THE AVERAGE CASE ANALYSIS OF ALGORITHMS

PHILIPPE FLAJOLET AND ROBERT SEDGEWICK
ROCQUENCOURT & PRINCETON

Saddle Point Asymptotics

October 22, 1994

&&&&&&&&&&&&

THE AVERAGE CASE ANALYSIS OF ALGORITHMS: *Saddle Point Asymptotics*

Philippe Flajolet[1] & Robert Sedgewick[2]

Foreword

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. The series should comprise 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;
9. Multivariate Asymptotics and Combinatorial Distributions.

Chapters 1–3 have been issued as INRIA Research Report 1888 ("The Average Case Analysis of Algorithms: Counting and Generating Functions", 116 pages, 1993). Chapters 4–5 as INRIA Research Report 2026 ("The Average Case Analysis of Algorithms: Complex Asymptotics and Generating Functions", 100 pages, 1993). The present report corresponds to Chapter 6 of the series.

---

[1]Algorithms Project, INRIA Rocquencourt, F-78153 Le Chesnay (France).
Work of this author has been supported in part by the ESPRIT III Basic Research Action Programme of the E.C. under contract ALCOM II (#7141).
[2]Department of Computer Science, Princeton University, Princeton, New-Jersey 08544 (USA)

Chapter 6

Saddle Point Asymptotics

Like a lazy hiker, the path crosses the ridge at a low point;
but unlike the hiker, the best path takes the steepest ascent to the ridge.
[$\cdots$] The integral will then be concentrated in a small interval.

— Daniel Greene and Donald Knuth [12, sec. 4.3.3]

A *saddle point* of a surface is a point reminiscent of the inner part of a saddle or of a geographical pass between two mountains. If the surface represents the modulus of an analytic function, saddle points are simply determined as the zeros of the derivative of that function.

In order to estimate *complex integrals* with an analytic integrand, it is a good strategy to take as a contour of integration a line that "crosses" one or several of the saddle points of the function. When applied to integrals depending on a large parameter —as is the case for Cauchy integrals giving coefficients of generating functions– this often provides useful asymptotic information.

The saddle point method, whenever it applies, leads to asymptotic estimates or even full asymptotic expansions depending on the nature of the problem. Its principle is to use a saddle point crossing path, then estimate the integrand locally near this saddle point (at which point the integrand achieves its maximum), and deduce by termwise integration an asymptotic expansion of the integral itself. Some sort of "concentration" property is required to ensure that this local contribution captures the essential part of the integral. A simplified form of the method, called the saddle point bound, permits to derive useful upper bounds by simply using trivial bounds on a saddle point crossing path.

The method applies well to rapidly varying functions, like entire functions and functions with singularities at a finite distance that exhibit some form of exponential growth. It is also suitable for the analysis of coefficients of large powers of some fixed function.

Applications are given here to the asymptotics of the Bell numbers, Stirling's formula and the asymptotic counting of integer partitions.

6.1 Introduction

Saddle point analysis is a general method suited to the estimation of integrals of analytic functions $F(z)$,

$$I = \int_A^B F(z)\, dz, \tag{6.1}$$

where $F(z) \equiv F_n(z)$ involves some large parameter $n \to +\infty$. The method is appropriate when the integrand F is subject to rather violent variations, typically when there occurs in it some exponential or some fixed function raised to a large power (for instance, n).

This situation covers a large number of Cauchy coefficient integrals of the form

$$g_n \equiv [z^n]g(z) = \frac{1}{2i\pi} \oint g(z)\, \frac{dz}{z^{n+1}}. \tag{6.2}$$

Two simple examples that we shall discuss throughout this introductory section are

$$J_n = \frac{1}{2i\pi} \oint (1+z)^{2n}\, \frac{dz}{z^{n+1}}, \qquad K_n = \frac{1}{2i\pi} \oint e^z\, \frac{dz}{z^{n+1}}, \tag{6.3}$$

giving the central binomial coefficient $\binom{2n}{n}$ and the inverse factorial $(n!)^{-1}$ respectively. In that case, with reference to Eq. (6.1), one can think of the end points A and B as coinciding and taken somewhat arbitrarily on the negative real axis while the contour naturally has to encircle the origin once and counter–clockwise.

Saddle point bounds. Considering the general form (6.1), we let $\mathcal{C}$ be a contour joining A and B and taken in a domain of the complex plane where $F(z)$ is analytic. By standard inequalities, we have

$$|I| \le ||\mathcal{C}|| \cdot \max_{z\in\mathcal{C}} |F(z)|, \tag{6.4}$$

with $||\mathcal{C}||$ representing the length of $\mathcal{C}$. This is the usual *trivial bound* from integration theory.

For an analytic integrand F with A and B inside the domain of analyticity, there is an infinite class P of acceptable paths to choose from, all in the analyticity domain of F. Thus, we may write

$$|I| \leq \min_{\mathcal{C} \in P} \left[||\mathcal{C}|| \cdot \max_{z \in \mathcal{C}} |F(z)| \right], \tag{6.5}$$

where the minimum is taken over all paths $\mathcal{C} \in P$. Broadly speaking, a bound of this type is called a *saddle point bound*. The irruption of saddles in the story of complex integrals is to be explained shortly.

Notice that the optimization problem need not be solved exactly, as any approximate solution to (6.5) still furnishes a valid upper bound because of the universal character of the trivial bound (6.4). In the particular case of Cauchy coefficient integrals (6.2) where $F(z) = g(z)z^{-n-1}$, it is convenient (and usually sufficient) to restrict attention to contours that are circles centered at the origin. In that case, the trivial bound for $g_n = [z^n]g(z)$ becomes

$$|g_n| \leq \frac{1}{R^n} \max_{|z|=R} |g(z)| \tag{6.6}$$

and the saddle point bound for this class of contours reads

$$|g_n| \leq \min_R \left[\frac{1}{R^n} \max_{|z|=R} |g(z)| \right]. \tag{6.7}$$

If in addition $g(z)$ has positive coefficients, the maximum in (6.7) is attained on the positive real line and

$$g_n \leq \min_R \frac{g(R)}{R^n},$$

where the minimum may be determined by cancellation of a derivative (see Theorem 6.1).

The quality of the saddle point bound may be checked by applying it to J_n and K_n where it provides for the inequalities valid for all n,

$$J_n \leq 2^n, \qquad K_n \leq \frac{e^n}{n^n},$$

as the minimum is achieved for $R = 1$ and $R = n$ respectively. These bounds are actually surprisingly good when compared to the corresponding precise asymptotic forms

$$J_n \sim \frac{2^n}{\sqrt{\pi n}}, \qquad K_n \sim \frac{e^n}{\sqrt{2\pi n}}. \tag{6.8}$$

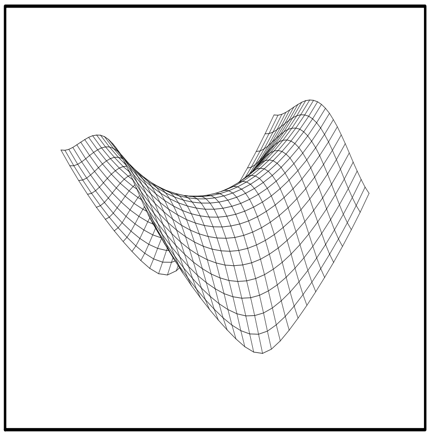

Figure 6.1: A region of a surface with a saddle point in its center.

It is seen on these two examples that the saddle point bounds catch the proper exponential growths, being off only by a factor of $\mathcal{O}(n^{-1/2})$. This is in fact a common phenomenon well explained by the saddle point method.

Saddle point paths. Understanding the general nature of contours satisfying minmax properties of the form (6.5) requires an incursion into general properties of the surface that represents the modulus of an analytic function. Let $F(z)$ denote a function analytic in an open set Ω. The quantity $|F(z)|$ as a function of z defines a surface Σ with specific properties.

- The surface Σ has no maxima; it has no minima except for isolated zeros of $F(z)$.
- A point z_0 such that $F(z_0) \neq 0$, $F'(z_0) \neq 0$ is called an ordinary point. It is traversed by two curves, the level curve and the steepest descent curve that cross each other at a right angle.
- A point z_0 such that $F(z_0) \neq 0$, $F'(z_0) = 0$ is called a *saddle point*. The name is due to the particular shape of Σ around that point. (See Fig. 6.1.)

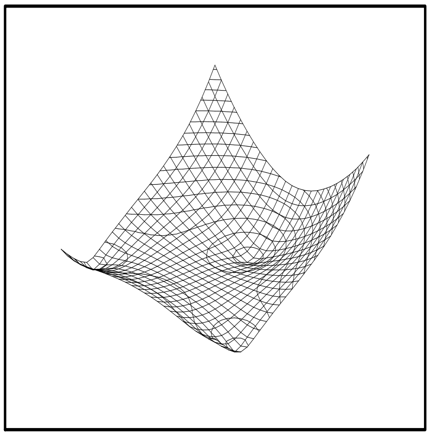

Figure 6.2: A plot of $|1 + z - z^2 + z^3|$ as a function of z reveals ordinary points, a zero, and a saddle point.

The assertions above are all consequences of the analyticity of $F(z)$ at z_0 which strongly constrains $|F(z)|$:

$$F(z) \approx F(z_0) + F'(z_0)(z - z_0) + \frac{1}{2}F''(z_0)(z - z_0)^2.$$

They are to be examined in greater detail in Section 6.3. Figure 6.2 shows the typical landscape induced by the modulus of an analytic function.

A path $\mathcal{C}$ that traverses a saddle point by connecting two points at a lower altitude on the surface and by following two steepest descent lines across the saddle point is clearly a local minimum for the path functional

$$\Phi(\mathcal{C}) = \max_{z \in \mathcal{C}} |F(z)|,$$

as neighbouring paths have a higher maximum. Such a path is called a *saddle–point path* or a *steepest descent path*. Thus, the search for a path realizing the minimum

$$\min_{\mathcal{C}} \left[\max_{z \in \mathcal{C}} |F(z)| \right],$$

naturally leads to considering saddle points and saddle-point paths.

Borrowing a metaphor of De Bruijn [5], the situation may be described as follows. Estimating a path integral is like estimating the difference of altitude between two villages in a mountain range. If the two villages are in different valleys, then the least effort[1] (the greatest accuracy in integral evaluations) should result from following paths that cross boundaries between valleys at passes, *i.e.*, saddle points.

The saddle point method. Given a fixed contour $\mathcal{C}$ traversing a saddle point along its axis, the saddle point corresponds locally to a maximum of the integrand along the path. It is therefore natural to expect that a small neighbourhood of the saddle point might provide the dominant contribution to the integral. The saddle point method is applicable precisely when this is the case and when this dominant contribution can be estimated by local expansions.

To proceed, it is convenient to set $F(z) = e^{f(z)}$and consider

$$I = \int_{\mathcal{C}} e^{f(z)}\, dz,$$

where $f(z)$ like $F(z)$ involves some large parameter. We assume that $\mathcal{C}$ connects the end points A and B and is a path traversing a unique saddle point $z_0 \in \mathcal{C}$. Thus, we also have the saddle point equation $F'(z_0) = 0$ or equivalently

$$f'(z_0) = 0,$$

and by assumption, $|e^{f(A)}| < |e^{f(z_0)}|$, $|e^{f(B)}| < |e^{f(z_0)}|$. The saddle point method is based on splitting $\mathcal{C}$ as $\mathcal{C} = \mathcal{C}^{(0)} \cup \mathcal{C}^{(1)}$, where $\mathcal{C}^{(0)}$ contains z_0, and estimating separately the integrals $\int_{\mathcal{C}^{(0)}}$ and $\int_{\mathcal{C}^{(1)}}$.

In order for the method to work, conflicting requirements must be satisfied. First, we assume that, by design, $\mathcal{C}^{(0)}$ captures most of the contribution to the integral; this forces $\mathcal{C}^{(0)}$ to be a sufficiently large portion of $\mathcal{C}$ (though usually $||\mathcal{C}^{(0)}||/||\mathcal{C}|| \to 0$) and requires $F(z)$ to be small compared to $F(z_0)$ on $\mathcal{C}^{(1)}$. Next we assume that $f(z)$ can be expanded locally near the saddle point z_0 in such a way that only the first two terms matter asymptotically. This forces $\mathcal{C}^{(0)}$ to be sufficiently small and implies

$$\int_{\mathcal{C}^{(0)}} e^{f(z)}\, dz \sim e^{f(z_0)} \int_{\mathcal{C}^{(0)}} \exp\left(\frac{1}{2} f''(z_0)(z-z_0)^2\right)\, dz.$$

By construction, the contour $\mathcal{C}^{(0)}$ is such that $f''(z_0)(z-z_0)^2$ is negative for $z \in \mathcal{C}^{(0)}$, in accordance with the requirement that z_0 is a local maximum.

[1]This is precisely what road networks do!

Then, up to normalizations and up to minor tail contributions, the integral on the right can be approximated by a complete Gaussian integral

$$\int_{-\infty}^{+\infty} e^{-ax^2/2}\,dx = \sqrt{\frac{2\pi}{a}}.$$

This discussion can now be summarized.

Saddle point principle. *For well-behaved functions $f(z) \equiv f_n(z)$,*

$$\frac{1}{2i\pi}\int_A^B e^{f(z)}\,dz \sim \frac{e^{f(z_0)}}{\sqrt{2\pi|f''(z_0)|}},$$

where z_0 is a saddle point of $f(z)$ on a path from A to B:

$$f'(z_0) = 0.$$

In asymptotic problems the saddle point z_0 usually depends on n.

