in $|u| < 1$, since $\sum_k p_k = 1$. The PGF completely determines the probability distribution,

$$p_k = \frac{1}{k!} \left. \frac{d^k}{du^k} p(u) \right|_{u=0}.$$

The expectation $E\{h\}$ of a function $h(Y)$ of the random variable $Y$ is defined by

$$E\{h(Y)\} = \sum_{k\geq 0} h(k) p_k,$$

and the moment of order $r$ of $Y$,

$$\mu^{(r)} = E\{Y^r\}, \tag{9.1}$$

is the expectation of $Y^r$ The PGF also determines the moments,

$$\mu^{(r)} = \left( u \frac{d}{du} \right)^r p(u) \bigg|_{u=1},$$

where it is understood that derivatives (if they are finite) are to be taken from the left, $p'(1) = p'(1^-)$, etc. In particular, the mean $\mu = \mu^{(1)}$ and variance $v = \mu^{(2)} - \mu^2 = E\{(X-\mu)^2\}$ are given by

$$\mu = p'(1), \qquad v = p''(1) + p'(1) - (p'(1))^2, \tag{9.2}$$

with the standard deviation

$$\sigma = \sqrt{v}.$$

The quantity

$$Q(s) = \log p(e^s)$$

is called the *cumulant generating function* and is related to the integral transforms introduced in later sections. The mean and standard deviation are then alternatively expressible as

$$\mu = \frac{d}{ds} Q(s) \bigg|_{s=0}, \qquad \sigma^2 = \frac{d^2}{ds^2} Q(s) \bigg|_{s=0},$$

a formula that is often useful in asymptotic studies as it is free of cancellations.

**Notation.** In order to express conveniently means and variances we introduce a general notation

$$\mathrm{Mean}(f) = \frac{f'(1)}{f(1)}, \qquad \mathrm{Var}(f) = \frac{f''(1)}{f(1)} + \frac{f'(1)}{f(1)} - \left(\frac{f'(1)}{f(1)}\right)^2,$$

for *any* function satisfying $f(1) \neq 0$, that is differentiable from the left at 1.

**Combinatorial parameters.** Throughout this chapter, we are given a combinatorial class $\mathcal{F}$, with generating function $F(z) = \sum_n f_n z^n$. Let $F_n = \mathrm{card}\{\mathcal{F}_n\}$. Thus, we have $f_n = F_n$ if $F$ is an OGF, $F_n = n! f_n$ if $F$ is an EGF. An integer valued parameter $\chi$ defined on $\mathcal{F}$ is to be analysed and estimates of the quantities

$$F_{n,k} = \mathrm{card}\{\omega \in \mathcal{F}_n \mid \chi[\omega] = k\}$$

are sought. In the language of probability theory, a discrete random variable $X_n$ is defined for each $n$ by

$$\Pr\{X_n = k\} = \frac{F_{n,k}}{F_n} = \frac{F_{n,k}}{\sum_k F_{n,k}}.$$

By its definition, $X_n$ represents the value of $\chi$ taken on a random object of $\mathcal{F}_n$, all such objects being taken with equal likelihood. The problem is to determine properties of the $X_n$, or equivalently of the doubly-indexed array $F_{n,k}$. For each $r$, the *moment* of order $r$ is defined by (9.1),

$$\mu_n^{(r)} \equiv E\{X_n^r\} = \sum_k \frac{F_{n,k}}{F_n} k^r, \tag{9.3}$$

from which the mean, variance, and standard deviation result.

Our main interest is in discrete random variables $X_n$ that correspond to an integer valued parameter $\chi$ of a combinatorial class $\mathcal{F}$. Let $F(z, u)$ be a bivariate generating function,

$$F(u, z) = \sum_{n,k} f_{n,k} u^k z^n.$$

either ordinary or exponential that is associated to a pair $(\mathcal{F}, \chi)$ of a combinatorial class and a parameter. Then, in the case of an OGF, we have $f_{n,k} = F_{n,k}$ with $F_{n,k}$ the number of $\omega \in \mathcal{F}_n$ such that $\chi[\omega] = k$, and in the case of an EGF, the counts and the coefficients are related by $F_{n,k} = n! f_{n,k}$. In both cases, we have

$$p_{n,k} := \Pr\{X_n = k\} = \frac{f_{n,k}}{f_n}, \qquad f_n = \sum_k f_{n,k}.$$

We define the quantities

$$f_n(u) = [z^n]F(z,u), \tag{9.4}$$

so that $f_n(1) = f_n$ is, up to a possible factor of $n!$, the number of objects in $\mathcal{F}_n$. Then, the probability generating function of $X_n$ is

$$p_n(u) := \frac{f_n(u)}{f_n(1)} = \frac{[z^n]F(z,u)}{[z^n]F(z,1)}. \tag{9.5}$$

From the general discussion of discrete random variables (9.1), the first two moments $\mu_n, \mu_n^{(2)}$ then result from coefficient extractions in derivatives. We summarize the computation in the following (easy!) theorem.

**Theorem 9.1 (Moments and BGF)** *Moments of a combinatorial parameter are expressed by coefficients in derivatives of the corresponding bivariate generating function $F(z,u)$, at $u=1$:*

$$\mu_n = \frac{[z^n]F_1(z)}{[z^n]F_0(z)}, \qquad \mu_n^{(2)} = \mu_n + \frac{[z^n]F_2(z)}{[z^n]F_0(z)}, \tag{9.6}$$

*where*

$$F_r(z) = \left.\frac{d^r}{du^r}F(z,u)\right|_{u=1}.$$

*The standard deviation is given by*

$$\sigma_n^2 = \mu_n^{(2)} - \mu_n^2.$$

Exercise 2. Let $g(z)$ be a nonlinear function that is analytic at the origin and has nonnegative coefficients there. For any positive $\alpha$ inside the disc of convergence of $g$, one has

$$\alpha^2 g''(\alpha)g(\alpha) + \alpha g'(\alpha)g(\alpha) - \alpha^2 g'(\alpha)^2 > 0.$$

[Hint. Consider the variable with PGF $g(\alpha u)/g(u)$.]

### 9.1.2 Moment inequalities

The purpose of this subsection is to demonstrate the usefulness of moment analysis as a first step in localizing "interesting" regions for the distribution of combinatorial parameters. For any random variable, the *moment inequalities* constrain the probability distribution, given the values of the mean and variance. They provide useful bounds on the distribution and in many cases even entail a property known as "concentration of distribution"; see Fig. 9.1, 9.2, 9.3.

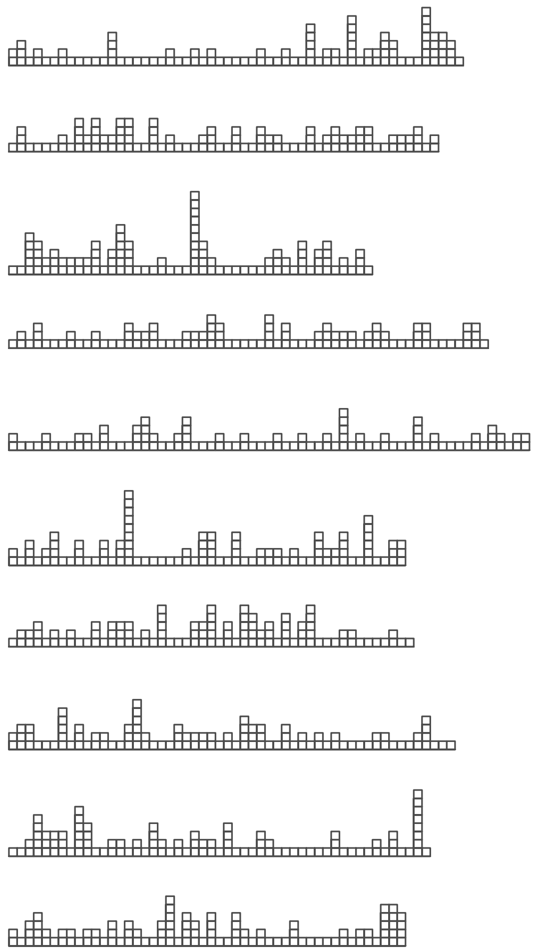

Figure 9.1: A sample of 10 integer compositions of size $n = 100$, drawn uniformly at random, illustrates the concentration property on the number of summands: a composition of size $n$ is likely to comprise a fairly large number —close to the mean, $\mu_n = n/2$— of summands that therefore each "tend" to be small. The corresponding BGF is $F(z,u) = (1-z)/(1-z(1+u))$.

**Lemma 9.1 (Moment inequalities)** *Let $Y$ be a random variable (not necessarily discrete) with mean $\mu$ and standard deviation $\sigma$.*

($i$) First moment inequality. *If $Y$ assumes only nonnegative values,*

$$\Pr\{Y \geq t\mu\} \leq \frac{1}{t}. \tag{9.7}$$

($ii$) Second moment inequality. *If $Y$ assumes arbitrary real values,*

$$\Pr\{|Y-\mu| \geq t\sigma\} \leq \frac{1}{t^2}. \tag{9.8}$$

The first relation is also known as Markov's inequality, the second one as Chebyshev's inequality; see [8, p. 283].

Proof. ($i$) Define a function $f(y)$ by $f(y) = 1$ if $y \geq t\mu$; then $\Pr\{Y \geq t\mu\} = E\{f(Y)\}$. Since $f(y) \leq y/(t\mu)$ for all $y$, we have

$$\Pr\{Y \geq t\mu\} = E\{f(Y)\} \leq E\{\frac{Y}{t\mu}\} = \frac{1}{t}.$$

Part ($ii$) follows from the first moment inequality applied to $(Y-\mu)^2$. □

The second moment inequality shows that the probability of values far from the mean (where distance is measured in the number of standard deviations) decays at least like $t^{-2}$.

Exercise 3. Assume that the "central absolute" moment of order $r$ of $Y$ exists, $\xi = E\{|Y-\mu|^r\}$. Then,

$$\Pr\{|Y-\mu| > t\xi^{1/r}\} \leq \frac{1}{t^r}.$$

**Definition 9.1** *A sequence of random variables $X_n$ with $\mu_n = E\{X_n\}$ has the* concentration property *if for any $\epsilon > 0$, one has*

$$\lim_{n\to\infty} \Pr\{1-\epsilon < \frac{X_n}{\mu_n} < 1+\epsilon\} = 1. \tag{9.9}$$

In the probabilistic literature, the concentration property is expressed by saying that $X_n/\mu_n$ converges to 1 *in probability.* We also say that the distribution of $\chi$ or the arrays $f_{n,k}$ and $f_{n,k}/f_n$ satisfy the concentration property.

**Theorem 9.2 (Concentration of distribution)** *Let $X_n$ be a sequence of random variables where $X_n$ has mean $\mu_n$ and standard deviation $\sigma_n$. Under the condition*

$$\lim_{n\to\infty} \frac{\sigma_n}{\mu_n} = 0,$$

*the $X_n$ satisfy the concentration property.*

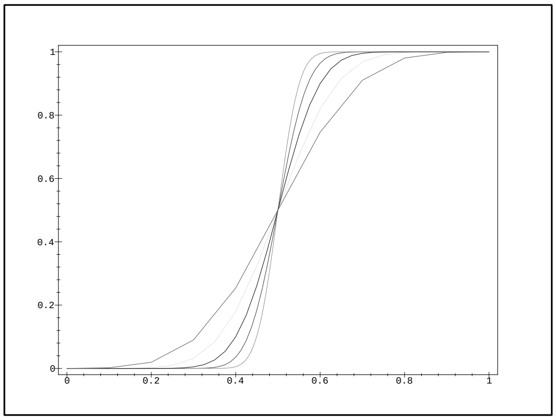


Figure 9.2: A rendering of the distribution function for the number of summands in random integer compositions of sizes $n = 10, 20, 40, 80, 160$. (The horizontal axis is scaled to $n$.) Concentration of distribution is manifest as higher values of $n$ lead to histograms that approach a step function. Here the mean is $\sim n/2$ and the standard deviation is $\mathcal{O}(n^{1/2})$. Quantitative estimates near the "threshold" of $n/2$ are the subject of later sections (9.4 and 9.5), where, after a suitable normalization, a Gaussian limit law is proved to hold.

PROOF. Apply the second moment inequality to the variable $X_n^* = X_n/\mu_n$ that has expectation equal to 1 and standard deviation $\sigma_n^* = \sigma_n/\mu_n = o(1)$. Then, one has

$$\Pr\{|X_n^* - 1| \geq \epsilon\} \leq \frac{(\sigma_n^*)^2}{\epsilon} = o(1),$$

for any fixed $\epsilon > 0$, as $n \to \infty$. □

### 9.1.3 Combinatorial applications

The multivariate version of the symbolic framework provides a very large number of bivariate generating functions. Moments are then accessible by differentiation and coefficient extraction, as summarized by Theorem 9.1. In addition, if computation reveals that the standard deviation $\sigma_n$ is of an asymptotic order smaller than $\mu_n$, then Theorem 9.2 applies and the distribution of $\chi$ is concentrated around its mean. Concentration of distribution is a frequently occurring phenomenon in analytic combinatorics, as several of the following examples demonstrate.

**Basic examples.** Previous chapters, especially Chapters 4–6 have demonstrated the usefulness of complex asymptotic methods in enumerating combinatorial structures. These basically univariate methods can be applied to generating functions of moments. For a given problem, it is often the case that the counting GF and the GF's associated to moments fall into the same broad category, namely

— *singularity analysis of the meromorphic type*, as discussed in Chapter 4;

— *singularity analysis of the algebraic–logarithmic type*, as discussed in Chapter 5;

— *saddle point analysis*, as discussed in Chapter 6.

Basic examples resorting to each of these analytic types are discussed below.

EXAMPLE 1. *Parts in unrestricted integer compositions.* Consider the number of parts in integer compositions. This corresponds to a sequence construction with bivariate OGF,

$$F(z,u) = \frac{1}{1-u\frac{z}{1-z}}.$$

The partial derivatives at $u = 1$ are rational functions,

$$F_0(z) = \frac{1-z}{1-2z}, \quad F_1(z) = \frac{z(1-z)}{(1-2z)^2}, \quad F_2(z) = \frac{2z^2(1-z)}{(1-2z)^3}.$$

Thus, we have $f_n = 2^{n-1}$, $\mu_n = (n+1)/2$ and $\sigma_n^2 = (n-1)/4$, for $n \geq 1$, so that $\sigma_n \sim \sqrt{n}/2$ and the distribution is concentrated near its mean. □

This example of integer compositions corresponds to a well-established analytic schema

$$F(z,u) = \frac{1}{1-ua(z)},$$

that translates sequences. The argument used here leads to concentration of distribution under general conditions on $a(z)$, called "supercriticality" (Section 9.3). Under these conditions, it is even the case that a limit Gaussian law holds for the number of components; see Section 9.5.

The next example is representative of an "exponential–logarithmic schema",

$$F(z,u) = \exp(ua(z)),$$

that translates sets. It applies to cycles in permutations, connected components in 2-regular graphs, factors of polynomials over finite fields, etc.

EXAMPLE 2. *Cycles in permutations.* From Sec. 3.2, the bivariate EGF of permutations with $\chi$ being the number of cycles is explicit,

$$F(z,u) = (1-z)^{-u}.$$

By the formulæ (9.6), one has

$$\mu_n = [z^n]\frac{1}{1-z}\log\frac{1}{1-z}, \qquad \mu_n^2 - \mu_n = [z^n]\frac{1}{1-z}\left(\log\frac{1}{1-z}\right)^2.$$

The coefficients are recovered directly by singularity analysis (Fig. 5.4),

$$\mu_n = \log n + \gamma + \mathcal{O}(\frac{1}{n}), \qquad \mu_n^{(2)} = \log^2 n + 2\gamma\log n + \gamma^2 - \frac{\pi^2}{6} + \mathcal{O}(\frac{\log n}{n}),$$

so that

$$\sigma_n^2 = \log n + \gamma - \frac{\pi^2}{6} + \mathcal{O}(\frac{\log n}{n}).$$

Thus, $\sigma_n \sim \sqrt{\log n}$. There is concentration of distribution for cycles in permutations and for the triangular array of Stirling numbers of the first kind. □

The last example of set partitions is a typical application of the saddle point method to

$$F(z,u) = \exp(ua(z)),$$

when $a(z)$ is either entire or fast growing at its singularity.

EXAMPLE 3. *Blocks in set partitions.* The bivariate EGF (Sec. 3.2) is

$$F(z,u) = \exp\left(u(e^z - 1)\right),$$

so that

$$F_1(z) = F_0(z)(e^z-1), \qquad F_2(z) = F_0(z)(e^z-1)^2,$$

where $F_0(z) = \exp(e^z - 1)$ is the EGF of the Bell numbers. The asymptotic analysis of the coefficients of $F_0$ is done by the saddle point method (Chapter 6). That method also applies to $F_1, F_2$, though the computations become intricate, as the variance involves cancellations. The end result is that the mean number of blocks is $\mu_n \sim n/(\log n)$, and $\sigma_n = n/(\log n)^2$. (Details of the computation may be found in Sachkov's book [52].) In particular, concentration of distribution holds for blocks in set partitions, that is to say for the triangular array of Stirling numbers of the second kind. □

**Trees and recursive structures.** The next examples deal with recursively specified combinatorial classes, especially trees. In this case, one has to cope with functional equations for GF's. It is only in the simplest situations that the bivariate GF admits a closed form. In general, one proceeds from the functional equation itself by successive differentiations, using simplifications that result at $u = 1$. A convenient path usually consists in expressing the corresponding GF's in terms of the basic univariate GF of counts, and use singularity analysis.

EXAMPLE 4. *Leaves in trees.* We discuss here the case of general Catalan trees and Cayley trees. For Catalan trees, *planarity* is reflected by a *sequence* construction, and the bivariate OGF $F(z, u)$ satisfies the functional equation,

$$F(z,u) = zu + \frac{zF(z,u)}{1 - F(z,u)}.$$

This reduces to a quadratic equation that admits an explicit solution, see Section 3.3,

$$F(z,u) = \frac{1}{2}\left(1 + (u-1)z - \sqrt{1 - 2(u+1)z + (u-1)^2 z^2}\right). \qquad (9.10)$$

The function $F_0(z)$ is the Catalan GF,

$$F_0(z) = \frac{1}{2}\left(1 - \sqrt{1-4z}\right),$$

and the number of trees of size $n$ is $\binom{2n-2}{n-1}/n$. In this approach, the partial derivatives are best computed with the help of a computer algebra system, and one finds

$$F_1(z) = \frac{1}{2}z + \frac{1}{2}\frac{z}{\sqrt{1-4z}}, \qquad F_2(z) = 2\frac{z^3}{(1-4z)^{3/2}}. \qquad (9.11)$$

There results that the number of leaves satisfies $\mu_n \sim n/4$; quadratic terms cancel in the computation of the variance and $\sigma_n = \mathcal{O}(n^{1/2})$ holds. Thus, once more, there is concentration of distribution.

Consider similarly Cayley trees. The functional equation for the bivariate EGF $F(z, u)$ of leaves in labelled nonplane trees is (Section 3.3)

$$F(z,u) - z(u - 1 + e^{F(z,u)}) = 0.$$

Here, *nonplanarity* is reflected by a *set* construction. The function $F_0(z)$ is the classical tree function that satisfies the functional equation $F_0 = ze^{F_0}$, and the number of Cayley trees is $n^{n-1}$. Differentiating the basic relation with respect to $u$ and simplifying at $u = 1$, yields

$$F_1 - z - zF_1e^{F_0} = 0, \qquad F_2 - ze^{F_0}(F_2 + F_1^2) = 0.$$

Thus, $F_1, F_2$ are expressible rationally in terms of $F_0$ by

$$F_1 = \frac{z}{1 - ze^{F_0}}, \qquad F_2 = \frac{zF_1^2e^{F_0}}{(1 - ze^{F_0})^3}. \tag{9.12}$$

Using the relation $F_0 = ze^{F_0}$, this further simplifies to

$$F_1 = \frac{z}{1 - F_0}, \qquad F_2 = \frac{z^2F_0}{(1 - F_0)^3}.$$

From singularity analysis,

$$F_0(z) = 1 - \sqrt{2}(1 - ez)^{1/2} + \mathcal{O}(1 - ez) \qquad (z \to e^{-1}),$$

there immediately result that $\mu_n \sim e^{-1}n$ and $\sigma_n = \mathcal{O}(\sqrt{n})$. Concentration holds like for general Catalan trees.

There are obvious similarities between (9.11) and 9.12), where the denominators are of respective degrees 1 and 3. Accordingly, the moments of orders 1, 2 are of asymptotic order $n$ and $n^2$. This is well explained by the general theory of simple families of trees. In this case, the univariate GF satisfies

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

where $\phi$ is the degree generator of the family. The bivariate GF satisfies

$$F(z, u) = z(u - 1) + z\phi(F(z, u)),$$

from which, upon differentiation,

$$F_1(z) = \frac{z}{1 - z\phi'(F(z))}, \qquad F_2(z) = \frac{z^3}{(1 - z\phi'(F(z)))^3},$$

which nicely accounts for the specific cases (9.11), (9.12). □

EXERCISE 4. Prove concentration of distribution for the number of leaves in binary Catalan trees. Generalize to simple families of trees corresponding to the equation $Y(z) = z\phi(Y(z))$.

EXERCISE 5. Analyse the distribution of leaves in nonplane *unlabelled* trees using suitable analytic properties of Pólya operators.

EXERCISE 6. Use the Lagrange inversion theorem in order to obtain explicit expressions for the distribution of the number of leaves in: ($i$) general Catalan trees, ($ii$) binary Catalan trees, ($iii$) Cayley trees. Evaluate in this way the factorial moments of all orders $E\{X_n(X_n - 1)\cdots(X_n - r + 1)\}$.

Discussions like the one relative to leaves can be carried out in full generality for simple families of trees. This is further illustrated by the next example relative to node types in such families.

EXAMPLE 5. *Node types in simple families of trees.* Consider a simply generated family of trees with generator $\phi(y)$, and examine the mean number of nodes of a fixed degree $r$ in a random tree of $n$ nodes. The BGF $F(z, u)$ satisfies the equation

$$Y = z\left(\phi(Y) + (u - 1)\phi_k Y^k\right).$$

The univariate GF of all trees is $F_0(z)$, the solution to $Y = z\phi(Y)$. Then, by differentiation with respect to $u$ at $u = 1$, we have

$$F_1(z) = \frac{z\phi_k F_0(z)^k}{1 - z\phi'(F_0(z))}.$$

On the other hand, we have by differentiation with respect to $z$, upon setting $u = 1$,

$$F_0'(z) = z\phi'(F_0(z))F_0'(z) + \frac{1}{z}F_0(z),$$

that is

$$\frac{1}{1 - z\phi'(F_0(z))} = \frac{zF_0'(z)}{F_0(z)},$$

and

$$F_1(z) = \frac{z^2 F_0'(z)}{F_0(z)} \cdot \phi_k (F_0(z))^k.$$

From the general analysis of Chapter 8, the function $F_0(z)$ tends to a limit $\tau$ as $z$ tends to its dominant singularity $\rho$. The quantity $\tau$ is defined by the "characteristic equation" $\phi(\tau) - \tau\phi'(\tau) = 0$, and one has $\rho = \tau/\phi(\tau)$. As a result, we have the singular equivalence,

$$F_1(z) \sim \frac{\phi_k \tau^k}{\phi(\tau)}\left(zF_0'(z)\right) \qquad (z \to \rho),$$

so that

$$\frac{[z^n]F_1(z)}{n[z^n]F_0(z)} \to \frac{\phi_k \tau^k}{\phi(\tau)}. \tag{9.13}$$

In other words, *the proportion of nodes of each type is directly determined by the shape of the generator $\phi$ and the singular constant $\tau$.* For instance, for Cayley trees ($\phi(y) = e^y$, $\tau = 1$) and general Catalan trees ($\phi(y) = (1-y)^{-1}$, $\tau = \frac{1}{2}$), we have that the asymptotic proportion of nodes of degree $k$ equals,

$$e^{-1}\frac{1}{k!}, \qquad 2^{-k-1} \tag{9.14}$$

respectively. Concentration of distributions also holds in all these cases by singularity analysis of the corresponding $F_2$.

The result expressed by Equation (9.13) is one about a collection of averages. It can also be viewed as a probability distribution over random nodes in random trees, where the underlying combinatorial set is now of cardinality $n \times [z^n]F(z)$. The corresponding PGF is then

$$\frac{\phi(\tau u)}{\phi(\tau)}.$$

This corresponds to a renormalization of the $\phi_k$ that, themselves, do not constitute in general a probability distribution. In the particular cases of (9.14), we obtain a Poisson or a geometric distribution in a random tree of the Cayley or Catalan family. (Such a property relates to the discussion of continuity theorems in Section 9.3.) □

Example 6. *Path length in general Catalan trees.* The problem corresponds to the functional equation (Section 3.4)

$$F(z,u) = \frac{z}{1 - F(uz,u)},$$

that does not admit of solutions in terms of simple elementary functions (see however the succinct discussion of Section 9.9). So, in such cases, the recourse is to differentiate the defining relation, solve, and pull out the derivatives $F_1, F_2$. One finds, with the usual simplifications at $u = 1$,

$$F_1 = \frac{z^2 F_0'}{(1-F_0)^2 - z},$$

where $F_0' = \frac{d}{dz}F_0(z)$. Thus,

$$F_1 = \frac{z}{2}\,\frac{1-\sqrt{1-4z}}{1-4z},$$

and the mean value satisfies

$$\mu_n \sim \frac{1}{2}\sqrt{\pi n^3}.$$

Cross derivatives proliferate upon successive differentiations since the functional equation involves $F(zu, u)$. The final result is that the standard deviation is $\mathcal{O}(n)$, so that concentration of distribution strikes again! □

EXERCISE 7. Analyse the mean and standard deviation of path length in binary Catalan trees and Cayley trees, and show that they are $\mathcal{O}(n^{3/2})$ and $\mathcal{O}(n)$ respectively. Generalize to simple families of trees.

EXAMPLE 7. *The Quicksort algorithm.* This is perhaps the most famous of all sorting algorithms. The analysis of the number of comparisons used is essentially equivalent to the analysis of path length in binary search trees. The latter problem is expressed by a bivariate generating function $F(z, u)$, where $[z^n u^k]F(z, u)$ is the probability that path length is equal to $k$ in a tree of size $n$ and $F(z, u)$ is defined by the functional equation [54],

$$\frac{\partial}{\partial z}F(z, u) = F(zu, u)^2, \qquad F(0, 0) = 1.$$

Moments are obtained by successively differentiating and solving at each stage a linear differential equation; for instance, initially, we have

$$\frac{d}{dz}F_0(z) = F_0(z)^2, \qquad F_0(0) = 1,$$

so that $F_0(z) = (1 - z)^{-1}$. The mean cost is found to be $2n \log n + \mathcal{O}(n)$ while the variance is $\mathcal{O}(n^2)$. Thus, concentration of distribution holds for Quicksort as well as for path length in binary search trees. The paper of McDiarmid and Hayward [48] provides precise quantitative information on large deviations from the mean. □

**Schemas.** Bivariate generating functions, once differentiated and instantiated at $u = 1$, yield univariate GFs. Coefficients can then be recovered using the complex analytic techniques of Chapters 4–6, namely: polar analysis of meromorphic functions, singularity analysis, or saddle point analysis of coefficients.

Several of the examples that we have just examined are typical of a general process. For instance, summands in compositions belong to a general combinatorial schema

$$F(z, u) = \frac{1}{1 - ua(z)},$$

that corresponds to the *sequence construction*. As seen in Section 4.8, the coefficient analysis of meromorphic functions is guaranteed to apply if $a(z)$ attains the value 1 at a point $\rho$ on the positive real line, strictly inside its disc of convergence. The schema is called supercritical (see also the discussion in the Section 9.3). In accordance with Section 4.8, the singular expansion of $F_0$ then reads

$$F_0(z) = \frac{1}{a'(\rho)(\rho - z)} + \mathcal{O}(1).$$

Since one has

$$F_1(z) = \frac{a(z)}{(1-a(z))^2}, \quad F_2(z) = \frac{2a(z)^2}{(1-a(z))^3},$$

the same reasoning applies to $F_1, F_2$, for which one finds mechanically,

$$F_1(z) = \frac{1}{a_1^2}(\rho - z)^{-2} + \frac{a_2 - a_1^2}{a_1^3}(\rho - z)^{-1} + \mathcal{O}(1),$$

$$F_2(z) = \frac{2}{a_1^3}(\rho - z)^{-3} + \frac{3a_2 - a_1^2}{a_1^4}(\rho - z)^{-2} + \mathcal{O}((\rho - z)^{-1}),$$

with $a_1 = a'(\rho), a_2 = a''(\rho)$, etc. Translating these singular expansions, one finds that cancellation occurs, so that $\mu_n = \mathcal{O}(n)$ and $\sigma_n = \mathcal{O}(n^{1/2})$.

**Proposition 9.1 (Concentration in supercritical sequences)** *In the bivariate sequence scheme,*

$$F(z, u) = \frac{1}{1 - ua(z)},$$

*assume that there exists $\rho > 0$ such that $a(z)$ has nonnegative coefficients is analytic in $|z| \leq \rho$, and $a(\rho) = 1$. Assume also that $a(0) = 0$ and $a(z)$ is* aperiodic[1] *Then, the moments of the number of components in a random sequence of size $n$ satisfy*

$$\mu_n = \frac{n}{\rho a'(\rho)} + \mathcal{O}(1), \qquad \sigma_n^2 = n \cdot \frac{\rho a''(\rho) + a'(\rho) - \rho a'(\rho)^2}{\rho^2 a'(\rho)^3} + \mathcal{O}(1).$$

*In particular, the number of components satisfies the concentration property.*

[1] *Aperiodicity of $a(z)$ means that for no $d \geq 2$ is $a(z^d)$ analytic at 0.*

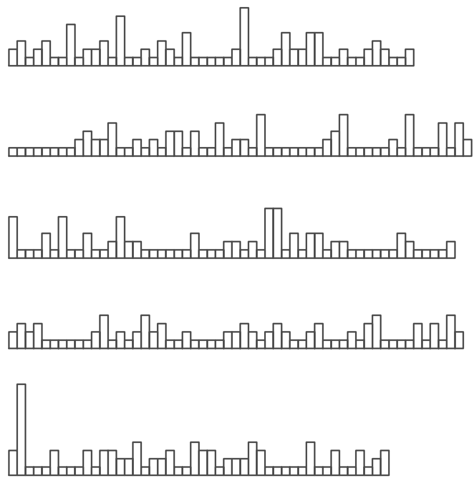

Figure 9.3: A sample of 5 surjections of size $n = 100$ (only the occupancy profile is shown) drawn uniformly at random reveals characteristics qualitatively similar to those of integer compositions as both combinatorial types resort to the supercritical sequence schema. Here the mean cardinality is $\sim n/(2\log 2) \doteq 0.72134\, n$ and the standard deviation is $\mathcal{O}(n^{1/2})$.

PROOF. See the discussion above. The aperiodicity condition is sufficient to ensure the existence of a unique dominant singularity, $\rho$, which is a simple pole since $a'(\rho) \neq 0$. □

We shall see later in Section 9.5 that a stronger Gaussian limit property holds for the distribution.

EXAMPLE 8. *Image cardinality in surjections.* The bivariate EGF is

$$F(z,u) = \frac{1}{1 - u(e^z - 1)},$$

so that the preceding scheme applies with a dominant singularity at $\rho = \log 2$. The mean and the standard deviation are

$$\mu_n \sim \frac{n}{2\log 2}, \qquad \sigma_n^2 \sim n\,\frac{1-\log 2}{4(\log 2)^2}.$$

□

Proposition 9.1, together with its companion results to be discussed later in this chapter, holds in a very general context, namely the subcriticality condition in sequence constructions. It may even be applied in circumstances when only partial information is available on generating functions entering a construction. The analysis of compositions with summands that are prime or even twin primes (!) illustrates the point.

Let

$$P(z) = z^2 + z^3 + z^5 + z^7 + z^{11} + z^{13} + \cdots$$

be the characteristic GF of numbers that are prime. Then

$$F(z,u) = \frac{1}{1 - uP(z)}$$

is the BGF of integer compositions in summands that are constrained to be prime, with $u$ marking the number of summands. Though many properties of the distribution of primes remain unknown, any rough version of the prime number theorem —the number of primes till $n$ is asymptotic to $n/\log n$— entails that $P(z)$ is analytic in $|z| < 1$. Let $p_j$ be the $j$th prime; since one has, for $0 < x < 1$ and any $m$,

$$\sum_{j=1}^{m} z^{p_j} < P(z) < \sum_{j=1}^{m} z^{p_j} + \frac{z^{p_m}}{1-z},$$

it is not hard to determine *rigorously* a bounding interval of any prescribed small width for the root of the equation $P(\rho) = 1$. Also, $P(z)$ clearly satisfies an aperiodicity condition, so that by a slightly amended form of Prop. 9.1, there is concentration of distribution. The quantity $\rho$ can be determined to great accuracy, for instance,

$$\rho = 0.67740177613066042797630631643196719199252141284195 \pm 10^{-50},$$

is already guaranteed by taking $m = 80$. Accordingly, the mean number of summands in a random composition of size $n$ is known to great accuracy and well approximated by

$$K \cdot n, \qquad K = 0.303655\cdots.$$

Perhaps even more surprising is the fact that a similar analysis can be carried out for twin prime (primes $p$ such that $p-2$ or $p+2$ is prime) for which it is still not proved that they form an infinite set. It suffices to replace $P(z)$ by

$$T(z) = z^3 + z^5 + z^7 + z^{11} + z^{13} + z^{17} + z^{19} + z^{29} + z^{31} + \cdots,$$

and one determines by the same device the characteristic root of $T(z) = 1$,

$$\rho_T = 0.7704180066328016185503714 \pm 10^{-25},$$

by considering all twin primes till 229. Similar analyses could be performed for Fermat primes or Mersenne primes. In all such cases, concentration of distribution holds and the involved constants are easily determined, even to several thousands of digits of accuracy.

