Chapter 7

Generating Functions

Generating functions are one of the most popular analytic tools in analysis of algorithms and combinatorics. They are helpful in establishing exact and recurrence formulas, deriving averages, variances and other statistical properties, finding asymptotic expansions, showing unimodality, convexity, and proving combinatorial identities. In this chapter we present the formal power series approach and analytic theory of ordinary generating functions, exponential generating functions, probability generating functions, and Dirichlet series. In particular, we construct generating functions for some popular combinatorial structures. In the application section, we present a fairly general approach to certain recurrence equations arising in the analysis of digital trees and find generating functions of the number of pattern occurrences in a random text.

THE GENERATING FUNCTION of a sequence $\{a_n\}_{n\geq 0}$ (e.g., representing the size of objects belonging to a certain class) is defined as

$$A(z) = \sum_{n\geq 0} a_n z^n$$

where the meaning of z is explained below. In the *formal power series* we assume that $A(z)$ (or even simpler A) is an algebraic object; more precisely, the set of such formal power series forms a *ring*. In this case, z does not have any value, but one can identify the coefficient at z^n. Moreover, we can manipulate formal power series to discover new identities, establish recurrences and exact formulas for the coefficients. Convergence of $A(z)$ is *not* an issue.

In *analytic theory* of generating functions, we assume that z is a complex number, and the issue of convergence is a pivotal one. In fact, singularity points of $A(z)$ (i.e., points where $A(z)$ is not defined) determine asymptotics of the coefficients. We study asymptotic methods in the next chapter.

We can define other generating functions. For example, the *exponential* generating function

$$A(z) = \sum_{n\geq 0} \frac{a_n}{n!} z^n$$

is a very popular one (e.g., when studying labeled combinatorial structures as we do in Section 7.2). An exotic one is the *tree-like* generating function defined as

$$A(z) = \sum_{n\geq 0} a_n \frac{n^{n-1}}{n!} z^n$$

that finds a myriad of applications in analysis of algorithms, coding theory, combinatorics and information theory. We discuss its applications in this and the forthcoming chapters.

In some situations (e.g., enumeration problems), a combinatorial view on generating functions is more convenient. Let $\mathcal{S}$ be a set of objects (e.g., words, apples, numbers, graphs), and let $\alpha \in \mathcal{S}$. We write $w(\alpha)$ for a weight function that we interpret here as the size of α, that is, $w(\alpha) = |\alpha|$. Define the counting generating function $A(z)$ as

$$A(z) = \sum_{\alpha \in \mathcal{S}} z^{w(\alpha)}.$$

Certainly, this definition is equivalent to the previous one if one sets a_n to be the number of objects α satisfying $w(\alpha) = n$. This approach is quite useful in studying combinatorial properties of objects (e.g., the number of connected components in a graph, the number of rooted trees).

This chapter is the first one in the last part of this book, and it is devoted to analytic methods. We shall discuss formal power series and analytic theory of ordinary generating functions, exponential generating functions, probability generating functions, and Dirichlet series. In particular, we construct generating functions for a class of combinatorial structures, explain the Lagrange formula, discuss the Borel transform and the Perron-Mellin summation formula. We illustrate our discussion on several examples and three major applications. First, we show how exponential generating functions can be used to solve exactly a certain class of recurrences arising in the analysis of strings. Then, we consider a given string, called *pattern*, and ask how many times it can occur (overlapping allowed) in a random string called the text. Finally, we use the theory of Dirichlet series to derive the total number of 1-digits in the binary representations of $1, 2, \ldots, n$.

There are a number of books dealing with generating function. Generating functions in combinatorics are thoroughly discussed in Comtet [56], Goulden and Jackson [145], Stanley [311] and Wilf [352]. Applications of generating functions to discrete mathematics and analysis of algorithms can be found in Sedgewick and Flajolet [302], and Graham, Knuth and Patashnik [136].

7.1 Ordinary Generating Functions

We shall discuss below elementary properties of ordinary generating functions (in short: OGF). In particular, we show how to manipulate formal power series, how to apply them to study combinatorial objects, and present elements of analytic theory of generating functions.

7.1.1 Formal Power Series

Define for a sequence $\{a_n\}_{n=0}^{\infty}$ an algebraic object

$$A(z) = \sum_{n \geq 0} a_n z^n$$

that we also write as

$$\{a_n\}_0^\infty \quad \longleftrightarrow \quad A(z).$$

It is convenient to introduce another notation, namely

$$[z^n]A(z) := a_n$$

which extracts the coefficient of $A(z)$ at z^n. In the *formal power series* approach there is no restriction on the domain of variation of z, and the series is not required to be convergent. In fact, we look at $A(z)$ (or simply A) as an algebraic object. The set of formal power series with standard addition and multiplication operations forms a *ring* $K[z]$. We recall that multiplication of A and B is defined as

$$AB \quad \longleftrightarrow \quad \left\{\sum_{k=0}^{n} a_k b_{n-k}\right\}_0^\infty .$$

With respect to multiplication, the ring $K[z]$ is a commutative semi group with neutral element $1 + \sum_{n\geq 1} 0 \cdot z^n = 1$. The *reciprocal* of series A is $1/A$ understood as

$$A \cdot (1/A) = 1$$

provided $a_0 \neq 0$. For example, $(1-z)$ is reciprocal to $A = \sum_{n\geq 0} z^n$ since

$$(1-z)(1+z+z^2+\cdots) = 1.$$

We denote by $1/(1-z)$ the series $A = \sum_{n\geq 0} z^n$. We must underline that $1/(1-z)$ does *not* have any analytical meaning in the formal power series theory; it is just a name of the power series generated by $\{1\}_0^\infty$. We know, however, that $\frac{1}{1-z} = \sum_{n\geq 0} z^i$ for $|z| < 1$, but for this to make sense we need analytic theory of generating functions that is discussed in the sequel.

We can define other operations and manipulation rules on the ring $K[z]$. The *derivative* $DA(z)$ of $A(z)$, also denoted as $A'(z)$, is

$$DA(z) = \sum_{n\geq 0} n a_n z^{n-1}.$$

The integration of $A(z)$ is

$$\int_0^z A(t)dt = \sum_{n\geq 1} \frac{a_{n-1}}{n} z^n.$$

Some manipulation rules are presented in Table 7.1. Rules (7.1)–(7.6) are easy to derive and left as a warm-up exercise. For example, (7.6) is a consequence of our definition (7.4) and the multiplication rule.

Example 7.1 *Harmonic Sum Power Series*

Table 7.1: Basic Properties of Formal Power Series

Let $\{a_n\}_{n=0}^{\infty}$ be a sequence and $A(z)$ or simply A its ordinary generating functions understood as a formal power series, that is, $\{a_n\}_0^\infty \longleftrightarrow A$. Let zD be a formal derivative operator multiplied by z, that is,

$$(zD)A \longleftrightarrow \{na_n\}_0^\infty.$$

The following relationships hold:

$$\{a_n\}_0^\infty + \{b_n\}_0^\infty \longleftrightarrow A + B, \tag{7.1}$$

$$\{a_{n+k}\}_0^\infty \longleftrightarrow \frac{A(z) - a_0 - \cdots - a_{k-1}z^{k-1}}{z^k}, \tag{7.2}$$

$$\{n^k a_n\}_0^\infty \longleftrightarrow (zD)^k A, \tag{7.3}$$

$$\{1\}_0^\infty \longleftrightarrow A := \frac{1}{1-z}, \tag{7.4}$$

$$\left\{\sum_{n_1+\cdots+n_k=n} a_{n_1}\cdots a_{n_k}\right\}_0^\infty \longleftrightarrow A^k, \tag{7.5}$$

$$\left\{\sum_{j=0}^{n} a_j\right\}_0^\infty \longleftrightarrow \frac{A}{1-z}. \tag{7.6}$$

Let $H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$ be the n-th harmonic number. What is the formal power series of H_n? By Rule (7.6) it is $A(z)/(1-z)$ where

$$A(z) = \sum_{n\geq 1} \frac{z^n}{n}.$$

Observe that the derivative $DA = 1/(1-z)$, so that

$$A = D^{-1}(1/(1-z)),$$

where D^{-1} is the inverse operator to D. It is convenient to give a name for such a power series. We shall see later that a good name is

$$D^{-1}(1/(1-z)) := \log\frac{1}{1-z}.$$

Thus

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

In fact, the reader may verify below identities that follow directly from the above and rule (7.6)

$$\begin{aligned} 1/(1-z)^2 \log(1/(1-z)) &= \sum_{n=1}^{\infty} z^n \sum_{k=1}^{n} H_k, \\ 1/(1-z)^2 \log(1/(1-z)) + 1/(1-z)^2 &= \sum_{n=0}^{\infty} (n+1) H_{n+1} z^n. \end{aligned}$$

