Asymptotic Enumeration Methods

A. M. Odlyzko

AT&T Bell Laboratories
Murray Hill, New Jersey 07974

1. Introduction

Asymptotic enumeration methods provide quantitative information about the rate of growth of functions that count combinatorial objects. Typical questions that these methods answer are: (1) How does the number of partitions of a set of n elements grow with n? (2) How does this number compare to the number of permutations of that set?

There do exist enumeration results that leave nothing to be desired. For example, if a_n denotes the number of subsets of a set with n elements, then we trivially have $a_n = 2^n$. This answer is compact and explicit, and yields information about all aspects of this function. For example, congruence properties of a_n reduce to well-studied number theory questions. (This is not to say that all such questions have been answered, though!) The formula $a_n = 2^n$ also provides complete quantitative information about a_n. It is easy to compute for any value of n, its behavior is about as simple as possible, and it holds uniformly for all n. However, such examples are extremely rare. Usually, even when there is a formula for the function we are interested in, it is a complicated one, involving summations or recurrences. The purpose of asymptotic methods is to provide simple explicit formulas that describe the behavior of a sequence for large values of indices. There is no satisfactory definition of what is meant by "simple" or by "explicit." However, we can illustrate this concept by some examples. The number of permutations of n letters is given by $b_n = n!$. This is a compact notation, but only in the sense that factorials are so widely used that they have a special symbol. The symbol $n!$ stands for $n \cdot (n-1) \cdot (n-2) \cdot \ldots \cdot 2 \cdot 1$, and it is the latter formula that has to be used to answer questions about the number of permutations. If one is after arithmetic information, such as the highest power of 7, say, that divides $n!$, one can obtain it from the product formula, but even then some work has to be done. For most quantitative purposes, however, $n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$ is inadequate. Since this formula is a product of n terms, most of them large, it is clear that $n!$ grows rapidly, but it is not obvious just how rapidly. Since all but the last term are ≥ 2, we have $n! \geq 2^{n-1}$, and since all but the last two terms are ≥ 3, we have $n! \geq 3^{n-2}$, and so on. On the other hand, each term is $\leq n$, so $n! \leq n^n$. Better bounds can clearly be obtained with greater care. The question such estimates raise is just how far can one go? Can one obtain an estimate for $n!$ that is easy to understand, compute, and manipulate? One answer provided by asymptotic methods is Stirling's formula: $n!$ is asymptotic to $(2\pi n)^{1/2}(n/e)^n$ as $n \to \infty$, which means that the limit as $n \to \infty$ of $n!(2\pi n)^{-1/2}(n/e)^{-n}$ exists and equals 1. This formula is concise and gives a useful representation of the growth rate of $n!$. It shows, for example, that for n large, the number of permutations on n letters is considerably larger than the number of subsets of a set with $\lfloor \frac{1}{2} n \log n \rfloor$ elements.

Another simple example of an asymptotic estimate occurs in the "problème des rencontres" [81]. The number d_n of *derangements* of n letters, which is the number of ways of handing back hats to n people so that no person receives his or her own hat, is given by

$$d_n = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \ . \tag{1.1}$$

This is a nice formula, yet to compute d_n exactly with it requires substantial effort, since the summands are large, and at first glance it is not obvious how large d_n is. However, we can obtain from (1.1) the asymptotic estimate

$$\frac{d_n}{n!} \to e^{-1} \quad \text{as} \quad n \to \infty \ . \tag{1.2}$$

To prove (1.2), we factor out $n!$ from the sum in (1.1). We are then left with a sum of rapidly decreasing terms that make up the initial segment of the series

$$e^{-1} = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \ ,$$

and (1.2) follows easily. It can even be shown that d_n is the nearest integer to $e^{-1}n!$ for all $n \geq 1$, see [81]. The estimate (1.2) does not allow us to compute d_n, but combined with the estimate for $n!$ cited above it shows that d_n grows like $(2\pi n)^{1/2} n^n e^{-n-1}$. Further, (1.2) shows that the fraction of all ways of handing out hats that results in every person receiving somebody else's hat is approximately $1/e$. Results of this type are often exactly what is desired.

Asymptotic estimates usually provide information only about the behavior of a function as the arguments get large. For example, the estimate for $n!$ cited above says only that the ratio of $n!$ to $(2\pi n)^{1/2}(n/e)^n$ tends to 1 as n gets large, and says nothing about the behavior of this ratio for any specific value of n. There are much sharper and more precise bounds for $n!$, and they will be presented in Section 3. However, it is generally true that the simpler the estimate, the weaker and less precise it is. There seems to be an unavoidable tradeoff between conciseness and precision. Just about the simplest formula that exactly expresses $n!$ is $n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1$. (We have to be careful, since there is no generally accepted definition of simplicity, and in many situations it is better to use other exact formulas for $n!$, such as the integral formula $n! = \int_0^\infty t^n e^{-t} dt$ for the Γ-function. There are also methods for evaluating $n!$ that are somewhat more efficient than the straightforward evaluation of the product.) Any other formula is likely to involve some loss of accuracy as a penalty for simplicity.

Sometimes, the tradeoffs are clear. Let $p(n)$ denote the number of partitions of an integer n. The Rademacher convergent series representation [13, 23] for $p(n)$ is valid for any $n \geq 1$:

$$p(n) = \pi^{-1} 2^{-1/2} \sum_{m=1}^{\infty} A_m(n) m^{1/2} \frac{d}{dv} (\lambda_v^{-1} \sinh(C m^{-1} \lambda_v)) \Big|_{v=n} , \tag{1.3}$$

where

$$C = \pi (2/3)^{1/2}, \quad \lambda_v = (v - 1/24)^{1/2} , \tag{1.4}$$

and the $A_m(n)$ satisfy

$$A_1(n) = 1, \quad A_2(n) = (-1)^n \quad \text{for all} \quad n \geq 1 ,$$

$$|A_m(n)| \leq m, \qquad \text{for all} \quad m, n \geq 1 ,$$

and are easy to compute. Remarkably enough, the series (1.3) does yield the exact integer value of $p(n)$ for every n, and it converges rapidly. (Although this is not directly relevant, we note that using this series to compute $p(n)$ gives an algorithm for calculating $p(n)$ that is close to optimal, since the number of bit operations is not much larger than the number of bits of $p(n)$.) By taking more and more terms, we obtain better and better approximations. The first term in (1.3) shows that

$$p(n) = \pi^{-1} 2^{-1/2} \frac{d}{dv} (\lambda_v^{-1} \sinh(C \lambda_v)) \Big|_{v=n} + O(n^{-1} \exp(C n^{1/2}/2)) , \tag{1.5}$$

and if we don't like working with hyperbolic sines, we can derive from (1.5) the simpler (but less precise) estimate

$$p(n) = \frac{1 + O(n^{-1/2})}{4 \cdot 3^{1/2} n} e^{C n^{1/2}} , \tag{1.6}$$

valid for all $n \geq 1$. Unfortunately, exact and rapidly convergent series such as (1.3) occur infrequently in enumeration, and in general we have to be content with poorer approximations.

The advantage of allowing parameters to grow large is that in surprisingly many cases, even when there do exist explicit expressions for the functions we are interested in, this procedure does yield simple asymptotic approximations, when the influence of less important factors falls off. The resulting estimates can then be used to compare numbers of different kinds of objects, decide what the most common objects in some category are, and so on. Even in situations where bounds valid for all parameter values are needed, asymptotic estimates can be used to suggest what form those bounds should take. Usually the error terms in asymptotic estimates can be made explicit (although good bounds often require substantial work), and can be used together with computations of small values to obtain universal estimates. It is common that already for n not much larger than 10 (where n is the basic parameter) the asymptotic estimate is accurate to within a few percent, and for $n \geq 100$ it is accurate to within a fraction of a percent, even though known proofs do not guarantee results as good as this. Therefore the value of asymptotic estimates is much greater than if they just provided a picture of what happens at infinity.

Under some conditions, asymptotic results can be used to prove completely uniform results. For example, if there were any planar maps that were not four-colorable, then almost every large planar map would not be four-colorable, as it would contain one of those small pathological maps. Therefore if it could be proved that most large planar maps are four-colorable, we would obtain a new proof of the four-color theorem that would be more satisfactory to many people than the original one of Haken and Appel. Unfortunately, while this is an attractive idea, no proof of the required asymptotic estimate for the normal chromatic number of planar maps has been found so far.

Asymptotic estimates are often useful in deciding whether an identity is true. If the growth rates of the two functions that are supposed to be equal are different, then the coincidence of initial values must be an accident. There are also more ingenious ways, such as that of Example 13.1, for deducing nonexistence of identities in a wide class from asymptotic information. Sometimes asymptotics is used in a positive way, to suggest what identities might hold.

Simplicity is an important advantage of asymptotic estimates. They are even more useful when no explicit formulas for the function being studied are available, and one has to deal with indirect relations. For example, let T_n be the number of rooted unlabeled trees with n vertices, so that $T_0 = 0$, $T_1 = T_2 = 1$, $T_3 = 2$, $T_4 = 4, \ldots$. No explicit formula for the T_n is known. However, if

$$T(z) = \sum_{n=1}^{\infty} T_n z^n \tag{1.7}$$

is the ordinary generating function of T_n, then Cayley and Pólya showed that

$$T(z) = z \exp \left(\sum_{k=1}^{\infty} T(z^k)/k \right) \ . \tag{1.8}$$

This functional equation can be derived using the general Pólya-Redfield enumeration method, an approach that is sketched in Section 15. Example 15.1 shows how analytic methods can be used to prove, starting with Eq. (1.8), that

$$T_n \sim C r^{-n} n^{-3/2} \quad \text{as} \quad n \to \infty \ , \tag{1.9}$$

where

$$C = 0.4399237 \ldots \ , \quad r = 0.3383219 \ldots \ , \tag{1.10}$$

are constants that can be computed efficiently to high precision. For $n = 20$, $T_n = 12,826,228$, whereas $Cr^{-20}20^{-3/2} = 1.274 \ldots \times 10^7$, so asymptotic formula (1.9) is accurate to better than 1%. Thus this approximation is good enough for many applications. It can also be improved easily by adding lower order terms.

Asymptotic enumeration methods are a subfield of the huge area of general asymptotic analysis. The functions that occur in enumeration tend to be of restricted form (often nonnegative and of regular growth, for example) and therefore the repertoire of tools that are commonly used is much smaller than in general asymptotics. This makes it possible to attempt a concise survey of the most important techniques in asymptotic enumeration. The task is not easy, though, as there has been tremendous growth in recent years in combinatorial enumeration and the closely related field of asymptotic analysis of algorithms, and the sophistication of the tools that are commonly used has been increasing rapidly.

In spite of its importance and growth, asymptotic enumeration has seldom been presented in combinatorial literature at a level other than that of a research paper. There are several books that treat it [43, 81, 175, 177, 235, 236, 237, 377], but usually only briefly. The only comprehensive survey that is available is the excellent and widely quoted paper of Bender [33]. Unfortunately it is somewhat dated. Furthermore, the last two decades have also witnessed a flowering of asymptotic analysis of algorithms, which was pioneered and popularized by Knuth. Combinatorial enumeration and analysis of algorithms are closely related, in that both deal with counting of particular structures. The methods used in the two fields are almost the same, and there has been extensive cross-fertilization between them. The literature on theoretical computer science, especially on average case analysis of algorithms, can therefore be used fruitfully in asymptotic enumeration. One notable survey paper in that area is that of Vitter and Flajolet [371]. There are also presentations of relevant methods in the books [177, 209, 235, 236, 237, 223]. Section 18 is a guide to the literature on these topics.

The aim of this chapter is to survey the most important tools of asymptotic enumeration, point out references for the results and methods that are discussed, and to mention additional relevant papers that have other techniques that might be useful. It is intended for a reader who has already used combinatorial, algebraic, or probabilistic methods to reduce a problem to that of estimating sums, coefficients of a generating function, integrals, or terms in a sequence satisfying some recursion. How such a reduction is to be accomplished will be dealt with sparingly, since it is a large subject that is already covered extensively in other chapters, especially [**?**]. We will usually assume that this task has been done, and will discuss only the derivation of asymptotic estimates.

The emphasis in this chapter is on elementary and analytic approaches to asymptotic problems, relying extensively on explicit generating functions. There are other ways to solve some of the problems we will discuss, and probabilistic methods in particular can often be used instead. We will only make some general remarks and give references to this approach in Section 16.

The only methods that will be discussed in detail are fully rigorous ones. There are also methods, mostly from classical applied mathematics (cf. [31]) that are powerful and often give estimates when other techniques fail. However, we do not treat them extensively (aside from some remarks in Section 16.4) since many of them are not rigorous.

Few proofs are included in this chapter. The stress is on presentation of basic methods, with discussions of their range of applicability, statements of general estimates derivable from them, and examples of their applications. There is some repetitiveness in that several functions, such as $n!$, are estimated several times. The purpose of doing this is to show how different methods compare in their power and ease of use. No attempt is made to present derivations starting from first principles. Some of the examples are given with full details of the asymptotic analysis, to explain the basic methods. Other examples are barely more than statements of results with a brief explanation of the method of proof and a reference to where the proof can be found. The reader might go through this chapter, possibly in a random order, looking for methods that might be applicable to a specific problem, or can look for a category of methods that might fit the problem and start by looking at the corresponding sections.

There are no prerequisites for reading most of this chapter, other than acquaintance with advanced calculus and elementary asymptotic estimates. Many of the results are presented so that they can be used in a cookbook fashion. However, many of the applications require knowledge of complex variables.

Section 2 presents the basic notation used throughout the chapter. It is largely the standard one used in the literature, but it seemed worthwhile summarizing it in one place. Section 3 is devoted to a brief discussion of identities and related topics. While asymptotic methods are useful and powerful, they can often be either augmented or entirely replaced by identities, and this section points out how to use them.

Section 4 summarizes the most important and most useful estimates in combinatorial enumeration, namely those related to factorials and binomial coefficients. Section 5 is the first one to feature an in-depth discussion of methods. It deals with estimates of sums in terms of integrals, summation formulas, and the inclusion-exclusion principle. However, it does not present the most powerful tool for estimation of sums, namely generating functions. These are introduced in Section 6, which presents some of the basic properties of, and tools for dealing with generating functions. While most generating functions that are used in combinatorial enumeration converge at least in some neighborhood of the origin, there are also many nonconvergent ones. Section 7 discusses some estimates that apply to all formal series, but are especially useful for nonconvergent ones.

Section 8 is devoted to estimates for convergent power series that do not use complex variables. While not as powerful as the analytic methods presented later, these techniques are easy to use and suffice in many applications.

Section 9 presents a variety of techniques for determining the asymptotics of recurrence relations. Many of these methods are based on generating functions, and some use analytic methods that are discussed later in the chapter. They are presented at this point because they are basic to combinatorial enumeration, and they also provide an excellent illustration of the power of generating functions.

Section 10 is an introduction to the analytic methods for estimating generating functions. Many of the results mentioned here are common to all introductory complex analysis courses. However, there are also many, especially those in Sections 10.4 and 10.5, are not as well known, and are of special value in asymptotics.

Sections 11 and 12 present the main methods used in estimation of coefficients of analytic functions in a single variable. The basic principle is that the singularities of the generating function that are closest to the origin determine the growth rate of the coefficients. If the function does not grow too fast as it approaches those singularities, the methods of Section 11 are usually applicable, while if the growth rate is high, methods of Section 12 are more appropriate.

Sections 13–15 discuss extensions of the basic methods of Sections 10–12 to multivariate generating functions, integral transforms, and problems that involve a combination of methods.

Section 16 is a collection of miscellaneous methods and results that did not easily fit into any other section, yet are important in asymptotic enumeration. Section 17 discusses the extent to which computer algebra systems can be used to derive asymptotic information. Finally, Section 18 is a guide to further reading on asymptotics, since this chapter does not provide complete coverage of the topic.

2. Notation

The symbols O, o, and $\sim$ will have the usual meaning throughout this paper:

$$f(z) = O(g(z)) \text{ as } z \to w \text{ means } f(z)/g(z) \text{ is bounded as } z \to w ;$$

$$f(z) = o(g(z)) \text{ as } z \to w \text{ means } f(z)/g(z) \to 0 \text{ as } z \to w ;$$

$$f(z) \sim g(z) \text{ as } z \to w \text{ means } f(z)/g(z) \to 1 \text{ as } z \to w .$$

When an asymptotic relation is stated for an integer variable n instead of z, it will implicitly be taken to apply only for integer values of $n \to w$, and then we will always have $w = \infty$ or $w = -\infty$. An introduction to the use of this notation can be found in [175]. Only a slight acquaintance with it is assumed, enough to see that $(1 + O(n^{-1/3}))^n = \exp(O(n^{2/3}))$ and $\log(n + n^{1/2}) = \log(n) + n^{-1/2} - (2n)^{-1} + O(n^{-3/2})$.

The notation $x \to w^-$ for real w means that x tends to w only through values $x < w$.

Some asymptotic estimates refer to *uniform convergence*. As an example, the statement that $f(z) \sim (1-z)^{-2}$ as $z \to 1$ uniformly in $|\text{Arg}(1-z)| < 2\pi/3$ means that for every $\epsilon > 0$, there is a $\delta < 0$ such that

$$|f(z)(1-z)^2 - 1| \le \epsilon$$

for all z with $0 < |1-z| < \delta$, $|\text{Arg}(1-z)| < 2\pi/3$. This is an important concept, since lack of uniform convergence is responsible for many failures of asymptotic methods to yield useful results.

Generating functions will usually be written in the form

$$f(z) = \sum_{n=0}^{\infty} f_n z^n \ , \tag{2.1}$$

and we will use the notation $[z^n]f(z)$ for the coefficient of z^n in $f(z)$, so that if $f(z)$ is defined by (2.1), $[z^n]f(z) = f_n$. For multivariate generating functions, $[x^m y^n]f(x,y)$ will denote the coefficient of $x^m y^n$, and so on. If a_n denotes a sequence whose asymptotic behavior is to be studied, then in combinatorial enumeration one usually uses either the *ordinary generating function* $f(z)$ defined by (2.1) with $f_n = a_n$, or else the *exponential generating function* $f(z)$ defined by (2.1) with $f_n = a_n/n!$. In this chapter we will not be concerned with the question of which type of generating function is best in a given context, but will assume that a generating function is given, and will concentrate on methods of extracting information about the coefficients from the form we have.

Asymptotic series, as defined by Poincaré, are written as

$$f_n \sim \sum_{k=0}^{\infty} a_k n^{-k} \ , \tag{2.2}$$

and mean that for every $K \geq 0$,

$$f_n = \sum_{k=0}^{K} a_k n^{-k} + O(n^{-K-1}) \quad \text{as} \quad n \to \infty \ . \tag{2.3}$$

The constant implied by the O-notation may depend on K. It is unfortunate that the same symbol is used to denote an asymptotic series as well as an asymptotic relation, defined in the first paragraph of this section. Confusion should be minimal, though, since asymptotic relations will always be written with an explicit statement of the limit of the argument.

The notation $f(z) \approx g(z)$ will be used to indicate that $f(z)$ and $g(z)$ are in some vague sense close together. It is used in this chapter only in cases where a precise statement would be cumbersome and would not help in explaining the essence of the argument.

All logarithms will be natural ones to base e unless specified otherwise, so that $\log 8 = 2.0794\ldots$, $\log_2 8 = 3$. The symbol $\lfloor x \rfloor$ denotes the greatest integer $\leq x$. The notation $x \to 1^-$ means that x tends to 1, but only from the left, and similarly, $x \to 0^+$ means that x tends to 0 only from the right, through positive values.

3. Identities, indefinite summations, and related approaches

Asymptotic estimates are useful, but often they can be avoided by using other methods. For example, the asymptotic methods presented later yield estimates for $\binom{n}{k}2^k$ as k and n vary, which can be used to estimate accurately the sum of $\binom{n}{k}2^k$ for n fixed and k running over the full range from 0 to n. That is a general and effective process, but somewhat cumbersome. On the other hand, by the binomial theorem,

$$\sum_{k=0}^{n} \binom{n}{k} 2^k = (1+2)^n = 3^n \ . \tag{3.1}$$

This is much more satisfactory and simpler to derive than what could be obtained from applying asymptotic methods to estimate individual terms in the sum. However, such identities are seldom available. There is nothing similar that can be applied to

$$\sum_{k \le n/5} \binom{n}{k} 2^k \ , \tag{3.2}$$

and we are forced to use asymptotic methods to estimate this sum.

Recognizing when some combinatorial identity might apply is not easy. The literature on this subject is huge, and some of the references for it are [172, 174, 186, 216, 336]. Many of the books listed in the references are useful for this purpose. Generating functions (see Section 6) are one of the most common and powerful tools for proving identities. Here we only mention two recent developments that are of significance for both theoretical and practical reasons. One is Gosper's algorithm for indefinite hypergeometric summation [171, 175]. Given a sequence $a_1, a_2, \ldots$, Gosper's algorithm determines whether the sequence of partial sums

$$b_n = \sum_{k=1}^{n} a_k \ , \quad n = 1, 2, \ldots \tag{3.3}$$

has the property that b_n/b_{n-1} is a rational function of n, and if it is, it gives an explicit form for b_n. We note that if b_n/b_{n-1} is a rational function of n, then so is

$$\frac{a_n}{a_{n-1}} = \frac{b_n/b_{n-1} - 1}{1 - b_{n-2}/b_{n-1}} \ . \tag{3.4}$$

Therefore Gosper's algorithm should be applied only when a_n/a_{n-1} is rational.

The other recent development is the Wilf-Zeilberger method for proving combinatorial identities [379, 380]. Given a conjectured identity, it provides an algorithmic procedure for verifying it. This method succeeds in a surprisingly wide range of cases. Typically, to prove an identity of the form

$$\sum_{k} U(n,k) = S(n) \ , \quad n \ge 0 \ , \tag{3.5}$$

where $S(n) \neq 0$, Wilf and Zeilberger define $F(n,k) = U(n,k)/S(n)$ and search for a rational function $R(n,k)$ such that if $G(n,k) = R(n,k)F(n,k-1)$, then

$$F(n+1,k) - F(n,k) = G(n,k+1) - G(n,k) \tag{3.6}$$

holds for all integers n, k with $n \geq 0$, and such that

1) for each integer k, the limit

$$f_k = \lim_{n\to\infty} F(n,k) \tag{3.7}$$

   exists and is finite.

2) for each integer $n \geq 0$, $\lim_{k\to\pm\infty} G(n,k) = 0$.

3) $\lim_{k\to-\infty} \sum_{n=0}^{\infty} G(n,k) = 0$.

If all these conditions are satisfied, and Eq. (3.5) holds for $n = 0$, then it holds for all $n \geq 0$.

Example 3.1. *Dixon's binomial sum identity.* This identity states that

$$\sum_k (-1)^k \binom{n+b}{n+k}\binom{b+c}{b+k}\binom{n+c}{c+k} = \frac{(n+b+c)!}{n!\,b!\,c!}\ . \tag{3.8}$$

This can be proved by the Wilf-Zeilberger method by taking

$$R(n,k) = \frac{(b+1-k)(c+1-k)}{2(n+k)(n+b+c+1)} \tag{3.9}$$

and verifying that the conditions above hold. ■

The Wilf-Zeilberger method requires finding a rational function $R(n,k)$ that satisfies the properties listed above. This is often hard to do, especially by hand. Gosper's algorithm leads to a systematic procedure for constructing such $R(n,k)$.

To conclude this section, we mention that a useful resource when investigating sequences arising in combinatorial settings is the book of Sloane [345, 346], which lists several thousand sequences and gives references for them. Section 17 mentions some software systems that are useful in asymptotics.

4. Basic estimates: factorials and binomial coefficients

No functions in combinatorial enumeration are as ubiquitous and important as the factorials and the binomial coefficients. In this section we state some estimates for these quantities, which will be used throughout this chapter and are of widespread applicability. Several different proofs of some of these estimates will be sketched later.

The basic estimate, from which many others follow, is that for the factorial. As was mentioned in the introduction, the basic form of Stirling's formula is

$$n! \sim (2\pi n)^{1/2} n^n e^{-n} \quad \text{as} \quad n \to \infty \ . \tag{4.1}$$

This is sufficient for many enumeration problems. However, when necessary one can draw on much more accurate estimates. For example Eq. 6.1.38 in [297] gives

$$n! = (2\pi n)^{1/2} n^n \exp(-n + \theta/(12n)) \tag{4.2}$$

for all $n \geq 1$, where $\theta = \theta(n)$ satisfies $0 < \theta < 1$. More generally, there is Stirling's asymptotic expansion:

$$\log\{n!(2\pi n)^{-1/2} n^{-n} e^n\} \sim \frac{1}{12n} - \frac{1}{360n^3} + \cdots \ . \tag{4.3}$$

(This is an asymptotic series in the sense of Eq. (2.2), and there is no convergent expansion for $\log\{n!(2\pi n)^{-1/2} n^{-n} e^n\}$ as a power series in n^{-1}.) Further terms in the expansion (4.3) can be obtained, and they involve Bernoulli numbers. In most references, such as Eq. 6.1.37 or 6.1.40 of [297], Stirling's formula is presented for $\Gamma(x)$, where Γ is Euler's gamma function. Expansions for $\Gamma(x)$ translate readily into ones for $n!$ because $n! = \Gamma(n+1)$.

Stirling's approximation yields the expansion

$$\binom{2n}{n} = \frac{4^n}{(\pi n)^{1/2}} \left\{ 1 - \frac{1}{8n} + \frac{1}{128n^2} + \frac{5}{1024n^3} + O(n^{-4}) \right\} \ . \tag{4.4}$$

A less precise but still useful estimate is

$$\binom{n}{\lfloor n/2 \rfloor} \sim \left(\frac{2}{\pi n} \right)^{1/2} 2^n \quad \text{as} \quad n \to \infty \ . \tag{4.5}$$

This estimate is used frequently. The binomial coefficients are *symmetric*, so that $\binom{n}{k} = \binom{n}{n-k}$ and *unimodal*, so that for a fixed n and k varying, the $\binom{n}{k}$ increase monotonically up to a peak at $k = \lfloor n/2 \rfloor$ (which is unique for n even and has two equal high points at $k = (n \pm 1)/2$ for n odd) and then decrease.

More important than Eq. (4.5) are expansions for general binomial coefficients. Eq. (4.2) shows that for $1 \le k \le n-1$,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \left\{\frac{n}{2\pi k(n-k)}\right\}^{1/2} \frac{n^n}{k^k(n-k)^{n-k}} \exp\left(O\left(\frac{1}{k}+\frac{1}{n-k}\right)\right)$$
$$= \left\{\frac{n}{2\pi k(n-k)}\right\}^{1/2} \exp\left(nH\left(\frac{k}{n}\right) + O\left(\frac{1}{k}+\frac{1}{n-k}\right)\right) , \quad (4.6)$$

where

$$H(x) = -x\log x - (1-x)\log(1-x) \quad (4.7)$$

is the entropy function. (We set $H(0) = H(1) = 0$ to make $H(x)$ continuous for $0 \le x \le 1$.) Simplifying further, we obtain

$$\binom{n}{k} = \exp(nH(k/n) + O(\log n)) , \quad (4.8)$$

an estimate that is valid for all $0 \le k \le n$. In many situations it suffices to use the weaker but simpler bound

$$\binom{n}{k} \le \left(\frac{ne}{k}\right)^k , \quad 0 \le k \le n . \quad (4.9)$$

Approximations of this form are used frequently in information theory and other fields.

A general estimate that can be derived by totally elementary methods, without recourse to Stirling's formula, is

$$\binom{n}{k}\binom{n}{\lfloor n/2\rfloor}^{-1} = \exp(-2(k-n/2)^2/n + O(|k-n/2|^3/n^2)) , \quad (4.10)$$

valid for $|k-n/2| \le n/4$, say. It is most useful for $|k-n/2| = o(n^{2/3})$, since the error term is small then. Similarly,

$$\binom{n}{k+r} \sim \binom{n}{k}\left(\frac{n-k}{k}\right)^r \quad \text{as} \quad n \to \infty , \quad (4.11)$$

uniformly in k provided r (which may be negative) satisfies $r^2 = o(k)$ and $r^2 = o(n-k)$. Further, we have

$$(n+k)! \sim n^k \exp(k^2/(2n))n! \quad \text{as} \quad n \to \infty , \quad (4.12)$$

again uniformly in k provided $k = o(n^{2/3})$.

5. Estimates of sums and other basic techniques

When encountering a combinatorial sum, the first reaction should always be to check whether it can be simplified by use of some identity. If no identity for the sum is found, the next step should be to try to transform the problem to eliminate the sum. Usually we are interested not in single isolated sums, but parametrized families of them, such as

$$b_n = \sum_k a_n(k) \ , \tag{5.1}$$

and it is the asymptotic behavior of the b_n as $n \to \infty$ that is desired. A standard and well-known technique (named the "snake-oil" method by Wilf [377]) for handling such cases is to form a generating function $f(z)$ for the b_n, use the properties of the $a_n(k)$ to obtain a simple form for $f(z)$, and then obtain the asymptotics of the b_n from the properties of $f(z)$. This method will be presented briefly in Section 6. In this section we discuss what to do if those two approaches fail. Sometimes the methods to be discussed can also be used in a preliminary phase to obtain a rough estimate for the sum. This estimate can then be used to decide which identities might be true, or what generating functions to form.

There are general methods for dealing with sums (cf. [234]), many of which are used in asymptotic enumeration. A basic technique of this type is summation by parts. Often sums to be evaluated can be expressed as

$$\sum_{j=1}^{n} a_j b_j \quad \text{or} \quad \sum_{j=1}^{\infty} a_j b_j \ ,$$

where the b_j, say, are known explicitly or behave smoothly, while the a_j by themselves might not be known well, but the asymptotics of

$$A(k) = \sum_{j=1}^{k} a_j \tag{5.2}$$

are known. Summation by parts relies on the identity

$$\sum_{j=1}^{n} a_j b_j = \sum_{k=1}^{n-1} A(k)(b_k - b_{k+1}) + A(n) b_n \ . \tag{5.3}$$

Example 5.1. *Sum of primes.* Let

$$S_n = \sum_{p \le n} p \ , \tag{5.4}$$

where p runs over the primes $\le n$. The Prime Number Theorem [23] states that the function

$$\pi(x) = \sum_{p \le x} 1 \tag{5.5}$$

satisfies

$$\pi(x) \sim \frac{x}{\log x} \quad \text{as} \quad x \to \infty \ . \tag{5.6}$$

(More precise estimates are available, but we will not use them.) We rewrite

$$S_n = \sum_{j=1}^{n} a_j b_j \ , \tag{5.7}$$

where

$$a_j = \begin{cases} 1 & j \text{ is prime} \ , \\ \\ 0 & \text{otherwise} \ , \end{cases} \tag{5.8}$$

and $b_j = j$ for all j. Then $A(k) = \pi(k)$ and summation by parts yields

$$S_n = \sum_{k=1}^{n-1} -\pi(k) + \pi(n)n \ . \tag{5.9}$$

Since

$$\sum_{k=1}^{n-1} \pi(k) \sim \sum_{k=2}^{n-1} \frac{k}{\log k} \sim \frac{n^2}{2\log n} \quad \text{as} \quad n \to \infty \ , \tag{5.10}$$

we have

$$S_n \sim \frac{n^2}{2\log n} \quad \text{as} \quad n \to \infty \ . \tag{5.11}$$

■

Summation by parts is used most commonly in situations like those of Example 5.1, to obtain an estimate for one sum from that of another.

Summation by parts is often easiest to carry out, both conceptually and notationally, by using integrals. If we let

$$A(x) = \sum_{k \le x} a_k \ , \tag{5.12}$$

then $A(x) = A(n)$ for $n \le x < n+1$. Suppose that $b_k = b(k)$ for some continuously differentiable function $b(x)$. Then

$$b_k - b_{k+1} = -\int_k^{k+1} b'(x)dx \ , \tag{5.13}$$

and we can rewrite Eq. (5.3) as

$$\sum_{j=1}^{n} a_j b_j = A(n)b(n) - \int_1^n A(x)b'(x)dx \ . \tag{5.14}$$

(One can apply similar formulas even when the b_j are not smooth, but this usually requires Riemann-Stieltjes integrals, cf. [14].) The approximation of sums by integrals that appears in (5.14) is common, and will be treated at length later.

5.1. Sums of positive terms

Sums of positive terms are extremely common. They can usually be handled with only a few basic tools. We devote substantial space to this topic because it is important and because the simplicity of the methods helps in illustrating some of the basic principles of asymptotic estimation, such as approximation by integrals, neglecting unimportant terms, and uniform convergence. For readers not familiar with asymptotic methods, working through the examples of this section is a good exercise that will make it easier to learn other techniques later.

Typical sums are of the form

$$b_n = \sum_k a_n(k) \ , \qquad a_n(k) \geq 0 \ , \tag{5.15}$$

where k runs over some range of summation, often $0 \leq k \leq n$ or $0 \leq k < \infty$, and the $a_n(k)$ may be given either explicitly or only through an asymptotic approximation. What is desired is the asymptotic behavior of b_n as $n \to \infty$. Usually the $a_n(k)$ for n fixed are unimodal, so that either i) $a_n(k) \leq a_n(k+1)$ for all k in the range, or ii) $a_n(k) \geq a_n(k+1)$ for all k, or iii) $a_n(k) \leq a_n(k+1)$ for $k \leq k_0$, and $a_n(k) \geq a_n(k+1)$ for $k > k_0$. The single most important task in estimating b_n is usually to find the maximal $a_n(k)$. This can be done either by combinatorial means (involving knowledge of where the $a_n(k)$ come from), by asymptotic estimation of the $a_n(k)$, or (most common when the $a_n(k)$ are expressed in terms of factorials or binomial coefficients) by finding where the ratio $a_n(k+1)/a_n(k)$ is close to 1. If $a_n(k+1)/a_n(k) < 1$ for all k, then we are in case ii) above, and if $a_n(k+1)/a_n(k) > 1$ for all k, we are in case i). If there is a k_0 in the range of summation such that $a_n(k_0+1)$ is close to $a_n(k_0)$, then we are almost certainly in case iii) and the peak occurs at some k close to k_0. The different cases are illustrated in the examples presented later in this section.

Once $\max a_n(k) = a_n(k_0)$ has been found, the next task is to show that most of the terms in the sum are insignificant. For example, if the sum in Eq. (5.15) is over $0 \leq k \leq n$, and if $a_n(0) = 1$ is the largest term, then

$$\sum_{\substack{k=0 \\ a_n(k) < n^{-2}}}^{n} a_n(k) < n^{-1} \ ,$$

which is negligible if we are only after a rough approximation to b_n, say of the form $b_n \sim c_n$ as $n \to \infty$, or even $b_n = c_n(1 + O(n^{-1}))$ as $n \to \infty$. Once the small terms have been discarded, we are usually left with a short range of summation. It can happen that this range is extremely short, and the maximal term $a_n(k_0)$ is much larger than any of its neighbors to the extent that $b_n \sim a_n(k_0)$ as $n \to \infty$. More commonly, the number of terms that contribute significantly to b_n does grow as $n \to \infty$, but slowly. Their contribution, relative to that of the maximal term $a_n(k_0)$, can usually be estimated by some simple function of $k - k_0$, and the sum of all of them approximated by an explicit integral. This method is sometimes referred to as Laplace's method for sums (in analogy to Laplace's method for estimating integrals, mentioned in Section 5.5, which proceeds in a similar spirit). There is extensive discussion of this method in [63].

Example 5.2. *Sums of the partition function.* We estimate

$$U_n = \sum_{k=1}^{n} p(k)^k \ , \tag{5.16}$$

where $p(k)$ is the number of partitions of k. Since any partition of $m-1$, say one with c_j parts of size j, can be transformed into a partition of m with $c_1 + 1$ parts of size 1, and c_j of size j for $j \geq 2$, we have $p(m) \geq p(m-1)$ for all $m \geq 2$. Therefore the largest term in the sum in (5.16) is the one with $k = n$. If the only estimate for $p(k)$ that we have is the one given by (1.6), then

$$p(n)^n = \exp(Cn^{3/2} - n \log(4 \cdot 3^{1/2} n) + O(n^{1/2})) \ . \tag{5.17}$$

Since the constant implied by the O-symbol is not specified, this estimate is potentially larger than $p(n)^n$ by a factor of $\exp(cn^{1/2})$, so we can only obtain asymptotics of $\log p(n)^n$, not of $p(n)^n$ itself. This also means that rough estimates of U_n follow easily from (5.17). Since $p(k)^k \leq p(n)^n$ for all $k < n$, and there are n terms in the sum, we have $p(n)^n \leq U_n \leq np(n)^n$, and because of the large error term in (5.17), we obtain

$$U_n = \exp(Cn^{3/2} - n \log(4 \cdot 3^{1/2} n) + O(n^{1/2})) \ . \tag{5.18}$$

Thus the use of the poor estimate (1.6) for $p(n)$ means that we can obtain only a crude estimate for U_n, and there is no need for careful analysis.

Instead of (1.6) we can use the more refined estimate (1.5). Let q_n denote first term on the right side of (1.5). Then we have

$$p(n) = q_n + O(n^{-1} \exp(Cn^{1/2}/2)) = q_n(1 + O(\exp(-Cn^{1/2}/2))) \ , \tag{5.19}$$

so

$$p(n)^n = q_n^n(1 + O(n \exp(-Cn^{1/2}/2))) = q_n^n(1 + O(\exp(-Cn^{1/2}/3))) \ , \tag{5.20}$$

say. Also, for some $\epsilon > 0$ we find from Eq. (1.5) (or Eq. 1.6) that for large n

$$q_{n-1} < q_n - \epsilon n^{-1/2} q_n \ .$$

Thus for large n,

$$\begin{aligned} q_{n-1}^{n-1} &< q_n^{n-1}(1-\epsilon n^{-1/2})^{n-1} \\ &< q_n^n \exp(-\epsilon n^{1/2}/2) \ , \end{aligned}$$

and therefore

$$\sum_{k=1}^{n-1} p(k)^k \leq (n-1)p(n-1)^{n-1} < q_n^n \exp(-\epsilon n^{1/2}/3) \ .$$

Thus we obtain

$$U_n = q_n^n (1 + O(\exp(-\delta n^{1/2}))) \tag{5.21}$$

for some $\delta > 0$.

The estimates of U_n presented above relied on the observation that the last term in the sum (5.16) defining U_n is much larger than the sum of all the other terms. This does not happen often. A more typical example is presented by

$$T_n = \sum_{k=1}^{n} p(k) \ . \tag{5.22}$$

As was noted before, $p(n)$ is larger than any of the other terms, but not by enough to dominate the sum. We therefore try the other approaches that were listed at the beginning of this section. We use only the estimate (1.6). Since $(1-x)^{1/2} < 1 - x/2$ for $0 \leq x \leq 1$, we find that for large n,

$$\begin{aligned} \sum_{k<n-n^{2/3}} p(k) &\leq np(n - \lceil n^{2/3} \rceil) \\ &\leq \exp(C(n - \lceil n^{2/3} \rceil)^{1/2}) \\ &\leq \exp(Cn^{1/2} - Cn^{1/6}/2) \\ &= O(p(n)\exp(-Cn^{1/6}/3)) \ . \end{aligned} \tag{5.23}$$

Thus most of the values of k contribute a negligible amount to the sum. For $k = n - j$, $0 \leq j \leq n^{2/3}$, we find that

$$p(n-j)/p(n) = (1 + O(n^{-1/3}))\exp(C(n-j)^{1/2} - Cn^{1/2}) \ .$$

Since

$$\begin{aligned} (n-j)^{1/2} &= n^{1/2} - jn^{-1/2}/2 + O(j^2 n^{-3/2}) \ , \\ p(n-j)/p(n) &= \exp(-Cjn^{-1/2}/2 + O(n^{-1/6})) \\ &= (1 + O(n^{-1/6}))\exp(-Cjn^{-1/2}/2) \ . \end{aligned} \tag{5.24}$$

Thus the ratios $p(n-j)/p(n)$ decrease geometrically, and so

$$p(n)^{-1} \sum_{0 \le j \le n^{2/3}} p(n-j) = \frac{(1+O(n^{-1/6}))}{1-\exp(-Cn^{-1/2}/2)} = 2C^{-1}n^{1/2}(1+O(n^{-1/6})) \ . \tag{5.25}$$

Therefore, combining all the estimates,

$$T_n = \sum_{k=1}^{n} p(k) = \frac{1+O(n^{-1/6})}{2 \cdot C \cdot 3^{1/2} \cdot n^{1/2}} \ e^{Cn^{1/2}} \ . \tag{5.26}$$

The $O(n^{-1/6})$ error term above can easily be improved with a little more care to $O(n^{-1/2})$, even if we continue to rely only on (1.6). ■

Before presenting further examples, we discuss some of the problems that can arise even in the simple setting of estimating positive sums. We then introduce the basic technique of approximating sums by integrals.

The lack of uniform convergence is a frequent cause of incorrect estimates. If $a_n(k) \sim c_n(k)$ for each k as $n \to \infty$, it does not necessarily follow that

$$b_n = \sum_k a_n(k) \sim \sum_k c_n(k) \quad \text{as} \quad n \to \infty \ . \tag{5.27}$$

A simple counterexample is given by $a_n(k) = \binom{n}{k}$ and $c_n(k) = \binom{n}{k}(1+k/n)$. To conclude that (5.27) holds, it is usually necessary to know that $a_n(k) \sim c_n(k)$ as $n \to \infty$ uniformly in k. Such uniform convergence does hold if we replace $c_n(k)$ in the counterexample above by $c'_n(k) = \binom{n}{k}(1+k/n^2)$, for example.

There is a general principle that sums of terms that vary smoothly with the index of summation should be replaced by integrals, so that for $\alpha > 0$, say,

$$\sum_{k=1}^{n} k^{\alpha} \sim \int_1^{n+1} u^{\alpha} du \quad \text{as} \quad n \to \infty \ . \tag{5.28}$$

The advantage of replacing a sum by an integral is that integrals are usually much easier to handle. Many more closed-form expressions are available for definite and indefinite integrals than for sums. We will discuss extensions of this principle of replacing sums by integrals further in Section 5.3, when we present the Euler-Maclaurin summation formula. Usually, though, we do not need anything sophisticated, and the application of the principle to situations like that of (5.28) is easy to justify. If $a_n = g(n)$ for some function $g(x)$ of a real argument x, then

$$\left| g(n) - \int_n^{n+1} g(u)du \right| \le \max_{n \le u \le n+1} |g(u) - g(n)| \ , \tag{5.29}$$

and so

$$\left|\sum_n g(n) - \int g(u)du\right| \le \sum_n \max_{n\le u\le n+1} |g(u)-g(n)| \ , \tag{5.30}$$

where the integral is over $[a, b+1]$ if the sum is over $a \le n \le b$, $a, b \in Z$. If $g(u)$ is continuously differentiable, then $|g(u) - g(n)| \le \max_{n\le v\le n+1} |g'(v)|$ for $n \le u \le n+1$. This gives the estimate

$$\left|\sum_{n=a}^{b} g(n) - \int_a^{b+1} g(u)du\right| \le \sum_{n=a}^{b} \max_{n\le v\le n+1} |g'(v)| \ . \tag{5.31}$$

Often one can find a simple explicit function $h(w)$ such that $|g'(v)| \le h(w)$ for any v and w with $|v - w| \le 1$, in which case Eq. (5.31) can be replaced by

$$\left|\sum_{n=a}^{b} g(n) - \int_a^{b+1} g(u)du\right| \le \int_a^{b+1} h(v)dv \ . \tag{5.32}$$

For good estimates to be obtained from integral approximations to sums, it is usually necessary for individual terms to be small compared to the sum.

Example 5.3. *Sum of* $\exp(-\alpha k^2)$. In the final stages of an asymptotic approximation one often encounters sums of the form

$$h(\alpha) = \sum_{k=-\infty}^{\infty} \exp(-\alpha k^2) \ , \qquad \alpha > 0 \ . \tag{5.33}$$

There is no closed form for the indefinite integral of $\exp(-\alpha u^2)$ (it is expressible in terms of the Gaussian error function only), but there is the famous evaluation of the definite integral

$$\int_{-\infty}^{\infty} \exp(-\alpha u^2)du = (\pi/\alpha)^{1/2} \ . \tag{5.34}$$

Thus it is natural to approximate $h(\alpha)$ by $(\pi/\alpha)^{1/2}$. If $g(u) = \exp(-\alpha u^2)$, then $g'(u) = -2\alpha u g(u)$, and so for $n \ge 0$,

$$\max_{n\le v\le n+1} |g'(v)| \le 2\alpha(n+1)g(n) \ . \tag{5.35}$$

For the integral in Eq. (5.30) to yield a good approximation to the sum we must show that the error term is smaller than the integral. The largest term in the sum occurs at $n = 0$ and equals 1. The error bound (5.35) that comes from approximating $g(0) = 1$ by the integral of $g(u)$ over $0 \le u \le 1$ is 2α. Therefore we cannot expect to obtain a good estimate unless $\alpha \to 0$. We find that

$$2\alpha(n+1)g(n) \le 4\alpha u g(u/2) \quad \text{for} \quad n \ge 1, \quad n \le u \le n+1 \ ,$$

so (integral approximation again!)

$$
\begin{aligned}
\sum_{n=1}^{\infty} 2\alpha(n+1)g(n) &\leq 4\alpha \int_1^{\infty} ug(u/2)du \\
&\leq 4\alpha \int_0^{\infty} ug(u/2)du = (8\alpha)^{1/2} .
\end{aligned}
\tag{5.36}
$$

Therefore, taking into account the error for $n = 0$ which was not included in the bound (5.36), we have

$$
\begin{aligned}
h(\alpha) &= \sum_{n=-\infty}^{\infty} \exp(-\alpha n^2) = \int_{-\infty}^{\infty} \exp(-\alpha u^2)du + O(\alpha^{1/2} + \alpha) \\
&= (\pi/\alpha)^{1/2} + O(\alpha^{1/2}) \quad \text{as} \quad \alpha \to 0^+ .
\end{aligned}
\tag{5.37}
$$

For this sum much more precise estimates are available, as will be shown in Example 5.9. For many purposes, though, (5.37) is sufficient. ■

Example 5.3 showed how to use the basic tool of approximating a sum by an integral. Moreover, the estimate (5.37) that it provides is ubiquitous in asymptotic enumeration, since many approximations reduce to it. This is illustrated by the following example.

Example 5.4. *Bell numbers* (cf. [63]). The Bell number, $B(n)$, counts the partitions of an n-element set. It is given by [81]

$$
B(n) = e^{-1} \sum_{k=1}^{\infty} \frac{k^n}{k!} . \tag{5.38}
$$

In this sum no single term dominates. The ratio of the $(k+1)$-st to the k-th term is

$$
\frac{(k+1)^n}{(k+1)!} \cdot \frac{k!}{k^n} = \frac{1}{k+1} \left(1 + \frac{1}{k}\right)^n . \tag{5.39}
$$

As k increases, this ratio strictly decreases. We search for the point where it is about 1. For $k \geq 2$,

$$
\left(1 + \frac{1}{k}\right)^n = \exp\left(n \log\left(1 + \frac{1}{k}\right)\right) = \exp(n/k + O(n/k^2)) , \tag{5.40}
$$

so the ratio is close to 1 for n/k close to $\log(k+1)$. We choose k_0 to be the closest integer to w, the solution to

$$
n = w \log(w + 1) . \tag{5.41}
$$

For $k = k_0 + j$, $1 \le j \le k_0/2$, we find, since $\log(1 + i/k_0) = i/k_0 - i^2/(2k_0^2) + O(i^3/k_0^3)$,

$$\begin{aligned} \frac{k^n}{k!} &= \frac{k_0^n}{k_0!} \frac{(1+j/k_0)^n}{k_0^j \Pi_{i=1}^j (1+i/k_0)} \\ &= \frac{k_0^n}{k_0!} \exp\left(jn/k_0 - j\log k_0 - j^2(n+k_0)/(2k_0^2) + O(nj^3/k_0^3 + j/k_0)\right) \ . \end{aligned} \tag{5.42}$$

The same estimate applies for $-k_0/2 \le j \le 0$. The term $jn/k_0 - j\log k_0$ is small, since $|k_0 - w| \le 1/2$ and w satisfies (5.41). We find

$$\begin{aligned} n/k_0 - \log k_0 &= n/w - \log(w+1) + O(n/w^2 + 1/w) \\ &= O(n/w^2 + 1/w) \ . \end{aligned} \tag{5.43}$$

By (5.41), $w \sim n/\log n$ as $n \to \infty$. We now further restrict j to $|j| \le n^{1/2}\log n$. Then (5.42) and (5.43) yield

$$\frac{k^n}{k!} = \frac{k_0^n}{k_0!} \exp(-j^2(n+k_0)/(2k_0^2) + O((\log n)^6 n^{-1/2})) \ . \tag{5.44}$$

Approximating the sum by an integral, as in Example 5.3, shows that

$$\sum_{\substack{k \\ |j| \le n^{1/2}\log n}} \frac{k^n}{k!} = \frac{k_0^n}{k_0!} k_0 (2\pi)^{1/2} (n+k_0)^{-1/2} (1 + O((\log n)^6 n^{-1/2})) \ . \tag{5.45}$$

(An easy way to obtain this is to apply the estimate of Example 5.3 to the sum from $-\infty$ to ∞, and show that the range $|j| > n^{1/2}\log n$ contributes little.) To estimate the contribution of the remaining summands, with $|j| > n^{1/2}\log n$, we observe that the ratio of successive terms is ≤ 1, so the range $1 \le k \le k_0 - \lfloor n^{1/2}\log n \rfloor$ contributes at most k_0 (the number of terms) times the largest term, which arises for $k = k_0 - \lfloor n^{1/2}\log n \rfloor$. By (5.44), this largest term is

$$O(k_0^n (k_0!)^{-1} \exp(-(\log n)^3)) \ .$$

For $k \ge k_1 \ge k_0 + \lfloor n^{1/2}\log n \rfloor$, we find that the ratio of the $(k+1)$-st to the k-th term is, for large n,

$$\begin{aligned} \le \frac{1}{k_1+1}\left(1 + \frac{1}{k_1}\right)^n &= \exp(n/k_1 - \log(k_1+1) - n/(2k_1^2) + O(n/k_1^3)) \\ &\le \exp(-(k_1 - k_0)n/k_1^2 + O(n/k_1^3)) \\ &\le \exp(-2n^{-1/2}) \ \le \ 1 - n^{-1/2} \ , \end{aligned} \tag{5.46}$$

and so the sum of these terms, for $k_1 \le k < \infty$, is bounded above by $n^{1/2}$ times the term for $k = k_1$. Therefore the estimate on the right-hand side of (5.45) applies even when we sum on all k, $1 \le k < \infty$.

To obtain an estimate for $B(n)$, it remains only to estimate $k_0^n/k_0!$. To do this, we apply Stirling's formula and use the property that $|k_0 - w| \le 1/2$ to deduce that

$$B(n) \sim (\log w)^{1/2} w^{n-w} e^w \quad \text{as} \quad n \to \infty \,, \tag{5.47}$$

where w is given by (5.41).

There is no explicit formula for w in terms of n, and substituting various asymptotic approximations to w, such as

$$w = \frac{n}{\log n} + O\left(\frac{n}{(\log n)^2}\right) \tag{5.48}$$

(see Example 5.10) yields large error terms in (5.47), so for accuracy it is usually better to use (5.47) as is. There are other approximations to $B(n)$ in the literature (see, for example, [33, 63]). They differ slightly from (5.47) because they estimate $B(n)$ in terms of roots of equations other than (5.41).

Other methods of estimating $B(n)$ are presented in Examples 12.5 and 12.6. ■

5.2. Alternating sums and the principle of inclusion-exclusion

At the beginning of Section 5, the reader was advised in general to search for identities and transformations when dealing with general sums. This advice is even more important when dealing with sums of terms that have alternating or irregularly changing coefficients. Finding the largest term is of little help when there is substantial cancellation among terms. Several general approaches for dealing with this difficulty will be presented later. Generating function methods for dealing with complicated sums are discussed in Section 6. Contour integration methods for alternating sums are mentioned in Section 10.3. The summation formulas of the next section can sometimes be used to estimate sums with regularly varying coefficients as well. In this section we present some basic elementary techniques that are often sufficient.

Sometimes it is possible to obtain estimates of sums with positive and negative summands by approximating separately the sums of the positive and of the negative summands. Methods of the preceding section or of the next section are useful in such situations. However, this approach is to be avoided as much as possible, because it often requires extremely precise estimates of the two sums to obtain even rough bounds on the desired sums. One method that often works and is much simpler consists of a simple pairing of adjacent positive and negative terms.

Example 5.5. *Alternating sum of square roots.* Let

$$S_n = \sum_{k=1}^{n} (-1)^k k^{1/2} \ . \tag{5.49}$$

We have

$$\begin{aligned}(2m)^{1/2} - (2m-1)^{1/2} &= (2m)^{1/2} \left\{ 1 - \left(1 - \frac{1}{2m} \right)^{1/2} \right\} \\ &= (2m)^{1/2} \left\{ 1 - \left(1 - \frac{1}{4m} + O(m^{-2}) \right) \right\} \\ &= (8m)^{-1/2} + O(m^{-3/2}) \ ,\end{aligned} \tag{5.50}$$

so

$$\begin{aligned}\sum_{k=1}^{2\lfloor n/2 \rfloor} (-1)^k k^{1/2} &= \sum_{m=1}^{\lfloor n/2 \rfloor} (8m)^{-1/2} + O(1) \\ &= n^{1/2}/2 + O(1) \ .\end{aligned} \tag{5.51}$$

Hence

$$S_n = \begin{cases} n^{1/2}/2 + O(1) & \text{if } n \text{ is even} \ , \\ -n^{1/2}/2 + O(1) & \text{if } n \text{ is odd} \ . \end{cases} \tag{5.52}$$

■

In Example 5.5, the sums of the positive terms and of the negative terms can easily be estimated accurately (for example, by using the Euler-Maclaurin formula of the next section) to obtain (5.52). In other cases, though, the cancellation is too extensive for such an approach to work. This is especially true for sums arising from the principle of inclusion-exclusion.