As an illustration, a blind application of this principle to J_n and K_n gives for $f(z)$

$$2n\log(1+z) - (n+1)\log z, \qquad z - (n+1)\log z,$$

for the saddle point equation,

$$\frac{2n}{1+z} - \frac{n+1}{z} = 0, \qquad 1 - \frac{n+1}{z} = 0,$$

and for z_0,

$$1 + \mathcal{O}(\frac{1}{n}), \qquad n+1,$$

which results in the correct asymptotic equivalents (6.8) for J_n and K_n.

In the sequel, we make use of these general principles but focus on the particular case of Cauchy coefficient integrals for generating functions with positive coefficients. The geometry of the problem is simpler in that case and it suffices to consider, as integration contours, circles with proper radii centered at the origin and passing through saddle points.

6.2 Saddle point bounds

This section implements the principles set forth in the introduction concerning saddle point bounds and details two situations that are important for applications: coefficient extraction for either a fixed function or powers of a fixed function.

Cauchy coefficient integrals. The saddle point bounds are especially easy to apply in the case of coefficient integrals relative to analytic functions with positive coefficients.

Theorem 6.1 (Saddle point bounds for Cauchy integrals) *Let $f(z)$ be a function analytic at 0 that has nonnegative Taylor coefficients and radius of convergence ρ, $0 < \rho \le +\infty$. Assume that $f(0) \neq 0$ and that $f(z)$ is not a polynomial.*

(i). The coefficients of f satisfy

$$f_n \equiv [z^n]f(z) \le \inf_{0<r<\rho} \frac{f(r)}{r^n} \tag{6.9}$$

(ii). If in addition $f(\rho) = +\infty$, then for n large enough, the equation

$$\zeta \frac{f'(\zeta)}{f(\zeta)} = n \tag{6.10}$$

has a unique positive solution $\zeta = \zeta(n)$ in $]0, \rho[$, and

$$f_n \le f(\zeta(n))(\zeta(n))^{-n}. \tag{6.11}$$

Proof. (i). By trivial bounds, any analytic function satisfies

$$|f_n| \le \frac{1}{2\pi} \left| \int_{|z|=r} f(z) \frac{dz}{z^{n+1}} \right| \le \frac{1}{r^n} \max_{|z|=r} |f(z)|.$$

Since $f(z)$ has nonnegative coefficients, the triangular inequality applies and

$$\max_{|z|=r} |f(z)| \le f(r) \qquad \text{so that} \qquad f_n \le \frac{f(r)}{r^n}. \tag{6.12}$$

This bound is valid for all $r < \rho$, and part (i) of the statement follows.

(ii). The equation giving ζ is

$$\frac{d}{d\zeta} f(\zeta)\zeta^{-n} = 0,$$

which determines a possible local extremum of the bound (6.9). We observe that $f(r)r^{-n}$ is decreasing near 0 and increasing near ρ. Thus it has at least one minimum in the interval $[0, \rho]$.

We next prove that its derivative only vanishes once, so that the zero of the derivative determines the position of that minimum. Unicity of the solution to the saddle point equation is guaranteed by the positivity of

$$\frac{d^2}{dz^2} f(z) z^{-n} = \frac{z^2 f''(z) - 2nz f'(z) + n(n+1) f(z)}{z^{n+2}}, \tag{6.13}$$

for $z > 0$ where the numerator,

$$\sum_k (n+1-k)(n-k) f_k z^k,$$

has only nonnegative coefficients. Thus $f(z)z^{-n}$ is convex for $z \in (0, \rho)$ and unicity of the minimum results. □

For Theorem 6.1, a direct argument based on the nonnegativity of coefficients is of course possible as $f_n r^n \le f(r)$. However the approach taken in the proof above has the advantage of pointing the way to a number of extensions of the method, to functions with negative coefficients and to the complete saddle point method.

Equation (6.10) gives, under mild restrictions, the solution to the minimization problem of (6.9). However the general character of (6.12) that is valid for any (legal) r permits to replace $\zeta(n)$ by an approximate value $\hat{\zeta}(n)$ while still retaining the validity of inequality (6.11). The tightness of the bound so obtained will naturally depend on the quality of the approximation, $\zeta(n) - \hat{\zeta}(n)$ as well as on the growth pattern of $f(z)$.

We have already indicated the application to the inverse factorial in the introduction. A few more applications follow.

1. Involutions. Let I_n be the number of involutions of $[1..n]$, that is the number of permutations τ such that τ^2 is the identity permutation. The egf of I_n is

$$I(z) = e^{z+z^2/2}.$$

From Theorem 6.1, the value of ζ is the positive root of $\zeta(1+\zeta) = n+1$, hence

$$\zeta = \frac{-1+\sqrt{4n+1}}{2} = \sqrt{n} - \frac{1}{2} + \frac{1}{8\sqrt{n}} + \mathcal{O}(n^{-3/2}).$$

By routine asymptotic computations, the bound (6.11) becomes

$$\frac{I_n}{n!} \le e^{-1/4} n^{-n/2-1/2} e^{n/2+\sqrt{n}} (1 + o(1)).$$

Notice that if we use the approximate saddle point value, $\hat{\zeta}(n) = \sqrt{n}$, we only lose the factor of $e^{-1/4} \cong 0.78$.

In agreement with the discussion in the introduction these bounds are quite good, and we shall see later that they only off by a factor of $\mathcal{O}(n^{1/2})$ from the true asymptotic form of I_n given in [17].

2. Bell numbers. The number of partitions of a set of n elements defines the Bell number B_n and one has

$$B_n = n! e^{-1} [z^n] f(z) \qquad \text{where} \qquad f(z) = e^{e^z}.$$

The saddle point equation for $f(z)$ is

$$\zeta e^{\zeta} = n.$$

This famous equation admits an asymptotic solution obtained by iteration ("bootstrapping")

$$\zeta(n) = \log n - \log\log n + \frac{\log\log n}{\log n} + \mathcal{O}\left(\frac{\log^2 \log n}{\log^2 n}\right),$$

and the saddle point bound reads

$$B_n \le n! \frac{e^{e^{\zeta}} - 1}{\zeta^n}.$$

With the approximate solution $\hat{\zeta}(n) = \log n$, this provides the upper bound

$$B_n < n! \frac{e^{n-1}}{(\log n)^n}.$$

In particular, there are much fewer set partitions than permutations, the ratio being roughly $e^{-n \log\log n}$.

3. A large singular function. Define

$$f(z) = \exp(\frac{z}{1-z}) \qquad \text{and} \qquad f_n = [z^n] f(z).$$

The saddle point equation is

$$\frac{\zeta}{(1-\zeta)^2} = n.$$

By the observation following Theorem 1, we need only solve it approximately the resulting bounds being still valid. Here it is natural to take

$$\hat{\zeta}(n) = 1 - \frac{1}{\sqrt{n}},$$

leading to

$$f_n \leq e^{-1/2} e^{2\sqrt{n}}(1 + o(1)).$$

which is only off the true value by a factor of $\mathcal{O}(n^{3/4})$.

4. Integer partitions. Let p_n denote the number of integer partitions of n, with OGF

$$p(z) = \prod_{j \geq 1} \frac{1}{1 - z^j}.$$

A form more amenable to bounding derives from the exp–log reorganization,

$$\begin{aligned} p(z) &= \exp \sum_{n=1}^{\infty} \log(1 - z^n)^{-1} \\ &= \exp\left(\frac{z}{1-z} + \frac{1}{2}\frac{z^2}{1-z^2} + \frac{1}{3}\frac{z^3}{1-z^3} \cdots\right) \\ &\leq \exp\left(\frac{\pi^2}{6(1-z)}\right), \end{aligned}$$

where the last equation results from the elementary bound valid for $z \in (0, 1)$,

$$\frac{z^k}{1 - z^k} \leq \frac{1}{k(1 - z)},$$

together with the identity $\sum k^{-2} = \pi^2/6$.

An approximate saddle point is

$$\hat{\zeta}(n) = 1 - \frac{\pi}{\sqrt{6n}},$$

which gives a saddle point bound of the form

$$p_n < (e^{\pi^2/12} + o(1)) e^{\pi\sqrt{2n/3}}.$$

Again, the true value is only a factor of $O(n^{-1})$ from this upper bound.

Large powers. Probabilistic analysis of algorithms and combinatorial structures often require extracting coefficients of large order in powers of large exponents of some fixed function. This is a particularly favorable situation for saddle point analysis.

We start with an entire function $g(z)$ that has nonnegative coefficients and satisfies $g(0) \neq 0$. If g is a polynomial, we let d denote its degree; in this chapter, we define the degree to be $d = \infty$ if g is entire but not a polynomial.

Theorem 6.2 (Saddle point bounds for large powers) *Let $g(z)$ be an entire function of degree $d \leq \infty$ with positive coefficients. Let λ be a positive number of the open interval $]0, d[$. Then, with $N = \lfloor \lambda n \rfloor$, one has*

$$g_N^{(n)} \equiv [z^N](g(z))^n \leq (g(\zeta))^n \zeta^{-N},$$

where ζ is the unique positive root of the equation

$$\zeta \frac{g'(\zeta)}{g(\zeta)} = \lambda.$$

Proof. Saddle point bounds applied to Cauchy integrals give directly

$$g_N^{(n)} \leq \frac{g(r)^n}{r^N},$$

and the best bound of this kind is obtained for ζ that cancels the derivative. The discussion is entirely similar to that of Theorem 6.1. □

This theorem is also effective for deriving bounds on powers of implicitly defined functions in conjunction with Lagrange inversion. Another of its important uses is for multivariate asymptotics and estimation of tail distributions.

1. Entropy bounds for binomial coefficients. Consider the problem of providing estimates on the binomial coefficients $\binom{n}{\lambda n}$ for some λ with $0 < \lambda < 1$. We assume for notational convenience that λn is an integer and set $N = \lambda n$. Theorem 6.2 provides

$$\binom{n}{\lambda n} = [z^N](1+z)^n \leq (1+\zeta)^n \zeta^{-N},$$

where

$$\frac{\zeta}{1+\zeta} = \lambda \qquad \text{i.e.,} \qquad \zeta = \frac{\lambda}{1-\lambda}.$$

A simple computation then shows that

$$\binom{n}{\lambda n} \le \exp(nH(\lambda)), \qquad \text{where} \qquad H(\lambda) = -\lambda\log\lambda - (1-\lambda)\log(1-\lambda)$$

is the *entropy function.*

Thus, for $\lambda \neq \frac{1}{2}$, the binomial coefficients $\binom{n}{\lambda n}$ are exponentially smaller than the central coefficient $\binom{n}{n/2}$, and the entropy function precisely quantifies this exponential gap.

EXERCISE 1. Develop similar bounds for dice players who are interested in

$$g(z) = \frac{1}{6}(1 + z + z^2 + z^3 + z^4 + z^5).$$

Which function plays the rôle of the entropy? Plot it for $\lambda \in [0, 5]$ and determine its behaviour as $\lambda \to 0$.

Develop a general theory of entropy bounds for any positive polynomial function.

2. Tail estimates and large deviations. Theorem 6.2 is only one of a large number of closely related applications of saddle point bounds in this range of problems. For instance, under the same conditions, consideration of the integral representation

$$\sum_{k\le\lambda n} g_k^{(n)} = \frac{1}{2i\pi}\oint (g(z))^n \frac{1-z^{N+1}}{1-z}\,\frac{dz}{z^{N+1}},$$

with $N = \lambda n$ assumed for simplicity to be integral, leads to bounds for the partial sums of the $g_k^{(n)}$ when $\frac{k}{n} \le \lambda$.

The idea is to use the same value of ζ as for the bound on $g_{\lambda n}^{(n)}$, this being an approximate saddle point. It is easy to check that $\zeta < 1$ when λ is less than the threshold

$$\lambda_0 = \frac{g'(1)}{g(1)}.$$

The bound obtained is then

$$\sum_{k\le\lambda n} g_k^{(n)} \le \frac{1}{1-\zeta}(g(\zeta))^N \zeta^{-N}.$$

Observe that λ_0 is the mean of the probability distribution with probability generating function $g(z)/g(1)$ and $\lambda_0 n$ the mean of the distribution with generating function $(g(z)/g(1))^n$. The constraint $\lambda < \lambda_0$ is then natural as exponentially small bounds are expected to hold only for a region excluding the mean. The formula above provides estimates relative to the left tail of the probability distribution. The exponential bound on the tail coincides with what probabilists call a large deviation law.

Large deviation bounds for right tails are developed in a similar manner from the integral representation,

$$\sum_{k \geq \lambda n} g_k^{(n)} = \frac{1}{2i\pi} \oint g(z)^n \frac{1}{1 - z^{-1}} \frac{dz}{z^{N+1}}.$$

With ζ satisfying the saddle point equation, for $\lambda > \lambda_0$ and $\lambda < d$, one has $\zeta > 1$ and

$$\sum_{k \geq \lambda n} g_k^{(n)} \leq \frac{\zeta}{\zeta - 1} (g(\zeta))^n \zeta^{-N}.$$

Corollary 6.1 (Large deviation bounds for powers) *Assume the conditions of Theorem 6.2, and define* $\lambda_0 = g'(1)/g(1)$. *Let* $\zeta \equiv \zeta(\lambda)$ *be the unique positive root of* $\zeta g'(\zeta)/g(\zeta) = \lambda$.