This discussion illustrates the fact that operators that translate combinatorial constructions are in many cases "analyticity improving", and only partial information of intervening GF's often suffices to obtain very precise information on combinatorial counts and probability distributions. We do not enter a categorization of schemas and concentration properties at this stage, since we shall see shortly much stronger conclusions —Gaussian limit laws, large deviations— that can be drawn under similar sets of conditions. The current discussion is only meant to illustrate of the kind of generality that can be attained and the power of symbolic methods in conjunction with complex asymptotics.

EXERCISE 8. Show that similar conclusions hold for the *labelled cycle* schema,

$$F(z,u) = \log \frac{1}{1 - ua(z)}.$$

EXERCISE 9. Concentration properties hold for the bivariate *labelled set* schema,

$$F(z,u) = \exp(ua(z)),$$

when $a(z)$ has a unique dominant singularity of the logarithmic type

$$a(z) = c_0 \log(1 - z/\rho) + c_1 + o\left((\log(1 - z/\rho))^{-1}\right).$$

Concentration properties also hold for the *unlabelled set* schema, when the generator $a(z)$ is logarithmic and $\rho < 1$.

EXERCISE 10. Discuss concentration properties for the *labelled set* schema,

$$F(z,u) = \exp(ua(z)),$$

when $a(z)$ is admissible in the sense of saddle point analysis.

## 9.2 Limit laws and combinatorial distributions

The main theme of this chapter is well illustrated by two parameters of random binary words for which explicit counts are available. We thus let

$\mathcal{W} = \{a, b\}^\star$ denote the class of binary words built over the alphabet $\{a, b\}$, with the size of a word being its length. Obviously, there are $W_n = 2^n$ such words of size $n$, and the OGF is

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

The two parameters considered are the following: $\alpha[w]$ is the number of initial $a$'s in the word $w$; $\beta[w]$ is the total number of $a$'s in $w$. The corresponding probability distributions,

$$q_{n,k} = \Pr\{\alpha[w] = k \;\big|\; w \in \mathcal{W}_n\}, \qquad r_{n,k} = \Pr\{\beta[w] = k \;\big|\; w \in \mathcal{W}_n\},$$

are immediately determined:

$$q_{n,k} = \frac{1}{2^{k+1}}\ (0 \le k < n), \quad q_{n,n} = \frac{1}{2^n}; \qquad\qquad r_{n,k} = \frac{1}{2^n}\binom{n}{k}.$$

Clearly, the distribution of the number of initial $a$'s (parameter $\alpha$) converges to a geometric law of parameter $1/2$,

$$\lim_{n\to\infty} q_{n,k} = q_k, \qquad q_k = \frac{1}{2^{k+1}}, \tag{9.15}$$

uniformly, with a "speed of convergence" that is $\mathcal{O}(2^{-n})$. We thus have a case of direct convergence of the probability distribution of a combinatorial parameter to a *discrete limit law*. Here the mean and the variance of the combinatorial distribution are both asymptotically constant.

The total number of $a$'s (parameter $\beta$) obeys a binomial law and a simple computation based on Stirling's formula (see, *e.g.*, [54, p. 196]) yields the approximation,

$$\lim_{\nu\to\infty} \frac{\binom{2\nu}{\nu+h}}{\binom{2\nu}{\nu}} = e^{-h^2/\nu}, \tag{9.16}$$

taking for simplicity $n = 2\nu$ to be even. This approximation holds for the "central range" $h = o(\nu^{3/4})$ uniformly, with error terms that are $O(\nu^{-1} + h^4\nu^{-3})$. In this case, in the asymptotic limit, the combinatorial distribution is approximated by a continuous function of the Gaussian type,

$$\xi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

with

$$r_{n,k} \sim \frac{1}{\sqrt{n/4}}\xi(x), \qquad x = \frac{k - n/2}{\sqrt{n/4}}. \tag{9.17}$$

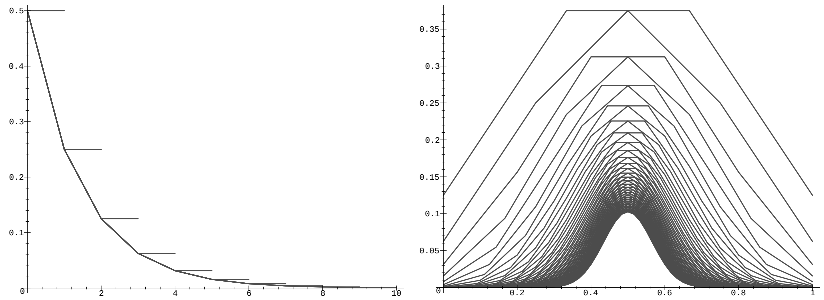


Figure 9.4: Histograms of probability distributions for the number of initial $a$'s in a random binary string (parameter $\alpha$, left) and the total number of $a$'s (parameter $\beta$, right). The histogram corresponding to $\alpha$ is not normalized and direct convergence to a discrete geometric law is apparent; for $\beta$, the horizontal axis is scaled to $n$, and the histograms quickly conform to the bell-shaped curve that is characteristic of the continuous gaussian limit.

In other words, the pattern is revealed after a normalization that consists in *centering* by the mean $n/2$ and *scaling* by the standard deviation $\sqrt{n/4}$. Figure 9.4 displays the histograms of the distributions of $\alpha$ and $\beta$. with convergence to the discrete geometric law for $\alpha$ (there, curves coincide till the last value where they branch off) or to the continuous Gaussian law (where the bell-shaped profile is apparent in the limit).

The different behaviours exhibited by $\alpha$ and $\beta$ are reflected by different behaviours of the moments. In the first case, the mean and standard deviation are both $\mathcal{O}(1)$ and there is a discrete limit law; in the second case, the mean and standard deviation both tend to infinity, each individual value is $o(1)$ and given asymptotically in terms of a continuous function. Although several combinations (asymptotic orders of moments and types of limit distributions) are possible in principle, these two are representative of what is observed most of the time in elementary combinatorics.

It is of interest to try and correlate these two different behaviours with analytic properties of bivariate generating functions (BGF's) as introduced in Chapter 3. We find easily, by the techniques of that chapter,

$$G(z,u) = \frac{1-z}{1-uz} \cdot \frac{1}{1-2z}, \qquad H(z,u) = \frac{1}{1-z(1+u)}.$$

Let $g_n(u) = [z^n]G(z,u)$ and $h_n(u) = [z^n]H(z,u)$. The probability generating functions (PGF's) of $\alpha, \beta$ over $\mathcal{W}_n$ are then

$$q_n(u) = \frac{g_n(u)}{g_n(1)}, \qquad r_n(u) = \frac{h_n(u)}{h_n(1)},$$

with $g_n(1) = h_n(1) = 2^n$. Now, for any fixed value of $u$, singularity analysis of meromorphic functions applies, and we have, for instance when $|u| < \frac{3}{2}$,

$$q_n(u) = \frac{\frac{1}{2}}{1 - \frac{u}{2}} + \mathcal{O}((\frac{2}{3})^n), \qquad r_n(u) = \left(\frac{1+u}{2}\right)^n.$$

In the first case, there is pointwise convergence of $q_n(u)$ to

$$q(u) = \frac{\frac{1}{2}}{1 - \frac{u}{2}} = \sum_{k \geq 0} \frac{u^k}{2^{k+1}},$$

which is precisely the PGF of a geometric variable with parameter $1/2$ found earlier to be asymptotic limit of the distribution of $\alpha$. This is no coincidence: *continuity theorems for PGF's* to be discussed in Section 9.3 show that convergence to a discrete limit law may be inferred from convergence of the corresponding PGF's.

In the second case, there is no direct convergence of the PGF $r_n(u)$ to a fixed limit, but convergence can be restored after a suitable *normalization*. Technically, this requires introducing Fourier or Laplace transforms of distributions as we are dealing with continuous laws in the limit and integral forms replace discrete sums expressing PGF's. However, the two cases are quite similar, and *continuity theorems for integral transforms* to be discussed in Section 9.4 show that convergence to a continuous limit law may be inferred from convergence of the corresponding integral transforms.

In the perspective of analytic combinatorics, the univariate counting problem corresponds to a common univariate GF

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

with a pole at $z = 1/2$. From the BGF point of view, the two cases have clear analytical counterparts, and the auxiliary parameter $u$ intervenes by modifying the singular behaviour of $G$ and $H$. In the case of $G$, there are two components in the BGF

$$G(z,u) = \frac{1}{1-uz} \cdot \boxed{\frac{1-z}{1-2z}},$$

and, in essence, all the dominant singular part —a simple pole at $z = 1/2$— is concentrated in the second component that does not change when $u$ varies. Thus, for instance with $|u| < 3/2$, one has

$$G(z,u) \underset{z\to 1}{\sim} \frac{1}{1-u/2} G(z,1),$$

the coefficients admit the estimate,

$$[z^n]G(z,u) \sim \frac{\frac{1}{2}}{1-\frac{u}{2}}\, 2^n,$$

hence a discrete limit law by the continuity theorem for PGF's of Section 9.3. In the second case, the auxiliary parameter modifies the location of the singularity,

$$H(z,u) = \frac{1}{1 - z\,\boxed{(1+u)}}.$$

Then, the singular behaviour is strongly dependent upon a singularity at

$$\rho(u) = \frac{1}{(1+u)}$$

that gets *displaced* when $u$ varies. Accordingly, the coefficients obey a "power law", here of an exact type,

$$[z^n]H(z,u) = \rho(u)^{-n}.$$

This analytical form is reminiscent of the central limit theorem of probability theory after which large powers —corresponding to sums of a large number of independent random variables— entail convergence to a Gaussian law. By continuity theorems for integral transforms (the analytic foundation of the central limit theorem) there results a continuous limit law of the Gaussian type in this case.

The preceding discussion is indicative of the overall structure of this chapter. We consider BGF's and examine the way the auxiliary variable $u$ induces a perturbation of the singular behaviour of the univariate GF. A minor perturbation mode is likely to lead to a discrete limit law; a major perturbation mode, *e.g.*, a "movable" singularity, is likely to lead to a continuous limit law via large power approximations like $\rho(u)^{-n}$. Globally, this approach to limit laws is called *singularity perturbation* analysis, a term introduced in [23]. The following sections provide precise validity conditions: discrete laws are treated first in Section 9.3, while continuous laws, with special emphasis on the ubiquitous Gaussian law, form the subject of Section 9.4 and of subsequent sections.

## 9.3 Discrete limit laws

Moment methods give useful qualitative information on global characteristics of a distribution, for instance where it tends to be concentrated. However, much more is usually true in decomposable structures: almost any "natural" parameter is likely to obey a common *limit law*. In simpler cases, a limit law is directly visible on the array of numbers $f_{n,k}/f_n$, if these quantities converge for each fixed $k$, as $n \to \infty$. The example of the length of initial runs in random words of Section 9.2 is typical.

**Definition 9.2** *If $X_n$ is a sequence of random variables such that there exist numbers $p_k$, with $\sum_k p_k = 1$, and*

$$p_k = \lim_{n\to\infty} \Pr\{X_n = k\},$$

*then $X_n$ is said to satisfy a* discrete limit law*.*

*Alternatively, one says that the $X_n$* converge in distribution *to a variable with the discrete distribution $\{p_k\}_{k=0}^{\infty}$.*

We also say that a combinatorial parameter $\chi$ that underlies $X_n$, the corresponding bivariate generating function $F(z,u)$, or the array of coefficients $f_{n,k} = [z^n u^k]F(z,u)$, obey a *discrete limit law*.

Convergence to a discrete limit law is always *uniform*, in the sense that

$$\epsilon_n := \sup_k |\Pr\{X_n = k\} - p_k|$$

tends to 0 as $n \to \infty$. The quantity $\epsilon_n$ is called the *speed of convergence* to the limit law. Uniform convergence is proved as follows. Fix a small $\epsilon > 0$. Then, there exists a $k_0$ such that $\sum_{k\geq k_0} p_k \leq \epsilon$, so that $\sum_{k<k_0} p_k > 1-\epsilon$. Now, by simple convergence, there exists an $n_0$ such that, for all $n$ larger than $n_0$ and $k < k_0$,

$$|p_{n,k} - p_k| < \frac{\epsilon}{k_0},$$

with $p_{n,k} = \Pr\{X_n = k\}$. Thus, we have

$$\sum_{k<k_0} p_{n,k} > 1-2\epsilon, \qquad \sum_{k\geq k_0} p_{n,k} \leq 2\epsilon.$$

In other words, $\sum_{k\geq k_0} p_k$ and $\sum_{k\geq k_0} p_{n,k}$ are both in $[0, 2\epsilon]$, while each difference $|p_{n,k} - p_k|$ for $k < k_0$ is at most $\epsilon/k_0$. This determines the $\mathcal{L}_1$–distance between the two distributions,

$$\sum_k |p_{n,k} - p_n| < 5\epsilon,$$

by the triangular inequality. In particular, the quantity $\sup_k |p_{n,k} - p_k|$ is at most $5\epsilon$.

EXERCISE 11. The sequence $X_n$, where

$$\Pr\{X_n = 0\} = \frac{1}{3}, \quad \Pr\{X_n = 1\} = \frac{1}{3}, \quad \Pr\{X_n = n\} = \frac{1}{3},$$

does not satisfy a discrete limit law in the sense above, although $\lim_k \Pr\{X_n = k\}$ exists for each $k$.

A highly plausible indication of the occurrence of a discrete law is the fact that $\mu_n = \mathcal{O}(1)$, $\sigma_n = \mathcal{O}(1)$. Examination of initial entries in the table of values will then help decide whether a limit law is likely to hold or not.

**Singleton cycles in permutations.** Consider for instance the number of singleton cycles in a random permutation. The bivariate EGF is

$$F(z, u) = \frac{\exp(z(u-1))}{1-z},$$

so that,

$$F_0(z) = \frac{1}{1-z}, \quad F_1(z) = \frac{z}{1-z}, \quad F_2(z) = \frac{z^2}{1-z}.$$

Thus, for $n \geq 2$, we have $\mu_n = \sigma_n = 1$. The distribution is *not* concentrated, and, in random experiments, we expect to observe a number of singleton cycles that, most of the time, takes only a few small integer values. while no single value predominates.

The table of numerical values of the probabilities $f_{n,k} = [z^n u^k]F(z, u)$ immediately tells what goes on. Here is a small sample.

| | $k = 0$ | $k = 1$ | $k = 2$ | $k = 3$ | $k = 4$ | $k = 5$ |
|---|---|---|---|---|---|---|
| $n = 4$ | 0.375 | 0.333 | 0.250 | 0.000 | 0.041 | |
| $n = 5$ | 0.366 | 0.375 | 0.166 | 0.083 | 0.000 | 0.008 |
| $n = 10$ | 0.367 | 0.367 | 0.183 | 0.061 | 0.015 | 0.003 |
| $n = 20$ | 0.367 | 0.367 | 0.183 | 0.061 | 0.015 | 0.003 |

The exact distribution is easily extracted from the bivariate GF,

$$f_{n,k} := [z^n u^k]F(z, u) = \frac{1}{k!}[z^{n-k}]\frac{e^{-z}}{1-z} = \frac{d_{n-k}}{k!},$$

where

$$d_n = [z^n]\frac{e^{-z}}{1-z} = \sum_{j=0}^{n} \frac{(-1)^j}{j!}$$

is such that $n!d_n$ is the number of derangements of size $n$. Asymptotically, one has $d_n \sim e^{-1}$. Thus, for fixed $k$, we have

$$\lim_{n\to\infty} f_{n,k} = p_k, \qquad p_k = \frac{e^{-1}}{k!}.$$

A distribution of the form

$$p_k = e^{-\lambda}\frac{\lambda^k}{k!}$$

is called a Poisson law of rate $\lambda$. Thus, the distribution of singleton cycles in a random permutation of large size tends to a Poisson law of rate 1.

As expected from the quality of approximations in meromorphic analyses, convergence is quite fast. Here is a table of differences,

$$\delta_{n,k} = f_{n,k} - \frac{e^{-1}}{k!}.$$

| | $k=0$ | $k=1$ | $k=2$ | $k=3$ | $k=4$ | $k=5$ |
|---|---|---|---|---|---|---|
| $n=10$ | $2.3\,10^{-8}$ | $-2.5\,10^{-7}$ | $1.2\,10^{-6}$ | $-3.7\,10^{-6}$ | $7.3\,10^{-6}$ | $1.0\,10^{-5}$ |
| $n=20$ | $1.8\,10^{-20}$ | $-3.9\,10^{-19}$ | $3.9\,10^{-18}$ | $-2.4\,10^{-17}$ | $1.1\,10^{-16}$ | $-3.7\,10^{-16}$ |

The speed of convergence is easily bounded. One has $d_n = e^{-1} + \mathcal{O}(1/n!)$, by the alternating series property, so that

$$f_{n,k} = \frac{e^{-1}}{k!} + O\left(\frac{1}{k!\,(n-k)!}\right) = \frac{e^{-1}}{k!} + O\left(\frac{1}{n!}\binom{n}{k}\right) = \frac{e^{-1}}{k!} + O\left(\frac{2^n}{n!}\right).$$

Accordingly, the numerical evaluation of $\Delta_n = \max_k |\delta_{n,k}|$ yields

$$\Delta_{10} = 1.0\,10^{-5}, \quad \Delta_{20} = 6.3\,10^{-15}, \quad \Delta_{50} = 1.5\,10^{-52}.$$

The upper bound $2^n/n!$ on the error evaluates to $3.7\,10^{-50}$ for $n = 50$, so that this analysis is indeed quite tight.

### 9.3.1 Continuity theorem for PGFs

A higher level approach to discrete limit laws in analytic combinatorics is based on asymptotic estimates of $p_n(u)$, the PGF of the random variable $X_n$. If, for sufficiently many values of $u$, one has

$$p_n(u) \to p(u) \qquad (n \to +\infty),$$

one can infer that the coefficients $p_{n,k} = [u^k]p_n(u)$ (for any fixed $k$) tend to the limit $p_k$ with generating function $p(u)$. A *continuity theorem* for characteristic functions describes precisely sets of conditions under which convergence of probability generating functions to a limit entails convergence of coefficients to a limit, that is to say the occurrence of a discrete limit law. We state here a continuity theorem with very general analytic conditions.

**Theorem 9.3 (Continuity of PGFs)** *Let $\Omega$ be an arbitrary set with at least one accumulation point in the interior of the unit disc. Assume that the PGFs $p_n(u) = \sum_{k\geq 0} p_{n,k}u^k$ and $p(u) = \sum_{k\geq 0} p_k u^k$ are such that there is convergence,*

$$\lim_{n\to+\infty} p_n(u) = p(u),$$

*pointwise for each $u$ in $\Omega$. Then a discrete limit law holds in the sense that, for each $k$,*

$$\lim_{n\to+\infty} p_{n,k} = p_k.$$

*If in addition, for some $r > 0$, the PGF's $p_n(u)$ and $p(u)$ are analytic in $|u| < r$ and continuous on $|u| = r$, with $p_n(u) \to p(u)$ on $|u| = r$, then the speed of convergence to the limit satisfies*

$$|p_{n,k} - p_k| \leq r^{-k} \sup_{|u|=r} |p_n(u) - p(u)|.$$

Proof. The $p_n(u)$ are *a priori* analytic in $|u| < 1$ and uniformly bounded by 1 in modulus, in $|u| \leq 1$. By Vitali's convergence theorem (see [60, p. 168] or [35]), there is convergence of $p_n(u)$ to $p(u)$ and this convergence is uniform in any subdisc of the unit disc, for instance, $|u| \leq \frac{1}{2}$. Then, by Cauchy's coefficient formula, we have

$$\begin{aligned} p_k &= \frac{1}{2i\pi}\int_{|u|=1/2} p(u)\,\frac{du}{u^{k+1}} \\ &= \lim_{n\to\infty} \frac{1}{2i\pi}\int_{|u|=1/2} p_n(u)\,\frac{du}{u^{k+1}} \\ &= \lim_{n\to\infty} p_{n,k}. \end{aligned} \tag{9.18}$$

Uniformity granted by Vitali's theorem is essential in justifying the second line of (9.18).

The second part of the theorem follows similarly from Cauchy's coefficient formula upon integrating along the circle $|u| = r$,

$$p_{n,k} - p_n = \frac{1}{2i\pi}\int_{|u|=r} (p_n(u) - p(u))\,\frac{du}{u^{k+1}},$$

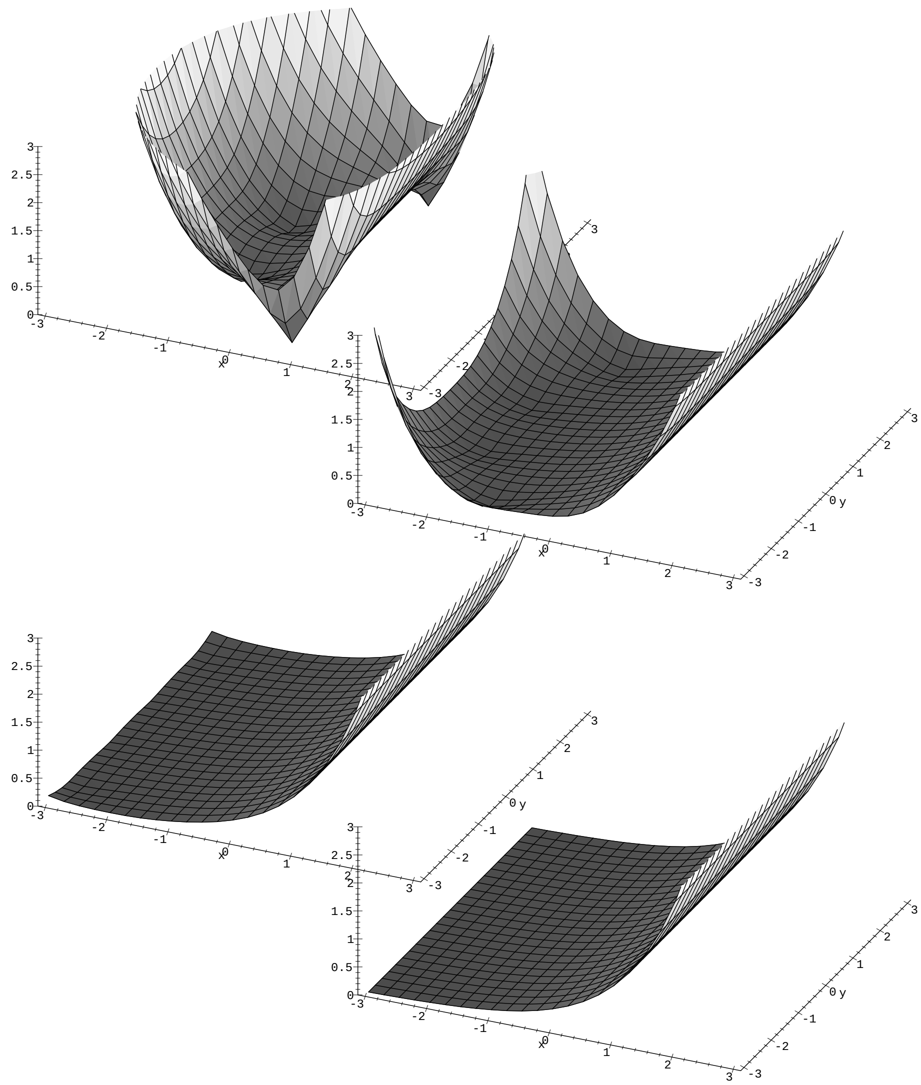

Figure 9.5: The PGFs of singleton cycles in random permutations of size $n = 4, 8, 12$ (left to right and top to bottom) illustrate convergence to the limit PGF of the Poisson(1) distribution (bottom right). Here the modulus of each PGF for $|\Re(u)|, |\Im(u)| \leq 3$ is displayed.

and trivial bounds. □

Feller gives the sufficient set of conditions: $p_n(u) \to p(u)$ pointwise for all real $u \in ]0,1[$; see [18, p. 280] for a proof that only involves elementary real analysis. It is perhaps surprising that very different sets can be taken, for instance,

$$\Omega = \left[-\frac{1}{3}, -\frac{1}{2}\right], \qquad \Omega = \{\frac{1}{n}\}, \qquad \Omega = \left\{\frac{\sqrt{-1}}{2} + \frac{1}{2^n}\right\}.$$

Equipped with this theorem, it is possible to revisit the analysis of singleton cycles in permutations. By singularity analysis of meromorphic functions,

$$p_n(u) \to e^{u-1},$$

pointwise for each $u$, hence convergence to the Poisson limit. The PGF's then satisfy

$$|p_n(u) - p(u)| = \mathcal{O}(R^{-n})$$

for any fixed $R > 1$ when $u$ stays in a bounded region of the complex plane. Hence, the PGF's $p_n(u)$ converge exponentially fast to their limit and the same property holds for the corresponding probability distributions. Eventually, *the method is based on two successive inversions by means of Cauchy's coefficient formula.* The first inversion is achieved via singularity analysis, the second one via the proof method encapsulated in Theorem 9.3.

Exactly the same method applies to the number of $m$–cycles of permutations for which the exponential BGF is

$$F(z,u) = \frac{e^{(u-1)z^m/m}}{1-z}.$$

Then,

$$\lim_{n\to\infty} [z^n]F(z,u) = e^{(u-1)/m},$$

which is the PGF of a Poisson law of rate $\lambda = \frac{1}{m}$. Thus, the number of $m$–cycles in a random permutation of large size obeys in the limit a Poisson law of rate $1/m$.

### 9.3.2 Combinatorial applications

**Combinatorial compositions.** The bivariate *composition schema*,

$$F(z,u) = g(uh(z)),$$

illustrates a situation in analytic combinatorics leading to discrete laws. This schema expresses over generating functions the combinatorial operation $\Gamma[\mathcal{H}]$ of *substitution* of components $\mathcal{H}$ enumerated by $h(z)$ inside "templates" $\Gamma$ enumerated by $g(z)$.

We assume globally that $g$ and $h$ have nonnegative coefficients and that $h(0) = 0$ so that the composition $g(h(z))$ is well-defined formally. We let $\rho_g$ and $\rho_h$ denote the radii of convergence of $f$ and $g$, and define

$$\tau_g = \lim_{x\to\rho_g^-} g(x) \qquad \text{and} \qquad \tau_h = \lim_{x\to\rho_h^-} h(x), \tag{9.19}$$

the (possibly infinite) limits exist because of nonnegativity of coefficients. Three cases are to be distinguished.

**Definition 9.3** *The composition schema $g(uh(z))$ is said to be*

(*i*) subcritical *if* $\tau_h < \rho_g$,

(*ii*) critical *if* $\tau_h = \rho_g$,

(*iii*) supercritical *if* $\tau_h > \rho_g$.

The analytic intuition is simple: the behaviour of $g(h(z))$ at its dominant singularity is dictated by the dominant singularity of $g$ (subcritical case), or by the dominant singularity of $f$ (supercritical case), or it should involve a mixture of the two (critical case). An instance of the supercritical schema has been discussed in Section 9.1. We examine here the subcritical case, and first state a general lemma about subcritical compositions.

**Lemma 9.2 (Subcritical composition)** *Consider the bivariate scheme*

$$F(z, u) = g(uh(z)).$$

*Assume that $g(z)$ and $h(z)$ satisfy the* subcriticality *condition $\tau_h < \rho_g$, and that $h(z)$ has a unique singularity at $\rho = \rho_h$ on its disc of convergence that is of the algebraic–logarithmic type*

$$h(z) = \tau - c(1 - \frac{z}{\rho})^\lambda + o((1 - \frac{z}{\rho}))^\lambda,$$

*where $\tau = \tau_h$, $c \in R^+$, $0 < \lambda < 1$. Then, a discrete limit law holds,*

$$\lim_{n\to\infty} \frac{f_{n,k}}{f_n} = p_k, \qquad p_k = \frac{k g_k \tau^{k-1}}{g'(\tau)},$$

*with probability generating function*

$$p(u) = \frac{ug'(\tau u)}{g'(\tau)}.$$

PROOF. First, we examine the univariate problem. By assumption, $g(z)$ is analytic at $\tau$, so that $g(h(z))$, which is singular at $\rho_h$, is analytic in a $\Delta$–domain. Its singular expansion is obtained by composing the regular expansion of $g(z)$ at $\tau$ with the singular expansion of $h(z)$ at $\rho_h$:

$$F_0(z) \equiv g(h(z)) = g(\tau) - cg'(\tau)(1 - z/\rho)^\lambda (1 + o(1)).$$

Thus, $F_0(z)$ satisfies the conditions of singularity analysis, and

$$f_n \equiv [z^n]F_0(z) = -\frac{cg'(\tau)}{\Gamma(-\lambda)} n^{-\lambda-1}(1 + o(1)). \tag{9.20}$$

For the bivariate problem, choose a fixed $r > 1$ such that $r < \rho_g/\tau_h$. For any $u$ with $|u| \le r$, the singular expansion of $g(uh(z))$ is similarly

$$\begin{aligned} F(z,u) = g(uh(z)) &= g(u\tau - cu(1 - z/\rho)^\lambda + o((1 - z/\rho)^\lambda)) \\ &= g(u\tau) - cug'(u\tau)(1 - z/\rho)^\lambda + o((1 - z/\rho)^\lambda). \end{aligned}$$

Thus, we have immediately (!), by singularity analysis,

$$\lim_{n\to\infty} \frac{[z^n]F(z,u)}{[z^n]F(z,1)} = \frac{ug'(u\tau)}{g'(\tau)}.$$

By the continuity theorem, this is enough to establish convergence to the discrete limit law with PGF $ug'(\tau u)/g'(\tau)$, and the lemma is established. □

In the labelled universe, the functional composition schema encompasses the sequence, set, and cycle constructions: take for $g$ the functions

$$Q(w) = \frac{1}{1-w}, \quad E(w) = e^w, \quad L(w) = \log\frac{1}{1-w}.$$

The lemma then specializes to the combinatorial constructions of sequences, sets, and cycles. The results involve discrete laws of the Poisson or geometric type and Fig. 9.6 summarizes the definitions as well as the PGFs of the laws that occur.

**Proposition 9.2 (Subcritical constructions)** *Consider the constructions of* sequence $(g(z) = Q(z))$, set $(g(z) = E(z))$ *and* cycle $(g(z) = L(z))$, *and assume the subcriticality conditions of the previous lemma, namely $\tau < 1$, $\tau < \infty$, $\tau < 1$, respectively. Then, the distribution of the number of components, $f_{n,k}/f_n$, admits a discrete limit law that is of shifted 2–geometric, Poisson, and geometric type respectively,*

$$p_k^{seq} = (1-\tau)^2 k\tau^{k-1}, \qquad p_k^{set} = e^{-\tau}\frac{\tau^{k-1}}{(k-1)!}, \qquad p_k^{cyc} = (1-\tau)\tau^{k-1},$$

*for $k \ge 1$.*

| Law | $p_k$ | PGF | |
|---|---|---|---|
| Geometric($a$) | $(1-a)a^k$ | $\frac{1-a}{1-au}$ | Cycle |
| 2–Geometric($a$) | $(1-a)^2(k+1)a^k$ | $\left(\frac{1-a}{1-au}\right)^2$ | Sequence |
| Poisson($\lambda$) | $e^{-\lambda}\frac{\lambda^k}{k!}$ | $\exp(\lambda(u-1))$ | Set |

Figure 9.6: A table of some basic discrete laws and their occurrence in a subcritical schema.

PROOF. The discussion immediately specializes to sequences, sets, and cycles, where the PGF of the discrete limit law involves the derivatives

$$Q'(w) = \frac{1}{(1-w)^2}, \quad E'(w) = e^w, \quad L'(w) = \frac{1}{1-w}.$$

The last two cases give rise to the classical Poisson and geometric law. The first case gives rise to the "2–geometric" law of Figure 9.6 whose name comes from the fact that its PGF is the square of the PGF of a geometric variable. (Alternatively, a 2–geometric variable is the sum of 2 independent geometric random variables.) □

EXAMPLE 9. *Root degrees in trees.* Consider first the number of components in a sequence (ordered forest) of general Catalan trees. The bivariate OGF is

$$F(z,u) = \frac{1}{1-uh(z)}, \quad h(z) = \frac{1}{2}\left(1-\sqrt{1-4z}\right).$$

We have $\tau_h = 1/2 < \rho_g = 1$, so that the composition schema is subcritical. Thus, for a forest of total size $n$, the number $X_n$ of tree components satisfies

$$\lim_{n\to\infty} \Pr\{X_n = k\} = \frac{k}{2^{k+1}} \qquad (k \geq 1).$$

Since a tree is equivalent to a node appended to a forest, this asymptotic estimate also holds for the root degree of a general Catalan tree.