Comparing these two expressions, we can obtain the following

$$\sum_{k=1}^{n} H_k = (n+1)H_{n+1} - n - 1.$$

Example 7.2 *Ordered Partition of an Integer*

Let $f_k(n)$ be the number of ways that the nonnegative number n can be partitioned as an *ordered* sum of k nonnegative integers. For example: $f_2(3) = 4$ since $3 = 3+0 = 2+1 = 1+2 = 0+3$. Observe that

$$f_k(n) = \sum_{n_1+\cdots+n_k=n} 1,$$

hence by Rule (7.5) we have

$$\sum_{n=0}^{\infty} f_k(n) z^n = 1/(1-z)^k.$$

We shall need analytic theory of generating functions to recover $f_k(n)$ from the above. □

7.1.2 Combinatorial Calculus

As we mentioned before, generating functions are often used in the enumeration of combinatorial structures. This is known as *combinatorial calculus*. In such enumerations, it is desirable to introduce classes of combinatorial objects. There are excellent books dealing with such problems (e.g., Flajolet and Sedgewick [118], Goulden and Jackson [145] and Stanley [311]), so we restrict our discussion to two simplest cases and a few examples. We shall follow here the presentation of Flajolet and Sedgewick [118].

Let $\mathcal{A}$ be a class of combinatorial objects (e.g., set of binary trees, strings, graphs). If $\mathcal{A}$ is a set of words, then we often call it a *language*. The associated generating function is defined as

$$A(z) = \sum_{\alpha \in \mathcal{A}} z^{|\alpha|},$$

where $|\alpha|$ is the size of the object α. Let now $\mathcal{B}$ be another class. We can combine $\mathcal{A}$ and $\mathcal{B}$ into a new class. We define "disjoint union" denoted as $\mathcal{A} + \mathcal{B}$ that gives the class consisting of disjoint copies of the members of $\mathcal{A}$ and $\mathcal{B}$. We can also define a "Cartesian product" denoted as $\mathcal{A} \times \mathcal{B}$ that produces a class of ordered pairs of objects, one from $\mathcal{A}$ and the other from $\mathcal{B}$. It is easy to see that

$$\mathcal{C} = \mathcal{A} + \mathcal{B} \quad \longleftrightarrow \quad C(z) = A(z) + B(z), \tag{7.7}$$
$$\mathcal{C} = \mathcal{A} \times \mathcal{B} \quad \longleftrightarrow \quad C(z) = A(z)B(z). \tag{7.8}$$

For example, to prove (7.8) we argue as follows

$$C(z) = \sum_{\gamma \in \mathcal{A} \times \mathcal{B}} z^{|\gamma|} = \sum_{\alpha \in \mathcal{A}} \sum_{\beta \in \mathcal{B}} z^{|\alpha| + |\beta|} = A(z)B(z).$$

Example 7.3 *Enumeration of Strings*

Let $\mathcal{A}$ be the set of all binary strings with no two consecutive 1 bits. Such strings are either an object of size zero that we denote as ϵ, or a single 1 or a recursive (composed) object consisting of 0 or 10 followed by a string with no two consecutive 1 bits (hence belonging to $\mathcal{A}$). In other words, the class $\mathcal{A}$ can be written formally as

$$\mathcal{A} = \epsilon + \{1\} + \{0, 01\} \times \mathcal{A}.$$

By (7.7) and (7.8), the above immediately translates into

$$A(z) = 1 + z + (z + z^2)A(z).$$

Example 7.4 *Enumeration of Binary Trees*

Let $\mathcal{B}$ be the class of (unlabeled) binary trees which are either empty or consist of a root and a left binary tree and a right binary tree. In symbolic calculus

$$\mathcal{B} = \epsilon + \{\text{node}\} \times \mathcal{B} \times \mathcal{B}$$

that translates into

$$B(z) = 1 + zB^2(z).$$

Example 7.5 *Partitions of Integers*

We consider now a little more sophisticated problem. As in Example 2, we partition the integer n into unlimited number of terms, but this time we ignore the order. In other words, we assume

$$n = n_1 + n_2 + \cdots, \qquad \text{subject to} \qquad 1 \le n_1 \le n_2 \le \cdots$$

Let p_n be the number of such solutions. For example, $p_4 = 5$ since $4 = 1 + 1 + 1 + 1 = 1 + 1 + 2 = 1 + 3 = 2 + 2 = 4$. To find p_n, we represent the class $\mathcal{C}$ of all partitions as a Cartesian product of simpler sets. Define $\{i^*\}$ as

$$\{i^*\} = \{\epsilon;\ i;\ i, i;\ i, i, i;\ \cdots\},$$

that is, $\{i^*\}$ is a set consisting of unlimited repetitions of i (including the zeroth repetition). Clearly,

$$\mathcal{C} = \{1^*\} \times \{2^*\} \times \cdots$$

For example, the partition $1+1+2=4$ takes the third element from $\{1^*\}$, the second element from $\{2^*\}$, and the first element from all others. If $I_i(z)$ denotes the generating function of $\{i^*\}$, then

$$I_i(z) = \sum_{n=0}^{\infty} z^{in} = 1/(1-z^i)$$

and therefore

$$C(z) = \sum_{n=0}^{\infty} p_n z^n = \prod_{i=1}^{\infty} \frac{1}{(1-z^i)}.$$

Now, once we know the formula we can find an even easier derivation. Observe that n can also be represented as

$$n = 1n_1 + 2n_2 + \cdots + nn_n$$

where n_i is the number of i's in the partitions. Then

$$\begin{aligned}
\sum_{n=0}^{\infty} p_n z^n &= \sum_{n=0}^{\infty} \sum_{1n_1+2n_2+\cdots+nn_n=n} x^{1n_1+2n_2+\cdots+nn_n} \\
&= (1+z+z^2+\cdots)\cdot(1+z^2+z^6+\cdots)\cdots(1+z^i+z^{2i}+\cdots)\cdots \\
&= \prod_{i=1}^{\infty} \frac{1}{(1-z^i)}.
\end{aligned}$$

This latter approach is quite useful in some situations. For example, in Section 3.3.3 we used (3.23), that we repeat below

$$\sum_{m=0}^{\infty} \sum_{n_1+2n_2+\cdots mn_m=m} \frac{x^m}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{m} c_i^{n_i} = \exp\left(\sum_{i=1}^{\infty} c_i x^i\right),$$

to prove the result of Karlin and Ost (cf. Theorem 3.8). We delayed the proof of (3.23), and now are are ready to derive it. We use the same approach as above. Starting from the right hand-side we have

$$\begin{aligned}
\exp\left(\sum_{i=1}^{\infty} c_i x^i\right) &= \left(1 + c_1 x + \frac{c_1 x^2}{2!} + \frac{c_1 x^3}{3!} + \cdots\right)\cdots\left(1 + c_i x^i + \frac{c_1 x^{2i}}{2!} + \frac{c_1 x^{3i}}{3!} + \cdots\right)\cdots \\
&= \sum_{m=0}^{\infty} \sum_{n_1+2n_2+\cdots mn_m=m} \frac{x^{n_1+2n_2+\cdots mn_m}}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{\infty} c_i^{n_i} \\
&= \sum_{m=0}^{\infty} \sum_{n_1+2n_2+\cdots mn_m=m} \frac{x^m}{n_1!n_2!\cdots n_m!} \prod_{i=1}^{m} c_i^{n_i}.
\end{aligned}$$

This proves the desired identity. $\square$

7.1.3 Elements of Analytic Theory

In the formal theory of power series, we viewed the ordinary generating function $A(z)$ as an algebraic element and z is added for a convenience. Convergence of the series was not an issue. If, however, the power series converges to an analytic function, then we may use a whole machinery of the theory of analytic functions to find analytic information about $A(z)$. In particular, such information can be used to obtain asymptotic expansion of the coefficients (cf. Chapter 8).