For fixed λ *and all* n, *with* $N = \lfloor \lambda n \rfloor$, *one has the left and right tail bounds*

$$\sum_{k \leq \lambda n} g_k^{(n)} \leq \frac{1}{1 - \zeta} (g(\zeta))^n \zeta^{-N}, \qquad \sum_{k \geq \lambda n} g_k^{(n)} \leq \frac{\zeta}{\zeta - 1} (g(\zeta))^n \zeta^{-N}$$

corresponding respectively to the two cases

$$\lambda < \lambda_0 \qquad \textit{and} \qquad \lambda > \lambda_0.$$

EXERCISE 2. Use the saddle point bounding technique to get an upper bound on the number of solutions of the equation in the $x_i \in \mathbb{Z}$

$$x_1^2 + x_2^2 + \cdots + x_n^2 \leq \lambda n,$$

which represents the number of integer lattice points in a sphere of radius $\sqrt{\lambda n}$ in high dimensional space.

The bounds should involve the function

$$\Theta(z) = \sum_{n \in \mathbb{Z}} z^{n^2} = 1 + 2 \sum_{n=1}^{\infty} z^{n^2}.$$

How does the result compare to the the volume of the hypersphere of radius $\sqrt{\lambda n}$ in n–dimensional space?
[Such bounds are useful in combinatorial optimization, the knapsack problem and cryptography.]

EXERCISE 3. Develop a similar theory for

$$[z^{\lambda n}]h(z)(g(z))^n$$

when $h(z)$ is an entire function with positive coefficients.

6.3 Saddle point paths

As a preparation for the full saddle point method, we briefly elaborate the indications given in the introduction relative to the classification of points on a surface $|f(z)|$ where $f(z)$ is an analytic function. The existence of an analytic expansion

$$f(z) = f(z_0) + f'(z_0)(z - z_0) + \frac{1}{2}f''(z_0)(z - z_0)^2 + \cdots$$

strongly constrains the "topography" of the surface. In the vicinity of z_0, we set $z - z_0 = \rho e^{i\theta}$ and consider the properties of $|f(z)|$ for small ρ, $\rho \to 0$.

An *ordinary point* is such that $f(z_0) \neq 0$ and $f'(z_0) \neq 0$. At such a point

$$\left|\frac{f(z)}{f(z_0)}\right| \sim |1 + c\rho e^{i(\theta - t)}| \qquad \text{where} \qquad \frac{f'(z_0)}{f(z_0)} = ce^{-it},$$

with $c > 0$. Along the half–line $\theta = t$ the function $|f(z)|$ increases at the fastest rate when z moves away from z_0, while along the half–line $\theta = t + \pi$ the function decreases at the fastest rate. The direction $\theta \equiv t \pmod{\pi}$ is a direction of steepest descent. Along the perpendicular direction $\theta \equiv t + \frac{\pi}{2}$, to the contrary, $|f(z)|$ is stationary (up to second order terms), so that this direction is tangent to a level curve $|f(z)| = |f(z_0)|$ on the surface.

A *zero* is defined by the condition $f(z_0) = 0$. It is a local minimum of the modulus of the function. The surface defined by $|f(z)|$ can neither have local maxima (this is the "maximum principle") nor local minima, apart from zeros. We just proved it for ordinary points, and the discussion that follows establishes the property for saddle points as well.

A *saddle point* is by definition a point such that $f(z_0) \neq 0$ and $f'(z_0) = 0$. We consider first simple saddle points at which $f''(z_0) \neq 0$. At such a point

$$\left|\frac{f(z)}{f(z_0)}\right| \sim |1 + c\rho e^{2i(\theta - t)}| \qquad \text{where} \qquad \frac{f''(z_0)}{2f(z_0)} = ce^{-2it},$$

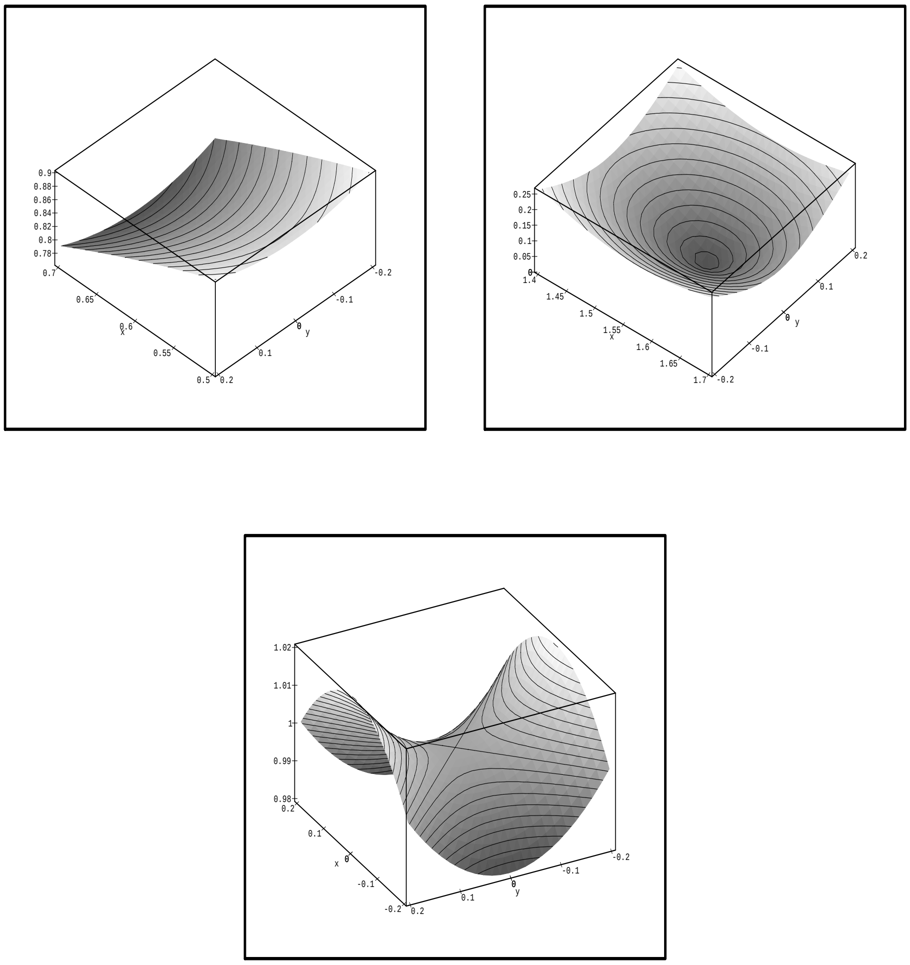

Figure 6.3: The different types of points on a surface $|f(z)|$: an ordinary point, a zero, a simple saddle point. Here $f(z) = \cos z$ and the points are an ordinary point at $\pi/4$ (upper left), a zero at $\pi/2$ (upper right), and a saddle point at 0 (bottom center). Level lines are shown on the surfaces.

with $c > 0$. Consider ρ small, $\rho \to 0$, and examine the various possibilities for θ taken in an interval of amplitude 2π when moving away from the saddle point. The direction $\theta = t$ is one of fastest increase; the direction $\theta = t + \pi/4$ is stationary; the direction $\theta = t + \pi/2$ is one of fastest decrease; the direction $\theta = t + 3\pi/4$ is again stationary. The same pattern repeats itself when t is changed to $t + \pi$.

In other words, at a simple saddle point, two level curves cross at a right angle; there is a direction of steepest descent and a perpendicular direction of steepest ascent away from the saddle point. The direction of steepest descent is also called the *axis* of the saddle point.

The discussion of multiple saddle points defined by cancellation of more derivatives is similar: there are m level curves and m steepest ascent/descent curves interlaced if all derivatives till order $m - 1$ inclusively vanish.

The various types of points are depicted on Figure 6.3.

EXERCISE 4. Describe precisely the topography of the surface $|f(z)| = 0$ at a multiple saddle point where more than one derivative vanishes. Show that the angle between consecutive level curves is $\frac{2\pi}{m}$ and that the steepest ascent/descent curves bissect the level curves.

EXERCISE 5. [Gauss] Use the classification of points to prove the fundamental theorem of algebra: *"A polynomial of degree d has exactly d complex zeros."*

6.4 Saddle point analysis of the exponential

The purpose of this section is to provide the basis for full saddle point analyses by working out in some detail the problem of estimating $[z^n]e^z$.

The starting point is the Cauchy coefficient integral

$$f_n = \frac{1}{2i\pi} \int_{|z|=r} e^z \frac{dz}{z^{n+1}},$$

where the contour of integration is taken to be a circle of radius r. Of course, it is known in advance that

$$f_n = \frac{1}{n!} \sim \frac{e^n}{n^n \sqrt{2\pi n}}.$$

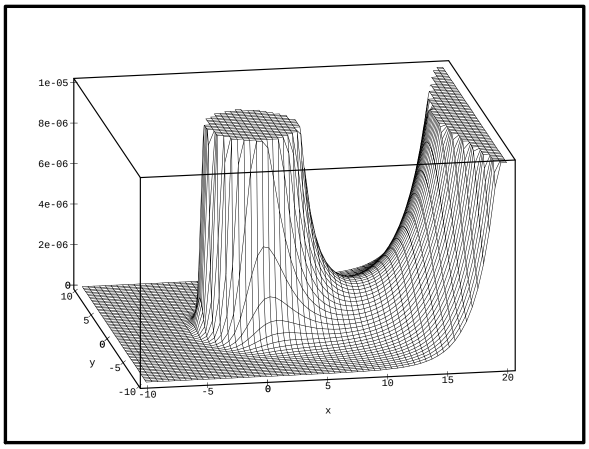

Figure 6.4: The modulus of e^z/z^{n+1} and the saddle point at $z = n+1$ (here $n = 9$).

The topography of the modulus of the integrand is displayed in Figure 6.4. It has a saddle point at $z = n + 1$ with an axis perpendicular to the real line. From the previous discussion of saddle point bounds, we thus expect good bounds to derive from adopting as integration contour a circle centered at the origin with radius $n + 1$ (or about!) as integration contour.

It proves convenient to switch to polar coordinates and set $z = re^{i\theta}$. The original integral becomes

$$f_n = \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{r\,e^{i\theta} - n\log r - ni\theta}\, d\theta.$$

In accordance with saddle point principles we adopt now the choice $r = n$ ($n+1$ would do equally well but would unnecessarily encumber calculations). The integral rewrites

$$f_n = \frac{e^n}{n^n} \cdot \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{n[e^{i\theta} - 1 - i\theta]}\, d\theta. \tag{6.14}$$

Set $h(\theta) = e^{i\theta} - 1 - i\theta$ with expansion

$$h(\theta) = -\frac{\theta^2}{2} - \frac{i\theta^3}{6} + \frac{\theta^4}{24} + \cdots.$$

The absence of a linear term in θ indicates a saddle point. The function

$$|e^{h(\theta)}| = e^{\cos\theta - 1},$$

is unimodal with its peak at $\theta = 0$ and the same property holds for $|e^{nh(\theta)}|$ which is even more strongly peaked at $\theta = 0$.

An estimation of f_n should naturally proceed by isolating a small portion of the contour (corresponding to z near the real axis). We thus set

$$I_n^{(0)} = \int_{-\delta}^{+\delta} e^{nh(\theta)}\, d\theta, \qquad I_n^{(1)} = \int_{\delta}^{2\pi-\delta} e^{nh(\theta)}\, d\theta,$$

and choose δ in such a manner that ($\ll$ means here "much smaller than")

(C1) $n\delta^2 \to \infty$, that is $\delta \gg n^{-1/2}$;

(C2) $n\delta^3 \to 0$, that is $\delta \ll n^{-1/3}$.

One way of realizing the compromise is to take $\delta = n^e$ where e is any number between $\frac{1}{2}$ and $\frac{1}{3}$, for instance

$$\delta \equiv \delta(n) = n^{-2/5}.$$

From (C1), there follows that $e^{nh(\delta)}$ is exponentially small, being dominated by a term of the form $e^{-\mathcal{O}(n^{1/5})}$. As $|e^{h(\theta)}|$ decreases on $[\delta, \pi]$, one has

$$\text{(C3)} \qquad |e^{h(\theta)}| \leq |e^{h(\delta)}| \qquad \text{for} \qquad \theta \in [\delta, 2\pi - \delta] \tag{6.15}$$

so that a similar upper bound also holds for the noncentral integral: since $|I_n^{(1)}| < 2\pi n e^{nh(\delta)}$. We also have $h(\delta) \sim \delta^2/2$, hence the bound

$$|I_n^{(1)}| = \mathcal{O}(e^{-\frac{1}{2}n^{1/5}}). \tag{6.16}$$

Thus, by (C1), δ has been taken large enough so that the central integral $I_n^{(0)}$ "captures" most of the contribution, while the remainder integral $I_n^{(1)}$ is exponentially small by (C3).

We now turn to the precise evaluation of the central integral $I_n^{(0)}$. Near $\theta = 0$, only the terms till order 2 matter in the expansion of $h(\theta)$ because of the condition (C2) which ensures $n\theta^3 \to 0$. One has:

$$\begin{aligned} I_n^{(0)} &\sim \int_{-\delta}^{+\delta} e^{-n\theta^2/2}\, d\theta \\ &\sim \frac{1}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{+\delta\sqrt{n}} e^{-t^2/2}\, dt \\ &\sim \frac{1}{\sqrt{n}} \int_{-\infty}^{+\infty} e^{-t^2/2}\, dt \\ &\sim \sqrt{\frac{2\pi}{n}}. \end{aligned} \tag{6.17}$$

The first line of (6.17) uses the fact that $n\theta^3 \to 0$ so that $h(\theta)$ can be reduced to its quadratic approximation, with error terms of order $n\delta^3 = n^{-1/5}$; the second line is based on the rescaling $t = \theta\sqrt{n}$. The third line is justified by the fact that the tails of the Gaussian integral are exponentially small so that the integral can be completed to the full range $(-\infty, +\infty)$, which induces error terms that are exponentially small. Finally, the complete Gaussian integral can be evaluated and the estimate follows. Thus condition (C2) has ensured that the integration interval is not "too large" so that a *local expansion* of the integral till quadratic terms suffices for asymptotic estimates, but large enough so that a valid estimate results by (C3) from completion of the integral.