Suppose that X is some set of objects and P is a set of properties. For $R \subseteq P$, let $N_=(R)$ be the number of objects in X that have exactly the properties in R and none of the properties in $P \setminus R$. We let $N_\geq(R)$ denote the number of objects in X that have all the properties in R and possibly some of those in $P \setminus R$. The principle of inclusion-exclusion says that

$$N_=(R) = \sum_{R \subseteq Q \subseteq P} (-1)^{|Q \setminus R|} N_\geq(Q) \ . \tag{5.53}$$

(This is a basic version of the principle. For more general results, proofs, and references, see [81, 173, 351].)

Example 5.6. *Derangements of n letters.* Let X be the set of permutations of n letters, and suppose that P_i, $1 \leq i \leq n$, is the property that the i-th letter is fixed by a permutation, and $P = \{P_1, \ldots, P_n\}$. Then d_n, the number of derangements of n letters, equals $N_=(\phi)$, where ϕ is the empty set, and so by (5.53)

$$d_n = \sum_{Q \subseteq P} (-1)^{|Q|} N_{\geq}(Q) \ . \tag{5.54}$$

However, $N_{\geq}(Q)$ is just the number of permutations that leave all letters specified by Q fixed, and thus

$$\begin{aligned} d_n &= \sum_{Q \subseteq P} (-1)^{|Q|} (n - |Q|)! \\ &= \sum_{k=0}^{n} (-1)^k (n-k)! \binom{n}{k} = \sum_{k=0}^{n} (-1)^k \frac{n!}{k!} \ , \end{aligned} \tag{5.55}$$

which is Eq. (1.1). ■

The formula (1.1) for derangements is easy to use because the terms decrease rapidly. Moreover, this formula is exceptionally simple, largely because $N_{\geq}(Q)$ depends only on $|Q|$. In general, the inclusion-exclusion principle produces complicated sums that are hard to estimate. A frequently helpful tool is provided by the *Bonferroni inequalities* [81, 351]. One form of these inequalities is that for any integer $m \geq 0$,

$$N_=(R) \geq \sum_{\substack{Q \\ R \subseteq Q \subseteq P \\ |Q \setminus R| \leq 2m}} (-1)^{|Q \setminus R|} N_{\geq}(Q) \tag{5.56}$$

and

$$N_=(R) \leq \sum_{\substack{Q \\ R \subseteq Q \subseteq P \\ |Q \setminus R| \leq 2m+1}} (-1)^{|Q \setminus R|} N_{\geq}(Q) \ . \tag{5.57}$$

Thus in general

$$\left| N_=(R) - \sum_{\substack{Q \\ R \subseteq Q \subseteq P \\ |Q \setminus R| \leq k}} (-1)^{|Q \setminus R|} N_{\geq}(Q) \right| \leq \sum_{\substack{Q \\ R \subseteq Q \subseteq P \\ |Q \setminus R| \leq k+1}} N_{\geq}(Q) \ . \tag{5.58}$$

These inequalities are frequently applied for $n = |X|$ increasing. Typically one chooses k that increases much more slowly than n, so that the individual terms $N_{\geq}(Q)$ in (5.58) can be estimated asymptotically, as the interactions of the different properties counted by $N_{\geq}(Q)$ is not too complicated to estimate. Bender [33] presents some useful general principles to be used in such estimates (especially the asymptotically Poisson distribution that tends to occur when the method is successful). We present an adaptation of an example from [33].

Example 5.7. *Balls and cells.* Given n labeled cells and m labeled balls, let $a_h(m,n)$ be the number of ways to place the balls into cells so that exactly h of the cells are empty. We consider h fixed. Let X be the ways of placing the balls into the cells (n^m in total), and $P = \{P_1, \ldots, P_n\}$, where P_i is the property that the i-th cell is empty. If $R = \{P_1, \ldots, P_h\}$, then $a_h(m,n) = \binom{n}{h} N_=(R)$. Now

$$N_\geq(Q) = (n - |Q|)^m \ , \tag{5.59}$$

so

$$\begin{aligned} \sum_{\substack{Q \\ R\subseteq Q\subseteq P \\ |Q\setminus R|=t}} N_\geq(Q) &= \binom{n-h}{t}(n-h-t)^m \\ &= n^m e^{-mh/n}(ne^{-m/n})^t (t!)^{-1}(1 + O((t^2+1)mn^{-2} + (t^2+1)n^{-1})) \ , \end{aligned} \tag{5.60}$$

provided $t^2 \leq n$ and $mt^2n^{-2} \leq 1$, say. In the range $0 \leq t \leq \log n$, $n \log n \leq m \leq n^2(\log n)^{-3}$, we find that the right-hand side of (5.60) is

$$n^m e^{-mh/n}(ne^{-m/n})^t (t!)^{-1}(1 + O(mn^{-2}(\log n)^2)) \ .$$

We now apply (5.58) with $k = \lfloor \log n \rfloor$, and obtain

$$\begin{aligned} a_h(m,n) &= \binom{n}{h} N_=(R) \sim \binom{n}{h} n^m \exp(-mh/n - ne^{-m/n}) \\ &\sim n^m (h!)^{-1}(ne^{-m/n})^h \exp(-ne^{-m/n}) \end{aligned} \tag{5.61}$$

as $m, n \to \infty$, provided $n \log n \leq m \leq n^2(\log n)^{-3}$. Since $a_h(m,n)n^{-m}$ is the probability that there are exactly h empty cells, the relation (5.61) (which we have established only for fixed h) shows that this probability is asymptotically distributed like a Poisson random variable with parameter $n \exp(-m/n)$.

Many additional results on random distributions of balls into cells, and references to the extensive literature on this subject can be found in [241]. ■

Bonferroni inequalities include other methods for estimating $N_=(R)$ by linear combinations of the $N_\geq(Q)$. Recent approaches and references (phrased in probabilistic terms) can be found in [152]. For bivariate Bonferroni inequalities (where one asks for the probability that at least one of two sets of events occurs) see [153, 249].

The Chen-Stein method [75] is a powerful technique that is often used in place of the principle of inclusion-exclusion, especially in probabilistic literature. Recent references are [17, 27].

5.3. Euler-Maclaurin and Poisson summation formulas

Section 5.0 showed that sums can be successfully approximated by integrals if the summands are all small compared to the total sum and vary smoothly as functions of the summation index. The approximation (5.29), though crude, is useful in a wide variety of cases. Sometimes, though, more accurate approximations are needed. An obvious way is to improve the bound (5.29). If $g(x)$ is really smooth, we can expect that the difference

$$a_n - \int_n^{n+1} g(u)du$$

will vary in a regular way with n. This is indeed the case, and it is exploited by the Euler-Maclaurin summation formula. It can be found in many books, such as [63, 175, 297, 298]. There are many formulations, but they do not differ much.

Euler-Maclaurin summation formula. Suppose that $g(x)$ has $2m$ continuous derivatives in $[a, b]$, $a, b \in Z$. Then

$$\begin{aligned} \sum_{k=a}^{b} g(k) &= \int_a^b g(x)dx + \sum_{r=1}^{m} \frac{B_{2r}}{(2r)!} \left\{ g^{(2r-1)}(b) - g^{(2r-1)}(a) \right\} \\ & \quad + \frac{1}{2}\{g(a) + g(b)\} + R_m \ , \end{aligned} \tag{5.62}$$

where

$$R_m = -\int_a^b g^{(2m)}(x) \frac{B_{2m}(x - \lfloor x \rfloor)}{(2m)!} dx \ , \tag{5.63}$$

and so

$$|R_m| \le \int_a^b |g^{(2m)}(x)| \frac{|B_{2m}(x - \lfloor x \rfloor)|}{(2m)!} dx \ . \tag{5.64}$$

In the above formulas, the $B_n(x)$ denote the Bernoulli polynomials, defined by

$$\frac{ze^{xz}}{e^z - 1} = \sum_{n=0}^{\infty} B_n(x) \frac{z^n}{n!} \ . \tag{5.65}$$

The B_n are the Bernoulli numbers, defined by

$$\frac{z}{e^z - 1} = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} \ , \tag{5.66}$$

so that $B_n = B_n(0)$, and

$$\begin{aligned} B_0 &= 1 \ , \quad B_1 = -1/2 \ , \quad B_2 = 1/6 \ , \\ B_3 &= B_5 = B_7 = \cdots = 0 \ , \\ B_4 &= -1/30 \ , \quad B_6 = 1/42 \ , \quad B_8 = -1/30, \ \ldots \ . \end{aligned} \tag{5.67}$$

It is known that

$$|B_{2m}(x - \lfloor x \rfloor)| \leq |B_{2m}| \ , \tag{5.68}$$

so we can simplify (5.64) to

$$|R_m| \leq |B_{2m}|((2m)!)^{-1} \int_a^b |g^{(2m)}(x)| dx \ . \tag{5.69}$$

There are many applications of the Euler-Maclaurin formula. One of the most frequently cited ones is to estimate factorials.

Example 5.8. *Stirling's formula.* We transform the product in the definition of $n!$ into a sum by taking logarithms, and find that for $g(x) = \log x$ and $m = 1$ we have

$$\log n! = \sum_{k=1}^{n} \log k = \int_1^n (\log x) dx + \frac{1}{2} \log n + \frac{1}{2} B_2 \left\{ \frac{1}{n} - 1 \right\} + R_1 \ , \tag{5.70}$$

where

$$R_1 = \int_1^n \frac{B_2(x - \lfloor x \rfloor)}{2x^2} dx = C + O(n^{-1}) \tag{5.71}$$

for

$$C = \int_1^\infty \frac{B_2(x - \lfloor x \rfloor)}{2x^2} dx \ . \tag{5.72}$$

Therefore

$$\log n! = n \log n - n + \frac{1}{2} \log n + C + 13/12 + O(n^{-1}) \ , \tag{5.73}$$

which gives

$$n! \sim C' n^{1/2} n^n e^{-n} \quad \text{as} \quad n \to \infty \ . \tag{5.74}$$

To obtain Stirling's formula (4.1), we need to show that $C' = (2\pi)^{1/2}$. This can be done in several ways (cf. [63]). In Examples 12.1, 12.4, and 12.5 we will see other methods of deriving (4.1). ■

There is no requirement that the function $g(x)$ in the Euler-Maclaurin formula be positive. That was not even needed for the crude approximation of a sum by an integral given in Section 5.0. The function $g(x)$ can even take complex values. (After all, Eq. (5.62) is an identity!) However, in most applications this formula is used to derive an asymptotic estimate with a small error term. For that, some high order derivatives have to be small, which means that $g(x)$ cannot change sign too rapidly. In particular, the Euler-Maclaurin formula usually is not very useful when the $g(k)$ alternate in sign. In those cases one can sometimes use the differencing trick (cf. Example 5.5) and apply the Euler-Maclaurin formula to $h(k) = g(2k) + g(2k+1)$. There is also Boole's summation formula for alternating sums that can be applied. (See Chapter 2, §3 and Chapter 6, §6 of [298], for example.) Generalizations to other periodic patterns in the coefficients have been derived by Berndt and Schoenfeld [47].

The bounds for the error term R_m in the Euler-Maclaurin formula that were stated above can often be improved by using special properties of the function $g(x)$. For example, when $g(x)$ is analytic in x, there are contour integrals for R_m that sometimes give good estimates (cf. [315]).

The Poisson summation formula states that

$$\sum_{n=-\infty}^{\infty} f(n+a) = \sum_{m=-\infty}^{\infty} \exp(2\pi i m a) \int_{-\infty}^{\infty} f(y) \exp(-2\pi i m y) dy \tag{5.75}$$

for "nice" functions $f(x)$. The functions for which (5.75) holds include all continuous $f(x)$ for which $\int |f(x)|dx < \infty$, which are of bounded variation, and for which $\sum_n f(n+a)$ converges for all a. For weaker conditions that ensure validity of (5.75), we refer to [63, 365]. The Poisson summation formula often converts a slowly convergent sum into a rapidly convergent one. Generally it is not as widely applicable as the Euler-Maclaurin formula as it requires extreme regularity for the Fourier coefficients to decrease rapidly. On the other hand, it can be applied in some situations that are not covered by the Euler-Maclaurin formula, including some where the coefficients vary in sign.

Example 5.9. *Sum of* $\exp(-\alpha k^2)$. We consider again the function $h(\alpha)$ of Example 5.3. We let $f(x) = \exp(-\alpha x^2)$, $a = 0$. Eq. (5.15) then gives

$$h(\alpha) = \sum_{n=-\infty}^{\infty} \exp(-\alpha n^2) = (\pi/\alpha)^{1/2} \sum_{m=-\infty}^{\infty} \exp(-\pi^2 m^2/\alpha) \ . \tag{5.76}$$

This is an identity, and the sum on the right-hand side above converges rapidly for small α. Many applications require the evaluation of the sum on the left in which α tends to 0. Eq. (5.76) offers a method of converting a slowly convergent sum into a tractable one, whose asymptotic behavior is explicit. ■

5.4. Bootstrapping and other basic methods

Bootstrapping is a useful technique that uses asymptotic information to obtain improved estimates. Usually we start with some rough bounds, and by combining them with the relations defining the function or sequence that we are studying, we obtain better bounds.

Example 5.10. *Approximation of Bell numbers.* Example 5.4 obtained the asymptotics of the Bell numbers B_n, but only in terms of w, the solution to Eq. (5.41). We now show how to obtain asymptotic expansions for w. As n increases, so does w. Therefore $\log(w+1)$ also increases, and so $w < n$ for large n. Thus

$$n = w \log(w+1) < w \log(n+1) \ ,$$

and so

$$n(\log(n+1))^{-1} < w < n \ . \tag{5.77}$$

Therefore

$$\log(w+1) = \log n + O(\log\log n) \ , \tag{5.78}$$

and so

$$w = \frac{n}{\log(w+1)} = \frac{n}{\log n} + O\left(\frac{n\log\log n}{(\log n)^2}\right) \ . \tag{5.79}$$

To go further, note that by (5.79),

$$\begin{aligned} \log(w+1) &= \log\left(\frac{n}{\log n}\left(1 + O\left(\frac{\log\log n}{\log n}\right)\right)\right) \\ &= \log n - \log\log n + O((\log\log n)(\log n)^{-1}) \ , \end{aligned} \tag{5.80}$$

and so by applying this estimate in Eq. (5.41), we obtain

$$w = \frac{n}{\log n} + \frac{n\log\log n}{(\log n)^2} + \frac{n(\log\log n)^2}{(\log n)^3} + O\left(\frac{n\log\log n}{(\log n)^3}\right) \ . \tag{5.81}$$

This procedure can be iterated indefinitely to obtain expansions for w with error terms $O(n(\log n)^{-\alpha})$ for as large a value of α as desired. ■

In the above example, w can also be estimated by other methods, such as the Lagrange-Bürmann inversion formula (cf. Example 6.7). However, the bootstrapping method is much more widely applicable and easy to apply. It will be used several times later in this chapter.

5.5. Estimation of integrals

In some of the examples in the preceding sections integrals were used to approximate sums. The integrals themselves were always easy to evaluate. That is true in most asymptotic enumeration problems, but there do occur situations where the integrals are more complicated. Often the hard integrals are of the form

$$f(x) = \int_{\alpha}^{\beta} g(t)\exp(xh(t))dt \ , \tag{5.82}$$

and it is necessary to estimate the behavior of $f(x)$ as $x \to \infty$, with the functions $g(t)$, $h(t)$ and the limits of integration α and β held fixed. There is a substantial theory of such integrals, and good references are [54, 63, 100, 315]. The basic technique is usually referred to as Laplace's method, and consists of approximating the integrand by simpler functions near its maxima. This approach is similar to the one that is discussed at length in Section 5.1 for estimating sums. The contributions of the approximations are then evaluated, and it is shown that the remaining ranges of integration, away from the maxima, contribute a negligible amount. By breaking up the interval of integration we can write the integral (5.82) as a sum of several integrals of the same type, with the property that there is a unique maximum of the integrand and that it occurs at one of the endpoints. When $\alpha > 0$, the maximum of the integrand occurs for large x at the maximum of $h(t)$ (except in rare cases where $g(t) = 0$ for that t for which $h(t)$ is maximized). Suppose that the maximum occurs at $t = \alpha > 0$. It often happens that

$$h(t) = h(\alpha) - c(t-\alpha)^2 + O(|t-\alpha|^3) \tag{5.83}$$

for $\alpha \le t \le \beta$ and $c = -h''(\alpha)/2 > 0$, and then one obtains the approximation

$$f(x) \sim g(\alpha)\exp(xh(\alpha))[-\pi/(4xh''(\alpha))]^{1/2} \quad \text{as} \quad x \to \infty \;, \tag{5.84}$$

provided $g(\alpha) \neq 0$. For precise statements of even more general and rigorous results, see for example Chapter 3, §7 of [315]. Those results cover functions $h(t)$ that behave near $t = \alpha$ like $h(\alpha) - c(t-\alpha)^\mu$ for any $\mu > 0$.

When the integral is highly oscillatory, as happens when $h(t) = iu(t)$ for a real-valued function $u(t)$, still other techniques (such as the stationary phase method), are used. We will not present them here, and refer to [54, 63, 100, 315] for descriptions and applications. In Section 12.1 we will discuss the saddle point method, which is related to both Laplace's method and the stationary phase method.

Laplace integrals

$$F(x) = \int_0^\infty f(t)\exp(-xt)dt \tag{5.85}$$

can often be approximated by integration by parts. We have (under suitable conditions on $f(t)$)

$$\begin{aligned} F(x) &= x^{-1}f(0) + x^{-1}\int_0^\infty f'(t)\exp(-xt)dt \\ &= x^{-1}f(0) + x^{-2}f'(0) + x^{-2}\int_0^\infty f''(t)\exp(-xt)dt \;, \end{aligned} \tag{5.86}$$

and so on. There are general results, usually associated with the name of Watson's Lemma, for deriving such expansions. For references, see [100, 315].

6. Generating functions

6.1. A brief overview

Generating functions are a wonderfully powerful and versatile tool, and most asymptotic estimates are derived from them. The most common ones in combinatorial enumeration are the ordinary and exponential generating functions. If $a_0, a_1, \ldots,$ is any sequence of real or complex numbers, the *ordinary generating function* is

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \ , \tag{6.1}$$

while the *exponential generating function* is

$$f(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{n!} \ . \tag{6.2}$$

Doubly-indexed arrays, for example $a_{n,k}$, $0 \le n < \infty$, $0 \le k \le n$, are encoded as two-variable generating functions. Depending on the array, sometimes one uses

$$f(x,y) = \sum_{n=0}^{\infty} \sum_{k=0}^{n} a_{n,k} x^k y^n \ , \tag{6.3}$$

and sometimes other forms that might even mix ordinary and exponential types, as in

$$f(x,y) = \sum_{n=0}^{\infty} \frac{y^n}{n!} \sum_{k=0}^{n} a_{n,k} x^k \ . \tag{6.4}$$

For example, the Stirling numbers of the first kind, $s(n,k)$ ($(-1)^{n+k} s(n,k)$ is the number of permutations on n letters with k cycles) have the generating function (see pp. 50, 212–213, and 234–235 in [81])

$$1 + \sum_{n=1}^{\infty} \frac{y^n}{n!} \sum_{k=1}^{n} s(n,k) x^k = (1+y)^x \ . \tag{6.5}$$

In general, a generating function is just a formal power series, and questions of convergence do not arise in the definition. However, some of the main applications of generating functions in asymptotic enumeration do rely on analyticity or other convergence properties of those functions, and there the domain of convergence is important.

A generating function is just another form for the sequence that defines it. There are many reasons for using it. One is that even for complicated sequences, generating functions are frequently simple. This might not be obvious for the partition function $p(n)$, which has the ordinary generating function

$$f(z) = \sum_{n=0}^{\infty} p(n) z^n = \prod_{k=1}^{\infty} (1 - z^k)^{-1} \ . \tag{6.6}$$

The sequence $p(n)$, which is complicated, is encoded here as an infinite product. The terms in the product are simple and vary in a regular way with the index, but it is not clear at first what is gained by this representation. In other cases, though, the advantages of generating functions are clearer. For example, the exponential generating function for derangements (Eq. (1.1) and Example 5.6) is

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} \frac{d_n}{n!} z^n = \sum_{n=0}^{\infty} \frac{z^n}{n!} \sum_{n=0}^{n} (-1)^k \frac{n!}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \sum_{n=k}^{\infty} z^n = \frac{e^{-z}}{1-z} \ , \end{aligned} \tag{6.7}$$

which is extremely compact.

Reasons for using generating functions go far beyond simplicity. The one that matters most for this chapter in that generating functions can be used to obtain information about the asymptotic behavior of sequences they encode, information that often cannot be obtained in any other way, or not as easily. Methods such as those of Section 10.2 can be used to obtain immediately from Eq. (6.7) the asymptotic estimate $d_n \sim e^{-1} n!$ as $n \to \infty$. This estimate can also be derived easily by elementary methods from Eq. (1.1), so here the generating function is not essential. In other cases, though, such as that of the partition function $p(n)$, all the sharp estimates, such as that of Hardy and Ramanujan given in (1.5), are derived by exploiting the properties of the generating function. If there is any main theme to this chapter, it is that generating functions are usually the easiest, most versatile, and most powerful way to study asymptotic behavior of sequences. Especially when the generating function is analytic, its behavior at the dominant singularities (a term that will be defined in Section 10) determines the asymptotics of the sequence. When the generating function is simple, and often even when it is not simple, the contribution of the dominant singularity can often be determined easily, although the sequence itself is complicated.

There are many applications of generating functions, some related to asymptotic questions. Averages can often be studied using generating functions. Suppose, for example, that $a_{n,k}$, $0 \le k \le n$, $0 \le n < \infty$, is the number of objects in some class of size n, which have weight k

(for some definition of size and weight), and that we know, either explicitly or implicitly, the generating function $f(x, y)$ of $a_{n,k}$ given by (6.4). Then

$$g(y) = f(1, y) = \sum_{n=0}^{\infty} \frac{y^n}{n!} \sum_{k=0}^{n} a_{n,k} \tag{6.8}$$

is the exponential generating function of the number of objects of size n, while

$$h(y) = \frac{\partial}{\partial x} f(x, y)\Big|_{x=1} = \sum_{n=0}^{\infty} \frac{y^n}{n!} \sum_{k=0}^{n} k a_{n.k} \tag{6.9}$$

is the exponential generating function of the sum of the weights of objects of size n. Therefore the average weight of an object of size n is

$$\frac{[y^n]h(y)}{[y^n]g(y)} \ . \tag{6.10}$$

The wide applicability and power of generating functions come primarily from the structured way in which most enumeration problems arise. Usually the class of objects to be counted is derived from simpler objects through basic composition rules. When the generating functions are chosen to reflect appropriately the classes of objects and composition rules, the final generating function is derivable in a simple way from those of the basic objects. Suppose, for example, that each object of size n in class C can be decomposed uniquely into a pair of objects of sizes k and $n - k$ (for some k) from classes A and B, and each pair corresponds to an object in C. Then c_n, the number of objects of size n in C, is given by the convolution

$$c_n = \sum_{k=0}^{n} a_k b_{n-k} \ , \tag{6.11}$$

(where a_k is the number of objects of size k in A, etc.). Hence if $A(z) = \sum a_n z^n$, $B(z) = \sum b_n z^n$, $C(z) = \sum c_n z^n$ are the ordinary generating functions, then

$$C(z) = A(z)B(z) \ . \tag{6.12}$$

Thus ordered pairing of objects corresponds to multiplication of ordinary generating functions. If $A(z) = \sum a_n z^n$ and

$$b_n = \sum_{k=0}^{n} a_k \ ,$$

then $B(z) = \sum b_n z^n$ is given by

$$B(z) = \frac{A(z)}{1 - z} \ , \tag{6.13}$$

so that the ordinary generating function of cumulative sums of coefficients is obtained by dividing by $1 - z$. There are many more such general correspondences between operations on combinatorial objects and on the corresponding generating functions. They are present, implicitly or explicitly, in most books that cover combinatorial enumeration, such as [81, 173, 351, 377]. The most systematic approach to developing and using general rules of this type has been carried out by Flajolet and his collaborators [139]. They develop ways to see immediately (cf. [134]) that if we consider mappings of a set of n labeled elements to itself, so that all n^n distinct mappings are considered equally likely, then the generating function for the longest path length is given by

$$f(z) = \sum_{k=0}^{\infty} \left(\frac{1}{1 - t(z)} - e^{v_k(z)} \right) , \tag{6.14}$$

where

$$v_k(z) = t_{k-1}(z) + \frac{1}{2} t_{k-2}(z)^2 + \cdots + \frac{1}{k} t_0(k)^k , \tag{6.15}$$

with

$$t_0(z) = z , \quad t_{h+1}(z) = z \exp(t_h(z)) , \tag{6.16}$$

and $t(z) = \lim_{h \to \infty} t_h(z)$ (in the sense of formal power series, so convergence is that of coefficients). Furthermore, as is mentioned in Section 17, many of these rules for composition of objects and generating functions can be implemented algorithmically, automating some of the chores of applying them.

We illustrate some of the basic generating function techniques by deriving the generating function for rooted labeled trees, which will occur later in Examples 6.6 and 10.8. (The rooted unlabeled trees, with generating function given by (1.8), are harder.)

Example 6.1. *Rooted labeled trees.* Let t_n be the number of rooted labeled trees on n vertices, so that $t_1 = 1$, $t_2 = 2$, $t_3 = 9$. (It will be shown in Example 6.6 that $t_n = n^{n-1}$.) Let

$$t(z) = \sum_{n=1}^{\infty} t_n \frac{z^n}{n!} \tag{6.17}$$

be the exponential generating function. If we remove the root of a rooted labeled tree with n vertices, we are left with $k \geq 0$ rooted labeled trees that contain a total of $n - 1$ vertices. The total number of ways of arranging an ordered selection of k rooted trees with a total of $n - 1$ vertices is

$$[z^{n-1}] t(z)^k .$$

Since the order of the trees does not matter, we have

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

different trees of size n that have exactly k subtrees, and so

$$\begin{aligned} t_n &= \sum_{k=0}^{\infty} \frac{1}{k!}[z^{n-1}]t(z)^k \\ &= [z^{n-1}]\sum_{k=0}^{\infty} t(z)^k/k! \;=\; [z^n]z\exp(t(z)) \;, \end{aligned} \tag{6.18}$$

which gives

$$t(z) = z\exp(t(z)) \;. \tag{6.19}$$

As an aside, the function $t_h(z)$ of Eq (6.16) is the exponential generating function of rooted labeled trees of height $\leq h$. ■

The key to the successful use of generating functions is to use a generating function that is of the appropriate form for the problem at hand. There is no simple rule that describes what generating function to use, and sometimes two are used simultaneously. In combinatorics and analysis of algorithms, the most useful forms are the ordinary and exponential generating functions, which reflects how the classes of objects that are studied are constructed. Sometimes other forms are used, such as the double exponential form

$$f(z) = \sum_{n=0}^{\infty} \frac{a_n z^n}{(n!)^2} \tag{6.20}$$

that occurs in Section 7, or the Newton series

$$f(z) = \sum_{n=0}^{\infty} a_n z(z-1)\cdots(z-n+1) \;. \tag{6.21}$$

Also frequently encountered are various q-analog generating functions, such as the Eulerian

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n z^n}{(1-q)(1-q^2)\cdots(1-q^n)} \;. \tag{6.22}$$

In multiplicative number theory, the most common are Dirichlet series

$$f(z) = \sum_{n=1}^{\infty} a_n n^{-z} \;, \tag{6.23}$$

which reflect the multiplicative structure of the integers. If a_n is a multiplicative function (so that $a_{mn} = a_m a_n$ for all relatively prime positive integers m and n) then the function (6.23) has an Euler product representation

$$f(z) = \prod_p (1 + a_p p^{-z} + a_{p^2} p^{-2z} + \cdots) \ , \tag{6.24}$$

where p runs over the primes. This allows new tools to be used to study $f(z)$ and through it a_n. Additive problems in combinatories and number theory often are handled using functions such as functions such as

$$f(z) = \sum_{n=1}^{\infty} z^{a_k} \ , \tag{6.25}$$

where $0 \leq a_1 < a_2 < \cdots$ is a sequence of integers. Addition of two such sequences then corresponds to a multiplication of the generating functions of the form (6.25).

We next mention the "snake oil method." This is the name given by Wilf [377] to the use of generating functions for proving identities, and comes from the surprising power of this technique. The typical application is to evaluation of sequences given by sums of the type

$$a_n = \sum_k b_{n,k} \ . \tag{6.26}$$

The standard procedure is to form a generating function of the a_n and manipulate it through interchanges of summation and other tricks to obtain the final answer. The generating function can be ordinary, exponential, or (less commonly) of another type, depending on what gives the best results. We show a simple application of this principle that exhibits the main features of the method.

Example 6.2. *A binomial coefficient sum* [377]. Let

$$a_n = \sum_{k=0}^{n} \binom{n+k}{2k} 2^{n-k} \ , \quad n \geq 0 \ . \tag{6.27}$$

We define $A(z)$ to be the ordinary generating function of a_n. We find that

$$\begin{aligned} A(z) &= \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} \binom{n+k}{2k} 2^{n-k} \\ &= \sum_{k=0}^{\infty} 2^{-k} \sum_{n=k}^{\infty} 2^n z^n \binom{n+k}{2k} = \sum_{k=0}^{\infty} 2^{-k} (2z)^{-k} \sum_{n=0}^{\infty} \binom{n+k}{2k} (2z)^{n+k} \\ &= \sum_{k=0}^{\infty} 2^{-k} (2z)^{-k} \frac{(2z)^{2k}}{(1-2z)^{2k+1}} = \frac{1}{1-2z} \sum_{k=0}^{\infty} \left(\frac{z}{1-2z} \right)^k \\ &= \frac{1-2z}{(1-4z)(1-z)} = \frac{2}{3(1-4z)} + \frac{1}{3(1-z)} \ . \end{aligned} \tag{6.28}$$

Therefore we immediately find the explicit form

$$a_n = (2^{2n+1} + 1)/3 \quad \text{for} \quad n \geq 0 \; . \tag{6.29}$$

■

We next present some additional examples of how generating functions are derived. We start by considering linear recurrences with constant coefficients.

The first step in solving a linear recurrence is to obtain its generating function. Suppose that a sequence $a_0, a_1, a_2, \ldots$ satisfies the recurrence

$$a_n = \sum_{i=1}^{d} c_i a_{n-i}, \quad n \geq d \; . \tag{6.30}$$

Then

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} a_n z^n = \sum_{n=0}^{d-1} a_n z^n + \sum_{n=d}^{\infty} z^n \sum_{i=1}^{d} c_i a_{n-i} \\ &= \sum_{n=0}^{d-1} a_n z^n + \sum_{i=1}^{d} c_i z^i \sum_{n=d}^{\infty} a_{n-i} z^{n-i} \\ &= \sum_{n=0}^{d-1} a_n z^n + \sum_{i=1}^{d} c_i z^i \left(f(z) - \sum_{n=0}^{d-i-1} a_n z^n \right) \; , \end{aligned} \tag{6.31}$$

and so

$$f(z) = \frac{g(z)}{1 - \sum_{i=1}^{d} c_i z^i} \; , \tag{6.32}$$

where

$$g(z) = \sum_{n=0}^{d-1} a_n z^n - \sum_{i=1}^{d} c_i z^i \sum_{n=0}^{d-i-1} a_n z^n \tag{6.33}$$

is a polynomial of degree $\leq d-1$. Eq. (6.32) is the fundamental relation in the study of linear recurrences, and $1 - \sum c_i z^i$ is called the *characteristic polynomial* of the recursion.

Example 6.3. *Fibonacci numbers.* We let $F_0 = 0$, $F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$, and

$$F(z) = \sum_{n=0}^{\infty} F_n z^n \; .$$

Then by (6.32) and (6.33),

$$F(z) = \frac{z}{1 - z - z^2} \; . \qquad ■ \tag{6.34}$$

Often there is no obvious recurrence for the sequence a_n being studied, but there is one involving some other auxiliary function. Usually if one can obtain at least as many recurrences as there are sequences, one can obtain their generating functions by methods similar to those used for a single sequence. The main additional complexity comes from the need to solve a system of linear equations with polynomial coefficients. We illustrate this with the following example.

Example 6.4. *Sequences with forbidden subwords.* Let $A = a_1 a_2 \cdots a_k$ be a binary string of length k. Define $f_A(n)$ to be the number of binary strings of length n that do not contain A as a subword of k adjacent characters. (Subsequences do not count, so that if $A = 1110$, then A is contained in 1101110010, but not in 101101.) We introduce the correlation polynomial $C_A(z)$ of A:

$$C_A(z) = \sum_{j=0}^{k-1} c_A(j) z^j \ , \tag{6.35}$$

where $c_A(0) = 1$ and for $1 \leq j \leq k-1$,

$$c_A(j) = \begin{cases} 1 & \text{if} \ \ a_1 a_2 \cdots a_{k-j} = a_{j+1} a_{j+2} \cdots a_k \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{6.36}$$

As examples, we note that if $A = 1000$, then $C_A(z) = 1$, whereas $C_A(z) = 1 + z + z^2 + z^3$ if $A = 1111$. The generating function

$$F_A(z) = \sum_{n=0}^{\infty} f_A(n) z^n \tag{6.37}$$

then satisfies

$$F_A(z) = \frac{C_A(z)}{z^k + (1-2z) C_A(z)} \ . \tag{6.38}$$

To prove this, define $g_A(n)$ to be the number of binary sequences $b_1 b_2 \cdots b_n$ of length n such that $b_1 b_2 \cdots b_k = A$, but such that $b_j b_{j+1} \cdots b_{j+k-1} \neq A$ for any j with $2 \leq j \leq n-k+1$; i.e., sequences that start with A but do not contain it any place else. We then have $g_A(n) = 0$ for $n < k$, and $g_A(k) = 1$. We also define

$$G_A(z) = \sum_{n=0}^{\infty} g_A(n) z^n \ . \tag{6.39}$$

We next obtain a relation between $G_A(z)$ and $F_A(z)$ that will enable us to determine both.

If $b_1 b_2 \cdots b_n$ is counted by $f_A(n)$, then for x either 0 or 1, the string $x b_1 b_2 \cdots b_n$ either does not contain A at all, or if it does contain it, then $A = x b_1 b_2 \cdots b_{k-1}$. Therefore for $n \geq 0$,

$$2 f_A(n) = f_A(n+1) + g_A(n+1) \tag{6.40}$$

and multiplying both sides of Eq. (6.40) by z^n and summing on $n \geq 0$ yields

$$2F_A(z) = z^{-1}(F_A(z) - 1) + z^{-1}G_A(z) \ . \tag{6.41}$$

We need one more relation, and to obtain it we consider any string $B = b_1b_2 \cdots b_n$ that does not contain A any place inside. If we let C be the concatenation of A and B, so that $C = a_1a_2 \cdots a_kb_1b_2 \cdots b_n$, then C starts with A, and may contain other occurrences of A, but only at positions that overlap with the initial A. Therefore we obtain,

$$f_A(n) = \sum_{\substack{j=1 \\ c_A(k-j)=1}}^{k} g_A(n+j) \quad \text{for} \quad n \geq 0 \ , \tag{6.42}$$

and this gives the relation

$$F_A(z) = z^{-k}C_A(z)G_A(z) \ . \tag{6.43}$$

Solving the two equations (6.41) and (6.43), we find that $F_A(z)$ satisfies (6.38), while

$$G_A(z) = \frac{z^k}{z^k + (1-2z)C_A(z)} \ . \tag{6.44}$$

The proof above follows that in [182], except that [182] uses generating functions in z^{-1}, so the formulas look different. Applications of the formulas (6.38) and (6.44) will be found later in this chapter, as well as in [182, 130]. Other approaches to string enumeration problems are referenced there as well. Other approaches and applications of string enumerations are given in the references to [182] and in papers such as [18]. ■

The above example can be generalized to provide generating functions that enumerate sequences in which any of a given set of patterns are forbidden [182].

Whenever one has a finite system of linear recurrences with constant coefficients that involve several sequences, say $a_n^{(i)}$, $1 \leq i \leq k$, $n \geq 0$, one can translate these recurrences into linear equations with polynomial coefficients in the generating functions $A^{(i)}(z) = \sum a_n^{(i)} z^n$ for these sequences. To obtain the $A^{(i)}(z)$, one then needs to solve the resulting system. Such solutions will exist if the matrix of polynomial coefficients is nonsingular over the field of rational functions in z. In particular, one needs at least as many equations (i.e., recurrence relations) as k, the number of sequences, and if there are exactly as many equations as sequences, then the determinant of the matrix of the coefficients has to be a nonzero polynomial.

One interesting observation is that when a system of recurrences involving several sequences is solved by the above method, each of the generating functions $A^{(i)}(z)$ is a rational function in z. What this means is that each of the sequences $a_n^{(i)}$, $1 \leq i \leq k$, satisfies a linear recurrence with constant coefficients that does not involve any of the other $a_n^{(j)}$ sequences! In principle, therefore, that recurrence could have been found right at the beginning by combinatorial methods. However, usually the degree of the recurrence for an isolated $a_n^{(j)}$ sequence is high, typically about k times as large as the average degree of the k recurrences involving all the $a_n^{(j)}$. Thus the use of several sequences $a_n^{(j)}$ leads to much simpler and combinatorially more appealing relations.

That generating functions can significantly simplify combinatorial problems is shown by the following example. It is taken from [349], and is a modification of a result of Klarner [229] and Pólya [321]. This example also shows a more complicated derivation of explicit generating functions than the simple ones presented so far.

Example 6.5. *Polyomino enumeration* [349]. Let a_n be the number of n-square polyominoes P that are inequivalent under translation, but not necessarily under rotation or reflection, and such that each row of P is an unbroken line of squares. Then $a_1 = 1$, $a_2 = 2$, $a_3 = 6$. We define $a_0 = 0$. It is easily seen that

$$a_n = \sum (m_1 + m_2 - 1)(m_2 + m_3 - 1) \cdots (m_{s-1} + m_s - 1) \ , \tag{6.45}$$

where the sum is over all ordered partitions $m_1 + \cdots + m_s = n$ of n into positive integers m_i. Let $a_{r,n}$ be the sum of terms in (6.45) with $m_1 = r$, where we set $a_{n,n} = 1$, and $a_{r,n} = 0$ if $r > n$ or $n < 0$. Then

$$a_n = \sum_{r=1}^{\infty} a_{r,n} \ , \tag{6.46}$$

$$a_{r,n} = \sum_{i=1}^{\infty} (r + i - 1) a_{i,n-r} \ , \qquad r < n \ . \tag{6.47}$$

Define

$$A(x,y) = \sum_{n=1}^{\infty} \sum_{r=1}^{\infty} a_{r,n} x^r y^n \ , \tag{6.48}$$

so that

$$A(1,y) = \sum_{n=1}^{\infty} a_n y^n \tag{6.49}$$

is the generating function of the a_n, which are what we need to estimate.

By (6.47), we find that

$$A(x,y) = \sum_{n=1}^{\infty} x^n y^n + \sum_{n=1}^{\infty} \sum_{r=1}^{\infty} \sum_{i=1}^{\infty} (r + i - 1) a_i(n - r) x^r y^n$$

$$ \tag{6.50} $$

$$ = \quad \frac{xy}{1-xy} + \frac{x^2y^2}{(1-xy)^2}A(1,y) + \frac{xy}{1-xy}G(x,y) \ , \tag{6.51} $$

where

$$ G(y) = \sum_{n=1}^{\infty}\sum_{i=1}^{\infty} i a_{i,n} y^n = \frac{\partial}{\partial x}A(x.y)\bigg|_{x=1} \ , \tag{6.52} $$

We now set $x = 1$ in (6.50) and obtain an equation involving $A(1, y)$ and $G(y)$, namely

$$ A(1,y) = \frac{y}{1-y} + \frac{y^2}{(1-y)^2}A(1,y) + \frac{y}{1-y}G(y) \ . \tag{6.53} $$

We next differentiate (6.50) with respect to x, and set $x = 1$. This gives us a second equation,

$$ G(y) = \frac{y}{(1-y)^2} + \frac{2y^2}{(1-y)^3}A(1,y) + \frac{y}{(1-y)^2}G(y) \ . \tag{6.54} $$

We now eliminate $G(y)$ from (6.53) and (6.54) to obtain

$$ A(1,y) = \frac{y(1-y)^3}{1-5y+7y^2-4y^3} \ . \tag{6.55} $$

This formula shows that

$$ a_{n+3} = a_{n+2} - 7a_{n+1} + 4a_n \quad \text{for} \quad n \ge 2 \ . \tag{6.56} $$

Using the results of Section 10 we can easily obtain from (6.55) an asymptotic estimate

$$ a_n \sim c\alpha^n \quad \text{as} \quad n \to \infty \ , \tag{6.57} $$

where c is a certain constant and $\alpha = 3.205569\ldots$ is the inverse of the smallest zero of $1 - 5y + 7y^2 - 4y^3$. ■

For other methods and results related to polyomino enumeration, see [326, 327].

6.2. Composition and inversion of power series

So far we have only discussed simple operations on generating functions, such as multiplication. What happens when we do something more complicated? There are several frequently occurring operations on generating functions whose results can be described explicitly.

Faà di Bruno's formula [81]. Suppose that

$$ A(z) = \sum_{m=0}^{\infty} a_m \frac{z^m}{m!} \ , \quad B(z) = \sum_{n=0}^{\infty} b_n \frac{z^n}{n!} \ , \tag{6.58} $$

are two exponential generating functions with $b_0 = 0$. Then the formal composition $C(z) = A(B(z))$ is well-defined, and

$$C(z) = \sum_{n=0}^{\infty} c_n \frac{z^n}{n!} \tag{6.59}$$

with

$$c_0 = 0, \quad c_n = \sum_{k=1}^{n} a_k B_{n,k}(b_1, b_2, \ldots, b_{n-k+1}) \ , \tag{6.60}$$

where the $B_{n,k}$ are the exponential Bell polynomials defined by

$$\sum_{n,k=0}^{\infty} B_{n,k}(x_1, \ldots, x_{n-k+1}) \frac{t^n u^k}{n!} = \exp\left(u \sum_{m=1}^{\infty} x_m \frac{t^m}{m!} \right) \ , \tag{6.61}$$

with the x_j independent variables.

Faà di Bruno's formula makes it possible to compute successive derivatives of functions such as $\log A(z)$ in terms of the derivatives of $A(z)$. For further examples, see [81, 335, 336]. Faà di Bruno's formula is derivable in a straightforward way from the multinomial theorem.

Composition of generating functions occurs frequently in combinatorics and analysis of algorithms. When it yields the desired generating function as a composition of several known generating functions, the basic problem is solved, and one can work on the asymptotics of the coefficients using Faà di Bruno's formula or other methods. A more frequent event is that the composition yields a functional equation for the generating function, as in Example 6.1, where the exponential generating function $t(z)$ for labeled rooted trees was shown to satisfy $t(z) = z\exp(t(z))$. General functional equations are hard to deal with. (Many examples will be presented later.) However, there is a class of them for which an old technique, the Lagrange-Bürmann inversion formula, works well. We start by noting that if

$$f(z) = \sum_{n=0}^{\infty} f_n z^n \tag{6.62}$$

is a formal power series with $f_0 = 0$, $f_1 \neq 0$, then there is an inverse formal power series $f^{\langle -1 \rangle}(z)$ such that

$$f(f^{\langle -1 \rangle}(z)) = f^{\langle -1 \rangle}(f(z)) = z \ . \tag{6.63}$$

The coefficients of $f^{\langle -1 \rangle}(z)$ can be expressed explicitly in terms of the coefficients of $f(z)$. More generally, we have the following result.

Lagrange-Bürmann inversion formula. Suppose that $f(z)$ is a formal power series with $[z^0]f(z) = 0$, $[z^1]f(z) \neq 0$, and that $g(z)$ is any formal power series. Then for $n \geq 1$,

$$[z^n]\{g(f^{\langle -1 \rangle})(z)\} = n^{-1}[z^{n-1}]\{g'(z)(f(z)/z)^{-n}\} \ . \tag{6.64}$$

In particular, for $g(z) = z$, we have

$$[z^n]f^{\langle -1\rangle}(z) = n^{-1}[z^{n-1}](f(z)/z)^{-n} \ . \tag{6.65}$$

Example 6.6. *Rooted labeled trees.* As was shown in Example 6.1, the exponential generating function of rooted labeled trees satisfies $t(z) = z\exp(t(z))$. If we rewrite it as $z = t(z)\exp(-t(z))$, we see that $t(z) = f^{\langle -1\rangle}(z)$, where $f(z) = z\exp(-z)$. Therefore Eq. (6.65) yields

$$\begin{aligned} [z^n]t(z) &= n^{-1}[z^{n-1}]\exp(-nz) \\ &= n^{-1}n^{n-1}/(n-1)! \ = \ n^{n-1}/n! \ , \end{aligned} \tag{6.66}$$

which shows that t_n, the number of rooted labeled trees on n nodes, is n^{n-1}. ■

Proof of a form of the Lagrange-Bürmann theorem is given in Chapter ?. Extensive discussion, proofs, and references are contained in [81, 173, 205, 375]. Some additional recent references are [159, 208]. There exist generalizations of the Lagrange-Bürmann formula to several variables [173, 169, 208].

The Lagrange-Bürmann formula, as stated above, is valid for general formal power series. If $f(z)$ is analytic in a neighborhood of the origin, then so are $f^{\langle -1\rangle}(z)$ and $g(f^{\langle -1\rangle})(z)$, provided $g(z)$ is also analytic near 0 and $f'(0) \neq 0$, $f(0) = 0$. Most of the presentations of this inversion formula in the literature assume analyticity. However, that is not a real restriction. To prove (6.65), say, in full generality, it suffices to prove it for any n. Given n, if we let

$$F(z) = \sum_{k=0}^{n} f_k z^k \ , \quad G(z) = \sum_{k=0}^{n} g_k z^k \ ,$$

then we see that

$$[z^n]\{g(f^{\langle -1\rangle})(z)\} = [z^n]G(F^{\langle -1\rangle})(z) \ , \tag{6.67}$$

and $F(z)$ and $G(z)$ are analytic, so the formula (6.65) can be applied. Thus combinatorial proofs of the Lagrange-Bürmann formula do not offer greater generality than analytic ones.

While the analytic vs. combinatorial distinction in the proofs of the Lagrange-Bürmann formula does not matter, it is possible to use analyticity of the functions $f(z)$ and $g(z)$ to obtain useful information. Example 6.6 above was atypical in that a simple explicit formula was derived. Often the quantity on the right-hand side of (6.64) is not explicit enough to make clear its asymptotic behavior. When that happens, and $g(z)$ and $f(z)$ are analytic, one can use the contour integral representation

$$[z^{n-1}]\{g'(z)(f(z)/z)^{-n}\} = \frac{1}{2\pi i}\int_{\Gamma} g'(z)f(z)^{-n}dz \ , \tag{6.68}$$

where Γ is a positively oriented simple closed contour enclosing the origin that lies inside the region of analyticity of both $g(z)$ and $f(z)$. This representation, which is discussed in Section 10, can often be used to obtain asymptotic information about coefficients $[z^n]g(f^{\langle -1\rangle})(z)$ (cf. [273]).

The Lagrange-Bürmann formula can provide numerical approximations to roots of equations and even convergent infinite series representations for such roots. An important case is the trinomial equation $y = z(1 + y^r)$, and there are many others.

Example 6.7. *Dominant zero for forbidden subword generating functions.* The generating functions $F_A(z)$ and $G_A(z)$ of Example 6.4 both have denominators

$$h(z) = z^k + (1-2z)C(z) \ , \tag{6.69}$$

where $C(z)$ is a polynomial of degree $\leq k$, with coefficients 0 and 1, and with $C(0) = 1$. It will be shown later that $h(z)$ has only one zero ρ of small absolute value, and that this zero is the dominant influence on the asymptotic behavior of the coefficients of $F_A(z)$ and $G_A(z)$. Right now we obtain accurate estimates for ρ.

For simplicity, we will consider only large k. Since $C(z)$ has nonnegative coefficients and $C(0) = 1$, $h(3/4) \leq (3/4)^k - 1/2 < 0$ for $k \geq 3$. On the other hand, $h(1/2) = 2^{-k}$. Therefore $h(z)$ has a real zero ρ with $1/2 < \rho < 3/4$. As $k \to \infty$, $\rho \to 1/2$, since

$$\rho^k = (2\rho - 1)C(\rho) \ , \tag{6.70}$$

and $\rho^k \to 0$ as $k \to \infty$ for $1/2 < \rho < 3/4$, while $2\rho - 1$ and $C(\rho)$ are bounded. We can deduce from (6.69) that

$$2\rho - 1 \sim 2^{-k}C(1/2)^{-1} \quad \text{as} \quad k \to \infty \ , \tag{6.71}$$

uniformly for all polynomials $C(z)$ of the prescribed type. By applying the bootstrapping technique (see Section 5.4) we can find even better approximations. By (6.71),

$$C(\rho) = C(1/2) + O(|\rho - 1/2|) = C(1/2) + O(2^{-k}) \ , \tag{6.72}$$

$$\rho^k = 2^{-k}(1 + O(2^{-k}))^k = 2^{-k}(1 + O(k2^{-k})) \ , \tag{6.73}$$

so (6.70) now yields

$$\rho = 1/2 + 2^{-k-1}C(1/2)^{-1} + O(k2^{-2k}) \ . \tag{6.74}$$

Even better approximations can be obtained by repeating the process using (6.74). At the next stage we would apply the expansion

$$\begin{aligned} C(\rho) &= C(1/2) + (\rho - 1/2)C'(1/2) + O((\rho - 1/2)^2) \\ &= C(1/2) + 2^{-k-1}C'(1/2) + O(k2^{-2k}) \end{aligned} \tag{6.75}$$

and a similar one for ρ^k.

A more systematic way to obtain a rapidly convergent series for ρ is to use the inversion formula. If we set $u = \rho - 1/2$, then (6.70) can be rewritten as $w(u) = 1$, where

$$w(u) = 2uC(1/2 + u)(1/2 + u)^{-k} = \sum_{j=1}^{\infty} a_j u^j \ , \tag{6.76}$$

with

$$a_1 = 2^{k+1}C(1/2) \neq 0 \ . \tag{6.77}$$

Hence $u = w^{\langle -1 \rangle}(1)$, and the Lagrange-Bürmann inversion formula (6.65) yields the coefficients of $w^{\langle -1 \rangle}(z)$. In particular, we find that

$$\rho = 1/2 + u \approx 1/2 + 2^{-k-1}C(1/2)^{-1} + k2^{-2k-1}C(1/2)^{-2} - 2^{-2k-2}C'(1/2)C(1/2)^{-3} + \cdots \tag{6.78}$$

as a Poincaré asymptotic series. With additional work one can show that the series (6.78) converges, and that

$$\begin{aligned} \rho = 1/2 &+ 2^{-k-1}C(1/2)^{-1} + k2^{-2k-1}C(1/2)^{-2} \\ &- 2^{-2k-2}C'(1/2)C(1/2)^{-3} + O(k^2 2^{-3k}) \ , \end{aligned} \tag{6.79}$$

for example. The same estimate can be obtained by the bootstrapping technique. ■

6.3. Differentiably finite power series

Homogeneous recurrences with constant coefficients are the nicest large set of sequences one can imagine, with rational generating functions, and well-understood asymptotic behavior. The next class in complexity consists of the polynomially-recursive or, *P-recursive sequences*, $a_0, a_1, \ldots$, which satisfy recurrences of the form

$$p_d(n)a_{n+d} + p_{d-1}(n)a_{n+d-1} + \cdots + p_0(n)a_n = 0, \quad n \geq 0 \ , \tag{6.80}$$

where d is fixed and $p_0(n), \ldots, p_d(n)$ are polynomials in n. Such sequences are common in combinatorics, with $a_n = n!$ a simple example. Normally P-recursive sequences do not have explicit forms for their generating functions. In this section we briefly summarize some of their main properties. Asymptotic properties of P-recursive sequences will be discussed in Section 9.2. The main references for the results quoted here are [254, 350].

A formal power series

$$f(z) = \sum_{k=0}^{\infty} a_k z^k \tag{6.81}$$

is called differentiably finite, or D-*finite*, if the derivatives $f^{(n)}(z) = \frac{d^n f(z)}{dz^n}$, $n \geq 0$, span a finite-dimensional vector space over the field of rational functions with complex coefficients. The following three conditions are equivalent for a formal power series $f(z)$:

i) $f(z)$ is D-finite.

ii) There exist finitely many polynomials $q_0(z), \ldots, q_k(z)$ and a polynomial $q(z)$, not all 0, such that

$$q_k(z) f^{(k)}(z) + \cdots + q_0(z) f(z) = q(z) \ . \tag{6.82}$$

iii) There exist finitely many polynomials $p_0(z), \ldots, p_m(z)$, not all 0, such that

$$p_m(z) f^{(m)}(z) + \cdots + p_0(z) f(z) = 0 \ . \tag{6.83}$$

The most important result for combinatorial enumeration is that a sequence $a_0, a_1, \ldots,$ is P-recursive if and only if its ordinary generating function $f(z)$, defined by (6.81), is D-finite. This makes it possible to apply results that are more easily proved for D-finite power series.

If $f(z)$ is D-finite, then so is the power series obtained by changing a finite number of the coefficients of $f(z)$. If $f(z)$ is algebraic (i.e., there exist polynomials $q_0(z), \ldots, q_d(z)$, not all 0, such that $q_d(z) f(z)^d + \cdots + q_0(z) f(z) + q_0(z) = 0$), then $f(z)$ is D-finite. The product of two D-finite power series is also D-finite, as is any linear combination with polynomial coefficients. Finally, the Hadamard product of two D-finite series is D-finite. The proofs rely on elementary linear algebra constructions. An important feature of the theory is that identity between D-finite series is decidable.

The concept of a D-finite power series can be extended to several variables [254, 405], and there are generalizations of P-recursiveness [254, 405]. (See also [161].) Zeilberger [405] has used the word *holonomic* to describe corresponding sequences and generating functions.