The first problem we face is to see under what conditions the series

$$A(z) = \sum_{n=0}^{\infty} a_n z^n$$

converges for z complex. Fortunately, following Hadamard we can give a fairly complete answer to this question.

Theorem 7.1 (Hadamard) *There exists a number $0 \leq R \leq \infty$ such that the series $A(z)$ converges for $|z| < R$ and diverges for $|z| > R$. The radius of convergence R can be expressed as*

$$R = \frac{1}{\limsup_{n\to\infty} |a_n|^{\frac{1}{n}}}, \tag{7.9}$$

where by convention $1/0 = \infty$ and $1/\infty = 0$. The function $A(z)$ is analytic for $|z| < R$.

Proof. The cases $R = 0$ and $R = \infty$ are left to the reader. Let $0 < R < \infty$ and $|z| < R$. We prove that the series $A(z)$ converges. Choose $\varepsilon > 0$ such that

$$|z| < \frac{R}{1 + \varepsilon R}.$$

By definition of $\limsup$, there exists N such that for all $n > N$ we have from (7.9)

$$|a_n|^{\frac{1}{n}} < R^{-1} + \varepsilon.$$

Thus

$$|a_n||z|^n < \Big(|z|(R^{-1} + \varepsilon)\Big)^n \leq \alpha^n$$

for $\alpha < 1$. The series converges absolutely for $|z| < R$.

To prove the second part of the theorem, let now $|z| > R$. We show that then $a_n z^n \not\to 0$, hence the series cannot converge. Let for $\varepsilon > 0$

$$\theta = |z/R - \varepsilon z|.$$

Since $|z| > R$ we can choose $\varepsilon > 0$ such that $\theta > 1$. But, by (7.9) and the definition of lim sup for *infinitely* many n we have

$$|a_n|^{\frac{1}{n}} > R^{-1} - \varepsilon,$$

hence for infinitely many n

$$|a_n||z|^n > \left|(R^{-1} - \varepsilon)z\right|^n = \theta^n,$$

which increases exponentially. This completes the proof. ■

Hadamard's theorem sheds light on asymptotics of the coefficients a_n. We will devote the whole Chapter 8 to this problem. Here we only observe that Hadamard's theorem tells us what is happening with $A(z)$ for $|z| < R$ and $|z| > R$, but nothing about $|z| = R$. Really? In fact, from the theorem we should conclude that the function $A(z)$ has at least one *singularity* (a point where the function ceases to be well defined) on the circle $|z| = R$. To see this, we argue by contradiction. If there is no singularity on $|z| = R$, then we can cover it with a finite number of disks (by the Heine-Borel theorem) on which $A(z)$ is analytic. This would imply that $A(z)$ converges in a larger disk $|z| < R + \varepsilon$, which is the desired contradiction.

So far we were concerned with the calculation of generating functions from their coefficients. But often we need to find the coefficients knowing the generating function. One way to solve this problem is to have a list of generating functions together with their corresponding coefficients and find $[z^n]A(z)$ by inspection. In Table 7.2 we list some classic generating functions. We illustrate their usage on several examples.

Example 7.6 *A Simple Recurrence*

We start with a simple recurrence. Let

$$a_{n+1} = 2a_n + n, \qquad n \geq 0,$$

with $a_0 = 1$. Multiplying both sides by z^n and summing over all $n \geq 0$, we obtain, after using rules (7.2)–(7.3),

$$\frac{A(z) - 1}{z} = 2A(z) + \frac{z}{(1-z)^2}$$

which can be solved to yield

$$\begin{aligned} A(z) &= \frac{1 - 2z + 2z^2}{(1-z)^2(1-2z)} \\ &= \frac{2}{1-2z} - \frac{1}{(1-z)^2}. \end{aligned}$$

The last line is a partial fraction expansion of the first line. By Entry 1 of Table 7.2 we immediately obtain

$$a_n = 2^{n+1} - n - 1$$

1. Binomial coefficients	$\frac{1}{1-z-uz}$	$=$	$\sum_{n,k\geq 0} \binom{n}{k} u^k z^n$
	$\frac{z^k}{(1-z)^{k+1}}$	$=$	$\sum_{n\geq k} \binom{n}{k} z^n$
	$\frac{1}{(1-z)^{k+1}}$	$=$	$\sum_{n\geq 0} \binom{n+k}{n} z^n$
	$(1+z)^\alpha$	$=$	$\sum_{n\geq 0} \binom{\alpha}{n} z^n$
2. Bernoulli numbers	$\frac{z}{(e^z-1)}$	$=$	$\sum_{n=0}^{\infty} B_n \frac{z^n}{n!}$
3. Catalan numbers	$\frac{1-\sqrt{1-4z}}{2z}$	$=$	$\sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} z^n$
4. Reciprocals	$\ln \frac{1}{1-z}$	$=$	$\sum_{n\geq 1} \frac{z^n}{n}$
5. Harmonic numbers	$\frac{1}{1-z}\ln \frac{1}{1-z}$	$=$	$\sum_{n\geq 1} H_n z^n$
6. Fibonacci numbers	$\frac{z}{1-z-z^2}$	$=$	$\sum_{n\geq 0} F_n z^n$
7. Tree function	$ze^{T(z)} = T(z)$	$=$	$\sum_{n\geq 1} \frac{n^{n-1}}{n!} z^n$
8. Stirling numbers of the first kind	$\frac{1}{(1-z)^u}$	$=$	$\sum_{n,k\geq 0} \left[{n \atop k}\right] u^k \frac{z^n}{n!}$
	$u(u+1)\cdots(u+n-1)$	$=$	$\sum_{k\geq 0} \left[{n \atop k}\right] u^k$
9. Stirling numbers of the second kind	$e^{u(e^z-1)}$	$=$	$\sum_{n,k\geq 0} \left\{{n \atop k}\right\} u^k \frac{z^n}{n!}$
	$(e^z-1)^n$	$=$	$n!\sum_{k\geq n} \left\{{k \atop n}\right\} \frac{z^k}{k!}$
	$\frac{z^k}{(1-z)(1-2z)\cdots(1-kz)}$	$=$	$\sum_{n\geq 0} \left\{{n \atop k}\right\} z^n$

Table 7.2: Classic Generating Functions

for all $n \geq 0$.

Example 7.7 *Enumeration of Unlabeled Trees – Catalan Numbers*

In Example 4 we proved that the ordinary generating function, $B(z)$, of the number of *unlabeled binary trees* satisfies $B(z) - 1 = zB^2(z)$. It can be solved to yield

$$B(z) = \frac{1 - \sqrt{1-4z}}{2z} .$$

Let us now find $[z^n]B(z)$, that is, the nth Catalan number. We shall prove that

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

Observe that by the binomial series (cf. Table 7.2) we have

$$zB(z) = -\frac{1}{2}\sum_{n\geq 1}\binom{\frac{1}{2}}{n}(-4z)^n.$$

Extracting coefficients of both sides of the above, we obtain

$$\begin{aligned}
[z^n]B(z) &= -\frac{1}{2}\binom{\frac{1}{2}}{n+1}(-4)^{n+1} \\
&= -\frac{1}{2}\frac{\frac{1}{2}(\frac{1}{2}-1)\cdots(\frac{1}{2}-n)(-4)^n}{(n+1)!} \\
&= \frac{1\cdot 3\cdot 5\cdots(2n-1)2^n}{(n+1)!} \\
&= \frac{1}{n+1}\frac{1\cdot 3\cdot 5\cdots(2n-1)}{n!}\frac{2\cdot 4\cdot\cdot 6\cdots 2n}{1\cdot 2\cdot 3\cdots n} \\
&= \frac{1}{n+1}\binom{2n}{n}.
\end{aligned}$$

Example 7.8 *Stirling Numbers of Second Kind*

Let us enumerate the number of set partitions (i.e., a collection of nonempty pairwise disjoint subsets whose union is the original set) of the set $[n] := \{1, 2, \ldots, n\}$ into k subsets known also as classes. For example, the set [3] is partitioned into $k = 2$ classes as follows:

$$[3] = \{\{1\}\{2,3\}; \{1,3\}\{2\}; \{1,2\}\{3\}\}.$$

This number is known as the *Stirling number of the second kind*, and it is denoted as $\left\{ {n \atop k} \right\}$. We first find a recurrence on it, and then derive an explicit formula.