Putting everything together, we have obtained

$$I_n^{(0)} + I_n^{(1)} \sim \sqrt{\frac{2\pi}{n}}.$$

Hence the final result

$$f_n \equiv \frac{1}{n!} \sim \frac{e^n}{n^n\sqrt{2\pi n}}.$$

We have thus established Stirling's formula by the saddle point method.

In summary the process of saddle point analysis is made possible by a fundamental *split* of the integration contour —here, a circle— into a small arc centered on the real axis. The small arc has to satisfy two conflicting requirement: to be large enough by (C1) so as to capture the essential contribution of the integral; and to be small enough to allow the function to be well approximated locally by its quadratic terms. In addition, the estimation was made possible because the function decays appropriately, away from the real axis, so that the integrand stays small on the noncentral part of the contour as expressed by (C3) of (6.15).

6.5 Admissibility

It is possible to encapsulate the conditions that render possible the analysis of $[z^n]e^z$ into a general definition. This leads to the notion of admissible functions. By design, saddle point analysis applies to such functions and asymptotic forms for their coefficients can be systematically determined. Such an approach was initiated by Hayman [15] whose steps we closely follow in this section. A crisp account is also given in Section II.7 of Wong's book [30].

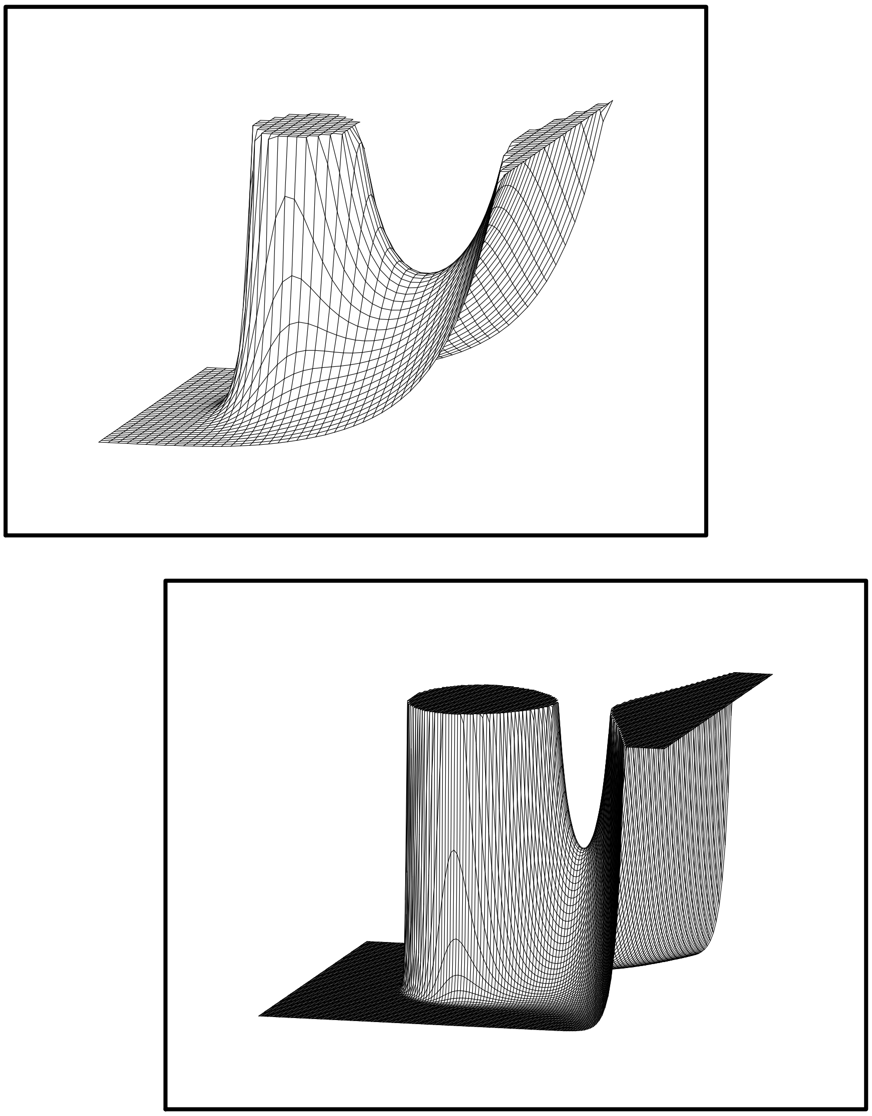

Figure 6.5: Plots of $|e^z z^{-n-1}|$ for $n = 3$ and $n = 30$ (scaled according to the value of the saddle point) illustrate the essential concentration condition as higher values of n produce steeper saddle point paths.

In addition, admissible functions satisfy useful closure properties so that an infinite class of admissible functions relevant to combinatorial applications can be constructed. Such a class includes OGF's and EGF's for integer partitions, set partitions, involutions, etc.

We consider here a function $f(z)$ that is analytic at the origin. As was done with the exponential, we switch to polar coordinates and examine the expansion of $f(re^{i\theta})$ when the argument is near the real axis. The basic expansion is

$$\log f(re^{i\theta}) = \log f(r) + \sum_{\nu=1}^{\infty} \alpha_\nu(r) \frac{(i\theta)^\nu}{\nu!}. \tag{6.18}$$

One has, in general, $\alpha_\nu(r) = r\frac{d}{dr}\alpha_{\nu-1}(r)$.

The basic quantities for saddle point analysis are the first two terms, $a(r) = \alpha_1(r)$ and $b(r) = \alpha_2(r)$. It proves convenient to operate with $f(z)$ put into exponential form, $f(z) = e^{h(z)}$, and a simple computation yields

$$\begin{aligned} a(r) &= rh'(r) \\ b(r) &= r^2h''(r) + rh'(r). \end{aligned} \tag{6.19}$$

In terms of f, itself, one has:

$$\begin{aligned} a(r) &= r\frac{f'(r)}{f(r)} \\ b(r) &= r\frac{f'(r)}{f(r)} + r^2\frac{f''(r)}{f(r)} - r^2\left(\frac{f'(r)}{f(r)}\right)^2. \end{aligned}$$

When $f(z)$ has nonnegative Taylor coefficients, $a(r)$ and $b(r)$ are always positive for $r > 0$, by an argument already encountered in (6.13).

Exercise 6. Let $f(z)$ have nonnegative coefficients. Relate the mean and variance of a discrete distribution with generating function

$$g(x) = \frac{f(rx)}{f(x)}$$

to $a(r)$ and $b(r)$. As a variance cannot be negative, one must have $b(r) \geq 0$, with strict inequality, except in degenerate cases.

Definition 6.1 *Let $f(z)$ have radius of convergence ρ with $0 < \rho \leq \infty$ and be always positive on some subinterval $]R_0, \rho[$ of $]0, \rho[$. The function $f(z)$ is said to be* admissible *if it satisfies the following three conditions.*

H1. [Capture condition] $\lim_{r \to \rho} b(r) = +\infty$.

H2. [Locality condition] For some function $\delta(r)$ *defined over* $]R_0, \rho[$ *and satisfying* $0 < \delta < \pi$, *one has*

$$f(re^{i\theta}) \sim f(r) e^{i\theta a(r) - \theta^2 b(r)/2} \qquad \text{as } r \to R_0,$$

uniformly in $|\theta| \leq \delta(r)$.

H3. [Decay condition] Uniformly in $\delta(r) \leq |\theta| < \pi$

$$f(re^{i\theta}) = o\left(\frac{f(r)}{\sqrt{b(r)}}\right).$$

Admissible functions in the above sense are also (rightly!) called *Hayman admissible* functions in the literature.

Theorem 6.3 (Coefficients of admissible functions) *Let* $f(z)$ *be an admissible function and* $\zeta \equiv \zeta(n)$ *be the unique solution in the interval* $]R_0, \rho[$ *of the the saddle point equation*

$$\zeta \frac{f'(\zeta)}{f(\zeta)} = n.$$

The Taylor coefficients of $f(z)$ *satisfy*

$$f_n \equiv [z^n] f(z) \sim \frac{f(\zeta)}{\zeta^n \sqrt{2\pi b(\zeta)}} \qquad \text{as } n \to \infty \tag{6.20}$$

with $b(z) = z^2 h''(z) + z h'(z)$ *and* $h(z) = \log f(z)$.

Proof. *1.* The first step involves proving a more general result that describes the shape of the individual terms $f_n r^n$ in the Taylor expansion of $f(z)$ as r gets closer to its limit value ρ. The terms turn out to exhibit a bell-shaped profile. The asymptotic form (6.20) will then then results from a proper choice of r.

Lemma 6.1 *As* r *tends to* ρ, *one has*

$$f_n r^n = \frac{f(r)}{\sqrt{2\pi b(r)}} \left[\exp\left(-\frac{(a(r) - n)^2}{b(r)} \right) + o(1) \right], \tag{6.21}$$

where the error term $o(1)$ *is uniform for all integers* n.

The coefficients f_n are given by Cauchy's formula,

$$f_n r^n = \frac{1}{2\pi}\int_{-\delta}^{2\pi-\delta} f(re^{i\theta})e^{-in\theta}\,d\theta,$$

where $\delta = \delta(n)$ is as specified by the admissibility definition. The estimation of this integral is once more based on a fundamental split

$$f_n r^n = I^{(0)} + I^{(1)} \qquad \text{where} \qquad I^{(0)} = \frac{1}{2\pi}\int_{-\delta}^{+\delta},\ I^{(1)} = \frac{1}{2\pi}\int_{+\delta}^{2\pi-\delta}.$$

From condition H3 (the "decay" condition), uniformly in n:

$$I^{(1)} = \frac{o(f(r))}{b(r)^{1/2}}. \tag{6.22}$$

On the other hand, condition H2 (the "locality" condition) gives uniformly in n:

$$\begin{aligned} I^{(0)} &= \frac{f(r)}{2\pi}\int_{-\delta}^{+\delta} e^{i(a(r)-n)\theta-\frac{1}{2}b(r)\theta^2}(1+o(1))\,d\theta \\ &= \frac{f(r)}{2\pi}\left[\int_{-\delta}^{+\delta} e^{i(a(r)-n)\theta-\frac{1}{2}b(r)\theta^2}\,d\theta + o\left(\int_{-\infty}^{+\infty} e^{-\frac{1}{2}b(r)\theta^2}\right)\right]. \end{aligned} \tag{6.23}$$

The second integral in the last line of (6.23) is $O(b(r)^{-1/2})$ as $r \to \rho$. Rescaling the first integral and setting $(a(r)-n)(2/b(r))^{1/2} = c$, we obtain

$$I^{(0)} = \frac{f(r)}{\pi\sqrt{2b(r)}}\left[\int_{-\delta\sqrt{b(r)/2}}^{+\delta\sqrt{b(r)/2}} e^{-t^2+ict} + o(1)\right]. \tag{6.24}$$

Now, it follows from conditions H2 and H3, both taken at $\theta = \delta(r)$ that $b(r)\delta^2 \to \infty$ as $r \to \rho$. Thus the integral in (6.24) can be extended to a complete Gaussian integral, introducing only $o(1)$ error terms. This entails

$$I^{(0)} = \frac{f(r)}{\pi\sqrt{2b(r)}}\left[\int_{-\infty}^{+\infty} e^{-t^2+ict} + o(1)\right], \tag{6.25}$$

and the Gaussian integral evaluates to $\sqrt{\pi}e^{-c^2/4}$ (by completing the square and shifting vertically the integration line). Thus, combining the estimate (6.25) for the central integral $I^{(0)}$ and the estimate (6.22) for the remainder integral, we obtain the estimate (6.21).

2. To establish the theorem, we first observe that $a(r)$ is increasing (as its derivative $b(r)/r$ is positive) and, in addition tends to infinity (this results from setting $n = 0$ in formula (6.21)). Thus $\zeta(n)$ is well–defined. Setting $r = \zeta(n)$ in (6.21) then completes the proof of the theorem. □

The rôle of the various conditions should be clear from the preceding discussion and from the study of the exponential function. The choice of the function $\delta(n)$ for a particular problem is to be guided by consideration of the expansion (6.18). We must have

$$\alpha_2(r)\delta^2 \to \infty \qquad \text{and} \qquad \alpha_3(r)\delta^3 \to 0.$$

This is because the method requires a nearly complete integral to arise while the error terms after the quadratic part of $\log f(re^{i\theta})$ should be kept small enough. Thus, in order to work, the method necessitates *a priori*

$$\frac{(\alpha_3(r))^2}{(\alpha_2(r))^3} \to 0.$$

Then, δ should be taken in such a way that ($\ll$ still means "much smaller than")

$$\frac{1}{\alpha_2^{1/2}} \ll \delta \ll \frac{1}{\alpha_3^{1/3}}, \tag{6.26}$$

a possible choice being the geometric mean of the two bounds

$$\delta(r) = \alpha_2^{-1/4}\alpha_3^{-1/6}. \tag{6.27}$$

Non-admissible functions. Before showing cases of applications, we briefly comment on functions that fail to satisfy admissibility conditions.

The function $f(z) = \frac{1}{1-z}$ cannot be admissible as the asymptotic form that Theorem 6.3 would imply is the erroneous

$$[z^n]\frac{1}{1-z} \overset{!!}{\sim} \frac{e^{-1}}{\sqrt{2\pi}},$$

corresponding to a saddle point near $1 - \frac{1}{n}$. The expansion (6.18) has

$$\log f(r) = \log\frac{1}{1-r}, \quad \alpha_1(r) = \frac{r}{1-r}, \quad \alpha_2(r) = \frac{r}{(1-r)^2}, \quad \alpha_3(r) = \frac{r(1+r)}{(1-r)^3},$$

and the coefficient of order ν involves $(1-r)^{-\nu}$. The locality condition and the decay condition cannot be simultaneously satisfied (see also the discussion around Eq. (6.26) and (6.27)) and the order of growth requirements discussed above are not satisfied.