When we investigate a sequence $\{a_n\}$, sometimes the combinatorial context yields only relations for more complicated object with several indices. While we might like to obtain the generating function $f(z) = \sum a_n z^n$, we might instead find a formula for a generating function

$$F(z_1, z_2, \ldots, z_k) = \sum_{n_1,\ldots,n_k} b_{n_1}, \ldots, n_k z_1^{n_1}, \ldots, z_k^{n_k} \ , \tag{6.84}$$

where $a_n = b_{n,n,\ldots,n}$, say. When this happens, we say that $f(z)$ is a *diagonal* of $F(z_1, \ldots, z_k)$. (There are more general definitions of diagonals in [90, 253, 254, 255], which are recent references for this topic.) Diagonals of D-finite power series in any number of variables are D-finite. Diagonals of two-variable rational functions are algebraic, but there are three-variable rational functions whose diagonals are not algebraic [151].

6.4. Unimodality and log-concavity

A finite sequence $a_0, a_1, \ldots, a_n$ of real numbers is called *unimodal* if for some index k, $a_0 \leq a_1 \leq \cdots \leq a_k$ and $a_k \geq a_{k+1} \geq \cdots \geq a_n$. A sequence $a_0, \ldots, a_n$ of nonnegative elements is called *log-concave* (short for logarithmically concave) if $a_j^2 \geq a_{j-1}a_{j+1}$ holds for $1 \leq j \leq n-1$. Unimodal and log-concave sequences occur frequently in combinatorics and are objects of intensive study. We present a brief review of some of their properties because asymptotic methods are often used to prove unimodality and log-concavity. Furthermore, knowledge that a sequence is log-concave or unimodal is often helpful in obtaining asymptotic information. For example, some methods provide only asymptotic estimates for summatory functions of sequences, and unimodality helps in obtaining from those estimates bounds on individual coefficients. This approach will be presented in Section 13, in the discussion of central and local limit theorems.

The basic references for unimodality and log-concavity are [222, 352]. For recent results, see also [56] and the references given there. All the results listed below can be found in those sources and the references they list.

In the rest of this subsection we will consider only sequences of nonnegative elements. A sequence $a_0, \ldots, a_n$ will be said to *have no internal zeros* if there is no triple of integers $0 \leq i < j < k \leq n$ such that $a_j = 0$, $a_i a_k \neq 0$. It is easy to see that a log-concave sequence with no internal zeros is unimodal, but there are sequences of positive elements that are unimodal but not concave. The convolution of two unimodal sequences does not have to be unimodal. However, it is unimodal if each of the two unimodal sequences is also symmetric.

Convolution of two log-concave sequences is log-concave. The convolution of a log-concave and a unimodal sequence is unimodal. A log-concave sequence is even characterized by the property that its convolution with any unimodal sequence is unimodal. This last property is related to the variation-diminishing character of log-concave sequences (see [222]), which we will not discuss at greater length here except to note that there are more restrictive sets of sequences (the Pólya frequency classes, see [56, 222]) which have stronger convolution properties.

The binomial coefficients $\binom{n}{k}$, $0 \le k \le n$, are log-concave, and therefore unimodal. The q-binomial coefficients $\begin{bmatrix} n \\ k \end{bmatrix}_q$ are log-concave for any $q \ge 1$. On the other hand, if we write a single coefficient $\begin{bmatrix} n \\ k \end{bmatrix}_q$ for fixed n and k as a polynomial in q, the sequence of coefficients is unimodal, but does not have to be log-concave.

The most frequently used method for showing that a sequence $a_0, \ldots, a_n$ is log-concave is to show that all the zeros of the polynomial

$$A(z) = \sum_{k=0}^{n} a_k z^k \tag{6.85}$$

are real and ≤ 0. In that case not only are the a_k log-concave, but so are $a_k \binom{n}{k}^{-1}$. Absolute values of the Stirling numbers of both kinds were first shown to be log-concave by this method [195]. There are many unsolved conjectures about log-concavity of combinatorial sequences, such as the Read-Hoggar conjecture that coefficients of chromatic polynomials are log-concave (cf. [57]).

A variety of combinatorial, algebraic, and geometric methods have been used to prove unimodality of sequences, and we refer the reader to [352] for a comprehensive and insightful survey. In Section 12.3 we will discuss briefly some proofs of unimodality and log-concavity that use asymptotic methods. The basic philosophy is that since the Gaussian distribution is log-concave and unimodal (when we extend the definition of these concepts to continuous distributions), these properties should also hold for sequences that by the central limit theorem or its variants are asymptotic to the Gaussian. Therefore one can expect high-order convolutions of sequences to be log-concave at least in their central region, and there are theorems that prove this under certain conditions.

6.5. Moments and distributions

The second moment method is a frequently used technique in probabilistic arguments, as is shown in Chapter ? and [55, 108, 348]. It is based on *Chebyshev's inequality*, which says that if X is a real-valued random variable with finite second moment $E(X^2)$, then

$$\text{Prob}\left(|X - E(X)| \geq \alpha |E(X)|\right) \leq \frac{E(X^2) - E(X)^2}{\alpha^2 E(X)^2} \ . \tag{6.86}$$

An easy corollary of inequality (6.86) that is often used is

$$\text{Prob}\left(X = 0\right) \leq \frac{E(X^2) - E(X)^2}{E(X)^2} \ . \tag{6.87}$$

(There is a slightly stronger version of the inequality (6.87), in which $E(X)^2$ in the denominator is replaced by $E(X^2)$.) The inequalities (6.86) and (6.87) are usually applied for $X = Y_1 + \cdots + Y_n$, where the Y_j are other random variables. The helpful feature of the inequalities is that they require only knowledge of the pairwise dependencies among the Y_j, which is easier to study than the full joint distribution of the Y_j. For other bounds on distributions that can be obtained from partial information about moments, see [343].

The reason moment bounds are mentioned at all in this chapter is that asymptotic methods are often used to derive them. Generating functions are a common and convenient method for doing this.

Example 6.8. *Waiting times for subwords.* In a continuation and application of Example 6.4, let A be a binary string of length k. How many tosses of a fair coin (with sides labeled 0 and 1) are needed on average before A appears as a block of k consecutive outcomes? By a general observation of probability theory, this is just the sum over $n \geq 0$ of the probability that A does not appear in the first n coin tosses, and thus equals

$$\sum_{n=0}^{\infty} f_A(n) 2^{-n} = F_A(1/2) = 2^k C_A(1/2) \ , \tag{6.88}$$

where the last equality follows from Eq. (6.38). Another, more general, way to derive this is to use $G_A(z)$. Note that $g_A(n)2^{-n}$ is the probability that A appears in the first n coin tosses, but not in the first $n-1$. Hence the r-th moment of the time until A appears is

$$\sum_{n=0}^{\infty} n^r g_A(n) 2^{-n} = \left(z \frac{d}{dz} \right)^r G_A(z) \bigg|_{z=1/2} \ . \tag{6.89}$$

If we take $r = 1$, we again obtain the expected waiting time given by (6.88). When we take $r = 2$, we find that the second moment of the time until the appearance of A is

$$\sum_{n=0}^{\infty} n^2 g_A(n) 2^{-n} = 2^{2k+1} C_A(1/2)^2 - (2k-1) 2^k C_A(1/2) + 2^k C_A'(1/2) \ , \tag{6.90}$$

and therefore the variance is

$$\begin{aligned} & 2^{2k} C_A(1/2)^2 - (2k-1)2^k C_A(1/2) + 2^k C'_A(1/2) \\ &= 2^{2k} C_A(1/2)^2 + O(k2^k) \,, \end{aligned} \tag{6.91}$$

since $1 \le C_A(1/2) \le 2$. Higher moments can be used to obtain more detailed information. A better approach is to use the method of Example 9.2, which gives precise estimates for the tails as well as the mean of the distribution. ■

Information about moments of distribution functions can often be used to obtain the limiting distribution. If $F_n(x)$ is a sequence of distribution functions such that for every integer $k \ge 0$, the k-th moment

$$\mu_n(k) = \int x^k dF_n(x) \tag{6.92}$$

converges to $\mu(k)$ as $n \to \infty$, then there is a limiting measure with distribution function $F(x)$ whose k-th moment is $\mu(k)$. If the moments $\mu(k)$ do not grow too rapidly, then they determine the distribution function $F(x)$ uniquely, and the $F_n(x)$ converge to $F(x)$ (in the weak star sense [50]). A sufficient condition for the $\mu(k)$ to determine $F(x)$ uniquely is that the generating function

$$U(x) = \sum_{k=0}^{\infty} \frac{\mu(2k) x^k}{(2k)!} \tag{6.93}$$

should converge for some $x > 0$. In particular, the standard normal distribution with

$$F(x) = (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-u^2/2) du \tag{6.94}$$

has $\mu(2k) = 1 \cdot 3 \cdot 5 \cdot 7 \cdot \ldots \cdot (2k-1)$ (and $\mu(2k+1) = 0$), so it is determined uniquely by its moments. On the other hand, there are some frequently encountered distributions, such as the log-normal one, which do not have this property.

7. Formal power series

This section discusses generating functions $f(z)$ that might not converge in any interval around the origin. Sequences that grow rapidly are common in combinatorics, with $a_n = n!$ the most obvious example for which

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{7.1}$$

does not converge for any $z \neq 0$. The usual way to deal with the problem of a rapidly growing sequence a_n is to study the generating function of a_n/b_n, where b_n is some sequence with known asymptotic behavior. When $b_n = n!$, the ordinary generating function of a_n/b_n is then the exponential generating function of a_n. For derangements (Eqs. (1.1) and (6.7)) this works well, as the exponential generating function of d_n converges in $|z| < 1$ and has a nice form. Unfortunately, while we can always find a sequence b_n that will make the ordinary generating function $f(z)$ of a_n/b_n converge (even for all z), usually we cannot do it in a way that will yield any useful information about $f(z)$. The combinatorial structure of a problem almost always severely restricts what forms of generating function can be used to take advantage of the special properties of the problem. This difficulty is common, for example, in enumeration of labeled graphs. In such cases one often resorts to formal power series that do not converge in any neighborhood of the origin. For example, if $c(n,k)$ is the number of connected labeled graphs on n vertices with k edges, then it is well known (cf. [349]) that

$$\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} c(n,k)\frac{x^k y^n}{n!} = \log\left(\sum_{m=0}^{\infty}\frac{(1+x)^{\binom{m}{2}}y^m}{m!}\right) . \tag{7.2}$$

While the series inside the log in (7.2) does converge for $-2 \le x \le 0$, and any y, it diverges for any $x > 0$ as long as $y \neq 0$, and so this is a relation of formal power series.

There are few methods for dealing with asymptotics of formal power series, at least when compared to the wealth of techniques available for studying analytic generating functions. Fortunately, combinatorial enumeration problems that do require the use of formal power series often involve rapidly growing sequences of positive terms, for which some simple techniques apply. We start with an easy general result that is applicable both to convergent and purely formal power series.

Theorem 7.1. ([33]) *Suppose that $a(z) = \sum a_n z^n$ and $b(z) = \sum b_n z^n$ are power series with radii of convergence $\alpha > \beta \ge 0$, respectively. Suppose that $b_{n-1}/b_n \to \beta$ as $n \to \infty$. If $a(\beta) \neq 0$, and $\sum c_n z^n = a(z)b(z)$, then*

$$c_n \sim a(\beta) b_n \quad \textit{as} \quad n \to \infty . \tag{7.3}$$

The proof of Theorem 7.1, which can be found in [33], is simple. The condition $\alpha > \beta$ is important, and cannot be replaced by $\alpha = \beta$. We can have $\beta = 0$, and that is indeed the only possibility if the series for $b(z)$ does not converge in a neighborhood of $z = 0$.

Example 7.1. *Double set coverings* [33, 80]. Let v_n be the number of choices of subsets $S_1, \ldots, S_r$ of an n-element set T such that each $t \in T$ is in exactly two of the S_i. There is no restriction on r, the number of subsets, and some of the S_i can be repeated. Let c_n be the corresponding number when the S_i are required to be distinct. We let $C(z) = \sum c_n z^n/n!$, $V(z) = \sum v_n z^n/n!$ be the exponential generating functions. Then it can be shown that

$$C(z) = \exp(-1 - (e^z - 1)/2) A(z) \ , \qquad (7.4)$$

$$V(z) = \exp(-1 + (e^z - 1)/2) A(z) \ , \qquad (7.5)$$

where

$$A(z) = \sum_{k=0}^{\infty} \exp(k(k-1)z/2)/k! \ . \qquad (7.6)$$

We see immediately that $A(z)$ does not converge in any neighborhood of the origin. We have

$$a_n = [z^n]A(z) = 2^{-n} \sum_{k=2}^{\infty} \frac{k^n (k-1)^n}{k!} \ . \qquad (7.7)$$

By considering the ratio of consecutive terms in the sum in (7.7), we find that the largest term occurs for $k = k_0$ with $k_0 \log k_0 \sim 2n$, and by the methods of Section 5.1 we find that

$$a_n \sim \frac{\pi^{1/2} k_0^n (k_0 - 1)^n}{n^{1/2} 2^n (k_0 - 1)!} \quad \text{as} \quad n \to \infty \ . \qquad (7.8)$$

Therefore $a_{n-1}/a_n \to 0$ as $n \to \infty$, and Theorem 7.1 tells us that

$$c_n \sim v_n \sim e^{-1} n! a_n \quad \text{as} \quad n \to \infty \ . \qquad (7.9)$$

■

Usually formal power series occur in more complicated relations than those covered by Theorem 7.1. For example, if f_n is the number of connected graphs on n labeled vertices which have some property, and F_n is the number of graphs on n labeled vertices each of whose connected components has that property, then (cf. [394])

$$1 + \sum_{n=1}^{\infty} F_n \frac{x^n}{n!} = \exp\left(\sum_{n=1}^{\infty} f_n \frac{x^n}{n!} \right) \ . \qquad (7.10)$$

Theorem 7.2. ([34]) *Suppose that*

$$\begin{aligned} a(x) &= \sum_{n=1}^{\infty} a_n x^n \ , \quad F(x,y) = \sum_{h,k \geq 0} f_{hk} x^h y^k \ , \\ b(x) &= \sum_{n=0}^{\infty} b_n x^n = F(x, a(x)) \ , \quad D(x) = F_y(x, a(x)) \ , \end{aligned} \qquad (7.11)$$

where $F_y(x,y)$ is the partial derivative of $F(x,y)$ with respect to y. Assume that $a_n \neq 0$ and

(i)

$$a_{n-1} = o(a_n) \quad \textit{as} \quad n \to \infty \ , \tag{7.12}$$

(ii)

$$\sum_{k=r}^{n-r} |a_k a_{n-k}| = O(a_{n-r}) \quad \textit{for some} \quad r > 0 \ , \tag{7.13}$$

(iii) for every $\delta > 0$ there are $M(\delta)$ and $K(\delta)$ such that for $n \geq M(\delta)$ and $h + k > r + 1$,

$$|f_{hk} a_{n-h-k+1}| \leq K(\delta) \delta^{h+k} |a_{n-r}| \ . \tag{7.14}$$

Then

$$b_n = \sum_{k=0}^{r-1} d_k a_{n-k} + O(a_{n-r}) \ . \tag{7.15}$$

Condition (iii) of Theorem 7.2 is often hard to verify. Theorem 2 of [34] shows that this condition holds under certain simpler hypotheses. It follows from that result that (iii) is valid if $F(x, y)$ is analytic in x and y in a neighborhood of $(0, 0)$. Hence, if $F(x, y) = \exp(y)$ or $F(x, y) = 1 + y$, then Theorem 7.2 becomes easy to apply. One can also deduce from Theorem 2 of [34] that Theorem 7.2 applies when (i) and (ii) hold, $b_0 = 0$, $b_n \geq 0$, and

$$1 + a(z) = \exp\left(\sum_{k=1}^{\infty} b(z^k)/k\right) \ , \tag{7.16}$$

another relation that is common in graph enumeration (cf. Example 15.1). There are also some results weaker than Theorem 7.2 that are easier to apply [393].

Example 7.2. *Indecomposable permutations* [81]. For every permutation σ of $\{1, \ldots, n\}$, let $\{1, \ldots, n\} = \cup I_h$, where the I_h are the smallest intervals such that $\sigma(I_h) = I_h$ for all h. For example, $\sigma = (134)(2)(56)$ corresponds to $I_1 = \{1, 2, 3, 4\}$, $I_2 = \{5, 6\}$, and the identity permutation has n components. A permutation is said to be indecomposable if it has one component. For example, if σ has the 2-cycle $(1n)$, it is indecomposable. Let c_n be the number of indecomposable permutations of $\{1, \ldots, n\}$. Then [81]

$$\sum_{n=1}^{\infty} c_n z^n = 1 - \frac{1}{1 + \sum_{n=1}^{\infty} n! z^n} \ . \tag{7.17}$$

We apply Theorem 7.2 with $a_n = n!$ for $n \geq 1$ and $F(x, y) = 1 - (1 + y)^{-1}$. We easily obtain

$$c_n \sim n! \quad \text{as} \quad n \to \infty \ , \tag{7.18}$$

so that almost all permutations are indecomposable. ∎

Some further useful expansions for functional inverses and computations of formal power series have been obtained by Bender and Richmond [40].

8. Elementary estimates for convergent generating functions

The word "elementary" in the title of this section is a technical term that means the proofs do not use complex variables. It does not necessarily imply that the proofs are simple. While some, such as those of Section 8.1, are easy, others are more complicated. The main advantage of elementary methods is that they are much easier to use, and since they impose much weaker requirements on the generating functions, they are more widely applicable. Usually they only impose conditions on the generating function $f(z)$ for $z \in \mathbb{R}^+$.

The main disadvantage of elementary methods is that the estimates they give tend to be much weaker than those derived using analytic function approaches. It is easy to explain why that is so by considering the two generating functions

$$f_1(z) = \sum_{n=0}^{\infty} z^n = (1-z)^{-1} \tag{8.1}$$

and

$$f_2(z) = 3/2 + \sum_{n=1}^{\infty} 2z^{2n} = 3/2 + 2z^2(1-z^2)^{-1} \ . \tag{8.2}$$

Both series converge for $|z| < 1$ and diverge for $|z| > 1$, and both blow up as $z \to 1$. However,

$$f_1(z) - f_2(z) = -\frac{1-z}{2(1+z)} \to 0 \quad \text{as} \quad z \to 1 \ . \tag{8.3}$$

Thus these two functions behave almost identically near $z = 1$. Since $f_1(z)$ and $f_2(z)$ are both $\sim (1-z)^{-1}$ as $z \to 1^-$, $z \in \mathbb{R}^+$, and their difference is $O(|z-1|)$ for $z \in \mathbb{R}^+$, it would require exceptionally delicate methods to detect the differences in the coefficients of the $f_j(z)$ just from their behavior for $z \in \mathbb{R}^+$. There is a substantial difference in the behavior of $f_1(z)$ and $f_2(z)$ for real z if we let $z \to -1$, so our argument does not completely exclude the possibility of obtaining detailed information about the coefficients of these functions using methods of real variables only. However, if we consider the function

$$f_3(z) = 2 + \sum_{n=1}^{\infty} 3z^{3n} = 2 + 3z^3(1-z^3)^{-1} \ , \tag{8.4}$$

then $f_1(z)$ and $f_3(z)$ are both $\sim (1-z)^{-1}$ as $z \to 1^-$, $z \in \mathbb{R}^+$, yet now

$$|f_1(z) - f_3(z)| = O(|z-1|) \quad \text{for all} \quad z \in \mathbb{R} \ .$$

This difference is comparable to what would be obtained by modifying a single coefficient of one generating function. To determine how such slight changes in the behavior of the generating functions affect the behavior of the coefficients we would need to know much more about the functions if we were to use real variable methods. On the other hand, analytic methods, discussed in Section 10 and later, are good at dealing with such problems. They require less precise knowledge of the behavior of a function on the real line. Instead, they impose weak conditions on the function in a wider domain, namely that of the complex numbers.

For reasons discussed above, elementary methods cannot be expected to produce precise estimates of individual coefficients. They often do produce good estimates of summatory functions of the coefficients, though. In the examples above, we note that

$$\sum_{n=1}^{N} [z^n] f_j(z) \sim N \quad \text{as} \quad N \to \infty \tag{8.5}$$

for $1 \le j \le 3$. This holds because the $f_j(z)$ have the same behavior as $z \to 1^-$, and is part of a more general phenomenon. Good knowledge of the behavior of the generating function on the real axis combined with weak restrictions on the coefficients often leads to estimates for the summatory function of the coefficients.

There are cases where elementary methods give precise bounds for individual coefficients. Typically when we wish to estimate f_n, with ordinary generating function $f(z) = \sum f_n z^n$ that converges for $|z| < 1$ but not for $|z| > 1$, we apply the methods of this section to

$$g_n = f_n - f_{n-1} \quad \text{for} \quad n \ge 1, \quad g_0 = f_0 \tag{8.6}$$

with generating function

$$g(z) = \sum_{n=0}^{\infty} g_n z^n = (1-z) f(z) \ . \tag{8.7}$$

Then

$$\sum_{k=0}^{n} g_k = f_n \ , \tag{8.8}$$

and so estimates of the summatory function of the g_k yield estimates for f_n. The difficulty with this approach is that now $g(z)$ and not $f(z)$ has to satisfy the hypotheses of the theorems, which requires more knowledge of the f_n. For example, most of the Tauberian theorems apply only to power series with nonnegative coefficients. Hence to use the differencing trick above to obtain estimates for f_n we need to know that $f_{n-1} \le f_n$ for all n. In some cases (such as that of $f_n = p_n$, the number of ordinary partitions of n) this is easily seen to hold through combinatorial arguments. In other situations where one might like to apply elementary methods, though, $f_{n-1} \leq f_n$ is either false or else is hard to prove. When that happens, other methods are required to estimate f_n.

8.1. Simple upper and lower bounds

A trivial upper bound method turns out to be widely applicable in asymptotic enumeration, and is surprisingly powerful. It relies on nothing more than the nonnegativity of the coefficients of a generating function.

Lemma 8.1. *Suppose that $f(z)$ is analytic in $|z| < R$, and that $[z^n]f(z) \geq 0$ for all $n \geq 0$. Then for any x, $0 < x < R$, and any $n \geq 0$,*

$$[z^n]f(z) \leq x^{-n} f(x) \ . \tag{8.9}$$

Example 8.1. *Lower bound for factorials.* Let $f(z) = \exp(z)$. Then Lemma 8.1 yields

$$\frac{1}{n!} = [z^n]e^z \leq x^{-n}e^x \tag{8.10}$$

for every $x > 0$. The logarithm of $x^{-n}e^x$ is $x - n\log x$, and differentiating and setting it equal to 0 shows that the minimum value is attained at $x = n$. Therefore

$$\frac{1}{n!} = [z^n]e^z \leq n^{-n}e^n \ , \tag{8.11}$$

and so $n! \geq n^n e^{-n}$. This lower bound holds uniformly for all n, and is off only by an asymptotic factor of $(2\pi n)^{1/2}$ from Stirling's formula (4.1). ■

Suppose that $f(z) = \sum f_n z^n$. Lemma 8.1 is proved by noting that for $0 < x < R$, the n-th term, $f_n x^n$, in the power series expansion of $f(x)$, is $\leq f(x)$. As we will see in Section 10, it is often possible to derive a similar bound on the coefficients f_n even without assuming that they are nonnegative. However, the proof of Lemma 8.1 shows something more, namely that

$$f_0 x^{-n} + f_1 x^{-n+1} + \cdots + f_{n-1}x^{-1} + f_n \leq x^{-n} f(x) \tag{8.12}$$

for $0 < x < R$. When $x \leq 1$, this yields an upper bound for the summatory function of the coefficients. Because (8.12) holds, we see that the bound of Lemma 8.1 cannot be sharp in general. What is remarkable is that the estimates obtainable from that lemma are often not far from best possible.

Example 8.2. *Upper bound for the partition function.* Let $p(n)$ denote the partition function. It has the ordinary generating function

$$f(z) = \sum_{n=0}^{\infty} p(n) z^n = \prod_{k=1}^{\infty} (1 - z^k)^{-1} . \tag{8.13}$$

Let $g(s) = \log f(e^{-s})$, and consider $s > 0$, $s \to 0$. There are extremely accurate estimates of $g(s)$. It is known [13, 23], for example, that

$$g(s) = \pi^2/(6s) + (\log s)/2 - (\log 2\pi)/2 - s/24 + O(\exp(-4\pi^2/s)) . \tag{8.14}$$

If we use (8.14), we find that $x^{-n} f(x)$ is minimized at $x = \exp(-s)$ with

$$s = \pi/(6n)^{1/2} - 1/(4n) + O(n^{-3/2}) , \tag{8.15}$$

which yields

$$p(1) + p(2) + \cdots + p(n) \le 2^{-3/4} e^{-1/4} n^{-1/4} (1 + o(1)) \exp(2\pi 6^{-1/2} n^{1/2}) . \tag{8.16}$$

Comparing this to the asymptotic formula for the sum that is obtainable from (1.6) (see Example 5.2), we see that the bound of (8.16) is too high by a factor of $n^{1/4}$. If we use (8.16) to bound $p(n)$ alone, we obtain a bound that is too large by a factor of $n^{3/4}$.

The application of Lemma 8.1 outlined above depended on the expansion (8.14), which is complicated to derive, involving modular transformation properties of $p(n)$ that are beyond the scope of this survey. (See [13, 23] for derivations.) Weaker estimates that are still useful are much easier to derive. We obtain one such bound here, since the arguments illustrate some of the methods from the preceding sections.

Consider

$$g(s) = \sum_{k=1}^{\infty} - \log(1 - e^{-ks}) . \tag{8.17}$$

If we replace the sum by the integral

$$I(s) = \int_1^{\infty} - \log(1 - e^{-us}) du , \tag{8.18}$$

we find on expanding the logarithm that

$$I(s) = \int_1^{\infty} \left(\sum_{m=1}^{\infty} m^{-1} e^{-mus} \right) du = s^{-1} \sum_{m=1}^{\infty} m^{-2} e^{-ms} , \tag{8.19}$$

since the interchange of summation and integration is easy to justify, as all the terms are positive. Therefore as $s \to 0^+$,

$$sI(s) \to \sum_{m=1}^{\infty} m^{-2} = \pi^2/6 \ , \tag{8.20}$$

so that $I(s) \sim \pi^2/(6s)$ as $s \to 0^+$. It remains to show that I is indeed a good approximation to $g(s)$. This follows easily from the bound (5.32), since it shows that

$$g(s) = I(s) + O\left(\int_1^{\infty} \frac{se^{-vs}}{1-e^{-vs}} dv\right) \ . \tag{8.21}$$

We could estimate the integral in (8.21) carefully, but we only need rough upper bounds for it, so we write it as

$$\begin{aligned}
\int_1^{\infty} \frac{se^{-vs}}{1-e^{-vs}} dv &= \int_s^{\infty} \frac{e^{-u}}{1-e^{-u}} du \\
&= \int_s^{1} \frac{e^{-u}}{1-e^{-u}} du + \int_1^{\infty} \frac{e^{-u}}{1-e^{-u}} du \\
&= \int_s^{1} \frac{du}{e^u - 1} + c \leq \int_s^1 \frac{du}{u} + c = c - \log s
\end{aligned} \tag{8.22}$$

for some constant c. Thus we find that

$$g(s) = I(s) + O(\log(s^{-1})) \quad \text{as} \quad s \to 0^+ \ . \tag{8.23}$$

Combining (8.23) with (8.20) we see that

$$g(s) \sim \pi^2/(6s) \quad \text{as} \quad s \to 0^+ \ . \tag{8.24}$$

Therefore, choosing $s = \pi/(6n)^{1/2}$, $x = \exp(-s)$ in Lemma 8.1, we obtain a bound of the form

$$p(n) \leq \exp((1+o(1))\pi(2/3)^{1/2} n^{1/2}) \quad \text{as} \quad n \to \infty \ . \quad \blacksquare \tag{8.25}$$

Lemma 8.1 yields a lower bound for $n!$ that is only a factor of about $n^{1/2}$ away from optimal. That is common. Usually, when the function $f(z)$ is reasonably smooth, the best bound obtainable from Lemma 8.1 will only be off from the correct value by a polynomial factor of n, and often only by a factor of $n^{1/2}$.

The estimate of Lemma 8.1 can often be improved with some additional knowledge about the f_n. For example, if $f_{n+1} \geq f_n$ for all $n \geq 0$, then we have

$$x^{-n} f(x) \geq f_n + f_{n+1}x + f_{n+2}x^2 + \cdots \geq f_n(1-x)^{-1} \ . \tag{8.26}$$

For $f_n = p(n)$, the partition function, then yields an upper bound for $p(n)$ that is too large by a factor of $n^{1/4}$.

To optimize the bound of Lemma 8.1, one should choose $x \in (0, R)$ carefully. Usually there is a single best choice. In some pathological cases the optimal choice is obtained by letting $x \to 0^+$ or $x \to R^-$. However, usually we have $\lim_{x \to R^-} f(x) = \infty$, and $[z^m]f(z) > 0$ for some m with $0 \leq m < n$ as well as for some $m > n$. Under these conditions it is easy to see that

$$\lim_{x \to 0^+} x^{-n} f(x) = \lim_{x \to R^-} x^{-n} f(x) = \infty \ . \tag{8.27}$$

Thus it does not pay to make x too small or too large. Let us now consider

$$g(x) = \log(x^{-n} f(x)) = \log f(x) - n \log x \ . \tag{8.28}$$

Then

$$g'(x) = \frac{f'}{f}(x) - \frac{n}{x} \ , \tag{8.29}$$

and the optimal choice must be at a point where $g'(x) = 0$. For most commonly encountered functions $f(x)$, there exists a constant $x_0 > 0$ such that

$$\left(\frac{f'}{f}\right)'(x) > 0 \tag{8.30}$$

for $x_0 < x < R$, and so $g''(x) > 0$ for all $x \in (0, R)$ if n is large enough. For such n there is then a unique choice of x that minimizes the bound of Lemma 8.1. However, one major advantage of Lemma 8.1 is that its bound holds for all x. To apply this lemma, one can use any x that is convenient to work with. Usually if this choice is not too far from the optimal one, the resulting bound is fairly good.

We have already remarked above that the bound of Lemma 8.1 is usually close to best possible. It is possible to prove general lower bounds that show this for a wide class of functions. The method, originated in [277] and developed in [305], relies on simple elementary arguments. However, the lower bounds it produces are substantially weaker than the upper bounds of Lemma 8.1. Furthermore, to apply them it is necessary to estimate accurately the minimum of $x^{-n} f(x)$, instead of selecting any convenient values of x. A more general version of the bound below is given in [305].

Theorem 8.1. *Suppose that $f(x) = \sum f_n x^n$ converges for $|x| < 1$, $f_n \geq 0$ for all n, $f_{m_0} > 0$ for some m_0, and $\sum f_n = \infty$. Then for $n \geq m_0$, there is a unique $x_0 = x_0(n) \in (0, 1)$ that* *minimizes* $x^{-n} f(x)$. *Let* $s_0 = -\log x_0$, *and*

$$A = \frac{\partial^2}{\partial s^2} \log f(e^{-s}) \Big|_{s=s_0} . \tag{8.31}$$

If $A \geq 10^6$ *and for all* t *with*

$$s_0 \leq t \leq s_0 + 20A^{-1/2} \tag{8.32}$$

we have

$$\left| \frac{\partial^3}{\partial s^3} \log f(e^{-s}) \Big|_{s=t} \right| \leq 10^{-3} A^{3/2} , \tag{8.33}$$

then

$$\sum_{k=0}^{n} f_k \geq x_0^{-n} f(x_0) \exp(-30 s_0 A^{1/2} - 100) . \tag{8.34}$$

As is usual for Tauberian theorems, Theorem 8.1 only provides bounds on the sum of coefficients of $f(z)$. As we mentioned before, this is unavoidable when one relies only on information about the behavior of $f(z)$ for z a positive real number. The conditions that Theorem 8.1 imposes on the derivatives are usually satisfied in combinatorial enumeration applications and are easy to verify.

Example 8.3. *Lower bound for the partition function.* Let $f(z)$ and $g(s)$ be as in Example 8.2. We showed there that $g(s)$ satisfies (8.24) and similar rough estimates show that $g'(s) \sim -\pi^2/(6s^2)$, $g''(s) \sim \pi^2/(3s^3)$, and $g'''(s) \sim -\pi^2/s^4$ as $s \to 0^+$. Therefore the hypotheses of Theorem 8.1 are satisfied, and we obtain a lower bound for $p(0) + p(1) + \cdots + p(n)$. If we only use the estimate (8.24) for $g(s)$, then we can only conclude that for $x = e^{-s}$,

$$\log(x^{-n} f(x)) = ns + g(s) \sim ns + \pi^2/(6s) \quad \text{as} \quad s \to 0 , \tag{8.35}$$

and so the minimum value occurs at $s \sim \pi/(6n)^{1/2}$ as $n \to \infty$. This only allows us to conclude that for every $\epsilon > 0$ and n large enough,

$$\log(p(0) + \cdots + p(n)) \geq (1 - \epsilon) \pi (2/3)^{1/2} n^{1/2} . \tag{8.36}$$

However, we can also conclude even without further computations that this lower bound will be within a multiplicative factor of $\exp(cn^{1/4})$ of the best upper bound that can be obtained from Lemma 8.1 for some $c > 0$ (and therefore within a multiplicative factor of $\exp(cn^{1/4})$ of the correct value). In particular, if we use the estimate (8.14) for $g(s)$, we find that for some $c' > 0$,

$$p(0) + \cdots + p(n) \geq \exp(\pi (2/3)^{1/2} n^{1/2} - c' n^{1/4}) . \tag{8.37}$$

Since $p(k) \le p(k+1)$, the quantity on the right-hand side of (8.37) is also a lower bound for $p(n)$ if we increase c', since $(n+1)p(n) \ge p(0) + \cdots + p(n)$. ■

The differencing trick described at the introduction to Section 8 could also be used to estimate $p(n)$, since Theorem 8.1 can be applied to the generating function of $p(n+1) - p(n)$. However, since the error term is a multiplicative factor of $\exp(cn^{1/4})$, it is simpler to use the approach above, which bounds $p(n)$ below by $(p(0) + \cdots + p(n))/(n+1)$.

Brigham [58] has proved a general theorem about asymptotics of partition functions that can be derived from Theorem 8.1. (For other results and references for partition asymptotics, see [13, 23, 150].)

Theorem 8.2. *Suppose that*

$$f(z) = \prod_{k=1}^{\infty} (1 - z^k)^{-b(k)} = \sum_{n=0}^{\infty} a(n) z^n \ , \tag{8.38}$$

where the $b(k) \in \mathbf{Z}$, $b(k) \ge 0$ for all k, and that for some $C > 0$, $u > 0$, we have

$$\sum_{k \le x} b(k) \sim C x^u (\log x)^v \quad \text{as} \quad x \to \infty \ . \tag{8.39}$$

Then

$$\begin{aligned} \log\left(\textstyle\sum_{n \le m} a(n)\right) &\sim u^{-1} \{C u \Gamma(u+2) \zeta(u+1)\}^{1/(u+1)} \\ &\quad \cdot (u+1)^{(u-v)/(u+1)} m^{u/(u+1)} (\log m)^{v/(u+1)} \end{aligned} \tag{8.40}$$

as $m \to \infty$.

If $b(k) = 1$ for all k, $a(n)$ is p_n, the ordinary partition function. If $b(k) = k$ for all k, $a(n)$ is the number of plane partitions of n. Thus Brigham's theorem covers a wide class of interesting partition functions. The cost of this generality is that we obtain only the asymptotics of the logarithm of the summatory function of the partitions being enumerated. (For better estimates of the number of plane partitions, for example, see [9, 170, 387]. For ordinary partitions, we have the expansion (1.3).)

Brigham's proof of Theorem 8.2 first shows that

$$f(e^{-w}) \sim C w^{-u} (-\log w)^v \Gamma(u+1) \zeta(u+1) \quad \text{as} \quad w \to 0^+ \tag{8.41}$$

and then invokes the Hardy-Ramanujan Tauberian theorem [328]. Instead, one can obtain a proof from Theorem 8.1. The advantage of using Theorem 8.1 is that it is much easier to generalize. Hardy and Ramanujan proved their Tauberian theorem only for functions whose growth rates are of the form given by (8.41). Their approach can be extended to other functions, but this is complicated to do. In contrast, Theorem 8.1 is easy to apply. The conditions of Theorem 8.1 on the derivatives are not restrictive. For a function $f(z)$ defined by (8.38) we have $B \to \infty$ if $\sum b(k) = \infty$, and the condition (8.33) can be shown to hold whenever there are constants c_1 and c_2 such that for all $w > 1$, and all sufficiently large m,

$$\sum_{k \leq mw} b(k) \leq c_1 w^{c_2} \sum_{k \leq m} b(k) \ , \tag{8.42}$$

say. The main difficulty in applying Theorem 8.1 to generalizations of Brigham's theorem is in accurately estimating the minimal value in Lemma 8.1.

There are many other applications of Lemma 8.1 and Theorem 8.1. For example, they can be used to prove the results of [158] on volumes of spheres in the Lee metric.

Lemma 8.1 can be generalized in a straightforward way to multivariate generating functions. If

$$f(x, y) = \sum_{m,n \geq 0} a_{m,n} x^m y^n \tag{8.43}$$

and $a_{m,n} \geq 0$ for all m and n, then for any $x, y > 0$ for which the sum in (8.43) converges we have

$$a_{m,n} \leq x^{-m} y^{-n} f(x, y) \ . \tag{8.44}$$

Generalizations of the lower bound of Theorem 8.1 to multivariate functions can also be derived, but are again harder than the upper bound [289].

8.2. Tauberian theorems

The Brigham Tauberian theorem for partitions [58], based on the Hardy-Ramanujan Tauberian theorem [328], was quoted already in Section 8.1. It applies to certain generating functions that have (in notation to be introduced in Section 10) a large singularity and gives estimates only for the logarithm of the summatory function of the coefficients. Another theorem that is often more precise, but is again designed to deal with rapidly growing partition functions, is that of Ingham [212], and will be discussed at the end of this section. Most of the Tauberian theorems in the literature apply to functions with small singularities (i.e., ones that do not grow rapidly as the argument approaches the circle of convergence) and give asymptotic relations for the sum of coefficients. References for Tauberian theorems are [117, 154, 190, 212, 325]. Their main advantage is generality and ease of use, as is shown by the applications made to 0-1 laws in [77, 78, 79]. They can often be applied when the information about generating functions is insufficient to use the methods of Sections 11 and 12. This is especially true when the circle inside which the generating function converges is a natural boundary beyond which the function cannot be continued.

One Tauberian theorem that is often used in combinatorial enumeration is that of Hardy, Littlewood, and Karamata. We say a function $L(t)$ varies slowly at infinity if, for every $u > 0$, $L(ut) \sim L(t)$ as $t \to \infty$.

Theorem 8.3. *Suppose that* $a_k \geq 0$ *for all* k, *and that*

$$f(x) = \sum_{k=0}^{\infty} a_k x^k$$

converges for $0 \leq x < r$. *If there is a* $\rho \geq 0$ *and a function* $L(t)$ *that varies slowly at infinity such that*

$$f(x) \sim (r-x)^{-\rho} L\left(\frac{1}{r-x}\right) \quad \text{as} \quad x \to r- \quad , \tag{8.45}$$

then

$$\sum_{k=0}^{n} a_k r^k \sim (n/r)^{\rho} L(n)/\Gamma(\rho+1) \quad \text{as} \quad n \to \infty \ . \tag{8.46}$$

Example 8.4. *Cycles of permutations* ([33]). If S is a set of positive integers, and f_n the probability that a random permutation on n letters will have all cycle lengths in S (i.e., $f_n = a_n/n!$, where a_n is the number of permutations with cycle length in S), then

$$f(z) = \sum_{n=0}^{\infty} f_n z^n = \prod_{k \in S} \exp(z^k/k) = (1-z)^{-1} \prod_{k \notin S} \exp(-z^k/k) \ . \tag{8.47}$$

If $|\mathbb{Z}^+ \setminus S| < \infty$, then the methods of Sections 10.2 and 11 apply easily, and one finds that

$$f_n \sim \exp\left(-\sum_{k \notin S} 1/k\right) \quad \text{as} \quad n \to \infty \ . \tag{8.48}$$

This estimate can also be proved to apply for $|\mathbb{Z}^+ \setminus S| = \infty$, provided $|\{1, \ldots, m\} \setminus S|$ does not grow too rapidly when $m \to \infty$. If $|S| < \infty$ (or when $|\{1, \ldots, m\} \cap S|$ does not grow rapidly), the methods of Section 12 apply. When $S = \{1, 2\}$, one obtains, for example, the result of Moser and Wyman [292] that the number of permutations of order 2 is

$$\sim (n/e)^{n/2} 2^{-1/2} \exp(n^{1/2} - 1/4) \quad \text{as} \quad n \to \infty \ . \tag{8.49}$$

(For sharper and more general results, see [292, 376].) The methods used in these cases are different from the ones we are considering in this section.

We now consider an intermediate case, with

$$|\{1,\ldots,m\}\cap S| \sim \rho m \quad \text{as} \quad m \to \infty \ . \tag{8.50}$$

for some fixed ρ, $0 \le \rho \le 1$. This case can be handled by Tauberian techniques. To apply Theorem 8.3, we need to show that $L(t) = f(1-t^{-1})t^{-\rho}$ varies slowly at infinity. This is equivalent to showing that for any $u \in (0,1)$,

$$f(1-t^{-1}) \sim f(1-t^{-1}u)u^{\rho} \quad \text{as} \quad t \to \infty \ . \tag{8.51}$$

Because of (8.47), it suffices to prove that

$$\sum_{k\in S} k^{-1}\{(1-t^{-1})^k - (1-t^{-1}u)^k\} = \rho \log u + o(1) \quad \text{as} \quad t \to \infty \ , \tag{8.52}$$

but this is easy to deduce from (8.50) using summation by parts (Section 5). Therefore we find from Theorem 8.3 that

$$\sum_{n=0}^{m} f_n \sim f(1-1/n)\Gamma(\rho+1)^{-1} \quad \text{as} \quad n \to \infty \ . \tag{8.53}$$

(For additional results and references on this problem see [317].) ■

As the above example shows, Tauberian theorems yield estimates under weak assumptions. These theorems do have some disadvantages. Not only do they usually estimate only the summatory function of the coefficients, but they normally give no bounds for the error term. (See [154] for some Tauberian theorems with remainder terms.) Furthermore, they usually apply only to functions with nonnegative coefficients. Sometimes, as in the following theorem of Hardy and Littlewood, one can relax the nonnegativity condition slightly.

Theorem 8.4. *Suppose that* $a_k \ge -c/k$ *for some* $c > 0$*,*

$$f(z) = \sum_{k=1}^{\infty} a_k x^k \ , \tag{8.54}$$

and that $f(x)$ *converges for* $0 < x < 1$*, and that*

$$\lim_{x\to 1-} f(x) = A \ . \tag{8.55}$$

Then

$$\lim_{n\to\infty} \sum_{k=1}^{n} a_k = A \ . \tag{8.56}$$

Some condition such as $a_k \geq -c/k$ on the a_k is necessary, or otherwise the theorem would not hold. For example, the function

$$f(x) = \frac{1-x}{1+x} = 1 - 2x + 2x^2 \cdots \tag{8.57}$$

satisfies (8.55) with $A = 0$, but (8.56) fails.

We next present an example that shows an application of the above results in combination with other asymptotic methods that were presented before.

Example 8.5. *Permutations with distinct cycle lengths.* The probability that a random permutation on n letters will have cycles of distinct lengths is $[z^n]f(z)$, where

$$f(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z^k}{k}\right) . \tag{8.58}$$

Greene and Knuth [177] note that this is also the limit as $p \to \infty$ of the probability that a polynomial of degree n factors into irreducible polynomials of distinct degrees modulo a prime p. It is shown in [177] that

$$[z^n]f(z) = e^{-\gamma}(1 + n^{-1}) + O(n^{-2}\log n) \quad \text{as} \quad n \to \infty , \tag{8.59}$$

where $\gamma = 0.577\ldots$ is Euler's constant. A simplified version of the argument of [177] will be presented that shows that

$$[z^n]f(z) \sim e^{-\gamma} \quad \text{as} \quad n \to \infty . \tag{8.60}$$

Methods for obtaining better expansions, even more precise than that of (8.59), are discussed in Section 11.2. For related results obtained by probabilistic methods, see [20].

We have, for $|z| < 1$,

$$\begin{aligned} f(z) &= (1+z)\exp\left(\sum_{k=2}^{\infty} \log(1 + z^k/k)\right) \\ &= (1+z)\exp\left(\sum_{k=2}^{\infty} z^k/k + g(z)\right) \\ &= (1+z)(1-z)^{-1}\exp(g(z)) , \end{aligned} \tag{8.61}$$

where

$$g(z) = -z + \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} \sum_{k=2}^{\infty} \frac{z^{mk}}{k^m} . \tag{8.62}$$

Since the coefficients of $g(z)$ are small, the double sum in (8.62) converges for $z = 1$, and we have

$$\begin{aligned} g(1) &= \lim_{z\to 1-} g(z) = -1 + \sum_{k=2}^{\infty}\sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} k^{-m} \\ &= -1 + \sum_{k=2}^{\infty} \{\log(1+k^{-1}) - k^{-1}\} \\ &= -\log 2 + \lim_{n\to\infty} (\log(n+1) - H_n) = -\log 2 - \gamma \ , \end{aligned} \tag{8.63}$$

where $H_n = 1 + 1/2 + 1/3 + \cdots + 1/n$ is the n-th *harmonic number*. Therefore, by (8.61), we find from Theorem 8.4 that if $f_n = [z^n]f(z)$, then

$$f_0 + f_1 + \cdots + f_n \sim ne^{-\gamma} \quad \text{as} \quad n \to \infty \ . \tag{8.64}$$

To obtain asymptotics of f_n, we note that if $h_n = [z^n]\exp(g(z))$, then by (8.61),

$$f_n = 2h_0 + 2h_1 + \cdots + 2h_{n-1} + h_n \ . \tag{8.65}$$

We next obtain an upper bound for $|h_n|$. There are several ways to proceed. The method used below gives the best possible result $|h_n| = O(n^{-2})$.

Since $g(z)$ has the power series expansion (8.62), and $h_n = [z^n]\exp(g(z))$, comparison of terms in the full expansion of $\exp(g(z))$ and $\exp(v(z))$ shows that $|h_n| \le [z^n]\exp(v(z))$, where $v(z)$ is any power series such that $|[z^n]g(z)| \le [z^n]v(z)$. For $n \ge 2$,

$$[z^n]g(z) = \sum_{\substack{m|n \\ m\ge 2 \\ m<n}} \frac{(-1)^{m-1}}{m}\left(\frac{m}{n}\right)^m \ . \tag{8.66}$$

The term $(m/n)^m$ is monotone decreasing for $1 \le m \le n/e$, since its derivative with respect to m is ≤ 0 in that range. Therefore

$$|[z^n]g(z)| \le \frac{1}{2}\left(\frac{2}{n}\right)^2 + \sum_{3\le m\le n/3} \frac{1}{m}\left(\frac{3}{n}\right)^3 + \frac{2}{n}2^{-n/2} \le 10n^{-2} \ , \tag{8.67}$$

say. Hence we can take

$$v(z) = 10\sum_{n=1}^{\infty} n^{-2}z^n \ , \tag{8.68}$$

and then we need to estimate

$$w_n = [z^n]\exp(v(z)) \ . \tag{8.69}$$

We let $w(z) = \exp(v(z))$, and note that

$$w'(z) = v'(z)w(z) \ , \tag{8.70}$$

so for $n \geq 1$,

$$nw_n = 10 \sum_{k=0}^{n-1} w_k (n-k)^{-1} \ . \tag{8.71}$$

Further, since $v(1) < \infty$, and $w_n \geq 0$ for all n, we have $w_n \leq A = w(1) = \exp(v(1))$ for all n. Let $B = 10^6 A$ and note that $w_n \leq Bn^{-2}$ for $1 \leq n \leq 10^3$. Suppose now that $w_m \leq Bm^{-2}$ for $1 \leq m < n$ for some $n \geq 10^3$. We will prove that $w_n \leq Bn^{-2}$, and then by induction this inequality will hold for all $n \geq 1$. We apply Eq. (8.70). For $0 \leq k \leq 100$, we use $w_k \leq A$, $(n-k)^{-1} \leq 2n^{-1}$. For $100 < k \leq n/2$,

$$w_k (n-k)^{-1} \leq Bk^{-2}(n-k)^{-1} \leq 2Bk^{-2}n^{-1} \ , \tag{8.72}$$

and so

$$\sum_{100 \leq k \leq n/2} w_k (n-k)^{-1} \leq B(40n)^{-1} \ . \tag{8.73}$$

Finally,

$$\sum_{n/2 < k \leq n-1} w_k (n-k)^{-1} \leq 4Bn^{-2} \sum_{n/2 < k \leq n-1} (n-k)^{-1} \leq 4Bn^{-2} H_n \ . \tag{8.74}$$

Therefore, by (8.71),

$$nw_n \leq 2000An^{-1} + B(4n)^{-1} + 4BH_n n^{-2} \leq Bn^{-1} \ , \tag{8.75}$$

which completes the induction step and proves that $w_n \leq Bn^{-2}$ for all $n \geq 1$. ■

There are Tauberian theorems that apply to generating functions with rapidly growing coefficients but are more precise than Brigham's theorem or the estimates obtainable with the methods of Section 8.1. One of the most useful is Ingham's Tauberian theorem for partitions [212]. The following result is a corollary of the more general Theorem 2 of [212].

Theorem 8.5. *Let* $1 \leq u_1 < u_2 < \ldots$ *be positive integers such that*

$$|\{u_j : u_j \leq x\}| = Bx^{\beta} + R(x) \ , \tag{8.76}$$

where $B > 0$, $\beta > 0$, *and*

$$\int_1^y x^{-1} R(x) dx = b \log y + c + o(1) \quad \text{as} \quad y \to \infty \ . \tag{8.77}$$

Let

$$a(z) = \sum_{n=1}^{\infty} a_n z^n = \prod_{j=1}^{\infty} (1 - z^{u_j})^{-1} , \tag{8.78}$$

$$a^*(z) = \sum_{n=1}^{\infty} a_n^* z^n = \prod_{j=1}^{\infty} (1 + z^{u_j}) . \tag{8.79}$$

Then, as $m \to \infty$,

$$\sum_{n=1}^{m} a_n \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} e^c V^{-\alpha(b+1/2)} m^{(b+1/2)(1-\alpha)-1/2} \exp(\alpha^{-1}(Vm)^{\alpha}) , \tag{8.80}$$

$$\sum_{n=1}^{m} a_n^* \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} 2^b (V^* m)^{-\alpha/2} \exp(\alpha^{-1}(V^* m)^{\alpha}) , \tag{8.81}$$

where

$$\alpha = \beta(\beta+1)^{-1}, \quad V = \{B\beta\Gamma(\beta+1)\zeta(\beta+1)\}^{1/\beta} , \quad V^* = (1-2^{-\beta})^{1/\beta} V . \tag{8.82}$$

If $u_1 = 1$, *then as* $n \to \infty$

$$a_n \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} e^c V^{-\alpha(b-1/2)} n^{(b-1/2)(1-\alpha)-1/2} \exp(\alpha^{-1}(Vn)^{\alpha}) , \tag{8.83}$$

and if $1, 2, 4, 8, \ldots$ *all belong to* $\{u_j\}$, *then*

$$a_n^* \sim (2\pi)^{-1/2}(1-\alpha)^{1/2} 2^b (V^*)^{\alpha/2} n^{\alpha/2-1} \exp(\alpha^{-1}(V^* n)^{\alpha}) . \tag{8.84}$$

Theorem 8.5 provides more precise information than Brigham's Theorem 8.2, but under more restrictive conditions. It is derived from Ingham's main result, Theorem 1 of [212], which can be applied to wider classes of functions. However, that theorem cannot be used to derive Theorem 8.2. The disadvantage of Ingham's main theorem is that it requires knowledge of the behavior of the generating function in the complex plane, not just on the real axis. On the other hand, the region where this behavior has to be known is much smaller than it is for the analytic methods that give more accurate answers, and which are presented in Sections 10–12. Only behavior of the generating functions $\Pi(1 - z^{\lambda_j})^{-1}$ or $\Pi(1 + z^{\lambda_j})$ in an angle $|\text{Arg}|(1 - z)| \le \pi/2 - \delta$ for some $\delta > 0$ needs to be controlled.

Ingham's paper [212] contains an extended discussion of the relations between different Tauberian theorems and of the necessity for various conditions.

9. Recurrences

This section presents some basic methods for handling recurrences. The title is slightly misleading, since almost all of this chapter is devoted to methods that are useful in this area. Almost all asymptotic estimation problems concern quantities that are defined through implicit or explicit recurrences. Furthermore, the most common and most effective method of solving recurrences is often to determine its generating function and then apply the methods presented in the other sections. However, there are many recurrences, and those discussed in Sections 9.4 and 9.5 require special methods that do not fit into other sections. These methods deserve to be included, so it seems preferable to explain them after treating some of the more common types of recurrences, even though those could have been covered elsewhere in this chapter.

Since generating functions are the most powerful tool for handling combinatorial recurrences, all the books listed in Section 18 that help in dealing with combinatorial identities and generating functions are also useful in handling recurrences. Methods for recurrences that are not amenable to generating function methods are presented in [175, 177]. Lueker [264] is an introductory survey to some recurrence methods.

Wimp's book [382] is concerned primarily with numerical stability problems in computing with recurrences. Such problems are important in computing values of orthogonal polynomials, for example, but seldom arise in combinatorial enumeration. However, there are sections of [382] that are relevant to our topic, for example to the discussion of differential equations in Section 9.2.

9.1. Linear recurrences with constant coefficients

The most famous sequence that satisfies a linear recurrence with constant coefficients is that of the Fibonacci numbers, defined by $F_0 = F_1 = 1$, $F_n = F_{n-1} + F_{n-2}$ for $n \geq 2$. There are many others that are only slightly less well known. Fortunately, the theory of such sequences is well developed, and from the standpoint of asymptotic enumeration their behavior is well understood. (For a survey of number theoretic results, together with a list of many unsolved problems about such sequences that arise in that area, see [73].) There are even several different approaches to solving linear recurrences with constant coefficients. The one we emphasize here is that of generating functions, since it fits in best with the rest of this chapter. For other approaches, see [287, 298], for example.