Consider next the number of components in a set (unordered forest) of Cayley trees. The bivariate EGF is

$$F(z,u) = e^{uh(z)}, \quad h(z) = ze^{h(z)}.$$

We have $\tau_h = 1 < \rho_g = +\infty$, again a subcritical composition schema. Thus the number $X_n$ of tree components in a random unordered forest of size $n$ admits the limit distribution

$$\lim_{n\to\infty} \Pr\{X_n = k\} = e^{-1}/(k-1)!, \qquad (k \geq 1),$$

a shifted Poisson law of parameter 1; asymptotically, the same property also holds for the root degree of a random Cayley tree □

The root degree in a random labelled nonplane tree (Cayley tree) admits in the asymptotic limit a Poisson law, while the root degree of a large plane tree (a Catalan tree) tends to a 2–geometric distribution. Proposition 9.2 shows, in a precise technical sense, that the 2–geometric law for Catalan trees is a direct reflection of planarity specified by a sequence construction, while the Poisson law arises from the set construction attached to nonplanarity.

Exercise 12. Discuss the schema "sets-of-sets",

$$F(z,u) = \exp(e^{uh(z)} - 1),$$

and exhibit a discrete limit law that involves the Bell numbers.
Discuss the schema of "sets-of-sequences",

$$F(z,u) = \exp\left(\frac{1}{1-uh(z)} - 1\right),$$

and exhibit a discrete limit law that involves Laguerre polynomials.

Exercise 13. Discuss the speed of convergence to the discrete limit law in the subcritical composition schema.

Another direct application of continuity of PGFs is the distribution of the number of $\mathcal{H}$-components of a fixed size $m$ in a composition $\Gamma[\mathcal{H}]$ with GF $g(h(z))$, again under the *subcriticality* condition. The bivariate GF is then

$$F(z,u) = g(h(z) + (u-1)h_m z^m),$$

with $h_m = [z^m]h(z)$. The singular expansion at $z = \rho$ is

$$F(z,u) = g(\tau+(u-1)h_m\rho^m) - cg'(\tau+(u-1)h_m\rho^m)(1-z/\rho)^\lambda) + o((1-z/\rho)^\lambda).$$

Thus, the PGF $p_n(u)$ for objects of size $n$ satisfies

$$\lim_{n\to\infty} p_n(u) = \frac{g'(\tau+(u-1)h_m\rho^m)}{g'(\tau)}. \qquad (9.21)$$

Like before this specializes in the case of sequences, sets, and cycles giving a result analogous to Proposition 9.2.

**Proposition 9.3 (Fixed size components)** *Under the subcriticality conditions of Lemma 9.2 and Proposition 9.2, the number of components of a fixed size $m$ in a random sequence, set, or cycle construction applied to a class with GF $h(z)$ admits a discrete limit law. With $h_m = [z^m]h(z)$, $\rho$ the radius of convergence of $h(z)$, and $\tau = h(\rho)$, we have the following.*

*For sequences, the limit law is 2-geometric of parameter*

$$a = \frac{h_m \rho^m}{1 - \tau + h_m \rho^m}.$$

*For sets, the limit law is Poisson with parameter*

$$\lambda = h_m \rho^m.$$

*For cycles, the limit is geometric of parameter*

$$a = \frac{h_m \rho^m}{1 - \tau + h_m \rho^m}.$$

EXAMPLE 10. *Root subtrees of size $m$.* In a Cayley tree, the number of root subtrees of size $m$ has, in the limit, a Poisson distribution,

$$p_k = e^{-\lambda} \frac{\lambda^k}{k!}, \qquad \lambda = \frac{m^{m-1} e^{-m}}{m!}.$$

In a general Catalan tree, the distribution is a 2–geometric,

$$p_k = (1-a)^2 (k+1) a^k, \qquad a^{-1} = 1 + \frac{m 2^{2m-1}}{\binom{2m-2}{m-1}}.$$

□

**Other combinatorial schemas.** Arbitrarily many schemas leading to discrete limit laws could be listed. Roughly, conditions are that the auxiliary variable $u$ does not affect the location nor the nature of the dominant singularity of $F(z,u)$. Such conditions are met by the subcritical schemas, since the auxiliary variable only appears as a multiplicative coefficient in a local singular expansion. An even simpler example is the product schema,

$$F(z,u) = A(uz) \cdot B(z),$$

that corresponds to a product construction, $\mathcal{F} = \mathcal{A} \times \mathcal{B}$, with $u$ marking the size of the $\mathcal{A}$–component in the product.

Assume that the radii of convergence satisfy $\rho_A > \rho_B$ and that $B(z)$ has a unique dominant singularity of the algebraic-logarithmic type. Then, by singularity analysis, we have

$$[z^n]F(z,u) \sim A(u\rho)b_n,$$

with $b_n = [z^n]B(z)$, and $\rho = \rho_B$. Since

$$[z^n]F(z,1) \sim A(\rho)b_n,$$

the size of the $\mathcal{A}$ component has a discrete limit law with PGF,

$$p(u) = \frac{A(\rho u)}{A(\rho)}.$$

The analysis of $m$–cycles in permutations can also be cast into this category. In effect, one can regard the class of permutations as a labelled product $\mathcal{A}\star\mathcal{B}$, where $\mathcal{A}$ is the set of $m$-cycles, and $\mathcal{B}$ is the class of $m$-derangements; see Section 3.3. Then, the bivariate GF for the number of $m$-cycles is

$$F(u,z) = e^{uz^m/m} \cdot \frac{e^{-z^m/m}}{1-z}.$$

Up to an unessential simplification (with $u^m$ being replaced by $u$, for convenience), this GF is also of the product type, with $\rho = 1$.

> Exercise 14. Give a direct elementary proof of the limit law for products under the weaker requirement that the coefficient of $B$ satisfy $b_{n+1}/b_n \to \rho^{-1}$.

As a last example here, we discuss the length of the longest initial run of $a$'s in random binary words satisfying various types of constraints. This discussion completes the informal presentation of Section 9.2. The basic combinatorial objects are the set $\mathcal{W} = \{a,b\}^\star$ of binary words. A word $w \in \{\mathcal{W}\}$ can also be viewed as describing a walk in the plane, provided one interprets $a$ and $b$ as the vectors $(+1,+1)$ and $(+1,-1)$ respectively. Such walks in turn describe fluctuations in coin tossing games [18]. What is especially interesting here is to observe the complete chain where a specific constraint leads in succession to a combinatorial decomposition , a specific analytic type of BGF, and a local singular structure that is then reflected by a particular limit law.

Example 11. *Initial runs in random walks.* We consider here four types of walks:

- unconstrained walks corresponding to $\mathcal{W}$;
- gambler's ruin sequences $\mathcal{W}_1$, whose abscissa is always nonnegative, and whose final altitude is 0;
- balanced walks $\mathcal{W}_2$, whose final altitude is 0.
- gambler's no-credit sequences $\mathcal{W}_3$, whose abscissa is always nonnegative;

The parameter $\chi$ of interest is in all cases the length of the longest initial run of $a$'s.

First, the unconstrained walks obey the decomposition (already used in the analysis of success runs in earlier chapters),

$$\mathcal{W} = a^\star (b a^\star)^\star.$$

Thus, the BGF is

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

By singularity analysis of the pole at $\rho = 1/2$, the PGF of $\chi$ on random words of $\mathcal{W}_n$ satisfies

$$p_n(u) \sim \frac{\frac{1}{2}}{1-\frac{u}{2}},$$

and, as expected (Section 9.2), this corresponds to a limit geometric law of parameter $\frac{1}{2}$.

As is well-known, the ruin sequences $\mathcal{W}_1$ play an important rôle in combinatorial decompositions. A ruin sequence decomposes into "arches" that are ruin sequences encapsulated by a pair $a, b$,

$$\mathcal{W}_1 = (a\mathcal{W}_1 b)^\star,$$

which yields a GF of the Catalan domain,

$$W_1(z) = \frac{1}{1-z^2 W_1(z)}, \qquad W_1(z) = \frac{1-\sqrt{1-4z^2}}{2z^2}.$$

To extract the initial run of $a$'s, we observe that a word whose initial $a$-run is $a^k$ contains $k$ components of the form $b\mathcal{W}_1$. This corresponds to a decomposition in terms of the first traversal of altitudes $k-1, \ldots, 1, 0$:

$$\mathcal{W}_1 = \sum_{k \geq 0} a^k (b\mathcal{W}_1)^k.$$

Thus, the BGF is

$$F_1(z,u) = \frac{1}{1 - z^2 u W_1(z)}.$$

This is an even function of $z$. In terms of the singular element, $\delta = (1 - 4z)^{1/2}$, one finds

$$F_1(z^{1/2}, u) = \frac{2}{2-u} - \frac{2u}{(2-u)^2}\delta + \mathcal{O}(\delta^2),$$

as $z \to 1/4$. Thus, the PGF of $\chi$ on random words of $\mathcal{W}_{1,2n}$ satisfies

$$p_{1,2n}(u) \sim \frac{u}{(2-u)^2},$$

which is the PGF of a shifted 2–geometric of parameter 1/2. (Naturally, in this case, explicit expressions for the combinatorial distribution are available, as this is equivalent to the classical ballot problem.)

As a balanced walk decomposes into a sequence of arches, either positive or negative, we have

$$W_2(z) = \frac{1}{1 - 2z^2 W_1(z)} = \frac{1}{\sqrt{1-4z^2}}.$$

The walks $\mathcal{W}_2^+$ that start with at least one $a$ have a decomposition similar to $\mathcal{W}_1$:

$$\mathcal{W}_2^+ = \left(\sum_{k\geq 1} a^k b(\mathcal{W}_1 b)^{k-1}\right) \cdot \mathcal{W}_2,$$

since they factor uniquely as a $\mathcal{W}_1$ component that hits 0 for the first time followed by a $\mathcal{W}_2$ oscillation. Thus,

$$W_2^+(z) = \frac{z^2}{1 - z^2 W_1(z)} W_2(z).$$

The walks $\mathcal{W}_2^- = \mathcal{W}_2 \setminus \mathcal{W}_2^+$ consist of either the empty word or a sequence of positive or negative arches starting with a negative arch, so that

$$W_2^-(z) = 1 + \frac{z^2 W_1(z)}{1 - 2z^2 W_1(z)}.$$

The BGF results from these decompositions:

$$F_2(z,u) = \frac{z^2 u}{1 - z^2 u W_1(z)} W_2(z) + 1 + \frac{z^2 W_1(z)}{1 - 2z^2 W_1(z)}.$$

Again, the singular expansion is obtained mechanically,

$$F_2(z^{1/2}, u) = \frac{1}{2-u}\frac{1}{\delta} + \mathcal{O}(1),$$

where $\delta = (1-4z)^{1/2}$. Thus, the PGF of $\chi$ on random words of $\mathcal{W}_{2,2n}$ satisfies

$$p_{2,2n}(u) \sim \frac{1}{2-u}.$$

The limit law is a geometric one with parameter 1/2.

A no-credit gambler's sequence of type $\mathcal{W}_3$ decomposes into an initial run $a^k$, a succession of descents with their companion (positive) arches in some number $\ell \leq k$, and a succession of ascents with their corresponding (positive) arches. The computations are similar to the previous cases, more intricate, but still "automatic". One finds that

$$F_3(z,u) = \left(\frac{XY}{(1-X)(1-Y)} - \frac{XY^2}{(1-XY)(1-Y)}\right)\frac{1}{1-Y} + \frac{1}{1-X},$$

with $X = zu$, $Y = zW_1(z)$, so that

$$F_3(z,u) = 2\,\frac{1-u-2\,z+2\,uz^2+(u-1)\sqrt{1-4\,z^2}}{(1-zu)\left(1-2\,z-\sqrt{1-4\,z^2}\right)\left(2-u+u\sqrt{1-4\,z^2}\right)}.$$

There are now two singularities at $z = \pm\frac{1}{2}$, with singular expansions,

$$F_3(z,u) \underset{z\to 1/2}{=} \frac{u\sqrt{2}}{(2-u)^2}\frac{1}{\sqrt{1-2z}} + \mathcal{O}(1), \qquad F_3(z,u) \underset{z\to -1/2}{=} \frac{4-u}{4-u^2} + o(1),$$

so that only the singularity at 1/2 matters asymptotically. Then, we have

$$p_{3,n}(u) \sim \frac{u}{(2-u)^2},$$

and the limit law is a shifted 2–geometric of parameter 1/2. □

EXERCISE 15. Carry out similar analyses for the "ternary" ballot problem, where $a$ is interpreted as the vector $(1,2)$ and $b$ as $(1,-1)$.

## 9.4 Continuous limit laws

Throughout this chapter, we are interested in sequences of random variables that arise from an integer valued combinatorial parameter $\chi$ defined on a combinatorial class $\mathcal{F}$. It is a fact that, when the mean $\mu_n$ and the standard deviation $\sigma_n$ of $\chi$ on $\mathcal{F}_n$ tend to infinity as $n$ gets large, then a continuous limit law *usually* holds.

A random variable $Y$ is specified by its *distribution function*,

$$\Pr\{Y \leq x\} = G(x),$$

and is said to be *continuous* if $G(x)$ is continuous. In that case, $G(x)$ has no jump, and there is no single value in the range of $Y$ that bears nonzero probability mass. If in addition $G(x)$ is differentiable, the random variable $Y$ is said to have a *density*, $g(x) = G'(x)$, so that

$$\Pr\{x < Y \leq x + dx\} = g(x)\, dx.$$

A function $f(Y)$ of the random variable $y$ has its expectation defined by

$$E\{f(Y)\} = \int_{-\infty}^{+\infty} f(x)\, dG(x),$$

where the integral is to be taken in the sense of *Stieltjes integration*. If $Y$ has a density, then a standard (Riemann or Lebesgue) integral may be used,

$$E\{f(Y)\} = \int_{-\infty}^{+\infty} f(x) g(x)\, dx.$$

Moments are then defined as in the discrete case by $\mu^{(r)} = E\{X^r\}$.

A particularly important case is the standard *Gaussian* or *normal* distribution function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw,$$

also called the *error function* (erf), the corresponding density being

$$\xi(x) \equiv \Phi'(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

### 9.4.1 Convergence of distributions

Convergence of probability distributions is defined as convergence of distribution functions, after a possible normalization. Graphically, this convergence is illustrated by Fig. 9.7 (top), which may be compared to Fig. 9.2

where a coarser normalization has been used in the context of concentration properties.

**Definition 9.4** *Let $Y$ be a continuous random variable with distribution function $G_Y(x)$. A sequence of random variables $Y_n$ with distribution functions $G_{Y_n}(x)$ is said to converge in distribution to $Y$ if,* pointwise, *for each $x$,*

$$\lim_{n\to\infty} G_{Y_n}(x) = G(x).$$

*In that case, one writes*

$$Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y, \qquad G_{Y_n} \stackrel{\mathcal{D}}{\Longrightarrow} G_Y.$$

This convergence of distribution functions is also called *weak convence* in the probabilistic literature. (The definition above is in fact the mildly restrictive specialization to the case of a continuous limit.)

In combinatorial applications, one tends to deal with integer–valued variables $X_n$ that represent combinatorial parameters, $\chi$. Convergence in distribution to a continuous limit may then arise only after normalization. If $X$ is a random variable with mean $\mu$ and standard deviation $\sigma$, then its *normalized* version $X^*$ is defined as

$$X^* = \frac{X-\mu}{\sigma}, \qquad X = \mu + \sigma X^*$$

with distribution functions satisfying

$$G_{X^*}(x) = G_X(\mu + \sigma x), \qquad G_X(x) = G_{X^*}(\frac{x-\mu}{\sigma}).$$

Normalization is a simple way to reduce random variables to zero mean and unit variance. It is from there that limit laws can usually be found[2].

**Definition 9.5** *If $X_n$ is a sequence of random variables and there exists a continuous variable $Y$ such that the normalized sequence $X_n^*$ converges in distribution to $Y$,*

$$X_n^* \stackrel{\mathcal{D}}{\Longrightarrow} Y,$$

*then $X_n$ is said to satisfy a* continuous limit law *of type $Y$. The distribution functions $G_{X_n}$ then satisfy*

$$\lim_{n\to\infty} G_{X_n^*}(x) \equiv \lim_{n\to\infty} G_{X_n}(\mu_n + x\sigma_n) = G_Y(x),$$

*for each $x$.*

[2] In the case of the "direct" occurrence of discrete limit laws in the previous section 9.3, normalization could also have been introduced but it would unnecessarily complicate the presentation. In the case of continuous laws, normalization becomes a necessity. More general normalizations where the centering and scaling constants differ from the mean and standard deviation, though useful in some contexts, will not be needed here.

The definition does not require uniform convergence. If uniformity holds, then

$$\epsilon_n := \sup_x |G_{X_n^*}(x) - G_Y(x)|$$

tends to 0 as $n \to \infty$. The quantity $\epsilon_n$ is called the *speed of convergence* to the limit law. It is a known fact that convergence to a Gaussian limit is always uniform.

Discrete limit laws can be established via convergence of probability generating functions to a common limit, as asserted by the continuity theorem for PGFs, Theorem 9.3. In the case of continuous limit laws, one first needs an analogue of the PGF. Not too unexpectedly, this analogue involves an integral rather than a sum expressing a PGF. Also, for determinacy, extending terms like $u^k$ to nonintegral values of $k$ requires an exponential form of the argument, like $u = e^s$.

The two integral transforms associated to a random variable $Y$ and that play the rôle of the PGF are

— the *Laplace transform* —also called the *moment generating function*— $\lambda_Y(s)$ defined by

$$\lambda_Y(s) := E\{e^{sY}\} = \int_{-\infty}^{+\infty} e^{sx}\, dG(x);$$

— the *Fourier transform* —also called the *characteristic function*— $\phi_Y(t)$ defined by

$$\phi_Y(t) := E\{e^{itY}\} = \int_{-\infty}^{+\infty} e^{itx}\, dG(x).$$

These two transforms are formal variants of each other, since

$$\phi_Y(t) = \lambda_Y(it).$$

One usually takes $s$ and $t$ to be real in these two definitions. With this restriction, the Fourier transform $\phi(t)$ is then always defined, for any distribution, since $\int_R dG(x) = 1$ and $|e^{itx}| \le 1$. The Laplace transform need not exist for $s$ real and $s \neq 0$ depending upon the particular distribution. Its existence requires a fast enough (exponential) decay of the "tails" of the distribution $G(x)$ as $x \to \pm\infty$. For the Gaussian distribution, we have

$$\lambda(s) = e^{s^2/2}, \qquad \phi(t) = e^{-t^2/2}.$$

These definitions generalize the PGF: if $Y$ is discrete and nonnegative with PGF $p_Y(u)$, then

$$\lambda_Y(s) = p_Y(e^s), \qquad \phi_Y(t) = p(e^{it}).$$

In this particular case, the Laplace transform always exists for $s < 0$, and the Fourier transform is periodic of period $2\pi$.

The moments are accessible from either transform,

$$\mu^{(r)} := E\{Y^r\} = \frac{d^r}{ds^r}\lambda(s)\bigg|_{s=0} = (-i)^r \ \frac{d^r}{dt^r}\phi(t)\bigg|_{t=0}.$$

In particular, we have

$$\begin{aligned} \mu &= \frac{d}{ds}\lambda(s)\bigg|_{s=0} = -i\ \frac{d}{dt}\phi(t)\bigg|_{t=0} \\ \mu^{(2)} &= \frac{d^2}{ds^2}\lambda(s)\bigg|_{s=0} = -\ \frac{d}{dt}\phi(t)\bigg|_{t=0} \\ \sigma^2 &= \frac{d^2}{ds^2}\log\lambda(s)\bigg|_{s=0} = -\ \frac{d^2}{dt^2}\log\phi(t)\bigg|_{t=0}. \end{aligned} \tag{9.22}$$

The direct expression of the standard deviation in terms of $\log\lambda(s)$, called the *cumulant generating function*, often proves computationally handy. Centering and scaling is easily performed on transforms: if $X$ has mean $\mu$ and variance $\sigma^2$, then $X^*$ has transforms

$$\phi_{X^*}(t) = E\{\exp(it\frac{X-\mu}{\sigma})\} = e^{-i\mu t/\sigma}\phi_X(\frac{t}{\sigma}),$$

$$\lambda_{X^*}(s) = E\{\exp(s\frac{X-\mu}{\sigma})\} = e^{-i\mu s/\sigma}\lambda_X(\frac{s}{\sigma}).$$

### 9.4.2 Continuity theorem for integral transforms

There are two classical versions of the continuity theorem, one for characteristic functions, the other for Laplace transforms. Both may be viewed as extensions of the continuity theorem for PGF's. Characteristic functions always exist and the corresponding continuity theorem gives a necessary and sufficient condition for convergence of distributions. As they are a universal tool, characteristic functions are therefore often favoured in the probabilistic literature. In the context of this book, strong analyticity properties go along with combinatorial constructions and both transforms usually exist.

**Theorem 9.4 (Continuity of integral transforms)** *Let $Y, Y_n$ be random variables with Fourier* *transforms (characteristic functions)* $\phi(t), \phi_n(t)$, *and assume that* $Y$ *has a continuous distribution function. A necessary and sufficient condition for the convergence in distribution,*

$$Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y,$$

*is that,* pointwise, *for each real* $t$,

$$\lim_{n\to\infty} \phi_n(t) = \phi(t).$$

*Let* $Y, Y_n$ *be random variables with Laplace transforms* $\lambda(s), \lambda_n(s)$ *that exist in a common interval* $[-s_0, s_0]$. *If,* pointwise *for each real* $s \in [-s_0, s_0]$,

$$\lim_{n\to\infty} \lambda_n(s) = \lambda(s),$$

*then the* $Y_n$ *converge in distribution to* $Y$,

$$Y_n \stackrel{\mathcal{D}}{\Longrightarrow} Y,$$

Proof. See Billingsley's book [8, Sec. 26], for Fourier transforms, and [8, p. 408], for Laplace transforms. □

> Exercise 16. Construct a sequence $Y_n$ that converges in distribution to a standard Gaussian limit $Y$, but for which $\lambda_n(s)$ only exists at $s = 0$.

The continuity theorem for PGFs eventually relies on continuity of the Cauchy coefficient formula that realizes the inversion needed in recovering coefficients from PGFs. Similarly, the continuity theorem for integral transforms may be viewed as expressing the continuity of inverse Laplace or Fourier transforms, in the specific context of probability distribution functions.

The next theorem is an effective version of the Fourier inversion theorem that proves especially useful for characterizing speeds of convergence. It bounds in a constructive manner the sup-norm distance between two distribution functions by a special metric distance between their characteristic functions. Recall that $||f||_\infty := \sup_{x\in R} |f(x)|$.

**Theorem 9.5 (Berry-Esseen inequality)** *Let* $F, G$ *be distribution functions with characteristic functions* $\phi(t), \gamma(t)$. *Assume that* $G$ *has a bounded derivative. There exist absolute constants* $c_1, c_2$ *such that for any* $T > 0$,

$$||F - G||_\infty \le c_1 \int_{-T}^{+T} \left| \frac{\phi(t) - \gamma(t)}{t} \right| dt + c_2 \frac{||G'||_\infty}{T}.$$

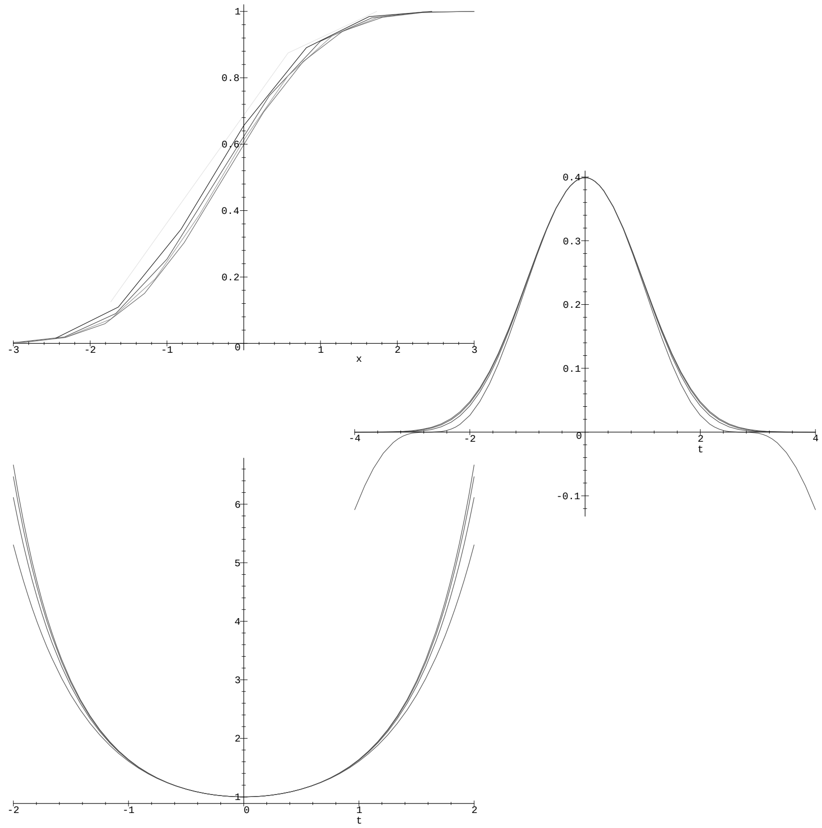


Figure 9.7: The normalized distribution function of the binomial law (top), the corresponding Fourier transforms (middle), and the Laplace transforms (bottom), for $n = 3, 6, 9, 12, 15$.

The distribution functions centred around the mean $\mu_n = n/2$ and scaled according to the standard deviation $\sigma_n = n^{1/2}/2$ quickly converge to a limit which is the Gaussian error function,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw.$$

Accordingly, the corresponding Fourier transforms —or characteristic functions— converge to

$$\phi(t) = e^{-t^2/2},$$

while the Laplace transforms —or moment generating functions— converge to

$$\lambda(s) = e^{s^2/2}.$$

PROOF. See Feller [19, p. 538] who gives

$$c_1 = \frac{1}{\pi}, \qquad c_2 = \frac{24}{\pi}$$

as possible values for the constants. □

This theorem is typically used with $G$ being the limit distribution function (often a Gaussian for which $||G'||_\infty = (2\pi)^{-1/2}$) and $F = F_n$ a distribution that belongs to a sequence converging to $G$. The quantity $T$ may be assigned an arbitrary value; the one giving the best bound in a specific application context is then normally chosen.

**Powers and the central limit theorem.** The binomial distribution is defined as the distribution of a random variable $X_n$ with PGF

$$p_n(u) = \left(\frac{1}{2} + \frac{u}{2}\right)^n,$$

and characteristic function, $\phi_n(t) = p_n(e^{it})$. The mean is $\mu_n = n/2$ and the variance is $\sigma_n^2 = n/4$. Therefore, the normalized variable $X_n^* = (X_n - \mu_n)/\sigma_n$ has characteristic function

$$\phi_n^*(t) = \left(\cosh(\frac{it}{\sqrt{n}}\right)^n. \tag{9.23}$$

The asymptotic form is easily found by taking logarithms, and one has

$$\log \phi_n^*(t) = n \log \left(1 - \frac{t^2}{2n} + \frac{t^4}{6n^2} + \cdots\right) = -\frac{t^2}{2} + \mathcal{O}(\frac{1}{n}), \tag{9.24}$$

pointwise, for any fixed $t$, as $n \to \infty$. This establishes convergence to the Gaussian limit. In addition, the Berry-Esseen inequalities show that the speed of convergence is $\mathcal{O}(n^{-1/2})$, a fact that is easily verified directly using Stirling's formula [54].

EXERCISE 17. Use characteristic functions to extend the normal limit law to the Bernoulli distribution with PGF $(p + qu)^n$, for fixed $p, q$ with $p + q = 1$.

Show that if $q$ depends on $n$ in such a way that $q = \lambda/n$ for some fixed $\lambda$, then the limit law becomes Poisson of rate $\lambda$.

The *central limit theorem* of probability theory concerns sums $S_n = T_1 + \cdots + T_n$ of random variables. It is assumed that the $T_j$ are independent with a common distribution of mean $\mu$ and standard deviation $\sigma$. Then the normalized variable $S_n^*$ converges to the standard normal distribution,

$$S_n^* \equiv \frac{S_n - \mu n}{\sigma\sqrt{n}} \overset{\mathcal{D}}{\Longrightarrow} \mathcal{N}(0,1).$$

The proof is like before based on local expansions of characteristic functions. It suffices to consider the case of zero-mean variables. We have, pointwise for each $t$,

$$\phi\left(\frac{t}{\sigma\sqrt{n}}\right)^n = \left(1 - \frac{t^2}{2n} + o(\frac{t^2}{2n})\right)^n \to e^{-t^2/2},$$

like in Eq. (9.23,9.24). See [19, p. 259] or [8, Sec. 27] for details.

There are many extensions of the central limit theorem, to variables that are independent but not necessarily identically distributed (the Lindeberg–Lyapounov conditions) or variables that are only dependent in some weak sense (mixing conditions); see the discussion in [8, Sec. 27].

### 9.4.3 Quasi-powers

The central limit theorem of probability theory asserts that the sum of a large number of independent identically distributed (i.i.d.) random variables converges to a Gaussian limit. Analytically, the characteristic function of the sum is a large power of a fixed function that, after normalization, converges to the characteristic function of the Gaussian limit, namely $e^{-t^2/2}$. In a combinatorial context, this approach admits a fruitful extension. As we now show, it suffices that the PGF of a combinatorial parameter behaves "nearly" like a large power of a fixed function to ensure convergence to a Gaussian limit. We first illustrate this point by considering the Stirling cycle distribution.

Consider the Stirling cycle numbers and let $X_n$ be the random variable with PGF,

$$p_n(u) = \binom{n+u-1}{n} = \frac{u(u+1)(u+2)\cdots(u+n-1)}{n!} = \frac{\Gamma(u+n)}{\Gamma(u)\Gamma(n+1)}.$$

We have for fixed $u$ near 1,

$$p_n(u) = \frac{n^{u-1}}{\Gamma(u)}\left(1 + \mathcal{O}(\frac{1}{n})\right) = \frac{1}{\Gamma(u)}\left(e^{(u-1)}\right)^{\log n}\left(1 + \mathcal{O}(\frac{1}{n})\right). \quad (9.25)$$

As results from Stirling's formula for the Gamma function or from singularity analysis of $[z^n](1-z)^{-u}$ (Chapter 5), the error term in (9.25) is

$\mathcal{O}(n^{-1})$ when $u$ stays around 1, for instance $|u-1| \leq \frac{1}{2}$. Thus, as $n \to +\infty$, $p_n(u)$ is approximately a "large power" of $e^{u-1}$ taken with exponent $\log n$, multiplied by a fixed function, $(\Gamma(u))^{-1}$. By analogy to the central limit theorem, we may expect a Gaussian law.

The mean satisfies $\mu_n = \log n + \gamma + o(1)$, the standard deviation satisfies $\sigma_n = \sqrt{\log n} + o(1)$. We thus consider the normalized random variable,

$$X_n^* = \frac{X_n - L - \gamma}{\sqrt{L}}, \qquad L = \log n,$$

whose characteristic function is

$$\phi_n^*(t) = \frac{e^{-it(L^{1/2}+\gamma L^{-1/2})}}{\Gamma(e^{it/\sqrt{L}})} \exp\left(L(e^{it/\sqrt{L}} - 1)\right)\left(1 + \mathcal{O}(\frac{1}{n})\right).$$

For fixed $t$, with $L \to \infty$, the logarithm is then found mechanically to satisfy

$$\log \phi_n^*(t) = -\frac{t^2}{2} + \mathcal{O}((\log n)^{-1/2}).$$

This is sufficient to establish a Gaussian limit law,

$$\lim_{n\to\infty} \Pr\left\{X_n \leq \log n + \gamma + x\sqrt{\log n}\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw. \qquad (9.26)$$

This result was obtained by Goncharov, as early as 1943.

The cycle example is characteristic of the occurrence of Gaussian laws in analytic combinatorics. What happens is that the approximation (9.25) by a power with "large" exponent $\beta_n = \log n$ leads after normalization, to the characteristic function of a Gaussian variable, namely $e^{-t^2/2}$. From there, the limit distribution (9.26) results by the continuity theorem. This is in fact a very general phenomenon, as demonstrated by the following theorem of Hwang [36] that builds upon earlier statements of Bender and Richmond [5].