More generally, functions of the form $(1-z)^{-\alpha}$ are typical instances with too slow a growth to be admissible. In a sense, singularity analysis salvages the situation by using a larger part of the contour and by normalizing to a "global" Hankel Gamma integral instead of a more "local" Gaussian integral. (This is also in accordance with the fact that the saddle point formula gives for the coefficient a fraction 0.14676 of the true value which is 1.)

Other functions failing to satisfy the decay condition alone are e^{z^2} and $e^{z^2}+e^z$ as they are also large, away from the central arc and near the negative real axis.

EXERCISE 7. Show that

$$f_n \equiv [z^n]\frac{1}{(1-z)^\alpha} = \frac{1}{2i\pi}\int_{1-1/n-i\infty}^{1-1/n+i\infty} \frac{1}{(1-z)^\alpha}\frac{dz}{z^{n+1}}.$$

Get the asymptotic form of f_n by a suitable normalization reminiscent of singularity analysis.

Closure properties. A valuable characteristic of Hayman's work is that it leads to general theorems guaranteeing that large classes of functions are admissible.

Theorem 6.4 (Closure of admissible functions) *Let $f(z)$ and $g(z)$ be admissible functions and let $P(z)$ be a polynomial with real coefficients. Then:*

- *(i) The product $f(z)g(z)$ and the exponential $e^{f(z)}$ are admissible functions.*
- *(ii) The sum $f(z)+P(z)$ is admissible. If the leading coefficient of $P(z)$ is positive then $f(z)P(z)$ and $P(f(z))$ are admissible.*
- *(iii) If the Taylor coefficients of $e^{P(z)}$ are eventually positive, then $e^{P(z)}$ is admissible.*

Proof. We refer to Hayman's original paper [15] for full proofs that are not difficult. They essentially reduce to making an inspired guess for the choice of the δ function, which may be guided by Equations (6.26) and (6.27), and then checking the conditions of the admissibility definition. For instance, in the case of the exponential, $F(z) = e^{f(z)}$, the conditions H1, H2, H3 are satisfied if one takes

$$\delta(r) = (f(r))^{-2/5}.$$

$\square$

6.6 Combinatorial enumeration and admissibility

Admissible functions involve some sort of exponential growth, so that their domain of application is closely related to various forms of set constructions.

1. Involutions. The involution numbers have EGF $f(z) = e^{z+z^2/2}$ which is the exponential of a polynomial with positive coefficients, hence an admissible function by Theorem 6.4. The saddle point $\zeta(n)$ was already analyzed:

$$\zeta(n) = \sqrt{n} - \frac{1}{2} + \mathcal{O}(n^{-1/2}).$$

For the involution numbers, the saddle point bound needs only be multiplied by a factor of

$$\frac{1}{\sqrt{2\pi\zeta(2\zeta+1)}}.$$

This quantity "measures" the quality of the saddle point bound and is only $\mathcal{O}(n^{-1/2})$. Substitution of the expansion of ζ into the saddle point formula finally yields an asymptotic equivalent

Proposition 6.1 *The number I_n of involutive permutations satisfies*

$$I_n \sim n! \frac{e^{-1/4}}{\sqrt{2\pi n}} n^{-n/2} e^{n/2+\sqrt{n}}.$$

This result is originally due to Moser and Wyman.

EXERCISE 8. Find an asymptotic equivalent for the number of permutations of $[1\,..\,n]$ of order m ($\sigma^m = 1$) that have EGF

$$\exp\left(\sum_{d \mid m} \frac{z^d}{d}\right).$$

EXERCISE 9. Find an asymptotic equivalent for the number of permutations of $[1\,..\,n]$ with longest cycle of size $\le m$ that have EGF

$$\exp\left(\sum_{1\le d\le m} \frac{z^d}{d}\right).$$

2. Bell numbers and set partitions. The Bell numbers have $f(z) = \exp(e^z - 1)$ as EGF. Theorem 6.4 provides all that is needed: $f(z)$ is the exponential of $e^z - 1$; the latter function is admissible being the sum of e^z known to be admissible and of the polynomial -1. Hence the saddle point formula of Theorem 6.3 applies. The computation of ζ has been given already in Section 6.2.

Proposition 6.2 *The number B_n of set partitions satisfies*

$$B_n \sim n! \frac{e^{e^\zeta - 1}}{\zeta^n \sqrt{2\pi\zeta(\zeta+1)e^\zeta}}. \tag{6.28}$$

where ζ is defined implicitly by

$$\zeta e^\zeta = n, \qquad \textit{so that} \qquad \zeta = \log n - \log\log n + o(1).$$

This example is probably the most famous application of saddle point techniques to combinatorics, see [5].

We observe here that the asymptotic form in terms of ζ itself is the proper one as no back substitution of an asymptotic expansion of ζ (in terms of n and $\log n$) can provide an asymptotic expansion for B_n solely in terms of n. It is often the case that *saddle point estimates involve implicitly defined quantities.*

EXERCISE 10. For involution numbers and Bell numbers, find a direct proof by the saddle point method.

EXERCISE 11. Using explicit sums for I_n and B_n rederive their asymptotic forms by the Laplace method.

3. A large singular function. Take $f(z) = \exp(z/(1-z))$, a function closely related to the number of increasing subsequences in permutations.

The saddle point bounds suggest a growth of the coefficient driven by a term of the form $e^{2\sqrt{n}}$. Proving $f(z)$ to be admissible essentially reduces to finding a proper δ–function. We have

$$\log f(re^{i\theta}) = \frac{1}{1-r}+\frac{r}{(1-r)^2}(i\theta)+\frac{1}{2}\frac{r(1+r)}{(1-r)^3}(i\theta)^2+\frac{1}{6}\frac{r(1+4r+r^2)}{(1-r)^4}(i\theta)^3+\cdots.$$

The condition $\alpha_3^2\alpha_2^{-3} \to 0$ of Eq. (6.26) are satisfied here, and one may choose $\delta(r) = (1-r)^e$ for any $e \in]\frac{4}{3}, \frac{3}{2}[$; for instance in accordance with (6.27) the value $\delta(r) = (1-r)^{\frac{17}{12}}$ is adequate. Checking admissibility from this point is a routine mater left to the reader.

We know already that the saddle point is $\sim 1 - n^{-1/2}$. An application of Theorem 6.3 furnishes the estimate

$$[z^n]f(z) \sim \frac{e^{-1/2}e^{2\sqrt{n}}}{2\sqrt{\pi}n^{3/4}},$$

which is only $\mathcal{O}(n^{-3/4})$ of the corresponding saddle point bound.

Exercise 12. Interpret $f(z)$ as the generating function of a combinatorial class. [Hint: fragmentations of a permutation into pieces.]

Exercise 13. Show that

$$[z^n]\exp\left(\frac{c}{1-z}\right)$$

grows like $\exp(2\sqrt{cn})$.

4. Integer partitions. This final example involves a function with a finite radius of convergence that arises from an unlabelled-set construction.

Proposition 6.3 *The number p_n of partitions of integer n satisfies*

$$p_n \equiv [z^n]\prod_{k=1}^{\infty}\frac{1}{1-z^k} \sim \frac{1}{4n\sqrt{3}}e^{\pi\sqrt{2n/3}} \tag{6.29}$$

The asymptotic formula (6.29) is only the first term of a complete expansion involving decreasing exponentials (!) that was discovered by Hardy and Ramanujan in 1917. While the full Hardy–Ramanujan expansion necessitates considering infinitely many saddle–points near the unit circle and requires the so–called modular transformation [1], the first term (6.29) follows from the admissibility theorem. We shall not prove this here, but the form (6.29) only requires the asymptotic expansion of the partition generating function near $z = 1$ the major steps of which were given in Section 6.2. The singular behaviour along and near to the real line is comparable to that of $\exp((1-z)^{-1})$, which explains a growth like $e^{\sqrt{n}}$ for the number of integer partitions.

Exercise 14. Find an expansion till terms of order $o(1)$ for

$$\log p(z) = \sum_{n=1}^{\infty} \frac{1}{n} \frac{z^n}{1-z^n},$$

as $z \to 1^-$. [Hint: this may be done by setting $z = e^{-t}$ and using Mellin transform techniques from the next chapter.]

Exercise 15. Complete the proof that the partition OGF $p(z)$ is admissible and find an asymptotic expansion of $p(re^{i\theta})$ when $r \to 1$ and θ is small. Complete the proof of (6.29).

Exercise 16. Analyse asymptotically the number of partitions of n into distinct summands:

$$q_n = [z^n] \prod_{k \geq 1} (1 + z^k).$$

Exercise 17. Let r_n be the number of partitions of n into sumands that are squares. Find an asymptotic equivalent of $\log r_n$. Generalize to cubes, etc.

6.7 Large powers

Saddle point analysis of coefficients of large powers has the happy feature that integrals are always taken in a bounded part of the complex plane. In this way, saddle point conditions analogous to those of admissibility are directly verified and strong uniformity of estimates results. We treat here the analysis of dominant terms; a simple variation of the proof to be given in the next section then suffices to derive full expansions.

The context of the analysis is similar to Theorem 6.2 though an additional aperiodicity condition has to be imposed. Here, a function $g(z)$ analytic at 0 with $g(z) \neq 0$ is *periodic* if it can be put under the form $g(z) = \gamma(z^m)$ for some analytic γ and some $m \geq 2$; otherwise, it is said to be *aperiodic*. Some aperiodicity condition is a necessity for asymptotic expansions as otherwise $[z^n]g(z) = 0$ when $n \not\equiv 0 \bmod m$.

Theorem 6.5 (Saddle point analysis for large powers) *Let $g(z)$ be an entire function of degree d with positive coefficients assumed to be aperiodic and such that $g(0) \neq 0$. Let λ be a positive number of some subinterval $[\lambda_a, \lambda_b]$ of the open interval $]0, d[$. Then, with $N = \lfloor \lambda n \rfloor$, one has uniformly for $\lambda \in [\lambda_a, \lambda_b]$*

$$g_N^{(n)} \equiv [z^N](g(z))^n = \frac{(g(\zeta))^n}{\zeta^{N+1}\sqrt{2\pi n V}}(1 + o(1)),$$

where ζ is the unique positive root of the equation

$$\zeta\frac{g'(\zeta)}{g(\zeta)} = \lambda \qquad \text{and} \qquad V = \frac{d^2}{d\zeta^2}[\log g(\zeta) - \lambda \log \zeta]. \tag{6.30}$$

Proof. We may freely assume that λn is an integer, $\lambda n = N$. Then

$$g_N^{(n)} = \frac{1}{2i\pi}\oint e^{nh(z)}\,\frac{dz}{z} \qquad \text{where} \qquad h(z) = \log g(z) - \lambda \log z. \tag{6.31}$$

Integration along a circle passing through the saddle point of $e^{nh(z)}$ leads to taking $z = \zeta e^{i\theta}$, where ζ depends on λ only:

$$\zeta\frac{g'(\zeta)}{g(\zeta)} = \lambda.$$

The expansion of $h(z)$ around ζ is

$$h(z) - h(\zeta) = \frac{h''(\zeta)}{2}(z-\zeta)^2 + \mathcal{O}((z-\zeta)^3). \tag{6.32}$$

According to general saddle point principles, the integral of (6.31) has to be split. We set

$$I^{(0)} = \frac{1}{2i\pi}\int_{\zeta e^{-i\delta}}^{\zeta e^{+i\delta}} e^{n(h(z)-h(\zeta))}\,\frac{dz}{z}, \quad I^{(1)} = \frac{1}{2i\pi}\int_{\zeta e^{+i\delta}}^{\zeta e^{i(2\pi-\delta)}} e^{n(h(z)-h(\zeta))}\,\frac{dz}{z}. \tag{6.33}$$

In order for $I^{(0)}$ to represent most of the contribution to the full integral, one should take $\delta \equiv \delta(n)$ such that ($\ll$ means again "much smaller than")

$$n\delta^2 \to \infty, \quad n\delta^3 \to 0 \qquad \Longrightarrow \qquad \frac{1}{n^{1/2}} \ll \delta \ll \frac{1}{n^{1/3}},$$

which is satisfied for instance by the geometric mean $\delta = n^{-5/12}$.

The integrand in (6.31) satisfies a form of decay condition similar to condition H3 of admissibility by virtue of the following lemma:

Lemma 6.2 (Periodicity lemma) *Let $f(z)$ be analytic in $|z| < R$, have nonnegative Taylor coefficients and satisfy $f(0) \neq 0$. If for some r with $0 < r < R$ and some $\theta \in]0, 2\pi[$ one has $|f(re^{i\theta})| = f(r)$ then $f(z)$ is periodic of some period $m \geq 2$ and $e^{i\theta}$ is an mth root of unity.*

Proof. As $f(z)$ has positive coefficients, the triangular inequality implies

$$|f(re^{i\theta})| \leq \sum_n f_n r^n = f(r).$$

Equality can be realized only if $f_n e^{ni\theta} = f_n$ for all n. An easy *a contrario* argument shows that, for $\theta \neq 0$, this is only possible if $f(z)$ is periodic and $e^{i\theta}$ a root of unity. □

We now return to the proof of Theorem 6.5. From the lemma above, an aperiodic entire function attains its maximum only on the positive real axis. For n large enough, which means δ small enough, one then has

$$\sup_{|\theta| \geq \delta} |g(\zeta e^{i\theta})| = |g(\zeta e^{i\delta})|. \tag{6.34}$$

Under these conditions,

$$I^{(1)} = \mathcal{O}(e^{-n^{1/5}}) \qquad \text{since} \qquad n(h(\zeta e^{i\delta}) - h(\zeta)) = \mathcal{O}(n\delta^2) = \mathcal{O}(n^{1/4}). \tag{6.35}$$

This bound disposes automatically of the noncentral part of the integral.