Suppose that we have a linear recurrence or a system of recurrences and have found that the generating function $f(z)$ we are interested in has the form

$$f(z) = \frac{G(z)}{h(z)} \ , \tag{9.1}$$

where $G(z)$ and $h(z)$ are polynomials. The basic tool for obtaining asymptotic information about $[z^n]f(z)$ is the partial fraction expansion of a rational function [205]. Dividing $G(z)$ by $h(z)$ we obtain

$$f(z) = p(z) + \frac{g(z)}{h(z)} \ , \tag{9.2}$$

where $p(z)$, $g(z)$, and $h(z)$ are all polynomials in z and $\deg g(z) < \deg h(z)$. We can assume that $h(0) \neq 0$, since if that were not the case, we would have $g(0) = 0$ (as in the opposite case $f(z)$ would not be a power series in z, but would have terms such as z^{-1} or z^{-2}) and we could cancel a common factor of z from $g(z)$ and $h(z)$. Therefore, if $d = \deg h(z)$, we can write

$$h(z) = h(0) \prod_{j=1}^{d'} \left(1 - \frac{z}{z_j}\right)^{m_j} \ , \tag{9.3}$$

where z_j, $1 \leq j \leq d'$ are the distinct roots of $h(z) = 0$, z_j has multiplicity $m_j \geq 1$, and $\sum m_j = d$. Hence we find [175, 205] that for certain constants $c_{j,k}$,

$$\begin{aligned} f(z) &= p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} \frac{c_{j,k}}{(1 - z/z_j)^k} \\ &= p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} c_{j,k} \sum_{h=0}^{\infty} \binom{h+k-1}{k-1} z^h z_j^{-h} \ . \end{aligned} \tag{9.4}$$

Thus

$$a_n = [z^n]p(z) + \sum_{j=1}^{d'} \sum_{k=1}^{m_j} c_{j,k} \binom{h+k-1}{k-1} z_j^{-n} \ . \tag{9.5}$$

When $m_j = 1$,

$$c_{j,1} = \frac{-g(z_j)}{z_j h'(z_j)} \ , \tag{9.6}$$

and explicit formulas for the $c_{j,k}$ when $m_j > 1$ can also be derived [175], but are unwieldy and seldom used.

Example 9.1. *Fibonacci numbers.* As was noted in Example 6.3,

$$F(z) = \sum_{n=0}^{\infty} F_n z^n = \frac{z}{1 - z - z^2} \ .$$

Now

$$h(z) = 1 - z - z^2 = (1 + \phi^{-1} z)(1 - \phi z) \ , \tag{9.7}$$

where $\phi = (1 + 5^{1/2})/2$ is the golden ratio. Therefore

$$F(z) = \frac{1}{\sqrt{5}} \left(\frac{1}{1 - \phi z} - \frac{1}{1 + \phi^{-1} z} \right) \tag{9.8}$$

and for $n \geq 0$,

$$F_n = [z^n] F(z) = 5^{-1/2} (\phi^n - (-\phi)^{-n}) \ . \tag{9.9}$$

■

The partial fraction expansion (9.4) shows that the first-order asymptotics of sequence a_n satisfying a linear recurrence of the form (6.30) are determined by the smallest zeros of the characteristic polynomial $h(z)$. The full asymptotic expansion is given by (9.5), and involves all the zeros. In practice, using (9.5) presents some difficulties, in that multiplicities of zeros are not always easy to determine, and the coefficients $c_{j,k}$ are often even harder to deal with. Eventually, for large n, their influence becomes negligible, but when uniform estimates are required they present a problem. In such cases the following theorem is often useful.

Theorem 9.1. *Suppose that $f(z) = g(z)/h(z)$, where $g(z)$ and $h(z)$ are polynomials, $h(0) \neq 0$, $\deg g(z) < \deg h(z)$, and that the only zeros of $h(z)$ in $|z| < R$ are $\rho_1, \ldots, \rho_k$, each of multiplicity 1. Suppose further that*

$$\max_{|z|=R} |f(z)| \leq W \ , \tag{9.10}$$

and that $R - |\rho_j| \geq \delta$ for some $\delta > 0$ and $1 \leq j \leq k$. Then

$$\left| [z^n] f(z) + \sum_{j=1}^{k} \frac{g(\rho_j)}{h'(\rho_j)} \rho_j^{-n-1} \right| \leq W R^{-n} + \delta^{-1} R^{-n} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)| \ . \tag{9.11}$$

Theorem 9.1 is derived using methods of complex variables, and a proof is sketched in Section 10. That section also discusses how to locate all the zeros $\rho_1, \ldots, \rho_k$ of a polynomial $h(z)$ in a disk $|z| < R$. In general, the zero location problem is not a serious one in enumeration problems. Usually there is a single positive real zero that is closer to the origin than any other, it can be located accurately by simple methods, and R is chosen so that $|z| < R$ encloses only that zero.

Example 9.2. *Sequences with forbidden subblocks.* We continue with the problem presented in Examples 6.4 and 6.8. Both $F_A(z)$ and $G_A(z)$ have as denominators

$$h(z) = z^k + (1 - 2z) C_A(z) \ , \tag{9.12}$$

which is a polynomial of degree exactly k. Later, in Example 10.6, we will show that for $k \geq 9$, $h(z)$ has exactly one zero ρ in $|z| \leq 0.6$, and that for $|z| = 0.55$, $|h(z)| \geq 1/100$. Furthermore, by Example 6.7, $\rho \to 1/2$ as $k \to \infty$. On $|z| = 0.55$,

$$|F_A(z)| \leq 100 \cdot (0.55)^k \ . \tag{9.13}$$

Theorem 9.1 then shows, for example, that for $n > k \geq k_0$,

$$\begin{aligned} \left| [z^n]F_A(z) + \frac{C_A(\rho)\rho^{-n-1}}{h'(\rho)} \right| &\leq 100(0.55)^{k-n} + 40(0.55)^{-n}|h'(\rho)|^{-1} \\ &\leq 50(0.55)^{-n} \ , \end{aligned} \tag{9.14}$$

since by Example 6.7, as $k \to \infty$,

$$h'(\rho) = k\rho^{k-1} - 2C_A(\rho) + (1-2\rho)C'_A(\rho) \sim -2C_A(\rho) \sim -\rho^{-1} \ . \tag{9.15}$$

The estimate (9.14), when combined with the expansions of Example 6.7, gives accurate approximations for p_n, the probability that A does not appear as a block among the first n coin tosses. We have

$$\begin{aligned} p_n &= 2^{-n}[z^n]F_z(z) \\ &= -2^{-n}C_A(\rho)\rho^{-n-1}(h'(\rho))^{-1} + O(\exp(-0.09n)) \ . \end{aligned} \tag{9.16}$$

We now estimate $h'(\rho)$ as before, in (9.15), but more carefully, putting in the approximation for ρ from Example 6.7. We find that

$$h'(\rho) = -\rho^{-1} + O(k2^{-k}) \ , \tag{9.17}$$

and

$$\rho^{-n} = 2^n \exp(-n(2^k C_A(1/2))^{-1} + O(nk2^{-2k})) \ . \tag{9.18}$$

Therefore

$$p_n = \exp(-n(2^k C_A(1/2))^{-1} + O(nk2^{-2k})) + O(\exp(-n/12)) \ . \tag{9.19}$$

This shows that p_n has a sharp transition. It is close to 1 for $n = o(2^k)$, and then, as n increases through 2^k, drops rapidly to 0. (The behavior on the two sides of 2^k is not symmetric, as the drop towards 0 beyond 2^k is much faster than the increases towards 1 on the other side.) For further results and applications of such estimates, see [180, 181]. Estimates such as (9.19) yield results sharper than those of Example 6.8. They also prove (see

Example 14.1) that the expected lengths of the longest run of 0's in a random sequence of length n is $\log_2 n + u(\log_2 n) + o(1)$ as $n \to \infty$, where $u(x)$ is a continuous function that is not constant and satisfies $u(x+1) = u(x)$. (See also the discussion of carry propagation in [236].) For other methods and results in this area, see [18]. ■

Inhomogeneous recurrences with constant coefficients, say,

$$a_n = \sum_{i=1}^{d} c_i a_{n-i} + b_n, \quad n \geq d \ , \tag{9.20}$$

are not covered by the techniques discussed above. One can still use the basic generating function approach to derive the ordinary generating function of a_n, but this time it is in terms of the ordinary generating function of b_n. If b_n does not grow too rapidly, the "subtraction of singularities" method of Section 10.2 can be used to derive the asymptotics of a_n in a form similar to that given by (9.26).

9.2. Linear recurrences with varying coefficients

Linear recurrences with constant coefficients have a nice and complete theory. That is no longer the case when one allows coefficients that vary with the index. This is not a fault of mathematicians in not working hard enough to derive elegant results, but reflects the much more complicated behavior that can occur. The simplest case is when the recurrence has a finite number of terms, and the coefficients are polynomials in n.

Example 9.3. *Two-sided generalized Fibonacci sequences.* Let t_n be the number of integer sequences $(b_j, \ldots, b_2, b_1, 1, 1, a_1, a_2, \ldots, a_k)$ with $j + k + 2 = n$ in which each b_i is the sum of one or more contiguous terms immediately to its right, and each a_i is likewise the sum of one or more contiguous terms immediately to its left. It was shown in [120] that $t_1 = t_2 = 1$ and that

$$t_{n+1} = 2nt_n - (n-1)^2 t_{n-1} \quad \text{for} \quad n \geq 2 \ . \tag{9.21}$$

If we let

$$t(z) = \sum_{n=1}^{\infty} \frac{t_n z^{n-1}}{(n-1)!} \tag{9.22}$$

be a modified exponential generating function, then the recurrence (9.21) shows that

$$t'(z)(1-z)^2 - t(z)(2-z) = 1 \ . \tag{9.23}$$

Standard methods for solving ordinary differential equations, together with the initial conditions $t_1 = t_2 = 1$, then yield the explicit solution

$$t(z) = (1-z)^{-1} \exp((1-z)^{-1}) \left[C + \int_z^1 (1-w)^{-1} \exp(-(1-w)^{-1}) dw \right] , \tag{9.24}$$

where

$$C = e^{-1} - \int_0^1 (1-w)^{-1} \exp(-(1-w)^{-1}) dw = 0.148495\ldots \ . \tag{9.25}$$

Once the explicit formula (9.24) for $t(z)$ is obtained, the methods of Section 12 give the estimate

$$t_n \sim C(n-1)!(e/\pi)^{1/2} \exp(2n^{1/2})(2n^{1/4})^{-1} \quad \text{as} \quad n \to \infty \ . \tag{9.26}$$

It is easy to show that the absolute value of

$$(1-z)^{-1} \exp((1-z)^{-1}) \int_z^1 (1-w)^{-1} \exp(-(1-w)^{-1}) dw \tag{9.27}$$

is small for $|z| < 1$. Therefore the asymptotics of the t_n are determined by the behavior of coefficients of

$$C(1-z)^{-1} \exp((1-z)^{-1}) \ , \tag{9.28}$$

and that can be obtained easily. The estimate (9.26) then follows. ■

To see just how different the behavior of linear recurrences with polynomial coefficients can be from those with constant coefficients, compare the behavior of the sequences in Example 9.3 above and Example 9.4 (given below). The existence of such differences should not be too surprising, since after all even the first order recurrence $a_n = na_{n-1}$ for $n \geq 2$, $a_1 = 1$, has the obvious solution $a_n = n!$, which is not at all like the solutions to constant coefficient recurrences. However, when $a_n = na_{n-1}$, a simple change of variables, namely $a_n = b_n n!$, transforms this recurrence into the trivial one of $b_n = b_{n-1} = \cdots = b_1 = 1$ for all n. Such rescaling is among the most fruitful methods for dealing with nonlinear recurrences, even though it is seldom as simple as for $a_n = n!$.

Example 9.3 is typical in that a sequence satisfying a linear recurrence of the form

$$a_n = \sum_{j=1}^{r} c_j(n) a_{n-j} \ , \qquad n \geq r \ , \tag{9.29}$$

where r is fixed and the $c_j(n)$ are rational functions (a P-recursive sequence in the notation of Section 6.3) can always be transformed into a differential equation for a generating function. Whether anything can be done with that generating function depends strongly on the recurrence and the form of the generating function. Example 9.3 is atypical in that there is an explicit solution to the differential equation. Further, this explicit solution is a nice analytic function. This is due to the special choice of the form of the generating function. An exponential generating function seems natural to use in that example, since the recurrence (9.21) shows immediately that $t_n \leq (2n-2)(2n-4)\ldots 2 = 2^{n-1}(n-1)!$, and a slightly more involved induction proves that t_n grows at least as fast as a factorial. If we tried to use an ordinary generating function

$$u(z) = \sum_{n=1}^{\infty} t_n z^n \ , \tag{9.30}$$

then the recurrence (9.21) would yield the differential equation

$$z^4 u''(z) + z^3 u'(z) + (1 - 2z^2)u(z) = z - z^2 \ , \tag{9.31}$$

which is not as tractable. (This was to be expected, since $u(z)$ is only a formal power series.) Even when a good choice of generating function does yield an analytic function, the differential equation that results may be hard to solve. (One can always find a generating function that is analytic, but the structure of the problem may not be reflected in the resulting differential equation, and there may not be anything nice about it.)

There is an extensive literature on analytic solutions of differential equations (cf. [205, 206, 207, 272, 368, 372]), but it is not easy to apply in general. Singularities of analytic functions that satisfy linear differential equations with analytic coefficients are usually of only a few basic forms, and so the methods of Sections 11 and 12 suffice to determine the asymptotic behavior of the coefficients. The difficulty is in locating the singularities and determining their nature. We refer to [206, 207, 272, 368, 372] for methods for dealing with this difficulty, since they are involved and so far have been seldom used in combinatorial enumeration. There will be some further discussion of differential equations in Section 15.3.

Some aspects of the theory of linear recurrences with constant coefficients do carry over to the case of varying coefficients, even when the coefficients are not rational functions. For example, there will in general be r linearly independent solutions to the recurrence (9.29) (corresponding to the different starting conditions). Also, if a solution a_n has the property that a_{n+1}/a_n tends to a limit α as $n \to \infty$, then $1/\alpha$ is a limit of zeros of

$$1 - \sum_{j=1}^{r} c_j(n) z^j \ , \tag{9.32}$$

and therefore is often a root of

$$1 - \sum_{j=1}^{r} \left(\lim_{n\to\infty} c_j(n) \right) z^j \ . \tag{9.33}$$

Whether there are exactly r linearly independent solutions is a difficult problem. Extensive research was done on this topic 1920's and 1930's [2, 29], culminating in the work of Birkhoff and Trjitzinsky [51, 52, 53, 366, 367]. This work applies to recurrences of the form (9.29) where the $c_j(n)$ have Poincaré asymptotic expansions

$$c_j(n) \sim n^{k_j/k}\{c_{j,0} + c_{j,1}n^{-1/k} + c_{j,2}n^{-2/k} + \cdots\} \quad \text{as} \quad n \to \infty \ , \tag{9.34}$$

where the k_j and k are integers and $c_{j,0} \neq 0$ if $c_j(n)$ is not identically 0 for all n. It follows from this work that solutions to the recurrence are expressible as linear combinations of elements of the form

$$(n!)^{p/q} \exp(P(n^{1/m}))n^{\alpha}(\log n)^h \ , \tag{9.35}$$

where h, m, p, and q are integers, $P(z)$ a polynomial, and α a complex number. An exposition of this theory and how it applies to enumeration has been given by Wimp and Zeilberger [384]. (There is a slight complication in that most of the literature cited above is concerned with recurrences for functions of a real argument, not sequences, but this is not a major difficulty.) There is still a problem in identifying which linear combination provides the derived solution. Wimp and Zeilberger point out that it is usually easy to show that the largest of the terms of the form (9.35) does show up with a nonzero coefficient, and so determines the asymptotics of a_n up to a multiplicative constant. However, the Birkhoff-Trjitzinsky method does not in general provide any techniques for determining that constant.

The major objection to the use of the Birkhoff-Trjitzinsky results is that they may not be rigorous, since gaps are alleged to exist in the complicated proofs [211, 383]. Furthermore, in almost all combinatorial enumeration applications the coefficients are rational, and so one can use the theory of analytic differential equations.

When there is no way to avoid linear recurrences with coefficients that vary but are not rational, one can sometimes use the work of Kooman [243, 244], which develops the theory of second order linear recurrences with almost-constant coefficients.

Example 9.4. *An oscillating sequence.* Let

$$a_n = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} \ , \qquad n = 0, 1, \ldots \ . \tag{9.36}$$

Then a_n satisfies the linear recurrence

$$a_{n+2} - \left(2 - \frac{2}{n}\right) a_{n+1} + \left(1 - \frac{1}{n}\right) a_n = 0, \quad n \geq 0 \ . \tag{9.37}$$

The methods of [244] can be used to show that for some constants c and ϕ

$$a_n = cn^{-1/4} \sin(2n^{1/2} + \phi) + o(n^{-1/4}) \quad \text{as} \quad n \to \infty \ , \tag{9.38}$$

which is a much more precise estimate than the crude one mentioned in Example 10.1.

Another, in some ways preferable method for obtaining asymptotic expansions for a_n is mentioned in Example 12.8. It is based on an explicit form for the generating function of a_n, $f(z) = \sum a_n z^n$. An interchange of orders of summation (easily justified for $|z|$ small, say $|z| < 1/2$) shows that

$$\begin{aligned} f(z) &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \sum_{n=k}^{\infty} \binom{n}{k} z^n \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} \frac{z^k}{(1-z)^{k+1}} = \frac{1}{1-z} \exp\left(-\frac{z}{1-z}\right) \ . \end{aligned} \tag{9.39}$$

The saddle point method can then be applied to obtain asymptotic expansions for a_n. ■

9.3. Linear recurrences in several variables

Linear recurrences in several variables that have constant coefficients can be attacked by methods similar to those used in a single variable. If we have

$$a_{m,n} = \sum_{\substack{i=0 \\ i+j>0}}^{d} \sum_{i=0}^{d} c_{i,j} a_{m-i,n-j} \tag{9.40}$$

for $m, n \geq d$, say, then the generating function

$$f(x,y) = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{m,n} x^m y^n \tag{9.41}$$

satisfies the relation

$$f(x,y)\left(1-\sum_{\substack{i=0\\i+j>0}}^{d}\sum_{i=0}^{d}c_{i,j}x^iy^j\right) = \sum_{\substack{m=0\\m>d}}^{\infty}\sum_{\substack{n=0\\\text{or }n>d}}^{\infty}a_{m,n}x^my^n \tag{9.42}$$

$$-\sum_{\substack{i=0\\i+j>0}}^{d}\sum_{i=0}^{d}c_{ij}x^iy^j\sum_{\substack{m,n\\m\le d-i\\\text{or }n\le d-i}}a_{m,n}x^my^n\ .$$

If $a_{m,n}=0$ for $0\le m<d$ and $n\ge d$ as well as for $0\le n<d$ and $m\ge d$ (so that all the $a_{m,n}$ are fully determined by $a_{m,n}$ for $0\le m<d$, $0\le n<d$), then $f(x,y)$ is a rational function. If this condition does not hold, $f(x,y)$ can be complicated.

The paragraph above shows that under common conditions, constant coefficient recurrences lead to generating functions that are rational even in several variables. However, even when the rational function is determined, there is no equivalent of partial fraction decomposition to yield elegant asymptotics of the coefficients. Coefficients of multivariate generating functions are much harder to handle than those of univariate functions. There are tools (discussed in Section 13), that are usually adequate to handle rational generating functions, but they are not simple.

When the coefficients of the multivariate recurrences vary, available knowledge is extremely limited. Even if the coefficients are polynomials, we obtain a partial differential equation for the generating function. Sometimes there are tricks that lead to a simple solution (cf. Example 15.6), but this is not common.

9.4. Nonlinear recurrences

Nonlinear recurrences come in a great variety of shapes, and the methods that are used to solve them are diverse, depending on the nature of the problem. This section presents a sample of the most useful techniques that have been developed.

Sometimes a nonlinear recurrence has a simple solution because of a nice algebraic factorization. For example, suppose that z_0 is any given complex number, and

$$z_{n+1}=z_n^2-2 \quad \text{for} \quad n\ge 0\ . \tag{9.43}$$

If we set

$$w=(z_0+(z_0^2-4)^{1/2})/2\ , \tag{9.44}$$

we have $z_0 = w + w^{-1}$, and more generally

$$z_n = w^{2^n} + w^{-2^n} \quad \text{for} \quad n \geq 0 \; . \tag{9.45}$$

Eq. (9.45) is easily established through induction. However, this is an exceptional instance, and already recurrences of the type $z_{n+1} = z_n^2 + c$ for c a complex constant lead to deep questions about the Mandelbrot set and chaotic behavior [91].

Since linear recurrences are well understood, the best that one can hope for when confronted with a nonlinear recurrence is that it might be reducible to a linear one. This works in many situations.

Example 9.5. *Planted plane trees.* Let $a_{n,h}$ be the number of planted plane trees with n nodes and height $\leq h$ [64, 177], and let

$$A_h(z) = \sum_{n=0}^{\infty} a_{n,h} z^n \; . \tag{9.46}$$

Since a tree of height $\leq h+1$ has a root and any number of subtrees, each of height $\leq h$,

$$\begin{aligned} A_{h+1}(z) &= z(1 + A_h(z) + A_h(z)^2 + \cdots) \\ &= z(1 - A_h(z))^{-1} \; . \end{aligned} \tag{9.47}$$

Iterating this recurrence, we obtain a finite continued fraction that looks like

$$A_{h+1}(z) = \frac{z}{1 - \frac{z}{1 - \frac{z}{1 \ldots}}} \; . \tag{9.48}$$

The general theory of continued functions represents a convergent as a quotient of two sequences satisfying recurrences involving the partial quotients. (For references, see [218, 319].) After playing with this idea, one finds that the substitution

$$A_h(z) = \frac{z P_h(z)}{P_{h+1}(z)} \tag{9.49}$$

gives

$$P_{h+1}(z) = P_h(z) - z P_{h-1}(z) \; , \qquad h \geq 2 \; ,$$

where $P_0(z) = 0$, $P_1(z) = 1$. This is a linear recurrence when we regard z as fixed, and so the theory presented before leads to the explicit representation

$$P_h(z) = (1 - 4z)^{-1/2} \left\{ \left(\frac{1 + (1 - 4z)^{1/2}}{2} \right)^h - \left(\frac{1 - (1 - 4z)^{1/2}}{2} \right)^h \right\} \; . \tag{9.50}$$

De Bruijn, Knuth, and Rice [64] use this representation to determine the average height of plane trees. ■

Greene and Knuth (p. 30 of [177]) note that the continued fraction method of replacing a convergent by a quotient of elements of two sequences in general leads not to a single sequence of polynomials like the $P_h(z)$ of Example 9.5, but to two sequences. This is only slightly harder to handle, and allows one to linearize more complicated recurrences.

There are many additional ways to linearize a recurrence. (A small list is given on p. 31 of [177].) For example, a purely multiplicative relation $a_n = a_{n-1}^2/a_{n-2}$ is transformed into the linear $\log a_n = 2\log a_{n-1} - \log a_{n-2}$ by taking logarithms. One of the most fruitful tricks of this type is taking inverses. Thus $a_n = a_{n-1}/(1+a_{n-1})$ is equivalent to

$$\frac{1}{a_n} = \frac{1}{a_{n-1}} + 1 \ , \tag{9.51}$$

which has the obvious solution $a_n^{-1} = a_0^{-1} + n$. (This assumes $a_0 \neq -1/k$ for any $k \in \mathbb{Z}^+$.)

Linearization works well, but is limited in applicability. More widely applicable, but producing answers that are not as clear, is approximate linearization, where a given nonlinear recurrence is close to a linear one. The following example combines approximate linearization with bootstrapping.

Example 9.6. *A quadratic recurrence.* The study of the average height of binary trees in [132] involves the recurrence

$$a_n = a_{n-1}(1 - a_{n-1}) \quad \text{for} \quad n \geq 1 \ , \tag{9.52}$$

with $a_0 = 1/2$. The a_n are monotone decreasing, so we try the inverse trick. We find

$$\frac{1}{a_n} = \frac{1}{a_{n-1}(1-a_{n-1})} = \frac{1}{a_{n-1}} + 1 + \frac{a_{n-1}}{1-a_{n-1}} \ . \tag{9.53}$$

Iterating this recurrence (but applying it only to the first term on the right-hand side of Eq. (9.53)) we obtain

$$\begin{aligned} \frac{1}{a_n} &= \frac{1}{a_{n-2}} + 2 + \frac{a_{n-2}}{1-a_{n-2}} + \frac{a_{n-1}}{1-a_{n-1}} \\ &= \cdots \\ &= \frac{1}{a_0} + n + \sum_{j=0}^{n-1} \frac{a_j}{1-a_j} \\ &= n + 2 + \sum_{j=0}^{n-1} \frac{a_j}{1-a_j} \ . \end{aligned} \tag{9.54}$$

Equation (9.54) shows that $a_n^{-1} > n$, so $a_n < 1/n$. Applying this bound to a_j for $2 \le j \le n-1$ in the sum on the right-hand side of Eq. (9.54), we find that

$$n \le a_n^{-1} \le n + O(\log n) \ . \tag{9.55}$$

When we substitute this into (9.54), we find that $a_n^{-1} = n + \log n + o(\log n)$, and further iterations produce even more accurate estimates. ■

Approximate linearization also works well for some rapidly growing sequences.

Example 9.7. *Doubly exponential sequences.* Many recurrences are of the form

$$a_{n+1} = a_n^2 + b_n \ , \tag{9.56}$$

where b_n is much smaller than a_n^2 (and may even depend on the a_n for $k \le n$, as in $b_n = a_n$ or $b_n = a_{n-1}$). Aho and Sloane [3] found that surprisingly simple solutions to such recurrences can often be found. The basic idea is to reduce to approximate linearization by taking logarithms. We find that if a_0 is the given initial value, and $a_n > 0$ for all n, then the transformation

$$u_n = \log a_n \ , \tag{9.57}$$

$$\delta_n = \log(1 + b_n a_n^{-2}) \ , \tag{9.58}$$

reduces (9.56) to

$$u_{n+1} = 2u_n + \delta_n \ , \qquad n \ge 0 \ . \tag{9.59}$$

Therefore

$$\begin{aligned} u_n &= \delta_{n-1} + 2u_{n-1} = \delta_{n-1} + 2\delta_{n-2} + 4u_{n-2} \\ &= \dots \\ &= \sum_{j=1}^{n} 2^{j-1}\delta_{n-j} + 2^n u_0 \\ &= 2^n (u_0 + \delta_0/2 + \delta_1/4 + \cdots + \delta_{n-1}/2^n) \ . \end{aligned} \tag{9.60}$$

If we assume that the δ_k are small, then

$$\alpha = u_0 + \sum_{k=0}^{\infty} \delta_k 2^{-k-1} \tag{9.61}$$

exists, and

$$r_n = u_n - 2^n \alpha = 2^n \sum_{k=n}^{\infty} \delta_k 2^{-k-1} \ . \tag{9.62}$$

If the δ_k are sufficiently small, the difference r_n in (9.62) will be small, and

$$a_n = \exp(u_n) = \exp(2^n \alpha - r_n) \ . \tag{9.63}$$

The expression (9.63) might not seem satisfactory, since both a_n and r_n are expressed in terms of all the a_k, for $k < n$ and for $k \geq n$. The point of (9.63) is that for many recurrences, r_n is negligibly small, while α is given by the rapidly convergent series (9.61), so that only the first few a_n are needed to obtain a good estimate for the asymptotic behavior of a_n. We next discuss a particularly elegant case.

Suppose that $a_n \geq 1$ and $|b_n| < a_n/4$ for all $n \geq 0$. Then $a_{n+1} \geq a_n$ and $|\delta_{n+1}| \leq |\delta_n|$ for $n \geq 0$, and so $|r_n| \leq |\delta_n|$. Hence

$$a_n \exp(-|\delta_n|) \leq \exp(2^n \alpha) \leq a_n \exp(|\delta_n|) \tag{9.64}$$

and since

$$\begin{aligned} \exp(|\delta_n|) \quad &\leq \quad 1 + |b_n| a_n^{-2} \ < \ 1 + (4a_n)^{-1} \ , \\ \exp(-|\delta_n|) \quad &\geq \quad (1 + (4a_n)^{-1})^{-1} \ \geq \ 1 - (3a_n)^{-1} \ , \end{aligned} \tag{9.65}$$

we find that

$$|a_n - \exp(2^n \alpha)| < (2a_n)^{-1} \leq 1/2 \ . \tag{9.66}$$

If a_n is an integer, then we can assert that it is the closest integer to $\exp(2^n \alpha)$.

The restriction $|b_n| < a_n/4$ is severe. The basic method applies even without it, and the expansion (9.63) is valid, for example, if we only require that $|\delta_{n+1}| \leq |\delta_n|$ for $n \geq n_0$. However, we will not in general obtain results as nice as (9.66) if we only impose these weak conditions.

The method outlined above can be applied to recurrences that appear to be of a slightly different form. Sometimes only a trivial transformation is required. For example, Golomb's nonlinear recurrence,

$$a_{n+1} = a_0 a_1 \cdots a_n + b, \quad a_0 = 1 \ , \tag{9.67}$$

for b a constant, is easily seen to be equivalent to

$$a_{n+1} = (a_n - b) a_n + b, \quad a_0 = 1, \quad a_1 = b + 1 \ . \tag{9.68}$$

The substitution

$$x_n = a_n - b/2 \tag{9.69}$$

transforms (9.68) into

$$x_{n+1} = x_n^2 + (2 - b) b/4 \ , \tag{9.70}$$

which is of the form treated above. (If the x_n are integers, the inequality (9.66) with x_n replacing a_n might not apply to the x_n because the condition $|(2-b)b/4| < |x_k|/4$ might fail for some k. The trick to use here is to start the recurrence with some x_k, say x_{k_0}, so that the condition $|(2-b)b/4| < |x_k|/4$ applies for $k \geq k_0$. The new α for which (9.66) holds will then be defined in terms of $x_{k_0}, x_{k_0+1}, \ldots$.)

In some situations the results presented above cannot be applied, but the basic method can still be extended. That is the case for the recurrence

$$a_{n+1} = a_n a_{n-1} + 1, \quad a_0,\ a_1 \geq 1 \tag{9.71}$$

of [3]. The result is that a_n is the nearest integer to

$$\alpha^{F_n} \beta^{F_{n-1}} , \tag{9.72}$$

where α and β are positive constants, and the F_k are the Fibonacci numbers. What matters is that the recurrence leads to doubly exponential (and regular) growth of a_n. Example 15.3 shows how this principle can be applied even when the a_n are not numbers, but polynomials whose coefficients need to be estimated. ■

9.5. Quasi-linear recurrences

This section mentions some methods and results for studying recurrences that have linearity properties, but are not linear. The most important of them are recurrences involving minimization or maximization. They arise frequently in problems that use dynamic programming approaches and in divide and conquer methods. An important example, treated in [147], is that of a sequence f_n, given by $f_0 = 1$ and

$$f_{n+1} = g_{n+1} + \min_{0 \leq k \leq n} (\alpha f_k + \beta f_{n-k}) \quad \text{for} \quad n \geq 0 , \tag{9.73}$$

where $\alpha, \beta > 0$, and g_n is some given sequence. Fredman and Knuth showed that if $g_n = 0$ for $n \geq 1$ and $\alpha + \beta < 1$, then

$$f_n \geq c n^{1+1/\gamma} \quad \text{for some} \quad c = c(\alpha, \beta) > 0 , \tag{9.74}$$

where γ is the solution to

$$\alpha^{-\gamma} + \beta^{-\gamma} = 1 . \tag{9.75}$$

They proved that $\lim_{n\to\infty} f_n n^{-1-1/\gamma}$ exists if and only if $(\log\alpha)/(\log\beta)$ is irrational. They also presented analyses of this recurrence for $\alpha+\beta \geq 1$, as well as of several recurrences that have different g_n.

The value of the Fredman-Knuth paper is less in the precise results they obtain for several recurrences of the type (9.73) than in the methods they develop, which allow one to analyze related problems. A crucial role in their approach is played by the observation that for the g_n they consider, the minimum in (9.73) can be located rather precisely. The conditions for such localization are applicable to many other sequences as well.

Further work on the recurrence (9.73) was done by Kapoor and Reingold [220], who obtained a complete solution under certain conditions. Their solution is complicated, expressed in terms of the weighted external path length of a binary tree. It is sufficiently explicit, though, to give a complete picture of the continuity, convexity, and oscillation properties of f_n. In some cases their solution simplifies dramatically.

Another class of quasi-linear recurrences involves the greatest integer function. Following [104], consider recurrences of the form

$$a(0) = 1, \quad a(n) = \sum_{i=1}^{s} r_i a(\lfloor n/m_i \rfloor), \quad n \geq 1, \tag{9.76}$$

where $r_i > 0$ for all i, and the m_i are integers, $m_i \geq 2$ for all i. Let $\tau > 0$ be the (unique) solution to

$$\sum_{i=1}^{s} r_i m_i^{-\tau} = 1 \ . \tag{9.77}$$

If there is an integer d and integers u_i such that $m_i = d^{u_i}$ for $1 \leq i \leq s$, then $\lim\ a(n)n^{-\tau}$ as $n \to \infty$ does not exist, but the limit of $a(d^k)d^{-k\tau}$ as $k \to \infty$ does exist. If there is no such d, then the limit of $a(n)n^{-\tau}$ as $n \to \infty$ does exist, and can readily be computed. For example, when

$$a(n) = a(\lfloor n/2 \rfloor) + a(\lfloor n/3 \rfloor) + a(\lfloor n/6 \rfloor) \quad \text{for} \quad n \geq 1 \ ,$$

this limit is $12(\log 432)^{-1}$. Convergence to the limit is extremely slow, as is shown in [104]. The method of proof used in [104] is based on renewal theory. Several other methods for dealing with recurrences of the type (9.76) are mentioned in [104] and the references listed in that paper. There are connections to other recurrences that are linear in two variables, such as

$$b(m,n) = b(m,n-1) + b(m-1,n) + b(m-1,n-1), \quad m,n \geq 1 \ . \tag{9.78}$$

Consider an infinite sequence of integers $2 \le a_1 < a_2 < \ldots$ such that

$$\sum_{j=1}^{\infty} a_j^{-1} \log a_j < \infty ,$$

and define $c(0) = 0$,

$$c(n) = \sum_{j=1}^{\infty} c(\lfloor n/a_j \rfloor) + 1, \qquad n \ge 1 . \tag{9.79}$$

If ρ is the (unique) positive solution to

$$\sum_{j=1}^{\infty} a_j^{-\rho} = 1 ,$$

then Erdös [103] showed that

$$c(n) \sim cn^{\rho} \quad \text{as} \quad n \to \infty \tag{9.80}$$

for a positive constant c. Although the recurrence (9.79) is similar to that of Eq. (9.76), the results are different (no oscillations can occur for a recurrence given by Eq. (9.79)) and the methods are dissimilar.

Karp [221] considers recurrences of the type $T(x) = a(x) + T(h(x))$, where x is a nonnegative real variable, $a(x) \ge 0$, and $h(x)$ is a random variable, $0 \le h(x) \le x$, with $m(x)$ being the expectation of $h(x)$. Such recurrences arise frequently in the analysis of algorithms, and Karp proves several theorems that bound the probability that $T(x)$ is large. For example, he obtains the following result.

Theorem 9.2. *Suppose that $a(x)$ is a nondecreasing continuous function that is strictly increasing on $\{x : a(x) > 0\}$, and $m(x)$ is a continuous function. Then for all $x \in \mathbb{R}^+$ and $k \in \mathbb{Z}^+$,*

$$\text{Prob}\,(T(x) \ge u(x) + ka(x)) \le (m(x)/x)^k ,$$

where $u(x)$ is the unique least nonnegative solution to the equation $u(x) = a(x) + u(m(x))$.

Another result, proved in [176], is the following estimate.

Theorem 9.3. *Suppose that $r, a_1, \ldots, a_N \in \mathbb{R}^+$ and that $b \ge 0$. For $n > N$, define*

$$a_n = 1 + \max_{1 \le k \le n-1} \frac{b + a_{n-1} + a_{n-2} + \cdots + a_{n-k}}{k + r} . \tag{9.81}$$

Then

$$a_n \sim (n/r)^{1/2} \qquad as \qquad n \to \infty . \tag{9.82}$$

Theorem 9.3 is proved by an involved induction on the behavior of the a_n.

10. Analytic generating functions

Combinatorialists use recurrence, generating functions, and such transformations as the Vandermonde convolution; others, to my horror, use contour integrals, differential equations, and other resources of mathematical analysis.

J. Riordan [336]

The use of analytic methods in combinatorics did horrify Riordan. They are widespread, though, because of their utility, which even Riordan could not deny. About half of this chapter is devoted to such methods, as they are extremely flexible and give very precise estimates.

10.1. Introduction and general estimates

This section serves as an introduction to most of the remaining sections of the paper, which are concerned largely with the use of methods of complex variables in asymptotics. Many of the results to be presented later can be used with little or no knowledge of analytic functions. However, even some slight knowledge of complex analysis is helpful in getting an understanding of the scope and limitations of the methods to be discussed. There are many textbooks on analytic functions, such as [205, 364]. This chapter assumes that the reader has some knowledge of this field, but not a deep one. It reviews the concepts that are most relevant in asymptotic enumeration, and how they affect the estimates that can be obtained. It is not a general introduction to the subject of complex analysis, and the choices of topics, their ordering, and the decision of when to include proofs were all made with the goal of illustrating how to use complex analysis in asymptotics.

There are several definitions of analytic functions, all equivalent. For our purposes, it will be most convenient to call a function $f(z)$ of one complex variable *analytic* in a connected open set $S \subseteq \mathbb{C}$ if in a small neighborhood of every point $w \in S$, $f(z)$ has an expansion as a power series

$$f(z) = \sum_{n=0}^{\infty} a_n (z-w)^n, \quad a_n = a_n(w), \tag{10.1}$$

that converges. Practically all the functions encountered in asymptotic enumeration that are analytic are analytic in a disk about the origin. A necessary and sufficient condition for $f(z)$, defined by a power series (6.1), to be analytic in a neighborhood of the origin is that $|a_n| \leq C^n$ for some constant $C > 0$. Therefore there is an effective dichotomy, with common generating functions either not converging near 0 and being only formal power series, or else converging and being analytic.

A function $f(z)$ is called *meromorphic* in S if it is analytic in S except at a (countable isolated) subset $S' \subseteq S$, and in a small neighborhood of every $w \in S'$, $f(z)$ has an expansion of the form

$$f(z) = \sum_{n=-N(w)}^{\infty} a_n (z-w)^n \ , \qquad a_n = a_n(w) \ . \tag{10.2}$$

Thus meromorphic functions can have poles, but nothing more. Alternatively, a function is meromorphic in S if and only if it is the quotient of two functions analytic in S. In particular, z^{-5} is meromorphic throughout the complex plane, but $\sin(1/z)$ is not. In general, functions given by nice expressions are analytic away from obvious pathological points, since addition, multiplication, division, and composition of analytic functions usually yield analytic or meromorphic functions in the proper domains. Thus $\sin(1/z)$ is analytic throughout $\mathbb{C} \setminus \{0\}$, and so is z^{-5}, while $\exp(1/(1-z))$ is analytic throughout $\mathbb{C} \setminus \{1\}$, but is not meromorphic because of the essential singularity at $z = 1$. Not all functions that might seem smooth are analytic, though, as neither $f(z) = \bar{z}$ ($\bar{z}$ denoting the complex conjugate of z) nor $f(z) = |z|$ is analytic anywhere. The smoothness condition imposed by (10.1) is very stringent.

Analytic continuation is an important concept. A function $f(z)$ may be defined and analytic in S, but there may be another function $g(z)$ that is analytic in $S' \supset S$ and such that $g(z) = f(z)$ for $z \in S$. In that case we say that $g(z)$ provides an analytic continuation of $f(z)$ to S', and it is a theorem that this extension is unique. A simple example is provided by

$$\sum_{n=0}^{\infty} z^n = \frac{1}{1-z} \ . \tag{10.3}$$

The power series on the left side converges only for $|z| < 1$, and defines an analytic function there. On the other hand, $(1-z)^{-1}$ is analytic throughout $\mathbb{C} \setminus \{1\}$, and so provides an analytic continuation for the power series. This is a common phenomenon in asymptotic enumeration. Typically a generating function will converge in a disk $|z| < r$, will have a singularity at r, but will be continuable to a region of the form

$$\{z : |z| < r + \delta, \ |\mathrm{Arg}(z-r)| > \pi/2 - \epsilon\} \tag{10.4}$$

for $\delta, \ \epsilon > 0$. When this happens, it can be exploited to provide better or easier estimates of the coefficients, as is shown in Section 11.1. That section explains the reasons why continuation to a region of the form (10.4) is so useful.

If $f(z)$ is analytic in S, z is on the boundary of S, but $f(z)$ cannot be analytically continued to a neighborhood of z, we say that z is a *singularity* of $f(z)$. Isolated singularities that are not poles are called essential, so that $z = 1$ is an essential singularity of $\exp(1/(1-z))$, but not of $1/(1-z)$. (Note that $z = 1$ is an essential singularity of $f(z) = (1-z)^{1/2}$ even though $f(1) = 0$.) Throughout the rest of this chapter we will often refer to *large singularities* and *small singularities*. These are not precise concepts, and are meant only to indicate how fast the function $f(z)$ grows as $z \to z_0$, where z_0 is a singularity. If $z_0 = 1$, we say that $(1-z)^{1/2}$, $\log(1-z)$, $(1-z)^{-10}$ have small singularities, since $|f(z)|$ either decreases or grows at most like a negative power of $|1-z|$ as $z \to 1$. On the other hand, $\exp(1/(1-z))$ or $\exp((1-z)^{-1/5})$ will be said to have large singularities. Note that for $z = 1+iy$, $y \in \mathbb{R}$, $\exp(1/(1-z))$ is bounded, so the choice of path along which the singularity is approached is important. In determining the size of a singularity z_0, we will usually be concerned with real z_0 and generating functions $f(z)$ with nonnegative coefficients, and then usually will need to look only at z real, $z \to z_0^-$. When the function $f(z)$ is *entire* (that is, analytic throughout $\mathbb{C}$), we will say that ∞ is a singularity of $f(z)$ (unless $f(z)$ is a constant), and will use the large vs. small singularity classification depending on how fast $f(z)$ grows as $|z| \to \infty$. The distinction between small and large singularities is important in asymptotics because different methods are used in the two cases.

A simple closed contour Γ in the complex plane is given by a continuous mapping $\gamma : [0,1] \to \mathbb{C}$ with the properties that $\gamma(0) = \gamma(1)$, and that $\gamma(s) \neq \gamma(t)$ whenever $0 \leq s < t \leq 1$ and either $s \neq 0$ or $t \neq 1$. Intuitively, Γ is a closed path in the complex plane that does not intersect itself. For most applications that will be made in this chapter, simple closed contours Γ will consist of line segments and sections of circles. For such contours it is easy to prove that the complex plane is divided by the contour into two connected components, the inside and the outside of the curve. This result is true for all simple closed curves by the Jordan curve theorem, but this result is surprisingly hard to prove.

In asymptotic enumeration, the basic result about analytic functions is the Cauchy integral formula for their coefficients.

Theorem 10.1. *If $f(z)$ is analytic in an open set S containing* 0*, and*

$$f(z) = \sum_{n=0}^{\infty} a_n z^n \tag{10.5}$$

in a neighborhood of 0*, then for any* $n \geq 0$*,*

$$a_n = [z^n]f(z) = (2\pi i)^{-1} \int_\Gamma f(z) z^{-n-1} dz \ , \tag{10.6}$$

where Γ *is any simple closed contour in* S *that contains the origin inside it and is positively oriented (i.e., traversed in counterclockwise direction).*

An obvious question is why should one use the integral formula (10.6) to determine the coefficient a_n of $f(z)$. After all, the series (10.5) shows that

$$n!\, a_n = \frac{d^n}{dz^n} f(z) \Big|_{z=0} \ . \tag{10.7}$$

Unfortunately the differentiation involved in (10.7) is hard to control. Derivatives involve taking limits, and so even small changes in a function can produce huge changes in derivatives, especially high order ones. The special properties of analytic functions are not reflected in the formula (10.7), and for nonanalytic functions there is little that can be done. On the other hand, Cauchy's integral formula (10.6) does use special properties of analytic functions, which allow the determination of the coefficients of $f(z)$ from the values of $f(z)$ along any closed path. This determination involves integration, so that even coarse information about the size of $f(z)$ can be used with it. The analytic methods that will be outlined exploit the freedom of choice of the contour of integration to relate the behavior of the coefficients to the behavior of the function near just one or sometimes a few points.

If the power series (10.5) converges for $|z| < R$, and for the contour Γ we choose a circle $z = r\exp(i\theta)$, $0 \leq \theta \leq 2\pi$, $0 < r < R$, then the validity of (10.6) is easily checked by direct computation, since the power series converges absolutely and uniformly so one can interchange integration and summation. The strength of Cauchy's formula is in the freedom to choose the contour Γ in different ways. This freedom yields most of the powerful results to be discussed in the following sections, and later in this section we will outline how this is achieved. First we discuss some simple applications of Theorem 10.1 obtained by choosing Γ to be a circle centered at the origin.

Theorem 10.2. *If* $f(z)$ *is analytic in* $|z| < R$*, then for any* r *with* $0 < r < R$ *and any* $n \in Z$*,* $n \geq 0$*,*

$$|[z^n]f(z)| \leq r^{-n} \max_{|z|=r} |f(z)| \ . \tag{10.8}$$

The choice of Γ in Theorem 10.1 to be the circle of radius r gives Theorem 10.2. If $f(z)$, defined by (10.5), has $a_n \geq 0$ for all n, then

$$|f(z)| \leq \sum_{n=0}^{\infty} a_n |z|^n = f(|z|)$$

and therefore we obtain Lemma 8.1 as an easy corollary to Theorem 10.2. The advantage of Theorem 10.2 over Lemma 8.1 is that there is no requirement that $a_n \geq 0$. The bound of Theorem 10.2 is usually weaker than the correct value by a small multiplicative factor such as $n^{1/2}$.

If $f(z)$ is analytic in $|z| < R$, then for any $\delta > 0$, $f(z)$ is bounded in $|z| < R - \delta$, and so Theorem 10.2 shows that $a_n = [z^n]f(z)$ satisfies $|a_n| = O((R-\delta)^{-n})$. On the other hand, if $|a_n| = O(S^{-n})$, then the power series (10.5) converges for $|z| < S$ and defines an analytic function in that disk. Thus we obtain the easy result that if $f(z)$ is analytic in a disk $|z| < R$ but in no larger disk, then

$$\limsup |a_n|^{1/n} = R^{-1} \ . \tag{10.9}$$

Example 10.1. *Oscillating sequence.* Consider the sequence, discussed already in Example 9.4, given by

$$a_n = \sum_{k=0}^{n} \binom{n}{k} \frac{(-1)^k}{k!} \ , \qquad n = 0, 1, \dots \ . \tag{10.10}$$

The maximal term in the sum (10.10) is of order roughly $\exp(cn^{1/2})$, so a_n cannot be much larger. However, the sum (10.10) does not show that a_n cannot be extremely small. Could we have $|a_n| \leq \exp(-n)$ for all n, say? That this is impossible is obvious from (9.39), though, by the argument above. The generating function $f(z)$, given by Eq. (9.39), is analytic in $|z| < 1$, but has an essential singularity at $z = 1$, so we immediately see that for any $\epsilon > 0$, $|a_n| < (1+\epsilon)^n$ for all sufficiently large n, and that $|a_n| > (1-\epsilon)^n$ for infinitely many n. (More powerful methods for dealing with analytic generating functions, such as the saddle point method to be discussed in Section 12, can be used to obtain the asymptotic relation for a_n given in Example 9.4.) ■

There is substantial literature dealing with the growth rate of coefficients of analytic functions. The book of Evgrafov [110] is a good reference for these results. However, the estimates presented there are not too useful for us, since they apply to wide classes of often pathological functions. In combinatorial enumeration we usually encounter much tamer generating functions for which the crude bounds of [110] are obvious or easy to derive. Instead, we need to use the tractable nature of the functions we encounter to obtain much more delicate estimates.

The basic result, derived earlier, is that the power series coefficients a_n of a generating function $f(z)$, defined by (10.5), grow in absolute value roughly like R^{-n}, if $f(z)$ is analytic in $|z| < R$. A basic result about analytic functions says that if the Taylor series (10.5) of $f(z)$ converges for $|z| < R$ but for every $\epsilon > 0$ there is a z with $|z| = R + \epsilon$ such that the series (10.5) diverges at z, then $f(z)$ has a singularity z with $|z| = R$. Thus the exponential growth rate of the a_n is determined by the distance from the origin of the nearest singularity of $f(z)$, with close singularities giving large coefficients. Sometimes it is not obvious what R is. When the coefficients of $f(z)$ are positive, as is common in combinatorial enumeration and analysis of algorithms, there is a useful theorem of Pringsheim [364]:

Theorem 10.3. *Suppose that $f(z)$ is defined by Eq. (10.5) with $a_n \geq 0$ for all $n \geq n_0$, and that the series (10.5) for $f(z)$ converges for $|z| < R$ but not for any $|z| > R$. Then $z = R$ is a singularity of $f(z)$.*

As we remarked above, the exponential growth rate of the a_n is determined by the distance from the origin of the nearest singularity. Theorem 10.3 says that if the coefficients a_n are non-negative, it suffices to look along the positive real axis to determine the radius of convergence R, which is also the desired distance to the singularity. There can be other singularities at the same distance from the origin (for example, $f(z) = (1 - z^2)^{-1}$ has singularities at $z = \pm 1$), but Theorem 10.3 guarantees that none are closer to 0 than the positive real one.

Since the singularities of smallest absolute value of a generating function exert the dominant influence on the asymptotics of the corresponding sequence, they are called the *dominant singularities*. In the most common case there is just one dominant singularity, and it is almost always real. However, we will sometimes speak of a large set of singularities (such as the k first order poles in Theorem 9.1, which are at different distances from the origin) as dominant ones. This allows some dominant singularities to be more influential than others.

Many techniques, including the elementary methods of Section 8, obtain bounds for summatory functions of coefficients even when they cannot estimate the individual coefficients. These methods succeed largely because they create a dominant singularity. If $f(z) = \sum f_n z^n$ converges for $|z| < 1$, diverges for $|z| > 1$, and has $f_n \geq 0$, then the singularity at $z = 1$ is at least as large as any other. However, there could be other singularities on $|z| = 1$ that are just as large. (This holds for the functions $f_2(z)$ and $f_3(z)$ defined by (8.2) and (8.4).) When we consider the generating function of $\sum_{k \le n} f_k$, though, we find that

$$h(z) = \sum_{n=0}^{\infty} \left(\sum_{k=0}^{n} f_k \right) z^n = (1-z)^{-1} f(z) \ , \tag{10.11}$$

so that $h(z)$ has a singularity at $z = 1$ that is much larger than any other one. That often provides enough of an extra boost to push through the necessary technical details of the estimates.

Most generating functions $f(z)$ have their coefficients $a_n = [z^n]f(z)$ real. If $f(z)$ is analytic at 0, and has real coefficients, then $f(z)$ satisfies the reflection principle,

$$f(z) = \overline{f(\overline{z})} \ . \tag{10.12}$$

This implies that zeros and singularities of $f(z)$ come in complex conjugate pairs.

The success of analytic methods in asymptotics comes largely from the use of Cauchy's formula (10.6) to estimate accurately the coefficients a_n. At a more basic level, this success comes because the behavior of an analytic function $f(z)$ reflects precisely the behavior of the coefficients a_n. In the discussion of elementary methods in Section 8, we pointed out that the behavior of a generating function for real arguments does not distinguish between functions with different coefficients. For example, the functions $f_1(z)$ and $f_3(z)$ defined by (8.1) and (8.4) are almost indistinguishable for $z \in \mathbb{R}$. However, they differ substantially in their behavior for complex z. The function $f_1(z)$ has only a first order pole at $z = 1$ and no other singularities, while $f_3(z)$ has poles at $z = 1$, $\exp(2\pi i/3)$, and $\exp(4\pi i/3)$. The three poles at the three cubic roots of unity reflect the modulo 3 periodicity of the coefficients of $f_3(z)$. This is a general phenomenon, and in the next section we sketch the general principle that underlies it. (The degree to which coefficients of an analytic function determine the behavior at the singularities is the subject of Abelian theorems. We will not need to delve into this subject to its full depth. For references, see [190, 364].)

Analytic methods are extremely powerful, and when they apply, they often yield estimates of unparalleled precision. However, there are tricky situations where analytic methods seem as if they ought to apply, but don't (at least not easily), whereas simpler approaches work.