Let us start with a recurrence on $\left\{ {n \atop k} \right\}$. The set of partitions of $[n]$ into k classes can be split into two groups: In the first group, you will have all partitions in which n is in a separate class itself, that is, we have a subset of the form $\{n\}$. The second group contains all other partitions. It is easy to notice that the number of partitions in the first group is $\left\{ {n-1 \atop k-1} \right\}$. So, it remains to enumerate the second group.

Observe that the second group contains all partitions of $[n]$ in which the element n is not listed separately but together with other elements. Let us now delete n from all such partitions. We definitely obtain a partition of $[n-1]$ into k classes, but some classes will be repeated. Which one and how many of them? Since we have k classes, and n could appear in any of them, hence after the deletion of n we will end up with k copies of the same class. Thus, the second group contains $k\left\{ {n-1 \atop k} \right\}$ partitions. This proves the following recurrence

$$\left\{ {n \atop k} \right\} = \left\{ {n-1 \atop k-1} \right\} + k\left\{ {n-1 \atop k} \right\}, \qquad (n,k) \neq (0,0). \tag{7.10}$$

We also postulate that $\left\{ {0 \atop 0} \right\} = 1$ and $\left\{ {n \atop 0} \right\} = 0$ for $n \neq 0$.

To find the generating function

$$S_k(z) = \sum_{n \geq k} \left\{ {n \atop k} \right\} z^n$$

and an explicit formula for $\left\{ {n \atop k} \right\}$, we proceed as in the previous examples. The recurrence (7.10) translates into

$$S_k(z) = zS_{k-1}(z) + kzS_k(z)$$

for $k \geq 1$ and $S_0(z) = 1$. Thus,

$$S_k(z) = \frac{z}{1-kz} S_{k-1}(z),$$

and finally

$$S_k(z) = \sum_{n \geq k} \left\{ {n \atop k} \right\} z^n = \frac{z^k}{(1-z)(1-2z)\cdots(1-kz)}$$

which is Entry 9 of Table 7.2.

Now, we find an explicit formula for $\left\{ {n \atop k} \right\} = [z^n]S_k(z)$. As before, we need a partial fractional expansion that looks like this

$$\frac{1}{(1-z)(1-2z)\cdots(1-kz)} = \sum_{j=1}^{k} \frac{a_j}{(1-jz)}.$$

After multiplying both sides by $(1-rz)$ and evaluating at $z=1/r$ we obtain

$$a_r = (-1)^{k-r}\frac{r^{k-1}}{(r-1)!(k-r)!}.$$

In view of this, we immediately find

$$\begin{aligned}
\left\{ n \atop k \right\} &= [z^n]\frac{z^k}{(1-z)(1-2z)\cdots(1-kz)} \\
&= [z^{n-k}]\sum_{r=1}^{k}\frac{a_r}{(1-rz)} = \sum_{r=1}^{k} a_k r^{n-k} \\
&= \frac{1}{k!}\sum_{r=1}^{k}(-1)^{k-r}\binom{k}{r} r^n
\end{aligned}$$

which is not much simpler than the recurrence itself. □

7.1.4 Generating Functions Over an Arithmetic Progression

Finally, we deal with the following problem: How to compute sums over an arithmetic progression; e.g., $\sum_{k\geq 0}\binom{n}{3k}x^{3k}$. In general, let $A(z)$ be the ordinary generating function of a sequence a_n, and for fixed r define

$$A_r(z) = \sum_{n=rk,\ k\geq 0} a_n z^n = \sum_{k=0}^{\infty} a_{rk} z^{rk}.$$

Can we estimate $A_r(z)$ knowing $A(z)$? We can, and here is how it works. Let $\omega^r = 1$ be the rth root of unity. There are r such roots that we denote as ω_k for $k=0,1,\ldots,r-1$. In fact, $\omega_k = e^{(2\pi ik)/r}$. Observe that for r not dividing n (i.e., $r \not| n$) we have

$$\frac{1}{r}\sum_{k=0}^{r-1}\omega_k^n = \frac{1}{r}\sum_{k=0}^{r-1} e^{(2\pi ikn)/r} = \frac{1}{r}\frac{1-e^{2\pi in}}{1-e^{(2\pi in)/r}} = 0.$$

If $r|n$, then $\omega_k^n = 1$ for all $k=0,1,\ldots,r-1$. In summary,

$$\frac{1}{r}\sum_{\omega^r=1}\omega^n = \begin{cases} 1 & \text{if } r|n \\ 0 & \text{otherwise.}\end{cases} \tag{7.11}$$

This immediately implies

$$A_r(z) = \frac{1}{r}\sum_{k=0}^{r-1} A(z\omega_k). \tag{7.12}$$

Example 7.9 *A Combinatorial Sum*
We will find

$$R_n = \sum_k (-1)^k \binom{n}{3k}$$

applying (7.12) with $r = 3$. We first observe that

$$A(z) = \sum_k \binom{n}{k} z^k = (1+z)^n.$$

Since $R_n = A_3(-1)$, we obtain

$$\begin{aligned} A_3(-1) &= \frac{1}{3}\left(A(z) + A(z\omega_1) + A(z\omega_2)\right)\big|_{z=-1} \\ &= \frac{1}{3}\left((1-\omega_1)^n + (1-\omega_2)^n\right) \\ &= \frac{1}{3}\left(\left(\frac{3-\sqrt{3}i}{2}\right)^n + \left(\frac{1+\sqrt{3}i}{2}\right)^n\right) \\ &= \frac{2}{3}3^{\frac{n}{2}}\cos\left(\frac{n\pi}{3}\right) \end{aligned}$$

which would be hard to find without (7.12). □

7.2 Exponential Generating Functions

It is easy to construct a sequence a_n for which the ordinary generating function (OGF) either does not exist or cannot be computed in a closed form. Think of $a_n = n!$ or $a_n = n^n/n!$. Often, a very simple trick solves the problem. Namely, one defines a weighted generating function such that

$$A(z) = \sum_{n=0}^{\infty} a_n \frac{z^n}{w_n}$$

where w_n is the weight. The most popular and widely used weight is $w_n = n!$ that produces the *exponential* generating function. But other weights work remarkably well in a variety of applications; e.g., $w_n = n!/n^n$ leads to a *tree-like* generating function that we shall discuss in the sequel.

7.2.1 Elementary Properties

Let us first concentrate on exponential generating functions. We sometimes must distinguish between ordinary and exponential generating functions, therefore, we often write (lower-case

Table 7.3: Basic Properties of Exponential Generating Functions

Let $\{a_n\}_{n=0}^{\infty}$ be a sequence and $a(z)$ its exponential generating function. We write

$$\{a_n\}_0^\infty \quad \longleftrightarrow \quad a(z).$$

We shall also write $Da(z)$ or $a'(z)$ for the derivative of $a(z)$.
The following relationships hold:

$$\{a_n\}_0^\infty + \{b_n\}_0^\infty \quad \longleftrightarrow \quad a(z) + b(z), \tag{7.13}$$

$$\{a_{n+k}\}_0^\infty \quad \longleftrightarrow \quad D^k a(z) = \frac{d^k}{dz^k} a(z), \tag{7.14}$$

$$\{n^k a_n\}_0^\infty \quad \longleftrightarrow \quad (zD)^k a(z), \tag{7.15}$$

$$\left\{ \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k} \right\}_{n=0}^{\infty} \quad \longleftrightarrow \quad a(z)b(z). \tag{7.16}$$

letter) $a(z)$ for the exponential generating function of $\{a_n\}_{n=0}^{\infty}$, that is,

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

Elementary properties of exponential generating functions are listed in Table 7.3. As in the case of ordinary generating functions, we can view exponential generating functions as a formal power series, but we skip this formalism and assume that $a(z)$ is an analytic function. Clearly, what we said above about analytic theory of OGF applies here with a_n replaced by $a_n/n!$.

The properties (7.13)–(7.16) presented in Table 7.3 are easy to prove, and left for the reader. We only show here how to derive (7.16) from its ordinary generating function counterpart (7.5). If $a(z)$ and $b(z)$ are exponential generating functions, then

$$n![z^n]a(z)b(z) = \left[\frac{z^n}{n!}\right] a(z)b(z) = n! \sum_{k=0}^{n} \frac{a_k}{k!} \frac{b_{n-k}}{(n-k)!} = \sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}$$

which yields (7.16).