**Theorem 9.6 (Quasi-Powers)** *Let the $X_n$ be nonnegative discrete random variables with probability generating function $p_n(u)$. Assume that,* uniformly *in a fixed complex neighbourhood of $u = 1$,*

$$p_n(u) = A(u)\,(B(u))^{\beta_n}\left(1 + \mathcal{O}(\frac{1}{\kappa_n})\right), \qquad (9.27)$$

*where $A(u), B(u)$ are analytic at $u = 1$ $A(1) = B(1) = 1$, and $B(u)$ satisfies the so-called* "variability condition"*,*

$$Var(B(u)) \equiv B''(1) + B'(1) - B'(1) \neq 0.$$

*Under these conditions, the distribution of $X_n$ is asymptotically Gaussian, and the speed of convergence to the Gaussian limit is $\mathcal{O}(\kappa_n^{-1} + \beta_n^{-1/2})$:*

$$\Pr\left\{\frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}} \le x\right\} = \Phi(x) + O\left(\frac{1}{\kappa_n} + \frac{1}{\sqrt{\beta_n}}\right).$$

*The mean and variance of $X_n$ satisfy*

$$\begin{aligned} \mu_n &= \beta_n\, Mean(B(u)) + Mean(A(u)) + \mathcal{O}(\tfrac{1}{\kappa_n}) \\ \sigma_n &= \beta_n\, Var(B(u)) + Var(A(u)) + \mathcal{O}(\tfrac{1}{\kappa_n}) \end{aligned} \qquad (9.28)$$

This theorem is a direct application of the following lemma, also due to Hwang, that applies more generally to arbitrary discrete or continuous distributions, and is thus entirely phrased in terms of integral transforms.

**Lemma 9.3 (Quasi-Powers: general case)** *Assume that the Laplace transforms $\lambda_n(s) = E\{e^{sX_n}\}$ of a sequence of random variables $X_n$ are analytic in a disc $|s| < \rho$, for some $\rho > 0$, and satisfy there an expansion of the form*

$$\lambda_n(s) = e^{\beta_n U(s) + V(s)}\left(1 + \mathcal{O}(\frac{1}{\kappa_n})\right), \qquad (9.29)$$

*with $\beta_n, \kappa_n \to +\infty$ as $n \to +\infty$, and $U(s), V(s)$ analytic in $|s| \le \rho$. Assume also the variability condition,*

$$U''(0) \ne 0.$$

*Under these assumptions, the mean and variance of $X_n$ satisfy*

$$\begin{aligned} E\{X_n\} &= \beta_n U'(0) + V'(0) + \mathcal{O}(\kappa_n^{-1}), \\ Var\{X_n\} &= \beta_n U''(0) + V''(0) + \mathcal{O}(\kappa_n^{-1}). \end{aligned} \qquad (9.30)$$

*The distribution of $X_n$ is asymptotically Gaussian and the speed of convergence to the Gaussian limit is $\mathcal{O}(\kappa_n^{-1} + \beta_n^{-1/2})$.*

Proof. This follows the lines of Hwang's thesis [36]. First, we estimate the mean and variance. The variable $s$ is *a priori* restricted to a small neighbourhood of 0. By assumption, the function

$$\log \lambda_n(s) = \beta_n U(s) + V(s) + \mathcal{O}(\frac{1}{\kappa_n})$$

is analytic at 0. This asymptotic expansion carries over, with the same type of error term, to derivatives at 0 because of analyticity. This fact can be checked directly from Cauchy integral representations,

$$\frac{d^r}{ds^r}\log\lambda_n(s)\Big|_{s=0} = \frac{1}{2i\pi}\int_\gamma \log\lambda_n(s)\,\frac{ds}{s^{r+1}},$$

upon using a small but fixed integration contour $\gamma$ and taking advantage of the basic expansion of $\log\lambda_n(s)$. Thus, the mean and variance satisfy the estimates of (9.30).

Next, we consider the normalized variable,

$$X_n^\star = \frac{X_n - \beta_n U'(0)}{\sqrt{\beta_n U''(0)}}, \qquad \lambda_n^\star(s) = E\{e^{sX_n^\star}\}.$$

We have

$$\log\lambda_n^\star = -\frac{\beta_n U'(0)}{\sqrt{\beta_n U''(0)}}s + \log\lambda_n(\frac{s}{\sqrt{\beta_n U''(0)}}).$$

Local expansions to third order based on the assumption (9.29) show that

$$\log\lambda_n^\star = \frac{s^2}{2} + \mathcal{O}(\frac{|s|+|s|^3}{\beta_n^{1/2}}) + \mathcal{O}(\frac{1}{\kappa_n}), \tag{9.31}$$

uniformly with respect to $s$ in a disc of radius $\mathcal{O}(\beta_n^{1/2})$, and in particular in any fixed neighbourhood of 0. This is enough to conclude to convergence in distribution to a Gaussian limit, by the continuity theorem of either Laplace transforms (restricting $s$ to be real) or of Fourier transforms (taking $s = it$).

Finally, the speed of convergence results from the Berry-Esseen inequalities. Take $T \equiv T_n = c\beta_n^{1/2}$, where $c$ is taken sufficiently small but nonzero, in such a way that local expansions at 0 apply. Then, the expansion (9.31) instantiated at $s = it$ entails that the quantity

$$J_n := \int_{-T_n}^{T_n} \left|\frac{\lambda_n^\star(it) - e^{-t^2/2}}{t}\right| dt$$

satisfies

$$J_n = \mathcal{O}(\beta_n^{-1/2} + \kappa_n^{-1}),$$

and the statement follows by the Berry-Esseen theorem. □

Theorem 9.6 applies immediately to the Stirling cycle distribution for which the estimate (9.25) was derived. It shows in addition that the speed of convergence is $\mathcal{O}((\log n)^{-1/2})$ for this distribution.

The Quasi-Powers Theorem under either form (9.27) or (9.29) can be read *formally* as expressing the distribution of a random variable

$$Z = Y + T_1 + T_2 + \cdots + T_{\beta_n},$$

where $Y$ "corresponds" to $e^{V(s)}$ (or $A(u)$) and each $T_j$ to $e^{U(s)}$ (or $B(u)$). However, there is no *a priori* requirement that $\beta_n$ should be an integer, nor that $e^{U(s)}, e^{V(s)}$ be Laplace transforms of distribution functions. In a way, the theorem recycles the intuition that underlies the central limit theorem and makes use of the analytic machinery behind it. But, in applications, functions like $e^{U(s)}, e^{V(s)}$ do not necessarily admit an *a priori* probabilistic interpretation.

It is of particular importance to note that the conditions of Theorem 9.6 and Lemma 9.3 are purely local: *what is required is* local *analyticity of the quasi-power approximation at $u = 1$ for PGF's or, equivalently, $s = 0$ for Laplace transforms.* This important feature is ultimately due the normalization of random variables and transforms that goes along with limit laws

### 9.4.4 Singularity perturbation and Gaussian laws.

The main thread of this chapter is *bivariate generating functions.* In general, we are given a BGF $F(z, u)$ and aim at extracting a limit distribution from it. The quasi-power paradigm in the form (9.27) is what one should look for, in the case where the mean and the standard deviation both tend to infinity with the size $n$ of the model.

We proceed heuristically in this informal discussion. Start from the BGF and consider $u$ as a parameter. If singularity analysis applies to the counting generating function $F(z, 1)$, it leads to an approximation,

$$f_n \approx C \cdot \rho^{-n} n^{\alpha},$$

where $\rho$ is the dominant singularity of $F(z, 1)$ and $\alpha$ is related to the critical exponent of $F(z, 1)$ at $\rho$. A similar type of analysis is often applicable for $u$ near 1. Then, it is reasonable to expect an approximation for the $z$-coefficients of the bivariate GF,

$$f_n(u) \approx C(u)\rho(u)^{-n} n^{\alpha(u)}.$$

In this perspective, the corresponding PGF is of the form

$$p_n(u) \approx \frac{C(u)}{C(1)} \left(\frac{\rho(u)}{\rho(1)}\right)^{-n} n^{\alpha(u)-\alpha(1)}.$$

This strategy is thus a perturbation analysis of singular expansions with the auxiliary parameter $u$ being restricted to a small neighbourhood of 1.

In particular if only the singularity moves, we have a rough form

$$p_n(u) \approx \frac{C(u)}{C(1)} \left(\frac{\rho(u)}{\rho(1)}\right)^{-n},$$

suggesting a Gaussian law with mean and variance that are both $\mathcal{O}(n)$. If only the exponent moves,

$$p_n(u) \approx \frac{C(u)}{C(1)} n^{\alpha(u)-\alpha(1)},$$

suggesting again a Gaussian law but with mean and variance that are both $\mathcal{O}(\log n)$.

These cases point to the fact that a rather simple perturbation of the univariate analysis eventually yields limiting distributions. The quantitative analysis of the way singular expansions of a bivariate generating function $F(z,u)$ evolve when $u$ lies in an infinitesimal neighbourhood of 1 thus resorts to singularity perturbation asymptotics. For reasons just explained, such an analysis is an essential ingredient in the derivation of limit laws.

## 9.5 Explicit schemas

Each major coefficient extraction method of Chapters 4–6 plays a rôle in singularity perturbation analyses and the derivation of Gaussian limit laws. The next two subsections illustrate this point in the following contexts:

— *meromorphic analysis* for functions with polar singularities;

— *singularity analysis* for functions with algebraic–logarithmic singularity;

Saddle point analysis for functions with fast growth at their singularity is discussed later in Section 9.8.

Roughly, the decomposable character of many elementary combinatorial structures is reflected by strong analyticity properties of bivariate GF's that can then be subjected to the Quasi-Powers Theorem (Theorem 9.6). The coefficients extraction methods, especially singularity analysis, being based on contour integration supply the necessary uniformity required by the Quasi-Powers Theorem. (In contrast, Darboux's method or Tauberian theorems, being nonconstructive, are *not* normally applicable in this context.)

### 9.5.1 The meromorphic schema

This section discusses schemas that relie on the analysis of coefficients of meromorphic functions, as discussed in Chapter 4. It is largely based on works of Bender who, starting with [2], was the first to propose abstract analytic schemas leading to Gaussian laws in analytic combinatorics.

**The surjection distribution.** We revisit the distribution of image cardinality in surjections for which the concentration property has been established before. This example serves to introduce bivariate asymptotics in the meromorphic case. Consider the distribution of image cardinality in surjections, with BGF

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

Restrict $u$ near 1, for instance $|u-1| \le \frac{1}{10}$. The function $F(z,u)$, as a function of $z$, is meromorphic with singularities at

$$\rho(u) + 2ik\pi, \qquad \rho(u) = \log(1+\frac{1}{u}).$$

The principal determination of the logarithm is used (with $\rho(u)$ near $\log 2$ when $u$ is near 1). It is then seen that $\rho(u)$ stays within 0.06 from $\log 2$, for $|u-1| \le \frac{1}{10}$. Thus $\rho(u)$ is the unique dominant singularity of $F$, the next nearest one being $\rho(u) \pm 2i\pi$ with modulus larger than 5.

From the coefficient analysis of meromorphic functions (Chapter 4), the quantities $f_n(u) = [z^n]F(z,u)$ are estimated as follows,

$$\begin{aligned} f_n(u) &= \operatorname{Res}\left(F(z,u)z^{-n-1}\right)_{z=\rho(u)} + \frac{1}{2i\pi}\int_{|z|=5} F(z,u)\,\frac{dz}{z^{n+1}} \\ &= \frac{1}{u\rho(u)e^{\rho(u)}}\,\rho(u)^{-n} + \mathcal{O}(5^{-n}). \end{aligned} \tag{9.32}$$

It is important to note that the error term is *uniform* in $u$, for $u$ close enough to 1, here $|u-1| \le 0.1$. This fact derives from the coefficient extraction method, since, in the remainder Cauchy integral of (9.32), the denominator of $F(z,u)$ stays bounded away from 0.

The second estimate in Equation (9.32), constitutes a typical case of application of the quasi-power schema. Thus, the number $X_n$ of image points in a random surjection of size $n$ obeys in the limit a Gaussian law. The local expansion of $\rho(u)$,

$$\rho(u) \equiv \log(1+u^{-1}) = \log 2 - \frac{1}{2}(u-1) + \frac{3}{8}(u-1)^2 + \cdots,$$

yields

$$\frac{\rho(1)}{\rho(u)} = 1 + \frac{1}{2\log 2}(u-1) - \frac{3\ln(2)-2}{8(\log 2)^2}(u-1)^2 + \mathcal{O}\left((u-1)^3\right),$$

so that the mean and standard deviation satisfy

$$\mu_n \sim C_1 n, \qquad \sigma_n \sim \sqrt{C_2\, n}.$$

There, the constants $C_1, C_2$ are determined by the three-term expansion of $\rho(u)$,

$$C_1 = \frac{1}{2\log 2}, \qquad C_2 = \frac{1-\log 2}{4(\log 2)^2},$$

and finally [2],

$$\Pr\{X_n \le C_1 n + x\sqrt{C_2 n}\} = \Phi(x) + O\left(\frac{1}{\sqrt{n}}\right).$$

This result can be viewed alternatively as a purely asymptotic statement on Stirling partition numbers: *uniformly for all real $x$,*

$$\frac{\displaystyle\sum_{k\le C_1 n + x\sqrt{C_2 n}} k!\left\{ n \atop k \right\}}{\displaystyle\sum_k k!\left\{ n \atop k \right\}} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-w^2/2}\,dw + O\left(\frac{1}{\sqrt{n}}\right).$$

**General meromorphic conditions.** The following schema vastly generalizes the case of surjections. It is inspired by the works of Bender [2].

**Theorem 9.7 (Meromorphic schema)** *Let $F(z,u)$ be a bivariate function that is analytic in a domain*

$$\mathcal{D}_0 = \left\{(z,u) \;\middle|\; |z| < \rho, |u| < 1\right\},$$

*and has nonnegative coefficients at $(0,0)$. Assume that there exists $\epsilon > 0$ and $r > \rho$ such that in the domain,*

$$\mathcal{D} = \left\{(z,u) \;\middle|\; |z| \le r, |u-1| < \epsilon\right\},$$

*the function $F(z,u)$ admits the representation*

$$F(z,u) = \frac{B(z,u)}{C(z,u)},$$

*where $B(z,u), C(z,u)$ are analytic for $(z,u) \in \mathcal{D}$. Assume also that the equation*

$$C(\zeta, 1) = 0 \tag{9.33}$$

*has only one (simple) root $\zeta = \rho$ in $|z| \leq r$ and that $B(\rho, 1) \neq 0$. Assume finally the "variability condition",*

$$0 < \liminf \frac{\sigma_n^2}{n}.$$

*Then, the variable that has probability generating function*

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

*converges in distribution to a Gaussian variable with a speed of convergence that is $\mathcal{O}(n^{-1/2})$. The mean $\mu_n$ and the standard deviation $\sigma_n$ are symptotically linear in $n$.*

Under these conditions (see the proof below), there exists a unique root $\rho(u)$ of the implicit equation

$$C(\rho(u), u) = 0$$

that is analytic at $u = 1$ and such that $\rho(1) = \rho$. Then, the mean $\mu_n$ and variance $\sigma_n^2$ satisfy

$$\mu_n = \text{Mean}\left(\frac{\rho(1)}{\rho(u)}\right) n + \mathcal{O}(1), \qquad \sigma_n^2 = \text{Var}\left(\frac{\rho(1)}{\rho(u)}\right) n + \mathcal{O}(1), \tag{9.34}$$

and the variability condition is satisfied as soon as

$$\text{Var}\left(\frac{\rho(1)}{\rho(u)}\right) > 0. \tag{9.35}$$

The quantities in (9.34–9.35) can themselves be expressed in terms of partial derivatives of $C(z,u)$ alone, by series reversion,

$$\rho(u) = \rho - \frac{c_{0,1}}{c_{1,0}}(u-1) - \frac{c_{1,0}^2 c_{0,2} - 2c_{1,0}c_{1,1}c_{0,1} + c_{2,0}c_{0,1}^2}{2c_{1,0}^3}(u-1)^2 + \mathcal{O}((u-1)^3), \tag{9.36}$$

where

$$c_{i,j} := \left.\frac{\partial^{i+j}}{\partial z^i \partial u^j} C(z,u)\right|_{(\rho,1)}.$$

In particular, the coefficients in the mean and variance are entirely determined by (9.36, hence computable from $C(z,u)$.

Proof. Consider a domain $|u-1| \le \delta$ inside the region of analyticity of $B, C$. Then, one has

$$f_n(u) := [z^n]F(z,u) = \frac{1}{2i\pi} \oint F(z,u) \frac{dz}{z^{n+1}},$$

where the integral is taken along a small enough contour encircling the origin. We use the analysis of polar singularities described in Chapter 4, like in (9.32). As $\rho$ exists and is simple, it can be lifted to a locally analytic function $\rho(u)$, for $u$ close enough to 1, say $|u-1| \le \delta$. (This is a consequence of the implicit function theorem, or alternatively of the Weierstrass preparation theorem.) As $F(z,u)$ has at most one (simple) pole in $|z| \le r$, we have

$$f_n(u) = \operatorname{Res}\left(\frac{B(z,u)}{C(z,u)} z^{-n-1}\right)_{z=\rho(u)} + \frac{1}{2i\pi} \int_{|z|=r} F(z,u) \frac{dz}{z^{n+1}}, \qquad (9.37)$$

where we may suitably restrict $\delta$ so that $|r-\rho(u)| < \frac{1}{2}(r-\rho)$.

The modulus of the second term in (9.37) is bounded from above by

$$\frac{K}{r^n} \qquad \text{where} \qquad K = \frac{\sup_{|z|=r, |u-1| \le \delta} |B(z,u)|}{\inf_{|z|=r, |u-1| \le \delta} |C(z,u)|}. \qquad (9.38)$$

Since the domain $|z| = r, |u-1| \le \delta$ is closed, $C(z,u)$ attains its minimum that must be nonzero, given the unicity of the zero of $C$. At the same time, $B(z,u)$ being analytic, its modulus is bounded from above. Thus, the constant $K$ in (9.38) is finite.

A residue computation of the first term, in accordance with the analysis of meromorphic functions, then yields

$$f_n(u) = \frac{B(\rho(u), u)}{C'(\rho(u), u)} \rho(u)^{-n-1} + \mathcal{O}(r^{-n}),$$

uniformly for $u$ in a small enough fixed neighbourhood of 1. The mean and variance then satisfy (9.34), with the coefficient in the leading term of the variance term that is, by assumption, nonzero. Thus, the conditions of the Quasi-Powers Theorem in the form (9.27) are satisfied, and the law is Gaussian in the asymptotic limit. □

Some form of condition regarding variability and simplicity of the dominant pole is a necessity. For instance, the functions

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

each fail to satisfy the variability condition since the variance is identically 0. The variance is $\mathcal{O}(1)$ for a related function like

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

where a discrete limit law that is geometric holds. Yet another situation arises when considering

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

There is a double pole at 1 when $u = 1$ that arises from "confluence" at $u = 1$ of two analytic branches $\rho_1(u) = 1$ and $\rho_2(u) = 1/u$. In this particular case, the limit law is continuous but non-Gaussian; in fact, the limit is the uniform distribution over the interval $[0, 1]$, since

$$F(z,u) = 1 + z(1+u) + z^2(1+u+u^2) + z^3(1+u+u^2+u^3) + \cdots.$$

In addition, for this case, the mean is $\mathcal{O}(n)$ but the variance is $\mathcal{O}(n^2)$.

EXERCISE 18. Extend the analysis to the case of a pole of order $m+1$. See [2].

EXAMPLE 12. *Binomial coefficients and the Central Limit Theorem.* The function

$$F(z,u) = \frac{1}{1 - z(1+u)} = \sum_{n,k} \binom{n}{k} u^k z^n$$

is the BGF of binomial coefficients and also of binary $\{0,1\}$-strings with $u$ marking the number of 0's. We have $\rho(1) = \frac{1}{2}$ and generally

$$\rho(u) = (1+u)^{-1}.$$

Thus, the limit Gaussian law results from Theorem 9.6,

$$\sum_{j \le k} \frac{1}{2^n} \binom{n}{k} = \Phi(x) + \mathcal{O}\left(\frac{1}{\sqrt{n}}\right),$$

uniformly with respect to $k$.

For the very same reasons, if $g(u)$ is any nondegenerate PGF ($\mathrm{Var}(g) \neq 0$) that is analytic at 1, then

$$F(z,u) = \frac{1}{1 - zg(u)}$$

has a singularity at $1/g(u)$ that is a simple pole,

$$\rho(u) = \frac{1}{g(u)}.$$

We thus rederive in this way a version of the central limit theorem for discrete probability distributions with PGFs that are analytic at 1. In such a case, a refined Gaussian convergence property (a local limit law, see Section 9.7 below) also derives from the saddle point method. □

EXAMPLE 13. *Runs in permutations and Eulerian numbers.* The exponential BGF of Eulerian numbers (that count runs in permutations) is

$$F(z,u) = \frac{u(1-u)}{e^{(u-1)z} - u},$$

where, for $u = 1$, we have $F(z,1) = (1-z)^{-1}$. The roots of the denominator are then

$$\rho_k(u) := \rho(u) + \frac{2ik\pi}{u-1}, \qquad \rho(u) = \frac{\log u}{u-1},$$

where $k$ is an arbitrary element of $Z$. As $u$ is close to 1, $\rho(u)$ is close to 1, while the other poles $\rho_k(u)$ with $k \neq 0$ escape to infinity. This fact is also consistent with the limit form $F(z,1) = (1-z)^{-1}$ which has only one pole at 1. If one restricts $u$ to $|u| \leq 2$, there is clearly at most one root of the denominator in $|z| \leq 2$ that is given by $\rho(u)$. Thus, we have for $u$ close enough to 1,

$$F(z,u) = \frac{1}{\rho(u) - z} + \mathcal{O}(2^{-n}),$$

and

$$[z^n]F(z,u) = \rho(u)^{-n-1} + \mathcal{O}(2^{-n}).$$

The variability conditions are satisfied since

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

Therefore, the Eulerian distribution is asymptotically Gaussian. The mean and variance are given by

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

This example is a famous one that has been known for a long time [11] and our derivation here follows Bender's seminal paper [2]. There are also

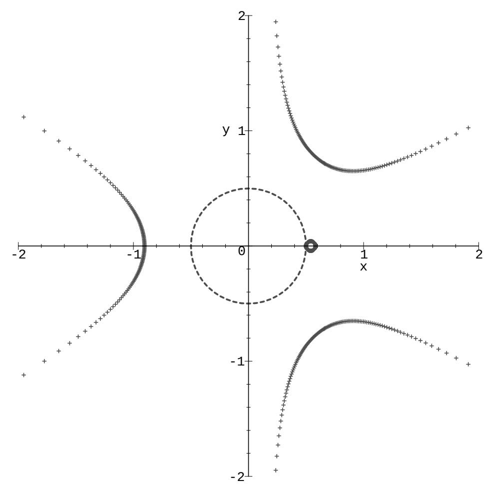


Figure 9.8: The diagrams of poles of the BGF $F(z,u)$ associated to the pattern *abaa* with correlation polynomial $c(z) = 1 + z^3$ when $u$ varies on the unit circle. The denominator is of degree 4 in $z$: one branch, $\rho(u)$ clusters near the dominant singularity $\rho = \frac{1}{2}$ of $F(z,1)$ while three other singularities stay away from the disc $|z| \le \frac{1}{2}$ and escape to infinity as $u \to 1$.

interesting connections with probability theory: if $U_j$ are independent random variables that are uniformly distributed over the interval $[0,1]$, then one has

$$[z^n u^k]F(z,u) = \Pr\{\lfloor U_1 + \cdots + U_n \rfloor < k\}.$$

Because of this fact, the normal limit is thus often derived a consequence of the central limit theorem of probability theory, after one takes care of unimportant details relative to the integer part $\lfloor \cdot \rfloor$ function; see [11, 52]. □

EXAMPLE 14. *Patterns in strings.* Consider the class $\mathcal{F}$ of binary strings (the "texts"), and fix a "pattern" $w$ of length $k$. Let $\chi$ be the number of (possibly overlapping) occurrences of $w$. (The pattern $w$ occurs if it is a factor, *i.e.*, if its letters occur contiguously in the text.) Let $F(z,u)$ be the BGF relative to the pair $(\mathcal{F},\chi)$. The Guibas-Odlyzko correlation polynomial[3] $c(z) \equiv c_w(z)$ relative to $w$ is defined for instance in [54], where

[3] The correlation polynomial has coefficients in $\{0,1\}$, with $[z^j]c(z) = 1$ iff $w$ matches its left shift by $j$ positions.

it is shown that the OGF of words with pattern $w$ excluded is

$$F(z,0) = \frac{c(z)}{z^k + (1-2z)c(z)}.$$

By similar string decompositions, the full BGF is found to be [22, p. 145]

$$F(z,u) = \frac{1-(c(z)-1)(u-1)}{1-2z-(u-1)(z^k+(1-2z)(c(z)-1))}.$$

Let $D(z,u)$ be the denominator. Then $D(z,u)$ depends analytically on $z$, for $u$ near 1 and $z$ near $1/2$. In addition, the partial derivative $D'_z(\frac{1}{2},1)$ is nonzero. Thus, $\rho(u)$ is analytic at $u=1$, with $\rho(1)=1/2$. The local expansion of the root $\rho(u)$ of $D(\rho(u),u)$ follows from local series reversion,

$$2\rho(u) = (1-2^{-k}(u-1) + (k2^{-2k} - 2^{-k}c(\frac{1}{2}))\,(u-1)^2 + \mathcal{O}((u-1)^3).$$

Theorem 9.7 applies. Hence: *the number of occurrences of a fixed pattern in a random large string is asymptotically normal.* The number of occurrences has a mean

$$\mu_n = \frac{n}{2^k} + \mathcal{O}(1)$$

that does not depend on the pattern but just on its length (a known fact, see [54]) and the standard deviation is $\mathcal{O}(\sqrt{n})$, with the coefficient determined by the three-term expansion of $\rho(u)$,

$$\sigma_n^2 = \left(2^{-k}(1+2c(\frac{1}{2})) + 2^{-2k}(1-2k)\right) n + \mathcal{O}(1).$$

□

Exercise 19. Show that asymptotic normality also holds when letters in strings are chosen independently but with an arbitrary probability distribution.

Example 15. *The discrete renewal problem.* How many times must one throw a die in order to reach a total of $n$? Consider the BGF

$$F(z,u) = \frac{1}{1-ug(z)}, \qquad g(z) = \frac{z(1-z^6)}{6(1-z)}.$$

Then, $[z^n]F(z,1)$ is the probability of reaching $n$ in some unspecified number of throws. This probability converges fast to $\alpha = \frac{2}{7} \doteq 0.28571$. (Note that each throw makes "progress" by an amount of $\frac{7}{2}$ on average.) The coefficient $[z^n u^k]F(z,u)$ is then the joint probability that $n$ is reached and the number of throws is $k$. Here, $g(z)$ being a polynomial, it is analytic and also invertible at 1, since $g'(1) \neq 0$. Thus $\rho(u)$ defined by $g(\rho(u)) = u^{-1}$ exists and is analytic in a complex neighbourhood of 1,

$$\rho(u) = 1 - \frac{u-1}{g'(1)} - \frac{g''(1) - 2g'(1)^2}{2g'(1)^3}(u-1)^3 + \cdots .$$

Theorem 9.7 applies and the distribution of the number of throws (conditioned upon the fact that $n$ is attained) is Gaussian in the asymptotic limit.

The argument works for an arbitrary PGF $g$ that is analytic at 1, and the number of "throws" (conditioned upon the event that the "target" $n$ is attained) is again Gaussian with mean and variance $\mu_n, \sigma_n^2$ that satisfy

$$\mu_n \sim \frac{n}{g'(1)}, \qquad \sigma_n^2 \sim n\, \frac{g''(1) + g'(1) - g'(1)^2}{g'(1)^3}.$$

We thus obtain a version of the *renewal theorem* [8] under two assumptions: $(i)$ the basic random variable is discrete; $(ii)$ its PGF is analytic at 1. This example resembles integer compositions, with however the additional normalization, $g(1) = 1$. □

**Corollary 9.1 (Supercritical sequence schema)** *Let $F(z,u)$ be a bivariate generating function of the form*

$$F(z,u) = \frac{1}{1 - ug(z)},$$

*where $g(z)$ is a function analytic at the origin, satisfying $g(0) = 0$ and $g(z)$ aperiodic. Assume that there exists a value $\rho > 0$ strictly in the interior of the disc of convergence of $g(z)$ such that $g(\rho) = 1$. Then, the distribution associated to $F(z,u)$ converges to a Gaussian limit.*

PROOF. A direct application of Theorem 9.7 with $B(z,u) = 1$, $C(z,u) = 1 - ug(z)$, and uniqueness of the dominant singularity granted by the aperiodicity assumption. '□

EXAMPLE 16. *Alignments and Stirling cycle numbers.* Recall that alignments are sequences of cycles. In that case, we have from Chapters 2 and 3 the exponential BGF,

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

The function $\rho(u)$ is explicit,

$$\rho(u) = 1 - e^{-1/u},$$

and in particular $\rho(1) = 1 - e^{-1} \doteq 0.63$ is the radius of convergence of the EGF $F(z,1)$. It is easy to check that in $|z| \le r = \frac{9}{10}$, there is at most one pole of the denominator of $F(z,u)$. The distribution of the number of cycle components in a random alignment is thus asymptotic normal.

Again, this can be viewed as a purely asymptotic statement on Stirling cycle numbers: *uniformly for all real $x$*,

$$\frac{\displaystyle\sum_{k \le C_1 n + x\sqrt{C_2 n}} k! \begin{bmatrix} n \\ k \end{bmatrix}}{\displaystyle\sum_k k! \begin{bmatrix} n \\ k \end{bmatrix}} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-w^2/2}\, dw + O\left(\frac{1}{\sqrt{n}}\right),$$

where the two constants $C_1, C_2$ are

$$C_1 = \frac{1}{e-1}, \qquad C_2 = \frac{1}{(e-1)^2}.$$

□

EXAMPLE 17. *Integer compositions.* Consider integer compositions where the summands are constrained to belong to a set $\Gamma \subseteq N^+$, and let $X_n$ be the number of summands in a random composition of integer $n$. The ordinary BGF is

$$F(z,u) = \frac{1}{1 - ug(z)}, \qquad g(z) = \sum_{\gamma \in \Gamma} z^\gamma.$$

Assume that $\Gamma$ contains at least two relatively prime elements, so that $g(z)$ is aperiodic. The radius of convergence of $G(z)$ can only be $\infty$ ($g(z)$ is a polynomial) or 1 ($g(z)$ comprises infinitely many terms but is dominated by $(1-z)^{-1}$). Then, there is a unique value $\rho < 1$ such that $g(\rho) = 1$ since $g(z)$ increases starting from 0 and $1 < g(1) \le \infty$. Also, by the implicit function theorem, there exists a solution to the equation $g(\rho(u)) = u^{-1}$, for $u$ sufficiently close to 1. Thus, the conditions of the corollary are satisfied and the distribution of $X_n$ is asymptotically normal.

For instance, the case $g(z) = z + z^2$, corresponds to Fibonacci coverings —*i.e.*, coverings of an interval by sticks of lengths $1, 2$— for which asymptotic normality holds. The case $g(z) = z/(1-z)$ corresponds to the class of unrestricted integer compositions. In that case, one has

$$\rho(u) = u(1+u)^{-1},$$

and

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

is asymptotically normal, with

$$\mu_n \sim \frac{n}{2}, \qquad \sigma_n \sim \frac{\sqrt{n}}{2}.$$

(Of course, this last fact results also directly from the Gaussian law of binomial coefficients.) Finally, a Gaussian limit holds for compositions into prime or twin-prime summands, even though the information available on these sets and their GF is somewhat partial. □

**Polyominos.** Polyominos are plane diagrams that are closely related to models of statistical physics, while having been the subject of a vast combinatorial literature. This example has the merit of illustrating a level of difficulty somewhat higher than in previous examples and typical of many "real–life" applications. Our presentation follows [3] and a recent paper of Louchard [46]. We consider here the variety of polyominos called *parallelograms*. A parallelogram is a sequence of segments,

$$[a_1, b_1],\ [a_2, b_2],\ \ldots, [a_m, b_m], \qquad a_1 \le a_2 \cdots \le a_m,\ b_1 \le b_2 \le \cdots \le b_m,$$

where the $a_j$ and $b_j$ are integers with $b_j - a_j \ge 1$, and one takes $a_1 = 0$ for definiteness. A parallelogram can thus be viewed as a stack of segments (with $[a_{j+1}, b_{j+1}]$ placed on top of $[a_j, b_j]$) that leans smoothly to the right.