Example 10.2. *Set partitions with distinct block sizes.* Let a_n be the number of partitions of a set of n elements into blocks of distinct sizes. Then $a_n = b_n \cdot n!$, where $b_n = [z^n]f(z)$, with

$$f(z) = \prod_{k=1}^{\infty} \left(1 + \frac{z^k}{k!}\right) . \tag{10.13}$$

The function $f(z)$ is entire and has nonnegative coefficients, so it might appear as an ideal candidate for an application of some of the methods for dealing with large singularities (such as the saddle point technique) that will be presented later. However, on circles $|z| = (n + 1/2)/e$, $n \in \mathbb{Z}^+$, $f(z)$ does not vary much, so there are technical problems in applying these analytic methods. On the other hand, combinatorial estimates can be used to show [233] that the b_n behave in a "regularly irregular" way, so that, for example,

$$b_{m(m+1)/2-1} \sim b_{m(m+1)/2} \quad \text{as} \quad m \to \infty , \tag{10.14}$$

$$b_{m(m+1)/2} \sim m b_{m(m+1)/2+1} \quad \text{as} \quad m \to \infty . \tag{10.15}$$

These estimates are obtained by expanding the product in Eq. (10.13) and noting that

$$b_n = \sum_{\substack{1 \le k_1 < \cdots < k_r \\ \sum_{i}^{r} k_i = n}} \frac{1}{\prod_{i=1}^{r} k_i!} . \tag{10.16}$$

Since factorials grow rapidly, the only terms in the sum in (10.16) that are significant are those with small k_i. The term $b_n z^n$ for $n = m(m+1)/2$ for example, comes almost entirely from the product of $z^k/k!$, $1 \le k \le m$, all other products contributing an asymptotically negligible amount. ■

10.2. Subtraction of singularities

An important basic tool in asymptotics of coefficients of analytic functions is that of subtraction of singularities. If we wish to estimate $[z^n]f(z)$, and we know $[z^n]g(z)$, and the singularities of $f(z) - g(z)$ are smaller than those of $f(z)$, then we can usually conclude that $[z^n]f(z) \sim [z^n]g(z)$ as $n \to \infty$. In practice, given a function $f(z)$, we find the dominant singularities of $f(z)$ (usually poles), and construct a simple function $g(z)$ with those singularities. We illustrate this approach with several examples. The basic theme will recur in other sections.

Example 10.3. *Bernoulli numbers.* The Euler-Maclaurin summation formula, introduced in Section 5.3, involves the Bernoulli numbers B_n with exponential generating function

$$f(z) = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} = \frac{z}{e^z - 1} . \tag{10.17}$$

The denominator $\exp(z) - 1$ has zeros at 0, $\pm 2\pi i$, $\pm 4\pi i, \ldots$. The zero at 0 is canceled by the zero of z, so $f(z)$ is analytic for $|z| < 2\pi$, but has first order poles at $z = \pm 2\pi i$, $\pm 4\pi i, \ldots$. Consider

$$g(z) = 2\pi i \left(\frac{1}{z - 2\pi i} - \frac{1}{z + 2\pi i} \right) . \tag{10.18}$$

Then $f(z) - g(z)$ is analytic for $|z| < 4\pi$, so

$$|[z^n](f(z) - g(z))| = O((4\pi - \epsilon)^{-n}) \quad \text{as} \quad n \to \infty \tag{10.19}$$

for every $\epsilon > 0$. On the other hand,

$$[z^n]g(z) = \begin{cases} 0 & n \text{ odd} , \\ 2(2\pi)^{-n} & n \text{ even} . \end{cases} \tag{10.20}$$

This gives the leading term asymptotics of B_n. By taking more complicated $g(z)$, we can subtract more of the singularities of $f(z)$ and obtain more accurate expansions for B_n. It is even possible to obtain an exponentially rapidly convergent series for B_n. ■

Example 10.4. *Rational function asymptotics.* As another example of the subtraction of singularities principle, we sketch a proof of Theorem 9.1. Suppose that the hypotheses of that theorem are satisfied. Let

$$u(z) = \sum_{j=1}^{k} \frac{-g(\rho_j)}{\rho_j h'(\rho_j)(1 - z/\rho_j)} . \tag{10.21}$$

Then $f(z) - u(z)$ has no singularities in $|z| \le R$, and for $|z| = R$,

$$|f(z) - u(z)| \le |f(z)| + |u(z)| \le W + \delta^{-1} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)| . \tag{10.22}$$

Hence, by Theorem 10.2,

$$\Big| [z^n](f(z) - u(z)) \Big| \le W R^{-n} + \delta^{-1} R^{-n} \sum_{j=1}^{k} |g(\rho_j)/h'(\rho_j)| . \tag{10.23}$$

On the other hand,

$$[z^n]u(z) = -\sum_{j=1}^{k} \rho_j^{-n-1} g(\rho_j)/h'(\rho_j) . \tag{10.24}$$

The last two estimates yield Theorem 9.1. ■

The reader may have noticed that the proof of Theorem 9.1 presented above does not depend on $f(z)$ being rational. We have proved the following more general result.

Theorem 10.4. *Suppose that $f(z)$ is meromorphic in an open set containing $|z| \le R$, that it is analytic at $z = 0$ and on $|z| = R$, and that the only poles of $f(z)$ in $|z| < R$ are at $\rho_1, \ldots, \rho_k$, each of multiplicity 1. Suppose further that*

$$\max_{|z|=R} |f(z)| \le W \tag{10.25}$$

and that $R - |\rho_j| \ge \delta$ for some $\delta > 0$ and $1 \le j \le k$. Then

$$\left| [z^n]f(z) + \sum_{j=1}^{k} r_j \rho_j^{-n-1} \right| \le W R^{-n} + \delta^{-1} R^{-n} \sum_{j=1}^{k} |r_j| \ , \tag{10.26}$$

where r_j is the residue of $f(z)$ at ρ_j.

In the examples above, the dominant singularities were separated from other ones, so their contributions were larger than those of lower order terms by an exponential factor. Sometimes the singularity that remains after subtraction of the dominant one is on the same circle, and only slightly smaller. Section 11 presents methods that deal with some cases of this type, at least when the singularity is not large. What makes those methods work is the subtraction of singularities principle. Next we illustrate another application of this principle where the singularity is large. (The generating function is entire, and so the singularity is at infinity.)

Example 10.5. *Permutations without long increasing subsequences.* Let $u_k(n)$ be the number of permutations of $\{1, 2, \ldots, n\}$ that have no increasing subsequence of length $> k$. Logan and Shepp [257] and Vershik and Kerov [370] established by calculus of variations and combinatorics that the average value of the longest increasing subsequence in a random permutation is asymptotic to $2n^{1/2}$. Frieze [149] has proved recently, using probabilistic methods, a stronger result, namely that almost all permutations have longest increasing subsequences of length close to $2n^{1/2}$. Here we consider asymptotics of $u_k(n)$ for k fixed and $n \to \infty$. The Schensted correspondence and the hook formula express $u_k(n)$ in terms of Young diagrams with $\le k$ columns. For k fixed, there are few diagrams and their influence can be estimated explicitly using Stirling's formula, although Selberg-type integrals are involved and the analysis is complicated. This analysis was done by Regev [329], who proved more general results. Here we sketch another approach to the asymptotics of $u_k(n)$ for k fixed. It is based on a result of Gessel [161]. If

$$U_k(z) = \sum_{n=0}^{\infty} \frac{u_k(n) z^{2n}}{(n!)^2} \ , \tag{10.27}$$

then

$$U_k(z) = \det(I_{|i-j|}(2z))_{1 \le i,j \le k} \ , \tag{10.28}$$

where the $I_m(x)$ are Bessel functions (Chapter 9 of [297]). H. Wilf and the author have noted that one can obtain the asymptotics of the $u_k(n)$ by using known asymptotic results about the $I_m(x)$. Eq. (9.7.1) of [297] states that for every $H \in \mathbb{Z}^+$,

$$I_m(z) = (2\pi z)^{-1/2} e^z \left(\sum_{h=0}^{H-1} c(m,h) z^{-h} + O(|z|^{-H}) \right) \ , \tag{10.29}$$

where this expansion is valid for $|z| \to \infty$ with $|\mathrm{Arg}(z)| \le 3\pi/8$, say. The $c(m,h)$ are explicit constants with $c(m,0) = 1$. Let us consider $k = 4$ to be concrete. Then, taking $H = 7$ in (10.29) (higher values of H are needed for larger k) we find from (10.28) that

$$U_4(z) = e^{8z} (3(256\pi^2 z^8)^{-1} + O(|z|^{-9})) \quad \text{for} \quad |z| \ge 1 \ . \tag{10.30}$$

It is also known that $I_m(-z) = (-1)^m I_m(z)$ and $I_m(z)$ is relatively small in the angular region $|\pi/2 - \mathrm{Arg}(z)| < \pi/8$. Therefore $U_4(-z) = U_4(z)$, and one can show that

$$|U_4(z)| = O(|z|^{-1} U_4(|z|)) \tag{10.31}$$

for z away from the real axis.

To apply the subtraction of singularities principle, we need an entire function $f(z)$ that is even, is large only near the real axis, and such that for $x \in \mathbb{R}$, $x \to \infty$,

$$f(x) \sim 3(256\pi^2 x^8)^{-1} \exp(8x) \ . \tag{10.32}$$

The function

$$f^*(z) = 3(128\pi^2 z^8)^{-1} \cosh(8z)$$

is even and has the desired asymptotic growth, but is not entire. We correct this defect by subtracting the contribution of the pole at $z = 0$, and let

$$f(z) = 3(128\pi^2 z^8)^{-1} (\cosh(8z) - 1 - 32z^2 - 512z^4/3 - 16384z^6/45 - 131072z^8/315) \ . \tag{10.33}$$

(It is not necessary to know explicitly the first 8 terms in the Taylor expansion of $\cosh(8z)$ that we wrote down above, as they do not affect the final answer.) With this definition

$$|U_4(z) - f(z)| = O(|z|^{-1} f(|z|)) \tag{10.34}$$

uniformly for all z with $|z| \geq 1$, say, and so if we apply Cauchy's theorem on the circle $|z| = n/4$, say, we find that

$$[z^{2n}](U_4(z) - f(z)) = O(n^{-2n}e^{2n}16^n n^{-9}) \ . \tag{10.35}$$

(The choice of $|z| = n/4$ is made to minimize the resulting estimate.) On the other hand, by Stirling's formula,

$$\begin{aligned} [z^{2n}]f(z) &= 3(128\pi^2)^{-1} \cdot ([z^{2n+8}]\cosh(8z)) \\ &= 3(128\pi^2)^{-1}8^{2n+8}/(2n+8)! \\ &\sim 1536\pi^{-5/2}n^{-2n}16^n e^{2n}n^{-17/2} \quad \text{as} \quad n \to \infty \ . \end{aligned} \tag{10.36}$$

Comparing (10.35) and (10.36), we see that

$$\begin{aligned} u_4(n) = (n!)^2[z^{2n}]U_4(z) &\sim (n!)^2 1536\pi^{-5/2}n^{-2n}16^n e^{2n}n^{-17/2} \\ &\sim 1536\pi^{-3/2}n^{-15/2}16^n \quad \text{as} \quad n \to \infty \ . \end{aligned} \tag{10.37}$$

■

Other methods can be applied to Gessel's generating function to obtain asymptotics of $u_k(n)$ for wider ranges of k ([306]).

The above example obtains a good estimate because the remainder term in (10.30) is smaller than the main term by a factor of $|z|^{-1}$. Had it been smaller only by a factor of $|z|^{-1/2}$, the resulting estimate would have been worthless, and it would have been necessary to obtain a fuller asymptotic expansion of $U_4(z)$ or else use smoothness properties of the remainder term. This is due to the phenomenon, mentioned before, that crude absolute value estimates in either Cauchy's theorem, or the elementary approaches of Section 8, usually lose a factor of $n^{1/2}$ when estimating the n-th coefficient.

The subtraction of singularities principle can be applied even when the generating functions seem to be more complicated than those of Example 10.5. If we consider the problem of that example, but with $k = 5$, then we find that

$$U_5(z) = 3\exp(10z)(5 \cdot 2^{13} \cdot \pi^{5/2}z^{25/2})^{-1}(1 + O(|z|^{-1})) \tag{10.38}$$

as $|z| \to \infty$, with $|\text{Arg}(z)| \leq 3\pi/8$, $U_5(-z) = U_5(z)$, and $U_5(z)$ is entire. We now need an entire function with known coefficients that grows as $\exp(10z)z^{-25/2}$. This is not difficult to obtain, as

$$I_0(10z)z^{-12} - \sum_{j=1}^{12} c_j z^{-j} \tag{10.39}$$

for suitable coefficients c_j has the desired properties.

10.3. The residue theorem and sums as integrals

Sometimes sums that are not easily handled by other methods can be converted to integrals that can be evaluated explicitly or estimated by the residue theorem. If $t(z)$ is a meromorphic function that has first order poles at $z = a,\ a+1, \ldots, b$, with $a \in Z$, each with residue 1, then

$$\sum_{n=a}^{b} f(n) = \frac{1}{2\pi i} \int_{\Gamma} f(z)t(z)dz \ , \tag{10.40}$$

where Γ is a simple closed contour enclosing $a, a+1, \ldots, b$, provided $f(z)$ is analytic inside Γ and $t(z)$ has no singularities inside Γ aside from the first order poles at $a, a+1, \ldots, b$. If $t(z)$ is chosen to have residue $(-1)^n$ at $z = n$, then we obtain

$$\sum_{n=a}^{b} (-1)^n f(n) = \frac{1}{2\pi i} \int_{\Gamma} f(z)t(z)dz \ . \tag{10.41}$$

A useful example is given by the formula

$$\sum_{k=0}^{n} \binom{n}{k} (-1)^k f(k) = \frac{(-1)^n n!}{2\pi i} \int_{\Gamma} \frac{f(z)dz}{z(z-1)\cdots(z-n)} \ . \tag{10.42}$$

The advantage of (10.40) and (10.41) is that the integrals can often be manipulated to give good estimates. This is especially valuable for alternating sums such as (10.41). An analytic function $f(z)$ is extremely regular, so a sum such as that in (10.40) can often be estimated by methods such as the Euler-Maclaurin summation formula (Section 5.3). However, that formula cannot always be applied to alternating sums such as that of (10.41), because the sign change destroys the regularity of $f(n)$. (However, as is noted in Section 5.3, there are generalizations of the Euler-Maclaurin formula that are sometimes useful.) It is hard to write down general rules for applying this method, as most situations require appropriate choice of $t(z)$ and careful handling of the integral. For a detailed discussion of this method, often referred to as Rice's method, see Section 4.9 of [205]. A pair of popular functions to use as $t(z)$ are

$$t_1(z) = \pi/(\sin \pi z), \quad t_2(z) = \pi/(\tan \pi z) \ . \tag{10.43}$$

One can show (Theorem 4.9a of [205]) that if $r(z) = p(z)/q(z)$ with $p(z)$ and $q(z)$ polynomials such that $\deg q(z) \geq \deg p(z) + 2$, and $q(n) \neq 0$ for any $n \in Z$, then

$$\sum_{n=-\infty}^{\infty} r(n) \quad = \quad -\sum \text{Res}(r(z)t_1(z)) \ , \tag{10.44}$$

$$\sum_{n=-\infty}^{\infty} (-1)^n r(n) \quad = \quad -\sum \text{Res}(r(z)t_2(z)) \ , \tag{10.45}$$

where the sums on the right-hand sides above are over the zeros of $q(z)$.

Examples of applications of these methods to asymptotics of data structures are given in [141] and [360].

10.4. Location of singularities, Rouché's theorem, and unimodality

A recurrent but only implicit theme throughout the discussion in this section is that of isolation of zeros. For example, to apply Theorem 9.1 we need to know that the polynomial $h(z)$ has only k zeros, each of multiplicity one, in $|z| < R$. Proofs of such results can often be obtained with the help of Rouché's theorem [205, 364].

Theorem 10.5. *Suppose that $f_1(z)$ and $f_2(z)$ are functions that are analytic inside and on the boundary of a simple closed contour Γ. If*

$$|f_2(z)| < |f_1(z)| \quad \text{for all} \quad z \in \Gamma \ , \tag{10.46}$$

then $f_1(z)$ and $f_1(z) + f_2(z)$ have the same number of zeros (counted with multiplicity) inside Γ.

Example 10.6. *Sequences with forbidden subblocks.* We consider again the topic of Examples 6.4, 6.8, and 9.2, and prove the results that were already used in Example 9.2. We again set

$$h(z) = z^k + (1 - 2z)C_A(z) \ , \tag{10.47}$$

where the only fact about $C_A(z)$ we will use is that it is a polynomial of degree $< k$ and coefficients 0 and 1, and $C_A(0) = 1$. We wish to show that $h(z)$ has only one zero in $|z| \leq 0.6$ if k is large. Write

$$C_A(z) = 1 + \frac{1}{2}\sum_{j=1}^{\infty} z^j + \frac{1}{2}\sum_{j=1}^{\infty} \epsilon_j z^j \ , \tag{10.48}$$

where $\epsilon_j = \pm 1$ for each j. Then

$$C_A(z) = \frac{2 - z}{2(1 - z)} + u(z) \ , \tag{10.49}$$

where

$$|u(z)| \leq \frac{|z|}{2(1 - |z|)} \ .$$

For $|z| = r < 1$, we have $|u(z)| \leq r/(2(1 - r))$. On the other hand, $z \to (2 - z)/(1 - z)$ maps circles to circles, since it is a fractional linear transformation, so it takes the circle $|z| = r$ to the circle with center on the real axis that goes through the two points $(2-r)/(1-r)$ and $(2+r)/(1+r)$. Therefore for $|z| = r < 1$,

$$|C_A(z)| \geq \frac{2+r}{2(1+r)} - \frac{r}{2(1-r)} = \frac{1-r-r^2}{1-r^2} \ , \tag{10.50}$$

and so $|C_A(z)| \geq 1/16$ for $|z| = r \leq 0.6$. Hence, if $k \geq 9$, then on $|z| = 0.6$,

$$|(1-2z)C_A(z)| \geq 1/80 > (0.6)^k \ , \tag{10.51}$$

and thus $(1-2z)C_A(z)$ and $h(z)$ have the same number of zeros in $|z| \leq 0.6$. On the other hand, $C_A(z)$ has no zeros in $|z| \leq 0.6$ by (10.50), while $1-2z$ has one, so we obtain the desired result, at least for $k \geq 9$. (A more careful analysis shows that $h(z)$ has only one root inside $|z| = 0.6$ even for $4 \leq k < 9$. For $1 \leq k \leq 3$, there are cases where there is no zero inside $|z| \leq 0.6$.) Example 6.7 shows how to obtain precise estimates of the single zero.

We note that (10.50) shows that for $|z| = 0.55$, $k \geq 9$

$$|h(z)| \geq |1 - 1.1| 0.2 - (0.55)^k \geq 0.02 - 0.01 \geq 1/100 \ , \tag{10.52}$$

a result that was used in Example 9.2. ■

Example 10.7. *Coins in a fountain.* An (n,k) fountain is an arrangement of n coins in rows such that there are k coins in the bottom row, and such that each coin in a higher row touches exactly two coins in the next lower row. Let $a_{n,k}$ be the number of (n,k) fountains, and $a_n = \sum_k a_{n,k}$ the total number of fountains of n coins. The values of a_n for $1 \leq n \leq 6$ are $1, 1, 2, 3, 5, 9$. If we let $a_0 = 1$ then it can be shown [313] that

$$f(z) = \sum_{n=0}^{\infty} a_n z^n = \cfrac{1}{1 - \cfrac{z}{1 - \cfrac{z^2}{1 - \cfrac{z^3}{1 \ldots}}}} \ . \tag{10.53}$$

This is a famous continued fraction of Ramanujan. (Other combinatorial interpretations of this continued fraction are also known, see the references in [313]. For related results, see [326, 327].) Although one can derive the asymptotics of the a_n from the expansion (10.53), it is more convenient to work with another expansion, known from previous studies of Ramanujan's continued fraction:

$$f(z) = \frac{p(z)}{q(z)} \ , \tag{10.54}$$

where

$$p(z) = \sum_{r\geq 0}(-1)^r \frac{z^{r(r+1)}}{(1-z)(1-z^2)\dots(1-z^r)} , \quad (10.55)$$

$$q(z) = \sum_{r\geq 0}(-1)^r \frac{z^{r^2}}{(1-z)(1-z^2)\dots(1-z^r)} . \quad (10.56)$$

Clearly both $p(z)$ and $q(z)$ are analytic in $|z| < 1$, so $f(z)$ is meromorphic there. We will show that $q(z)$ has a simple real zero x_0, $0.57 < x_0 < 0.58$, and no other zeros in $|z| < 0.62$, while $p(x_0) > 0$. It will then follow from Theorem 10.4 that

$$a_n = c x_0^{-n} + O((5/3)^n) \quad \text{as} \quad n \to \infty , \quad (10.57)$$

where $c = -p(x_0)/(x_0 q'(x_0))$. Numerical computation shows that $c = 0.31236\dots$, $x_0 = 0.576148769\dots$.

To establish the claim about x_0, let $p_n(z)$ and $q_n(z)$ denote the n-th partial sums of the series (10.55) and (10.56), respectively. Write $a(z) = q_3(z)(1-z)(1-z^2)/(1-z^3)$, so that

$$a(z) = 1 - 2z - z^2 + z^3 + 3z^4 + z^5 - 2z^6 - z^7 - z^9 , \quad (10.58)$$

and consider

$$b(z) = \prod_{j=1}^{9}(z - z_j) ,$$

where the z_j are 0.57577, $-0.46997 \pm i0.81792$, $0.74833 \pm i0.07523$, $-1.05926 \pm i0.36718$, $0.49301 \pm i1.58185$, in that order. (The z_j are approximations to the zeros of $a(z)$, obtained from numerical library subroutines. How they were derived is not important for the verification of our proof.) An easy hand or machine computation shows that if $a(z) = \sum_k a_k z^k$, $b(z) = \sum b_k z^k$, then

$$\sum_{k=0}^{9} |a_k - b_k| \leq 1.7 \times 10^{-4} ,$$

and so $|a(z) - b(z)| \leq 1.7 \times 10^{-4}$ for all $|z| \leq 1$. Another computation shows that $|b(z)| \geq 8 \times 10^{-4}$ for all $|z| = 0.62$.

On the other hand, for $0 \leq u \leq 0.62$ and $|z| = u$, we have for $k \geq 5$

$$\left| \frac{z^{(k+1)^2 - k^2}}{1 - z^{k+1}} \right| \leq \frac{u^{2k+1}}{1 - u^{k+1}} \leq \frac{u^9}{1 - u^5} . \quad (10.59)$$

Therefore

$$\left| \sum_{k=4}^{\infty} (-1)^k \frac{z^{k^2}}{\Pi_{j=4}^{k}(1 - z^j)} \right| \leq \frac{u^{16}}{1 - u^4} \sum_{m\geq 0} \left(\frac{u^9}{1 - u^5} \right)^m \leq 6 \times 10^{-4} , \quad (10.60)$$

and so by Rouché's theorem, $q(z)$ and $b(z)$ have the same number of zeros in $|z| \le 0.62$, namely 1. Since $q(z)$ has real coefficients, its zero is real. This establishes the existence of x_0. An easy computation shows that $q(0.57) > 0$, $q(0.58) < 0$, so $0.57 < x_0 < 0.58$.

To show that $p(x_0) > 0$, note that successive summands in (10.55) decrease in absolute magnitude for each fixed real $z > 0$, and $p(z) > 1 - z^2/(1 - z) > 0$ for $0 < z < 0.6$. ■

The method used in the above example is widely applicable to generating functions given by continued fractions. Typically they are meromorphic in a disk centered at the origin, with a single dominant pole of order 1. Usually there is no convenient representation of the form (10.54) with explicit $p(z)$ and $q(z)$, and one has to work harder to establish the necessary properties about location of poles.

It was mentioned in Section 6.4 that unimodality of a sequence is often deduced from information about the zeros of the associated polynomial. If the zeros of the polynomial

$$A(z) = \sum_{k=0}^{n} a_k z^k$$

are real and ≤ 0, then the a_k are unimodal, and even the $a_k \binom{n}{k}^{-1}$ are log-concave. However, weaker properties follow from weaker assumptions on the zeros. If all the zeros of $A(z)$ are in the wedge-shaped region centered on the negative real axis $|\text{Arg}(-z)| \le \pi/4$, and the a_k are real, then the a_k are log-concave, but the $a_k \binom{n}{k}^{-1}$ are not necessarily log-concave. (This follows by factoring $A(z)$ into polynomials with real coefficients that are either linear or quadratic, and noting that all have log-concave coefficients, so their product does too.) One can prove other results that allow zeros to lie in larger regions, but then it is necessary to impose restrictions on ratios of their distances from the origin.

10.5. Implicit functions

Section 6.2 presented functions, such as $f^{\langle -1 \rangle}(z)$, that are defined implicitly. In this section we consider related problems that arise when a generating function $f(z)$ satisfies a functional equation $f(z) = G(z, f(z))$. Such equations arise frequently in graphical enumeration, and there is a standard procedure invented by Pólya and developed by Otter that is almost algorithmic [188, 189] and routinely leads to them. Typically $G(z, w)$ is analytic in z and w in a small neighborhood of $(0, 0)$. Zeros of analytic functions in more than one dimension are not isolated, and by the implicit function theorem $G(z, w) = w$ is solvable for w as a function of

z, except for those points where

$$G_w(z,w) = \frac{\partial}{\partial w} G(z,w) = 1 \ . \tag{10.61}$$

Usually for z in a small neighborhood of 0 the solution w of $G(z,w) = w$ will not satisfy (10.61), and so w will be analytic in that neighborhood. As we enlarge the neighborhood under consideration, though, a simultaneous solution to $G(z,w) = w$ and (10.61) will eventually appear, and will usually be the dominant singularity of $f(z) = w(z)$. The following theorem covers many common enumeration problems.

Theorem 10.6. *Suppose that*

$$f(z) = \sum_{n=1}^{\infty} f_n z^n \tag{10.62}$$

is analytic at $z = 0$, that $f_n \geq 0$ for all n, and that $f(z) = G(z, f(z))$, where

$$G(z,w) = \sum_{m,n \geq 0} g_{m,n} z^m w^n \ . \tag{10.63}$$

Suppose that there exist real numbers $\delta, r, s > 0$ such that

(i) $G(z,w)$ is analytic in $|z| < r + \delta$ and $|w| < s + \delta$,

(ii) $G(r,s) = s$, $G_w(r,s) = 1$,

(iii) $G_z(r,s) \neq 0$ and $G_{ww}(r,s) \neq 0$.

Suppose that $g_{m,n} \in \mathbb{R}^+ \cup \{0\}$ for all m and n, $g_{0,0} = 0$, $g_{0,1} = 1$, and $g_{m,n} > 0$ for some m and some $n \geq 2$. Assume further that there exist $h > j > i \geq 1$ such that $f_h f_i f_j \neq 0$ while the greatest common divisor of $j - i$ and $h - i$ is 1. Then $f(z)$ converges at $z = r$, $f(r) = s$, and

$$f_n = [z^n] f(z) \sim (r G_z(r,s)/(2\pi G_{ww}(r,s)))^{1/2} n^{-3/2} r^{-n} \quad \text{as} \quad n \to \infty \ . \tag{10.64}$$

Example 10.8. *Rooted labeled trees.* As was shown in Example 6.1, the exponential generating function $t(z)$ of rooted labeled trees satisfies $t(z) = z \exp(t(z))$. Thus we have $G(z,w) = z \exp(w)$, and Theorem 10.6 is easily seen to apply with $r = e^{-1}$, $s = 1$. Therefore we obtain the asymptotic estimate

$$t_n/n! = [z^n] t(z) \sim (2\pi)^{-1/2} n^{-3/2} e^n \quad \text{as} \quad n \to \infty \ . \tag{10.65}$$

On the other hand, from Example 6.6 we know that $t_n = n^{n-1}$, a much more satisfactory answer, so that the estimate (10.65) only provides us with another proof of Stirling's formula. ■

The example above involves an extremely simple application of Theorem 10.6. More complicated cases will be presented in Section 15.1.

The statement of Theorem 10.6 is long, and the hypotheses stringent. All that is really needed for the asymptotic relation (10.64) to hold is that $f(z)$ should be analytic on $\{z : |z| \leq r, z \neq r\}$ and that

$$f(z) = c(r-z)^{1/2} + o(|r-z|^{1/2}) \tag{10.66}$$

for $|z - r| \leq \epsilon$, $|\mathrm{Arg}(r-z)| \geq \pi/2 - \epsilon$ for some $\epsilon > 0$. If these conditions are satisfied, then (10.64) follows immediately from either the transfer theorems of Section 11.1 or (with stronger hypotheses) from Darboux's method of Section 11.2. The purpose of Theorem 10.6 is to present a general theorem that guarantees (10.66) holds, is widely applicable, and is stated to the maximum extent possible in terms of conditions on the coefficients of $f(z)$ and $G(z,w)$.

Theorem 10.6 is based on Theorem 5 of [33] and Theorem 1 of [284]. The hypotheses of Theorem 5 of [33] are simpler than those of Theorem 10.6, but, as was pointed out by Canfield [67], the proof is faulty and there are counterexamples to the claims of that theorem. The difficulty is that Theorem 5 of [33] does not distinguish adequately between the different solutions $w = w(z)$ of $w = G(z,w)$, and the singularity of the combinatorially significant solution may not be the smallest among all singularities of all solutions. The result of Meir and Moon [284] provides conditions that assure such pathological behavior does not occur. (The statement of Theorem 10.6 incorporates some corrections to Theorem 1 of [284] provided by the authors of that paper.) It would be desirable to prove results like (10.64) under a simpler set of conditions.

In many problems the function $G(z,w)$ is of the form

$$G(z,w) = g(z)\phi(w) + h(z) \ , \tag{10.67}$$

where $g(z)$, $\phi(w)$, and $h(z)$ are analytic at 0. For this case Meir and Moon have proved a useful result (Theorem 2 of [284]) that implies an asymptotic estimate of the type (10.64). The hypotheses of that result are often easier to verify than those of Theorem 10.6 above. (As was noted by Meir and Moon, the last part of the conditions (4.12a) of [284] has to be replaced by the condition that $y_i > h_i$, $y_j > h_j$, and $y_k > h_k$ for some $k > j > i \geq 1$ with $\gcd(j-i, k-i) = 1$.)

Whenever Theorem 10.6 applies, $f_n = [z^n]f(z)$ equals the quantity on the right-hand side of (10.64) to within a multiplicative factor of $1 + O(n^{-1})$. One can derive fuller expansions for the ratio when needed.

11. Small singularities of analytic functions

In most combinatorial enumeration applications, the generating function has a single dominant singularity. The methods used to extract asymptotic information about coefficients split naturally into two main classes, depending on whether this singularity is large or small.

In some situations the same generating function can be said to have either a large or a small singularity, depending on the range of coefficients that we are interested in. This is illustrated by the following example.

Example 11.1. *Partitions with bounded part sizes.* Let $p(n, m)$ be the number of (unordered) partitions of an integer n into integers $\le m$. It is easy to see that

$$P_m(z) = \sum_{n=0}^{\infty} p(n, m) z^n = \prod_{k=1}^{m} (1 - z^k)^{-1} \ . \tag{11.1}$$

The function $P_m(z)$ is rational, but has to be treated in different ways depending on the relationship of n and m. If n is large compared to m, it turns out to be appropriate to say that $P_m(z)$ has a small singularity, and use methods designed for this type of problems. However, if n is not too large compared to m, then the singularity of $P_m(z)$ can be said to be large. (Since the largest part in a partition of n is almost always $O(n^{1/2} \log n)$ [105], $p(n, m) \sim p(n)$ if m is much larger than $n^{1/2} \log n$.)

Although $P_m(z)$ has singularities at all the k-th roots of unity for all $k \le m$, $z = 1$ is clearly the dominant singularity, as $|P_m(r)|$ grows much faster as $r \to 1^-$ than $|P_m(z)|$ for $z = r \exp(i\theta)$ for any $\theta \in (0, 2\pi)$. If m is fixed, then the partial function decomposition can be used to obtain the asymptotics of $p(n, m)$ as $m \to \infty$. We cannot use Theorem 9.1 directly, since the pole of $P_m(z)$ at $z = 1$ has multiplicity 1. However, either by using the generalizations of Theorem 9.1 that are mentioned in Section 9.1, or by the subtraction of singularities principle, we can show that for any fixed m,

$$p(n, m) \sim [z^n] \left(\prod_{k=1}^{m} k! \right)^{-1} (1 - z)^{-m} \sim \left(\prod_{k=1}^{m} k! \right)^{-1} ((m-1)!)^{-1} \quad \text{as} \quad n \to \infty \ . \tag{11.2}$$

(See [23] for further details and estimates.) This approach can be extended for m growing slowly with n, and it can be shown without much effort that the estimate (11.2) holds for $n \to \infty$, $m \le \log \log n$, say. However, for larger values of m this approach becomes cumbersome, and other methods, such as those of Section 12, are necessary. ■

11.1. Transfer theorems

This section presents some results, drawn from [135], that allow one to translate an asymptotic expansion of a generating function around its dominant singularity into an asymptotic expansion for the coefficients in a direct way. These results are useful in combinatorial enumeration, since the conditions for validity are frequently satisfied. The proofs, which we do not present here, are based on the subtraction of singularities principle, but are more involved than in the cases treated in Section 10.2.

We start out with an application of the results to be presented later in this section.

Example 11.2. *2-regular graphs.* The generating function for 2-regular graphs is known [81] to be

$$f(z) = (1-z)^{-1/2} \exp\left(-\frac{1}{2}z - \frac{1}{4}z^2\right) \ . \tag{11.3}$$

(A simpler proof can be obtained from the exponential formula, cf. Eq. (3.9.1) of [377].) We see that $f(z)$ is analytic throughout the complex plane except for the slit along the real axis from 1 to ∞, and that near $z = 1$ it has the asymptotic expansion

$$f(z) = e^{-3/4}\left\{(1-z)^{-1/2} + (1-z)^{1/2} + \frac{1}{4}(1-z)^{3/2} + \cdots\right\} \ . \tag{11.4}$$

Theorem 11.1 below then shows that as $n \to \infty$,

$$\begin{aligned} [z^n]f(z) &\sim e^{-3/4}\left\{\binom{n-1/2}{n} + \binom{n-3/2}{n} + \frac{1}{4}\binom{n-5/2}{n} + \cdots\right\} \\ &\sim \frac{e^{-3/4}}{\sqrt{\pi n}}\left\{1 - \frac{5}{8n} - \frac{15}{128n^2} + \cdots\right\} \ . \qquad \blacksquare \end{aligned} \tag{11.5}$$

The basic transfer results will be presented for generating functions that have a single dominant singularity, but can be extended substantially beyond their circle of convergence. For r, $\eta > 0$, and $0 < \phi < \pi/2$, we define the closed domain $\Delta = \Delta(r, \phi, \eta)$ by

$$\Delta(r, \phi, \eta) = \{z : \ |z| \le r + \eta, \ |\mathrm{Arg}(z - r)| \ge \phi\} \ . \tag{11.6}$$

In the main result below we will assume that a generating function is analytic throughout $\Delta \setminus \{r\}$. Later in this section we will mention some results that dispense with this requirement. We will also explain why analyticity throughout $\Delta \setminus \{r\}$ is helpful in obtaining results such as those of Theorem 11.1 below.

One advantage to using Cauchy's theorem to recover information about coefficients of generating functions is that it allows one to prove the intuitively obvious result that small smooth changes in the generating function correspond to small smooth changes in the coefficients. We will use the quantitative notion of a function of slow variation at ∞ to describe those functions for which this notion can be made precise. (With more effort one can prove that the same results hold with a less restrictive definition than that below.)

Definition 11.1. *A function $L(u)$ is of slow variation at ∞ if*

i) There exist real numbers u_0 and ϕ_0 with $u_0 > 0$, $0 < \phi_0 < \pi/2$, such that $L(u)$ is analytic and $\neq 0$ in the domain

$$\{u : \ |\mathrm{Arg}(u - u_0)| \leq \pi - \phi_0\} \ . \tag{11.7}$$

ii) There exists a function $\epsilon(x)$, defined for $x \geq 0$ with $\lim_{x\to\infty} \epsilon(x) = 0$, such that for all $\theta \in [-(\pi - \phi_0), \pi - \phi_0]$ and $u \geq u_0$, we have

$$\left| \frac{L(ue^{i\theta})}{L(u)} - 1 \right| \ < \ \epsilon(u) \tag{11.8}$$

and

$$\left| \frac{L(u \log^2 u)}{L(u)} - 1 \right| \ < \ \epsilon(u) \ . \tag{11.9}$$

Theorem 11.1. *Assume that $f(z)$ is analytic throughout the domain $\Delta \setminus \{r\}$, where $\Delta = \Delta(r, \phi, \eta)$, $r, \eta > 0$, $0 < \phi < \pi/2$, and that $L(u)$ is a function of slow variation at ∞. If α is any real number, then*

A) If

$$f(z) = O\left((z - r)^{\alpha} L\left(\frac{1}{r - z}\right)\right)$$

uniformly for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) = O(r^{-n} n^{-\alpha-1} L(n)) \quad \text{as} \quad n \to \infty \ .$$

B) If

$$f(z) = o\left((z - r)^{\alpha} L\left(\frac{1}{r - z}\right)\right)$$

uniformly as $z \to r$ for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) = o(r^{-n} n^{-\alpha-1} L(n)) \quad \text{as} \quad n \to \infty \ .$$

C) If $\alpha \notin \{0, 1, 2, \ldots\}$ and

$$f(z) \sim (r - z)^{\alpha} L\left(\frac{1}{r - z}\right)$$

uniformly as $z \to r$ for $z \in \Delta \setminus \{r\}$, then

$$[z^n]f(z) \sim \frac{r^{-n} n^{-\alpha-1}}{\Gamma(-\alpha)} L(n) \ .$$

The restriction that there be only one singularity on the circle of convergence is easy to relax. If there are several (corresponding to oscillatory behavior of the coefficients), their contributions to the coefficients add. The crucial fact is that at each singularity the function $f(z)$ should be continuous except for an angular region similar to that of $\Delta(r, \phi, \eta)$.

The requirement that the generating function $f(z)$ be analytic in the interior of $\Delta(r, \phi, \eta)$ is in general harder to dispense with, at least by the methods of [135]. However, if the singularity at r is sufficiently large, one can obtain the same results with weaker assumptions that only require analyticity inside the disk $|z| < r$. The following result is implicit in [135].

Theorem 11.2. *Assume that $f(z)$ is analytic in the domain$\{z : |z| \leq r, z \neq r\}$ and that $L(u)$ is a function of slow variation at ∞. If α is any fixed real number with $\alpha < -1$, then the implications A), B), and C) of Theorem 11.1 are valid.*

Example 11.3. *Longest cycle in a random permutation.* The average length of the longest cycle in a permutation on n letters is $[z^n]f(z)$, where

$$f(z) = (1 - z)^{-1} \sum_{k \geq 0} \left[1 - \exp\left(-\sum_{j \geq k} j^{-1} z^j\right)\right] \ .$$

It is easy to see that $f(z)$ is analytic in $|z| < 1$, and a double application of the Euler-Maclaurin summation formula shows that $f(z) \sim G(1 - z)^{-2}$ as $z \to 1$, uniformly for $|z| \leq 1$, $z \neq 1$, where

$$G = \int_0^{\infty} \left[1 - \exp\left(-\int_x^{\infty} t^{-1} e^{-t} dt\right)\right] dx = 0.624 \ldots \ . \tag{11.10}$$

Therefore, by Theorem 11.2 with $L(u) = 1$,

$$[z^n]f(z) \sim Gn \quad \text{as} \quad n \to \infty \ , \tag{11.11}$$

a result first proved by Shepp and Lloyd [342] using Poisson approximations and Tauberian theorems. The derivation sketched above follows [134, 135]. The paper [134] contains many other applications of transfer theorems to random mapping problems. Additional recent papers on the cycle structure of random permutations are [19, 187]. They use probabilistic methods, not transfer theorems, and contain extensive references to other recent works. ■

In applying transfer theorems, it is useful to have explicit expansions and estimates for the coefficients of some frequently occurring functions. We state several asymptotic series:

$$[z^n](1-z)^\alpha \approx \frac{n^{-\alpha-1}}{\Gamma(-\alpha)} \left(1 + \sum_{k\geq 1} e_k^{(\alpha)} n^{-k}\right), \quad \alpha \neq 0, 1, 2, \ldots, \tag{11.12}$$

where

$$e_k^{(\alpha)} = \sum_{j=k}^{2k} (-1)^j \lambda_{k,j} (\alpha+1)(\alpha+2)\cdots(\alpha+j) \ , \tag{11.13}$$

and the $\lambda_{k,j}$ are determined by

$$e^t (1+vt)^{-1-1/v} = \sum_{k,j\geq 0} \lambda_{k,j} v^k t^j \ . \tag{11.14}$$

In particular,

$$\begin{aligned} e_1^{(\alpha)} &= \alpha(\alpha+1)/2, \\ e_2^{(\alpha)} &= \alpha(\alpha+1)(\alpha+2)(3\alpha+1)/24 \ . \end{aligned}$$

Also, for $\alpha, \beta \notin \{0, 1, 2, \ldots\}$,

$$[z^n](1-z)^\alpha(-z^{-1}\log(1-z))^\beta \approx \frac{n^{-\alpha-1}}{\Gamma(-\alpha)} (\log n)^\beta \left(1 + \sum_{k\geq 1} e_k^{(\alpha,\beta)} (\log n)^{-k}\right) \ , \tag{11.15}$$

where

$$e_k^{(\alpha,\beta)} = (-1)^k \binom{\beta}{k} \Gamma(-\alpha) \left(\frac{d^k}{ds^k} \Gamma(-s)^{-1} \Big|_{s=\alpha} \right) \ . \tag{11.16}$$

Further examples of asymptotic expansions are presented in [135].

Why is the analyticity of a function $f(z)$ throughout $\Delta(r, \phi, \eta) \setminus \{r\}$ so important? We explain this using as an example a function $f(z)$ that satisfies

$$f(z) = (1 + o(1))(1-z)^{1/2} \tag{11.17}$$

as $z \to 1$ with $z \in \Delta = \Delta(1, \pi/8, 1)$. We write

$$f(z) = (1-z)^{1/2} + g(z) \ , \tag{11.18}$$

so that

$$|g(z)| = o(|1-z|^{1/2}) \ . \tag{11.19}$$

Since $[z^n](1-z)^{1/2}$ grows like $n^{-3/2}$, we would like to show that

$$|[z^n]g(z)| = o(n^{-3/2}) \quad \text{as} \quad n \to \infty \ . \tag{11.20}$$

If $g(z)$ were analytic in a disk of radius $1+\delta$ for some $\delta > 0$, then we could conclude that $|[z^n]g(z)| < (1+\delta/2)^{-n}$ for large n, a conclusion much stronger than (11.20). However, if all we know is that $g(z)$ satisfies (11.19) in $|z| \le 1$, then we can only conclude from Cauchy's theorem that $[z^n]g(z) = O(1)$, since (11.19) implies that $|g(z)| \le C$ for all $|z| < 1$ and some $C > 0$. Then Theorem 10.2 gives

$$|[z^n]g(z)| \le Cr^{-n} \tag{11.21}$$

uniformly for all $n \ge 0$ and all $r < 1$, and hence $|[z^n]g(z)| \le C$ for all n, a result that is far from what is required. If we know that $g(z)$ can be continued to $\Delta \setminus \{r\}$ and satisfies (11.19) there, we can do a lot better. We choose the contour $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4$, pictured in Fig. 1, with

$$\Gamma_1 = \{z : \ |z-1| = 1/n, \ |\mathrm{Arg}(z-1)| \ge \pi/4\} \ , \tag{11.22}$$

$$\Gamma_2 = \{z : \ z = 1 + r\exp(\pi i/4), \ 1/n \le r \le \delta\} \ , \tag{11.23}$$

$$\Gamma_3 = \{z : \ |z| = |1+\delta\exp(\pi i/4)|, \ |\mathrm{Arg}(z-1)| \ge \pi/4\} \ , \tag{11.24}$$

$$\Gamma_4 = \{z : \ z = 1 + r\exp(-\pi i/4), \ 1/n \le r \le \delta\} \ , \tag{11.25}$$

where $0 < \delta < 1/2$. We will show that the integrals

$$g_j = \frac{1}{2\pi i}\int_{\Gamma_j} g(z) z^{-n-1} dz \tag{11.26}$$

on the Γ_j are small. On Γ_3, $g(z)$ is bounded, so we trivially obtain the exponential upper bound

$$|g_3| = O((1+\delta/2)^{-n}) \ . \tag{11.27}$$

On Γ_1, $|g(z)| = o(n^{-1/2})$, $|z^{-n-1}| \le (1-1/n)^{-n-1} = O(1)$, and the length of Γ_1 is $\le 2\pi/n$, so

$$|g_1| = o(n^{-3/2}) \quad \text{as} \quad n \to \infty \ . \tag{11.28}$$

Next, on Γ_2, for $z = 1 + r\exp(\pi i/4)$,

$$\begin{aligned} |z|^{-n} &= |1 + r2^{-1/2} + ir2^{-1/2}|^{-n} = (1 + r2^{1/2} + r^2)^{-n/2} \\ &\le (1+r)^{-n/2} \le \exp(-nr/10) \end{aligned} \tag{11.29}$$

for $0 \le r < 1$. Since $g(z)$ satisfies (11.19), for any $\epsilon > 0$ we have

$$|g(1 + r\exp(\pi i/4))| \le \epsilon r^{1/2} \tag{11.30}$$

if $0 < r \le \eta$ for some $\eta = \eta(\epsilon) \le \delta$. Therefore

$$\begin{aligned} |g_2| &\le \epsilon \int_0^\eta r^{1/2} \exp(-nr/10) dr + O\left(\int_\eta^\infty \exp(-nr/10) dr \right) \\ &\le \epsilon n^{-3/2} \int_0^\infty r^{1/2} \exp(-r/10) dr + O(\exp(-n\eta/10)) \ , \end{aligned} \tag{11.31}$$

and so

$$|g_2| = o(n^{-3/2}) \ . \tag{11.32}$$

Since $|g_4| = |g_2|$, inequalities (11.27), (11.28), and (11.32) show that (11.20) holds.

The critical factor in the derivation of (11.20) was the bound for (11.29) for $|z|^{-n}$ on the segment $z = 1 + r\exp(\pi i/4)$. Integrating on the circle $|z| = 1$ or even on the line $\mathrm{Re}\,(z) = 1$ does not give a bound for $|z|^{-n}$ that is anywhere as small, and the resulting bounds do not approach (11.20) in strength. The use of the circular arc Γ_1 in the integral is only a minor technical device used to avoid the singularity at $z = 1$.

When one cannot continue a function to a region like $\Delta \setminus \{1\}$, it is sometimes possible to obtain good estimates for coefficients by working with the generating function exclusively in $|z| \le 1$, provided some smoothness properties apply. This method is outlined in the next section.

11.2. Darboux's theorem and other methods

A singularity of $f(z)$ at $z = w$ is called algebraic if $f(z)$ can be written as the sum of a function analytic in a neighborhood of w and a finite number of terms of the form

$$(1 - z/w)^\alpha g(z) \ , \tag{11.33}$$

where $g(z)$ is analytic near w, $g(w) \neq 0$, and $\alpha \notin \{0, 1, 2, \ldots\}$. Darboux's theorem [87] gives asymptotic expansions for functions with algebraic singularities on the circle of convergence. We state one form of Darboux's result, derived from Theorem 8.4 of [354].

Theorem 11.3. *Suppose that $f(z)$ is analytic for $|z| < r$, $r > 0$, and has only algebraic singularities on $|z| = r$. Let a be the minimum of $Re\,(\alpha)$ for the terms of the form (11.33) at* *the singularities of $f(z)$ on $|z| = r$, and let w_j, α_j, and $g_j(z)$ be the w, α, and $g(z)$ for those terms of the form (11.33) for which $Re(\alpha) = a$. Then, as $n \to \infty$,*

$$[z^n]f(z) - \sum_j \frac{g_j(w_j)n^{-\alpha_j-1}}{\Gamma(-\alpha_j)w_j^n} + o(r^{-n}n^{-a-1}) \ . \tag{11.34}$$

Jungen [219] has extended Darboux's theorem to functions that have a single dominant singularity which is of a mixed algebraic and logarithmic form. His method can be applied also to functions that have several such singularities on their circle of convergence.

We do not devote much attention to Darboux's and Jungen's theorems because they can be obtained from the transfer theorems of Section 11.1. The only reason for stating Theorem 11.3 is that it occurs frequently in the literature.

Some functions, such as

$$f(z) = \prod_{k=1}^{\infty}(1 + z^k/k^2) \ , \tag{11.35}$$

are analytic in $|z| \leq 1$, cannot be continued outside the unit circle, yet are nicely behaved on $|z| = 1$. Therefore there is no dominant singularity that can be studied to determine the asymptotics of $[z^n]f(z)$. To minimize the size of the integrand, it is natural to move the contour of integration in Cauchy's formula to the unit circle. Once that is done, it is possible to exploit smoothness properties of $f(z)$ to bound the coefficients. The Riemann-Lebesgue lemma implies that if $f(z)$ is integrable on the unit circle, then as $n \to \infty$,

$$[z^n]f(z) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(e^{i\theta}) \exp(-ni\theta) d\theta = o(1) \ . \tag{11.36}$$

More can be said if the derivative of $f(z)$ exists on the unit circle. When we apply integration by parts to the integral in (11.36), we find

$$[z^n]f(z) = (2\pi n)^{-1} \int_{-\pi}^{\pi} f'(e^{i\theta}) \exp(-(n-1)i\theta) d\theta \ , \tag{11.37}$$

and so $|[z^n]f(z)| = o(n^{-1})$ if $f'(z)$ exists and is integrable on the unit circle. Existence of higher derivatives leads to even better estimates. We do not attempt to state a general theorem, but illustrate an application of this method with an example. The same technique can be used in other situations, for example in obtaining better error terms in Darboux's theorem [87].

Example 11.4. *Permutations with distinct cycle lengths.* Example 8.5 showed that for the function $f(z)$ defined by Eq. (8.58), $[z^n]f(z) \sim \exp(-\gamma)$ as $n \to \infty$. This coefficient is the probability that a random permutation on n letters has distinct cycle lengths. The more precise estimate (8.59) was derived by Greene and Knuth [177] by working with recurrences for the coefficients of $f(z)$ and auxiliary functions. Another approach to deriving fuller asymptotic expansions for $[z^n]f(z)$ is to use the method outlined above. It suffices to show that the function $g(z)$ defined by Eq. (8.62) has a nice expansion in the closed disk $|z| \le 1$. Since

$$g(z) = -z + \sum_{m=2}^{\infty} \frac{(-1)^{m-1}}{m} \{\mathrm{Li}_m(z^m) - z^m\} \ , \tag{11.38}$$

where the $\mathrm{Li}_m(w)$ are the polylogarithm functions [251], one can use the theory of the $\mathrm{Li}_m(w)$. A simpler way to proceed is to note, for example, that

$$\sum_{k=2}^{\infty} \frac{z^{2k}}{k^2} = \sum_{k=2}^{\infty} \frac{z^{2k}}{k(k-1)} + r(z) \ , \tag{11.39}$$

where

$$r(z) = -\sum_{k=2}^{\infty} \frac{z^{2k}}{k^2(k-1)} \ , \tag{11.40}$$

and so $r'(z)$ is bounded and continuous for $|z| \le 1$, as are the terms in (8.62) with $m \ge 3$. On the other hand,

$$\sum_{k=2}^{\infty} \frac{z^{2k}}{k(k-1)} = z^2 + (1-z^2)\log(1-z^2) \ , \tag{11.41}$$

so we can write $g(z) = g_1(z) + g_2(z)$, where $g_1(z)$ is an explicit function (given by Eq. (11.41)) such that the coefficients of $\exp(g_1(z))$ can be estimated asymptotically using transfer methods or other techniques, and $g_2(z)$ has the property that $g_2'(z)$ is bounded and continuous in $|z| \le 1$. Continuing this process, we can find, for every K, an expansion for the coefficients of $f(z)$ that has error term $O(n^{-K})$. To do this, we write $g(z) = G_1(z) + G_2(z)$. In this expansion $G_1(z)$ will be explicitly given and analytic inside $|z| < 1$ and analytically continuable to some region that extends beyond the unit disk with the exception of cuts from a finite number of points on the unit circle out to infinity. Further, $G_2(z)$ will have the property that $G_2^{(K)}(z)$ is bounded and continuous in $|z| \le 1$. This will then give the desired expansion for the coefficients of $f(z)$. ■

12. Large singularities of analytic functions

This section presents methods for asymptotic estimation of coefficients of generating functions whose dominant singularities are large.

12.1. The saddle point method

The saddle point method, also referred to as the method of steepest descent, is by far the most useful method for obtaining asymptotic information about rapidly growing functions. It is extremely flexible and has been applied to a tremendous variety of problems. It is also complicated, and there is no simple categorization of situations where it can be applied, much less of the results it produces. Given the purpose and limitations on the length of this chapter, we do not present a full discussion of it. For a complete and insightful introduction to this technique, the reader is referred to [63]. Many other books, such as [110, 115, 315, 385] also have extensive presentations. What this section does is to outline the method, show when and how it can be applied and what kinds of estimates it produces. Examples of proper and improper applications of the method are presented. Later subsections are then devoted to general results obtained through applications of the saddle point method. These results give asymptotic expansions for wide classes of functions without forcing the reader to go through the details of the saddle point method.

The saddle point method is based on the freedom to shift contours of integration when estimating integrals of analytic functions. The same principle underlies other techniques, such as the transfer method of Section 11.1, but the way it is applied here is different. When dealing with functions of slow growth near their principal singularity, as happens for transfer methods, one attempts to push the contour of integration up to and in some ways even beyond the singularity. The saddle point method is usually applied when the singularity is large, and it keeps the path of integration close to the singularity.

In the remainder of this section we will assume that $f(z)$ is analytic in $|z| < R \leq \infty$. We will also make the assumption that for some R_0, if $R_0 < r < R$, then

$$\max_{|z|=r} |f(z)| = f(r) \ . \tag{12.1}$$

This assumption is clearly satisfied by all functions with real nonnegative coefficients, which are the most common ones in combinatorial enumeration. Further, we will suppose that $z = r$ is the unique point with $|z| = r$ where the maximum value in (12.1) is assumed. When this assumption is not satisfied, we are almost always dealing with some periodicity in the asymptotics of the coefficients, and we can then usually reduce to the standard case by either changing variables or rewriting the generating function as a sum of several others, as was discussed in Section 10. (Such a reduction cannot be applied to the function of Eq. (9.39), though.)