Example 7.10 *Bell Numbers*

In Example 8 we computed Stirling numbers of the second kind $\left\{ n \atop k \right\}$ that represent the number of ways the set $[n] = \{1, \ldots, n\}$ can be partitioned into k classes. What about if we are interested in the total number of partitions of $[n]$ irrespectively of how many classes are there. These numbers are called the *Bell numbers* and are denoted as b_n. We prove that

$$b_{n+1} = \sum_{k=0}^{n} \binom{n}{k} b_k$$

with $b_0 = 1$ by convention. Indeed, let us consider a collection of $[n+1]$ partitions. Its cardinality is b_{n+1}. We split this collection into $k = 0, 1, \ldots, n$ groups such that the kth group consists of all partitions of $[n+1]$ in which the class that includes symbol $n+1$ contains exactly k others symbols. Certainly, the above follows.

To obtain the exponential generating function $b(z)$ of Bell numbers, we use (7.14) and (7.16) which yields

$$b'(z) = e^z b(z), \qquad b(0) = 1.$$

Solving it one arrives at

$$b(z) = \exp\left(e^z - 1\right).$$

A closed form formula for b_k can be derived, but this is left as an exercise.

Example 7.11 *A Binomial Sum*

In some computations, one needs to compute the generating function of the binomial sum s_n defined as

$$s_n = \sum_{k=0}^{n} \binom{n}{k} a_k$$

where a_k is a sequence. Clearly by (7.16) we have

$$\sum_{n \geq 0} s_n \frac{z^n}{n!} = e^z a(z)$$

where $a(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}$. The situation is a little more complex if one wants to compute the ordinary generating function of s_n. But

$$s_n = \sum_{k=0}^{n} \binom{n}{k} a_k \qquad \longleftrightarrow \qquad S(z) = \frac{1}{1-z} A\left(\frac{z}{1-z}\right). \tag{7.17}$$

Indeed, by Entry 1 of Table 7.2 we have

$$\begin{aligned} \frac{1}{1-z} A\left(\frac{z}{1-z}\right) &= \sum_{m=0}^{\infty} a_m z^m \frac{1}{(1-z)^{m+1}} = \sum_{m=0}^{\infty} a_m z^m \sum_{j=0}^{\infty} \binom{m+j}{j} z^j \\ &= \sum_{n=0}^{\infty} z^n \sum_{k=0}^{n} \binom{n}{k} a_k \end{aligned}$$

as desired. □

7.2.2 Labeled Combinatorial Structures

As in the case of ordinary generating functions, the exponential generating functions have a combinatorial interpretation. We consider here *labeled* structures (e.g., permutations, cycles in a permutation, set partitions, labeled trees, labeled graphs) where each "atom" of a structure (e.g., a node in graphs or trees) has a distinctive integer label. Let $\mathcal{A}$ be a class of labeled combinatorial objects, and let a_n be the counting sequence (i.e., the number of objects of size n). Then the exponential generating function (or in short: EGF) of $\mathcal{A}$ is

$$a(z) = \sum_{\alpha \in \mathcal{A}} \frac{z^{|\alpha|}}{|\alpha|!} = \sum_{n \geq 0} a_n \frac{z^n}{n!}$$

since there are $n!$ orderings of labels.

Foata presented in [127] a detailed formalization of labeled constructions. There are many excellent surveys and books that treat exponential combinatorial families in depth (cf. [102, 118, 160, 302, 352]), so here we restrict our attention only to a few examples. We shall follow the presentation of Flajolet and Sedgewick [118].

In *labeled combinatorial structures* each object consists of atoms that are marked by distinct integers. The *size* of a structure is the number of its atoms (e.g., number of nodes in a graph or a tree). We assume, however, that such structures are *well-labeled* by insisting that the labels are consecutive integers starting from 1. We allow *re-labeling* that *preserves* the order relations between labels. For example, a structure with labels $\{5, 9, 11\}$ can be relabeled as $\{1, 2, 3\}$, but not $\{1, 3, 2\}$. Knowing this, we can introduce two new *labeled constructions*, namely, *disjoint union* and *product*.

Disjoint union of structures $\mathcal{A}$ and $\mathcal{B}$ denoted as $\mathcal{A} + \mathcal{B}$ is defined as in the unlabeled case except that we need to relabel the composite structure to get well-labeled structures. For example, $\{1\} + \{1, 2\} = \{1, 2, 3\}$. In a similar fashion, we define the product $\mathcal{C} = \mathcal{A} \star \mathcal{B}$ of $\mathcal{A}$ and $\mathcal{B}$ by forming ordered pairs from $\mathcal{A} \times \mathcal{B}$ and performing all possible *order consistent relabeling* ensuring that the resulting pairs are well-labeled. Using the same argument as above for the unlabeled structures (ordinary generating functions), we easily prove the following translations

$$\mathcal{C} = \mathcal{A} + \mathcal{B} \quad \longleftrightarrow \quad c(z) = a(z) + b(z), \tag{7.18}$$

$$\mathcal{C} = \mathcal{A} \star \mathcal{B} \quad \longleftrightarrow \quad c(z) = a(z)b(z). \tag{7.19}$$

The last relation can be also proved by observing that the counting number c_n of $\mathcal{C}$ is related to the counting numbers a_n and b_n of $\mathcal{A}$ and $\mathcal{B}$ by

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

that leads directly to the product of the associated exponential generating functions.

We illustrate the above constructions on two interesting examples, namely, counting labeled trees (i.e., Cayley's trees) and the number of cycles in a permutation (i.e., Stirling numbers of the first kind).

Example 7.12 *Labeled or Cayley's Trees*

Let $\mathcal{T}$ be the class of all labeled rooted trees called also the Cayley trees. Think of such trees as labeled connected graphs without cycles and with one node marked as the root (e.g., subtrees are not ordered; there are no left and right subtrees as shown in Figure 7.1). Our goal is to find the number t_n of all labeled rooted trees of size n. In Figure 7.1 we show all $t_3 = 9$ labeled trees with $n = 3$ nodes. Let $T(z)$ be the EGF of t_n (we write $T(z)$ instead of $t(z)$ for historical reasons). Observe that a rooted labeled tree consists of the root and either one rooted labeled subtree or two rooted labeled subtrees or k rooted labeled subtrees, and so on. But k rooted labeled subtrees can be symbolically represented as a product of k rooted labeled trees $\mathcal{T}$, that is, $\mathcal{T}^k$. Finally, since the subtrees are *not* ordered we must divide the product $\mathcal{T}^k$ by $k!$ to construct the structure $\mathcal{T}$ (we recall that $\mathcal{T}^k$ is an order product). Having this in mind, and writing `root` for the root node, we can represent $\mathcal{T}$ as

$$\mathcal{T} = \texttt{root} \star \left(\epsilon + \mathcal{T} + \frac{1}{2!}\mathcal{T}^2 + \frac{1}{3!}\mathcal{T}^3 + \cdots + \frac{1}{k!}\mathcal{T}^k + \cdots \right).$$

This translates into

$$T(z) = ze^{T(z)}. \tag{7.20}$$

Thus $T(z)$ is given only implicitly. We will return to this problem in Section 7.3.2 where we show that $t_n = [z^n]T(z) = n^{n-1}$.

Example 7.13 *Cycles in Permutations or Stirling Numbers of the First Kind.*

Let $\mathcal{K}_n$ represent the subclass of circular permutations of size n. A permutation is circular if a circular shift leads to the same permutation (e.g., $(1,3,2)$ is the same circular permutation as $(3,2,1)$ but not the same as $(1,2,3)$). It is easy to see that the number of circular permutations k_n is $k_n = (n-1)!$. Indeed, we split $n!$ permutations of $[n]$ into $(n-1)!$ groups of size n such that two permutations belong to the same group if one is a circular shift of the other. But, $nk_n = n!$, hence the claim.