The quantity $m$ is called the height, the quantity $b_m - a_1$ the width, their sum is called the (semi)perimeter, and the grand total $\sum_j (b_j - a_j)$ is called the area. We examine parallelograms of fixed area and investigate the distribution of the perimeter. The answer involves the two functions,

$$J_0(q,u) := \sum_{n\ge 0} \frac{(-1)^n u^n q^{n(n+1)/2}}{(q;q)_n (uq;q)_n}, \qquad J_1(q,u) := \sum_{n\ge 1} \frac{(-1)^{n-1} u^n q^{n(n+1)/2}}{(q;q)_{n-1} (uq;q)_n},$$

where the "$q$–factorial" notation is used,

$$(a;q)_n = (1-a)(1-aq)\cdots(1-aq^{n-1}).$$

Such functions belong to the realm of "$q$–analogues" and $J_0, J_1$ generalize the classical Bessel functions to which they reduce, after normalization, when $q \to 1$.

The bivariate OGF of parallelograms, with $z$ marking area and $u$ marking perimeter is

$$F(z,u) = u\frac{J_1(z,u)}{J_0(z,u)}. \tag{9.39}$$

This GF results from a simple construction: a parallelogram is either an interval, or it is derived from an existing parallelogram by stacking on top a new interval. Let $G(w) \equiv G(x,y,z,w)$ be the OGF with $x,y,z,w$ marking width, height, area, and length of top segment, respectively. The GF of a parallelogram made of a single nonzero interval is

$$a(w) \equiv a(x,y,z,w) = \frac{xyzw}{1-xzw}.$$

The operation of piling up a new segment on top of a segment of length $m$ that is represented by a term $w^m$ is described by

$$y\left(\frac{z^m w^m}{1-xzw} + \cdots + \frac{zw}{1-xzw}\right) = xyzw\frac{1-x^m w^m}{(1-zw)(1-xzw)}.$$

Thus, $G$ satisfies the functional equation,

$$G(w) = \frac{xyzw}{1-xzw} + \frac{xyzw}{(1-zw)(1-xzw)}\left[G(1) - G(xzw)\right]. \tag{9.40}$$

The method of constructing parallelograms by "adding a slice" is thus reflected by the relation (9.40). Now, an equation of the form,

$$G(w) = a(w) + b(w)[G(1) - G(\lambda w)],$$

is solved by iteration:

$$\begin{aligned} G(w) &= a(w) + b(w)G(1) - b(w)G(\lambda(w)) \\ &= \left(a(w) - b(w)a(\lambda w) + b(w)b(\lambda w)a(\lambda^2 w) - \cdots\right) \\ &\qquad + G(1)\left(b(w) - b(w)b(\lambda w) + b(w)b(\lambda w)b(\lambda^2 w) - \cdots\right). \end{aligned}$$

One then isolates $G(1)$ by setting $w = 1$. This expresses $G(1)$ as the quotient of two similar looking series (formed with sums of products of $b$–values). Here, this gives $G(x,y,z,1)$, from which the form (9.39) of $F(z,u)$ derives, since $F(z,u) = G(u,u,z,1)$.

In such a seemingly difficult situation, one should first estimate $[z^n]F(z,1)$, the number of parallelogram of "size" (*i.e.*, area) equal to $n$. We have $F(z,1) = J_1(z,1)/J_0(z,1)$, where the denominator is

$$J_0(z,1) = 1 - \frac{z}{(1-z)^2} + \frac{z^3}{(1-z)^2(1-z^2)^2} - \frac{z^6}{(1-z)^2(1-z^2)^2(1-z^3)^2} + \cdots.$$

Clearly, $J_0(z,1)$ and $J_1(z,1)$ are analytic in $|z| < 1$, and it is not hard to see that $J_0(z,1)$ decreases from 1 to about $-0.24$ when $z$ varies between 0 and $\frac{1}{2}$, with a root at

$$\rho \doteq 0.43306\,19231\,29252,$$

where $J_0'(\rho,1) \doteq -3.76 \neq 0$, so that the zero is simple[4]. Since $F(z,1)$ is by construction meromorphic in the unit disc, and $J_1(\rho,1) \doteq 0.48 \neq 0$, the number of parallelograms satisfies

$$[z^n]F(z,1) \sim \frac{J_1(\rho,1)}{\rho J_0'(\rho,1)} \left(\frac{1}{\rho}\right)^n = \alpha_1 \cdot \alpha_2^n,$$

where

$$\alpha_1 \doteq 0.29745\,35058\,07786, \qquad \alpha_2 \doteq 2.30913\,85933\,31230.$$

As is common in meromorphic analyses, the approximation of coefficients is quite good; for instance, the relative error is only about $10^{-8}$ for $n = 35$.

We are now ready for bivariate asymptotics. Take $|z| \leq r = \frac{7}{10}$ and $|u| \leq \frac{11}{10}$. Because of the form of their general terms that involve $z^{n^2/2}u^n$ in the numerators while the denominators stay bounded away from 0, the functions $J_0(z,u)$ and $J_1(z,u)$ remain analytic there. Thus, $\rho(u)$ exists and is analytic for $u$ in a sufficiently small neighbourhood of 1 (by Weierstrass preparation or implicit functions). The nondegeneracy conditions are easily verified by numerical computations. There results that Theorem 9.7 applies.

> *The perimeter of a random parallelogram of area $n$ admits a limit law that is Gaussian with mean and variance that are each of order $\mathcal{O}(n)$.*

The constants are then determined by Formulæ (9.36), and it is found (computer algebra helps!) that $\mu_n \sim \mu n$, $\sigma_n \sim \sigma\sqrt{n}$, with

$$\mu \doteq 0.84176\,20156, \qquad \sigma \doteq 0.42420\,65326.$$

Similar methods show that the expected height and width are each $\mathcal{O}(n)$ on average, again with Gaussian limits. This indicates that a random parallelogram is most likely to resemble a slanted stack of fairly short segments.

> EXERCISE 20. Analyse precisely the distribution of height and width of a random parallelogram of area $n$.

---

[4] As usual, such computations can be easily validated by carefully controlled numerical evaluations coupled with Rouché's theorem (see Chapter 4).

EXERCISE 21. Define a coin fountain as a vector $v = (v_0, v_1, \ldots, v_\ell)$, such that $v_0 = 0$, $v_j \geq 0$ is an integer, $v_\ell = 0$ and $|v_{j+1} - v_j| = 1$. Take as size characteristic the quantity $n = \sum v_j$. Then the distribution of the length $\ell$ in a random coin fountain of size $n$ is asymptotically normal. (This amounts to considering all ruin sequences of a fixed area as equally likely, and considering the number of steps in the game as a random variable.)

Show similarly that the number of vector entries equal to 0 is asymptotically Gaussian.

### 9.5.2 The algebraic–logarithmic schema

In this subsection largely based on [29], we examine a scheme that arises when generating functions contain algebraic of the form $C(z,u)^{-\alpha}$, where $\alpha$ may be an arbitrary real number. For instance, trees often lead to singularities of the square-root type and this singular behaviour persists for a number of bivariate generating functions associated to parameters that are sufficiently decomposable. In such cases, one should appeal to the method of singularity analysis as detailed in Chapter 5.

An especially nice feature of the method of singularity analysis and of the associated Hankel contours is the fact that it preserves uniformity of expansions; for instance, if in a $\Delta$-domain indented at 1,

$$|f(z)| \leq K \cdot |1 - z|^{-\alpha},$$

then, for some explicitly known absolute constant $J(\alpha)$, one has

$$|[z^n]f(z)| \leq K \cdot J(\alpha) n^{\alpha - 1}. \tag{9.41}$$

It is even the case that $J(\alpha)$ is bounded by an absolute constant when $\alpha$ is restricted to a bounded region of the complex plane that avoids the set $\{0, -1, -2, \ldots\}$. This feature is crucial in translating bivariate expansion, where we need to estimate *uniformly* a coefficient $f_n(u) = [z^n]F(z,u)$ that depends on the parameter $u$, given some (uniform) knowledge on the singular structure of $F(z,u)$ in terms of $z$. The method is in the sense more general than the meromorphic scheme; only the error terms in estimates of PGFs tend to be naturally less good as we replace an exponentially small error term inherent to meromorphic functions by a term that is $\mathcal{O}(n^{-1})$ in the context of singularity analysis.

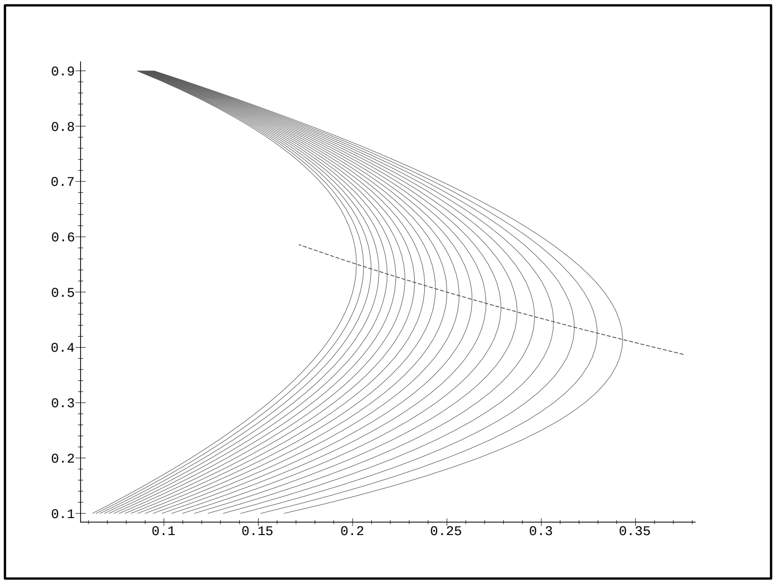


Figure 9.9: A display of the family of GF's $F(z, u_0)$ corresponding to leaves in general Catalan trees when $u_0 \in [\frac{1}{2}, \frac{3}{2}]$. It is seen that the singularities are all of the square root type (dashed line), with a movable singularity at $\widetilde{\rho}(u) = (1 + u^{1/2})^{-2}$.

**Leaves in general Catalan trees.** As an introductory example, let us briefly revisit the distribution of the number of leaves in general Catalan trees, see (9.10). Explicit expressions are otherwise known for the coefficients as products of two binomial coefficients, see Chapter 3. The computations are a little simpler if we adopt as BGF

$$G(z,u) = F(z,u^2) = \frac{1}{2}\left(1 + (u^2-1)z - \sqrt{1-2(u^2+1)z+(u^2-1)^2z^2}\right).$$

In other words, we are taking a BGF for a parameter equal to twice the number of leaves. In this case, the discriminant factors nicely:

$$1-2(u^2+1)z+(u^2-1)^2z^2 = (1-z(1+u)^2)(1-z(1-u)^2).$$

We thus have

$$G(z,u) = A(z,u) + B(z,u)\sqrt{C(z,u)}, \tag{9.42}$$

with

$$A(z,u) = \frac{1}{2}(1+(u^2-1)z), \quad B(z,u) = -\frac{1}{2}\sqrt{1-z(1-u)^2},$$

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

This decomposition clearly shows that, when $u$ is close enough to 1, the function $G(z,u)$ has a dominant singularity of the square-root type at

$$\rho(u) = \frac{1}{(1+u)^2}.$$

For fixed $u$, we have for the BGF, by (9.42),

$$G(z,u) = a_0(u)+b(u)\sqrt{1-z/\rho(u)}+a_1(u)(1-z/\rho(u))+O((1-z/\rho(u))^{3/2}), \tag{9.43}$$

for some functions $a_0, a_1, b, c$ that are analytic at $u=1$. Hence, by singularity analysis,

$$[z^n]G(z,u) = \frac{-2}{\sqrt{\pi}}\, B(\rho(u),u)\,\rho(u)^{-n} n^{-3/2}\left(1+O(\frac{1}{n})\right). \tag{9.44}$$

Since the expansion (9.43) is uniform when $u$ lies in a sufficiently small complex neighbourhood of 1, the expansion of the coefficient in (9.44) is also uniform by virtue of the general remark concerning singularity analysis. We are thus exactly in a case of application of the Quasi-Powers Theorem and the limit law for the number of leaves is asymptotically Gaussian.

**General algebraic-logarithmic conditions.** The example of leaves leads to simple computations. At the same times it already shows all the machinery needed for the general case.

**Theorem 9.8 (Algebraic–logarithmic schema)** *Let $F(z,u)$ be a bivariate function that is analytic in a domain*

$$\mathcal{D}_0 = \left\{(z,u) \;\middle|\; |z|<\rho, |u|<1\right\},$$

*and has nonnegative coefficients at $(0,0)$. Assume that there exists $\epsilon>0$, $\vartheta<\frac{\pi}{2}$, and $r>\rho$ such that in the domain,*

$$\mathcal{D} = \left\{(z,u) \;\middle|\; |z|\le r,\ Arg(z-\rho)\in[\vartheta, 2\pi-\vartheta],\ |u-1|<\epsilon\right\},$$

*the function $F(z,u)$ admits the representation*

$$F(z,u) = A(z,u) + B(z,u)C(z,u)^{-\alpha}\left(\log C(z,u)\right)^k, \tag{9.45}$$

*where $A(z,u), B(z,u), C(z,u)$ are analytic for $(z,u)\in\mathcal{D}$, $k$ is a nonnegative integer, and $\alpha\notin\{0,-1,-2,\ldots\}$. Assume also that the equation*

$$C(\zeta,1) = 0 \tag{9.46}$$

*has only one (simple) root $\zeta = \rho$ in $|z| \le r$ and that $B(\rho, 1) \neq 0$. Assume finally the "variability condition",*

$$0 < \liminf \frac{\sigma_n^2}{n}.$$

*Then, the variable with probability generating function*

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

*converges in distribution to a Gaussian variable with a speed of convergence that is $\mathcal{O}(n^{-1/2})$. The mean $\mu_n$ and the standard deviation $\sigma_n$ are symptotically linear in $n$.*

The remarks following the statement of Theorem 9.7 apply. There exists a unique root $\rho(u)$ of the implicit equation

$$C(\rho(u), u) = 0$$

that is locally analytic at 1 and such that $\rho(1) = \rho$. Then, the mean $\mu_n$ and variance $\sigma_n^2$ satisfy

$$\mu_n = \text{Mean}\left(\frac{\rho(1)}{\rho(u)}\right) n + \mathcal{O}(1), \qquad \sigma_n^2 = \text{Var}\left(\frac{\rho(1)}{\rho(u)}\right) n + \mathcal{O}(1),$$

and the variability condition is satisfied as soon as

$$\text{Var}\left(\frac{\rho(1)}{\rho(u)}\right) > 0. \tag{9.47}$$

The coefficients in these expansions are then expressible in terms of the derivatives of $C(z, u)$ at $(\rho, 1)$, as summarized in (9.36).

Proof. To simplify notations, assume first that $k = 0$. The function $F(z, 1)$ is by assumption analytic at the origin with nonnegative Taylor coefficients. Therefore, it has a dominant positive singularity by Pringsheim's theorem; this dominant singularity is unique and coincides with $\rho > 0$. By assumption (9.45), $F(z, 1)$ admits a singular expansion of the form

$$\begin{aligned} F(z,1) &= (a_0 + a_1(z-\rho) + \cdots) \\ &\quad + (b_0 + b_1(z-\rho) + \cdots)\left(c_1(z-\rho) + c_2(z-\rho)^2 + \cdots\right)^{-\alpha}. \end{aligned} \tag{9.48}$$

There, the $a_j, b_j, c_j$ represent the coefficients of the expansion in $z$ of $A, B, C$ for $z$ near $\rho$ when $u$ is instantiated at 1. We may consider $C(z, u)$ normalized by the condition that $-c_1\rho$ is positive real, so that $b_0$ is real. By singularity analysis, we thus derive the estimate

$$[z^n]F(z,1) = b_0(-c_1\rho)^{-\alpha}\rho^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+\mathcal{O}(\frac{1}{n})\right). \qquad (9.49)$$

All that is needed now is a "lifting" of relations (9.48,9.49), for $u$ in a small neighbourhood of 1. We may restrict this neighbourhood as we please, with $|u-1| \le \delta$ provided we keep $\epsilon \ge \delta > 0$. First, by Weierstrass preparation, there is for $u$ sufficiently near to 1, a unique simple root $\rho(u)$ near $\rho$ of the equation

$$C(\rho(u), u) = 0.$$

We have $\rho(1) = \rho$ with $\rho(u)$ being locally analytic at 1. One can then expand $A, B, C$ near $(\rho(u), u)$. This gives the bivariate expansion

$$\begin{aligned} F(z,u) &= (a_0(u) + a_1(z-\rho(u)) + \cdots) \\ &+ (b_0(u) + b_1(u)(z-\rho(u)) + \cdots)\left(c_1(z-\rho(u)) + c_2(u)(z-\rho(u))^2 + \cdots\right)^{-\alpha}. \end{aligned} \qquad (9.50)$$

There, by assumption, we have that $a_j(u), b_j(u), c_j(u)$ are analytic in $|u-1| \le \epsilon$, and are each $\mathcal{O}(r^{-n})$. In addition, $\rho(u)^\alpha$ and $(-c_1(u))^\alpha$ are well-defined by principal values, since their specializations at $u=1$ are positive. Thus, we have a singular expansion for $F(z,u)$; for instance, when $\alpha \in ]-1,0[$,

$$\begin{aligned} F(z,u) &= a_0(u) + a_1(u)(z-\rho(u)) \\ &\quad + b_0(u)(-c_1(u)\rho(u))^{-\alpha}(1-z/\rho(u))^{-\alpha} + R(z), \end{aligned} \qquad (9.51)$$

where

$$R(z) = \mathcal{O}\left((1-z/\rho(u))^{\alpha+1}\right),$$

and the $\mathcal{O}$–error term is uniform for $|u-1| < \delta$:

$$|R(z)| \le K \cdot |1 - z/\rho(u)|,$$

for some absolute constant $K$. We thus have

$$[z^n]F(z,u) = b_0(u)(-c_1(u)\rho(u))^{-\alpha}\rho(u)^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}\left(1+\mathcal{O}(\frac{1}{n})\right). \qquad (9.52)$$

where the error term is again uniform. An especially important fact for this argument is the following: *the singularity analysis analysis method is a uniform*[5] *coefficient extraction method.*

[5] *For instance, Darboux's method is based on the Riemann-Lebesgue lemma that only provides* non-uniform *error terms and it cannot be employed for such bivariate asymptotics.*

Equation (9.52) shows that $f_n(u) = [z^n]F(z,u)$ satisfies precisely the conditions of the Quasi-Powers Theorem. Therefore, the corresponding law with PGF $f_n(u)/f_n(1)$ is asymptotically normal with a mean and a standard deviation that are both $\mathcal{O}(n)$. Since the error term in (9.52) is $\mathcal{O}(1/n)$, the speed of convergence to the Gaussian limit is $\mathcal{O}(1/\sqrt{n})$.

If the bivariate GF $F(z,u)$ is logarithmic, then an almost identical argument applies. For instance, one has, by singularity analysis,

$$[z^n]F(z,1) = b_0(-c_1\rho)^{-\alpha}\rho^{-n}\frac{n^{\alpha-1}}{\Gamma(\alpha)}(\log n)^k\left(1+\mathcal{O}(\frac{1}{\log n})\right).$$

This estimates "lifts" to complex values of $u$ near 1. As a consequence, a Gaussian law holds, with a speed of convergence that is $\mathcal{O}((\log n)^{-1/2})$. The estimate of the speed of convergence can be improved to $O(n^{-1/2})$, provided one uses the modified expansion

$$\begin{aligned} [z^n]F(z,1) &= b_0(-c_1\rho)^{-\alpha}\rho^{-n}h_n^{(\alpha,k)}\left(1+\mathcal{O}(\frac{1}{n})\right), \\ h_n^{(\alpha,k)} &= [z^n](1-z)^{-\alpha}\left(\log(-c_1(1-z))\right)^k, \end{aligned}$$

as well as its lifting for complex values of $u$. □

EXAMPLE 18. *Leaves in trees.* We start with binary Catalan trees and with the BGF

$$F(z,u) = z(u + 2zF(z,u) + F(z,u)^2),$$

so that

$$F(z,u^2) = \frac{1}{2z}\left(1 - 2z - \sqrt{(1-2z(1+u))(1-2z(1-u))}\right).$$

This is almost the same as the BGF of leaves in general Catalan trees. The dominant singularity is

$$\rho(u) = \frac{1}{2(1+u)},$$

and the law is Gaussian. The asymptotic form of the mean and variance are immediately derived from consideration of

$$\phi(u) = \frac{\rho(1)}{\rho(u)} = \frac{1+u}{2},$$

with

$$\text{Mean}(\phi) = \frac{1}{2}, \qquad \text{Var}(\phi) = \frac{1}{4}.$$

Since this computation is relative to *twice* the number of leaves, we find that the number of leaves $X_n$ in a binary Catalan tree satisfies

$$E\{X_n\} = \frac{n}{4} + O(1), \qquad \sigma\{X_n\} = \frac{\sqrt{n}}{4} + O(n^{-1/2}).$$

In the case of Cayley trees, the BGF equation[6] is

$$F(z,u) = z(u - 1 + e^{F(z,u)}).$$

By Lagrange inversion, the distribution is related to the Stirling partition numbers. The functional equation admits an explicit solution in terms of Lambert's "$W$-function", which is such that $z = We^W$, with the branch choice that $W = 0$ when $z = 0$. Thus, $W(z) = -T(-z)$, where $T = ze^T$ is the classical "Cayley tree function". Here, we have

$$F(z,u) = z(u-1) - W(-ze^{z(u-1)}).$$

The function $W$ has a dominant singularity of the square-root type at $-e^{-1}$. Thus, one can solve for $\rho(u)$, again in terms of the $W$ function. Here, we find

$$\rho(u) = \frac{1}{u-1} W(e^{-1}(u-1)).$$

In particular, we get $\rho(1) = e^{-1}$, as we should. The expansion near $u = 1$ then comes automatically

$$\frac{\rho(u)}{\rho(1)} = 1 - e^{-1}(u-1) + \frac{3}{2}e^{-2}(u-1)^2 + \mathcal{O}((u-1)^3).$$

Hence the mean and the variance of the number $X_n$ of leaves in a random tree of size $n$ satisfy:

$$E\{X_n\} \sim e^{-1}\, n \approx 0.36787\, n, \qquad \sigma^2\{X_n\} \sim e^{-2}(e-2)\, n \approx 0.09720\, n,$$

and the limit law is a Gaussian. □

EXERCISE 22. Show that the number of leaves in a unary-binary (Motzkin) tree is asymptotically Gaussian.

[6] This example constitutes a typical application of symbolic manipulation systems.

**Patterns in binary Catalan trees.** We develop here a more sophisticated example coming from the analysis of pattern matching in trees [57, 27] that generalizes the problem of leaves. Fix a nonempty binary tree $w$ and let $\omega[t] \equiv \omega_w[t]$ be the number of occurrences of pattern $w$ in tree $t$. By this, we mean the number of internal nodes $\nu$ in $t$ such that the subtree of $t$ rooted at $\nu$ is isomorphic to $w$. The problem is of interest in the analysis of some symbolic manipulation algorithms and of "sharing" strategies; see [57, 27] for the algorithmic context.

A pattern occurs either in the left root subtree $t_0$ or in the right root subtree $t_1$ or at the root iself if $t$ coincides with $w$. This gives rise to the recursive definition

$$\omega[t] = \omega[t_0] + \omega[t_1] + [\![t = w]\!], \qquad \omega[\emptyset] = 0,$$

where $[\![P]\!]$ denotes the indicator function of $P$ whose value is 1 if $P$ is true, and 0 otherwise. The function $u^{\omega[t]}$ is almost multiplicative, and

$$u^{\omega[t]} = u^{[\![t=w]\!]} u^{\omega[t_0]} u^{\omega[t_1]} = u^{\omega[t_0]} u^{\omega[t_1]} + [\![t = w]\!] \cdot (u - 1).$$

Thus, the bivariate generating function $F(z, u)$ where $z$ marks internal nodes and $u$ marks the number of occurrences of $w$,

$$F(z, u) := \sum_t z^{|t|} u^{\omega[t]},$$

satisfies the algebraic equation,

$$F(z, u) = 1 + (u - 1)z^m + zF(z, u)^2,$$

with $m = |w|$ the number of internal nodes of $w$.

The quadratic equation for $F$ leads to

$$F(z, u) = \frac{1}{2z}\left(1 - \sqrt{1 - 4z - 4z^{m+1}(u - 1)}\right).$$

The discriminant has a unique root $\rho = 1/4$ when $u = 1$, while it has $m+1$ roots for $u \neq 1$. By general properties of implicit and algebraic functions (implicit function theorem, Weierstrass preparation), as $u$ tends to 1, one of these roots, call it $\rho(u)$ tends to $1/4$ while all the other ones $\{\rho_j(u)\}_{j=1}^m$ escape to infinity. We have

$$H(z, u) := \frac{1 - 4z - 4z^{m+1}(u - 1)}{1 - z/\rho(u)} = \prod_{j=1}^{m}(1 - z/\rho_j(u)),$$

which is an analytic function in $(z, u)$ for $(z, u)$ in a complex neighbourhood of $(1/4, 1)$. This results from the fact that the algebraic function $1/\rho(u)$ is analytic at $u = 1$. It gives the singular expansion of $G(z, u) = zF(z, u)$:

$$G(z,u) = \frac{1}{2} - \frac{1}{2}\sqrt{H(z,u)}\sqrt{1 - z/\rho(u)}.$$

Thus, we are exactly under the conditions of the theorem. The quantity $\omega$ taken over a random binary tree of size $n + 1$ has mean and variance given asymptotically by

$$\text{Mean}\left(\frac{1}{4\rho(u)}\right) n, \qquad \text{Var}\left(\frac{1}{4\rho(u)}\right) n.$$

The expansion of $\rho(u)$ at 1 is computed easily by iteration of the defining equation:

$$z = \frac{1}{4} - z^{m+1}(u-1) = \frac{1}{4} - (\frac{1}{4} - z^{m+1}(u-1))^{m+1}(u-1) = \cdots.$$

Thus,

$$\rho(u) = \frac{1}{4} - \frac{1}{4^{m+1}}(u-1) + \frac{m+1}{4^{2m+1}}(u-1)^2 + \cdots.$$

This shows that the mean $\mu_n$ and the variance $\sigma_n^2$ of the number of occurrences of a pattern of size $m$ in a random binary tree of size $n$ satisfy

$$\mu_n \sim \frac{n}{4^m}, \qquad \sigma_n^2 \sim n(\frac{1}{4^m} - \frac{2m+1}{4^{2m}});$$

also, the distribution is asymptotically Gaussian. In particular, the probability of occurrence of a pattern at a random node of a random trees decreases fast (the factor of $4^{-m}$ in the estimate of averages) with the size of the pattern, a property that was to be expected and that also holds for strings. The paper [57] shows that similar properties (equivalent to the mean value analysis) hold for any simply generated family. The expression of the BGF $F(z, u)$ is given in [27], where similar developments are used to show that the minimal "dag representation" of a random tree — identical subtrees are "shared" and represented only once— is of average size $O(n(\log n)^{-1/2})$.

EXERCISE 23. Discuss patterns in general Catalan trees, and more generally in simple families of trees.

### 9.5.3 The exponential–logarithmic schema

So far, the occurrence of a Gaussian law has been related to a *movable singularity* that causes coefficients of a bivariate generating function $F(z,u)$ to obey a rough power law of the form

$$f_n(u) = [z^n]F(z,u) \approx \rho(u)^{-n},$$

so that the Quasi-Powers Theorem applies with a scaling factor $\beta_n = n$. In this section, we discuss the situation of a fixed singularity and *variable exponent* in singular expansions. This means a somewhat stronger decomposition property for a BGF as the singularity remains constant when the auxiliary parameter $u$ varies, as in $F(z,u) = C(z)^{-\alpha(u)}$. Typical cases of application are to the set constructions, where the analysis of number of components can be rephrased as the estimation of coefficients in

$$F(z,u) = \exp\left(uG(z)\right),$$

when $G(z)$ is, roughly speaking, logarithmic. In this case, we have parameters whose mean and variance grow logarithmically, a typical instance being the number of cycles in permutations. Analytically, this comes from an approximate form

$$F(z,u) \approx (1 - z/\rho)^{-\alpha(u)},$$

so that

$$f_n(u) = [z^n]F(z,u) \approx \rho^{-n} n^{\alpha(u)-1} \equiv \frac{\rho^{-n}}{n} \exp(\alpha(u)\log n).$$

This is again a case of application of the Quasi-Powers Theorem, but now with a scaling factor $\beta_n = \log n$. The developments in this section are inspired by the paper [28].

**Theorem 9.9 (Exponential–logarithmic schema)** *Let $F(z,u)$ be a bivariate function that is analytic in a domain*

$$\mathcal{D}_0 = \left\{(z,u) \;\middle|\; |z| < \rho, |u| < 1\right\},$$

*and has nonnegative coefficients at $(0,0)$. Assume that there exist $\epsilon > 0$ and $r > \rho$ such that in the domain,*

$$\mathcal{D} = \left\{(z,u) \;\middle|\; |z| \le r, |u-1| \le \epsilon\right\},$$

*the function $F(z,u)$ admits the representation*

$$F(z,u) = A(z,u) + B(z,u)C(z)^{-\alpha(u)} \tag{9.53}$$

*where*

($C_1$) *$A(z,u), B(z,u)$ are analytic for $(z,u) \in \mathcal{D}$,*

($C_2$) *the function $\alpha(u)$ is analytic in $|u-1| \le \epsilon$ with $\alpha(1) \notin \{0,-1,-2,\ldots\}$, and the variability condition is satisfied,*

$$\alpha'(1) + \alpha''(1) \neq 0,$$

($C_3$) *$C(z)$ is analytic for $|z| \le r$, the equation*

$$C(\zeta) = 0 \tag{9.54}$$

*has a unique and simple root $\zeta = \rho$ in $|z| \le r$, and $B(\rho,1) \neq 0$.*

*Then the variable with probability generating function*

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

*converges in distribution to a Gaussian variable and the speed of convergence is $O((\log n)^{-1/2})$. The corresponding mean $\mu_n$ and variance $\sigma_n^2$ satisfy*

$$\mu_n \sim \alpha'(1)\log n, \qquad \sigma_n^2 \sim \alpha'(1)\log n.$$

Proof. Clearly, for the univariate problem, by singularity analysis, one has

$$[z^n]F(z,1) = B(\rho,1)(-\rho C'(\rho))^{-\alpha(1)}\rho^{-n}\frac{n^{\alpha(1)-1}}{\Gamma(\alpha(1))}\left(1+O(\frac{1}{n}\right).$$

For the bivariate problem, the contribution arising from $[z^n]A(z,u)$ is exponentially small, since $A(z,u)$ is $z$-analytic in $|z| \le r$.

Write next

$$B(z,u) = (B(z,u) - B(\rho,u)) + B(\rho,u).$$

The first term satisfies

$$B(z,u) - B(\rho,u) = \mathcal{O}((z-\rho)),$$

uniformly with respect to $u$, since

$$\frac{B(z,u) - B(\rho,u)}{z-\rho}$$

is analytic in $z$ and $u$, by division of power series representations. Let $A$ be an upper bound on $\alpha(u)$ on $|u-1| \le \epsilon$. Then, by singularity analysis and its companion uniformity,

$$[z^n](B(z,u) - B(\rho,u))C(z)^{-\alpha(u)} = \mathcal{O}(\rho^{-n}n^{A-2}).$$

By suitably restricting the domain of $u$ to $|u-1| \le \delta$, one may freely assume that $A - 2 < \alpha(1) - \frac{7}{4}$. Thus, the contribution from this part is small.

It only remains to analyse

$$[z^n]B(\rho, u)C(z)^{-\alpha(u)}.$$

This is done exactly like in the univariate case, again taking advantage of the uniformity afforded by singularity analysis. We find, uniformly for $u$ in a small neighbourhood of 1,

$$[z^n]F(z,u) = \frac{B(\rho,u)\rho^{-n}}{n\Gamma(\alpha(u))}(-\rho C'(\rho))^{-\alpha(u)} e^{\alpha(u)\log n}\left(1 + \mathcal{O}(n^{-1/2})\right).$$

Thus, the Quasi-Powers Theorem applies and the law is Gaussian in the limit. □

The bivariate EGF for *permutations* with $u$ marking the number of cycles is

$$F(z,u) = \sum \begin{bmatrix} n \\ k \end{bmatrix} u^k \frac{z^n}{n!} = (1-z)^{-u} = \exp\left(u \log \frac{1}{1-z}\right) \cdot,$$

so that we are in the simplest case of an exponential– logarithmic schema. Theorem 9.9 shows (once more!) that *the number of cycles in a random permutation of size $n$ converges to a Gaussian limiting distribution.* This classical result stating the asymptotically normal distribution of the Stirling numbers (of the first kind) constitutes Goncharov's Theorem.