The first step in estimating $[z^n]f(z)$ by the saddle point method is to find the saddle point. Under our assumptions, that will be a point $r \in (R_0, R)$ which minimizes $r^{-n}f(r)$. We have encountered this condition before, in Section 8.1. The minimizing $r = r_0$ will usually be unique, at least for large n. (If there are several $r \in (R_0, R)$ for which $r^{-n}f(r)$ achieves its minimum value, then $f(z)$ is pathological, and the standard saddle point method will not be applicable. For functions $f(z)$ with nonnegative coefficients, it is easy to show uniqueness of the minimizing r, as was already discussed in Section 8.1.) Cauchy's formula (10.6) is then applied with the contour $|z| = r_0$. The reason for this choice is that for many functions, on this contour the integrand is large only near $z = r_0$, the contributions from the region near $z = r_0$ do not cancel each other, and remaining regions contribute little. This is in contrast to the behavior of the integrand on other contours. By Cauchy's theorem, any simple closed contour enclosing the origin gives the correct answer. However, on most of them the integrand is large, and there is so much cancellation that it is hard to derive any estimates. The circle going through the saddle point, on the other hand, yields an integral that can be controlled well by techniques related to Laplace's method and the method of stationary phase that were mentioned in Section 5.5. We illustrate with an example, which is a totally self-contained application of the saddle point method to an extremely simple situation.

Example 12.1. *Stirling's formula.* We estimate $(n!)^{-1} = [z^n]\exp(z)$. The saddle point, according to our definition above, is that $r \in \mathbb{R}^+$ that minimizes $r^{-n}\exp(r)$, which is clearly $r = n$. Consider the contour $|z| = n$, and set $z = n\exp(i\theta)$, $-\pi \le \theta \le \pi$. Then

$$\begin{aligned} [z^n]\exp(z) &= \frac{1}{2\pi i}\int_{|z|=n}\frac{\exp(z)}{z^{n+1}}dz \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} n^{-n}\exp(ne^{i\theta} - ni\theta)d\theta \ . \end{aligned} \tag{12.2}$$

Since $|\exp(z)| = \exp(\mathrm{Re}(z))$, the absolute value of the integrand in (12.2) is $n^{-n}\exp(n\cos\theta)$, which is maximized for $\theta = 0$. Now

$$e^{i\theta} = \cos\theta + i\sin\theta = 1 - \theta^2/2 + i\theta + O(|\theta|^3) \ ,$$

so for any $\theta_0 \in (0, \pi)$,

$$\int_{-\theta_0}^{\theta_0} n^{-n}\exp(ne^{i\theta} - ni\theta)d\theta = \int_{-\theta_0}^{\theta_0} n^{-n}\exp(n - n\theta^2/2 + O(n|\theta|^3))d\theta \ . \tag{12.3}$$

(It is the cancellation of the $ni\theta$ term coming from $ne^{i\theta}$ and the $-ni\theta$ term that came from change of variables in z^{-n} that is primarily responsible for the success of the saddle point method.) The $O(n|\theta|^3)$ term in (12.3) could cause problems if it became too large, so we will select $\theta_0 = n^{-2/5}$, so that $n|\theta|^3 \leq n^{-1/5}$ for $|\theta| \leq \theta_0$, and therefore

$$\exp(n - n\theta^2/2 + O(n|\theta|^3)) = \exp(n - n\theta^2/2)(1 + O(n^{-1/5})) \ . \tag{12.4}$$

Hence

$$\int_{-\theta_0}^{\theta_0} n^{-n} \exp(ne^{i\theta} - ni\theta) d\theta = (1 + O(n^{-1/5})) n^{-n} e^n \int_{-\theta_0}^{\theta} \exp(-n\theta^2/2) d\theta \ .$$

But

$$\begin{aligned}\int_{-\theta_0}^{\theta_0} \exp(-n\theta^2/2) d\theta &= \int_{-\infty}^{\infty} \exp(-n\theta^2/2) d\theta - 2\int_{\theta_0}^{\infty} \exp(-n\theta^2/2) d\theta \\ &= (2\pi/n)^{1/2} - O(\exp(-n^{1/5}/2)) \ ,\end{aligned}$$

so

$$\int_{-\theta_0}^{\theta_0} n^{-n} \exp(ne^{i\theta} - ni\theta) d\theta = (1 + O(n^{-1/5}))(2\pi/n)^{1/2} n^{-n} e^n \ . \tag{12.5}$$

On the other hand, for $\theta_0 < |\theta| \leq \pi$,

$$\cos\theta \leq \cos\theta_0 = 1 - \theta_0^2/2 + O(\theta_0^4) \ ,$$

so

$$n\cos\theta \leq n - n^{1/5}/2 + O(n^{-3/5}) \ ,$$

and therefore for large n

$$\left| \int_{\theta_0}^{\pi} n^{-n} \exp(ne^{i\theta} - ni\theta) d\theta \right| \leq n^{-n} \exp(n - n^{1/5}/3) \ ,$$

and similarly for the integral from $-\pi$ to $-\theta_0$. Combining all these estimates we therefore find that

$$(n!)^{-1} = [z^n] \exp(z) = (1 + O(n^{-1/5}))(2\pi n)^{-1/2} n^{-n} e^n \ , \tag{12.6}$$

which is a weak form of Stirling's formula (4.3). (The full formula can be derived by using more precise expansions for the integrand.)

Suppose we try to push through a similar argument using the contour $|z| = 2n$. This time, instead of Eq. (12.2), we find

$$[z^n] \exp(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} 2^{-n} n^{-n} \exp(2ne^{i\theta} - ni\theta) d\theta \ . \tag{12.7}$$

At $\theta = 0$, the integrand is $2^{-n}n^{-n}\exp(2n)$, which is $\exp(n)$ times as large as the value of the integrand in (12.2). Since the two integrals do produce the same answer, and from the analysis above we see that this answer is close to $n^{-n}\exp(n)$ in value, the integral in (12.7) must involve tremendous cancellation. That is indeed what we see in the neighborhood of $\theta = 0$. We find that

$$\exp(2ne^{i\theta} - ni\theta) = \exp(2n - n\theta^2 + ni\theta + O(n|\theta|^3)) \ , \tag{12.8}$$

and the $\exp(ni\theta)$ term produces wild oscillations of the integrand even over small ranges of θ. Trying to work with the integral (12.7) and proving that it equals something exponentially smaller than the maximal value of its integrand is not a promising approach. By contrast, the saddle point contour used to produce Eq. (12.2) gives nice behavior of the integrand, so that it can be evaluated. ■

The estimates for $n!$ obtained in Example 10.1 came from a simple application of the saddle point method. The motivation for the choice of the contour $|z| = n$ is provided by the discussion at the end of the example; other choices lead to oscillating integrands that cannot be approximated by a Gaussian, nor by any other nice function. The example above treated only the exponential function, but it is easy to see that this phenomenon is general; a rapidly oscillating term $\exp(ni\alpha)$ for $\alpha \neq 0$ is present unless the contour passes through the saddle point. When we do use this contour, and the Gaussian approximation is valid, we find that for functions $f(z)$ satisfying our assumptions we have the following estimate.

Saddle point approximation

$$[z^n]f(z) \sim (2\pi b(r_0))^{-1/2} f(r_0) r_0^{-n} \quad \text{as} \quad n \to \infty \ , \tag{12.9}$$

where r_0 is the saddle point (where $r^{-n}f(r)$ is minimized, so that $r_0 f'(r_0)/f(r_0) = n$) and

$$b(r) = r\frac{f'(r)}{f(r)} + r^2\frac{f''(r)}{f(r)} - r^2\left(\frac{f'(r)}{f(r)}\right)^2 = r\left(r\frac{f'(r)}{f(r)}\right)' \ . \tag{12.10}$$

Example 12.2. *Bell numbers.* Example 5.4 showed how to estimate the Bell number B_n by elementary methods, starting with the representation (5.38). The exponential generating function

$$B(z) = \sum_{n=0}^{\infty} B_n \frac{z^n}{n!} \tag{12.11}$$

satisfies

$$B(z) = \exp(\exp(z) - 1) \ ,$$

as can be seen from (5.38) or by other methods (cf. [81]). The saddle point occurs at that $r_0 > 0$ that satisfies

$$r_0 \exp(r_0) = n \ , \tag{12.12}$$

and

$$b(r_0) = r_0(1 + r_0)\exp(r_0) \ , \tag{12.13}$$

so the saddle point approximation says that as $n \to \infty$,

$$B_n \sim n!(2\pi r_0^2 \exp(r_0))^{-1/2} \exp(\exp(r_0) - 1) r_0^{-n} \ . \tag{12.14}$$

The saddle point approximation can be justified even more easily than for the Stirling estimate of $n!$. ■

The above approximation is widely applicable and extremely useful, but care has to be exercised is applying it. This is shown by the next example.

Example 12.3. *Invalid application of the saddle point method.* Consider the trivial example $f(z) = (1-z)^{-1}$, so that $[z^n]f(z) = 1$ for all $n \geq 0$. Then $f'(r)/f(r) = (1-r)^{-1}$, and so the saddle point is $r_0 = n/(n+1)$, and $b(r_0) = r_0/(1-r_0)^2 = n(n+1)$. Therefore if the approximation (12.9) were valid, it would give

$$\begin{aligned} [z^n]f(z) \quad &\sim \quad (2\pi n(n+1))^{-1/2}(n+1)\left(1 + \frac{1}{n}\right)^n \\ &\sim \quad (2\pi)^{-1/2} e \quad \text{as} \quad n \to \infty \ . \end{aligned} \tag{12.15}$$

Since $(2\pi)^{-1/2} e = 1.0844\ldots \neq 1 = [z^n]f(z)$, something is wrong, and the estimate (12.9) does not apply to this function. ■

The estimate (12.9) gave the wrong result in Example 12.3 because the Gaussian approximation on the saddle point method contour used so effectively in Example 12.1 (and in almost all cases where the saddle point method applies) does not hold over a sufficiently large region for $f(z) = (1-z)^{-1}$. In Example 12.1 we used without detailed explanation the choice $\theta_0 = n^{-2/5}$, which gave the approximation (12.5) for $|\theta| \leq \theta_0$, and yet led to an estimate for the integral over $\theta_0 < |\theta| \leq \pi$ that was negligible. This was possible because the third order term (i.e., $n|\theta|^3$) in Eq. (12.5) was small. When we try to imitate this approach for $f(z) = (1-z)^{-1}$, we fail, because the third order term is too large. Instead of $ne^{i\theta} - ni\theta$, we now have

$$-\log(1 - r_0 e^{i\theta}) - ni\theta = -\log(1 - r_0) - \frac{1}{2}n(n+1)\theta^2 - \frac{i}{6}n^2(n+1)\theta^3 + \cdots \ . \tag{12.16}$$

More fundamentally, the saddle point method fails here because the function $f(z) = (1-z)^{-1}$ does not have a large enough singularity at $z = 1$, so that when one traverses the saddle point contour $|z| = r_0$, the integrand does not drop off rapidly enough for a small region near the real axis to provide the dominant contribution.

When can one apply the saddle point approximation (12.9)? Perhaps the simplest, yet still general, set of sufficient conditions for the validity of (12.9) is provided by requiring that the function $f(z)$ be Hayman-admissible. Hayman admissibility is described in Definition 12.1, in the following subsection. Generally speaking, though, for the saddle point method to apply we need the function $f(z)$ to have a large dominant singularity at R, so that $f(r)$ grows at least as fast as $\exp((\log(R-r))^2)$ as $r \to R^-$ for $R < \infty$, and as fast as $\exp((\log r)^2)$ as $r \to \infty$ for $R = \infty$. The faster the growth rate, the easier it usually is to apply the method, so that $\exp(1/(1-z))$ or $\exp(\exp(1/(1-z)))$ can be treated easily.

In our application of the saddle point method to $\exp(z)$ in Example 12.1 we were content to obtain a poor error term, $1 + O(n^{-1/5})$, in Stirling's formula for $n!$. This was done to simplify the presentation and concentrate only on the main factors that make the saddle point method successful. With more care devoted to the integral one can obtain the full asymptotic expansion of $n!$. (Only the range $|\theta| \le \theta_0$ has to be considered carefully.) This is usually true when the saddle point method is applicable.

This section provided a sketchy introduction to the saddle point method. For a much more thorough presentation, including a discussion of the topographical view of the integrand and the "hill-climbing" interpretation of the contour of integration, see [63].

12.2. Admissible functions

The saddle point method is a powerful and flexible tool, but in its full generality it is often cumbersome to apply. In many situations it is possible to apply general theorems derived using the saddle point method that give asymptotic approximations that are not the sharpest possible, but which allow one to avoid the drudgery of applying the method step by step. The general theorems that we present were proved by Hayman [204] and by Harris and Schoenfeld

[198]. We next describe the classes of functions to which these theorems apply, and then present the estimates one obtains for them. It is not always easy to verify that these definitions hold, but it is almost always easier to do this than to apply the saddle point method from scratch. It is worth mentioning, furthermore, that for many generating functions, there are conditions that guarantee that they satisfy the hypotheses of the Hayman and the Harris-Schoenfeld theorems. These conditions are discussed later in this section.

The definition below is stated somewhat differently than the original one in [204], but can be shown to be equivalent to it.

Definition 12.1. *A function*

$$f(z) \;=\; \sum_{n=0}^{\infty} f_n z^n \tag{12.17}$$

is admissible in the sense of Hayman (or H-admissible) if

i) $f(z)$ *is analytic in* $|z| < R$ *for some* $0 < R \leq \infty$,

ii) $f(z)$ *is real for* z *real,* $|z| < R$,

iii) for $R_0 < r < R$,

$$\max_{|z|=r} |f(z)| \;=\; f(r)\;, \tag{12.18}$$

iv) for

$$a(r) \;=\; r\,\frac{f'(r)}{f(r)}\;, \tag{12.19}$$

$$b(r) \;=\; ra'(r) = r\,\frac{f'(r)}{f(r)} + r^2\frac{f''(r)}{f(r)} - r^2\left(\frac{f'(r)}{f(r)}\right)^2\;, \tag{12.20}$$

and for some function $\delta(r)$, *defined in the range* $R_0 < r < R$ *to satisfy* $0 < \delta(r) < \pi$, *the following three conditions hold:*

$$a) \quad f(re^{i\theta}) \sim f(r)\ \exp(i\theta a(r) - \theta^2 b(r)/2)$$
$$\text{as } r \to R \text{ uniformly for } |\theta| < \delta(r), \tag{12.21}$$

$$b) \quad f(re^{i\theta}) \;=\; o(f(r)b(r)^{-1/2})$$
$$\text{as } r \to R \text{ uniformly for } |\theta| < \delta(r), \tag{12.22}$$

$$c) \quad b(r) \to \infty \quad \text{as} \quad r \to R. \tag{12.23}$$

For H-admissible functions, Hayman [204] proved a basic result that gives the asymptotics of the coefficients.

Theorem 12.1. *If $f(z)$, defined by Eq. (12.17), is H-admissible in $|z| < R$, then*

$$f_n = (2\pi b(r))^{-1/2} f(r) r^{-n} \left\{ \exp\left(-\frac{(a(r)-n)^2}{b(r)} \right) + o(1) \right\} \tag{12.24}$$

as $r \to R$, with the $o(1)$ term uniform in n.

If we choose $r = r_n$ to be a solution to $a(r_n) = n$, then we obtain from Theorem 12.1 a simpler result. (The uniqueness of r_n follows from a result of Hayman [204] which shows that $a(r)$ is positive increasing in some range $R_1 < r < R$, $R_1 > R_0$.)

Corollary 12.1. *If $f(z)$, defined by Eq. (12.17), is H-admissible in $|z| < R$, then*

$$f_n \sim (2\pi b(r_n))^{-1/2} f(r_n) r_n^{-n} \quad \text{as} \quad n \to \infty \ , \tag{12.25}$$

where r_n is defined uniquely for large n by $a(r_n) = n$, $R_0 < r_n < R$.

Corollary 12.1 is adequate for most situations. The advantage of Theorem 12.1 is that it gives a uniform estimate over the approximate range $|a(r) - n| \precsim b(r)^{1/2}$. (Note that the estimate (12.24) is vacuous for $|a(r) - n| \; b(r)^{-1/2} \to \infty$.) Theorem 12.1 shows that the $f_n r^n$ are approximately Gaussian in the central region.

There are many direct applications of the above results.

Example 12.4. *Stirling's formula.* Let $f(z) = \exp(z)$. Then $f(z)$ is H-admissible for $R = \infty$; conditions i)–iii) of Definition 12.1 are trivially satisfied, while $a(r) = r$, $b(r) = r$, so iv) also holds for $R_0 = 0$, $\delta(r) = r^{-1/3}$, say. Corollary 12.1 then shows that

$$f_n = \frac{1}{n!} \sim (2\pi n)^{-1/2} e^n n^{-n} \quad \text{as} \quad n \to \infty \ , \tag{12.26}$$

since $r_n = n$, which gives a weak form of Stirling's approximation to $n!$. ■

In many situations the conditions of H-admissibility are much harder to verify than for $f(z) = \exp(z)$, and even in that case there is a little work to be done to verify that condition iv) holds. However, many of the generating functions one encounters are built up from other, simpler generating functions, and Hayman [204] has shown that often the resulting functions are guaranteed to be H-admissible. We summarize some of Hayman's results in the following theorem.

Theorem 12.2. *Let $f(z)$ and $g(z)$ be H-admissible for $|z| < R \leq \infty$. Let $h(z)$ be analytic in $|z| < R$ and real for real z. Let $p(z)$ be a polynomial with real coefficients.*

i) If the coefficients a_n of the Taylor series of $\exp(p(z))$ are positive for all sufficiently large n, then $\exp(p(z)))$ is H-admissible in $|z| < \infty$.

ii) $\exp(f(z))$ and $f(z)g(z)$ are H-admissible in $|z| < R$.

iii) If, for some $\eta > 0$, and $R_1 < r < R$,

$$\max_{|z|=r} |h(z)| = O(f(r)^{1-\eta}) , \tag{12.27}$$

then $f(z) + h(z)$ is H-admissible in $|z| < R$. In particular, $f(z) + p(z)$ is H-admissible in $|z| < R$ and, if the leading coefficient of $p(z)$ is positive, $p(f(z))$ is H-admissible in $|z| < R$.

Example 12.5. *H-admissible functions.* a) By i) Theorem 12.2, $\exp(z)$ is H-admissible, so we immediately obtain the estimate (12.26), which yields Stirling's formula. b) Since $\exp(z)$ is H-admissible, part iii) of Theorem 12.2 shows that $\exp(z) - 1$ is H-admissible. c) Applying part ii) of Theorem 12.2, we next find that $\exp(\exp(z) - 1)$ is H-admissible, which yields the asymptotics of the Bell numbers. ■

Hayman's results give only first order approximations for the coefficients of H-admissible functions. In some circumstances it is desirable to obtain full asymptotic expansions. This is possible if we impose additional restrictions on the generating function. We next state some results of Harris and Schoenfeld [198].

Definition 12.2. *A function $f(z)$ defined by Eq. (12.17) is* HS-admissible *provided it is analytic in $|z| < R$, $0 < R \leq \infty$, is real for real x, and satisfies the following conditions:*

A) There is an R_0, $0 < R_0 < R$ and a function $d(r)$ defined for $r \in (R_0, R)$ such that

$$\begin{aligned} 0 < d(r) < 1 \ , \\ r\{1 + d(r)\} < R \ , \end{aligned} \tag{12.28}$$

and such that $f(z) \neq 0$ for $|z - r| < rd(r)$.

B) If we define, for $k \geq 1$,

$$A(z) = \frac{f'(z)}{f(z)}, \quad B_k(z) = \frac{z^k}{k!} A^{(k-1)}(z), \quad B(z) = \frac{z}{2} B_1(z) \ , \tag{12.29}$$

then we have

$$B(r) > 0 \ \ for \ \ R_0 < r < R \ \ and \ \ B_1(r) \to \infty \ \ as \ \ r \to R \ .$$

C) For sufficiently large R_1 and n, there is a unique solution $r = u_n$ to

$$B_1(r) = n + 1, \quad R_1 < r < R \ . \tag{12.30}$$

Let

$$C_j(z, r) = \frac{-1}{B(r)} \left\{ B_{j+2}(z) + \frac{(-1)^j}{j+2} B_1(r) \right\} \ . \tag{12.31}$$

There exist nonnegative D_n, E_n, and n_0 such that for $n \geq n_0$,

$$|C_j(u_n, u_n)| \leq E_n D_n^j, \quad j = 1, 2, \ldots \ . \tag{12.32}$$

D) As $n \to \infty$,

$$\begin{aligned} &B(u_n) d(u_n)^2 \to \infty \ , \\ &D_n E_n B(u_n) d(u_n)^3 \to 0 \ , \\ &D_n d(u_n) \to 0 \ . \end{aligned} \tag{12.33}$$

For *HS*-admissible functions, Harris and Schoenfeld obtain complete asymptotic expansions.

Theorem 12.3. *If $f(z)$, defined by (12.17), is* HS-*admissible, then for any $N \geq 0$,*

$$f_n = 2(\pi \beta_n)^{-1/2} f(u_n) u_n^{-n} \left\{ 1 + \sum_{k=1}^{N} F_k(n) \beta_n^{-k} + O(\phi_N(n; d)) \right\} \ \ as \ n \to \infty \ , \tag{12.34}$$

where

$$\beta_n = B(u_n) \ , \tag{12.35}$$

$$F_k(n) = \frac{(-1)^k}{\sqrt{\pi}} \sum_{m=1}^{2k} \frac{\Gamma(m + k + \frac{1}{2})}{m!} \sum_{\substack{j_1 + \cdots + j_m = 2k \\ j_1, \ldots, j_m \geq 1}} \gamma_{j_1}(n) \cdots \gamma_{j_m}(n) \ , \tag{12.36}$$

$$\gamma_j(n) = C_j(u_n, u_n) \ , \tag{12.37}$$

and

$$\phi_N(n; d) = \max\{\mu(u_n, d), E'_n (D_n E''_n \beta_n^{-1/2})^{2N+2}\} \ ,$$

with

$$E'_n = \min(1, E_n), \quad E''_n = \max(1, E_n) \ , \tag{12.38}$$

$$\mu(r,d) = \max\left\{\lambda(r;d)B(r)^{1/2},\ \frac{\exp(-B(r)d(r)^2)}{d(r)B(r)^{1/2}}\right\} \ , \tag{12.39}$$

where $\lambda(r;d)$ *is the maximum value of* $|f'(z)/f(z)|$ *for* z *on the oriented path* $Q(r)$ *consisting of the line segment from* $r + ird(r)$ *to* $(1 - d(r)^2)^{1/2} + ird(r)$ *and of the circular arc from the last point to* ir *to* $-r$*.*

The conditions for *HS*-admissibility are often hard to verify. However, there is a theorem [311] which guarantees that they do hold for a large class of interesting functions.

Theorem 12.4. *If* $g(z)$ *is* H*-admissible, then* $f(z) = \exp(g(z))$ *is* HS*-admissible. Furthermore, the error term* $\phi_N(n;d)$ *of Theorem 12.3 is then* $o(\beta_n^{-N})$ *as* $n \to \infty$ *for every fixed* $N \geq 0$*.*

Example 12.6. *Bell numbers and* HS*-admissibility.* Since $\exp(x) - 1$ is H-admissible, as we saw in Example 12.5, we find that $\exp(\exp(z) - 1)$ is HS-admissible, and Theorem 12.3 yields a complete asymptotic expansion of the Bell numbers. ■

Theorem 12.4 does not apply when $g(z)$ is a polynomial. As is pointed out by Schmutz [339], for $g(z) = z^4 - z^3 + z^2$ the function $f(z) = \exp(g(z))$ is *HS*-admissible, but Theorem 12.3 does not give an asymptotic expansion because the error term $\phi_N(n;d)$ is too large. Schmutz [339] has obtained necessary and sufficient conditions for Theorem 12.3 to give an asymptotic expansion for the coefficients of $f(z) = \exp(g(z))$ when $g(z)$ is a polynomial.

12.3. Other saddle point applications

Section 12.1 presented the basic saddle point method and discussed its range of applicability. Section 12.2 was devoted to results derived using this method that are general and yet can be applied in a cook-book style, without a deep understanding of the saddle point technique. Such a cook-book approach is satisfactory in many situations. However, often one encounters asymptotic estimation problems that are not covered by any of general results mentioned in Section 12.2, but can be solved using the saddle point method. This section mentions several such results of this type that illustrate the range of problems to which this method is applicable. Additional applications will be presented in Section 15, where other techniques are combined with the saddle point method.

Example 12.7. *Stirling numbers.* The Stirling numbers of the first kind, $s(n,k)$, satisfy (6.5) as well as [81]

$$\sum_{k=0}^{n} s(n,k)z^k = z(z-1)\cdots(z-n+1) \ . \tag{12.40}$$

Since $(-1)^{n+k}s(n,k) > 0$, (which is reflected in the behavior of the generating function (12.40), which grows faster along the negative real axis than along the positive one), we rewrite it as

$$\sum_{k=0}^{n}(-1)^{n+k}s(n,k)z^k = z(z+1)\cdots(z+n-1) \ . \tag{12.41}$$

The function on the right-hand side behaves like a good candidate for an application of the saddle point method. For details, see [295, 296]. ■

The estimates mentioned in Example 12.7 are far from best possible in either the size of the error term or (more important) in the range of validity. References for the best currently known results about Stirling numbers of both the first and second kind are given in [363]. Some of the results in the literature are not rigorous. For example, [363] presents elegant and uniform estimates based on an application of the saddle point method. They are likely to be correct, but the necessary rigorous error analysis has not been performed yet, although it seems that this should be doable. Other results, like those of [232] are obtained by methods that there does not seem to be any hope of making rigorous in the near future. Some of the results, though, such as the original ones of Moser and Wyman [295, 296], and the more recent one of Wilf [378], are fully proved.

The saddle point method can be used to obtain full asymptotic expansions. These expansions are usually in powers of $n^{-1/2}$ when estimating $[z^n]f(z)$, and they hardly ever converge, but are asymptotic expansions as defined by Poincaré (as in Eq. (2.2)). The usual forms of the saddle point method are incapable of providing expansions similar to the Hardy-Ramanujan-Rademacher convergent series for the partition function $p(n)$ (Eq. (3.1)). However, the saddle point method can be applied to estimate $p(n)$. There are technical difficulties, since the generating function

$$f(z) = \sum_{n=0}^{\infty} p(n)z^n = \prod_{k=1}^{\infty}(1-z^k)^{-1} \tag{12.42}$$

has a large singularity at $z = 1$, but in addition has singularities at all other roots of unity. The contribution of the integral for z away from 1 can be crudely estimated to be $O(n^{-1}\exp(Cn^{1/2}/2))$ (the last term in Eq. (1.5)). A simple estimate of the integral near $z = 1$ yields the asymptotic expansion of Eq. (1.6). A more careful treatment of the integral, but one that follows the conventional saddle point technique, replaces the $1 + O(n^{-1/2})$ term in Eq. (1.6) by an asymptotic (in the sense of Poincare, so nonconvergent) series $\sum c_k n^{-k/2}$. To obtain Eq. (1.5), one needs to choose the contour of integration near $z = 1$ carefully and use precise estimates of $f(z)$ near $z = 1$.

De Bruijn [63] also discusses applications of the saddle point method when the saddle point is not on the real axis, and especially when there are several saddle points that contribute comparable amounts. This usually occurs when there are oscillations in the coefficients. When the oscillations are irregular, the tricks mentioned in Section 10 of changing variables do not work, and the contributions of the multiple saddle points have to be evaluated.

Example 12.8. *Oscillating sequence.* Consider the sequence a_n of Examples 9.4 and 10.1. As is shown in Example 9.4, its ordinary generating function is given by (9.39). It has an essential singularity at $z = 1$, but is analytic every place else. This function is not covered by our earlier discussion. For example, its maximal value is in general not taken on the positive real axis. It can be shown that the Cauchy integral has two saddle points, at approximately $z = 1-(2n)^{-1} \pm in^{-1/2}(1-(4n)^{-1})^{1/2}$. Evaluating $[z^n]f(z)$ by using Cauchy's theorem with the contour chosen to pass through the two points in the correct way yields the estimate (9.38). ■

In applying the saddle point method, a general principle is that multiplying a generating function $f(z)$ with dominant singularity at R by another function $g(z)$ which is analytic in $|z| < R$ and has much lower growth rate near $z = R$ yields a function $f(z)g(z)$ whose saddle point is close to that of $f(z)$. Usually one can obtain a relation of the form

$$[z^n](f(z)g(z)) \sim g(r_0)([z^n]f(z)) \ , \tag{12.43}$$

where r_0 is the saddle point for $f(z)$. This principle (which is related to the one behind Theorem 7.1) is useful, but has to be applied with caution, and proofs have to be provided for each case. For fuller exposition of this principle and general results, see [157]. The advantage of this approach is that often $f(z)$ is easy to manipulate, so the determination of a saddle point for it is easy, whereas multiplying it by $g(z)$ produces a messy function, and the exact saddle point for $f(z)g(z)$ is difficult to determine.

Example 12.9. *Boolean lattice of subsets of* $\{1, \ldots, n\}$. The number a_n of Boolean sublattices of the Boolean lattice of subsets of $\{1, \ldots, n\}$ has the exponential generating function [162]

$$A(z) = \sum_{n=0}^{\infty} a_n \frac{z^n}{n!} = \exp(2z + \exp(z) - 1) \ . \tag{12.44}$$

We can write $A(z) = \exp(2z)B(z)$, where $B(z)$ is the exponential generating function for the Bell numbers (Example 12.2). Since $B(z)$ grows much faster than $\exp(2z)$, it is easy to show that (12.43) applies, and so

$$a_n \sim \exp(2r_0)B_n \quad \text{as} \quad n \to \infty \ , \tag{12.45}$$

where r_0 is the saddle point for $B(z)$. Using the approximation (12.12) of Example 12.2, we find that

$$a_n \sim (n/\log n)^2 B_n \quad \text{as} \quad n \to \infty \ . \tag{12.46}$$

■

The insensitivity of the saddle point approximation to slight perturbations is reflected in slightly different definitions of a saddle point that are used. The saddle point approximation (12.9) for $[z^n]f(z)$ is stated in terms of r_0, the point that minimizes $f(r)r^{-n}$. The discussion of the saddle point emphasized minimization of the peak value of the integrand in Cauchy's formula, which is the same as minimizing $f(r)r^{-n-1}$, since the contour integral (10.6) involves $f(z)z^{-n-1}$. Some sources call the point minimizing $f(r)r^{-n-1}$ the saddle point. It is not important which definition is adopted. The asymptotic series coefficients look slightly differently in the two cases, but the final asymptotic series, when expressed in terms of n, are the same. The reason for slightly preferring the definition that minimizes $f(r)r^{-n}$ is that when the change of variable $z = r\exp(i\theta)$ is made in Cauchy's integral, there is no linear term in θ, and the integrand involves $\exp(-cn\theta^2 + O(|\theta|^3))$. If we minimized $f(r)r^{-n-1}$, we would have to deal with $\exp(-c'i\theta - c''n\theta^2 + O(|\theta|^3))$, which is not much more difficult to handle but is less elegant.

The same principle can be applied when the exact saddle point is hard to determine, and it is awkward to work with an implicit definition of this point. When that happens, there is often a point near the saddle point that is easy to handle, and for which the saddle point approximation holds. We refer to [157] for examples and discussion of this phenomenon.

12.4. The circle method and other techniques

As we mentioned in Section 12.3, the saddle point method is a powerful method that estimates the contribution of the neighborhood of only a single point, or at most a few points. The convergent series of Eq. (1.3) for the partition function $p(n)$ (as well as the earlier non-convergent but asymptotic and very accurate expansion of Hardy and Ramanujan) is obtained by evaluating the contribution of the other singularities of $f(z)$ to the integral. The m-th term in Eq. (1.3) comes from the primitive m-th roots of unity. To obtain this expansion one needs to use a special contour of integration and detailed knowledge of the behavior of $f(z)$. The details of this technique, called the circle method, can be found in [13, 23].

Convergent series can be obtained from the circle method only when the generating function is of a special form. For results and references, see [8, 10].

Nonconvergent but accurate asymptotic expansions can be derived from the circle method in a much wider variety of applications. It is especially useful when there is no single dominant singularity. For the partition function $p(n)$, all the singularities away from $z = 1$ contribute little, and it is $z = 1$ that creates the dominant term and yields Eq. (1.6). For other functions this is often false. For example, when dealing with additive problems of Waring's type, where one studies $N_{k,m}(n)$, the number of representations of a nonnegative integer n as

$$n = \sum_{j=1}^{m} x_j^k, \quad x_j \in \mathbb{Z}^+ \cup \{0\} \quad \text{for all} \quad j \ , \tag{12.47}$$

the natural generating function to study is

$$\sum_{n=0}^{\infty} N_{k,m}(n) z^n = g(z)^m \ , \tag{12.48}$$

where

$$g(z) = \sum_{h=0}^{\infty} z^{h^k} \ . \tag{12.49}$$

The function $g(z)$ has a natural boundary at $|z| = 1$, but it again grows fastest as z approaches a root of unity from within $|z| < 1$, so it is natural to speak of $g(z)$ having singularities at the roots of unity. The singularity at $z = 1$ is still the largest, but not by much, as other roots of unity contribute comparable amounts, with the contribution of other roots of unity ζ diminishing as the order of ζ increases. All the contributions can be estimated, and one can obtain solutions to Waring's problem (which was to show that for every k, there is an integer m such that $N_{k,m}(n) > 0$ for all n) and other additive problems. For details of this method see [23]. We mention here that for technical reasons, one normally works with generating functions of the form $G_n(z)^m$, where

$$G_n(z) = \sum_{h=0}^{\lfloor n^{1/k} \rfloor} z^{h^k} \ , \tag{12.50}$$

(so that the generating function depends on n), and analyzes them for $|z| = 1$ (since they are now polynomials), but the basic explanation above of why this process works still applies.

13. Multivariate generating functions

A major difficulty in estimating the coefficients of multivariate generating functions is that the geometry of the problem is far more difficult. It is harder to see what are the critical regions where the behavior of the function determines the asymptotics of the coefficients, and those regions are more complicated. Singularities and zeros are no longer isolated, as in the one-dimensional case, but instead form $(k-1)$-dimensional manifolds in k variables. Even rational multivariate functions are not easy to deal with.

One basic tool in one-dimensional complex analysis is the residue theorem, which allows one to move a contour of integration past a pole of the integrand. (We derived a form of the residue theorem in Section 10, in the discussion of poles of generating functions.) There is an impressive generalization by Leray [4, 250] of this theory to several dimensions. Unfortunately, it is complicated, and with few exceptions (such as that of [252], see also [49]) so far it has not been applied successfully to enumeration problems. On the other hand, there are some much simpler tools that can frequently be used to good effect.

An important tool in asymptotics of multivariate generating functions is the multidimensional saddle point method.

Example 13.1. *Alternating sums of powers of binomial coefficients.* Consider

$$S(s,n) = \sum_{k=0}^{2n} (-1)^{k+n} \binom{2n}{k}^s , \tag{13.1}$$

where s and n are positive integers. It has been known for a long time that $S(1,n) = 0$, $S(2,n) = (2n)!(n!)^{-2}$, $S(3,n) = (3n)!(n!)^{-3}$. However, no formula of this type has been known for $s > 3$. De Bruijn (see Chapter 4 of [63]) showed that $S(s,n)$ for integer $s > 3$ cannot be expressed as a ratio of products of factorials. Although his proof is not presented as an application of the multidimensional saddle point method, it is easy to translate it into those terms. $S(s,n)$ is easily seen to equal the constant term in

$$F(z_1, \ldots, z_{s-1}) = (-1)^n (1+z_1)^{2n} \ldots (1+z_{s-1})^{2n} (1-(z_1 \ldots z_{s-1})^{-1})^{2n} , \tag{13.2}$$

and so

$$S(s,n) = (2\pi i)^{-s+1} \int \cdots \int F(z_1, \ldots, z_{s-1}) z_1^{-1} \ldots z_{s-1}^{-1} dz_1 \ldots dz_{s-1} , \tag{13.3}$$

where the integral is taken with each z_j traversing a circle, say. De Bruijn's proof in effect shows that for s fixed and $n \to \infty$, there are two saddle points at $z_1 = \cdots = z_{s-1} = \exp(2i\alpha)$, with $\alpha = \pm(2s)^{-1}$, and this leads to the estimate

$$S(s,n) \sim \left\{2\cos\left(\frac{\pi}{2s}\right)\right\}^{2ns+s-1} 2^{2-s}(\pi n)^{(1-s)/2} s^{-1/2} \quad \text{as} \quad n \to \infty \ , \tag{13.4}$$

valid for any fixed integer $s \geq 2$. Since $\cos(\pi(2s)^{-1})$ is algebraic but irrational for $s \geq 4$, the asymptotic estimate (13.4) shows that $S(s,n)$ cannot be expressed as a ratio of finite products of $(a_j n)!$ for any fixed finite set of integers a_j.

In Chapter 6 of [63], de Bruijn derives the asymptotics of $S(s,n)$ as $n \to \infty$ for general real s. The approach sketched above no longer applies, and de Bruijn uses the integral representation

$$S(s,n) = \int_C \left(\frac{\Gamma(2n+1)}{\Gamma(n+z+1)\Gamma(n-z+1)}\right)^s \frac{dz}{2i\sin\pi z} \ ,$$

where C is a simple closed curve that contains the points $-n, -n+1, \ldots, -1, 0, 1, \ldots, n$ in its interior and has no other integer points on the real axis in its closure. A complicated combination of analytic techniques, including the one-dimensional saddle point method, then leads to the final asymptotic estimate of $S(s,n)$. ■

The multidimensional saddle point method works best when applied to large singularities. Just as for the basic one-dimensional method, it does not work when applied to small singularities, such as those of rational functions. Fortunately, there is a trick that often succeeds in converting a small singularity in n dimensions into a large one in $n-1$ dimensions. The main idea is to expand the generating function with respect to one of the variables through partial fraction expansions or other methods. It is hard to write down a general theorem, but the next example illustrates this technique.

Example 13.2. *Alignments of k sequences.* Let $f(k,n)$ denote the number of $k \times m$ matrices of 0's and 1's such that each column sum is ≥ 1 and each row sum is exactly n. (The number of columns, m, can vary, although obviously $k \leq m \leq kn$.) We consider k fixed, $n \to \infty$ [178]. If we let $N(r_1, \ldots, r_k)$ denote the number of $0, 1$ matrices with k rows, no columns of all 0's, and row sums $r_1, \ldots, r_k$, then it is easy to see [178] that

$$F(z_1, \ldots, z_k) = \sum_{r_1,\ldots,r_k \geq 0} N(r_1, \ldots, r_k) z_1^{r_1} \cdots z_k^{r_k} = \left(2 - \prod_{j=1}^{k}(1+z_j)\right)^{-1} . \tag{13.5}$$

We have $f(k,n) = N(n, \ldots, n)$, and so we need the diagonal terms of $F = F(z_1, \ldots, z_k)$. The function F is rational, so its singularity is small. Moreover, the singularities of F are difficult to visualize. However, in any single variable F is simple. We take advantage of this feature. Let

$$A(z) = \prod_{j=1}^{k-1} (1+z_j) \ , \tag{13.6}$$

where z stands for $(z_1, \ldots, z_{k-1}) \in \mathbb{C}^{k-1}$, and expand

$$\left(2 - \prod_{j=1}^{k}(1+z_j)\right)^{-1} = (2 - A(z)(1+z_k))^{-1} = \sum_{m=0}^{\infty} \frac{A(z)^m z_k^m}{(2-A(z))^{m+1}} \ . \tag{13.7}$$

Therefore

$$N(r_1, \ldots, r_{k-1}, m) = \frac{1}{(2\pi i)^{k-1}} \int \cdots \int \frac{A(z)^m}{(2-A(z))^{m+1}} \frac{dz_1}{z_1^{r_1+1}} \cdots \frac{dz_{k-1}}{z_{k-1}^{r_{k-1}}} \ . \tag{13.8}$$

The function whose coefficients we are trying to extract is now $A(z)^m/(2-A(z))^{m+1}$, which is still rational. However, the interesting case for us is $m \to \infty$, which transforms the singularity into a large one. We are interested in the case $r_1 = r_2 = \cdots = r_{k-1} = r = n$. Then the integral in (13.8) can be shown to have a saddle point at $z_j = \rho$, $1 \leq j \leq k-1$, where $\rho = 2^{1/k} - 1$, and one obtains the estimate [178]

$$f(k,n) = r^n n^{-(k-1)/2} \{ (\rho \pi^{(k-1)/2} k^{1/2})^{-1} 2^{(k^2-1)/(2k)} + O(n^{-1/2}) \} \quad \text{as} \quad n \to \infty \ . \quad \blacksquare \tag{13.9}$$

The examples above of applications of the multidimensional saddle point method all dealt with problems in a fixed dimension as various other parameters increase. A much more challenging problem is to apply this method when the dimension varies. A noteworthy case where this has been done successfully is the asymptotic enumeration of graphs with a given degree sequence by McKay and Wormald [279].

Example 13.3. *Simple labeled graphs of high degree.* Let $G(n; d_1, \ldots, d_n)$ be the number of labeled simple graphs on n vertices with degree sequence $d_1, d_2, \ldots, d_n$. Then $G(n; d_1, \ldots, d_n)$ is the coefficient of $z_1^{d_1} z_2^{d_2} \cdots z_n^{d_n}$ in

$$F = \prod_{\substack{j,k=1 \\ j<k}}^{n} (1 + z_j z_k) \ , \tag{13.10}$$

and so by Cauchy's theorem

$$G(n; d_1, \ldots, d_N) = (2\pi i)^{-n} \int \cdots \int F z_1^{-d_1-1} \cdots z_n^{-d_n-1} dz_1 \cdots dz_n \ , \tag{13.11}$$

where each integral is on a circle centered at the origin. Let all the radii be equal to some $r > 0$. The integrand takes on its maximum absolute value on the product of these circles at precisely the two points $z_1 = z_2 = \cdots = z_n = r$ and $z_1 = z_2 = \cdots = z_n = -r$. If $d_1 = d_2 = \cdots = d_n$, so that we consider only regular graphs, McKay and Wormald [279] show that for an appropriate choice of the radius r, these two points are saddle points of the integrand, and succeed through careful analysis in proving that if dn is even, and $\min(d, n-d-1) > cn(\log n)^{-1}$ for some $c > 2/3$, then

$$G(n, d, d, \ldots, d) = 2^{1/2}(2\pi n \lambda^{d+1}(1-\lambda)^{n-d})^{-n/2} \exp\left(\frac{-1 + 10\lambda - 10\lambda^2}{12\lambda(1-\lambda)} + O(n^{-\zeta})\right) \quad (13.12)$$

as $n \to \infty$ for any $\zeta < \min(1/4, 1/2 - 1/(3c))$, where $\lambda = d/(n-1)$.

McKay and Wormald [279] also succeed in estimating the number of irregular graphs, provided that all the degrees d_j are close to a fixed d that satisfies conditions similar to those above. The proof is more challenging because different radii are used for different variables and the result is complicated to state. ■.

The McKay-Wormald estimate of Example 13.3 is a true tour de force. The problem is that the number of variables is n and so grows rapidly, whereas the integrand grows only like $\exp(cn^2)$ at its peak. More precisely, after transformations that remove obvious symmetries are applied the integrand near the saddle point drops off like $\exp(-n \sum \theta_j^2)$. This is just barely to allow the saddle point method to work, and the symmetries in the problem are exploited to push the estimates through. This approach can be applied to other problems (cf. [278]), but it is hard to do. On the other hand, when the number of variables grows more slowly, multidimensional saddle point contributions can be estimated without much trouble.

So far this section has been devoted primarily to multivariate functions with large singularities. However, there is also an extensive literature on small singularities. The main thread connecting most of these works is that of central and local limit theorems. Bender [32] initiated this development in the setting of two-variable problems. We present some of his results, since they are simpler than the later and more general ones that will be mentioned at the end of this section.

Consider a double sequence of numbers $a_{n,k} \geq 0$. (Usually the $a_{n,k}$ are $\neq 0$ only for $0 \leq k \leq n$.) We will assume that

$$A_n = \sum_k a_{n,k} < \infty \quad (13.13)$$

for all n, and define the normalized double sequence

$$p_n(k) = a_{n,k}/A_n \ . \tag{13.14}$$

We will say that $a_{n,k}$ satisfies a central limit theorem if there exist functions σ_n and μ_n such that

$$\lim_{n\to\infty} \sup_x \left| \sum_{k\le\sigma_n x+\mu_n} p_n(k) - (2\pi)^{-1/2} \int_{-\infty}^{x} \exp(-t^2/2)dt \right| = 0 \ . \tag{13.15}$$

Equivalently, $p_n(k)$ is asymptotically normal with mean μ_n and variance σ_n^2.

Theorem 13.1. [32]. *Let $a_{n,k} \ge 0$, and set*

$$f(z,w) = \sum_{n,k\ge 0} a_{n,k} z^n w^k \ . \tag{13.16}$$

Suppose that there are (i) a function $g(s)$ that is continuous and $\neq 0$ near $s = 0$, (ii) a function $r(s)$ with bounded third derivative near $s = 0$, (iii) an integer $m \ge 0$, and (iv) $\epsilon, \delta > 0$ such that

$$\left(1 - \frac{z}{r(s)}\right)^m f(z, e^s) - \frac{g(z)}{1 - z/r(s)} \tag{13.17}$$

is analytic and bounded for

$$|z| < \epsilon, \quad |z| < |r(0)| + \delta \ . \tag{13.18}$$

Let

$$\mu = -r'(0)/r(0), \quad \sigma^2 = \mu^2 - r''(0)/r(0) \ . \tag{13.19}$$

If $\sigma \neq 0$, then (13.15) holds with $\mu_n = n\mu$ and $\sigma_n^2 = n\sigma^2$.

A central limit theorem is useful, but it only gives information about the cumulative sums of the $a_{n,k}$. It is much better to have estimates for the individual $a_{n,k}$. We say that $p_n(k)$ (and $a_{n,k}$) satisfy a local limit theorem if

$$\lim_{n\to\infty} \sup_x \left| \sigma_n p_n(\lfloor \sigma_n x + \mu_n \rfloor) - (2\pi)^{-1/2} \exp(-x^2/2) \right| = 0 \ . \tag{13.20}$$

In general, we cannot derive (13.20) from (13.15) without some additional conditions on the $a_{n,k}$, such as unimodality (see [32]). The other approach one can take is to derive (13.20) from conditions on the generating function $f(z,w)$.

Theorem 13.2. [32]. *Suppose that* $a_{n.k} \geq 0$*, and let* $f(z,w)$ *be defined by (13.16). Let* $-\infty < a < b < \infty$*. Define*

$$R(\epsilon) = \{z : a \leq Re(z) \leq b,\ |Im(z)| \leq \epsilon\}\ . \tag{13.21}$$

Suppose there exist $\epsilon > 0$*,* $\delta > 0$*, an integer* $m \geq 0$*, and function* $g(s)$ *and* $r(s)$ *such that*

(i) $g(s)$ *is continuous and* $\neq 0$ *for* $s \in R(\epsilon)$*,*

(ii) $r(s) \neq 0$ *and has a bounded third derivative for* $s \in R(\epsilon)$*,*

(iii) for $s \in R(\epsilon)$ *and* $|z| \leq |r(s)|(1+\delta)$*, the function defined by (13.17) is analytic and bounded,*

(iv)

$$\left(\frac{r'(\alpha)}{r(\alpha)}\right)^2 \neq \frac{r''(\alpha)}{r(\alpha)} \quad \mathit{for} \quad a \leq \alpha \leq b\ , \tag{13.22}$$

(v) $f(z, e^s)$ *is analytic and bounded for*

$$|z| \leq |r(Re(s))|(1+\delta) \quad \mathit{and} \quad s \leq |Im(s)| \leq \pi\ .$$

Then

$$a_{n,k} \sim \frac{n^m e^{-\alpha k} g(\alpha)}{m! r(\alpha)^m \sigma_\alpha (2\pi)^{1/2}} \quad \mathit{as} \quad n \to \infty \tag{13.23}$$

uniformly for $a \leq \alpha \leq b$*, where*

$$\frac{k}{n} = -\frac{r'(\alpha)}{r(\alpha)}\ , \tag{13.24}$$

$$\sigma_\alpha^2 = \left(\frac{k}{n}\right)^2 - \frac{r''(\alpha)}{r(\alpha)}\ . \tag{13.25}$$

There have been many further developments of central and local limit theorems for asymptotic enumeration since Bender's original work [32]. Currently the most powerful and general results are those of Gao and Richmond [155]. They apply to general multivariate problems, not only two-variable ones. Other papers that deal with central and local limit theorems or other multivariate problems with small singularities are [38, 42, 65, 96, 142, 143, 183, 227].

14. Mellin and other integral transforms

When the best generating function that one can obtain is an infinite sum, integral transforms can sometimes help. There is a large variety of integral transforms, such as those of

Fourier and Laplace. The one that is most commonly used in asymptotic enumeration and analysis of algorithms is the Mellin transform, and it is the only one we will discuss extensively below. The other transforms do occur, though. For example, if $f(x) = \sum a_n x^n / n!$ is an exponential generating function of the sequence a_n, then the ordinary generating function of a_n can be derived from it using the Laplace transform

$$\begin{aligned} \int_0^\infty f(xy) \exp(-x) dx &= \sum_n a_n y^n (n!)^{-1} \int_0^\infty x^n \exp(-x) dx \\ &= \sum_n a_n y^n \ . \end{aligned} \tag{14.1}$$

(This assumes that the a_n are small enough to assure the integrals above converge and the interchange of summation and integration is valid.) Related integral transforms can be used to transform generating functions into other forms. For example, to transform an ordinary generating function $F(u) = \sum a_n u^n$ into an exponential one, we can use

$$\frac{1}{2\pi i} \int_{|u|=r} F(u) \exp(w/u) du \ . \tag{14.2}$$

The basic references for asymptotics of integral transforms are [89, 95, 299, 347]. This section will only highlight some of the main properties of Mellin transforms and illustrate how they are used. For a more detailed survey, especially to analysis of algorithms, see [137].

Let $f(t)$ be a measurable function defined for real $t \geq 0$. The *Mellin transform* $f^*(z)$ of $f(t)$ is a function of the complex variable z defined by

$$f^*(z) = \int_0^\infty f(t) t^{z-1} dt \ . \tag{14.3}$$

If $f(t) = O(t^\alpha)$ as $t \to 0^+$ and $f(t) = O(t^\beta)$ as $t \to \infty$, then the integral in (14.3) converges and defines $f^*(z)$ to be an analytic function inside the "fundamental domain" $-\alpha < \text{Re}(z) < -\beta$. As an example, for $f(t) = \exp(-t)$, we have $f^*(z) = \Gamma(z)$ and $\alpha = 0$, $\beta = -\infty$. There is an inversion formula for Mellin transforms which states that

$$f(t) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} f^*(z) t^{-z} dz \ , \tag{14.4}$$

and the integral is over the vertical line with $\text{Re}(z) = c$. The inversion formula (14.4) is valid for $-\alpha < c < -\beta$, but much of its strength in applications comes from the ability to shift the contour of integration into wider domains to which $f^*(z)$ can be analytically continued.

The advantage of the Mellin transform is due largely to a simple property, namely that if $g(t) = af(bx)$ for b real, $b > 0$, then

$$g^*(z) = ab^{-z} f^*(z) \ . \tag{14.5}$$

This readily extends to show that if

$$F(t) = \sum_k \lambda_k f(\eta_k t) \tag{14.6}$$

(where the λ_k and $\eta_k > 0$ are such that the sum converges and $F(t)$ is well behaved), then

$$F^*(z) = \left(\sum_k \lambda_k \eta_k^{-z} \right) f^*(z) \ . \tag{14.7}$$

In particular, if

$$F(t) = \sum_{k=1}^{\infty} f(kt) \ , \tag{14.8}$$

then

$$F^*(z) = \left(\sum_{k=1}^{\infty} k^{-z} \right) f^*(z) = \zeta(z) f^*(z) \ , \tag{14.9}$$

where $\zeta(z)$ is the Riemann zeta function.

Example 14.1. *Runs of heads in coin tosses.* What is R_n, the expected length of the longest run of heads in n tosses of a fair coin? Let $p(n,k)$ be the probability that there is no run of k heads in a coin tosses. Then

$$R_n = \sum_{k=1}^{n} k(p(n, k+1) - p(n,k)) \ . \tag{14.10}$$

We now apply the estimates of Example 9.2. To determine $p(n,k)$, we take $A = 00 \cdots 0$, and then $C_A(z) = z^{k-1} + z^{k-2} + \cdots + z + 1$, so $C_A(1/2) = 1 - 2^{-k}$. Hence (9.19) shows easily that in the important ranges where k is of order $\log n$, we have

$$p(n,k) \cong \exp(-n2^{-k}) \ , \tag{14.11}$$

and there R_n is approximated well by

$$r(n) = \sum_{k=0}^{\infty} k(\exp(-n2^{-k-1}) - \exp(-n2^{-k})) \ . \tag{14.12}$$

The function $r(t)$ is of the form (14.6) with

$$\lambda_k = k, \quad \eta_k = 2^{-k}, \quad f(t) = \exp(-t/2) - \exp(-t) \ , \tag{14.13}$$

is easily seen to be well behaved, and so for $-1 < \mathrm{Re}(z) < 0$,

$$r^*(z) = \left(\sum_{k=0}^{\infty} k 2^{kz} \right) f^*(z) = 2^z (1 - 2^z)^{-2} f^*(z) \ . \tag{14.14}$$

Next, to determine $f^*(z)$, we note that for $\mathrm{Re}(z) > 0$ we have

$$\begin{aligned} f^*(z) &= \int_0^\infty f(t)t^{z-1}dt = \int_0^\infty e^{-t/2}t^{z-1}dt - \int_0^\infty e^{-t}t^{z-1}dt \\ &= (2^z - 1)\Gamma(z) \ . \end{aligned} \tag{14.15}$$

By analytic continuation this relation holds for $-1 < \mathrm{Re}(z)$, and we find that for $-1 < \mathrm{Re}(z) < 0$,

$$r^*(z) = 2^z(2^z - 1)^{-1}\Gamma(z) \ . \tag{14.16}$$

We now apply the inversion formula to obtain

$$r(t) = \frac{1}{2\pi i}\int_{-1/2-i\infty}^{-1/2+i\infty} 2^z(2^z-1)^{-1}\Gamma(z)t^{-z}dz \ . \tag{14.17}$$

The integrand is a meromorphic function in the whole complex plane that drops off rapidly on any vertical line. We move the contour of integration to the line $\mathrm{Re}(z) = 1$. The new integral is $O(t^{-1})$, and the residues at the poles (all on $\mathrm{Re}(z) = 0$) will give the main contribution to $r(t)$. There are first order poles at $z = 2\pi i m \log 2$ for $m \in \mathbb{Z} \setminus \{0\}$ coming from $2^z = 1$, and a single second order pole at $z = 0$, since $\Gamma(z)$ has a first order pole there as well. A short computation of the residues gives

$$r(t) = \log_2 t - \sum_{h=-\infty}^{\infty} (\log 2)^{-1}\Gamma(-2\pi i h(\log 2)^{-1})\exp(2\pi i h \log_2 t) + O(t^{-1}) \ . \tag{14.18}$$

■

There are other ways to obtain the same expansion (14.18) for $r(t)$ (cf. [181]). The periodic oscillating component in $r(t)$ is common in problems involving recurrences over powers of 2. This happens, for example, in studies of register allocation and digital trees [136, 138, 141]. The periodic function is almost always the same as the one in Eq. (14.18), even when the combinatorics of the problem varies. Technically this is easy to explain, because of the closely related recurrences leading to similar Mellin transforms for the generating functions.