Now, let us consider a more interesting question. We consider a permutation of $[n]$ and identify all its cycles. For example, the permutation

$$\begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 10 & 1 & 7 & 9 & 3 & 5 & 6 & 4 & 8 & 2 \end{matrix}$$

contains three cycles, namely: $(1,10,2)$, $(3,7,6,5)$ and $(4,9,8)$ that can be represented as cycles digraphs (cf. Figure 7.2). Let $\mathcal{P}_n^r$ be a subclass of all permutations of $[n]$ having r cycles. If $\mathcal{K}$ and $\mathcal{P}^{(r)}$ denote the classes of all circular permutations and permutations with r cycles, respectively, then a little thought reveals that

$$\mathcal{P}^{(r)} = \frac{1}{r!}\mathcal{K}^r.$$

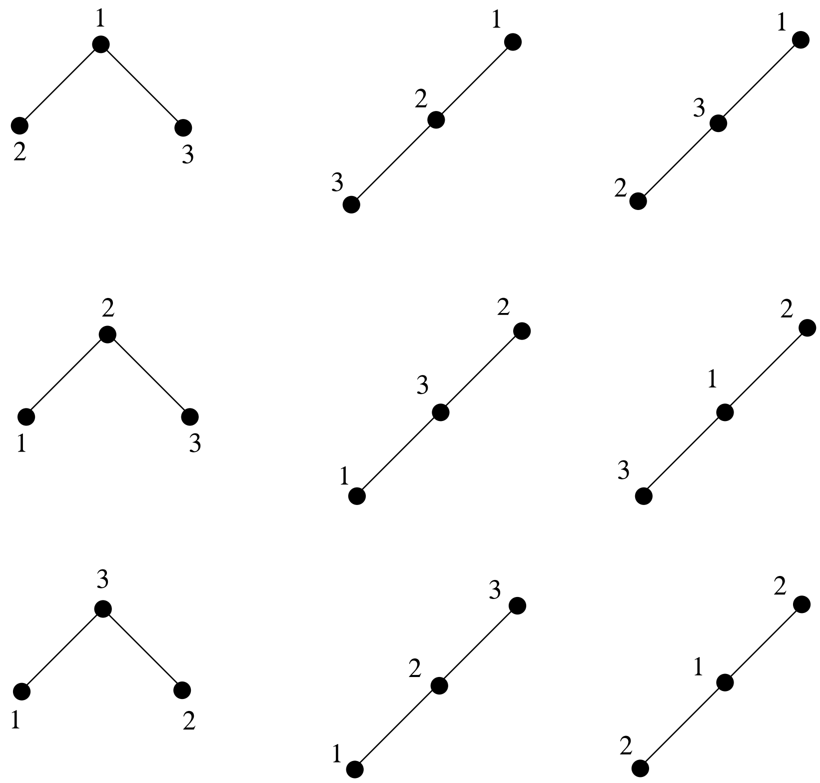

Figure 7.1: Nine rooted labeled trees built on $n = 3$ nodes with roots at the top.

This translates into the following EGF equation

$$P^{(r)}(z) = \frac{1}{r!} K^r(z).$$

But

$$K(z) = \sum_{n \geq 1} (n-1)! \frac{z^n}{n!} = \log \frac{1}{1-z}.$$

The numbers

$$\begin{bmatrix} n \\ r \end{bmatrix} := \frac{n!}{r!} [z^n] \left(\log \frac{1}{1-z} \right)^r \tag{7.21}$$

are called *Stirling number of the first kind* and represent the number of permutations with r cycles.

Let us see if we can get a more explicit formula for the generating function $P^{(r)}(z)$. We

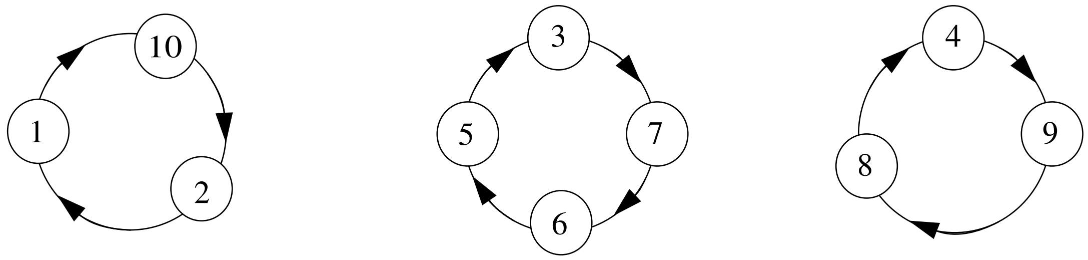

Figure 7.2: Three Cycles of the Permutation of Example 13.

first introduce a *bivariate* generating function

$$P(z,u) = \sum_{r\geq 0} P^{(r)}(z)u^r.$$

Observe that

$$P(z,u) = \sum_{r\geq 0}\left(\log\frac{1}{1-z}\right)^r \frac{u^r}{r!} = \exp\left(-u\log(1-z)\right) = \frac{1}{(1-z)^u}.$$

Using Entry 1 of Table 7.2, we obtain

$$\begin{aligned}\left[\frac{z^n}{n!}\right] P(z,u) &= \sum_{r=0}^{n} \left[{n \atop r}\right] u^r \\ &= \left[\frac{z^n}{n!}\right](1-z)^{-u} = n!\binom{u+n-1}{n} \\ &= u(u+1)\cdots(u+n-1).\end{aligned}$$

This agrees with Entry 8 of Table 7.2. $\square$

7.3 Cauchy, Lagrange and Borel Formulas

We have been lucky so far since in order to find $[z^n]A(z)$ we only need to look up an entry in Table 7.2. But this is not always the case as already seen in Example 12. We discuss below two formulas, one by Cauchy and the other one by Lagrange, that provide a guidance for finding $[z^n]A(z)$. We learn how to extract coefficients from generating functions given either explicitly (cf. Cauchy's formula) or implicitly (cf. Lagrange's Inversion Formula). We shall also discuss Borel transform that allows to transfer an exponential generating function into an ordinary generating function.

7.3.1 Cauchy Coefficient Formula

We start with the Cauchy formula that we already discussed in Chapter 2 (cf. Theorem 2.6). Observe that if $A(z)$ is analytic in the vicinity of $z = 0$, then

$$a_n = [z^n]A(z) = \frac{1}{2\pi i} \oint \frac{A(z)}{z^{n+1}} dz \tag{7.22}$$

where the integral is around any simple curve encircling $z = 0$. Indeed, since $A(z)$ is analytic at $z = 0$, it has a convergent series representation, say

$$A(z) = \sum_{k \geq 0} a_k z^k.$$

Thus

$$\oint \frac{A(z)}{z^{n+1}} dz = \sum_{k \geq 0} a_k \oint z^{-(n+1-k)} dz = 2\pi i a_n,$$

where the last equality follows from (2.22) which we repeat here

$$\oint z^{-n} dz = \int_0^1 e^{-2\pi i n} dt = \begin{cases} 2\pi i & \text{for } n = 1 \\ 0 & \text{otherwise.} \end{cases}$$

The interchange of the integral and sum above is justified since the series converges uniformly. Using Cauchy's formula and Cauchy's residue theorem we can evaluate exactly or asymptotically the coefficient a_n. Since the Cauchy formula (7.22) is mostly used in the asymptotic evaluation of the coefficients, we delay our discussion till the next chapter.

Finally, we should mention that Cauchy's formula is often used to establish a bound on the coefficients a_n. Let

$$M(R) = \max_{|z| \leq R} |A(z)|.$$

Then for all $n \geq 0$

$$|a_n| \leq \frac{M(R)}{R^n} \tag{7.23}$$

which follows immediately from (7.22) and a trivial majorization under the integral.

7.3.2 Lagrange Inversion Formula

Cauchy's formula can also be used to prove Lagrange's formula that allows to extract coefficient of *implicitly* defined generating functions. We already saw examples of generating functions that are defined by functional equations. We were lucky so far to be able to solve explicitly such functional equations, We cannot count on this forever. For example, while enumerating *unlabeled* trees we found an explicit formula discussed in Example 7, but this is not any more true for enumerating *labeled trees*. In the latter case, the generating function $T(z)$ of the number of *labeled trees* satisfies $T(z) = ze^{T(z)}$ (cf. Example 12). This function does not have a simple, explicit solution, but with the help of the Lagrange formula we shall find easily the coefficients.

We first digress and examine the inversion problem which is at the heart of the Lagrange formula. Define explicitly a function $x(z)$ as a solution of

$$x(z) = z\Phi(x(z)) \tag{7.24}$$

provided it exists. We assume that $\Phi(u)$ is analytic at some point x_0, and $\Phi(0) \neq 0$ for the problem to be well posed. To study the existence issue, let us rephrase (7.24) as

$$\Theta(x(z)) = z, \qquad \text{where} \qquad \Theta(u) = \frac{u}{\Phi(u)}. \tag{7.25}$$

In other words, we are looking for conditions under which the inverse function $x(z) = \Theta^{-1}(z)$ exists. We shall see that for the existence of the inverse function in a vicinity of x_0 we must require that $\Theta'(x_0) \neq 0$.