EXAMPLE 19. *Cycles in derangements.* The number of cycles is asymptotically normal in generalized derangements where a finite set $S$ of cycle lengths are forbidden. This results immediately from the BGF

$$F(z,u) = \exp(uG(z)), \qquad G(z) = \log \frac{1}{1-z} - \sum_{s \in S} \frac{z^s}{s}.$$

The classical derangement problem corresponds to $S = \{1\}$; see [10]. □

EXAMPLE 20. *Clouds and 2-regular graphs.* "Clouds" are defined in [10, p. 274]: let $n$ straight lines in the plane be given in general position, so that there are $\binom{n}{2}$ intersecting points; a cloud of size $n$ is a (maximal) set of $n$ intersection points, no three of which are collinear. By duality, there is a one–to–one correspondence between clouds and *2–regular graphs.* A 2–regular graph of size $n$ is an undirected graph with $n$ edges, such that each vertex has degree exactly 2. Any 2–regular graph may be decomposed into a product of connected components that are (undirected) cycles of length at least 3. Hence the bivariate EGF for 2–regular graphs, with $u$ marking the number of connected components, is:

$$F(z,u) = \exp\left(u(\frac{1}{2}\log\frac{1}{1-z} - \frac{z}{2} - \frac{z^2}{4})\right) = \frac{e^{-uz/2-uz^2/4}}{(1-z)^{u/2}}.$$

The function $\exp(u(z/2 + z^2/4))$ is entire, so that the conditions of Theorem 9.9 are satisfied. Thus, the number of connected components in a 2–regular graph, (this is equivalent to the number of polygons in a cloud) has a Gaussian limiting distribution. □

EXAMPLE 21. *Random mappings.* Let $f$ denote a function that maps the set $N = \{1, 2, \cdots, n\}$ into itself. Such a function $f$ may be represented by a directed graph $G_f$ with vertex set $N$ and edge set $\{(i, f(i)); i \in N\}$. Such graphs, in which every point has out–degree one, are called *functional digraphs*; see [33, p. 68]. A functional digraph may be viewed as a set of components that are themselves cycles of rooted labelled trees. The bivariate EGF for functional digraphs with $u$ marking connected components is

$$F(z,u) = \exp\left(u(\log\frac{1}{1-T(z)})\right),$$

where the generating function of rooted labelled trees $T(z)$ is the Cayley tree function defined implicitly by the relation $T(z) = z\exp(T(z))$. By the inversion theorem for implicit functions we have

$$T(z) = 1 - \sqrt{2(1-ez)} + \sum_{k\geq 2} c_k(1-ez)^{k/2}.$$

Thus,

$$F(z,u) = \exp\left\{u\left(\frac{1}{2}\log\frac{1}{1-ez} + H((1-ez)^{1/2})\right)\right\},$$

where $H(v)$ is analytic at $v = 0$. From this form and Theorem 9.9, we obtain a theorem of Stepanov [56]: the number of components in functional digraphs has a limiting Gaussian distribution.

This approach extends to functional digraphs satisfying various degree constraints as considered in [1]. This analysis and similar ones are of some relevance to integer factorization, using Pollard's "rho" method [25, 42, 54]. □

**Unlabelled constructions.** In the case of unlabelled structures, the class $\mathcal{F}$ of multisets over a class $\mathcal{G}$ have OGF,

$$\sum_{n\geq 0} F_n z^n = \prod_{n\geq 1} (1-z^n)^{-G_n}.$$

By taking logarithms and reorganizing the corresponding series, we get the alternative form

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

Similarly, in the bivariate case, where $u$ marks the number of components, the bivariate GF is (see Chapter 3),

$$F(z,u) = \sum_{n,k\geq 0} F_{n,k} u^k z^n = \exp\left(\frac{u}{1}G(z) + \frac{u^2}{2}G(z^2) + \frac{u^3}{3}G(z^3) + \cdots\right),$$

which is of the form $\exp(G(z))^u \cdot B(z,u)$. Here, we are interested in structures such that $G(z)$ has a logarithmic singularity, in which case Theorem 9.9 applies, as soon as $G(z)$ has radius of convergence $\rho < 1$.

EXAMPLE 22. *Polynomial factorization.* Fix a finite field $K = GF(q)$ and consider the class $\mathcal{P}$ of monic polynomials (having leading coefficient 1) in $K[z]$, with $\mathcal{I}$ the subclass of irreducible polynomials. Obviously, $P_n = q^n$, so that

$$P(z) = (1-qz)^{-1}.$$

Because of the unique factorization property, a polynomial is a multiset of irreducible polynomial, whence the relation

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

The preceding relation can be inverted using Moebius inversion. If we set $L(z) = \log P(z)$, then we have

$$I(z) = \sum_{k\geq 1} \mu(k)\frac{L(z^k)}{k} = \log\frac{1}{1-qz} + \sum_{k\geq 2}\mu(k)\frac{L(z^k)}{k},$$

where $\mu$ is the Moebius function.

Since $L(z^k)$ is analytic for $|z| < q^{-1/2}$ whenever $k \geq 2$, and $|L(z^k)| < c^{st}|z|^k$, the sum $\sum_{k\geq 2}\mu(k)L(z^k)/k$ is analytic for $|z| \leq \tau$, with $q^{-1} <$

$\tau < q^{-1/2}$. Hence $I(z)$ has an isolated singularity of logarithmic type at $z = q^{-1} < 1$.

Thus the average number of irreducible factors in a polynomial, and its variance, are both asymptotically $\log n + O(1)$ (this result appears in [42, Ex. 4.6.2.5]). Let $\Omega_n$ be the random variable representing the number of irreducible factors of a random polynomial of degree $n$ over $GF(q)$, each factor being counted with its order of multiplicity. Then as $n$ tends to infinity, for any two real constants $\lambda < \mu$, we have

$$\Pr\{\log n + \lambda\sqrt{\log n} < \Omega_n < \log n + \mu\sqrt{\log n}\} \to \frac{1}{\sqrt{2\pi}} \int_{\lambda}^{\mu} e^{-t^2/2}\, dt.$$

This statement [28] is a counterpart of the famous Erdös–Kac Theorem (1940) for the number of prime divisors of natural numbers (with here $\log n$ that replaces $\log\log n$ when dealing with integers at most $n$). A similar result holds for the parameter $\omega_n$ that represents the number of *distinct* irreducible factors in a random polynomial of degree $n$. □

It is perhaps instructive to re-examine this last example at an abstract level, in the light of general principles of analytic combinatorics.

> *A polynomial over a finite field is determined by the* sequence *of its coefficients. Hence, the class of all polynomials, as a sequence class, has a polar singularity. On the other hand, unique factorization entails that a polynomial is also a* multiset *of irreducible factors ("primes"). Thus, the class of irreducible polynomials, that is implicitly determined, is logarithmic, since the multiset construction to be inverted is in essence an exponential operator. Consequently, the number of irreducible factors obeys the exponential-logarithmic scheme, so that it is asymptotically Gaussian.*

Eventually, the limit law arises because of the purely analytic character of the generating functions involved, together with permanence of analytic relations implied by combinatorial constructions.

EXAMPLE 23. *Mapping patterns.* Let $f$ and $g$ be two functions mapping the set $\{1, 2, \cdots, n\}$ into itself. Mappings $f$ and $g$ are said to be equivalent if there exists a permutation $\pi$ of $\{1, 2, \cdots, n\}$ such that $f(i) = j$ iff $g(\pi(i)) = \pi(j)$. Mapping patterns are thus equivalence classes of mapping functions, or equivalently functional digraphs on unlabelled points. They correspond to multisets of cycles of rooted unlabelled trees. The OGF for rooted unlabelled trees satisfies the implicit relation

$A(z) = z \exp(\sum \frac{1}{k} A(z^k))$, and Otter [50] proved that

$$A(z) = 1 - c_1 \sqrt{(1 - z/\eta)} + \sum_{k \geq 2} c_k (1 - z/\eta)^k.$$

for some $\eta < 1$.

On the other hand, by the translation of the cycle construction, if $\mathcal{G}$ is the unlabelled cycle construction applied to $\mathcal{A}$, then (see Chapter 3),

$$G(z) = \sum_{k \geq 1} \frac{\phi(k)}{k} \log \frac{1}{1 - A(z^k)},$$

where $\phi(k)$ is the Euler totient function. In the present context, since $A(z)$ has radius of convergence $\eta$ strictly less than 1,

$$G(z) = \log \frac{1}{1 - A(z)} + S(z),$$

where $S(z)$ is analytic at $\eta$. Finally the bivariate OGF for random mapping patterns satisfies

$$\begin{aligned} F(z,u) &= \exp\big(\sum_{k \geq 1} u^k \frac{G(z^k)}{k}\big) \\ &= \exp\big(u \log \frac{1}{1 - A(z)} + uS(z) + T(z,u)\big) \\ &= \exp\left(\frac{u}{2} \log \frac{1}{1 - z/\eta} + u\big((1 - z/\eta)^{1/2}\big) + uS(z) + T(z,u)\right), \end{aligned}$$

where $S(z)$ is analytic at $\eta$, $T(z,u)$ is analytic for $z = \eta$ and $u = 1$, and $H$ is analytic around 0, with $H(0) = 0$. Thus conditions for applying Theorem 9.9 are satisfied and the number of components in random mapping patterns has a Gaussian limiting distribution. The mean value is asymptotic to $\frac{1}{2} \log n$ (this result appears in [49] and the variance is $\frac{1}{2} \log n + O(1)$. □

Example 24. *Arithmetical semigroups.* Knopfmacher [40] defines an arithmetical semigroup as a semigroup with unique factorization, and a size function (or degree) such that

$$|xy| = |x| + |y|,$$

where the number of elements of a fixed size is finite. If $\mathcal{P}$ is an arithmetical semigroup and $\mathcal{I}$ its set of 'primes' (irreducible elements), axiom $A^{\#}$ of Knopfmacher asserts the condition

$$\text{card}\{x \in \mathcal{P} \ \ / \ \ |x| = n\} = cq^n + O(q^{\alpha n}) \quad (\alpha < 1).$$

It is shown by Knopfmacher that several algebraic structures forming arithmetical semigroups satisfy axiom $A^{\#}$, and thus the conditions of Theorem 9.9 are automatically satisfied. Therefore, the results deriving from Theorem 9.9 fit into the framework of Knopfmacher's "abstract analytic number theory", since they provide general conditions under which theorems of the Erdös–Kac type must hold true. Examples of application are Galois polynomial rings (the example of polynomial factorization), finite modules or semisimple finite algebras over a finite field $K = GF(q)$, integral divisors in algebraic function fields, ideals in the principal order of a algebraic function field, finite modules, or semisimple finite algebras over a ring of integral functions. □

## 9.6 Implicit schemas and functional equations

There is usually a fairly transparent approach to the analysis of BGFs defined implicitly as solutions of functional equations. One should start with the analysis at $u = 1$ and then examine the effect on singularities when $u$ varies in a very small neighbourhood of 1. In accordance with what we have already seen many times, the process is a perturbation analysis of the solution to a functional equation near a singularity; the singularity may *move*, or it may be stable but the corresponding exponent may *vary*.

In this section, we mostly illustrate by way of examples the application of our previous main theorems to functions defined implicitly. It would be possible to encapsulate several of these principles into general theorems; however, the sets of conditions become soon rather "heavy". For the purpose of preserving a simple perspective, we limit ourselves to sketchy proofs and refer to the original literature for details.

### 9.6.1 Rational functions.

Positive rational functions arise in connection with problems that can be described by finite state devices, by paths in graphs, and by Markov chains. The bivariate problem is then expressed by a linear equation

$$Y(z,u) = V(z,u) + T(z,u) \cdot Y(z,u), \tag{9.55}$$

where $T = T(z,u)$ is an $m \times m$ matrix with polynomial entries in $z, u$ having nonnegative coefficients, $Y = Y(z,u)$ is an $m \times 1$ column vector of unknowns, and $V = V(z,u)$ is a column vector of nonnegative initial conditions.

We first revisit the univariate problem

$$Y(z) = V(z) + T(z) \cdot Y(z). \tag{9.56}$$

A matrix $T(z)$ with polynomial entries in $z$ and nonnegative coefficients is said to be

— *transitive*, if there exists an exponent $e$ such that $T(z)^e$ has all its $m^2$ entries that are nonzero; the smallest such $e$ is called the transitivity exponent;

— *proper*, if it is transitive and for the transitivity exponent $e$, $T(z)^e$ has all its entries that are of $z$-valuation strictly greater than 0;

— *primitive*, if $T(z) = U(z^d)$ for a matrix with polynomial entries $U(z)$ entails $d = 1$.

In this case (see the developments below and Chapter 8), an extended Perron-Frobenius theory holds for the univariate matrix $T(z)$. In other words, the function

$$C(z) = \det(I - T(z))$$

has a unique dominant root that is a simple zero at some positive point $\rho > 0$. Accordingly, any component $F(z) = Y_i(z)$ of a solution to the system (9.56) has a unique dominant singularity at $z = \rho$ that is a simple pole,

$$F(z) = \frac{B(z)}{C(z)},$$

with $B(\rho) \neq 0$.

In the bivariate case, each component of the solution to the system (9.55) can be put under the form

$$F(z, u) = \frac{B(z, u)}{C(z, u)}, \qquad C(z, u) = \det(I - T(z, u)).$$

Since $B(z, u)$ is a polynomial, it does not vanish for $(z, u)$ in a sufficiently small neighbourhood of $(\rho, 1)$. Similarly, by Weierstrass preparation, there exists a function $\rho(u)$ locally analytic near $u = 1$, such that

$$C(\rho(u), u) = 0, \qquad \rho(1) = \rho.$$

Thus, it is sufficient that the variability conditions (9.36) be satisfied to infer a limit Gaussian distribution.

**Corollary 9.2 (Positive rational systems)** *Let $F(z,u)$ be a bivariate function that is analytic at $(0,0)$ and has nonnegative coefficients. Assume that $F(z,u)$ coincides with the component $Y_1$ of a system of linear equations in $Y = (Y_1, \ldots, Y_m)^T$,*

$$Y = V + T \cdot Y,$$

*where $V = (V_1(z,u), \ldots, V_m(z,u))$, $T = (T_{i,j}(z,u))_{i,j=1}^m$, and each of $V_j, T_{i,j}$ is a polynomial in $z, u$ with nonnegative coefficients. Assume also that $T(z,1)$ is transitive, proper, and primitive, and let $\rho(u)$ be the unique solution of*

$$\det(I - T(\rho(u), u)) = 0,$$

*that is analytic at 1 and such that $\rho(1) = \rho$. Then, provided the variability condition is satisfied,*

$$Var\left(\frac{\rho(1)}{\rho(u)}\right) > 0,$$

*a Gaussian Limit Law holds for the coefficients of $F(z,u)$ with mean and variance that are $\mathcal{O}(n)$ and speed of convergence that is $\mathcal{O}(n^{-1/2})$.*

Proof. As usual, we first examine the univariate problem[7]. Consider the matrix $T(z) = T(z,1)$. The underlying graph, with an edge between $i$ and $j$ iff $T(z)_{i,j}$ is strongly connected and aperiodic, like the graph of an irreducible aperiodic Markov chain. For any real $x > 0$, the matrix $T(x)$ is thus of the Perron-Frobenius type: it has a maximal (simple) eigenvalue $\lambda(x)$ that is strictly positive and such that the modulus of any other eigenvalue is strictly smaller. (See for instance [39, p. 542–551] for a crisp survey of Perron-Frobenius theory.) Also, since $\lambda(x)$ is equal to the spectral radius of $T(x)$,

$$\lambda(x) = \lim_{n\to\infty} ||T(x)^n||^{1/n},$$

and since the entries of $T(x)$ are positive by assumption, $\lambda(x)$ is nondecreasing. Next, $\lambda(x)$ is piecewise analytic (and even algebraic), being a solution to the polynomial equation

$$\det(\lambda I - T(x)) = 0.$$

Furthermore, $\lambda(x)$ cannot be constant over an interval: otherwise, one of the algebraic branches would be a positive constant, a contradiction with the fact that all branches reduce to 0 when $x = 0$, by properness ($T(0)$ being nilpotent, its spectral radius equals 0). Thus, $\lambda(x)$ is strictly increasing. Finally, from the spectral radius formula, $T(x) \to \infty$ as $x \to \infty$.

[7] For the sake of being self–contained, some of these developments duplicate the treatment given in Chapter 8.

From these considerations, there results the existence of a unique real $\rho > 0$ such that

$$\lambda(\rho) = 1.$$

This uniquely determined value $\rho$ is then also the smallest positive root of

$$C(z) \equiv \det(I - T(x)),$$

that is to say, the dominant positive singularity of $1/C(z)$. In particular, the coefficients $[z^n]C(z)^{-1}$ are of exponential order $\rho^{-n}$.

There remains to prove that the numerator of $F(z)$ does not vanish at $z = \rho$, that is, $B(\rho) \neq 0$. The Jacobi formula,

$$\log \det M = \mathrm{Tr}\ \log M,$$

for a matrix $M$, extends the functional property of logarithms from scalars to matrices. Here, it gives

$$\log \det(1 - T(z))^{-1} = \mathrm{Tr}\ \log(1 - T(z))^{-1} = \sum_{j=1}^{\infty} \left(\frac{1}{j}\, \mathrm{Tr}\ T(z)^j\right),$$

a function equal to $1/C(z)$ whose radius of convergence is exactly $\rho$, by the previous discussion. Thus, by domination of power series, $[z^n]\,\mathrm{Tr}\,(I - T(z))^{-1}$ must be of exponential order exactly $\rho^{-n}$, not smaller. The property then carries out to any nondegenerate positive system, since the solution to (9.56) is

$$F(z) = (I - T(z))^{-1} V(z),$$

with $V(z)$ nonnegative. Thus, $F(z)$ itself has radius of convergence exactly $\rho$. This implies in turn $B(\rho) \neq 0$.

The Gaussian limit law then results directly from Theorem 9.7. □

The constants $\mu, \sigma$ involved in estimates of the mean and standard deviation, $\mu_n \sim \mu n, \quad \sigma_n \sim \sigma\sqrt{n}$, are then determined from $C(z, u) = \det(I - T(z, u))$ by Eq. (9.36). Thus, in any particular application, one can determine by computation whether the variability condition is satisfied. It may be however more difficult to check these conditions for a whole classes of problems.

EXAMPLE 25. *Limit theorem for Markov chains.* Assume that $M$ is the transition matrix of an irreducible aperiodic Markov chain, and consider the parameter $\chi$ that records the number of passages through state 1 in a path of length $n$ that starts in state 1. Then, the theorem applies with

$$V = (1, 0, \ldots, 0)^T, \quad T_{i,j}(z, u) = zM_{i,j} + z(u-1)M_{i,0}\delta_{j,0}.$$

We therefore derive the classical limit theorem for Markov chains: *In an transitive aperiodic Markov chain, the number of times a designated state is reached when $n$ transitions are effected is asymptotically Gaussian.* The result also applies to paths in any strongly connected aperiodic digraph. □

EXAMPLE 26. *Sets of patterns in words.* The example 14 is clearly covered by the corollary. More generally, fix a finite set of binary strings $S \subset \{0,1\}^\star$, and a parameter $\chi$ that is equal to the number of occurrences of members of $S$ in binary strings of length $n$. This is a typical instance of a problem that can be recognized by a finite automaton that is essentially a digital tree built on $S$ and equipped with "return edges". Thus, the problem fits into the framework of Corollary 9.2. Bender and Kochman have proved that the limit law of $\chi$ is Gaussian; see [4] for a general discussion. Alternatively, the problem may be viewed as enumerating paths by length and number of traversals of a distinguished set of nodes in a De Bruijn graph [22]. □

EXAMPLE 27. *Tilings.* Take an $(2 \times n)$ chessboard of 2 rows and $n$ columns, and consider coverings with "monomer tiles" that are $(1 \times 1)$-pieces, and "dimer tiles" that are either of the horizontal $(1 \times 2)$ or vertical $(2 \times 1)$ type. Consider then the collection of all "partial coverings" in which each column is covered exactly, except for the last one. The partial coverings are then of one of 4 types and the legal transitions are described by a compatibility graph. For instance, if the previous column contained one horizontal dimer and one monomer, the current column has one occupied cell, and one free cell that may be occupied either by a monomer or a dimer. This finite state description corresponds to a set of linear equations over BGFs (with $z$ marking the area covered and $u$ marking the total number of tiles). The transition matrix is

$$T(z,u) = z \begin{pmatrix} u & u^2 & u^2 & u^2 \\ 1 & 0 & 0 & 0 \\ u & 0 & 0 & 0 \\ u & 0 & 0 & 0 \end{pmatrix}.$$

In particular, we have

$$\det(I - T(z,u)) = 1 - zu - z^2(u^2 + u^3).$$

Then, Corollary 9.2 applies. The method, called *transfer matrix method*, extends to $(k \times n)$ chessboards, for any fixed $k$. See [6] for details. □

EXERCISE 24. Consider integer compositions where consecutive summands add up to at least 4. The number of summands in such a composition of large size is asymptotically normal. [Hint: see [6]]

EXERCISE 25. Consider general Catalan trees of width less than a fixed bound $w$. (The width is the maximum number of nodes at any level in the tree.) Analyse the distribution of height in such trees.

### 9.6.2 Algebraic functions

Many combinatorial problems, especially as regards paths and trees, lead to descriptions by context–free languages. Accordingly, the GF's are algebraic functions. The most frequent situation is that of univariate GF's having singularities of the square-root type.

**Corollary 9.3 (Algebraic functions)** *Let $F(z,u)$ be a bivariate function that is analytic at $(0,0)$ and has nonnegative coefficients. Assume that $F(z,u)$ is one of the solutions $y$ of a polynomial equation*

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

*where $\Phi$ is an irreducible polynomial of total degree $m$, of degree $d \geq 2$ in $y$. Assume that $F(z,1)$ is has a unique dominant singularity at $\rho > 0$, with a singular behaviour of the square-root type there. Define the resultant polynomial,*

$$\Delta(z,u) = \mathit{result}_y \left( \Phi(z,u,y), \frac{\partial}{\partial y}\Phi(z,u,y) \right),$$

*and assume that $\rho$ is a simple root of $\Delta(z,1)$. Let $\rho(u)$ be the unique root of the equation*

$$\Delta(\rho(u),u),$$

*analytic at 1, such that $\rho(1) = 1$. Then, provided the variability condition*

$$\mathit{Var}\left(\frac{\rho(1)}{\rho(u)}\right) > 0,$$

*is satisfied, a Gaussian Limit Law holds for the coefficients of $F(z,u)$.*

PROOF. The assumption of a square-root singularity (see Chapter 8) means that the polynomial $\Phi(\rho, 1, y)$ has a double zero at $y = \tau$, where $\tau = \lim_{z\to\rho^-} F(z, 1)$. Equivalently, we have

$$\left(\frac{\partial}{\partial y}\Phi(\rho, 1, y)\right)_{y=\tau} = 0, \qquad \left(\frac{\partial^2}{\partial y^2}\Phi(\rho, 1, y)\right)_{y=\tau} \neq 0.$$

Thus, Weierstrass preparation gives the local factorization

$$\Phi(z, u, y) = (y^2 + c_1(z, u)y + c_2(z, u))H(z, u, y),$$

where $H(z, u, y)$ is analytic and nonzero at $(\rho, 1, \tau)$ while $c_1(z, u), c_2(z, u)$ are analytic at $(z, u) = (\rho, \tau)$.

From the solution of the quadratic equation, we must have locally

$$y = \frac{1}{2}\left(-c_1(z, u) \pm \sqrt{c_1(z, u)^2 - 4c_2(z, u)}\right).$$

Consider first $(z, u)$ restricted by $0 \leq z < \rho$ and $0 \leq u < 1$. Since $F(z, u)$ is real there, we must have $c_1(z, u)^2 - 4c_2(a, u)$ also real and nonnegative. Since $F(z, u)$ is continuous and increasing with $z$ for fixed $u$, and since the discriminant $c_1(z, u)^2 - 4c_2(a, u)$ vanishes at 0, the determination with the minus sign has to be constantly taken. In summary, we have

$$F(z, u) = \frac{1}{2}\left(-c_1(z, u) - \sqrt{c_1(z, u)^2 - 4c_2(z, u)}\right).$$

The function $C(z, u) = c_1^2(z, u) - 4c_2(z, u)$ has a simple real zero at $(\rho, 1)$. Thus there is locally a unique analytic branch of the solution to $C(\rho(u), u) = 0$ such that $\rho(1) = \rho.$. This branch is also by necessity a root of the resultant equation $\Delta(\rho(u), u) = 0$. The conditions of Theorem 9.8 therefore apply and the Gaussian law follows. □

This theorem asserts that, under suitable conditions, the only possible dominant singularity of the BGF is a "lifting" of the singularity of the univariate GF $F(z, 1)$ and the nature of the singularity —the square-root type— does not change. The result generalizes to the case of a function $\Phi$ that is analytic in sufficiently large bounded domains, *e.g.*, an entire function. The condition is that the analytic curves

$$\Phi(z, u, y) = 0, \qquad \frac{\partial}{\partial y}\Phi(z, u, y) = 0$$

have an intersection that "moves analytically" and nontrivially for $u$ near 1, and a sufficient condition for this is the nonvanishing of the Jacobian determinant

$$J(z, u, y) := \begin{vmatrix} \frac{\partial}{\partial z}\Phi(z, u, y) & \frac{\partial}{\partial y}\Phi(z, u, y) \\ \frac{\partial^2}{\partial zy}\Phi(z, u, y) & \frac{\partial^2}{\partial y^2}\Phi(z, u, y) \end{vmatrix} \tag{9.57}$$

and its first derivative with respect to $u$ at $(\rho, 1, \tau)$,

$$J(\rho, 1, \tau) \neq 0, \qquad \left.\frac{\partial}{\partial u} J(z, u, y)\right|_{(\rho,1,\tau)} \neq 0. \tag{9.58}$$

In the case of Corollary 9.3 and of these extensions, the expansion of $\rho(u)$ at $u = 1$, hence the mean and variance of the distribution, are computable explicitly from $\Phi$, its derivatives, and the quantities $\rho$ and $\tau = F(\rho, 1)$.

The corollary applies to a great variety of decomposable parameters of context–free languages, tree like objects, and more generally many recursively defined combinatorial types. Examples of parameters covered are leaves, node types, and various sorts of patterns in combinatorial tree models. Drmota has worked out a different set of conditions for asymptotic normality. In particular, one of Drmota's results [15] yields asymptotic normality, under minor technical restrictions, for a polynomial *system* with positive coefficients that is "irreducible", meaning that the dependency graph between nonterminals is strongly connected.

### 9.6.3 Differential equations

*Ordinary differential equations* (ODE's, for short) in one variable, when *linear* and with analytic coefficients, have solutions whose singularities occur at well-defined places, namely those that entail a reduction of order. The possible singular exponents of solutions are then obtained as roots of a polynomial equation, the indicial equation. Such ordinary differential equations are usually a reflection of a combinatorial decomposition and suitably parametrized versions then open access to a number of combinatorial parameters. In this case, the ODE normally remains an ODE in the main variable $z$ that records size, while the auxiliary variable $u$ only affects the coefficients but not the global shape of the original ODE.

Three cases may then occur for a linear ODE parametrized by $u$.

- *Movable singularity:* the location of the dominant singularity $\rho(u)$ changes with $u$ but the singular exponent does not change; the analysis is then similar to that of algebraic-logarithmic singularities.

- *Movable exponent:* the dominant singularity does not move but the singular exponent $\alpha(u)$ changes; the analysis then resorts to the exponential-logarithmic schema.

- *Movable singularity and movable exponent:* in this case, the singular behaviour is essentially dictated by the movable singularity but with

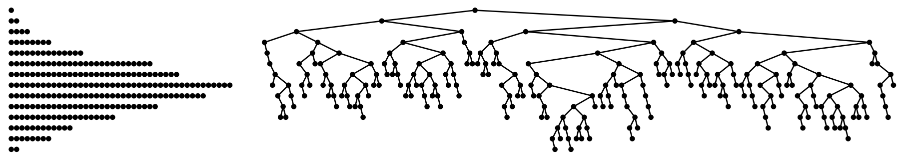

Figure 9.10: The distribution of internal node levels in binary search trees illustrates, even on a single random instance of size $n = 256$, the Gaussian law of node levels.

> an auxiliary contribution arising from the movable exponent; the analysis of this mixed case then requires an extension of the quasi-power framework, as developed by Gao in Richmond in [30].

Here, we focus on the important case of a fixed singularity and a movable exponent. The required singularity perturbation analysis is inspired by the treatment of Flajolet and Lafforgue in [23]. The corresponding univariate problems resort to holonomic asymptotics discussed in Chapter 8.

**Linear differential equations.** The example of the distribution of levels of nodes in random binary search trees or heap–ordered trees illustrates well the situation of a fixed singularity and movable exponent. A heap–ordered tree (HOT) is a plane binary increasing tree. HOTs constitute an unambiguous tree representation of permutations [54]. The EGF of HOTs is

$$F(z) = \frac{1}{1-z} = \sum_{n \geq 0} n! \, \frac{z^n}{n!},$$

as results either from the combinatorial bijection with permutations or from the root decomposition of increasing trees that translates into the functional equation,

$$F(z) = 1 + \int_0^z F^2(t)\, dt, \tag{9.59}$$

a Riccati equation in disguise. Let $F(z,u)$ be the BGF of HOT's where $u$ records the depth of external nodes. In other words, $f_{n,k} = [z^n u^k] F(z,u)$ is such that $\frac{1}{n} f_n n, k$ represents the probability that a random external node in a random tree of size $n$ is at depth $k$ in a random tree. The probability space is then a product set of cardinality $(n+1) \cdot n!$, as there are $n!$ trees each containing $(n+1)$ external nodes. By a standard equivalence principle, the quantities $\frac{1}{n} f_n n, k$ also give the probability that a random unsuccessful search in a random binary search tree of size $n$ necessitates $k$ comparisons.

Since the depth of a node is inherited from subtrees, the function $F(z,u)$ satisfies the *linear* integral equation,

$$F(z,u) = 1 + 2u \int_0^z F(t,u)\,\frac{dt}{1-t}, \qquad (9.60)$$

or, after differentiation,

$$\frac{\partial}{\partial z}F(z,u) = \frac{2u}{1-z}F(z,u), \qquad F(0,u) = 1.$$

This equation is in fact a linear ODE with $u$ entering as a parameter,

$$\frac{d}{dz}y(z) - \frac{2u}{1-z}y(z) = 0, \qquad y(0) = 0.$$

The solution of any separable first-oder ODE is obtained by quadratures, here,

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

From singularity analysis, provided $u$ avoids $\{0, -\frac{1}{2}, -1, \ldots\}$, we have

$$f_n(u) := [z^n]F(z,u) = \frac{n^{2u-1}}{\Gamma(2u)}\left(1 + \mathcal{O}(\frac{1}{n})\right),$$

and the error term is uniform in $u$ provided, say, $|u-1| \le \frac{1}{4}$. Thus, Theorem 9.9 applies, and the law with PGF $f_n(u)/f_n(1)$ converges to a Gaussian limit.

A similar result holds for levels of internal nodes, and is proved by similar devices. Figure 9.10 display the a random binary search tree of size 256, together with a diagram of internal node levels (left). The Gaussian profile is even perceptible on this single instance, which actually suggests a stronger "functional limit theorem" for these objects.

Naturally, explicit expressions are available in such a simple case,

$$\frac{f_n(u)}{f_n(1)} = \frac{2u\cdot(2u+1)\cdots(2u+n-1)}{(n+1)!},$$

so that more direct proofs of the Gaussian limit are possible: see [47, Ch. 2], for another proof of this result that is originally due to Louchard. What is interesting here is the fact that $F(z,u)$ viewed as a function of $z$ has a singularity at $z=1$ that does not move and, in a way, originates in the combinatorics of the problem — the EGF $(1-z)^{-1}$ of permutations. The auxiliary parameter $u$ appears here directly in the exponent, so that the application of singularity analysis or of the more sophisticated Theorem 9.9 is immediate.