Mellin transforms are useful in dealing with problems that combine combinatorial and arithmetic aspects. For example, if $S(n)$ denotes the total number of 1's in the binary representations of $1, 2, \ldots, n-1$, then it was shown by Delange that

$$S(n) = \frac{1}{2}n\log_2 n + nu(\log_2 n) + o(n) \quad \text{as} \quad n \to \infty \ , \tag{14.19}$$

where $u(x)$ is a continuous, nowhere differentiable function that satisfies $u(x) = u(x+1)$. The Fourier coefficients of $u(x)$ are known explicitly. Perhaps the best way to obtain these results is by using Mellin transforms. See [129, 353] for further information and references.

Mellin transforms are often combined with other techniques. For example, sums of the form $s_n = \sum a_k \binom{n}{k}$ with oscillating a_k lead to generating functions

$$s(z) = \sum_k a_k w(z)^k \ . \tag{14.20}$$

The asymptotic behavior of $s(z)$ near its dominant singularity can sometimes be determined by applying Mellin transforms. For a detailed explanation of the approach, see [137]. Examples of the application of this technique can be found in [13, 280].

15. Functional equations, recurrences, and combinations of methods

Most asymptotic enumeration results are obtained from combinations of techniques presented in the previous sections. However, it is only rarely that the basic asymptotic techniques can be applied directly. This section describes a variety of methods and results that are not easy to categorize. They use combinations of methods that have been presented before, and sometimes develop them further. In most of the examples that will be presented, some relations for generating functions are available, but no simple closed-form formulas, and the problem is to deduce where the singularities lie and how the generating functions behave in their neighborhoods. Once that task is done, previous methods can be applied to obtain asymptotics of the coefficients.

15.1. Implicit functions, graphical enumeration, and related topics

Example 15.1. *Rooted unlabeled trees.* We sketch a proof that T_n, the number of rooted unlabeled trees with n vertices, satisfies the asymptotic relation (1.9). The functional equation (1.8) holds with $T(z)$ regarded as a formal power series. The first step is to show that $T(z)$ is analytic in a neighborhood of 0. This can be done by working exclusively with Eq. (1.8). (There is an argument of this type in Section 9.5 of [188].) Another way to prove analyticity of $T(z)$ is to use combinatorics to obtain crude upper bounds for T_n. We use a combination of these approaches. If a tree with $n \geq 2$ vertices has at least two subtrees at the root, we can decompose it into two trees, the first consisting of one subtree at the root, the other of the root and the remaining subtrees. This shows that

$$T_n \leq T_{n-1} + \sum_{k=1}^{n-1} T_k T_{n-k} \ , \qquad n \geq 2 \ . \tag{15.1}$$

Therefore, if we define $a_1 = 1$, and

$$a_n = a_{n-1} + \sum_{k=1}^{n-1} a_k a_{n-k} \ , \qquad n \geq 2 \ , \tag{15.2}$$

then we have $T_n \leq a_n$. Now if

$$A(z) = \sum_{n=1}^{\infty} a_n z^n \ ,$$

then the defining relation (15.2) yields the functional equation

$$A(z) - z = zA(z) + A(z)^2 \ , \tag{15.3}$$

so that

$$A(z) = (1 - z - (1 - 6z + z^2)^{1/2})/2 \ . \tag{15.4}$$

Since $A(z)$ is analytic in $|z| < 3 - 2\sqrt{2} = 0.17157\ldots$, we have

$$0 \leq T_n \leq a_n = O(6^n) \ . \tag{15.5}$$

It will also be convenient to have an exponential lower bound for T_n. Let b_n be the number of rooted unlabeled trees in which every internal vertex has ≤ 2 subtrees. Then $b_1 = 1$, $b_2 = 1$, and

$$b_n \geq \sum_{k=1}^{\lfloor (n-1)/2 \rfloor} b_k b_{n-k-1} \quad \text{for} \quad n \geq 3 \ . \tag{15.6}$$

We use this to show that $b_n \geq (6/5)^n$ for $n \geq 7$. Direct computation establishes this lower bound for $7 \leq n \leq 14$, and for $n \geq 15$ we use induction and $b_n \geq b_k b_{n-k-1}$ with $k = \lfloor (n-1)/2 \rfloor$.

Since $T_n \geq b_n \geq (6/5)^n$, $T(z)$ converges only in $|z| < r$ for some r with $r < 1$. Since $T(0) = 0$, $|T(z)| \leq C_\delta |z|$ in $|z| \leq r - \delta$ for every $\delta > 0$, and therefore

$$u(z) = \sum_{k=2}^{\infty} T(z^k)/k \tag{15.7}$$

is analytic in $|z| < r^{1/2}$, and in particular at $z = r$. Therefore, although we know little about r and $u(z)$, we see that $T(z)$ satisfies $G(z, T(z)) = T(z)$, where

$$G(z, w) = z \exp(w + u(z)) \tag{15.8}$$

is analytic in z and w for all w and for $|z| < r^{1/2}$.

We will apply Theorem 10.6. First, though, we need to establish additional properties of $T(z)$. We have

$$T(z) \exp(-T(z)) = z \exp(u(z)) \to r \exp(u(r)) \quad \text{as} \quad z \to r^- \ , \tag{15.9}$$

and $0 < r\exp(u(r)) < \infty$. Since $T(z)$ is positive and increasing for $0 < z < r$, $T(r)$, the limit of $T(z)$ as $z \to r^-$ must exist and be finite.

We next show that $T(r) = 1$. We have

$$\frac{\partial}{\partial w} G(z, w) = G(z, w) \ . \tag{15.10}$$

We know that $G(z, T(z)) = T(z)$ for $|z| < r$, and in particular for some z arbitrarily close to r. If $T(r) \neq 1$, then by (15.10)

$$\frac{\partial}{\partial w}(G(z, w) - w)\bigg|_{w=T(z)} \neq 0 \tag{15.11}$$

in a neighborhood of $z = r$, and therefore $T(z)$ could be continued analytically to a neighborhood of $z = r$. This is impossible, since r is the radius of convergence of $T(z)$, and $T_n \geq 0$ implies by Theorem 10.3 that $T(z)$ has a singularity at $z = r$. Therefore we must have $T(r) = 1$, and $G_w(r, T(r)) = 1$.

We have now shown that conditions (i) and (ii) of Theorem 10.6 hold with the r of that theorem the same as the r we have defined and $s = T(r) = 1$, $\delta = r^{1/2} - r$. Condition (iii) is easy to verify. Finally, the conditions on the coefficients of $T(z)$ and $G(z, w)$ are clearly satisfied.

Since Theorem 10.6 applies, we do obtain an asymptotic expansion for T_n of the form (1.9), with C given by the formula (10.64). It still remains to determine r and C. No closed-form expressions are known for these constants. They are conjectured to be transcendental and algebraically independent of standard constants such as π and e, but no proof is available. Numerically, however, they are simple to compute. Note that

$$\begin{aligned} G_z(r, 1) &= \exp(1 + u(r))(1 + ru'(r)) \\ &= r^{-1} + u'(r) \ , \end{aligned} \tag{15.12}$$

$$G_{ww}(r, 1) = 1 \ , \tag{15.13}$$

so we only need to compute r and $u'(r)$. These quantities can be computed along with $u(r)$ in the same procedure. The basic numerical procedure is to determine r as the positive solution to $T(r) = 1$. To determine $T(x)$ for any positive x, we take any approximation to the $T(x^k)$, $k \geq 1$ (starting initially with x^k as an approximation to $T(x^k)$, say), and combine it with (1.8) (applied with $z = x^m$, $m \geq 1$) to obtain improved approximations. This procedure can be made rigorous. Upper bounds for r, $u(r)$, and $u'(r)$ are especially easy. Since $T_1 = 1$, $T(x) \geq x$

for $0 < x < 1$, and therefore, $T(x^k) \geq x^k$ for $k \geq 1$. Suppose that we start with a fixed value of x and derive some lower bounds of the form $T(x^k) \geq u_k^{(1)} \geq 0$ for $k \geq 1$. Then the functional equation (1.8) implies

$$T(x^m) \geq u_m^{(2)} = x \exp\left(\sum_{k=1}^{\infty} u_{km}/k\right) \qquad m \geq 1 \ . \tag{15.14}$$

This process can be iterated several more times, and to keep the computation manageable, we can always set $u_k^{(j)} = 0$ for $k \geq k_0$. If we ever find a lower bound $T(x) > 1$ by this process, then we know that $r < x$, since $T(r) = 1$. Lower bounds for r are slightly more complicated. ■

We mention here that if U_n denotes the number of unlabeled trees, then the ordinary generating function $U(z) = \sum U_n z^n$ satisfies

$$U(z) = T(z) - T(z)^2/2 + T(z^2)/2 \ . \tag{15.15}$$

Using the results from Example 15.1 about the analytic behavior of $T(z)$, it can be shown that

$$U_n \sim C' r^{-n} n^{-5/2} \ , \tag{15.16}$$

where $r = 0.3383219\ldots$ is the same as before, while $C' = 0.5349485\ldots$.

Example 15.2. *Leftist trees.* Let a_n denote the number of leftist trees of size n (i.e., rooted planar trees with n leaves, such that in any subtree S, the leaf nearest to the root of S is in the right subtree of S [237]). Then $a_1 = a_2 = a_3 = 1$, $a_4 = 2$, $a_5 = 4$. No explicit formula for a_n is known. Even the recurrences for the a_n are complicated, and involve auxiliary sequences. If

$$f(z) = \sum_{n=1}^{\infty} a_n z^n \tag{15.17}$$

denotes the ordinary generating function of a_n, then the combinatorially derived recurrences for the a_n show that [224]

$$f(z) = z + \frac{1}{2} f(z)^2 + \frac{1}{2} \sum_{m=1}^{\infty} g_m(z)^2 \ , \tag{15.18}$$

where the auxiliary generating functions $g_m(z)$ (which enumerate leftist trees with the leftmost leaf at distance $m-1$ from the root) satisfy

$$g_1(z) = z, \quad g_2(z) = zf(z), \quad g_{m+1}(z) = g_m(z) \left[f(z) - \sum_{j=1}^{m-1} g_j(z) \right] , \quad m \geq 2 \ , \tag{15.19}$$

and

$$f(z) = \sum_{m=1}^{\infty} g_m(z) \ . \tag{15.20}$$

These generating function relations might not seem promising. If r is the smallest singularity of $f(z)$, then $\sum g_m(z)^2$ is not analytic at r, so we cannot apply Theorem 10.6 in the way it was used in Example 15.1. However, Kemp [224] has sketched a proof that the analytic behavior of $f(z)$ is of the same type as that involved in functions covered by Theorem 10.6, so that it has a dominant square root singularity, and therefore

$$a_n = \alpha c^n n^{-3/2} + O(c^n n^{-5/2}) \ , \tag{15.21}$$

where

$$\alpha = 0.250363429 \ \ldots, \quad c = 2.749487902 \ \ldots \ . \tag{15.22}$$

The constants α and c are not known explicitly in terms of other standard numbers such as π or e, but they can be computed efficiently. The $\alpha c^n n^{-3/2}$ term in (15.21) gives an approximation to a_n that is accurate to within 4% for $n = 10$, and within 0.4% for $n = 100$. Thus asymptotic methods yield an approximation to a_n which is satisfactory for many applications. Further results about leftist trees can be found in [225]. ■

15.2. Nonlinear iteration and tree parameters

Example 15.3. *Heights of binary trees.* A binary tree [DEK] is a rooted tree with unlabeled nodes, in which each node has 0 or 2 successors, and left and right successors are distinguished. The size of a binary tree is the number of internal nodes, i.e., the number of nodes with two successors. We let B_n denote the number of binary trees of size n, so that $B_0 = 1$ (by convention), $B_1 = 1$, $B_2 = 2$, $B_3 = 5, \ldots$. Let

$$B(z) = \sum_{n=0}^{\infty} B_n z^n \ . \tag{15.23}$$

Since each nonempty binary tree consists of the root and two binary trees (the left and right subtrees), we obtain the functional equation

$$B(z) = 1 + zB(z)^2 \ . \tag{15.24}$$

This implies that

$$B(z) = \frac{1 - (1 - 4z)^{1/2}}{2z} \ , \tag{15.25}$$

so that

$$B_n = \frac{1}{n+1}\binom{2n}{n} , \tag{15.26}$$

and the B_n are the Catalan numbers. The formula (4.4) (easily derivable from Stirling's formula (4.1)) shows that

$$B_n \sim \pi^{-1/2} n^{-3/2} 4^n \quad \text{as} \quad n \to \infty . \tag{15.27}$$

The height of a binary tree is the number of nodes along the longest path from the root to a leaf. The distribution of heights in binary trees of a given size does not have exact formulas like that of (15.26) for the number of binary trees of a given size. There are several problems on heights that have been answered only asymptotically, and with varying degrees of success. The most versatile approach is through recurrences on generating functions. Let $B_{h,n}$ be the number of binary trees of size n and height $\le h$, and let

$$b_h(z) = \sum_{n=0}^{\infty} B_{h,n} z^n . \tag{15.28}$$

Then

$$b_0(z) = 0, \quad b_1(z) = 1 , \tag{15.29}$$

and an extension of the argument that led to the relation (15.24) yields

$$b_{h+1}(z) = 1 + z b_h(z)^2 , \quad h \ge 0 . \tag{15.30}$$

The $b_h(z)$ are polynomials in z of degree $2^{h-1} - 1$ for $h \ge 1$. Unfortunately there is no simple formula for them like Eq. (15.25) for $B(z)$, and one has to work with the recurrence (15.30) to obtain many of the results about heights of binary trees. Different problems involve study of the recurrence in different ranges of values of z, and the behavior of the recurrence varies drastically.

For any fixed z with $|z| \le 1/4$, $b_h(z) \to B(z)$ as $h \to \infty$. For $|z| > 1/4$ the behavior of $b_h(z)$ is more complicated, and is a subject of of nonlinear dynamics [91]. (It is closely related to the study of the Mandelbrot set.) For any real z with $z > 1/4$, $b_h(z) \to \infty$ as $h \to \infty$. To study the distribution of the $B_{h,n}$ as n varies for h fixed, but large, it is necessary to investigate this range of rapid growth. It can be shown [133] that for any λ_1 and λ_2 with $0 < \lambda_1 < \lambda_2 < 1/2$,

$$B_{h,n} = \frac{\exp(2^{h-1}(\beta(r) - r\beta'(r)\log r))}{2^{(h-1)/2}(2\pi(r^2\beta''(r) + r\beta'(r)))^{1/2}}(1 + O(2^{-h/2})) \tag{15.31}$$

uniformly as $h, n \to \infty$ with

$$\lambda_1 < n/2^h < \lambda_2 \ , \tag{15.32}$$

where the function $\beta(x)$ is defined for $1/4 < x < \infty$ by

$$\beta(x) = \log x + \sum_{j=1}^{\infty} 2^{-j} \log \left(1 + \frac{1}{b_j(x) - 1} \right) \ , \tag{15.33}$$

and r is the unique solution in $(1/4, \infty)$ to

$$r\beta'(r) = n2^{-h+1} \ . \tag{15.34}$$

The formula (15.31) might appear circular, in that it describes the behavior of the coefficients $\beta_{h,n}$ of the polynomial $b_h(z)$ in terms of the function $\beta(z)$, which is defined by $b_h(z)$ and all the other $b_j(z)$. However, the series (15.33) for $\beta(z)$ converges rapidly, so that only the first few of the $b_h(z)$ matter in obtaining approximate answers, and computation using (15.33) is efficient. The function $\beta(z)$ is analytic in a region containing the real half-line $x > 1/4$, so the behavior of the $B_{h,n}$ is smooth. It is also known [133] that the behavior of $B_{h,n}$ as a function of n is Gaussian near the peak, which occurs at $n \sim 2^{h-1} \cdot 0.628968 \ldots$. The distribution of $B_{h,n}$ is not Gaussian throughout the range (15.32), though.

The proof of the estimate (15.31) is derived from the estimate

$$b_h(z) = \exp(2^{h-1}\beta(z) - \log z)(1 + O(\exp(-\epsilon 2^h))) \ , \tag{15.35}$$

valid in a region along the half-axis $x > 1/4$. The estimates for the coefficients $B_{h,n}$ are obtained by applying the saddle point method. Because of the doubly-exponential rate of growth of $b_h(z)$ for z close to the real axis, it is easy to show that on the circle of integration, the region away from the real axis contributes a negligible amount to $B_{h,n}$. The relation (15.35) is sufficient, together with the smoothness properties of $\beta(z)$, to estimate the contribution of the integral near the real axis. To prove (15.35), one proceeds as in Example 9.7. However, greater care is required because of the complex variables that occur and the need for estimates that are uniform in the variables. The basic recurrence (15.30) shows that

$$\begin{aligned} \log b_{h+1}(z) &= 2 \log b_h(z) + \log z + \log \left(1 + \frac{1}{z b_h(z)^2} \right) \\ &= 2 \log b_h(z) + \log z + \log \left(1 + \frac{1}{b_{h+1}(z) - 1} \right) \ . \end{aligned} \tag{15.36}$$

Iterating this relation, we find that for $h \geq 1$,

$$\begin{aligned} \log b_{h+1}(z) &= 2^{h+1} \log b_1(z) + (2^h - 1) \log z + \sum_{k=0}^{h-1} 2^k \log \left(1 + \frac{1}{b_{h+1-k}(z) - 1}\right) \\ &= 2^h \left\{ \log z + \sum_{j=1}^{h+1} 2^{-j} \log \left(1 + \frac{1}{b_j(z) - 1}\right) \right\} - \log z \ . \end{aligned} \tag{15.37}$$

The basic equation (15.35) then follows. The technical difficulty is in establishing rigorous bounds for the error terms in the approximations. Details are presented in [133].

Most of the binary trees of a given height h are large, with about $0.3 \cdot 2^h$ internal nodes. This might give the misleading impression that most binary trees are close to the full binary tree of a similar size. However, if we consider all binary trees of a given size n, the average height is on the order of $n^{1/2}$, so that they are far from the full balanced binary trees. The methods that are used to study the average height are different from those used for trees of a fixed height. The basic approach of [133] is to let

$$H_n = \sum_{\substack{T \\ |T|=n}} \mathrm{ht}(T) \ ,$$

where the sum is over the binary trees T of size n, and $\mathrm{ht}(T)$ is the height of T. Then the average height is just H_n/B_n.

The generating function for the H_n is

$$H(z) = \sum_{n=0}^{\infty} H_n z^n = \sum_{h \geq 0} (B(z) - b_h(z)) \ , \tag{15.38}$$

and the analysis of [133] proceeds by investigating the behavior of $H(z)$ in a wedge-shaped region of the type encountered in Section 11.1. If we let

$$\epsilon(z) = (1 - 4z)^{1/2} \ , \tag{15.39}$$

$$e_h(z) = (B(z) - b_h(z))/(2B(z)) \ , \tag{15.40}$$

then the recurrence (15.30) yields

$$e_{h+1}(z) = (1 - \epsilon(z)) e_h(z) (1 - e_h(z)) \ , \quad e_0(z) = 1/2 \ . \tag{15.41}$$

Extensive analysis of this relation yields an approximation to $e_h(z)$ of the form

$$e_h(z) \approx \frac{\epsilon(z)(1 - \epsilon(z))^h}{1 - (1 - \epsilon(z))^h} \ , \tag{15.42}$$

valid for $|\epsilon(z)|$ sufficiently small, $|\text{Arg } \epsilon(z)| < \pi/4 + \delta$ for a fixed $\delta > 0$. (The precise error terms in this approximation are complicated, and are given in [133].) This then leads to an expansion for $H(z)$ in a sector $|z - 1/4| < \alpha$, $\pi/2 - \beta < |\text{Arg}(z - 1/4)| < \pi/2 + \beta$ of the form

$$H(z) = -2\log(1 - 4z) + K + O(|1 - 4z|^v) \ , \tag{15.43}$$

where v is any constant, $v < 1/4$, and K is a fixed constant. Transfer theorems of Section 11.1 now yield the asymptotic estimate

$$H_n \sim 2n^{-1}4^n \quad \text{as} \quad n \to \infty \ . \tag{15.44}$$

When we combine (15.44) with (15.27), we obtain the desired result that the average height of a binary tree of size n is $\sim 2(\pi n)^{1/2}$ as $n \to \infty$.

Distribution results about heights of binary trees can be obtained by investigating the generating functions

$$\sum_{h \geq 0} h^r (B(z) - b_h(z)) \ . \tag{15.45}$$

This procedure, carried out in [133] by using modifications of the approach sketched above for the average height, obtains asymptotics of the moments of heights. The method mentioned in Section 6.5 then leads to a determination of the distribution. However, the resulting estimates do not say much about heights far away from the mean. A more careful analysis of the behavior of $e_h(z)$ can be used [126] to show that if $x = h/(2n^{1/2})$, then

$$\frac{B_{h,n} - B_{h-1,n}}{B_n} \sim 2xn^{-1/2} \sum_{m=1}^{\infty} m^2(2m^2x^2 - 3)e^{-m^2x^2} \tag{15.46}$$

as $n, h \to \infty$, uniformly for $x = o((\log n)^{1/2})$, $x^{-1} = o((\log n)^{1/2})$.

For extremely small and large heights, different methods are used. It follows from [126] that

$$\frac{B_{h,n} - B_{h-1,n}}{B_n} \leq \exp(-c(h^2/n + n/h^2)) \tag{15.47}$$

for a constant $c > 0$, which shows that extreme heights are infrequent. (The estimates in [126] are more precise than (15.47).) Bounds of the above form for small heights are obtained in [126] by studying the behavior of the $b_h(z)$ almost on the boundary between convergence and divergence, using the methods of [399]. Let x_h be the unique positive root of $b_h(z) = 2$. Note that $B(1/4) = 2$, and each coefficient of the $b_h(z)$ is nondecreasing as $h \to \infty$. Therefore $x_2 > x_3 > \cdots > 1/4$. More effort shows [126] that x_h is approximately $1/4 + \alpha h^{-2}$ for a certain

$\alpha > 0$. This leads to an upper bound for $B_{h,n}$ by Lemma 8.1. Bounds for trees of large heights are even easier to obtain, since they only involve upper bounds for the $b_h(z) - b_{h-1}(z)$ inside the disk of convergence $|z| < 1/4$. ■

In addition to the methods of [132, 133, 126] that were mentioned above, there are also other techniques for studying heights of trees, such as those of [60, 331]. However, there are problems about obtaining fully rigorous proofs that way. (See the remarks in [126] on this topic.) Most of these methods can be extended to study related problems, such as those of diameters of trees [357].

The results of Example 15.3 can be extended to other families of trees (cf. [132, 133, 126]). What matters in obtaining results such as those of the above example are the form of the recurrences, and especially the positivity of the coefficients.

Example 15.4. *Enumeration of 2,3-trees* [300]. Height-balanced trees satisfy different functional equations than unrestricted trees, which results in different analytic behavior of the generating functions, and different asymptotics. Consider $2,3$-trees; i.e., rooted, oriented trees such that each nonleaf node has either two or three successors, and in which all root-to-leaf paths have the same length. If a_n is the number of $2,3$-trees with exactly n leaves, then $a_1 = a_2 = a_3 = a_4 = 1$, $a_5 = 2, \ldots$, and the generating function

$$f(z) = \sum_{n=1}^{\infty} a_n z^n \tag{15.48}$$

satisfies the functional equation

$$f(z) = z + f(z^2 + z^3) \ . \tag{15.49}$$

Iteration of the recurrence (15.49) leads to

$$f(z) = \sum_{k=0}^{\infty} Q_k(z) \ , \tag{15.50}$$

where $Q_0(z) = z$, $Q_{k+1}(z) = Q_k(z^2 + z^3)$, provided the series (15.50) converges. The Taylor series (15.48) converges only in $|z| < \phi^{-1}$, where $\phi = (1 + 5^{1/2})/2$ is the "golden ratio." Study of the polynomials $Q_k(z)$ shows that the expansion (15.50) converges in a region

$$D = \{z : |z| < \phi^{-1} + \delta, \ |\mathrm{Arg}(z - \phi^{-1})| > \pi/2 - \epsilon\} \tag{15.51}$$

for certain δ, $\epsilon > 0$, and that inside D,

$$f(z) = -c \log(\phi^{-1} - z) + w(\log(\phi^{-1} - z)) + O(|\phi^{-1} - z|) \ , \tag{15.52}$$

where $c = [\phi \log(4 - \phi)]^{-1}$, and $w(t)$ is a nonconstant function, analytic in a strip $|\mathrm{Im}\,(t)| < \eta$ for some $\eta > 0$, such that $w(t + \log(4 - \phi)) = w(t)$. The expression (15.52) only has to be proved in a small vicinity of ϕ^{-1} (intersected with D, of course). Since

$$Q(\phi^{-1} + \nu) = \phi^{-1} + (4 - \phi)\nu + O(|\nu|^2) \tag{15.53}$$

(so that ϕ^{-1} is a repelling fixed point of Q), behavior like that of (15.52) is to be expected, and with additional work can be rigorously shown to hold. Once the expansion (15.52) is established, singularity analysis techniques can then be applied to deduce that

$$a_n \sim \frac{\phi^n}{n} u(\log n) \quad \text{as} \quad n \to \infty \ , \tag{15.54}$$

where $u(t)$ is a positive nonconstant continuous function that satisfies $u(t) = u(t + \log(4 - \phi))$, and has mean value $(\phi \log(4 - \phi))^{-1}$. For details, see [300].

The same methods can be applied to related families of trees, such as those of B-trees. ■

The results of Example 15.3 and the generalizations mentioned above all apply only to the standard counting models, in which all trees with a fixed value of some simple property, such as size or height, are equally likely. Often, especially in computer science applications, it is necessary to study trees produced by some algorithm, and consider all outputs of this algorithm as equally likely. For example, in sorting it is natural to consider all permutations of n elements as equally probable. If random permutations are used to construct binary search trees, then the distribution of heights will be different from that in the standard model, and the two trees of maximal height will have probability of $2/n!$ of occurring. The average height turns out to be $\sim c \log n$ as $n \to \infty$, for $c = 4.311\ldots$ a certain constant given as a solution to a transcendental equation. This was shown by Devroye [92] (see also [93]) by an application of the theory of branching processes. For a detailed exposition of this method and other applications to similar problems, see [270]. The basic generating function approach that we have used in most of this chapter leads to functional iterations which have not been solved so far.

15.3. Differential and integral equations

Section 9.2 showed that differential equations arise naturally in analyzing linear recurrences of finite order with rational coefficients. There are other settings when they arise even more naturally. As is true of nonlinear iterations in the previous section and the functional equations of the next one, differential and integral equations are typically used to extract information about singularities of generating functions. We have already seen in Example 9.3 and other cases that differential equations can yield an explicit formula for the generating function, from which it is easy to deduce what the singularities are and how they affect the asymptotics of the coefficients. Most differential equations do not have a closed-form solution. However, it is often still possible to derive the necessary information about analytic behavior even when there is no explicit formula for the solution. We demonstrate this with a brief sketch of a recent analysis of this type [131]. Other examples can be found in [270].

Example 15.5. *Search costs in quadtrees* [131]. Quadtrees are a well-known data structure for multidimensional data storage [168]. Consider a d-dimensional data space, and let n points be drawn independently from the uniform distribution in the d-dimensional unit cube. We take d fixed and $n \to \infty$. Suppose that the first $n-1$ points have already been inserted into the quadtree, and let D_n be the search cost (defined as the number of internal nodes traversed) in inserting the n-th item. The result of Flajolet and Lafforgue [131] is that D_n converges in distribution to a Gaussian law when $n \to \infty$. If μ_n and σ_n denote the mean and standard deviation of D_n, respectively, then

$$\mu_n \sim 2d^{-1} \log n, \quad \sigma_n \sim d^{-1}(2 \log n)^{1/2} \quad \text{as} \quad n \to \infty \ , \tag{15.55}$$

and for all real $\alpha < \beta$, as $n \to \infty$,

$$Pr(\alpha\sigma_N < D_n - \mu_n < \beta\sigma_n) \sim (2\pi)^{-1/2} \int_\alpha^\beta \exp(-x^2/2) dx \ . \tag{15.56}$$

The results for μ_n and σ_n had been known before, and required much simpler techniques for their solution, see [270]. It was only necessary to study asymptotics of ordinary differential equations in a single variable. To obtain distribution results for search costs, it was necessary to study bivariate generating functions. The basic relation is

$$\sum_k Pr\{D_n = k\} u^k = (2^d u - 1)^{-1} (\phi_n(u) - \phi_{n-1}(u)) \ , \tag{15.57}$$

where the polynomials $\phi_n(u)$ have the bivariate generating function

$$\Phi(u,z) = \sum_{n=0}^{\infty} \phi_n(u) z^n \ . \tag{15.58}$$

which satisfies the integral equation

$$\begin{aligned} \Phi(u,z) = 1 + 2^d u \int_0^z \frac{dx_1}{x_1(1-x_1)} \int_0^{x_1} \frac{dx_2}{x_2(1-x_2)} \int_0^{x_2} \frac{dx_3}{x_3(1-x_3)} \cdots \\ \int_0^{x_{d-2}} \frac{dx_{d-1}}{x_{d-1}(1-x_{d-1})} \int_0^{x_{d-1}} \Phi(u,x_d) \frac{dx_d}{1-x_d} \ . \end{aligned} \tag{15.59}$$

This integral equation can easily be reduced to an equivalent differential equation, which is what is used in the analysis. For $d = 1$ there is an explicit solution

$$\Phi(u,z) = (1-z)^{-2u} \ , \tag{15.60}$$

which shows that D_n can be expressed in terms of Stirling numbers. This is not surprising, since for $d = 1$ the quadtree reduces to the binary search tree, for which these results were known before. For $d = 2$, $\Phi(u,z)$ can be expressed in terms of standard hypergeometric functions. However, for $d \geq 3$ there do not seem to be any explicit representations of $\Phi(u,z)$. Flajolet and Lafforgue use a singularity perturbation method to study the behavior of $\Phi(u,z)$. They start out with the differential system derivable in standard way from the differential equation associated to (15.59) (i.e., a system of d linear differential equations in z with coefficients that are rational in z). Since only values of u close to 1 are important for the distribution results, they regard u as a perturbation parameter of this system. For every fixed u, they determine the dominant singularity of the linear differential system in the variable z, using the indicial equations that are standard in this setting. It turns out that the dominant singularity is a regular one at $z = 1$, and

$$\Phi(u,z) \approx c(u)(1-z)^{-2u^{1/d}} \ , \tag{15.61}$$

at least for z and u close to 1. This behavior of $\Phi(u,z)$ is then used (in its more precise form, with explicit error terms) to deduce, through the transfer theorem methods explained in Section 11, the behavior of $\phi_n(u)$:

$$\phi_n(u) \approx c(u)\Gamma(2u^{1/d})^{-1} n^{2u^{1/d}-1} \ . \tag{15.62}$$

This form, again in a more precise formulation, is then used to deduce that the behavior of D_n is normal near its peak, and that the tails of the distribution are small. ■

15.4. Functional equations

One area that needs and undoubtedly will receive much more attention is that of complicated nonlinear relations for generating functions. Even in a single variable our knowledge is limited. Some of the work of Mahler [267, 268, 269], devoted to functions $f(z)$ satisfying equations of the form $p(f(z), f(z^g)) = 0$, where $p(u,v)$ is a polynomial, shows that it is possible to extract information about the analytic behavior of $f(z)$ near its singularities. This can then be used to study the coefficients.

Sometimes seemingly complicated functional equations do have easy solutions.

Example 15.6. *A pebbling game.* In a certain pebbling game [76], minimal configurations of size n are counted by $T_n(0)$, where $T_n(x)$ is a polynomial that satisfies $T_n(x) = 0$ for $0 \le n \le 2$, $T_3(x) = 4x + 2x^2$, and for $n \ge 3$,

$$T_{n+1}(x) = x^{-1}(1+x)^2 T_n(x) - x^{-1} T_n(0) + x T_n'(0) + n x^n \ . \tag{15.63}$$

The coefficients of $T_n(x)$ are ≥ 0, and

$$T_{n+1}(1) \le 4T_n(1) + T_n(1) + 1 \le 6T_n(1) \ , \tag{15.64}$$

so clearly each coefficient of $T_n(x)$ is $\le 6^n$, say. Let

$$f(x,y) = \sum_{n=0}^{\infty} T_n(x) y^n \ . \tag{15.65}$$

The bound on $T_n(1)$ shows that $f(x,y)$ is analytic in x and y for $|x| < 1$, $|y| < 1/6$, say, with x and y complex. Then the recurrence (15.63) leads to the functional equation

$$\begin{aligned}(x - y(1+x)^2) f(x,y) &= 2x^2(2+x)y^3 + x^2y^2(1-2x^2y^2)(1-xy)^{-2} \\ &\quad - y f(0,y) + x^2 y f_x(0,y) \ ,\end{aligned} \tag{15.66}$$

where $f_x(x,y)$ is the partial derivative of $f(x,y)$ with respect to x. We now differentiate the equation (15.66) with respect to x and set $x = 0$. We find that

$$(1-2y) f(0,y) = y f_x(0,y) \ , \tag{15.67}$$

and therefore

$$\begin{aligned}(x - y(1+x)^2) f(x,y) &= 2x^2(2+x)y^3 + x^2y^2(1-2x^2y^2)(1-xy)^{-2} \\ &\quad - [y + (2y-1)x^2] f(0,y) \ .\end{aligned} \tag{15.68}$$

When

$$x = y(1+x)^2 \ , \tag{15.69}$$

the left side of Eq. (15.68) vanishes, and Eq. (15.68) yields the value of $f(0,y)$. Now Eq. (15.69) holds for

$$x = (2y)^{-1}(1 - 2y \pm (1-4y)^{1/2}) \ .$$

To ensure that (15.69) holds for x and y both in a neighborhood of 0, we set

$$g(y) = (2y)^{-1}(1 - 2y - (1-4y)^{1/2}) \ . \tag{15.70}$$

Then $g(y) = y(1+g(y))^2$, $g(y)$ is analytic for $|y|$ small, and so substituting $x = g(y)$ in Eq. (15.68) yields

$$\begin{aligned} [y + (2y-1)g(y)^2]f(0,y) &= 2g(y)^2(2+g(y))y^3 \\ &\quad + y^2 g(y)^2(1 - 2y^2 g(y)^2)(1 - yg(y))^{-2} \ . \end{aligned} \tag{15.71}$$

Thus $f(0,y)$ is an algebraic function of y. Eq. (15.71) was proved only for $|y|$ small, but it can now be used to continue $f(0,y)$ analytically to the entire complex plane with the exception of a slit from $1/4$ to infinity along the positive real axis. There is a first order pole at $y = 1/r$, with $r = 4.1478990357\ldots$ the positive root of

$$r^3 - 7r^2 + 14r - 9 = 0 \ , \tag{15.72}$$

and no other singularities in $|y| < 1/4$. Hence we obtain

$$T_n(0) = [y^n]f(0,y) = cr^n + O((4+\epsilon)^n) \tag{15.73}$$

as $n \to \infty$, for every $\epsilon > 0$, where c is an algebraic number that can be given explicitly in terms of r.

The value of $f(0,y)$ is determined by Eq. (15.71), and together with Eq. (15.68) gives $f(x,y)$ explicitly as an algebraic function of x and y. The resulting expression can then be used to determine other coefficients of the polynomials $T_n(x)$. ■

Example 15.6 was easy to present because of the special structure of the functional equation. The main trick was to work on the variety defined by Eq. (15.69), on which the main term vanishes, so that one can analyze the remaining terms. The same basic approach also works in more complicated situations. The analysis of certain double queue systems leads to two-variable generating functions for the equilibrium probabilities that satisfy equations such as the following one, obtained by specializing the problem treated in [145]:

$$Q(z,w)f(z,w) = 2z(w-1)f(z,0) + 3w(z-1)f(0,w) \ , \tag{15.74}$$

valid for complex z and w with $|z|, |w| \leq 1$, where

$$Q(z,w) = 6zw - 3w - 2z - z^2w^2 \ . \tag{15.75}$$

The generating function $f(z,w)$ is analytic in z and w. What makes this problem tractable is that on the algebraic curve in two-dimensional complex space defined by $Q(z,w) = 0$, the quantity on the right-hand side of Eq. (15.74) has to vanish, and this imposes stringent conditions on $f(z,0)$ and $f(0,w)$, which leads to their determination. Once $f(z,0)$ and $f(0,w)$ are found, $f(z,w)$ is defined by Eq. (15.74), and one can determine the asymptotics of its coefficients. Treatment of functional equations of the type (15.74) was started by Malyshev [274]. For recent work and references to other papers in this area, see [144, 145]. This approach has so far been successful only for two-variable problems with $Q(z,w)$ of low degree. Moreover, the mathematics of the solution is far deeper than that used in Example 15.6.

16. Other methods

This section mentions a variety of methods that are not covered elsewhere in this chapter but are useful in asymptotic enumeration. Most are discussed briefly, since they belong to large and well developed fields that are beyond the scope of this survey.

16.1. Permanents

Van der Waerden's conjecture, proved by Falikman [113] and Egorychev [98], can be used to obtain lower bounds for certain enumeration problems. It states that if A is an $n \times n$ matrix that is doubly stochastic (entries ≥ 0, all row and column sums equal to 1) then the permanent of A satisfies $\mathrm{per}(A) \geq n^{-n} n!$. (For most asymptotic problems it is sufficient to rely on an earlier result of T. Bang [26] and S. Friedland [148] which gives a lower bound of $\mathrm{per}(A) \geq e^{-n}$ that is worse only by a factor of $n^{1/2}$.) There is also an upper bound for permanents. Minc's conjecture, proved first by Bragman and in a simpler way by Schrijver [340] states that an

$n \times n$ matrix A with $0, 1$ entries and row sums $r_1, \ldots, r_n$ has

$$\mathrm{per}(A) \leq \prod_{j=1}^{n} (r_j!)^{1/r_j} \ .$$

We now show how these results can be applied.

Example 16.1. *Latin rectangles.* Suppose we are given a $k \times n$ Latin rectangle, $k < n$, so that the symbols are $1, 2, \ldots, n$, and no symbol appears twice in any row or column. In how many ways can we extend this rectangle to a $(k+1) \times n$ Latin rectangle? To get a lower bound, form an $n \times n$ matrix $B = (b_{ij})$, with $b_{ij} = 1$ if i does not appear in column j of the rectangle, and $b_{ij} = 0$ otherwise. Then the row and column sums of B are all equal to $n - k$, so $(n-k)^{-1}B$ is doubly stochastic. Therefore $\mathrm{per}(B)$, which equals the desired number of ways of extending the rectangle, is $\geq (n-k)^n n^{-n} n!$ by van der Waerden's conjecture. By Minc's conjecture, we also have $\mathrm{per}(B) \leq ((n-k)!)^{n/(n-k)}$. If we let $L(k, n)$ denote the number of $k \times n$ Latin rectangles, then $L(1, n) = n!$, and the bounds derived above for the number of ways to extend any given rectangle give

$$L(k,n) \ \geq \ \prod_{j=0}^{k-1} \{(n-j)^n n^{-n} n!\} = n^{-kn} (n!)^{2n} ((n-k)!)^{-n} \ , \tag{16.1}$$

$$L(k,n) \ \leq \ \prod_{j=0}^{k-1} \{(n-j)!\}^{n/(n-j)} \ . \tag{16.2}$$

Sharper estimates for $L(k, n)$ have been obtained through more powerful and complicated methods by Godsil and McKay [163]. They obtain an asymptotic relation for $L(k, n)$ that is valid for $k = o(n^{6/7})$, and improved estimates for other k. (It is known that for any fixed k, the sequence $L(k, n)$ satisfies a linear recurrence with polynomial coefficients [160].) ■

There are problems in which inequalities for permanents give the correct asymptotic estimates. One such example is presented in [318] which discusses a variation on the "problème des rencontres."

16.2. Probability theory and branching process methods

Many combinatorial enumeration results can be phrased in probabilistic language, and a few probabilistic techniques have appeared in the preceding sections. However, the stress throughout this chapter has been on elementary and generating function approaches to asymptotic enumeration problems. Probabilistic methods provide another way to approach many of these problems. This has been appreciated more in the former Soviet Union than in the West, as can be seen in the books [240, 241, 338].

The last few years have seen a great increase in the applications of probabilistic methods to combinatorial enumeration and analysis of algorithms. Many powerful tools, such as martingales, branching processes, and Brownian motion asymptotics have been brought to bear on this topic. General introductions and references to these topics can be found in Chapter ? as well as in [5, 11, 20, 21, 27, 92, 93, 108, 258, 260, 262, 270].

16.3. Statistical physics

There is an extensive literature in mathematical physics concerned with asymptotic enumeration, especially in Ising models of statistical mechanics and percolation methods. Many of the methods are related to combinatorial enumeration. For an introduction to them, see Chapter ? or the books [30, 226].

16.4. Classical applied mathematics

There are many techniques, such as the ray method and the WKB method, that have been developed for solving differential and integral equations in what we might call classical applied mathematics. An introduction to them can be found in [31]. They are powerful, but they have the disadvantage that most of them are not rigorous, since they make assumptions about the form or the stability of the solution that are likely to be true, but have not been established. Therefore we have not presented such methods in this survey. For some examples of the nonrigorous applications of these methods to asymptotic enumeration, see the papers of Knessl and Keller [231, 232]. It is likely that with additional work, more of these methods will be rigorized, which will increase their utility.

17. Algorithmic and automated asymptotics

Deriving asymptotic expansions often involves a substantial amount of tedious work. However, much of it can now be done by computer symbolic algebra systems such as Macsyma, Maple, and Mathematica. There are many widely available packages that can compute Taylor series expansions. Several can also compute certain types of limits, and some have implemented Gosper's indefinite hypergeometric summation algorithm [171]. They ease the burden of carrying out the necessary but uninteresting parts of asymptotic analysis. They are especially useful in the exploratory part of research, when looking for identities, formulating conjectures, or searching for counterexamples.

Much more powerful systems are being developed. Given a sequence, there are algorithms that attempt to guess the generating function of that sequence [46, 162]. It is possible to go much further than that. Many of the asymptotic results in this chapter are stated in explicit forms. As an example, the asymptotics of a linear recurrence is derived easily from the characteristic polynomial and the initial conditions, as was shown in Section 9.1. One needs to compute the roots of the characteristic polynomial, and that is precisely what computer systems do well. It is therefore possible to write programs that will derive the asymptotics behavior from the specification of the recurrence. More generally, one can analyze asymptotics of a much greater variety of generating functions. Flajolet, Salvy, and Zimmermann [124, 139] have written a powerful program for just such computations. Their system uses Maple to carry out most of the basic analytic computations. It contains a remarkable amount of automated expertise in recognizing generating functions, computing their singularities, and extracting asymptotic information about their coefficients. For example, if

$$f(z) = -\log[1 + z\log(1-z^2)] + (1-z^3)^{-5} + \exp(ze^z) \ , \tag{17.1}$$

then the Flajolet-Salvy-Zimmermann system can determine that the singularity of $f(z)$ that is closest to the origin is at $z = \rho$, where ρ is the smallest positive root of

$$1 = -\rho\log(1-\rho^2) \ , \tag{17.2}$$

and then can deduce that

$$[z^n]f(z) = n^{-1}\rho^{-n} + O(n^{-2}\rho^{-n}) \quad \text{as} \quad n \to \infty \ . \tag{17.3}$$

The Flajolet-Salvy-Zimmermann system is even more powerful than indicated above, since it does not always require an explicit presentation of the generating function. Instead, often it can accept a formal description of an algorithm or data structure, derive the generating function from that, and then obtain the desired asymptotic information. For example, it can show that the average path length in a general planar tree with n nodes is

$$\frac{1}{2}\pi^{1/2}n^{3/2} + \frac{1}{2}n + O(n^{1/2}) \quad \text{as} \quad n \to \infty \ . \tag{17.4}$$

What makes systems such as that of [139] possible is the phenomenon, already mentioned in Section 6, that many common combinatorial operations on sets, such as unions and permutations, correspond in natural ways to operations on generating functions.

Further work extending that of [139] is undoubtedly going to be carried out. There are some basic limitations coming from the undecidability of even simple problems of arithmetic, which are already known to impose a limitation on the theories of indefinite integration. If we approximate a sum by an integral

$$\int_a^b x^{-\alpha} dx \ , \tag{17.5}$$

then as a next step we need to decide whether $\alpha = 1$ or not, since if $\alpha = 1$, this integral is $\log(b/a)$ (assuming $0 < a < b < \infty$), whereas if $\alpha \neq 1$, it is $(b^{1-\alpha} - a^{1-\alpha})/(1-\alpha)$. Deciding whether $\alpha = 1$ or not, when α is given implicitly or by complicated expressions, can be arbitrarily complicated. However, such difficulties are infrequent, and so one can expect substantial increase in the applicability of automated systems for asymptotic analysis.

The question of decidability of asymptotic problems and generic properties of combinatorial structures that can be specified in various logical frameworks has been treated by Compton in a series of papers [77, 78, 79]. There is the beautiful recent theory of 0-1 laws for random graphs, which says that certain (so-called first-order) properties are true with probability either 0 or 1 for random graphs. Compton proves that certain classes of asymptotic theories also have 0-1 laws, and describes general properties that have to hold for almost all random structures in certain classes. His analysis uses Tauberian theorems and Hayman admissibility to determine asymptotic behavior. For some further developments in this area, see also [35].

18. Guide to the literature

This section presents additional sources of information on asymptotic methods in enumeration and analysis of algorithms. It is not meant to be exhaustive, but is intended to be used as a guide in searching for methods and results. Many references have been presented already throughout this chapter. Here we describe only books that cover large areas relevant to our subject.

An excellent introduction to the basic asymptotic techniques is given in [175]. That book, intended to be an undergraduate textbook, is much more detailed than this chapter, and assumes no knowledge of asymptotics, but covers fewer methods. A less comprehensive and less elementary book that is oriented towards analysis of algorithms, but provides a good introduction to many asymptotic enumeration methods, is [177].

The best source from which to learn the basics of more advanced methods, including many of those covered in this chapter, is de Bruijn's book [63]. It was not intended particularly for those interested in asymptotic enumeration, but almost all the methods in it are relevant. De Bruijn's volume is extremely clear, and provides insight into why and how various methods work.

General presentations of asymptotic methods, although usually with emphasis on applications to applied mathematics (differential equations, special functions, and so on) are available in the books [54, 100, 114, 115, 315, 344, 354, 372, 382, 385]. Integral transforms are treated extensively in [89, 95, 116, 299, 365]. Books that deal with asymptotics arising in the analysis of algorithms or probabilistic methods include [11, 55, 108, 209, 223, 240, 241, 270, 338].

Nice general introductions to combinatorial identities, generating functions, and related topics are presented in [81, 351, 377]. Further material can be found in [13, 88, 99, 173, 188, 335, 336].

A very useful book is the compilation [168]. While it does not discuss methods in too much detail, it lists a wide variety of enumerative results on algorithms and data structures, and gives references where the proofs can be found.

Last, but not least in our listing, is Knuth's three-volume work [235, 236, 237]. While it is devoted primarily to analysis of algorithms, it contains an enormous amount of material on combinatorics, especially asymptotic enumeration.

Acknowledgements

The author thanks R. Arratia, E. A. Bender, E. R. Canfield, H. Cohen, P. Flajolet, Z. Gao, D. E. Knuth, V. Privman, L. B. Richmond, E. Schmutz, N. J. A. Sloane, D. Stark, S. Tavaré, H. S. Wilf, D. Zagier and D. Zeilberger for their helpful comments on preliminary drafts of this chapter.

References

[1] J. Aczél, *Lectures on Functional Equations and Their Applications*, Academic Press, 1966.

[2] C. R. Adams, On the irregular cases of linear ordinary difference equations, *Trans. Am. Math. Soc., 30* (1928), pp. 507–541.

[3] A. V. Aho and N. J. A. Sloane, Some doubly exponential sequences, *Fibonacci Quart., 11* (1973), 429–437.

[4] J. A. Aizenberg and A. P. Yuzhakov, *Integral Representations and Residues in Multidimensional Complex Analysis*, Trans. Math. Monographs, No. 58, Amer. Math. Soc., 1983.

[5] D. Aldous, *Probability approximations via the Poisson clumping heuristic*, Springer-Verlag, 1989.

[6] J.-P. Allouche and J. Shallit, The ring of k-regular sequences, *Theoretical Comp. Science, 98* (1992), 163–197.

[7] G. Almkvist, Proof of a conjecture about unimodal polynomials, *J. Number Theory, 32* (1989), 43–57.

[8] G. Almkvist, Exact asymptotic formulas for the coefficients of nonmodular functions, *J. Number Theory, 38* (1991), 145–160.

[9] G. Almkvist, A rather exact formula for the number of plane partitions, *A Tribute to Emil Grosswald*, M. Knapp and M. Sheingorn, eds., Amer. Math. Soc., to appear.

[10] G. Almkvist and G. E. Andrews, A Hardy-Ramanujan-Rademacher formula for restricted partitions, *J. Number Theory, 38* (1991), 135–144.

[11] N. Alon and J. H. Spencer, *The Probabilistic Method*, Wiley, 1992.

[12] G. E. Andrews, Applications of basic hypergeometric functions, *SIAM Review, 16* (1974), pp. 441–484.

[13] G. E. Andrews, *The Theory of Partitions*, Addison–Wesley, 1976.

[14] T. M. Apostol, *Mathematical Analysis*, Addison Wesley, 1957.

[15] T. M. Apostol, *Introduction to Analytic Number Theory*, Springer, 1976.

[16] J. Arney and E. D. Bender, Random mappings with constraints on coalescence and number of origins, *Pacific J. Mathematics*, *103* (1982), pp. 269–294.

[17] R. Arratia, L. Goldstein, and L. Gordon, Poisson approximation and the Chen-Stein method, *Statistical Science 5* (1990), 402–423.

[18] R. Arratia, L. Gordon, and M. S. Waterman, The Erdös-Rényi law in distribution for coin tossing and sequence matching, *Ann. Statist. 18* (1990), 539–570.

[19] R. Arratia and S. Tavaré, The cycle structure of random permutations, *Ann. Prob., 20* (1992), 1567–1591.

[20] R. Arratia and S. Tavaré, Limit theorems for combinatorial structures via discrete process approximation, *Random Structures Alg., 3* (1992), 321–345.

[21] R. Arratia and S. Tavaré, Independent process approximations for random combinatorial structures, *Adv. Math.* (1993), in press.

[22] F. C. Auluck and C. B. Haselgrove, On Ingham's Tauberian theorem for partitions, *Proc. Cambridge Philos. Soc., 48* (1952), pp. 566–570.

[23] R. Ayoub, *An Introduction to the Analytic Theory of Numbers*, Amer. Math. Soc., 1963.

[24] R. Baeza-Yates, R. Casas, J. Diaz, and C. Martinez, On the average size of the intersection of binary trees, *SIAM J. Comp. 21* (1992), 24–32.

[25] R. A. Baeza-Yates, A trivial algorithm whose analysis is not: a continuation, *BIT, 29* (1989), 378–394.

[26] T. Bang, Om matrixfunktioner som med et numerisk lille deficit viser v. d. Waerdens permanenthypotese, Proc. 1976 Turku Scand. Math. Congress.

[27] A. D. Barbour, L. Holst, and S. Janson, *Poisson Approximation*, Oxford University Press, 1992.

[28] E. W. Barnes, On the homogeneous linear difference equation of the second order with linear coefficients, *Messenger Math., 34* (1904), pp. 52–71.

[29] P. M. Batchelder, *An Introduction to Linear Difference Equations*, Harvard Univ. Press, 1927. Dover reprint, 1967.

[30] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics*, Academic Press, 1982.

[31] C. M. Bender and S. A. Orszag, *Applied Mathematical Methods for Scientists and Engineers*, McGraw-Hill, 1978.

[32] E. A. Bender, Central and local limit theorem applied to asymptotic enumeration, *J. Comb. Theory Ser. A., 15* (1973), pp. 91–111.

[33] E. A. Bender, Asymptotic methods in enumeration, *SIAM Review*, *16* (1974), pp. 485–515.

[34] E. A. Bender, An asymptotic expansion for the coefficients of some formal power series, *J. London Math. Soc.*, *9* (1975), pp. 451–458.

[35] E. A. Bender, Z.-C. Gao, and L. B. Richmond, Submaps of maps. I. General 0-1 laws, *J. Comb. Theory, B 55* (1992), 104–117.

[36] E. A. Bender and J. R. Goldman, Enumerative uses of generating functions, *Indiana Univ. Math. J., 20* (1971), pp. 753–765.

[37] E. A. Bender, A. M. Odlyzko, and L. B. Richmond, The asymptotic number of irreducible partitions, *European J. Combinatorics, 6* (1985), pp. 1–6.