Since $\Theta(x)$ is analytic at x_0, we have

$$\Theta(x) = \Theta(x_0) + (x - x_0)\Theta'(x_0) + \frac{1}{2}(x - x_0)^2\Theta''(x_0) + \cdots .$$

As a first order approximation, we find

$$x - x_0 \sim \frac{1}{\Theta'(x_0)}(z - z_0)$$

provided $\Theta'(x_0) \neq 0$. We can further iterate the above to derive a better approximation, namely

$$x - x_0 = \frac{1}{\Theta'(x_0)}(z - z_0) - \frac{1}{2}\frac{\Theta''(x_0)}{[\Theta'(x_0)]^3}(z - z_0)^2 + O((z - z_0)^3). \tag{7.26}$$

Iterating many times we can get a convergent series representation of the inverse $\Theta^{-1}(x)$ around x_0 provided $\Theta'(x_0) \neq 0$. We should mention, that the problem of finding the inverse function is known in literature as the *implicit-function theorem* or *inverse mapping theorem.* The reader is referred to any standard book on analysis or complex analysis (e.g., [285, 353]) for further reading.

We can continue the above iterative process, but unfortunately it becomes more and more complicated. We would like to extract coefficients of the inverse function $\Theta^{-1}(x)$ or $x(z)$ in a more direct and simpler way. Is it possible? The answer is emphatically *yes*, and it is known as the *Lagrange inversion formula.*

Theorem 7.2 (Lagrange Inversion Formula) *Let $\Phi(u)$ be analytic in some neighborhood of $u = 0$ with $[u^0]\Phi(u) \neq 0$, and let $X(z)$ be a solution of*

$$X = z\Phi(X). \tag{7.27}$$

The coefficients of $X(z)$ or in general $\Psi(X(z))$ where Ψ is an analytic function in a neighborhood of the origin satisfy

$$[z^n]X(z) = \frac{1}{n}[u^{n-1}]\left(\Phi(u)\right)^n, \tag{7.28}$$

$$[z^n]\Psi(X(z)) = \frac{1}{n}[u^{n-1}]\left(\Phi(u)^n\Psi'(u)\right). \tag{7.29}$$

Proof. We only prove (7.29) since (7.28) is a special case. Formally we can proceed as follows, taking $u = X(z)$ as an independent variable,

$$\begin{aligned}
[u^{n-1}]\left(\Phi(u)^n\Psi'(u)\right) &= [u^{n-1}]\left(\Psi'(u)(u/z)^n\right) = [u^{-1}]\left(\Psi'(u)/z^n\right) \\
&= \frac{1}{2\pi i}\oint \frac{\Psi'(u)}{z(u)^n}du \\
&= \frac{1}{2\pi i}\oint \frac{\Psi'(X(z))X'(z)}{z^n}dz \qquad u = X(z) \\
&= [z^n]\left(z\frac{d}{dz}\Psi(X(z))\right) \\
&= n[z^n]\Psi(X(z)).
\end{aligned}$$

Most of the above equalities are trivial to justify (the second one is just the residue theorem) except the third equality. But this one is a consequence of the inverse function theorem discussed above ($[u^0]\Phi(u) \neq 0$ is equivalent of the requirement that the first derivative of the inverse function is nonzero, as we shall see below), and the change of variables. ∎

Before we proceed and discuss more sophisticated examples, let check if (7.26) agrees with the Lagrange formula.

Example 7.14 *Checking (7.26).*
As in (7.25) we define $\Theta(u) = u/\Phi(u)$ (i.e., $X(z) = \Theta^{-1}(z)$), hence for $u = 0$ we have $\Theta'(0) = 1/\Phi(0)$. From Lagrange's formula (7.28) we obtain for $n = 1$

$$[z]X(z) = [u^0]\Phi(u) = 1/\Theta'(0),$$

which coincides with (7.26). To find $[z^2]X(z) = [z^2]\Theta^{-1}(z)$ we need a little bit of algebra. First of all, observe that by (7.28)

$$[z^2]X(z) = \frac{1}{2}[u]\Phi^2(u).$$

But

$$\begin{aligned}\Phi^2(u) &= \frac{u^2}{\Theta(u))^2} = \frac{u^2}{(u\Theta'(0) + \frac{1}{2}u^2\Theta''(0) + O(u^3))^2} \\ &= \frac{1}{[\Theta'(0)]^2} - u\frac{\Theta''(0)}{[\Theta'(0)]^3} + O(u^2).\end{aligned}$$

Thus,

$$[z^2]X(z) = -\frac{1}{2}\frac{\Theta''(0)}{[\Theta'(0)]^3}$$

which again agrees with (7.26). □

The most popular application of the Lagrange formula is to the tree function $T(z)$ defined in Example 12 that we repeat here

$$T(z) = ze^{T(z)}. \tag{7.30}$$

In fact, the tree function is closely related to the Lambert W-function $W(x)\exp(W(x)) = x$ (i.e., $T(z) = -W(-z)$). The interested reader is referred to [58] for a comprehensive survey on the Lambert function and its myriad applications.

Example 7.15 *Tree Function $T(z)$*

We compute here the coefficients $t_n = [z^n]T(z)$ that represents the number of labeled rooted trees. Setting $\Phi(u) = e^u$ and $\Psi(u) = u$ in the Lagrange formula, we obtain

$$\begin{aligned}[z^n]T(z) &= \frac{1}{n}[u^{n-1}]e^{nu} = \frac{1}{n}\frac{n^{n-1}}{(n-1)!} \\ &= \frac{n^{n-1}}{n!}\end{aligned} \tag{7.31}$$

which agrees with Entry 7 in Table 7.2.

We still need to know the radius of convergence of $T(z)$. We will show that $T(z)$ is analytic for all $|z| < e^{-1}$. Indeed, let us define as in (7.25) the function $\Theta(u) = u/\Phi(u) = ue^{-u}$, and observe that $T(z) = \Theta^{-1}(z)$. From our discussion proceeding the Lagrange formula (cf. (7.26)), we conclude that $T(z)$ or the inverse of $\Theta(u)$ exists as long as $\Theta'(u) \neq 0$. Let τ be the smallest real number such that

$$0 = \Theta'(\tau) = e^{-\tau}(1-\tau)$$

or equivalently

$$\Phi(\tau) - \tau\Phi'(\tau) = e^{\tau}(1-\tau).$$

We find $\tau = 1$ and $z = \Theta(\tau) = e^{-1}$. Since $\Theta(u)$ increases from $u = 0$ to $u = \tau$, the tree function $T(z)$ has it first singularity at $z = e^{-1}$. Thus it is analytic for $|z| < e^{-1}$. □

The tree function is so important that it deserves more discussions. One defines the *tree-like* generating function of a_n as

$$A(z) = \sum_{n=0}^{\infty} a_n \frac{n^n}{n!} z^n. \tag{7.32}$$

In particular, for $|z| < e^{-1}$

$$T(z) = \sum_{n=1}^{\infty} \frac{1}{n} \frac{n^n}{n!} z^n, \tag{7.33}$$

$$B(z) = \sum_{n=0}^{\infty} \frac{n^n}{n!} z^n = \frac{1}{1 - T(z)}. \tag{7.34}$$

To see how (7.34) follows from (7.33), observe that after differentiating (7.33) one finds $B(z) = 1 + zT'(z)$. On the other hand, differentiating the functional equation $T(z) = ze^{T(z)}$ we obtain

$$zT'(z) = \frac{T(z)}{1 - T(z)},$$

and the proof of (7.34) follows. In Exercise 3 the reader is asked to re-derive (7.34) from the Lagrange's formula.

Example 7.16 *Sum and Recurrence Arising in Coding Theory*
In information theory (e.g., channel coding, minimax redundancy of universal coding, and universal portfolios) the following sum is of interest

$$S_n = \sum_{i=0}^{n} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(1 - \frac{i}{n}\right)^{n-i}.$$

We can analyze it through the tree-like generating function $s(z) = \sum_{n \geq 0} S_n \frac{n^n}{n!} z^n$. Indeed, observe that $n^n S_n$ is a convolution of two identical sequences $\frac{n^n}{n!}$, hence

$$s(z) = [B(z)]^2$$

where $B(z)$ is defined in (7.34).