Evaluation of the central integral $I^{(0)}$ results directly from local expansions. Along the lines of Theorem 6.4, we have

$$\begin{aligned} I^{(0)} &\sim \frac{1}{2\pi} \int_{-\delta}^{+\delta} \exp(n(h(\zeta e^{i\theta}) - h(\zeta))\, d\theta \\ &\sim \frac{1}{2\pi} \int_{-\delta}^{+\delta} \exp(-n\zeta^2 h''(\zeta)\frac{\theta^2}{2})\, d\theta \\ &\sim \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(-n\zeta^2 h''(\zeta)\frac{\theta^2}{2})\, d\theta \\ &\sim \frac{1}{2\pi} \frac{1}{\zeta\sqrt{2\pi h''(\zeta)}}. \end{aligned} \tag{6.36}$$

The theorem 6.5 results from combining (6.35) and (6.36). Uniformity results from the fact that all estimates are uniform as one operates in a bounded region of the complex plane. □

We comment now on the uniformity property. Under the form stated in Theorem 6.5, this means that for all $\epsilon > 0$, there exists some n_0 such that for all $n > n_0$ the $o(1)$ term can be taken smaller than ϵ. Such uniformity considerations prove important for the derivation of Gaussian laws in a Section 6.10.

A simple example of application is to the central binomial coefficient. There, we have $\lambda = 1$ and $g(z) = (1+z)^2$, so that

$$h(z) = 2\log\frac{1}{1+z} - \log z, \quad h'(z) = \frac{2}{1+z} - \frac{1}{z}, \quad h''(z) = -\frac{2}{(1+z)^2} + \frac{1}{z^2}.$$

The saddle point is then at $\zeta = 1$ where $h''(\zeta) = \frac{1}{2}$ and a direct application of Theorem 6.5 yields

$$\binom{2n}{n} \equiv [z^n](1+z)^{2n} \sim \frac{4^n}{\sqrt{\pi n}}.$$

Exercise 18. Prove an adapted version of Theorem 6.5 for functions with a finite radius of convergence.

Exercise 19. With $h(z)$ an entire function with positive coefficients, analyze similarly

$$[z^{\lambda n}]h(z)(g(z))^n.$$

6.8 Combinatorial enumeration and large powers

We start with an estimation of the so-called central trinomial coefficients defined by

$$T_n = [z^n](1+z+z^2)^n.$$

A direct application of Theorem 6.5 with $\lambda = 1$ yields $\zeta = 1$ by the saddle point equation, so that

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

The Motzkin numbers count unary-binary trees, so that they may be defined by

$$M_n = [z^n]M(z) \qquad \text{where} \qquad M = z(1 + M + M^2).$$

The standard approach is the one seen earlier based on singularity analysis as the implicitly defined function $M(z)$ has an algebraic singularity of the $\sqrt{\ }$-type,

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

The Lagrange inversion formula provides an alternative route. It gives

$$M_n = \frac{1}{n}[z^{n-1}](1 + z + z^2)^n,$$

which is amenable to saddle point analysis using $\lambda = (n-1)/n$. Hence,

$$M_n \sim \frac{3^{n+1/2}}{2\sqrt{\pi n^3}}.$$

In general, the Lagrange inversion formula establishes an exact correspondence between two estimations problems relative to

— coefficients of large order in large powers and

— coefficients of implicitly defined functions.

Thus it can bring the evaluation of coefficients of implicit functions into the orbit of the saddle point method. The saddle point method is then sometimes more convenient to work with[2], especially when explicit or uniform upper bounds are required, since bounds are more easily obtained on fixed circles than on variable Hankel contours.

Exercise 20. Use Lagrange inversion to relate explicitly the analysis of

$$[z^n]Y(z) \qquad \text{where} \qquad Y(z) = z\phi(Y(z)).$$

by singularity analysis, and the analysis of large powers according to Theorem 6.5.

[2]In his original memoir [4], Darboux however goes in the converse direction and shows for instance that the Laplace method for integrals may be reduced to analysis of singularities at a finite distance of a generating function.

EXERCISE 21. Analyse the asymptotic behaviour of integrals

$$I_n = \int_0^1 f(x)(g(x))^n \, dx,$$

by taking the corresponding generating function. (This exercise properly belongs to Chapter 5.)

[Hint. See Darboux's memoir [4].]

Proceed similarly for contour integrals $\oint f(z)(g(z))^n \, dz$.

EXERCISE 22. In his first letter to Hardy, Ramanujan announces that

$$\frac{1}{2}e^n = 1 + \frac{n}{1!} + \frac{n^2}{2!} + \cdots + \frac{n^{n-1}}{(n-1)!} + \frac{n^n}{n!}\theta,$$

where

$$\theta = \frac{1}{3} + \frac{4}{135(n+k)},$$

and k lies between $\frac{8}{45}$ and $\frac{2}{21}$. Prove Ramanujan's assertion for all $n \geq 1$. [Hint. See [9].]

6.9 Full asymptotic expansions

Full asymptotic expansions for large powers derive from a simple modification of the proof of Theorem 6.5 via an analytic change of variables. For the case of coefficients of a fixed function, the situation is more intricate and we content ourselves with giving pointers to the literature.

Theorem 6.6 (Full asymptotic expansions for large powers) *Under the conditions of Theorem 6.5, there exists a full asymptotic expansion for $g_N^{(n)}$ in descending powers of n:*

$$g_N^{(n)} \equiv [z^N](g(z))^n \sim \frac{(g(\zeta))^n}{\zeta^{N+1}\sqrt{2\pi n V}}\left[1 + \sum_{k=1}^{\infty} \frac{e_k}{n^k}\right], \tag{6.37}$$

with ζ and V defined by (6.30).

In addition, the error terms in any terminating form of (6.37) are uniform, when λ satisfies the conditions of Theorem 6.5.

Proof. The proof implies an algorithm to compute the e_k. From the developments of Theorem 6.5, it suffices to derive a full asymptotic expansion of $I^{(0)}$ as $I^{(1)}$ is exponentially small. We start from

$$I^{(0)} = \int_{\zeta e^{-i\delta}}^{\zeta e^{+i\delta}} e^{n(h(z)-h(\zeta))} \frac{dz}{z}, \tag{6.38}$$

with

$$h(z) - h(\zeta) \sim \frac{h''(\zeta)}{2}(z-\zeta)^2.$$

The idea is to perform the change of variables

$$h(z) - h(\zeta) = \frac{h''(\zeta)}{2}t^2,$$

which defines t as an analytic function of z. From the expansion of $h(z)$, one has

$$t^2 = (z-\zeta)^2 \left[1 + \sum_{k=1}^{\infty} a_k (z-\zeta)^k\right],$$

or, equivalently,

$$t = (z-\zeta) \left[1 + \sum_{k=1}^{\infty} b_k (z-\zeta)^k\right],$$

for some computable coefficient sequences $\{a_k\}$ and $\{b_k\}$. The last expansion, upon inversion, provides

$$(z-\zeta) = t \left[1 + \sum_{k=1}^{\infty} c_k t^k\right],$$

which entails a similar expansion for the differential coefficient

$$\frac{dz}{z} = \frac{dt}{\zeta} \left[1 + \sum_{k=1}^{\infty} d_k t^k\right].$$

Thus, one has formally

$$I_n^{(0)} = \frac{1}{2\pi\zeta} \int_{-\tau}^{+\tau} e^{-nh''(\zeta)t^2/2} \left[1 + \sum_{k=1}^{\infty} d_k t^k\right] dt, \tag{6.39}$$

for some τ of the same order as δ.

A terminating form of order $2m$ of the expansion in (6.39) can justifiably be used and it yields

$$I_n^{(0)} = \frac{1}{2\pi\zeta}\int_{-\tau}^{+\tau} e^{-nh''(\zeta)t^2/2}\left[1+\sum_{k=1}^{2m-1} d_k t^k\right] dt + \mathcal{O}\left(\int_{-\tau}^{+\tau} e^{-nh''(\zeta)t^2/2} t^{2m}\, dt\right).$$

The integral in the remainder term is $\mathcal{O}(n^{-m})$. By a (now) standard argument, the integration range can be changed to $[-\infty, +\infty]$ introducing only exponentially small error terms. The integrals with a t^k of odd exponent k vanish. The integrals of even exponent are all expressible in terms of the Gamma function. One gets in this way

$$I_n^{(0)} = \frac{1}{\sqrt{2\pi n V}}\left[1+\sum_{k=1}^{m-1} d_k \frac{\Gamma(k-\frac{1}{2})}{\Gamma(-\frac{1}{2})}\frac{1}{(nV)^k}\right],$$

and the statement follows. □

We observe that the developments are a little simpler if one considers $N = \lambda n - 1$. Henrici develops a general derivation along these lines in Chapter 11 of his book [16].

A similar treatment can be inflicted to the exponential function as the the integral involved happens to be a pure power. In this way, a full asymptotic expansion of $1/n!$ equivalent to Stirling's formula comes out.

Exercise 23. Give explicit forms for the coefficients of the asymptotic expansion of $1/n!$ by using Lagrange inversion to support the change of variable $z \mapsto t$ in the proof of Theorem 6.6.

Discuss the general forms obtained in this way in Theorem 6.6 when $N = \lambda n - 1$.

Full asymptotic expansions can be determined for many admisssible functions. However, conditions stronger than plain admissibility are then required. The situation is also more intricate as several asymptotic scales interfere so that one does not generally obtain expansions in descending powers n or $\log n$, but rather in functions of the implicitly defined saddle point. This situation was already encountered in the dominant term analysis of Bell numbers.

Sets of conditions leading to full expansions have been given by Harris and Schoenfeld [14]. Odlyzko and Richmond showed in [24] that for any Hayman admissible function, a full expansion of $e^{f(z)}$ can be determined by Harris and Schoenfeld's method. The later situation covers for instance the

Bell numbers. In combinatorial practice, full expansions can be derived for most explicitly constructed generating functions.

Exercise 24. Show the existence of a full asymptotic expansion for the involution numbers and determine the corresponding asymptotic scale. Compute 20 terms of this expansion with a computer algebra system.

Exercise 25. Discuss the shape of a full asymptotic expansion for Bell numbers.

Exercise 26. Apply the method of Theorem 6.6 to an implicitly defined function $Y(z) = z\phi(Y(z))$ and compare the process with singularity analysis.

6.10 Gaussian distributions.

Saddle point analysis has consequences for multivariate asymptotics and is a direct way of proving that many discrete distributions tend to the Gaussian law in the asymptotic limit. For large powers, this property derives painlessly from our earlier developments, especially Theorem 6.5, by means of a "perturbation" analysis.

Theorem 6.7 (Gaussian limit law) *Assume that g satisfies the conditions of Theorem 6.4. Let λ_0 be such that $0 < \lambda_0 < d$ and define*

$$N_0 = \lambda_0 n, \quad N = \lambda_0 n + x\sqrt{n},$$

where x belongs to a fixed finite interval of the real line. Then, uniformly in x, one has

$$\frac{g_N^{(n)}}{g_{N_0}^{(n)}} = e^{-x^2/(2\sigma^2)}(1 + o(1))$$

where

$$\sigma^2 = \zeta\frac{g''(\zeta)}{g(\zeta)} + \frac{g'(\zeta)}{g(\zeta)} - \zeta\left(\frac{g'(\zeta)}{g(\zeta)}\right)^2 \quad \textit{with} \quad \zeta\frac{g'(\zeta)}{g(\zeta)} = \lambda_0.$$

Proof. [Sketch] Use the estimate of Theorem 6.5. The saddle point corresponding to $\lambda_0 + xn^{-1/2}$ is a function $\xi \equiv \xi(x, n)$. It is defined by the equation

$$\xi \frac{g'(\xi)}{g(\xi)} = \lambda_0 + \frac{x}{\sqrt{n}}$$

so that $\xi(0, n) = \zeta$ and $\xi(x, n) \to \zeta$ as $n \to \infty$. A complete expansion of $\xi(x, n)$ for fixed x and $n \to \infty$ is computed by standard devices. Propagation of this expansion into the main formula of Theorem 6.5 then yields the result. $\square$

In the particular case when $g(z)$ is the probability generating function of a discrete distribution, which implies $g(1) = 1$, and when λ_0 is taken to be the mean $g'(1)$ of the distribution, Theorem 6.7 describes a Gaussian approximation near the mean for the distribution. The results of Theorem 6.7 then simplify: the saddle point is at $\zeta = 1$, and the quantity σ^2 reduces to

$$\sigma^2 = g''(1) + g'(1) - (g'(1))^2,$$

which is the variance of the distribution.

The theorem then expresses the occurrence of a Gaussian law for the sum of n identicalldentify distributed random variables with probability generating function $g(z)$. In probability theory, this is usually called a *local limit theorem* as it is relative to the density of a probability distribution. In contrast estimates relative to the cumulative distribution function, which involves large "segments" of the probability distribution, are called *central limit theorems* or *integral limit theorems.*

Corollary 6.2 (Local limit theorem) *Let X be a discrete nonnegative random variable with probability generating function $g(z)$, mean $\mu = g'(1)$ and variance $\sigma^2 = g''(1) + g'(1) - (g'(1))^2$. Assume that $g(z)$ is aperiodic and analytic in $|z| < 1 + \epsilon$ for some $\epsilon > 0$. Then, the generating function $g^n(z)$ of a sum of n independent random variables equidistributed with X satisfies for integer $N = \mu n + x\sqrt{n}$*

$$[z^N](g(z))^n = \frac{1}{\sqrt{2\pi\sigma n}} e^{-x^2/(2\sigma^2)}(1 + o(1),$$

uniformly in n, for x in any fixed compact set of the real line.