[38] E. A. Bender and L. B. Richmond, Central and local limit theorems applied to asymptotic enumeration. II: Multivariate generating functions, *J. Comb. Theory B, 34* (1983), 255–265.

[39] E. A. Bender and L. B. Richmond, An aymptotic expansion for the coefficients of analytic generating functions, *Discrete Math.*, *50* (1984), pp. 135–141.

[40] E. A. Bender and L. B. Richmond, An asymptotic expansion for the coefficients of some power series II: Lagrange inversion, *Discrete Mathematics*, *50* (1984), pp. 135–141.

[41] E. A. Bender and L. B. Richmond, A survey of the asymptotic behaviour of maps, *J. Comb. Theory, Ser. B*, *40* (1986), pp. 297–329.

[42] E. A. Bender, L. B. Richmond, and S. G. Williamson, Central and local limit theorems applied to asymptotic enumeration. III. Matrix recursions, *J. Comb. Theory (A) 35* (1983), 263–278.

[43] E. A. Bender and S. G. Williamson, *Foundations of Applied Combinatorics*, Addison-Wesley, 1991.

[44] F. Bergeron and G. Cartier, Darwin: Computer algebra and enumerative combinatorics, *STACS–88*, R. Cori and M. Wirsing, eds., *LNCS*, *294* (1988), pp. 393–394.

[45] F. Bergeron and G. Labelle and P. Leroux, Functional equations for data structures, *STACS–88*, R. Cori and M. Wirsing, eds., *LNCS*, *294* (1988), pp. 73–80.

[46] F. Bergeron and S. Plouffe, Computing the generating function of a series given its first few terms, *Experimental Math. 1* (1992), 307–312.

[47] B. C. Berndt and L. Schoenfeld, Periodic analogues of the Euler-Maclaurin and Poisson summation formulas with applications to number theory, *Acta Arith. 28* (1975/76), 23–68.

[48] M. V. Berry and C. J. Howls, Hyperasymptotics, *Proc. Royal Soc. London A, 430* (1990), 653–667.

[49] A. Bertozzi and J. McKenna, Multidimensional residues, generating functions, and their application to queueing networks, *SIAM Review 35* (1993), 239–268.

[50] P. Billingsley, *Probability and Measure*, Wiley, 1979.

[51] G. D. Birkhoff, General theory of linear difference equations, *Trans. Amer. Math. Soc.*, *12* (1911), pp. 243–284.

[52] G. D. Birkhoff, Formal theory of irregular linear difference equations, *Acta Math., 54* (1930), pp. 205–246.

[53] G. D. Birkhoff and W. J. Trjitzinsky, Analytic theory of singular difference equations, *Acta Math.*, *60* (1932), pp. 1–89.

[54] N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals*, 2nd edition, Holt, Rinehart and Winston, New York, 1975.

[55] B. Bollobás, *Random Graphs*, Academic Press, 1985.

[56] F. Brenti, *Unimodal, Log-concave, and Pólya Frequency Sequences in Combinatorics*, Memoirs Amer. Math. Soc., no. 413 (1989).

[57] F. Brenti, G. F. Royle, and D. G. Wagner, Location of zeros of chromatic and related polynomials of graphs, to be published.

[58] N. A. Brigham, A general asymptotic formula for partition functions, *Proc. Amer. Math. Soc., 1* (1950), pp. 182–191.

[59] T. I'a. Bromwich, *An Introduction to the Theory of Infinite Series*, 2nd rev. ed., Macmillan, London, 1955.

[60] G. G. Brown and B. O. Shubert, On random binary trees, *Math. Oper. Res.*, *9* (1984), pp. 43–65.

[61] N. G. de Bruijn, On Mahler's partition problem, *Indagationes Math.*, *10* (1948), pp. 210–220.

[62] N. G. de Bruijn, The difference–differential equation $F'(x) = e^{\alpha x+\beta} F(x-1)$, *Indagationes Math.*, *15* (1953), pp. 449–458.

[63] N. G. de Bruijn, *Asymptotic Methods in Analysis*, North–Holland, Amsterdam, 1958.

[64] N. G. de Bruijn, D. E. Knuth, and S. O. Rice, The average height of planted plane trees, in *Graph Theory and Computing*, R.–C. Read, ed., Academic Press, New York, 1972, pp. 15–22.

[65] E. R. Canfield, Central and local limit theorems for the coefficients of polynomials of binomial type, *J. Comb. Theory*, Series A, *23* (1977), pp. 275–290.

[66] E. R. Canfield, The asymptotic behavior of the Dickman–de Bruijn function, *Congressus Numerantium*, *35* (1982), pp. 139–148.

[67] E. R. Canfield, Remarks on an asymptotic method in combinatorics, *J. Comb. Theory*, Series A, *37* (1984), pp. 348–352.

[68] M. Car, Factorisation dans $\mathbf{F}_q[X]$, *C. R. Acad. Sci. Paris Série I, 294* (1982), pp. 147–150.

[69] M. Car, Ensembles de polynômes irréductibles et thérèmes de densité, *Acta Arith, 44* (1984), pp. 323–342.

[70] L. Carlitz, Permutations, sequences and special functions, *SIAM Review*, *17* (1975), pp. 298–322.

[71] R. Casas, D. Diaz, and C. Martinez, Statistics on random trees, in *Automata, Languages, and Programming* (Proc. 18th ICALP, Madrid, 1991), J. Leach Albert, B. Monien, and M. Rodriguez Artalejo, eds., Springer LNCS #510, 1991, pp. 186–203.

[72] J. W. S. Cassels, On the representation of integers as the sums of distinct summands taken from a fixed set, *Acta Sci. Math. Hungar., 21* (1960), 111–124.

[73] L. Cerlienco, M. Mignotte, and F. Piras, Suites récurrentes linéaires, *L' Enseign. Math., 33* (1987), 67–108.

[74] Ch. A. Charalambides and A. Kyriakoussis, An asymptotic formula for the exponential polynomials and a central limit theorem for their coefficients, *Discrete Math.*, *54* (1985), pp. 259–270.

[75] L. H. Y. Chen, Poisson approximation for dependent trials, *Ann. Prob. 3* (1975) 534–545.

[76] F. R. K. Chung, R. L. Graham, J. A. Morrison, and A. M. Odlyzko, Pebbling a chessboard, *Am. Math. Monthly*, to appear.

[77] K. J. Compton, A logical approach to asymptotic combinatorics. I. First order properties, *Advances in Math.*, *65* (1987), pp. 65–96.

[78] K. J. Compton, 0–1 laws in logic and combinatorics, in *Proceedings NATO Advanced Study Institute on Algorithms and Order*, I. Rival, ed., Reidel, Dordrecht, 1988, pp. 353–383.

[79] K. J. Compton, A logical approach to asymptotic combinatorics. II. Monadic second–order properties, *J. Comb. Theory*, Series A, *50* (1989), pp. 110–131.

[80] L. Comtet, Birecouvrements et birevêtements d'un ensemble fini, *Studia Sci. Math. Hungar. 3* (1968), 137–152.

[81] L. Comtet, *Advanced Combinatorics*, Reidel, Dordrecht, 1974.

[82] C. N. Cooper and R. E. Kennedy, A partial asymptotic formul for the Niven numbers, *Fibonacci Quarterly*, *26* (1988), pp. 163–168.

[83] J. Coquet, A summation formula related to binary digits, *Inventiones math.*, *73* (1983), pp. 107–115.

[84] R. Courant and D. Hilbert, *Methods of Mathematical Physics*, Interscience, 1953 (vol. 1) and 1962 (vol. 2).

[85] T. W. Cusick, Recurrences for sums of powers of binomial coefficients, *J. Comb. Theory*, Series A, *52* (1989), pp. 77–83.

[86] H. E. Daniels, Saddlepoint approximations in statistics, *Annals Math. Statistics*, *25* (1954), pp. 631–650.

[87] G. Darboux, Mémoire sur l'approximation des fonctions de très-grands nombres, et sur une classe étendue de développements en série, *J. Math. Pures Appl., 4* (1878), 5–56, 377–416.

[88] F. N. David and D. E. Barton, *Combinatorial Chance*, Griffin, 1962.

[89] B. Davies, *Integral Transforms and Their Applications*, Springer, 1978.

[90] J. Denef and L. Lipshitz, Algebraic power series and diagonals, *J. Number Theory, 26* (1987), pp. 46–67.

[91] R. L. Devaney, *An Introduction to Chaotic Dynamical Systems*, 2nd ed., Addison-Wesley, 1989.

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

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

[94] P. Diaconis and D. Freedman, Finite exchangeable sequences, *Ann. Probab. 8* (1980), 745–764.

[95] G. Doetsch, *Handbuch der Laplace Transformation*, Birkhäuser, Basel, 1955.

[96] M. Drmota, Asymptotic distributions and a multivariate Darboux method in enumeration problems, *J. Comb. Theory A*, to appear.

[97] R. Durrett, *Probability: Theory and Examples*, Wadsworth and Brooks/Cole, 1991.

[98] G. P. Egorychev, The solution of van der Waerden's problem for permanents, *Adv. Math., 42* (1981), 299–305.

[99] G. P. Egorychev, *Integral Representation and the Computation of Combinatorial Sums*, Amer. Math. Soc. 1984.

[100] A. Erdélyi, *Asymptotic Expansions*, Dover reprint, 1956.

[101] A. Erdélyi, General asymptotic expansions of Laplace integrals, *Arch. Rational Mech. Anal.*, *7* (1961), pp. 1–20.

[102] A. Erdélyi and M. Wyman, The asymptotic evaluation of certain integrals, *Arch. Rational Mech. Anal.*, *14* (1963), pp. 217–260.

[103] P. Erdös, On some asymptotic formulas in the theory of 'Factorisatio numerorum', *Annals Math., 42* (1941), 989–993. (Corrections: *44* (1943), 647–651.)

[104] P. Erdös, A. Hildebrand, A. Odlyzko, P. Pudaite, and B. Reznick, The asymptotic behavior of a family of sequences, *Pacific J. Math., 126* (1987), pp. 227–241.

[105] P. Erdös and J. Lehner, The distribution of the number of summands in the partitions of a positive integer, *Duke Math. J., 8* (1941), 335–345.

[106] P. Erdös and J. H. Loxton, Some problems in partitio numerorum, *J. Austral. Math. Soc. (Ser. A) 27* (1979), 319–331.

[107] P. Erdös and B. Richmond, Concerning periodicity in the asymptotic behavior of partition functions, *J. Austral. Math. Soc. A 21* (1976), 447–456.

[108] P. Erdös and J. Spencer, *Probabilistic Methods in Combinatorics*, Academic Press and Akadémiai Kiado, New York, 1974.

[109] P. Erdös and P. Turán, On some problems of a statistical group–theory, I–IV; I: *Z. Wahrscheinlichkeitstheorie u. verw. Gebiete, 4* (1965), pp. 175–186; II–IV: *Acta Math. Acad. Sci. Hungar., 18* (1967), pp. 151–163 and 309–320, *19* (1968), pp. 413–435.

[110] M. A. Evgrafov, *Asymptotic Estimates and Entire Functions*, Gordon and Breach, New York, 1961.

[111] M. A. Evgrafov, *Analytic Functions*, Dover, New York, 1966.

[112] M. A. Evgrafov, Series and integral representations, pp. 1–81 in *Analysis I*, R. V. Gamkrelidze, ed., Springer 1989.

[113] D. I. Falikman, Proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki 29* (1981), 931–938. (In Russian.)

[114] M. V. Fedoryuk, *Asymptotics: Integrals and Series*, Nauka, Moscow 1987. (In Russian.)

[115] M. V. Fedoryuk, Asymptotic methods in analysis, pp. 83–191 in *Analysis I*, R. V. Gamkrelidze, ed., Springer 1989.

[116] M. V. Fedoryuk, Integral transforms, pp. 193–232 in *Analysis I*, R. V. Gamkrelidze, ed., Springer 1989.

[117] W. Feller, *An Introduction to Probability Theory*, vol. I, 3rd ed., vol. II, 2nd ed., John Wiley, New York, 1968, 1971.

[118] J. L. Fields, A uniform treatment of Darboux's method, *Arch. Rational Mech. Anal.*, *27* (1968), pp. 289–305.

[119] P. C. Fishburn and A. M. Odlyzko, Unique subjective probability on finite sets, *J. Ramanujan Math. Soc., 4* (1989), pp. 1–23.

[120] P. C. Fishburn, A. M. Odlyzko, and F. S. Roberts, Two–sided generalized Fibonacci sequences, *Fibonacci Quart., 27* (1989), pp. 352–361.

[121] P. Flajolet, Analyse d'algorithmes de manipulation de fichiers, *Institut de Recherche en Informatique et en Automatique*, No. 321, 1978.

[122] P. Flajolet, Combinatorial aspects of continued fractions, *Discrete Math.*, *32* (1980), pp. 125–161.

[123] P. Flajolet, Mathematical methods in the analysis of algorithms and data structures, *Trends in Theoretical Computer Science*, pp. 225–304, Egon Börger, ed., Computer Science Press, 1988.

[124] P. Flajolet, Analytic analysis of algorithms, *Proc. ICALP '92*, Springer Lecture Notes in Computer Science, 1992, to be published.

[125] P. Flajolet and J. Françon, Elliptic functions, continued fractions and doubled permutations, *European J. Combinatorics*, *10* (1989), pp. 235–241.

[126] P. Flajolet, Z. Gao, A. M. Odlyzko, and B. Richmond, The height of binary trees and other simple trees, *Combinatorics, Probability, and Computing* (1993), to appear.

[127] P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson, The analysis of multidimensional searching in quad-trees, pp. 100–109 in *Proc. 2nd ACM-SIAM Symp. Discrete Algorithms*, SIAM, 1991.

[128] P. Flajolet, G. Gonnet, C. Puech, and J. M. Robson, Analytic variations on quadtrees, *Algorithmica*, to appear.

[129] P. Flajolet, P. Grabner, P. Kirschenhofer, H. Prodinger, and R. F. Tichy, Mellin transforms and asymptotics: digital sums, *Theoretical Comp. Sci.*, to appear.

[130] P. Flajolet, P. Kirschenhofer, and R. Tichy, Deviations from normality in random strings, *Probability Theory and Related Fields, 80* (1988), 139–150.

[131] P. Flajolet and T. Lafforgue, Search costs in quadtrees and singularity perturbation asymptotics, to be published.

[132] P. Flajolet and A. M. Odlyzko, The average height of binary trees and other simple trees, *J. Comput. System Sci., 25* (1982), pp. 171–213.

[133] P. Flajolet and A. M. Odlyzko, Limit distributions for coefficients of iterates of polynomials with application to combinatorial enumeration, *Math. Proc. Cambridge Phil. Soc., 96* (1984), pp. 237–253.

[134] P. Flajolet and A. M. Odlyzko, Random mapping statistics, in *Advances in Cryptology: Proceedings of Eurocrypt '89*, J–J. Quisquater, ed., Springer Lecture Notes in Computer Science, *434* (1990), pp. 329–354.

[135] P. Flajolet and A. M. Odlyzko, Singularity analysis of generating function, *SIAM J. Discrete Math., 3* (1990), pp. 216–240.

[136] P. Flajolet, J.–C. Raoult, and J. Vuillemin, The number of registers required to evaluate arithmetic expressions, *Theoretical Computer Science*, *9* (1979), pp. 99–125.

[137] P. Flajolet, M. Régnier, and R. Sedgewick, Some uses of the Mellin integral transform in the analysis of algorithms, in *Combinatorial Algorithms on Words*, A. Apostolico and Z. Galil, eds., Springer, 1985, pp. 241–254.

[138] P. Flajolet and B. Richmond, Generalized digital trees and their difference-differential equations, *Random Structures Algor. 3* (1992), 305–320.

[139] P. Flajolet, B. Salvy, and P. Zimmermann, Automatic average–case analysis of algorithms, *Theoretical Computer Science*, *79* (1991), 37–109.

[140] P. Flajolet and R. Schott, Non–overlapping partitions, continued fractions, Bessel functions and a divergent series, *European J. Combinatorics*, *11* (1990).

[141] P. Flajolet and R. Sedgewick, Digital search trees revisited, *SIAM J. Comput., 15* (1986), 748–767.

[142] P. Flajolet and M. Soria, Gaussian limiting distributions for the number of components in combinatorial structures, *J. Combinatorial Theory,* Series A, *53* (1990), pp. 165–182.

[143] P. Flajolet and M. Soria, General combinatorial schemes with Gaussian limit distributions and exponential tails, *Discrete Math. 114* (1993), 159–180.

[144] L. Flatto, The longer queue model, *Prob. in Eng. Inform. Sci., 3* (1989), 537–559.

[145] L. Flatto and S. Hahn, Two parallel queues created by arrivals with two demands. I. *SIAM J. Appl. Math., 44* (1984), 1041–1053.

[146] G. W. Ford and G. E. Uhlenbeck, Combinatorial problems in the theory of graphs I, II, III, and IV, *Proc. Nat. Acad. Sci. U.S.A., 42* (1956), pp. 122–128, 203–208, 529–535 and *43* (1957), pp. 163–167. (Part II with R. Z. Norman.)

[147] M. L. Fredman and D. E. Knuth, Recurrence relations based on minimization, *J. Math. Anal. Appl., 48* (1974), 534–559.

[148] S. Friedland, A lower bound for the permanent of a doubly stochastic matrix, *Ann. Math. (2) 110* (1979), 167–176.

[149] A. Frieze, On the length of the longest monotone subsequence in a random permutation, *Ann. Appl. Prob., 1* (1991), 301–305.

[150] B. Fristedt, The structure of random partitions of large integers, *Trans. Amer. Math. Soc., 337* (1993), 703–735.

[151] H. Furstenberg, Algebraic function fields over finite fields, *J. Algebra, 7* (1967), pp. 271–272.

[152] J. Galambos, Bonferroni inequalities, *Annal. Prob., 5* (1977), 577–581.

[153] J. Galambos and Y. Xu, Some optimal bivariate Bonferroni-type bounds, *Proc. Amer. Math. Soc. 117* (1993), 523–528.

[154] T. H. Ganelius, *Tauberian Remainder Theorems*, Lecture Notes in Math. #232, Springer, 1971.

[155] Z. Gao and L. B. Richmond, Central and local limit theorems applied to asymptotic enumeration. IV: Multivariate generating functions, *J. Appl. Comp. Analysis*, *41* (1992), 177–186.

[156] D. Gardy, Méthodes de col et lois limites en analyse combinatoire, *Theoretical Computer Science, 94* (1992), 261–280.

[157] D. Gardy, Some results on the asymptotic behavior of coefficients of large powers of functions, to be published.

[158] D. Gardy and P. Solé, Saddle point techniques in asymptotic coding theory, pp. 75–81 in *Algebraic Coding*, Proc. 1st French-Soviet Workshop, 1991, G. Cohen, S. Litsyn, A. Lobstein, and G. Zémor, eds., Lecture Notes in Computer Science #573, Springer, 1992.

[159] A. M. Garsia and S. A. Joni, A new expansion for umbral operators and power series inversion, *Proc. Amer. Math. Soc., 64* (1977), 179–185.

[160] I. M. Gessel, Counting Latin rectangles, *Bull. Amer. Math. Soc., 16* (1987), 79-82.

[161] I. M. Gessel, Symmetric functions and P-recursiveness, *J. Comb. Theory (A) 53* (1990), 257–286.

[162] S. Getu, L. W. Shapiro, W.-J. Woan, and L. C. Woodson, How to guess a generating function, *SIAM J. Discrete Math., 5* (1992), 497–499.

[163] C. D. Godsil and B. D. McKay, Asymptotic enumeration of Latin rectangles, *J. Comb. Theory*, Series B, *48* (1990), pp. 19–44.

[164] W. M. Y. Goh and E. Schmutz, The expected order of a random permutation, to be published.

[165] W. M. Y. Goh and E. Schmutz, A central limit theorem on $GL_n(F_q)$, to be published.

[166] W. M. Y. Goh and E. Schmutz, Distribution of the number of distinct parts in a random partition, to be published.

[167] V. L. Goncharov, From the domain of combinatorial analysis, *Izv. Akad. Nauk SSSR Ser. Math., 8*, no. 1 (1944), 3–48. (In Russian. English translation in *Transl. Amer. Math. Soc., 19* (1962), 1–46.)

[168] G. H. Gonnet and R. Baeza-Yates, *Handbook of Algorithms and Data Structures*, 2nd ed., Addison-Wesley, 1991.

[169] I. J. Good, Generalizations to several variables of Lagrange's expansion, with applications to stochastic processes, *Proc. Cambridge Phil. Soc., 56* (1960), 367–380.

[170] B. Gordon and L. Houten, Notes on plane partitions. III, *Duke Math. J., 26* (1969), 801–824.

[171] R. W. Gosper, Jr., Decision procedure for indefinite hypergeometric summation, *Proc. Nat. Acad. Sci. USA 75* (1978), 40–42.

[172] H. W. Gould, *Combinatorial Identities*, 1972 (private printing).

[173] I. Goulden and D. Jackson, *Combinatorial Enumeration*, John Wiley, New York, 1983.

[174] I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products*, Academic Press, 1965.

[175] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, Addison Wesley, 1989.

[176] A. G. Greenberg, B. D. Lubachevsky, and A. M. Odlyzko, Simple, efficient asynchronous parallel algorithms for maximization, *ACM Trans. Programming Languages and Systems* (1988), pp. 313–337.

[177] D. H. Greene and D. E. Knuth, *Mathematics for the Analysis of Algorithms*, 2nd ed., Birkhäuser, Boston, 1982.

[178] J. R. Griggs, P. Hanlon, A. M. Odlyzko, and M. S. Waterman, On the number of alignments of k sequences, *Graphs and Combinatorics, 6* (1990), pp. 133–146.

[179] E. Grosswald, Generalization of a formula of Hayman and its application to the study of Riemann's zeta function, *Illinois J. Math.*, *10* (1966), pp. 9–23. Correction in *13* (1969), pp. 276–280.

[180] L. J. Guibas and A. M. Odlyzko, Maximal prefix–synchronized codes, *SIAM J. Appl. Math., 35* (1978), pp. 401–418.

[181] L. J. Guibas and A. M. Odlyzko, Long repetitive patterns in random sequences, *Z. Wahrscheinlichkeitstheorie u. verwandte Geb., 53* (1980), pp. 241–262.

[182] L. J. Guibas and A. M. Odlyzko, String overlaps, pattern matching, and nontransitive games, *J. Comb. Theory A, 30* (1981), pp. 183–208.

[183] W. J. Gutjahr, The variance of level numbers in certain families of trees, *Random Structures Alg. 3* (1992), 361–374.

[184] J. H. Halton, The properties of random trees, *Information Sciences 47* (1989), 95–133.

[185] R. A. Handelsman and J. S. Lew, Asymptotic expansion of Laplace transforms near the origin, *SIAM J. Math. Analysis*, *1* (1970).

[186] E. R. Hansen, *A Table of Series and Products*, Prentice-Hall, 1975.

[187] J. Hansen, Order statistics for decomposable combinatorial structures, *Rand. Struct. Alg.*, to appear.

[188] F. Harary and E. M. Palmer, *Graphical Enumeration*, Academic Press, 1973.

[189] F. Harary, R. W. Robinson, and A. J. Schwenk, Twenty–step algorithm for determining the asymptotic number of trees of various species, *J. Austral. Math. Soc.* (Series A), *20* (1975), pp. 483–503.

[190] G. H. Hardy, *Divergent Series*, Oxford University Press, London, 1949.

[191] G. H. Hardy and J. E. Littlewood, Tauberian theorems concerning power series and Dirichlet's series whose coefficients are positive, *Proc. London Math. Soc. (2) 13* (1914), 174-191. Reprinted in *Collected Papers of G. H. Hardy*, vol. 6, pp. 510–527.

[192] G. H. Hardy and J. E. Littlewood, Some theorems concerning Dirichlet's series, *Messenger Math., 43* (1914), 134-147. Reprinted in *Collected Papers of G. H. Hardy*, vol. 6, pp. 542–555.

[193] G. H. Hardy, J. E. Littlewood, and G. Pólya, *Inequalities*, 2nd ed., Cambridge Univ. Press, 1952.

[194] G. H. Hardy and S. Ramanujan, Asymptotic formulae for the distribution of integers of various types, *Proc. London Math. Soc. (2) 16* (1917), 112-132. Reprinted in *Collected Papers of G. H. Hardy*, vol. 1, pp. 277–293.

[195] L. H. Harper, Stirling behavior is asymptotically normal, *Ann. Math. Stat., 38* (1967), 410–414.

[196] B. Harris, Probability distributions related to random mappings, *Ann. Math. Statist., 31* (1960), pp. 1042–1062.

[197] B. Harris and C. J. Park, The distribution of linear combinations of the sample occupancy numbers, *Nederl. Akad. Wetensch. Proc. Ser. A, 74 = Indag. Math., 33* (1971), pp. 121–134.

[198] B. Harris and L. Schoenfeld, Asymptotic expansions for the coefficients of analytic functions, *Illinois J. Math., 12* (1968), pp. 264–277.

[199] T. E. Harris, *The Theory of Branching Processes*, Springer, 1963.

[200] W. A. Harris and Y. Sibuya, Asymptotic solutions of systems of nonlinear difference equations, *Arch. Rational Mech. Anal., 15* (1964), 277–395.

[201] W. A. Harris and Y. Sibuya, General solution of nonlinear difference equations, *Trans. Amer. Math. Soc., 115* (1965), 62–75.

[202] C. B. Haselgrove and H. N. V. Temperley, Asymptotic formulae in the theory of partitions, *Proc. Cambridge Phil. Soc.,50* (1954), 225–241.

[203] M. L. J. Hautus and D. A. Klarner, The diagonals of a double power series, *Duke Math. J., 38* (1971), 229–235.

[204] W. K. Hayman, A generalization of Stirling's formula, *J. reine angew. Math., 196* (1956), pp. 67–95.

[205] P. Henrici, *Applied and Computational Complex Analysis*, Wiley: Vol. 1, 1974; Vol. 2, 1977; Vol. 3, 1986.

[206] E. Hille, *Lectures on Ordinary Differential Equations*, Addison-Wesley, 1969.

[207] E. Hille, *Ordinary Differential Equations in the Complex Domain*, Wiley, 1976.

[208] J. J. Hofbauer, A short proof of the Lagrange-Good formula, *Discrete Math., 25* (1979), 135–139.

[209] M. Hofri, *Probabilistic Analysis of Algorithms*, Springer, 1987.

[210] C. Hunter, Asymptotic solutions of certain linear difference equations, with applications to some eigenvalue problems, *J. Math. Anal. Appl., 24* (1968), pp. 279–289.

[211] G. K. Immink, *Asymptotics of Analytic Difference Equations*, Lecture Notes in Math. #1085, Springer, 1984.

[212] A. E. Ingham, A Tauberian theorem for partitions, *Ann. of Math., 42* (1941), pp. 1075–1090.

[213] P. Jacquet and M. Régnier, Trie partitioning process: limiting distributions, pp. 196–210 in *CAAP '86*, P. Franchi-Zannettacci, ed., Lecture Notes in Computer Science #214, Springer, 1986.

[214] P. Jacquet and M. Régnier, Normal limiting distribution of the size of tries, pp. 209–223 in *Performance '87*, P.-J. Courtois and G. Latouche, eds., North-Holland, 1988.

[215] P. Jacquet and M. Régnier, Normal limiting distribution for the size and the external path length of tries, in preparation.

[216] L. B. W. Jolley, *Summation of Series*, 2nd ed., Dover, 1961.

[217] A. T. Jonassen and D. E. Knuth, A trivial algorithm whose analysis is not, *J. Comp. Sys. Sci.*, *16* (1978), pp. 301–322.

[218] W. B. Jones and W. J. Thron, *Continued Fractions: Analytic Theory and Applications*, Addison-Wesley, 1980.

[219] R. Jungen, Sur les séries de Taylor n'ayant que des singularitiés algébrico–logarithmiques sur leur cercle de convergence, *Comment. Math. Helv., 3* (1931), pp. 266–306.

[220] S. Kapoor and E. M. Reingold, Recurrence relations based on minimization and maximization, *J. Math. Anal. Appl., 109* (1985), 591–604.

[221] R. M. Karp, Probabilistic recurrence relations, *Proc. 23rd ACM Symp. Theory of Computing*, 1991, pp. 190–197.

[222] S. Karlin, *Total Positivity, Vol. 1*, Stanford Univ. Press, 1968.

[223] R. Kemp, *Fundamentals of the Average Case Analysis of Particular Algorithms*, Wiley, 1984.

[224] R. Kemp, A note on the number of leftist trees, *Inform. Proc. Letters 25* (1987), 227–232.

[225] R. Kemp, Further results on leftist trees, pp. 103–130 in *Random Graphs '87*, M. Karonski, J. Jaworski, and A. Rucinski, eds., Wiley, 1990.

[226] H. Kesten, *Percolation Theory for Mathematicians*, Birkhäuser, 1982.

[227] P. Kirschenhofer, A tree enumeration problem involving the asymptotics of the 'diagonals' of a power series, *Ann. Discrete Math. 33* (1987), 157–170.

[228] P. Kirschenhofer and H. Prodinger, On some applications of formulae of Ramanujan in the analysis of algorithms, *Mathematika, 38* (1991), 14–33.

[229] D. A. Klarner, A combinatorial formula involving the Fredholm integral equation, *J. Combinatorial Theory, 5* (1968), pp. 59–74.

[230] D. A. Klarner and R. L. Rivest, Asymptotic bounds for the number of convex n-ominoes, *Discrete Math., 8* (1974), 31–40.

[231] C. Knessl and J. B. Keller, Partition asymptotics for recursion equations, *SIAM J. Appl. Math., 50* (1990), 323–338.

[232] C. Knessl and J. B. Keller, Stirling number asymptotics from recursion equations using the ray method, *Studies Appl. Math., 84* (1991), 43–56.

[233] A. Knopfmacher, A. Odlyzko, B. Richmond, G. Szekeres, and N. Wormald, manuscript in preparation.

[234] K. Knopp, *Theory and Application of Infinite Series*, 2nd ed., reprinted by Hafner, 1971.

[235] D. E. Knuth, *The Art of Computer Programming Vol. 1: Fundamental Algorithms*, 2nd ed., Addison–Wesley, Reading, 1973.

[236] D. E. Knuth, *The Art of Computer Programming Vol. 2: Semi–Numerical Algorithms*, 2nd ed., Addison–Wesley, Reading, 1981.

[237] D. E. Knuth, *The Art of Computer Programming Vol. 3: Sorting and Searching*, Addison–Wesley, Reading, 1973.

[238] D. E. Knuth and B. Pittel, A recurrence related to trees, *Proc. Amer. Math. Soc., 105* (1989), 335–349.

[239] D. E. Knuth and A. Schönhage, The expected linearity of a simple equivalence algorithm, *Theoretical Comp. Sci., 6* (1978), 281–315.

[240] V. F. Kolchin, *Random Mappings*, Optimization Software Inc., New York, 1986.

[241] V. F. Kolchin, B. A. Sevast'yanov, and V. P. Chistyakov, *Random Allocations*, Wiley, 1978.

[242] J. Komlos, A. M. Odlyzko, L. H. Ozarow, and L. A. Shepp, On the properties of a tree–structured server process, *Ann. Appl. Prob., 1* (1990), 118–125.

[243] R. J. Kooman, *Convergence Properties of Recurrence Sequences*, Ph.D. Dissertation, Leiden, 1989.

[244] R. J. Kooman and R. Tijdeman, Convergence properties of linear recurrence sequences, *Nieuw Archief Wisk., Ser. 4, 4* (1990), 13–25.

[245] M. D. Kruskal, The expected number of components under a random mapping function, *Amer. Math. Monthly, 61* (1954), pp. 392–397.

[246] M. Kuczma, *Functional Equations in a Single Variable*, Polish Scientific Publishers, Warsaw, 1968.

[247] G. Labelle, Une nouvelle démonstration combinatoire des formules d'inversion de Lagrange, *Adv. Math., 42* (1981), 217–247.

[248] J. C. Lagarias, A. M. Odlyzko, and D. B. Zagier, On the capacity of disjointly shared networks, *Computer Networks and ISDN Systems, 10* (1985), pp. 275–285.

[249] M.-Y. Lee, Bivariate Bonferroni inequalities, *Aequationes Math. 44* (1992), 220–225.

[250] J. Leray, Le calcul différentiel et intégral sur une variété analytique complexe, *Bull. Soc. Math. France 87* (1959), 81–180.

[251] L. Lewin, *Polylogarithms and Associated Functions*, North Holland, 1981.

[252] B. Lichtin, The asymptotics of a lattice point problem associated to a finite number of polynomials. I, *Duke Math. J., 63* (1991), 139–192.

[253] L. Lipshitz, The diagonal of a D–finite power series is D–finite, *J. Algebra*, *113* (1988), pp. 373–378.

[254] L. Lipshitz, D–Finite Power Series, *J. Algebra*, *122* (1989), pp. 353–373.

[255] L. Lipshitz and A. van der Poorten, Rational functions, diagonals, automata and arithmetic, in *Number Theory*, Richard A. Mollin, ed., Walter de Gruyter, Berlin, 1990, pp. 339–358.

[256] B. F. Logan, J. E. Mazo, A. M. Odlyzko, and L. A. Shepp, On the average product of Gauss–Markov variables, *Bell System Tech. J., 62* (1983), pp. 2993–3006.

[257] B. F. Logan and L. A. Shepp, A variational problem for random Young tableaux, *Advances Math., 26* (1977), 206–222.

[258] G. Louchard, The Brownian motion: a neglected tool for the complexity analysis of sorted table manipulation, *RAIRO Theoretical Informatics*, *17* (1983), pp. 365–385.

[259] G. Louchard, The Brownian excursion: a numerical analysis, *Computers and Mathematics with Applications*, *10* (1984), pp. 413–417.

[260] G. Louchard, Brownian motion and algorithm complexity, *BIT 26* (1986), 17–34.

[261] G. Louchard, Exact and asymptotic distributions in digital and binary search trees, *RAIRO informatique théorique et applications*, *21* (1987), pp. 479–495.

[262] G. Louchard, B. Randrianarimanana, and R. Schott, Dynamic algorithms in D. E. Knuth's model; a probabilistic analysis, *Theoretical Comp. Sci., 93* (1992), 201–255.

[263] T. Luczak, The number of trees with a large diameter, to be published.

[264] G. S. Lueker, Some techniques for solving recurrences, *Computing Surveys, 12* (1980), 419–436.

[265] A. J. Macintyre and R. Wilson, Operational methods and the coefficients of certain power series, *Math. Ann., 127* (1954), 243–250.

[266] K. Mahler, On a special functional equation, *J. London Math. Soc. 15* (1940), pp. 115–123.

[267] K. Mahler, On a class of nonlinear functional equations connected with modular functions, *J. Austral. Math. Soc. Ser. A 22* (1976), 65–118.

[268] K. Mahler, On a special nonlinear functional equation, *Proc. Roy. Soc. London Ser. A 378* (1981), 155–178.

[269] K. Mahler, On the analytic relation of certain functional and difference equations, *Proc. Roy. Soc. London Ser. A 389* (1983), 1–13.

[270] H. S. Mahmoud, *Evolution of Random Search Trees*, Wiley, 1992.

[271] H. M. Mahmoud and B. Pittel, Analysis of the space of search trees under the random insertion algorithm, *J. Algorithms*, *10* (1989), pp. 52–75.

[272] B. Malgrange, Sur les points singuliers des équations différentielles, *L'Enseign. Math.*, *20* (1974), 147–176.

[273] C. L. Mallows, A. M. Odlyzko, and N. J. A. Sloane, Upper bounds for modular form, lattices, and codes, *J. Algebra, 36* (1975), 68–76.

[274] V. A. Malyshev, An analytic method in the theory of two-dimensional positive random walks, *Sibir. Mat. Zh., 13* (1972), 1314–1329 (in Russian).

[275] A. Maté and P. Nevai, Sublinear perturbations of the differential equation $y^{(n)} = 0$ and of the analogous difference equation, *J. Diff. Equations 53* (1984), 234–257.

[276] A. Maté and P. Nevai, Asymptotics for solutions of smooth recurrence relations, *Proc. Amer. Math. Soc., 93* (1985), 423–429.

[277] J. E. Mazo and A. M. Odlyzko, Lattice points in high–dimensional spheres, *Monatsh. Math., 110* (1990), pp. 47–61.

[278] B. D. McKay, The asymptotic numbers of regular tournaments, eulerian digraphs, and eulerian and oriented graphs, *Combinatorica, 10* (1990), 367–377.

[279] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs of high degree, *European J. Combinatorics, 11* (1990), 565–580.

[280] G. Meinardus, Asymptotische Aussagen über Partitionen, *Math. Z., 59* (1954), pp. 388–398.

[281] A. Meir and J. W. Moon, On the altitude of nodes in random trees, *Canadian J. Math.*, *30* (1978), pp. 997–1015.

[282] A. Meir and J. W. Moon, On random mapping patterns, *Combinatorica*, *4* (1984), pp. 61–70.

[283] A. Meir and J. W. Moon, Some asymptotic results useful in enumeration problems, *Aequationes Math., 33* (1987), 260–268.

[284] A. Meir and J. W. Moon, On an asymptotic method in enumeration, *J. Comb. Theory*, Series A, *51* (1989), pp. 77–89.

[285] A. Meir and J. W. Moon, The asymptotic behavior of coefficients of powers of certain generating functions, *European J. Comb., 11* (1990), 581–587.

[286] N. S. Mendelsohn, The asymptotic series for a certain class of permutation problems, *Canad. J. Math., 8* (1956), pp. 234–244.

[287] L. M. Milne-Thomson, *The Calculus of Finite Differences*, MacMillan, 1933.

[288] D. S. Mitrinović, *Analytic Inequalities*, Springer, 1970.

[289] D. Moews, Explicit Tauberian bounds for multivariate functions, to be published.

[290] J. W. Moon, Counting labeled trees, *Canad. Math. Monograph No. 1, Canad. Math. Congress*, 1970.

[291] J. W. Moon, Some enumeration results on series-parallel networks, *Annals Discrete Math., 33* (1987), 199–226. (*Random Graphs '85*, M. Karonski and Z. Palka, eds., North-Holland 1987.)

[292] L. Moser and M. Wyman, On the solutions of $x^d = 1$ in symmetric groups, *Canad. J. Math., 7* (1955), pp. 159–168.

[293] L. Moser and M. Wyman, Asymptotic expansions, *Canadian J. Math.*, *8* (1956), pp. 225–233.

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

[295] L. Moser and M. Wyman, Stirling numbers of the second kind, *Duke Math. J., 25* (1958), 29–43.

[296] L. Moser and M. Wyman, Asymptotic development of the Stirling numbers of the first kind, *J. London Math. Soc., 33* (1958), 133–146.

[297] National Bureau of Standards, *Handbook of Mathematical Functions*, M. Abramowitz and I. A. Stegun, eds., U.S. Gov. Printing Office, 9th printing, 1970.

[298] N. E. Nörlund, *Vorlesungen über Differenzenrechnung*, Springer, 1924. Dover reprint, 1954.

[299] F. Oberhettinger, *Tables of Mellin Transforms*, Springer, 1974.

[300] A. M. Odlyzko, Periodic oscillations of coefficients of power series that satisfy functional equations, *Adv. Math., 44* (1982), pp. 180–205.

[301] A. M. Odlyzko, Some new methods and results in tree enumeration, *Congressus Numerantium, 42* (1984), pp. 27–52.

[302] A. M. Odlyzko, On heights of monotonically labelled binary trees, *Congressus Numerantium, 44* (1985), pp. 305–314.

[303] A. M. Odlyzko, Enumeration of strings, in *Combinatorial Algorithms on Words*, A. Apostolico and Z. Galil, eds., Springer, 1985, pp. 205–228.

[304] A. M. Odlyzko, Applications of symbolic mathematics to mathematics, pp. 95–111 in *Applications of Computer Algebra*, R. Pavalle, ed., Kluwer, 1985.

[305] A. M. Odlyzko, Explicit Tauberian estimates for functions with positive coefficients, *J. Comput. Appl. Math., 41* (1992), 187–197.

[306] A. M. Odlyzko, B. Poonen, H. Widom, and H. S. Wilf, manuscript in preparation.

[307] A. M. Odlyzko and L. B. Richmond, *On the Compositions of an Integer, Combinatorial Mathematics VII*, R.–W. Robinson, G. W. Southern, and W. D. Wallis, eds., Springer–Verlag Lecture Notes in Mathematics #829, 1980, pp. 119–210.

[308] A. M. Odlyzko and L. B. Richmond, On the unimodality of some partition polynomials, *European J. Combinatorics, 3* (1982), pp. 69–84.

[309] A. M. Odlyzko and L. B. Richmond, On the unimodality of high convolutions of discrete distributions *Ann. Prob., 13* (1985), pp. 299–306.

[310] A. M. Odlyzko and L. B. Richmond, On the number of distinct block sizes in partitions of a set, *J. Combinatorial Theory A, 38* (1985) pp. 170–181.

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

[312] A. M. Odlyzko and H. S. Wilf, Bandwidths and profiles of trees, *J. Combinatorial Theory B, 42* (1987), pp. 348–370. (Condensed summary of results in *Graph Theory and its Applications to Algorithms and Computer Science*, Y. Alavi et al., eds., Wiley, 1985, pp. 605–622.)

[313] A. M. Odlyzko and H. S. Wilf, The editor's corner: n coins in a fountain, *Amer. Math. Monthly, 95* (1988), pp. 840–843.

[314] A. M. Odlyzko and H. S. Wilf, Functional iteration and the Josephus problem, *Glasgow Math. J., 33* (1991), pp. 235–240.

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

[316] R. Otter, The number of trees, *Ann. of Math., 49* (1948), pp. 583–599.

[317] A. I. Pavlov, On the number of substitutions with cycle lengths from a given set, *Discrete Appl. Math. 2* (1992), 445–459.

[318] S. G. Penrice, Derangements, permanents, and Christmas presents, *Amer. Math. Monthly, 98* (1991), 617–620.

[319] O. Perron, *Die Lehre von den Kettenbrüchen*, Chelsea reprint.

[320] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, *Acta Math., 68* (1937), pp. 145–254.

[321] G. Pólya, On the number of certain lattice polygons, *J. Combinatorial Theory 6* (1969), 102–105.

[322] G. Pólya and R. C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer, 1987.

[323] G. Pólya and G. Szegö, *Problems and Theorems in Analysis*, 2 volumes, English translation, Springer, 1972 and 1976.

[324] A. Popken, Asymptotic expansions from an algebraic standpoint, *Indagationes Math.*, *15* (1953), pp. 131–143.

[325] A. G. Postnikov, Tauberian theory and its applications, *Proc. Steklov Inst. Math., 144*; English translation, *Amer. Math. Soc. Transl.* (1980).

[326] V. Privman and N. M. Svrakic, Difference equations in statistical mechanics: I. Cluster statistics models and II. Solid-on-solid models in two dimensions, *J. Stat. Phys. 51* (1988), 1091–1110 and 1111–1126.

[327] V. Privman and N. M. Svrakic, *Directed Models of Polymers, Interfaces, and Clusters: Scaling and Finite-Size Properties*, Lecture Notes in Physics #338, Springer, 1989.

[328] H. Rademacher, On the partition function, *Proc. London Math. Soc., 43* (1937), pp. 241–254.

[329] A. Regev, Asymptotic values for degrees associated with strips of Young diagrams, *Advances Math., 41* (1981), 115–136.

[330] A. Rényi, Three more proofs and a generalization of a theorem of Irving Weiss, *Magyar Tud. Akad. Mat. Kutató Int. Közl., 7* (1962), 203–214.

[331] A. Rényi and G. Szekeres, On the height of trees, *J. Austral. Math. Soc., 7* (1967), pp. 497–507.

[332] L. B. Richmond, Asymptotic relations for partitions, *J. Number Theory, 4* (1975), 389–405.

[333] L. B. Richmond, The moments of partitions. II, *Acta Arith., 28* (1975), 229–243.

[334] L. B. Richmond, Asymptotic relations for partitions, *Trans. Amer. Math. Soc., 219* (1976), 379–385.

[335] J. Riordan, *Introduction to Combinatorial Analysis*, John Wiley, New York, 1958.

[336] J. Riordan, *Combinatorial Identities*, Wiley, 1968.

[337] K. F. Roth and G. Szekeres, Some asymptotic formulae in the theory of partitions, *Quart. J. Math. Oxford Ser., 5* (1954), pp. 241–259.

[338] V. N. Sachkov, *Probabilistic Methods in Combinatorial Analysis* (in Russian), Nauka, Moscow, 1978.

[339] E. Schmutz, Asymptotic expansions for the coefficients of $e^{P(z)}$, *Bull. London Math. Soc. 21* (1989), pp. 482–486.

[340] A. Schrijver, A short proof of Minc's conjecture, *J. Comb. Theory (A), 25* (1978), 80–83.

[341] R. Sedgewick, Data movement in odd–even merging, *SIAM J. Comput., 7* (1978), pp. 239–272.

[342] L. A. Shepp and S. P. Lloyd, Ordered cycle lengths in a random permutation, *Trans. Amer. Math. Soc., 121* (1966), pp. 340–357.

[343] J. A. Shohat and J. D. Tamarkin, *The Problem of Moments*, Amer. Math. Soc., 1943.

[344] L. Sirovich, *Techniques of Asymptotic Analysis*, Springer Verlag, 1971.

[345] N. J. A. Sloane, *A Handbook of Integer Sequences*, Academic Press, 1973. A revised and expanded edition is in press.

[346] N. J. A. Sloane, *The New Book of Integer Sequences*, Freeman, 1994, to be published.

[347] I. N. Sneddon, *The Uses of Integral Transforms*, McGraw, 1972.

[348] J. Spencer, *Ten Lectures on the Probabilistic Method*, SIAM, 1987.

[349] R. P. Stanley, Generating Functions, in *Studies in Combinatorics*, M.A.A. Studies in Mathematics, Vol. 17., G–C. Rota, ed., Math. Ass. of America, 1978, pp. 100–141.

[350] R. P. Stanley, Differentiably finite power series, *European J. Combinatorics, 1* (1980), 175–188.

[351] R. P. Stanley, *Enumerative Combinatorics*, Wadsworth and Brooks/Cole, Monterey, 1986.

[352] R. P. Stanley, Log-concave and unimodal sequences in algebra, combinatorics, and geometry, pp. 500–535 in *Graph Theory and its Applications: East and West*, *Annals New York Acad. Sci.*, no. 576, 1989.

[353] K. B. Stolarsky, Power and exponential sums of digital sums related to binomial coefficient parity, *SIAM J. Appl. Math., 32* (1977), 717–730.

[354] G. Szegö, *Orthogonal Polynomials*, Amer. Math. Soc. Coll. Publ., vol. 23, rev. ed., Amer. Math. Soc., New York, 1959.

[355] G. Szekeres, Some asymptotic formulae in the theory of partitions. II *Quart. J. Math. Oxford*, Ser. 2, *4* (1953), pp. 96–111.

[356] G. Szekeres, Regular iteration of real and complex functions, *Acta Math.*, *100* (1958), pp. 103–258.

[357] G. Szekeres, Distribution of labelled trees by diameter, pp. 392–397 in *Combinatorial Mathematics X*, Proc. 10-th Australian Conf. Comb. Math., Lecture Notes in Mathematics, Springer, 1982.

[358] G. Szekeres, Asymptotic distribution of the number and size of parts in unequal partitions, *Bull. Australian Math. Soc. 36* (1987), 89–97.

[359] G. Szekeres, Asymptotic distribution of partitions by number and size of parts, pp. 527–538 in *Number Theory*, vol. I, K. Györy and G. Halász, eds., Colloq. Math. Soc. J. Bolyai, No. 51, North-Holland, 1990.

[360] W. Szpankowski, The evaluation of an alternative sum with applications to the analysis of some data structures, *Inform. Proc. Letters, 28* (1988), 13–19.

[361] L. Takács, On the number of distinct forests, *SIAM J. Discr. Math., 3* (1990), 574–581.

[362] L. Takács, A Bernoulli excursion and its various applications, *Adv. Appl. Prob., 23* (1991), 557–585.

[363] N. M. Temme, Asymptotic estimates of Stirling numbers, *Studies Appl. Math. 89* (1993), 233–243.

[364] E. C. Titchmarsh, *The Theory of Functions*, 2nd ed., Oxford University Press, London, 1939.

[365] E. C. Titchmarsh, *Fourier Integrals*, 2nd ed., Oxford Univ. Press, 1948.

[366] W. J. Trjitzinsky, Analytic theory of linear q-difference equations, *Acta Math., 61* (1933), 1–38.

[367] W. J. Trjitzinsky, Analytic theory of linear differential equations, *Acta math., 62* (1933), 167–226.

[368] V. S. Varadarajan, Meromorphic differential equations, *Expositiones Math.. 9* (1991), 97–188.

[369] R. C. Vaughan, *The Hardy-Littlewood Method*, Cambridge Univ. Press, 1981.

[370] A. M. Vershik and C. V. Kerov, Asymptotics of the Plancherel measure of the symmetric group and a limiting form for Young tableau, *Dokl. Akad. Nauk USSR, 233* (1977), 1024–1027. (In Russian.)

[371] J. Vitter and P. Flajolet, Analysis of algorithms and data structures, *Handbook of Theoretical Computer Science*, Vol. A: Algorithms and Complexity, Ch. 9, pp. 432–524, J. Van Leeuwen, ed., North Holland, 1990.

[372] W. Wasow, *Asymptotic Expansions for Ordinary Differential Equations*, Wiley, 1965.

[373] W. D. Wei, Y. Z. Cai, C. L. Liu, and A. M. Odlyzko, Balloting labelling and personnel assignment, *SIAM J. Alg. Discr. Methods, 7* (1986), pp. 150–158

[374] E. A. Whitehead, Jr., *Four-discordant permutations*, J. Austral. Math. Soc. (Ser. A) 28 (1979), 369–377.

[375] E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th ed., Cambridge University Press, Cambridge, 1927.

[376] H. S. Wilf, The asymptotics of $e^{P(z)}$ and the number of elements of each order in S_n, *Bull. Am. Math. Soc.*, *15* (1986), pp. 228–232.

[377] H. S. Wilf, *Generatingfunctionology*, Academic Press, 1990.

[378] H. S. Wilf, The asymptotic behavior of the Stirling numbers of the first kind, *J. Comb. Theory Ser. A*, to appear.

[379] H. S. Wilf and D. Zeilberger, Rational functions certify combinatorial identities, *J. Amer. Math. Soc.*, *3* (1990), pp. 147–158.

[380] H. S. Wilf and D. Zeilberger, An algorithmic proof theory for hypergeometric (ordinary and "q") multisum/integral identities, *Inventiones math., 108* (1992), 575–633.

[381] R. Wilson, The coefficient theory of integral functions with dominant exponential parts, *Quart. J. Math. Oxford, ser. 2, 4* (1953), 142–149.

[382] J. Wimp, *Computation with Recurrence Relations*, Pitman, Boston, 1984.

[383] J. Wimp, Current trends in asymptotics: some problems and some solutions, *J. Computational Appl. Math., 35* (1991), 53–79.

[384] J. Wimp and D. Zeilberger, Resurrecting the asymptotics of linear recurrences, *J. Math. Anal. Appl.*, *111* (1985), pp. 162–176.

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

[386] R. Wong and M. Wyman, The method of Darboux, *J. Approx. Theory, 10* (1974), pp. 159–171.

[387] E. M. Wright, Asymptotic partition formulae, I: Plane partitions, *Quart. J. Math. Oxford Ser., 2* (1931), pp. 177–189.

[388] E. M. Wright, The coefficients of a certain power series, *J. London Math. Soc.*, *7* (1932), pp. 256–262.

[389] E. M. Wright, On the coefficients of power series having exponential singularities, *J. London Math. Soc.*, *24* (1949), pp. 304–309.

[390] E. M. Wright, Partitions of large bipartities, *Amer. J. Math., 80* (1958), 643–658.

[391] E. M. Wright, The asymptotic behavior of the generating functions of partitions of multipartities, *Quart. J. Math. Oxford (2) 10* (1959), 60–69.

[392] E. M. Wright, Partitions into k parts, *Math. Annalen*, *142* (1961), pp. 311–316.

[393] E. M. Wright, A relationship between two sequences, *Proc. London Math. Soc. 17* (1967), 296–304, 547–552.

[394] E. M. Wright, Asymptotic relations between enumerative functions in graph theory, *Proc. London Math. Soc. (3), 20* (1970) pp. 558–572.

[395] E. M. Wright, Graphs on unlabelled nodes with a given number of edges, *Acta Math., 126* (1971), pp. 1–9.

[396] E. M. Wright, The number of strong digraphs, *Bull. London Math. Soc., 3* (1971), 348–350.

[397] E. M. Wright, Graphs on unlabelled nodes with a large number of edges, *Proc. London Math. Soc. (3), 28* (1974), 577–594.

[398] E. M. Wright and B. G. Yates, The asymptotic expansion of a certain integral, *Quarterly J. Mathematics Oxford*, *1* (1950) pp. 41–53.

[399] R. A. Wright, L. B. Richmond, A. M. Odlyzko, and B. D. McKay, Constant time generation of free trees, *SIAM J. Comp., 15* (1986), pp. 540–548.

[400] M. Wyman, The asymptotic behavior of the Laurent coefficients, *Canad. J. Math., 11* (1959), pp. 534–555.

[401] M. Wyman, The method of Laplace, *Trans. Royal Soc. Canada*, *2* (1964), pp. 227–256.

[402] D. Zeilberger, Solutions of exponential growth to systems of partial differential equations, *J. Differential Eq.*, *31* (1979), pp. 287–295.

[403] D. Zeilberger, The algebra of linear partial difference operators and its applications, *SIAM J. Math. Analysis*, *11* (1980), pp. 919–932.

[404] D. Zeilberger, Six etudes in generating functions, *Intern. J. Computer Math.*, *29* (1989), pp. 201–215.

[405] D. Zeilberger, A holonomic approach to special function identities, *J. Comput. Appl. Math., 32* (1990), 321–368.

Fig. 1. Domain $\Delta(r, \phi, \eta)$ of Section 11.1 and the integration contour Γ.

Contents