**Corollary 9.4 (Linear differential equations)** *Let $F(z,u)$ be a bivariate generating function with nonnegative coefficients that satisfies a linear differential equation*

$$a_0(z,u)\frac{\partial^r F}{\partial z^r} + \frac{a_1(z,u)}{(\rho - z)}\frac{\partial^{r-1}F}{\partial z^{r-1}} + \cdots + \frac{a_r(z,u)}{(\rho - z)^r}F = 0,$$

*with $a_j(z,u)$ analytic at $\rho$, and $a_0(\rho,1) \neq 0$. Let $f_n(u) = [z^n]F(z,u)$, and assume the following conditions:*

- *[Nonconfluence] The indicial polynomial*

$$J(\alpha) = a_0(\rho,1)(\alpha)_{(r)} + a_1(\rho,1)(\alpha)_{(r-1)} + \cdots + a_r(\rho,1) \qquad (9.61)$$

  *has a unique root $\sigma > 0$ which is simple and such that all other roots $\alpha \neq \sigma$ satisfy $\Re(\alpha) < \sigma$;*
- *[Dominant growth] $f_n(1) \sim C \cdot \rho^{-n} n^{\sigma-1}$, for some $C > 0$.*
- *[Variability condition]*

$$\sup \frac{Var(f_n(u))}{\log n} > 0.$$

*Then the coefficients of $F(z,u)$ admit a limit Gaussian law.*

Proof. (See the paper by Flajolet and Lafforgue [23] for a detailed example, Chapter 8, or the books by Henrici [34] and Wasow [62] for a general treatment of singularities of linear ODEs.) We assume in this proof that no two roots of the indicial polynomial (9.61) differ by an integer. Consider first the univariate problem along the lines of Chapter 8. A differential equation,

$$a_0(z)\frac{d^r F}{dz^r} + \frac{a_1(z)}{(\rho - z)}\frac{d^{r-1}F}{dz^{r-1}} + \cdots + \frac{a_r(z)}{(\rho - z)^r}F = 0, \qquad (9.62)$$

with the $a_j(z)$ analytic at $\rho$ and $a_1(\rho) \neq 0$ has a basis of local singular solutions obtained by substituting $(\rho - z)^{-\alpha}$ and cancelling the terms of maximum order of growth. The candidate exponents are thus roots of the *indicial equation*,

$$J(\alpha) \equiv a_0(\rho)(\alpha)_{(r)} + a_1(\rho)(\alpha)_{(r-1)} + \cdots + a_r(\rho) = 0.$$

If there is a unique (simple) root of maximum real part, $\alpha_1$, then there exists a solution to (9.62) of the form

$$Y_1(z) = (\rho - z)^{-\alpha_1} h_1(\rho - z),$$

where $h_1(w)$ is analytic at 0 and $h_1(0) = 1$. (This results easily from a solution by indeterminate coefficients.) All other solutions are then of smaller growth and of the form

$$Y_j(z) = (\rho - z)^{-\alpha_j} h_j(\rho - z) \left(\log(z - \rho)\right)^{k_j},$$

for some integers $k_j$ and some functions $h_j(w)$ analytic and nonzero at $w = 0$. Then, $F(z)$ has the form

$$F(z) = \sum_{j=1}^{r} c_j Y_j(z).$$

Then, provided $c_1 \neq 0$,

$$[z^n]F(z) = \frac{c_1}{\Gamma(\sigma)} \rho^{-n} n^{\alpha_1 - 1}(1 + o(1)).$$

Under the assumptions of the theorem, we must have $\sigma = \alpha_1$, and $c_1 \neq 0$. The reality assumption is natural for a series $F(z)$ that has real coefficients.

When $u$ varies in a neighbourhood of 1, we have a uniform expansion

$$F(z, u) = c_1(u)(\rho - z)^{-\sigma(u)} H_1(\rho - z, u)(1 + o(1)), \qquad (9.63)$$

for some bivariate analytic function $H_1(w, u)$ with $H_1(0, u) = 1$, where $\sigma(u)$ is the algebraic branch that is a root of

$$J(\alpha, u) \equiv a_0(\rho, u)(\alpha)_{(r)} + a_1(\rho, u)(\alpha)_{(r-1)} + \cdots + a_r(\rho, u) = 0,$$

and coincides with $\sigma$ at $u = 1$. By singularity analysis, this entails

$$[z^n]F(z, u) = \frac{c_1(u)}{\Gamma(\sigma)} \rho^{-n} n^{\alpha_1(u) - 1}(1 + o(1)), \qquad (9.64)$$

uniformly for $u$ in a small neighbourhood of 1, with the error term being $\mathcal{O}(n^{-a})$ for some $a > 0$. Thus Theorem 9.9 applies and the limit law is Gaussian.

The crucial point in (9.63,9.64) is the uniform character of expansions with respect to $u$. This results from two facts: $(i)$ the solution to (9.62) may be specified by analytic conditions at a point $z_0$ such that $z_0 < \rho$ and there are no singularities of the equation between $z_0$ and $\rho$. $(ii)$ there is a suitable set of solutions with an analytic component in $z$ and $u$ and singular parts of the form $(\rho - z)^{-\alpha_j(u)}$, as results from the matrix theory of differential systems and majorant series. (This last point is easily verified if no two roots of the indicial equation differ by an integer; otherwise, see [23] for an alternative basis of solutions for $u$ near 1, $u \neq 1$.) □

EXAMPLE 28. *Node levels in quadtrees.* This example is taken from [23]. Quadtrees are one of the most versatile data structure for managing a collection of points in multidimensional space. They are based on a recursive decomposition similar to that of BSTs.

Here $d$ is the dimension of the data space. Let $f_{n,k}$ be the number of external nodes at level $k$ in a quadtree of size $n$ grown by random insertions, and let $F(z,u)$ be the corresponding BGF. Two integral operators play an essential rôle,

$$I\, g(z) = \int_0^z g(t)\,\frac{dt}{1-t} \qquad J\, g(z) = \int_0^z g(t)\,\frac{dt}{t(1-t)}.$$

The basic equation that reflects the recursive splitting process of quadtrees is then

$$F(z,u) = 1 + 2^d u J^{d-1} I\, F(z,u). \tag{9.65}$$

The integral equation (9.65) satisfied by $F$ then transforms into a differential equation of order $d$,

$$I^{-1} J^{1-d}\, F(z,u) = 2^d u F(z,u),$$

where

$$I^{-1} g(z) = (1-z)g'(z), \qquad J^{-1} g(z) = z(1-z)g'(z).$$

The linear ODE version of (9.65) has an indicial polynomial that is easily determined by examination of the reduced form of the ODE (9.65) at $z = 1$. There, one has

$$J^{-1} g(z) = I^{-1} g(z) - (z-1)^2 g'(z) \approx (1-z)g'(z).$$

Thus,

$$I^{-1} J^{1-d} (1-z)^{-\theta} = \theta^d (1-z)^{-\theta} + \mathcal{O}((1-z)^{-\theta+1}),$$

and the indicial polynomial is

$$J(\alpha, u) = \alpha^d - 2^d u.$$

In the univariate case, the root of largest real part is $\alpha_1 = 2$; in the bivariate case, we have

$$\alpha_1(u) = 2u^{1/d},$$

where the principal branch is chosen. Thus,

$$f_n(u) = \gamma(u) n^{\alpha_1(u)} (1 + o(1)).$$

By the combinatorial origin of the problem, $F(z,1) = (1-z)^{-2}$, so that the coefficient $\gamma(1)$ is nonzero. Thus, the conditions of the corollary are satisfied. The law is Gaussian in the limit, with mean and variance

$$\mu_n \sim \frac{2}{d}\log n, \qquad \sigma_n^2 \sim \frac{2}{d}\log n.$$

The same result applies to the cost of a random search, either successful or not, as shown in [23] by an easy combinatorial argument. □

**Nonlinear differential equations.** Though nonlinear differential equations do not obey a simple classification of singularities, there are a few examples in analytic combinatorics that can be treated by singularity perturbation methods. We detail here typical analysis of properties of binary search trees (BSTs), equivalently HOTs, that is taken from [20]. The Riccati equation involved reduces, by classical techniques, to a linear second order equation whose perturbation analysis is particularly transparent and akin to earlier analyses of ODEs. In this problem, the auxiliary parameter induces a movable singularity that directly resorts to the Quasi-Powers Theorem.

Example 29. *Paging of binary search trees.* Fix a "bucket size" parameter $b \geq 2$. Given a binary search tree $t$, its *b-index* is a tree that is constructed by retaining only those internal nodes of $t$ which correspond to subtrees of size $> b$. As a data structure, such an index is well-suited to "paging", where one has a two-level hierarchical memory structure: the index resides in main memory and the rest of the tree is kept in pages of capacity $b$ on peripheral storage, see for instance [47]. We let $\iota[t] = \iota_b[t]$ denote the size —number of nodes— of the $b$-index of $t$.

Like in Eq. (9.59), the bivariate generating function

$$F(z,u) := \sum_t \lambda(t) u^{\iota[t]} z^{|t|}$$

satisfies a Riccati equation that reflects the root decomposition of trees,

$$\frac{\partial}{\partial z}F(z,u) = uF^2(z,u) + (1-u)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right), \qquad F(0,u) = 1, \quad (9.66)$$

where the general quadratic relation (9.59) has to be corrected in its low order terms.

The GFs of moments are rational functions with a denominator that is a power of $(1-z)$, as results from differentiation at $u = 1$. Mean and variance follow:

$$\mu_n = \frac{2(n+1)}{b+2} - 1, \qquad \sigma_n^2 = \frac{2}{3}\,\frac{(b-1)b(b+1)}{(b+2)^2}\,(n+1).$$

(The result for the mean is well-known, refer to quantity $A_n$ in the analysis of quicksort on p. 122 of [41].)

Multiplying both sides of (9.66) by $u$ now gives an equation satisfied by $H(z,u) := uF(z,u)$,

$$\frac{\partial}{\partial z} H(z,u) = H^2(z,u) + u(1-u)\frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right),$$

that may as well be taken as a starting point since $H(z,u)$ is the bivariate GF of parameter $1+\iota_b$ (a quantity also equal to the number of external pages). The classical linearization transformation of Riccati equations,

$$H(z,u) = -\frac{X_z'(z,u)}{X(z,u)},$$

yields

$$\frac{\partial^2}{\partial z^2} X(z,u) + u(u-1)A(z)X(z,u) = 0, \qquad A(z) = \frac{d}{dz}\left(\frac{1-z^{b+1}}{1-z}\right), \tag{9.67}$$

with $X(0,u) = 1, X_z'(0,u) = -u$. By the classical existence theorem of Cauchy, the solution of (9.67) is an entire function of $z$ for each fixed $u$, as the linear differential equation has no singularity at a finite distance. Furthermore, the dependency of $X$ on $u$ is also everywhere analytic; see the remarks of [62, Sec. 24], for which a proof derives by inspection of the classical existence proof based on indeterminate coefficients and majorant series. Thus, $X(z,u)$ is actually an entire function of *both* complex variables $z$ and $u$. As a consequence, for any fixed $u = u_0$, the function $H(z,u_0)$ is a meromorphic function of $z$ whose coefficients are amenable to singularity analysis.

In order to proceed further, we need to prove that, in a sufficiently small neighbourhood of $u=1$, $X(z,u)$ has only one simple root, corresponding for $H(z,u)$ to a unique dominant and simple pole. This fact itself derives from general considerations surrounding the Preparation Theorem of Weierstrass: *in the vicinity of any point $(z_0,u_0)$ with $X(z_0,u_0)=0$, the roots of the bivariate analytic equation $X(z,u)=0$ are locally branches of an algebraic function.* Here, we have $X(z,1) \equiv 1-z$. Thus, as $u$ tends to 1, all solutions of $X(z,u)$ must escape to infinity except for one branch $\rho(u)$ that satisfies $\rho(1)=1$. By the nonvanishing of $X_u'(z,1)$ and the implicit function theorem, the function $\rho(u)$ is additionally an analytic function about $u = 1$.

The argument is now complete: for $u$ in a sufficiently complex neighbourhood of 1, we have a Quasi-Powers approximation,

$$[z^n]H(z,u) = \rho(u)^{-n-1}\left(1 + \mathcal{O}(K^{-n})\right),$$

for some fixed constant $K > 0$. The Gaussian limit results. □

As shown in [20], a similar analysis applies to patterns in binary search trees and heap–ordered trees. This is related to the analysis of local order patterns in permutations, for which gaussian limit laws have been obtained by Devroye [13] using extensions of the central limit theorem to weakly dependent random variables.

Similar displacements of singularity arise for node types in varieties of increasing trees, extending the case of HOTs that are binary. This is discussed in [7]. For instance, if $\phi(w)$ is the degree generator a family of increasing trees, the nonlinear ODE satisfied by the BGF of leaves is

$$\frac{\partial}{\partial z}F(z,u) = (u-1)\phi(0) + \phi(F(z,u)).$$

Whenever $\phi$ is a polynomial, there is a spontaneous singularity at some $\rho(u)$ that depends analytically on $u$. Thus, again the Quasi-Powers Theorem applies; see [7].

## 9.7 Local laws and large deviations

Under conditions similar to those of the Quasi-Powers Theorem, a cluster of conclusions may be drawn regarding densities of distributions and probabilities of large deviations from the mean. We examine here the occurrence of local limit laws, which corresponds to convergence of a discrete probability distribution to the *Gaussian density function* rather than convergence of distribution functions to the *Gaussian error function*, as we have seen so far. Such local laws hold very frequently, but their proofs require some sort of additional "smoothness" assumptions, either a combinatorial or analytic. Under assumptions of the Quasi-Powers Theorem, it is also possible to quantify precisely the exponential rate of decay for probabilities of rare events, far away from the center of the distribution. This section explores both aspects that fit well withing the general framework of quasi-powers. One aspects provides precise asymptotic information on values of the individual probabilities, especially near the mean; the other aspect quantifies the smallness of probabilities far away from the mean and, when conditions apply, it provides sharp quantitative versions of the concentration of distribution discussed at the beginning of this chapter.

### 9.7.1 Local limit laws

So far, we have examined the occurrence of continuous limit laws in the sense of convergence of distribution functions. Thus, a normalized $Y_n$ converges in distribution to $Y$, if

$$\Pr\{Y_n \le x\} \to \Pr\{Y \le x\}.$$

In the case of a Gaussian limit that arises from a sequence of discrete distributions of variables $X_n$ with mean and variance $\mu_n, \sigma_n^2$, such a property quantifies the probabilities over any nonempty interval scaled according to $\sigma_n$,

$$Pr\{\mu_n + a\sigma_n < X_n \le \mu_n + b\sigma_n\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx + o(1), \qquad (9.68)$$

for any $a, b$ with $a < b$. From there, it is however in general not possible to draw information on any individual probability,

$$p_{n,k} = \Pr\{X_n = k\},$$

by differencing, since the error terms in (9.68) will usually hide any nontrivial asymptotic information on individual $p_{n,k}$.

On the other hand, numerical examination of discrete probability distributions reveals that the histograms of the $p_{n,k}$ often assume a bell-shape profile in the asymptotic limit. For instance Figure 9.11, borrowed from our book [54], displays the $p_{n,k}$ that correspond to the Eulerian numbers. For a given value of $n$, the maximum probability $p_{n,k}$ is seen to occur "in the middle", near the mean, and to obey an approximate law,

$$p_{2n,n} \approx \frac{1.35}{\sqrt{2n}},$$

for values near $n = 60$. The standard deviation of the distribution is otherwise known to be $\sim \sqrt{n/12}$. Thus, the we expect an approximate formula of the form

$$p_{n,n/2+x\sqrt{n/12}} \approx \frac{C}{\sqrt{n}} e^{-x^2/2},$$

for integral values of the argument $k = n/2 + x\sqrt{n/12}$, with some constant $C$ about 1.35.

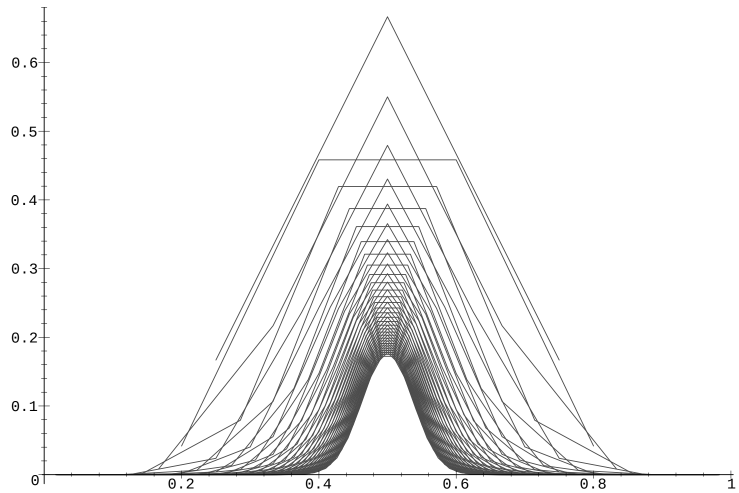


Figure 9.11: The histogram of the Eulerian distribution scaled to $(n+1)$ on the horizontal axis, for $n = 3\,.\,.\,60$. The distribution is seen to quickly converge to a bell-shaped curve corresponding to the Gaussian density $e^{-x^2/2}/(2\pi)^{1/2}$.

**Definition 9.6** *A sequence of discrete probability distributions,* $p_{n,k} = \Pr\{X_n = k\}$*, with mean* $\mu_n$ *and standard deviation* $\sigma_n$ *is said to obey a* local limit law of the Gaussian type *if, for some set* $S$ *of real numbers, and a sequence* $\epsilon \to 0$*,*

$$\sup \left| \sigma_n p_{n,\lfloor \mu_n + x\sigma_n \rfloor} - \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \right| \le \epsilon_n .$$

*The local limit law is said to hold on* $S$ *and the law is said to hold with speed of convergence* $\epsilon_n$*.*

When such a local limit law exists, it usually holds on arbitrary bounded intervals of the real line.

**Theorem 9.10 (Local limit law)** *Let* $X_n$ *be a sequence of nonnegative discrete random variables with probability generating function* $p_n(u)$*. Assume that* uniformly *in an annulus,*

$$1 - \epsilon \le u \le 1 + \epsilon, \qquad \epsilon > 0$$

*the PGFs satisfy*

$$p_n(u) = A(u)\,(B(u))^{\beta_n} \left(1 + \mathcal{O}(\frac{1}{\kappa_n})\right), \tag{9.69}$$

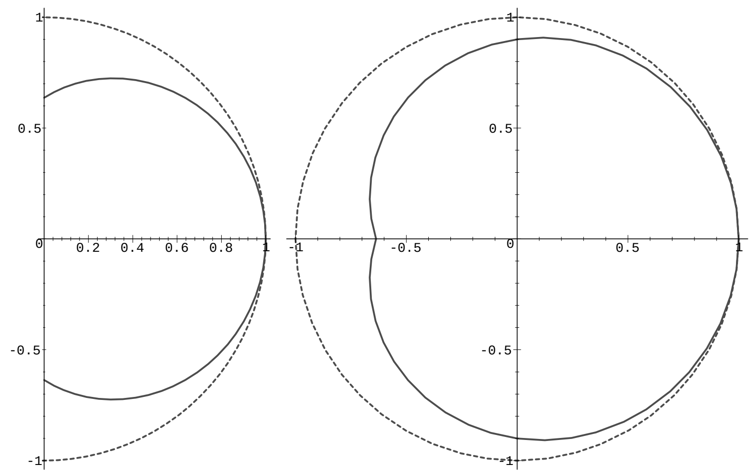


Figure 9.12: The values of the function $B(u)$ for the Eulerian distribution when $|u| = 1$ plotted by values (left) and by a polar plot of $|B(e^{i\theta})|$ on the ray of angle $\theta$ (right). (The dashed contours represent the relevant parts of the unit circle, for comparison.) The maximum is uniquely attained at $u = 1$, where $B(1) = 1$. This entails a local limit law for the Eulerian distribution.

*where* $A(u), B(u)$ *are analytic in the annulus and* $A(1) = B(1) = 1$, *Var*$(B(u)) = B''(1) + B'(1) - B'(1) \neq 0$. *Assume also that* $B(u)$ *attains uniquely in maximum on* $|u| = 1$ *at* $u = 1$*: for all* $v$*, with* $|v| = 1$ *and* $v \neq 1$*, one has* $|B(v)| < 1$.

*Under these conditions, the distribution of* $X_n$ *satisfies a local limit law of the Gaussian type on arbitrary bounded intervals of the real line.*

Note that the mean and variance of $X_n$ are given by Eq. (9.28).

Proof. A direct application of the saddle point method, as developed in Chapter 6. □

This theorem applies in particular to quasi-power expansions, whenever the dominant singularity $\rho(u)$, that is a perturbation of the dominant singularity $\rho$ of the univariate problem, is analytic at all points of $|u| = 1$ and uniquely attains its minimum at $u = 1$.

Example 30. *Local laws for sums of RV's.* The simplest application is to the binomial distribution, for which

$$B(u) = \frac{1+u}{2}.$$

In a precise technical sense, the local limit arises in the BGF,

$$F(z,u) = \frac{1}{1 - z(1+u)/2},$$

because the dominant singularity $\rho(u) = 2/(1+u)$ exists on the whole of the unit circle, $|u| = 1$, and it attains uniquely its minimum modulus at $u = 1$; accordingly, $B(u) = \rho(1)/\rho(u)$ is uniquely maximal at $u = 1$.

More generally, the theorem applies to any sum $S_n = T_1 + \cdots + T_n$ of independent, identically, random variables whose maximal span is equal to 1 and whose PGF is analytic on the unit circle. In that case, the BGF is

$$F(z,u) = \frac{1}{1 - zB(u)},$$

the PGF of $S_n$ is a pure power,

$$p_n(u) = B(u)^n,$$

and the fact that the minimal span of the $X_j$ is 1 entails that $B(u)$ attains uniquely its maximum at 1. Such cases have been known for a long time in probability theory. See Chapter 9 of [31]. □

At this stage, it is worth pointing an example *not* leading to a local law. Consider the binomial distribution restricted to even values,

$$p_{n,2k} = \frac{2}{2^n}\binom{n}{2k}, \qquad p_{n,2k+1} = 0.$$

The BGF is

$$F(z,u) = \frac{1}{1 - z(1+u)/2} + \frac{1}{1 - z(1-u)/2} - 1.$$

This has two poles,

$$\rho_1(u) = \frac{2}{1+u}, \qquad \rho_2(u) = \frac{2}{1-u},$$

and it is clearly not true that a single one dominates throughout the domain $|u| = 1$. Accordingly, the PGF satisfies

$$p_n(u) = (1+u)^n + (1-u)^n,$$

and no quasi-power law, with a unique analytic $B(u)$, holds uniformly for $u$ on the unit circle. In essence, a local limit law will be likely to hold when a PGF has a sharp peak near 1 and stays much smaller in modulus along the rest of the unit circle. In contrast, for the even binomial distribution, one has $p_n(1) = p_n(-1)$.

EXAMPLE 31. *Local law for the Eulerian distribution.* For Eulerian numbers, we have derived the approximate expression,

$$p_n(u) = B(u)^{-n-1} + \mathcal{O}(2^{-n}),$$

when $u$ is close enough to 1, with

$$B(u) = \rho(u)^{-1} = \frac{u-1}{\log u}.$$

The plot of the function $B(u)$ when $u$ varies over $|u| = 1$ is then displayed in Fig. 9.12.

This case requires in fact a minor extension of Theorem 9.10 since the principal determination of the logarithm cannot be extended to the whole of the unit circle, in particular at $u = -1$. However, it is easily realized that the quasi-power expansion holds with the possible exception of a small segment of the integration contour near $u = -1$. However, there, the integrand is anyway exponentially smaller than on the rest of the contour, and the proof of Theorem 9.10 is easily adjusted to cover such case.

From this enhanced argument, there results that a local limit law of the Gaussian type holds for the Eulerian distribution on any compact subset of the real line. □

With a similar care to be exercised regarding principal determinations and dominant singularities, many of our earlier analyses can be turned into local limit laws. What is needed is a dominant singularity $\rho(u)$ that yields the main asymptotic form of the PGF's on most of the unit disc and that achieves uniquely its minimum at 1, while the rest of the unit disc contributes negligibly. For instance, this covers the surjection distribution, for which

$$\rho(u) = \log(1 + u^{-1}), \qquad B(u) = \frac{\log 2}{\log(1 + u^{-1})},$$

leaves in general Catalan trees, where

$$B(u) = \frac{(1 + \sqrt{u})^2}{4},$$

or in binary Catalan trees.

The Stirling cycle distribution satisfies

$$p_n(u) = \frac{e^{(u-1)\log n}}{\Gamma(u)}\left(1 + \mathcal{O}(\frac{1}{n})\right).$$

This approximation remains uniform as long as $u$ avoids $-1$, but, there, $p_n(u)$ is small anyway (being $\mathcal{O}(n^{-2})$), so that again an extended form of Theorem 9.10 applies and a local limit law holds. The same argument applies to node levels in quadtrees of Example 28.

EXERCISE 26. Analyse asymptotically in detail the value of the peak of the Eulerian and Stirling cycle distributions.

EXERCISE 27. Discuss general conditions for occurrence of a local limit law in the supercritical sequence schema.

### 9.7.2 Large deviations

Moment inequalities constrain the shape of a distribution given its mean and variance. In particular, if $\sigma_n/\mu_n \to 1$, the concentration property holds. This property comes from Chebyshev's inequality according to which the probability of observing a value that deviates by more than $x$ standard deviations from the mean is $\mathcal{O}(x^{-2})$. Such general bounds, though sufficient to establish a concentration property, are much weaker than what holds under conditions of the quasi-power type, where the probabilities of deviation are in fact exponentially decreasing with in $x$.

Figure 9.13 displays the logarithms of the Eulerian distribution. As logarithms of probabilities are plotted, the distribution is seen to decay very rapidly away from the mean $\mu_n \sim n/2$. Consider for instance extreme cases. Clearly, there is a unique permutation that has a minimal number of rises, namely the fully sorted permutation with probability

$$p_{n,1} = \frac{1}{n!}.$$

In contrast, since $\mu_n \sim n/2$ and $\sigma_n^2 \sim n/12$, this extreme case is roughly at $x = \sqrt{3n}$ from the mean; thus, the Chebyshev inequalities only provides the very weak upper bound of $\sim \frac{1}{3n}$ for this extreme case. For $n = 40$, the Chebyshev upper bound on the probability is thus about 0.008 while the exact value $1/40!$ is of the order of $10^{-48}$.

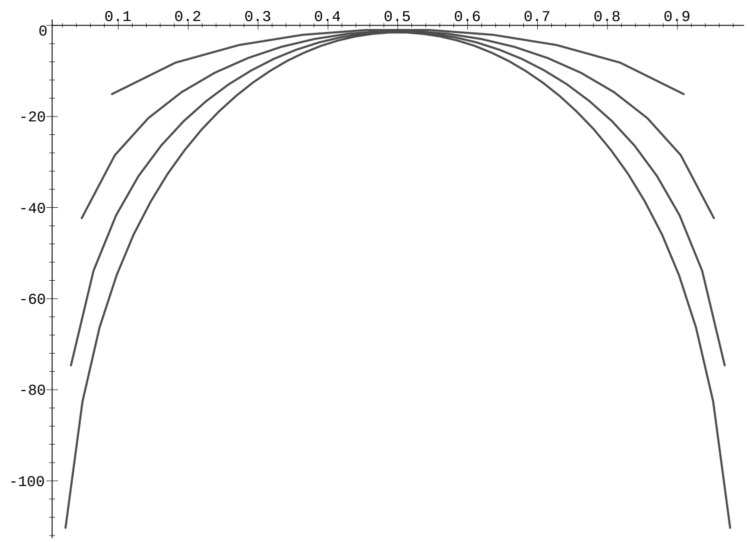


Figure 9.13: The quantities $\log p_{n,k}$ relative to the Eulerian numbers illustrate an extremely fast decay of the distribution away from the mean. Here, the diagrams corresponding to $n = 10, 20, 30, 40$ (top to bottom) are plotted. The common shape of the curves indicates a large deviation property.

Extensions of the quasi-power framework are once more well-suited to prove such exponentially small tails, as we now explain. It turns out that the ubiquitous functions $\rho(u), B(u)$ are directly related to *large deviation* estimates. Such estimates nicely supplement the already known limit laws, either central or local.

**Definition 9.7** *A sequence of discrete random variables* $\{X_n\}$ *with* $p_{n,k} = \Pr\{X_n = k\}$*, satisfies a* large deviation property *of type* $(\beta_n, W(x))$ *over the interval* $[x_0, x_1]$*, if for any* $x \in [x_0, x_1]$*,*

$$\frac{1}{\beta_n} \log p_{n,x\beta_n} \leq W(x) + \mathcal{O}(\beta_n^{-1}). \tag{9.70}$$

*The function* $W(x)$ *is called a* large deviation function *and* $\beta_n$ *is the* scaling factor*.*

The inequality (9.70) is *a priori* only meaningful if $x\beta_n$ is an integer, but it makes sense as well if it is taken that $p_{n,w} = 0$ for nonintegral values of $w$ and $\log 0 = -\infty$. Of course, the large deviation property is nontrivial when $W(x) \leq 0$, with $W(x)$ not identically 0.

**Theorem 9.11 (Large deviations)** *Consider a sequence of discrete random variables* $\{X_n\}$ *with PGF* $p_n(u)$. *Assume that there exist a functions* $A(u), B(u)$, *analytic in some interval* $[u_0, u_1]$ *with* $0 < u_0 < 1 < u_1$, *such that a quasi-power expansion holds,*

$$p_n(u) = A(u)B(u)^{\beta_n}\left(1 + \mathcal{O}(\kappa_n^{-1})\right), \tag{9.71}$$

*uniformly. Then the* $X_n$ *satisfy a large deviation property,*

$$\frac{1}{\beta_n}\log p_{n,x\beta_n} \leq W(x) + \mathcal{O}(\beta_n^{-1}), \tag{9.72}$$

*where the large deviation function* $W(x)$ *is given by*

$$W(x) = \min_{u\in[u_0,u_1]} \log\left(\frac{B(u)}{u^x}\right). \tag{9.73}$$

PROOF. The basic observation is that if $f(u) = \sum_n f_n u^n$ is an analytic function with nonnegative coefficients, then, for positive $u$,

$$f_k := [u^k]f(u) \leq \frac{f(u)}{u^k} \leq \min_{u>0} \frac{f(u)}{u^k}. \tag{9.74}$$

The first inequality holds for any positive $u$ in the disc of analyticity of $f(u)$; the second bound, with a similar condition, consists in taking the best possible value of $u$. See our earlier discussion of saddle point bounds.

The combination of the principle (9.74) applied to $f(u) = p_n(u)$, and of the assumption of the theorem (9.71) yields

$$\log p_{n,x\beta_n} \leq \beta_n \min_{u\in[u_0,u_1]} \log\left(\frac{B(u)}{u^x}\right) + \mathcal{O}(1).$$

Thus, a large deviation property holds with $W(x)$ given by (9.73). □

In general, the function $W(x)$ is computable from $B(u)$ and its derivatives. The minimum is attained at either an end-point or a point such that

$$\frac{d}{du}\left(\log B(u) - x\log u\right) = 0.$$

Let $\eta(x)$ be a value of $u \in [u_0, u_1]$ that cancels this derivative. Thus, $\eta$ is an inverse function of $uB'(u)/B(u)$,

$$\eta(x)\frac{B'(\eta(x))}{B(\eta(x)} = x.$$

Then, a large deviation function is

$$W(x) = \log B(\eta(x)) - x\log\eta(x). \tag{9.75}$$

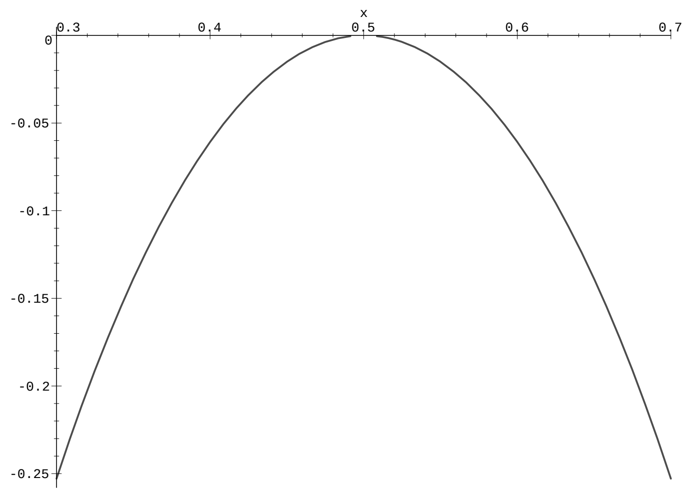


Figure 9.14: The large deviation function relative to the Eulerian distribution, for $u \in [0.3, 0.7]$.

EXERCISE 28. Prove similar types of bounds for the cumulative quantities

$$P_{n,k} = \sum_{j \le k} p_{n,j}, \qquad Q_{n,k} = \sum_{j \ge k} p_{n,j}.$$

EXAMPLE 32. *Large deviations for the Eulerian distribution.* In this case, the BGF has a unique dominant singularity for $u$ with $\epsilon < u < 1/\epsilon$, and any $\epsilon > 0$. Thus, there is a quasi-power expansion with

$$B(u) = \frac{(u-1)}{\log u},$$

on any interval $[\epsilon, 1/\epsilon]$. Then $\eta(x)$ is computable as the inverse function of

$$h(u) = \frac{u}{u-1} - \frac{1}{\log u}.$$

This function increases from 0 to 1 as $u$ increases from 0 to 1, so that the inverse function is well defined over any closed interval $[\epsilon, 1-\epsilon]$. The function $W(x)$ is then determined by (9.75); see Figure 9.14 for a plot of $W(x)$ that "explains" the data of Figure 9.13.

We find that

$$W(0.3) = W(0.7) = -0.252, \quad W(0.4) = W(0.6) = -0.061,$$

$$W(0.45) = W(0.55) = -0.015,$$

and $W(0.5) = 0$, as expected. For instance, the probability of deviating by 20% from the mean value $\mu_n \sim 0.5n$ is approximately $\exp(-0.061\,n)$. For $n = 100$, this upper bound is about $e^{-6.07}$, while the exact value of the probability gives $p_{100,60} \doteq e^{-8.58}$. In the same vein, there is probability less than $10^{-6}$ of deviating by 10% from the mean, when $n = 1,000$; the upper bound becomes less than $10^{-65}$, for $n = 10,000$, less than $10^{-653}$, for $n = 100,000$, etc. □

EXERCISE 29. Discuss the accuracy of large deviation bounds, by applying the saddle point method to the shifted distribution with PGF

$$\frac{p_n(uy)}{p_n(y)},$$

for suitable values of $y$. State conditions under which the bounds are optimal in their class. (Note: the technique is known as "shifting the mean".)