Exercise 27. Prove a central limit theorem by the methods and under the conditions of Theorem 6.5.

EXERCISE 28. Extend the Gaussian law, for suitable $h(z)$, to

$$[z^n]\frac{h(z)}{h(1)}\left(\frac{g(z)}{g(1)}\right)^n.$$

EXERCISE 29. Use the saddle point method to prove that Stirling's distribution

$$[z^k]\frac{z(z+1)\cdots(z+n-1)}{n!}$$

is asymptotically normal. [Hint: consider $k \approx \log n + x\sqrt{\log n}$.]

EXERCISE 30. Show that Theorem 6.7 can conversely be deduced from a local limit theorem by the technique of "shifting the mean". [Hint: see [12].]

6.11 Combinatorial averages, distributions, and saddle points

Saddle point methods are useful not only for estimating combinatorial counts (sections 6.6 and 6.8), but also for analyzing asymptotically characteristics of combinatorial structures. Their range of applications in this context can be categorized as follows:

— Estimations of moments, especially mean and variance, when the corresponding generating functions are admissible (increasing subsequences, set partitions).

— Large deviation bounds for combinatorial distributions that are associated with powers or "quasi-powers" (cycles in permutations, capacity in random allocations).

— Indirect bounds for parameters that can be related combinatorially to simpler parameters of one of the previous types (longest increasing subsequence, height of permutation trees).

Increasing sequences in permutations. Define a *tagged permutation* as a permutation together with one of its increasing subsequence distinguished. (We also consider the null subsequence as an increasing subsequence.) For instance,

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

is a tagged permutation with the increasing subsequence 3 6 8 that is distinguished. The vertical bars are used to identifty the tagged elements, but they may also be interpreted as decomposing the permutation into subpermutation fragments.

Let $\mathcal{P}$ be the class of all permutations, $\mathcal{P}^+$ the subclass of non empty permutations, and $\mathcal{T}$ the class of tagged permutations. Then, it is readily recognized that, up to isomorphism,

$$\mathcal{T} = \mathcal{P} \star set\,(\mathcal{P}^+),$$

since a tagged permutation can be reconstructed from its initial fragment and the *set* of its fragments (by ordering the set according to increasing values of initial elements).

Thus, for EGFs, one has

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

The quantity $t_n = T_n/n!$ is precisely the mean number of increasing subsequences in a random permutation of size n. Analytically, this function is a variant of the "large singular function" of Sections 6.2, 6.6. Admissibility conditions are again easily checked and Theorem 6.3 gives

$$t_n \equiv \frac{T_n}{n!} \sim \frac{e^{-1/2} e^{2\sqrt{n}}}{2\sqrt{\pi}\, n^{1/4}}. \tag{6.40}$$

The result is due to Lifschitz and Pittel [19] who obtained it using real analysis methods. (See exercises below).

This analysis provides indirectly information about the parameter $\lambda(\sigma)$ representing the *length of the longest increasing subsequence* in σ, a much less accessible parameter. If $\iota(\sigma)$ is the number of increasing subsequences, then clearly

$$2^{\lambda(\sigma)} \leq \iota(\sigma).$$

Let ℓ_n be the expectation of λ over permutations of size n. Then, by convexity of the function 2^x, one has

$$2^{\ell_n} \leq t_n, \qquad \text{so that} \qquad \ell_n \leq \frac{2}{\log 2}\sqrt{n}(1+o(1)), \tag{6.41}$$

by (6.40).

Proposition 6.4 *The expected length of the longest increasing subsequence in a random permutation of size n satisfies*

$$\ell_n \le \frac{2}{\log 2}\sqrt{n}(1+o(1)).$$

The upper bound obtained in this way is of the form $2.89\sqrt{n}$. In fact, Logan, Shepp, Vershik, Kirov established the much more difficult result

$$\ell_n \sim 2\sqrt{n}.$$

(Their proof is based on a detailed analysis of the profile of a random Young tableau.) The bound obtained here by a simple mixture of saddle point estimates and combinatorial approximations already provides the right order of magnitude.

EXERCISE 31. Let $t_n = \frac{1}{n!}T_n$ be the expected number of increasing subsequences in a random permutation of $[1\,..\,n]$. By decomposing according to the location of n in the permutation, establish directly the recurrence

$$t_n = t_{n-1} + \frac{1}{n}\sum_{k=0}^{n-1} t_k, \quad t_0 = 1.$$

Hence $T(z)$ satisfies the ODE

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

which can be solved.

EXERCISE 32. One has

$$T_n = \sum_{k=0}^{n} \binom{n}{k}\frac{n!}{k!}.$$

Perform an analysis by the Laplace method for sums.

Parts in random set partitions. The function

$$f(z,u) = e^{u(e^z-1)}$$

is the bivariate generating function of set partitions with u marking the number of parts. We set $f(z) = f(z,1)$ and define

$$g(z) = \frac{\partial}{\partial u} f(z,u)\bigg|_{u=1} = e^{e^z+z-1}.$$

Thus, the quantity

$$\frac{g_n}{f_n} = \frac{[z^n]g(z)}{[z^n]f(z)}$$

represents the mean number of parts in a random partition of $[1\,..\,n]$. We already know (Section 6.6) that $f(z)$ is admissible and so is $g(z)$ by closure properties. The saddle point for the coefficient integral of $f(z)$ occurs at ζ such that $\zeta e^\zeta = n$, and it is already known that $\zeta = \log n - \log\log n + o(1)$.

It would be possible to analyze $g(z)$ by means of Theorem 6.3 directly: the analysis then involves a saddle point $\zeta_1 \neq \zeta$ that is relative to $g(z)$. An analysis of the mean would then follow, albeit at some computational effort. It is however more transparent to appeal to Lemma 6.1 and analyse the coefficients of $g(z)$ at the saddle point of $f(z)$.

Let $a(r), b(r)$ and $a_1(r), b_1(r)$ be the functions of Eq. (6.19) relative to $f(z)$ and $g(z)$ respectively:

$$\begin{array}{lcl@{\qquad\qquad}lcl}
\log f(z) & = & e^z - 1 & \log g(z) & = & e^z + z - 1 \\
a(r) & = & re^r & a_1(r) & = & re^r + r = a(r) + r \\
b(r) & = & (r^2+r)e^r & b_1(r) & = & (r^2+r)e^r + r = b(r) + r.
\end{array}$$

Thus, estimating g_n by Lemma 6.1 with the formula taken at $r = \zeta$, one finds

$$g_n = \frac{e^\zeta f(\zeta)}{\sqrt{2\pi b_1(\zeta)}}\left[\exp\left(-\frac{\zeta^2}{b_1(\zeta)}\right) + o(1)\right],$$

while the corresponding estimate for f_n is

$$f_n = \frac{f(\zeta)}{\sqrt{2\pi b_1(\zeta)}}[1 + o(1)].$$

Given that $b_1(\zeta) \sim b(\zeta)$ and that ζ^2 is of much smaller order than $b_1(\zeta)$, one has

$$\frac{g_n}{f_n} = e^\zeta(1 + o(1)) = \frac{n}{\log n}(1 + o(1)).$$

A similar computation applies to the second moment of the number of parts which is found to be asymptotic to $e^{2\zeta}$ (the computation involves taking a second derivative). Thus, the standard deviation of the number of parts is of an order $o(e^{\zeta})$ that is smaller than the mean. This implies a concentration property for the distribution of the number of parts.

Proposition 6.5 *The variable X_n equal to the number of parts in a random partition of $[1\,..\,n]$ has expectation*

$$E\{X_n\} = \frac{n}{\log n}(1 + o(1)).$$

The distribution satisfies a "concentration" property: for any $\epsilon > 0$, one has

$$\Pr\left\{\left|\frac{X_n}{E\{X_n\}} - 1\right| > \epsilon\right\} \to 0 \qquad \text{as} \quad n \to +\infty.$$

EXERCISE 33. Generalize to bivariate generating functions $e^{uh(z)}$, assuming $e^{h(z)}$ to be admissible. Find conditions for the mean value to be asymptotic to

$$h(\zeta),$$

with ζ the saddle point relative to e^h.

EXERCISE 34. The mean number of parts in a random integer partition of size n is $O(n^{1/2})$.

Cycles in permutations. A random permutation of size n has k cycles with probability equal to

$$\pi_{n,k} = \frac{1}{n!}\begin{bmatrix} n \\ k \end{bmatrix},$$

where $\left[{n \atop k}\right]$ is a Stirling cycle number (Stirling number of the first kind). As is well-known, this distribution also gives the statistics of left-right maxima in permutations. The horizontal generating function,

$$\frac{1}{n!}\sum_{k=0}^{n}\begin{bmatrix} n \\ k \end{bmatrix} u^k = \frac{u(u+1)(u+2)\cdots(u+n-1)}{n!}, \tag{6.42}$$

provides the mean and the variance of the distribution as

$$H_n = \log n + O(1), \qquad H_n - H_n^{(2)} = \log n + O(1).$$

The standard deviation being of an order smaller than the mean, the distribution is concentrated. The saddle point bounds further quantify this fact. Though (6.42) is not a power, the occurrence of a large number of factors makes it share many analytic properties of powers. (We can think of it as a "quasi-power".)

Let us examine first the probability p_n that a random permutation has $k = \lfloor 2\log n \rfloor$ cycles. From trivial bounds, one has for any fixed $\zeta \in]0, +\infty[$

$$p_n \le \frac{\zeta(\zeta+1)(\zeta+2)\cdots(\zeta+n-1)}{n!}\zeta^{-k}.$$

Now, Stirling's formula shows that, as $n \to \infty$, for a fixed ζ,

$$\frac{\zeta(\zeta+1)(\zeta+2)\cdots(\zeta+n-1)}{n!} \sim \frac{n^{\zeta-1}}{\Gamma(\zeta)}.$$

Thus, for any fixed ζ, one has

$$p_n \le O(1)\exp[\log n(\zeta - 1 - 2\log\zeta)].$$

The optimal choice of ζ, which corresponds to a saddle point bound, is obtained when the coefficient of $\log n$ attains its minimum, and this occurs at $\zeta = 2$, giving

$$p_n = O\left(n^{1-2\log 2}\right). \tag{6.43}$$

Let q_n be the probability that a permutation has no more than $\lfloor 2\log n \rfloor$ cycles. The technique for right tails developed in Section 6.2 gives

$$\begin{aligned} q_n &\le O(1)n^{1-2\log 2}\left(1 + \frac{1}{\zeta} + \frac{1}{\zeta^2} + \cdots\right) \\ &= O\left(n^{1-2\log 2}\right). \end{aligned} \tag{6.44}$$

A similar computation applies to the number of permutations with no more than $x\log n$ cycles, for any fixed $x > 1$, the optimal choice of ζ being then $\zeta = x$.

Proposition 6.6 *For any fixed x, the probability that a permutation has at least $x\log n$ cycles is of the form*

$$O\left(n^{-1+x-x\log x}\right).$$

This technique has been put to use by Robson [27] who obtained similar bounds for the distribution of nodes in strata of random permutation trees (*i.e.*, heap-ordered trees and binary search trees). The distribution again involves the Stirling cycle numbers. In this way, Robson was able to derive an upper bound on the height of permutation trees of the form

$$C \log n \qquad \text{where} \qquad C \cong 4.31\,.$$

Once more, the upper bound obtained in this way is excellent as evidenced by the fact that it does in fact represent the exact order of growth. In effect, Devroye [6, 7] later showed by means of branching processes that the expected height is asymptotic to $C \log n$ with C being Robson's constant.

Exercise 35. Prove similarly left tail estimates bounding the probability that a permutation has a smaller number of cycles than expected.

Exercise 36. Show that the mean number of external nodes at altitude k in a random permutation tree of size n is

$$[z^n u^k]\frac{1}{(1-z)^{2u}}.$$

Prove Robson's upper bound on height of permutation trees.

Capacity in occupancy problems. A word of length n over an alphabet of cardinality m can be viewed as an allocation table that describes the way n balls are thrown into m urns. The size of the most filled urn, which corresponds to the maximum number of occurrences of any letter from the m-ary alphabet in the table, is called the *capacity*. We let $C_{n,m}$ denote this random variable, when all m^n tables are taken equally likely.

For many applications, m and n grow roughly proportionately. This is the case in hashing algorithms where n keys are thrown into m lists (buckets, urns) by means of a hash function; the constraint $n/m = O(1)$ ensures a constant retrieval time on average. (Knowledge of the capacity is useful, especially in the context of paging.)

Proposition 6.7 *Let n and m tend simultaneously to infinity, with the constraint that $\frac{n}{m} = \alpha$ remains constant. Then, the expected capacity satisfies*

$$\frac{1}{2}\frac{\log n}{\log\log n}(1+o(1)) \le E\{C_{n,m}\} \le 2\frac{\log n}{\log\log n}(1+o(1)).$$

Proof. We detail the proof when $\alpha = 1$ and abbreviate $C_n = C_{n,m}$. From Chapter 2, we know that

$$\begin{cases} \Pr\{C_n \le b\} &= \dfrac{n!}{n^n}[z^b](e_b(z))^n \\ \Pr\{C_n > b\} &= \dfrac{n!}{n^n}\left(e^{nz} - (e_b(z))^n\right), \end{cases} \tag{6.45}$$

where $e_b(z)$ is the truncated exponential:

$$e_b(z) = \sum_{j=0}^{b} \frac{z^j}{j!}.$$

The two equalities of (6.45) permit to bound the left part and right part of the distribution of capacity. We know already, from the study of large powers, that the central part of a distribution may be approached via the saddle point value $\zeta \approx 1$. Thus, taking saddle point bounds at 1, we get

$$\begin{cases} \Pr\{C_n \le b\} &\le \dfrac{n!e^n}{n^n}\left(\dfrac{e_b(1)}{e}\right)^n \\ \Pr\{C_n > b\} &\le \dfrac{n!e^n}{n^n}\left(1 - \left(\dfrac{e_b(1)}{e}\right)^n\right). \end{cases} \tag{6.46}$$

We set

$$\rho_b(n) = \left(\frac{e_b(1)}{e}\right)^n. \tag{6.47}$$

This quantity represents the probability that n Poisson variables of rate 1 all have value b or less. (We know for elementary probability theory that this should be a reasonable approximation of the problem at hand.) A weak form of Stirling's formula,

$$\frac{n!e^n}{n^n} < 2\sqrt{\pi n} \qquad (n \ge 1)$$

then yields an alternative form of (6.46),

$$\begin{cases} \Pr\{C_n \le b\} &\le 2\sqrt{\pi n}\rho_b(n) \\ \Pr\{C_n > b\} &\le 2\sqrt{\pi n}\,(1 - \rho_b(n)). \end{cases} \tag{6.48}$$

For fixed n, the function $\rho_b(n)$ increases steadily from e^{-n} to 1 as b varies from 0 to ∞. The problem is thus reduced to analyzing its variation. In particular, the "transition region" where $\rho_b(n)$ stays away from both 0 and 1 is expected to play a rôle. This suggests defining $b_0 \equiv b_0(n)$ such that

$$b_0! \le n < (b_0+1)!,$$

so that

$$b_0(n) = \frac{\log n}{\log\log n}(1+o(1)).$$

We also observe that, as $n, b \to \infty$,

$$\begin{aligned} \rho_b(n) &= (e^{-1}e_b(1))^n = \left(1 - \frac{e^{-1}}{(m+1)!} + O(\frac{1}{(m+2)!})\right)^n \\ &= \exp\left(-\frac{ne^{-1}}{(m+1)!} + O(\frac{n}{(m+2)!})\right). \end{aligned} \tag{6.49}$$

Left tail. We take $b = \lfloor \frac{1}{2} b_0 \rfloor$ and a simple computation from (6.49) shows that for n large enough,

$$\rho_b(n) \le \exp(-\sqrt[3]{n}).$$

Thus, by the first inequality of (6.48), the probability that the capacity be less than $\frac{1}{2}b_0$ is exponentially small:

$$\Pr\{C_n \le \frac{1}{2} b_0(n)\} \le 2\sqrt{\pi n}\exp(-\sqrt[3]{n}). \tag{6.50}$$

Right tail. Take $b = 2b_0$. Then, again from (6.49), for n large enough,

$$1 - \rho_b(n) \le 1 - \exp(-\frac{1}{n}) = \frac{1}{n}(1+o(1)).$$

Thus, the probability of observing a capacity that exceeds $2b_0$ is vanishingly small, and is $O(n^{-1/2})$.

Taking next $b = 2b_0 + r$ with $r > 0$, similarly gives the bound

$$\Pr\{C_n > 2b_0(n) + r\} \le 2\sqrt{\frac{\pi}{n}}\left(\frac{1}{b_0(n)}\right)^r. \tag{6.51}$$

Then Equations (6.50) and (6.51) imply

$$\left\{ \begin{aligned} \mathrm{E}\{C_n\} &\le 2b_0(n) + \sum_{r=0}^{\infty} 2\sqrt{\frac{\pi}{n}}(b_0(n))^{-r} = 2b_0(n)(1+o(1)) \\ \mathrm{E}\{C_n\} &\ge \sum_{r=0}^{\lfloor \frac{1}{2} b_0(n) \rfloor} \left[1 - 2\sqrt{\pi n}\exp(-\sqrt[3]{n})\right] = \frac{1}{2} b_0(n)(1+o(1)). \end{aligned} \right. \tag{6.52}$$

This justifies the claim of the proposition when $\alpha = 1$. The general case ($\alpha \neq 1$) follows similarly from saddle point bounds taken at $\zeta = \alpha$. □

The analysis in the proof above is not tight. Actually, the full saddle point method may be used, which eliminates the spurious $\sqrt{n}$ factors in the estimates above. In this way, it can be proved that the expected capacity satisfies, for any fixed $\alpha = n/m$:

$$\mathrm{E}\{C_{n,m}\} \sim \frac{\log n}{\log\log n}.$$

This result, in the context of longest probe sequences in hashing, was obtained by Gonnet [11] under the Poisson model. Many key estimates regarding random allocations (including capacity) are to be found in the book by Kolchin *et al.* [18]. Analysis of the type discussed above are also useful in evaluating various dynamic hashing algorithms by saddle point methods [8, 26].

6.12 Notes

Saddle point methods take their sources in applied mathematics, one of them being the asymptotic analysis by Debye (1909) of Bessel functions of large order. Saddle point analysis is sometimes called steepest descent analysis, especially when integration contours strictly coincide with steepest descent paths. Saddle points themselves are also called critical points (*i.e.*, points where a first derivative vanishes). Because of its roots in applied mathematics, the method is well covered by the literature in this area, and we refer to the books by Olver [25], Henrici [16] or Wong [30] for extensive discussions. A vivid introduction to the subject is to be found in De Bruijn's book [5]. We also recommend Odlyzko's impressive survey [23].

To a large extent, saddle point methods have made an irruption in combinatorial enumerations in the 1950's. Early combinatorial papers were concerned with permutations (involutions) or set partitions: this includes works by Moser and Wyman [20, 21, 22] that are mostly directed towards entire functions.

Hayman's approach [15] which we have exposed here (see also [30]) is notable in its generality as it considered saddle point analysis from a more abstract perspective by introducing general closure theorems. A similar thread was followed by Harris and Schoenfeld who gave stronger conditions then permitting full asymptotic expansions [14]; Odlyzko and Richmond [24] were successful in connecting this with Hayman admissibility. Another valuable work is Wyman's extension to nonpositive functions [31].

Interestingly enough, developments that parallel the ones in combinatorial analysis have taken place in other regions of mathematics.

In 1954, Daniels [3] introduced saddle point methods to obtain refined versions of the central limit theorem of probability theory. See for instance the description in Greene and Knuth's book [12]. Since then, the saddle point method has proved a useful tool for deriving Gaussian limiting distributions. We have given here some idea of this approach which is to be developed further in a later chapter, where we shall discuss some of Canfield's results [2]. In a related context, Flajolet and Odlyzko proved that iterates of positive polynomials lead to Gaussian laws [10], a property that is useful in the combinatorial analysis of balanced trees.

Analytic number theory also makes a heavy use of saddle point analysis. In additive number theory, the works by Hardy, Littlewood, and Ramanujan relative to integer partitions were especially influential, see for instance Andrews' book [1] and Hardy's book on Ramanujan [13] for a fascinating perspective. In multiplicative number theory, generating functions take the form of Dirichlet series while Perron's formula replaces Cauchy's formula. For saddle point methods in this context, we refer to Tenenbaum's book [29] and his seminar survey [28].

A more global perspective on limit probability distributions will be given in a later chapter.

Problems and Exercises

A large number of fast growing functions are amenable to saddle point analysis.

EXERCISE 37. Give an asymptotic equivalent for

$$[z^n]\exp(\frac{1}{(1-z)^\alpha}).$$

EXERCISE 38. Compare

$$[z^n]\tan(e^z-1) \qquad \text{and} \qquad [z^n]e^{\tan z - 1}.$$

EXERCISE 39. Discuss

$$[z^n](g(z))^n h(z)$$

when $g(z)$ is positive but h is not constrained. In particular discuss the shape of the asymptotic equivalent when several of the derivatives of h vanish at the saddle point.

Find an asymptotic equivalent for fixed r and $n \to +\infty$ of

$$[z^n](1-z)^r(1+z)^{2n}.$$

EXERCISE 40. Analyse

$$[z^n]\frac{1}{(1-z)^\alpha}e^{z/(1-z)}, \qquad [z^n]\log\frac{1}{1-z}e^{z/(1-z)}.$$

Devise an elementary perturbation theory of saddle points in order to cover this and similar cases.

Nonpositive functions can also be analysed by saddle point methods. In this case, selection of a saddle point path is dictated by the considerations of Section 6.3. Rather unexpected types of fluctuations may arise in such cases.

EXERCISE 41. Analyse

$$T_n = [z^n]\exp(-\frac{z}{1-z}) \qquad \text{and} \qquad U_n = \sum_k \binom{n}{k}\frac{(-1)^k}{k!}.$$

EXERCISE 42. Analyze

$$[z^n]\exp(\frac{z}{1+z^2}).$$

Bibliography

[1] Andrews, G. E. *The Theory of Partitions*, vol. 2 of *Encyclopedia of Mathematics and its Applications.* Addison–Wesley, 1976.

[2] 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.

[3] Daniels, H. E. Saddlepoint approximations in statistics. *Annals of Mathematical Statistics 25* (1954), 631–650.

[4] Darboux, G. Mémoire sur l'approximation des fonctions de très grands nombres, et sur une classe étendue de développements en série. *Journal de Mathématiques Pures et Appliquées* (Feb. 1878), 5–56,377–416.

[5] De Bruijn, N. G. *Asymptotic Methods in Analysis.* Dover, 1981. A reprint of the third North Holland edition, 1970 (first edition, 1958).

[6] Devroye, L. A note on the expected height of binary search trees. *Journal of the ACM 33* (1986), 489–498.

[7] Devroye, L. Branching processes in the analysis of the heights of trees. *Acta Informatica 24* (1987), 277–298.

[8] Flajolet, P. On the performance evaluation of extendible hashing and trie searching. *Acta Inf. 20* (1983), 345–369.

[9] Flajolet, P., Grabner, P., Kirschenhofer, P., and Prodinger, H. On Ramanujan's Q–function. Research Report 1760, Institut National de Recherche en Informatique et en Automatique, Oct. 1992. 13 pages. Accepted for publication by *Journal of Computational and Applied Mathematics.*

[10] Flajolet, P., and Odlyzko, A. Limit distributions for coefficients of iterates of polynomials with applications to combinatorial enumerations. *Mathematical Proceedings of the Cambridge Philosophical Society 96* (1984), 237–253.

[11] Gonnet, G. H. Expected length of the longest probe sequence in hash code searching. *Journal of the ACM 28*, 2 (1981), 289–304.

[12] Greene, D. H., and Knuth, D. E. *Mathematics for the analysis of algorithms*, second ed. Birkhauser, Boston, 1982.

[13] Hardy, G. H. *Ramanujan: Twelve Lectures on Subjects Suggested by his Life and Work*, third ed. Chelsea Publishing Company, New-York, 1978. Reprinted and Corrected from the First Edition, Cambridge, 1940.

[14] Harris, B., and Schoenfeld, L. Asymptotic expansions for the coefficients of analytic functions. *Illinois Journal of Mathematics 12* (1968), 264–277.

[15] Hayman, W. K. A generalization of Stirling's formula. *Journal für die reine und angewandte Mathematik 196* (1956), 67–95.

[16] Henrici, P. *Applied and Computational Complex Analysis*. John Wiley, New York, 1977. 3 volumes.

[17] Knuth, D. E. *The Art of Computer Programming*, vol. 3: Sorting and Searching. Addison-Wesley, 1973.

[18] Kolchin, V. F., Sevastyanov, B. A., and Chistyakov, V. P. *Random Allocations*. John Wiley and Sons, New York, 1978. Translated from the Russian original *Slučajnye Razmeščeniya*.

[19] Lifschitz, V., and Pittel, B. The number of increasing subsequences of the random permutation. *Journal of Combinatorial Theory, Series A 31* (1981), 1–20.

[20] Moser, L., and Wyman, M. On the solution of $x^d = 1$ in symmetric groups. *Canadian Journal of Mathematics 7* (1955), 159–168.

[21] Moser, L., and Wyman, M. Asymptotic expansions. *Canadian Journal of Mathematics 8* (1956), 225–233.

[22] Moser, L., and Wyman, M. Asymptotic expansions II. *Canadian Journal of Mathematics* (1957), 194–209.

[23] Odlyzko, A. M. Explicit Tauberian estimates for functions with positive coefficients. *Journal of Computational and Applied Mathematics 41* (1992), 187–197.

[24] Odlyzko, A. M., and Richmond, L. B. Asymptotic expansions for the coefficients of analytic generating functions. *Aequationes Mathematicae 28* (1985), 50–63.

[25] Olver, F. W. J. *Asymptotics and Special Functions*. Academic Press, 1974.

[26] Régnier, M. Analysis of grid file algorithms. *BIT 25* (1985), 335–357.

[27] Robson, J. M. The height of binary search trees. *Australian Computer Journal 11* (1979), 151–153.

[28] Tenenbaum, G. La méthode du col en théorie analytique des nombres. In *Séminaire de Théorie des Nombres, Paris 1985–1986* (1988), C. Goldstein, Ed., Birkhauser, pp. 411–441.

[29] Tenenbaum, G. *Introduction à la théorie analytique des nombres*, vol. 13. Institut Élie Cartan, Nancy, France, 1990.

[30] Wong, R. *Asymptotic Approximations of Integrals*. Academic Press, 1989.

[31] Wyman, M. The asymptotic behavior of the Laurent coefficients. *Canadian Journal of Mathematics 11* (1959), 534–555.

Contents