## 9.8 The saddle point schema

We shall be brief here, as the subject is excellently covered in Sachkov's book to which we refer for details. Entire functions and functions with a fast growth at their singularity do not in general lead to quasi-power expansions. As we known from univariate asymptotics (Chapter 6), the coefficient expansions involve a combination of large powers (that arise from the Cauchy kernel) and of the very fast singular behaviour of the function under consideration. Accordingly, bivariate asymptotic studies necessitate a perturbation of saddle point expansions. A framework more flexible than the Quasi-Powers Theorem is then needed.

Here, we base our brief discussion on a theorem taken from Sachkov's book [52].

**Theorem 9.12 (Generalized quasi-powers)** *Assume that the generating function $p_n(u)$ of a discrete random variable $X_n$ has a representation of the form*

$$p_n(u) = \exp\left(h_n(u)\right)\left(1 + o(1)\right),$$

*that holds uniformly, where each $h_n(u)$ is analytic in a fixed neighbourhood $\Omega$ of 1. Assume also the condition,*

$$\frac{h_n'''(u)}{(h_n'(1)+h_n''(1))^{3/2}} \to 0, \tag{9.76}$$

*uniformly for $u \in \Omega$. Then, the random variable*

$$X_n^* = \frac{X_n - h_n'(1)}{(h_n'(1)+h_n''(1))^{1/2}}$$

*converges in distribution to a normal law with parameters* $(0, 1)$.

PROOF. See [52, Sec. 1.4] for details. Set $\sigma^2 = h_n'(1) + h_n''(1)$, and expand the Laplace transform of $X_n$ at $t/\sigma$. This gives

$$h_n(e^{t/\sigma}) = h_n'(1)\frac{t}{\sigma} + (h_n'(1)+h_n''(1))\frac{t^2}{2\sigma} + o(1).$$

Thus, the Laplace transform of $X_n^*$ converges to the transform of a standard Gaussian. □

This theorem extends the quasi-power scheme. In effect, if

$$h_n(u) = \beta_n \log B(u) + A(u),$$

then the quantity (9.76) is $\mathcal{O}(\beta_n^{-1/2})$, uniformly. The application of this theorem to saddle point integrals is in principle routine, though the manipulation of asymptotic scales associated with expressions involving the saddle point value may become cumbersome. We detail here the case of singletons in random involutions for which the saddle point is an algebraic function of $n$ and $u$.

EXERCISE 30. Provide a metric version of the theorem, with error terms.

EXAMPLE 33. *Singletons in random involutions.* This example is again borrowed from Sachkov's book [52]. The BGF is

$$F(z,u) = \exp\left(zu + \frac{z^2}{2}\right).$$

The saddle point equation (see Chapter 6) is then

$$\left(\frac{d}{dz}uz + \frac{z^2}{2} - (n+1)\log z\right)_{z=\zeta} = 0.$$

This defines the saddle point $\zeta \equiv \zeta(n,u)$,

$$\begin{aligned}\zeta(n,u) &= -\frac{u}{2}+\frac{1}{2}\sqrt{4n+4+u^2}\\ &= \sqrt{n}-\frac{u}{2}+\frac{u^2+4}{8}\frac{1}{\sqrt{n}}+\mathcal{O}(n^{-1}),\end{aligned}$$

where the error term is uniform for $u$ near 1. By the saddle point formula, one has

$$[z^n]F(z,u)=\frac{1}{\sqrt{2\pi D(n,u)}}F(\zeta(n,u),u)\zeta(n,u)^{-n}.$$

The denominator is determined in terms of second derivatives, according to the classical saddle point formula (Chapter 6),

$$D(n,u)=\left(z^2\frac{\partial^2}{\partial z^2}+z\frac{\partial^2}{\partial z^2}\left[uz+\frac{z^2}{2}\right]\right)_{z=\rho},$$

and its main asymptotic order does not change when $u$ varies in a sufficiently small neighbourhood of 1,

$$D(n,u)=2n-u\sqrt{n}+\mathcal{O}(1),$$

again uniformly. Thus, the PGF of the number of singleton cycles satisfies

$$p_n(u)=\frac{F(\zeta(n,u),u)}{F(\zeta(n,1),1)}\left(\frac{\zeta(n,u)}{\zeta(n,1)}\right)^{-n}(1+o(1)),$$

uniformly, for $u$ near 1. This is of the form

$$p_n(u)=\exp\left(h_n(u)\right)(1+o(1)),$$

and local expansions then yield the centering constants

$$a_n := h_n'(1)=\sqrt{n}-\frac{1}{2}+\mathcal{O}(n^{-1/2}),\qquad b_n^2 := h_n'(1)+h_n''(1)=\sqrt{n}-1+\mathcal{O}(n^{-1/2}).$$

The theorem applies directly to this case and the variable

$$\frac{1}{b_n}(X_n-a_n)$$

is asymptotic to a standard normal.

A little care with the error terms in the asymptotic expansions shows that the mean and standard deviation $\mu_n, \sigma_n$ are asymptotic to $a_n, b_n$, respectively. Therefore, the number of singletons in a random involution of size $n$ has mean $\mu_n$ and standard deviation $\sigma_n$ that satisfy

$$\mu_n \sim n^{1/2}, \qquad \sigma_n \sim n^{1/4}.$$

This computation also determines the law of doubleton cycles and of all cycles, that are given by

$$\frac{1}{2}(n - X_n), \qquad \frac{1}{2}(n + X_n),$$

respectively. In particular, the number of doubleton cycles has average $\frac{1}{2}n - \frac{1}{2}n^{1/2}$. Thus, a random involution has a relatively small number of singleton cycles. □

> EXERCISE 31. Work out the law of the number of cycles of each length in random permutations all of which cycles are of length less than a fixed bound $b$.

EXAMPLE 34. *The Stirling partition numbers.* The numbers $\left\{ {n \atop k} \right\}$ correspond to the BGF

$$F(z, u) = \exp\left(u(e^z - 1)\right).$$

The saddle point $\zeta \equiv \zeta(n, u)$ is the positive root near $n/logn$ of the equation

$$\zeta e^{\zeta} = \frac{n+1}{u}.$$

The theorem applies and it is found that the Stirling partition distribution is asymptotically normal, with mean and variance that satisfy

$$\mu_n \sim \frac{n}{\log n}, \qquad \sigma_n^2 \sim \frac{n}{(\log n)^2}.$$

We refer once more to Sachkov's book for computational details. □

**Saddle point and functional equations.** The average-case analysis of the number of nodes in random digital trees or "tries" has been conducted in Chapter 7, using the Mellin transform technology. The corresponding distributional analysis is appreciably harder and due to Jacquet and Régnier [37]. A complete description is offered in Section 5.4 of Mahmoud's book which we follow. What is required is to analyse the BGF

$$F(z, u) = e^z T(z, u),$$

where the Poisson generating function $T(z,u)$ satisfies the nonlinear difference equation,

$$T(z,u) = uT^2(\frac{z}{2},u) + (1-u)(1+z)e^{-z}.$$

This equation is a direct reflection of the problem specification. At $u=1$, one has $T(z,1)=1$, $F(z,1)=e^z$. The idea is thus to analyse $[z^n]F(z,u)$ by the saddle point method.

The saddle point analysis of $F$ requires asymptotic information on $T(z,u)$ for $u=e^{it}$ (the original treatment of [37] is based on characteristic functions). There, the main idea is to "quais-linearize" the problem, setting

$$L(z,u) = \log T(z,u),$$

with $u$ a parameter. This function satisfies the approximate relation $L(z,u) \approx 2L(z/2,u)$, and a bootstrapping argument shows that, in suitable regions of the complex plane, $L(z,u) = \mathcal{O}(|z|)$, uniformly with respect to $u$. The function $L(z,u)$ is then expanded with respect to $u=e^{it}$ at $u=1$, *i.e.*, $t=0$, using a Taylor expansion, its companion integral representation, and the bootstrapping bounds. The moment-like quantities,

$$L_j(z) = \left. \frac{\partial^j}{\partial t^j} L(z,e^{it}) \right|_{t=0},$$

can be subjected to Mellin analysis for $j=1,2$ and bounded for $j \geq 3$. In this way, there results that

$$L(z,e^{it}) = L_1(z)t + \frac{1}{2}L_2(z)t^2 + \mathcal{O}(zt^3),$$

uniformly. The Gaussian law under a Poisson model immediately results from the continuity theorem of characteristic functions. Under the original Bernoulli model, the Gaussian limit follows from a saddle point analysis of

$$F(z,e^{it}) = e^z e^{L(z,e^{it})}.$$

An even more delicate analysis has been carried out by Jacquet and Szpankowski in [38]. It is relative to path length in digital search trees and involves the formidable non-linear bivariate difference-differential equation

$$\frac{\partial}{\partial z}F(z,u) = F^2(\frac{z}{2},u).$$

## 9.9 Extensions

We discuss here briefly the occurrence of non-Gaussian laws and multivariate extensions of the analytic framework.

**Special laws** Previous sections of this chapter have developed two basic paradigms for bivariate asymptotics:

— a "minor" singularity perturbation mode leading to discrete laws,

— a "major" singularity perturbation mode leading to continuous laws.

However, in both cases, the assumption is that the collection of singular expansions parametrized by the auxiliary variable all belong to a common analytic class and exhibit no sharp discontinuity.

Perhaps the simplest case of discontinuity in singular behaviour is the already discussed BGF,

$$F(z,u) = \frac{1-zu}{1-z},$$

where $u$ records the number of $a$'s in a random word of $a^\star b^\star$. The limit law is clearly the continuous uniform distribution over the interval $[0,1]$. From the point of view of the singular structure of $F(z,u)$, as a function of $z$, three distinct cases arise depending on the values of $u$:

- $u<1$: simple pole at $\rho(u)=1$;
- $u=1$: double pole at $\rho(1)=1$;
- $u>1$: simple pole at $\rho(u)=1/u$.

Thus both the singularity location at $\rho(u)$ and the singular exponent $\alpha(u)$ experience a nonanalytic transition at $u=1$. This arises from a "confluence" of two singular terms when $u=1$.

To visualize such cases, it is useful to introduce a simplified diagram representation: write $Z=\rho(u)-z$ and reduce the singular expansion to its dominant singular term $Z^{\alpha(u)}$. Then, the diagram representing $F(z,u)$ above is

| $u=1-\epsilon$ | $u=1$ | $u=1+\epsilon$ |
|---|---|---|
| $\rho(u)=1$ | $\rho(1)=1$ | $\rho(u)=1/u$ |
| $Z^{-1}$ | $Z^{-2}$ | $Z^{-1}$ |

A complete classification of such confluences and discontinuities is lacking and it is probably beyond reach given the vast diversity of real situations. We thus confine ourselves to discussing briefly a few representative examples.

EXAMPLE 35. *Arcsine law for unbiased random walks.* This problem is studies in detail by Feller [18, p. 94] who notes: "Contrary to intuition, the maximum accumulated gain is much more likely to occur towards the very beginning or the very end of a coin-tossing game than somewhere in the middle." In fact, if $X_n$ is the time of the first occurrence of the maximum, then in a random game (walk) of duration $n$,

$$\Pr\{X_n < xn\} \approx \frac{2}{\pi} \arcsin \sqrt{x},$$

a distribution function withe density

$$f(x) = \frac{1}{\pi\sqrt{x(1-x)}}.$$

The BGF results from the standard decomposition of positive walks. The walks decompose with respect to the maximum location in two walks constrained to be in a half-plane, with minor boundary conditions. The decomposition thus involves the GF of gambler's ruin sequences (Example 11),

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

Then the BGF of the location of the first maximum is

$$F(z,u) = \frac{1}{1-zuA(zu)} \cdot A(z) \cdot \frac{1}{1-zA(z)}.$$

Roughly, there is a sequence of steps ascending to the nonnegative maximum accompanied by "arches" (the left factor) followed by an excursion below than back to the maximum, followed by a sequence of descending steps with their companion arches. The dominant positive singularity is at $\rho(u) = 1/2$ if $u < 1$, and $\rho(u) = 1/(2u)$, if $u > 1$ and local expansions show that in each case, with $c_<(u), c_(u) >$ two computable functions,

$$F(z,u) \sim c_<(u)\frac{1}{\sqrt{1-2z}}, \qquad F(z,u) \sim c_>(u)\frac{1}{\sqrt{1-2z}},$$

while, at $u = 1$, all words are counted

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

Thus, the corresponding singularity diagram (negative singularities have a smaller weight and may be discarded in this discussion) is

| $u = 1 - \epsilon$ | $u = 1$ | $u = 1 + \epsilon$ |
|---|---|---|
| $\rho(u) = \frac{1}{2}$ | $\rho(1) = \frac{1}{2}$ | $\rho(u) = \frac{1}{2u}$ |
| $Z^{-1/2}$ | $Z^{-1}$ | $Z^{-1/2}$ |

Explicit expansions are available in this case and the arc-sine law results; see [18] for details. □

EXAMPLE 36. *Cyclic points in random mappings.* Another important example is that of the critical sequence construction. Consider a sequence construction with BGF

$$F(z,u) = \frac{1}{1 - uY(z)},$$

where $Y(z)$ has a dominant singular expansion

$$Y(z) = 1 - c\sqrt{1 - z/\rho} + \mathcal{O}(1 - z/\rho).$$

Then, the singularity diagram of $F(z,u)$ is immediate:

| $u = 1 - \epsilon$ | $u = 1$ | $u = 1 + \epsilon$ |
|---|---|---|
| $\rho(u) = \frac{1}{2}$ | $\rho(1) = -\frac{1}{2}$ | $\rho(u) = \frac{1}{2}$ |
| $Z^{1/2}$ | $Z^{-1/2}$ | $Z^{1/2}$ |

This situation occurs most notably when considering ordered forests of Cayley trees, or, equivalently, random mappings. In this case $F(z,u)$ is the exponential BGF with $u$ marking the number of cyclic points. This interesting situation has been extensively studied by Drmota and Soria [16, 17]. In such case, Hankel contours that underlie singularity analysis can be employed to derive a limit distribution, and saddle point related techniques yield a local limit law. The limit law of $X_n/\sqrt{n}$ is of the Rayleigh type $\mathcal{R}(\lambda)$, with density function,

$$g(x) = \lambda x e^{-\lambda x^2/2}.$$

□

EXAMPLE 37. *Path length in trees.* A final example is the distribution of path length in trees, that has been studied by Louchard, Takacs and others [44, 45, 58, 59]. The distribution is known *not* to be Gaussian as results from computation of the first few moments. In the case of general Catalan trees, the analysis reduces (*cf* Chapter 3) to that of the functional equation

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

This defines $F(z,u)$ as a formal continued fraction, which suggests setting

$$F(z,u) = \frac{A(z)}{B(z)},$$

the variable $u$ being viewed as a parameter. From the basic functional equation, there results

$$A(z) = B(zu), \qquad B(z) = B(zu) - zB(zu^2).$$

The functional equation for $B$ may now be solved by indeterminate coefficients:

$$B(z) = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{u^{n(n-1)} z^n}{(1-u)(1-u^2)\cdots(1-u^n)}.$$

This is a so–called "$q$-analogue". (The derivation above is in fact related to the classical Rogers-Ramanujan identities.) Because of the quadratic exponents involved, the functions $B(z)$ and $F(z,u)$ have radius of convergence 0 when $u > 1$, and are thus nonanalytic. In contrast, when $u < 1$, then $B(z,u)$ is an entire function of $z$, so that $F(z,u)$ is meromorphic in $z$. Hence the singularity diagram:

| $u = 1 - \epsilon$ | $u = 1$ | $u = 1 + \epsilon$ |
|---|---|---|
| $\rho(u) > \frac{1}{4}$ | $\rho(1) = \frac{1}{4}$ | $\rho(u) = 0$ |
| $Z^{-1}$ | $Z^{1/2}$ | — |

The limit law is an *Airy distribution*, that is related to Airy functions [44, 45, 58, 59]. By an analytical *tour de force*, Prellberg [51] has developed a method based on coalescing saddle points that permits us to extract such limits. As similar problems occur in relation to connectivity of random graphs, future years should see more applications of Prellberg's method. □

**Extremal parameters.** The focus of this chapter has been on parameters whose BGFs satisfy functional analytic relations. This leaves aside extremal parameters discussed in Section 3.6. Complex analytic methods are here again effective, but it is families of univariate GFs rather than BGFs that have to be studied.

Height leads to univariate relations,

$$Y_{h+1}(z) = \Phi(Y_h(z)),$$

where the functional $\Phi$ translates an admissible construction. The analysis requires uniform expansions, simultaneously, as $h \to \infty$ and $z \to \rho$, with $\rho$ the dominant singularity of $y_\infty$. This method has been particularly developed by Flajolet and Odlyzko in [24]; it leads to the existence of a limiting distribution for the height of binary trees and other simple trees expressed in terms of elliptic theta functions. (This generalizes the analysis of height in general Catalan trees for which explicit expressions are available; see Chapter 7.)

Maximum component sizes lead to functional relations

$$U_m(z) = \Phi(T_m[Y(z)]), \qquad T_m[Y(z)] = \sum_{j=0}^{m} Y_j z^j,$$

that involve a truncation operator $T_m$. The analysis is difficult, since moment GFs tend to have natural boundaries. However, Gourdon [32] has been able to analyse maximum and minimum parameters in general combinatorial schemas. The results often involve the extreme value distribution, as well as the Buchstab and Dickman functions of analytic function theory that describe the distribution of the smallest or largest prime factors in random natural numbers.

**Multivariate extensions.** There exist natural extensions of continuity theorems, both for PGFs and for integral transforms. Consider for instance the joint distribution of the numbers $\chi_1, \chi_2$ of singletons and doubletons in random permutations. Then, the parameter $\chi = (\chi_1, \chi_2)$ has a trivariate EGF

$$F(z, u_1, u_2) = \frac{\exp((u_1 - 1)z + (u_2 - 1)z^2/2)}{1 - z}.$$

Thus, the bivariate PGF satisfies, by meromorphic analysis,

$$p_n(u_1, u_2) = [z^n]F(z, u_1, u_2) \sim e^{(u_1-1)}\, e^{(u_2-1)/2}.$$

The joint distribution of $(\chi_1, \chi_2)$ is then a product of a Poisson(1) and a Poisson(1/2) distribution; in particular $\chi_1$ and $\chi_2$ are asymptotically independent. Such a fact results from an extension of the continuity theorem (Theorem 9.3) to multivariate PGF's that is proved by multiple Cauchy integration.

Consider next the joint distribution of $\chi = (\chi_1, \chi_2)$, where $\chi_j$ is the number of $j$-summands in a random integer composition. Each parameter individually obeys a limit Gaussian law, since the sequence construction is supercritical. The trivariate GF is

$$F(z, u_1, u_2) = \frac{1}{1 - z(1-z)^{-1} - (u_1 - 1)z - (u_2 - 1)z^2}.$$

By meromorphic analysis, a higher dimensional quasi-power approximation may be derived:

$$[z^n]F(z, u_1, u_2) \sim c(u_1, u_2)\rho(u_1, u_2)^{-n},$$

for some 3rd degree algebraic function $\rho(u_1, u_2)$. In such cases, multivariate versions of the continuity theorem for integral transforms can be applied. See the book by Gnedenko and Kolmogorov [31], and especially the treatment of Bender and Richmond in [5]. As a result, the joint distribution is, in the asymptotic limit, a bivariate Gaussian distribution. Such generalizations are typical and involve essentially no radically new concept, just natural technical adaptations.

A highly interesting approach to multivariate problems is that of *functional limit theorems.* There the goal is to characterize the joint distribution of a potentially infinite collections of parameters. The limit process is then a stochastic process. For instance, the joint distribution of all altitudes in random walks gives rise to Brownian motion. The joint distribution of all cycle lengths in random permutations is described explicitly by Cauchy's formula (Chapter 3), and DeLaurentis and Pittel [12] have also shown convergence to the standard Brownian motion process. A rather spectacular application of this context of ideas was provided in 1977 by Logan, Shepp, Vershik and Kerov [43, 61]. These authors show that the shape of the pair of Young tableaux [41] associated to a random permutation conforms, in the asymptotic limit and with high probability, to a deterministic trajectory defined as the solution to a variational problem. In particular, the width of a Young tableau associated to a permutation gives the length of the longest increasing sequence of the permutation. By specializing their results, the authors were able to show that the expected length in a random permutation of size $n$ is asymptotic to $2\sqrt{n}$, a long standing conjecture at the time.

## 9.10 Notes

This chapter is primarily inspired by the works of Bender and Richmond [2, 5, 6], Canfield [9], Flajolet, Soria, and Drmota [14, 15, 16, 17, 28, 29, 55] as well as Hwang [36].

Bender's seminal paper [2] initiated the study of bivariate analytic schemes that lead to Gaussian laws and the paper [2] may rightly be considered to be at the origin of the field. Canfield [9], building upon earlier works showed the approach to extend to saddle point schemas.

Tangible progress was next made possible by the development of the singularity analysis method [26]. Earlier works were mostly restricted to methods based on subtraction of singularities, as in [2], which is in particular effective for meromorphic cases. The extension to algebraic–logarithmic singularities was however difficult given that the classical method of Darboux does not provide for uniform error terms. In contrast, singularity analysis *does* apply to classes of analytic functions, since it allows for uniformity of estimates. The papers by Flajolet and Soria [28, 29] were the first to make clear the impact of singularity analysis on bivariate asymptotics. Gao and Richmond [30] were then able to extend the theory to cases where both a singularity and its singular exponent are allowed to vary.

From there, Soria developed considerably the framework of schemas in her doctorate [55]. Hwang extracted the very important concept of "quasi-powers" in his thesis [36] together with a wealth of properties like full asymptotic expansions, speed of convergence, and large deviations. Drmota established general existence conditions leading to Gaussian laws in the case of implicit, especially algebraic, functions [14, 15]. The "singularity perturbation" framework for solutions of linear differential equations first appears under that name in [23]. The presentation in this chapter is very liberally based on the survey paper [21]. Finally, the books by Sachkov, see [53] and especially [52], offer a modern perspective on bivariate asymptotics applied to classical combinatorial structures.

As pointed out in the introduction, the way combinatorial constructions induce limit laws via schemas based on a purely local perturbation of a singular structure is quite striking. Take for instance the principle that any fixed pattern occurs almost surely in a large random object and its number of occurrences is governed by Gaussian fluctuations. We have shown this property to hold true for strings, uniform tree models, and search trees. In a context that involves either a rational function, an algebraic function, or a solution to a nonlinear differential equation, it eventually reduces to a very simple property, a singularity that smoothly moves!

*I can see looming ahead one of those terrible exercises in probability where six men have white hats and six men have black hats and you have to work it out by mathematics how likely it is that the hats will get mixed up and in what proportion. If you start thinking about things like that, you would go round the bend. Let me assure you of that!*

—Agatha Christie
(*The Mirror Crack'd.* Toronto, Bantam Books, 1962.)

# Bibliography

[1] Arney, J., and Bender, E. D. Random mappings with constraints on coalescence and number of origins. *Pacific Journal of Mathematics 103* (1982), 269–294.

[2] Bender, E. A. Central and local limit theorems applied to asymptotic enumeration. *Journal of Combinatorial Theory 15* (1973), 91–111.

[3] Bender, E. A. Convex $n$-ominoes. *Discrete Mathematics 8* (1974), 219–226.

[4] Bender, E. A., and Kochman, F. The distribution of subword counts is usually normal. *European Journal of Combinatorics 14* (1993), 265–275.

[5] Bender, E. A., and Richmond, L. B. Central and local limit theorems applied to asymptotic enumeration II: Multivariate generating functions. *Journal of Combinatorial Theory, Series A 34* (1983), 255–265.

[6] Bender, E. A., Richmond, L. B., and Williamson, S. G. Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions. *Journal of Combinatorial Theory, Series A 35*, 3 (1983), 264–278.

[7] Bergeron, F., Flajolet, P., and Salvy, B. Varieties of increasing trees. In *CAAP'92* (1992), J.-C. Raoult, Ed., vol. 581 of *Lecture Notes in Computer Science*, pp. 24–48. Proceedings of the 17th Colloquium on Trees in Algebra and Programming, Rennes, France, February 1992.

[8] Billingsley, P. *Probability and Measure*, 2nd ed. John Wiley & Sons, 1986.

[9] Canfield, E. R. Central and local limit theorems for the coefficients of polynomials of binomial type. *Journal of Combinatorial Theory, Series A 23* (1977), 275–290.

[10] Comtet, L. *Advanced Combinatorics*. Reidel, Dordrecht, 1974.

[11] David, F. N., and Barton, D. E. *Combinatorial Chance*. Charles Griffin, London, 1962.

[12] DeLaurentis, J. M., and Pittel, B. G. Random permutations and brownian motion. *Pacific Journal of Mathematics 119*, 2 (1985), 287–301.

[13] Devroye, L. Limit laws for local counters in random binary search trees. *Random Structures and Algorithms 2*, 3 (1991), 302–315.

[14] Drmota, M. Asymptotic distributions and a multivariate Darboux method in enumeration problems. Manuscript, Nov. 1990.

[15] Drmota, M. Systems of functional equations. Manuscript, Aug. 1996. To appear in *Random Structures and Algorithms*, 1997.

[16] Drmota, M., and Soria, M. Marking in combinatorial constructions: Generating functions and limiting distributions. *Theoretical Computer Science 144*, 1–2 (June 1995), 67–99.

[17] Drmota, M., and Soria, M. Marking in combinatorial constructions: Images and preimages in random mappings. Preprint, 1997. To appear in *SIAM Journal on Discrete Mathematics*, 1997.

[18] Feller, W. *An Introduction to Probability Theory and its Applications*, third ed., vol. 1. John Wiley, 1968.

[19] Feller, W. *An Introduction to Probability Theory and Its Applications*, vol. 2. John Wiley, 1971.

[20] Flajolet, P., Gourdon, X., and Martínez, C. Patterns in random binary search trees. Research Report 2997, Institut National de Recherche en Informatique et en Automatique, Oct. 1996. 23 pages. To appear in *Random Structures & Algorithms*.

[21] Flajolet, P., Hwang, H.-K., and Soria, M. The ubiquitous Gaussian law in analytic combinatorics, 1997. In preparation.

[22] Flajolet, P., Kirschenhofer, P., and Tichy, R. F. Deviations from uniformity in random strings. *Probability Theory and Related Fields 80* (1988), 139–150.

[23] Flajolet, P., and Lafforgue, T. Search costs in quadtrees and singularity perturbation asymptotics. *Discrete and Computational Geometry 12*, 4 (1994), 151–175.

[24] Flajolet, P., and Odlyzko, A. The average height of binary trees and other simple trees. *Journal of Computer and System Sciences 25* (1982), 171–213.

[25] Flajolet, P., and Odlyzko, A. M. Random mapping statistics. In *Advances in Cryptology* (1990), J.-J. Quisquater and J. Vandewalle, Eds., vol. 434 of *Lecture Notes in Computer Science*, Springer Verlag, pp. 329–354. Proceedings of Eurocrypt'89, Houtalen, Belgium, April 1989.

[26] Flajolet, P., and Odlyzko, A. M. Singularity analysis of generating functions. *SIAM Journal on Discrete Mathematics 3*, 2 (1990), 216–240.

[27] Flajolet, P., Sipala, P., and Steyaert, J.-M. Analytic variations on the common subexpression problem. In *Automata, Languages, and Programming* (1990), M. S. Paterson, Ed., vol. 443 of *Lecture Notes in Computer Science*, pp. 220–234. Proceedings of the 17th ICALP Conference, Warwick, July 1990.

[28] Flajolet, P., and Soria, M. Gaussian limiting distributions for the number of components in combinatorial structures. *Journal of Combinatorial Theory, Series A 53* (1990), 165–182.

[29] Flajolet, P., and Soria, M. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics 114* (1993), 159–180.

[30] GAO, Z., AND RICHMOND, L. B. Central and local limit theorems applied to asymptotic enumerations IV: Multivariate generating functions. *Journal of Computational and Applied Mathematics 41* (1992), 177–186.

[31] GNEDENKO, B. V., AND KOLMOGOROV, A. N. *Limit Distributions for Sums of Independent Random Variables.* Addison-Wesley, 1968.

[32] GOURDON, X. *Combinatoire, Algorithmique et Géométrie des Polynômes.* PhD thesis, École Polytechnique, June 1996.

[33] HARARY, F., AND PALMER, E. M. *Graphical Enumeration.* Academic Press, 1973.

[34] HENRICI, P. *Applied and Computational Complex Analysis.* John Wiley, New York, 1977. 3 volumes.

[35] HILLE, E. *Analytic Function Theory.* Blaisdell Publishing Company, Waltham, 1962. 2 Volumes.

[36] HWANG, H.-K. *Théorèmes limites pour les structures combinatoires et les fonctions arithmetiques.* PhD thesis, École Polytechnique, Dec. 1994.

[37] JACQUET, P., AND RÉGNIER, M. Trie partitioning process: Limiting distributions. In *CAAP'86* (1986), P. Franchi-Zanetacchi, Ed., vol. 214 of *Lecture Notes in Computer Science*, pp. 196–210. Proceedings of the 11th Colloquium on Trees in Algebra and Programming, Nice France, March 1986.

[38] JACQUET, P., AND SZPANKOWSKI, W. Asymptotic behavior of the Lempel-Ziv parsing scheme and digital search trees. *Theoretical Computer Science 144*, 1–2 (1995), 161–197.

[39] KARLIN, S., AND TAYLOR, H. *A First Course in Stochastic Processes*, second ed. Academic Press, 1975.

[40] KNOPFMACHER, J. *Abstract Analytic Number Theory.* Dover, 1990.

[41] KNUTH, D. E. *The Art of Computer Programming*, vol. 3: Sorting and Searching. Addison-Wesley, 1973.

[42] KNUTH, D. E. *The Art of Computer Programming*, 2nd ed., vol. 2: Seminumerical Algorithms. Addison-Wesley, 1981.

[43] LOGAN, B. F., AND SHEPP, L. A. A variational problem for random Young tableaux. *Advances in Mathematics 26* (1977), 206–222.

[44] LOUCHARD, G. The Brownian excursion: a numerical analysis. *Computers and Mathematics with Applications 10*, 6 (1984), 413–417.

[45] LOUCHARD, G. Kac's formula, Lévy's local time and Brownian excursion. *Journal of Applied Probability 21* (1984), 479–499.

[46] LOUCHARD, G. Probabilsitic analysis of column-convex and directed diagonally-convex animals. Preprint, 1996.

[47] MAHMOUD, H. *Evolution of Random Search Trees.* John Wiley, New York, 1992.

[48] MCDIARMID, C. J. H., AND HAYWARD, R. B. Large deviations for quicksort. *Journal of Algorithms 21* (1996), 476–507.

[49] MEIR, A., AND MOON, J. W. On random mapping patterns. *Combinatorica 4* (1984), 61–70.

[50] Otter, R. The number of trees. *Annals of Mathematics 49*, 3 (1948), 583–599.

[51] Prellberg, T. Uniform $q$-series asymptotics for staircase polygons. *Journal of Physics A: Math. Gen. 28* (1995), 1289–1304.

[52] Sachkov, V. *Verojatnostnye Metody v Kombinatornom Analize.* Nauka, Moscow, 1978.

[53] Sachkov, V. N. *Combinatorial Methods in Discrete Mathematics*, vol. 55 of *Encyclopedia of Mathematics and its Applications.* Cambridge University Press, 1996.

[54] Sedgewick, R., and Flajolet, P. *An Introduction to the Analysis of Algorithms.* Addison-Wesley Publishing Company, 1996.

[55] Soria-Cousineau, M. *Méthodes d'analyse pour les constructions combinatoires et les algorithmes.* Doctorat ès sciences, Université de Paris–Sud, Orsay, July 1990.

[56] Stepanov, V. E. On the distribution of the number of vertices in strata of a random tree. *Theory of Probability and Applications 14*, 1 (1969), 65–78.

[57] Steyaert, J.-M., and Flajolet, P. Patterns and pattern-matching in trees: an analysis. *Information and Control 58*, 1–3 (July 1983), 19–58.

[58] Takacs, L. On a probability problem connected with railway traffic. *Journal of Applied Mathematics and Stochastic Analysis 4*, 1 (1991), 1–27.

[59] Takács, L. The asymptotic distribution of the total heights of random rooted trees. *Acta Scientifica Mathematica (Szeged) 57* (1993), 613–625.

[60] Titchmarsh, E. C. *The Theory of Functions*, second ed. Oxford University Press, 1939.

[61] Vershik, A. M., and Kerov, S. V. Asymptotics of the Plancherel measure of the symmetric group and the limiting form of Young tables. *Soviet Mathematical Doklady 18* (1977), 527–531.

[62] Wasow, W. *Asymptotic Expansions for Ordinary Differential Equations.* Dover, 1987. A reprint of the John Wiley edition, 1965.

# Contents