A more interesting example is the following. When analyzing the minimax redundancy of source coding for memoryless sources (cf. also Section 8.7.2), Shtarkov [308] came up with the following summation problem

$$D_n(m) = \sum_{i=1}^{m} \binom{m}{i} D_n^*(i), \tag{7.35}$$

where $D_0^*(1) = 0$, $D_n^*(1) = 1$ for $n \geq 1$, and $D_n^*(i)$ for $i > 1$ satisfies the following recurrence:

$$D_n^*(i) = \sum_{k=1}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} D_{n-k}^*(i-1) \tag{7.36}$$

provided that $D_n^*(i) = 0$ for $i > n$. We express $D_n(m)$ and $D_n^*(m)$ as tree-like generating functions. Let us introduce a new sequence $\widehat{D}_n^*(m)$ defined as

$$\widehat{D}_n^*(m) = \sum_{k=0}^{n} \binom{n}{k} \left(\frac{k}{n}\right)^k \left(1 - \frac{k}{n}\right)^{n-k} D_{n-k}^*(m-1) = D_n^*(m) + D_n^*(m-1)$$

and its tree-like generating functions is

$$\widehat{D}_m^*(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k \widehat{D}_k^*(m) \ .$$

Observe that it can be re-written as

$$\frac{n^n}{n!} \widehat{D}_n^*(m) = \sum_{k=0}^{n} \frac{k^k}{k!} \cdot \frac{(n-k)^{n-k}}{(n-k)!} D_{n-k}^*(m-1) \ .$$

We now multiply both sides of the above by z^n, sum it up, and then by the convolution formula of generating functions one arrives at

$$\widehat{D}_m^*(z) = B(z) D_{m-1}^*(z) \ .$$

Thus

$$D_m^*(z) = (B(z) - 1)^{m-1} D_1^*(z) = (B(z) - 1)^m \tag{7.37}$$

since $D_1^*(z) = B(z) - 1$. By (7.37) we obtain

$$D_m(z) = \sum_{i=1}^{m} \binom{m}{i} D_i^*(z) = \sum_{i=1}^{m} \binom{m}{i} (B(z) - 1)^i = B^m(z) - 1. \tag{7.38}$$

WE return to this problem in Section 8.7.2.

Finally, we are in a position to formulate an even more general recurrence. Under some initial conditions, define a sequence x_n^m satisfying for $n \geq 1$ and $m \geq 1$

$$x_n^m = a_n + \sum_{i=0}^{n} \binom{n}{i} \left(\frac{i}{n}\right)^i \left(1 - \frac{i}{n}\right)^{n-i} (x_i^{m-1} + x_{n-i}^{m-1}) \ , \tag{7.39}$$

where a_n is a given sequence (the so called *additive* term), and m is an additional parameter. Using the same arguments as above, one can solve this recurrence in terms of tree-like generating functions, namely:

$$X_m(z) = A(z) + 2B(z) X_{m-1}(z) \tag{7.40}$$

where

$$X_m(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k x_k^m \ , \qquad A(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} z^k a_k \ .$$

The recurrence (7.40) can be solved by telescoping in terms of m. □

7.3.3 Borel Transform

In some applications, it is better to work with ordinary generating functions while in others with exponential generating functions. Therefore, one would like to have an easy transfer formula from one generating function to another. In Example 11, we show how to obtain the ordinary generating function $S(z)$ for the binomial sum from its exponential generating function $s(z) = e^z a(z)$ (cf. (7.17)). This is an example of a transfer formula. A question arises whether one can find a transform that allows to recover the ordinary generating function from the exponential generating function. The answer is given by the **Borel transform** (cf. [337, 353]) that we discuss next.

Let a_n be a sequence such that its ordinary generating function $A(z)$ exists in $|z| < R$. Let also $a(z)$ be its exponential generating function. Define

$$f(z) = \int_0^\infty e^{-t} a(zt) dt$$

for $|z| < R$. We prove that this integral exists and $f(z) = A(z)$ for $|z| < R$. Indeed, observe that by Cauchy's bound (7.23) $|a_n R^n| < M$, hence

$$|a(z)| < M \sum_{n \ge 0} \frac{|z^n|}{R^n n!} = e^{|z|/R}.$$

Therefore, for $|z| < R$

$$\left| \int_0^\infty e^{-t} a(zt) dt \right| < M \int_0^\infty e^{-t(1-|z|/R)} dt,$$

and the integral exists for $|z| < R$. Thus, by Theorem 2.4(ii) we conclude that $f(z)$ is analytic, and the below interchange of the integral and the sum is justifiable for $|z| < R$:

$$\int_0^\infty e^{-t} \sum_{n \ge 0} a_n \frac{z^n t^n}{n!} dt = \sum_{n \ge 0} a_n z^n \int_0^\infty \frac{1}{n!} \int_0^\infty e^{-t} t^n dt = A(z)$$

where the last equation follows from $n! = \int_0^\infty e^{-t} t^n dt = \Gamma(n-1)$. In summary, we prove the following theorem.

Theorem 7.3 (Borel Transform) *Let $A(z) = \sum_{n\geq 0} a_n z^n$ converges for $|z| < R$. Then, for $|z| < R$*

$$A(z) = \int_0^\infty e^{-t} a(zt) dt \tag{7.41}$$

where $a(z) = \sum_{n\geq 0} a_n \frac{z^n}{n!}$ is the EGF of $\{a_n\}_{n=0}^\infty$.

7.4 Probability Generating Functions

Generating functions and other transforms are important and useful tools in probability theory. They are particularly handy in the analysis of algorithms and combinatorics, that is, *discrete* structures. If X is a discrete random variable with $p_k = \Pr\{X = k\}$ $(k = 0, 1, \ldots)$, then the *probability generating functions*, denoted as $G_X(z)$ (or in short: $G(z)$), is defined with respect to the sequence $\{p_k\}_{k\geq 0}$ as

$$G_X(z) = \sum_{k\geq 0} p_k z^k = \mathbf{E}[z^X].$$

We saw already in Chapter 5 moment generating functions in action when large deviations theory was discussed. Here, we first briefly review terminology and elementary facts in order to present later more interesting applications of probability generating functions.

The usefulness of probability generating functions stems from the fact that some quantities of interest for X are easily computable through the generating functions. For example, *moments* of X are related to the derivatives of $G_X(z)$ at $z = 1$. Indeed, if X possesses all moments, then

$$\mathbf{E}[X^r] = \left(z\frac{d}{dz}\right)^r G_X(z)\bigg|_{z=1},$$

as is easy to check. Often, *factorial moments* $\mathbf{E}[(X)_r] := \mathbf{E}[X(X-1)\cdots(X-r+1)]$ are easier to compute since

$$\mathbf{E}[X(X-1)\cdots(X-r+1)] = \frac{d^r}{dz^r} G_X(z)\bigg|_{z=1}.$$

In particular,

$$\begin{aligned}
\mathbf{E}[X] &= G'_X(1), \\
\mathbf{Var}[X] &= G''_X(1) + G'_X(1) - [G'_X(1)]^2.
\end{aligned}$$

Of course, one can recover the probability distribution p_k from the generating function since $p_k = [z^k]G_X(z)$, hence Cauchy and Lagrange formulas work fine. But, also

$$p_k = \frac{d^k}{dz^k} G_X(z)\bigg|_{z=0}.$$

Probability generating functions are commonly used in probability theory since they replace complicated operations like convolutions by algebraic equations. For example, if X and Y are independent, then

$$G_{X+Y}(z) = \mathbf{E}[z^{X+Y}] = \mathbf{E}[z^X]\mathbf{E}[z^Y] = G_X(z)G_Y(z).$$

Also, the distribution function $F(k) = \Pr\{X \leq k\} = \sum_{i=0}^{k} p_i$ can be represented in PGF terms, since by (7.6) we have

$$\sum_{k\geq 0} F(k)z^k = \frac{G_X(z)}{1-z}.$$

In Chapter 5 we used another transform for probability distributions, namely, the *moment generating function* defined as

$$M_X(s) = G_X(e^s) = \mathbf{E}[e^{sX}]$$

where s is assumed to be complex. The moment generating function $M_X(s)$ is well defined at least for $\Re(s) \leq 0$. We saw in Chapter 5.4 that if the definition of $M_X(s)$ can be extended to a vicinity of $s = 0$, say for $\Re(s) \in [-c, c]$ for some $c > 0$, then X has an exponential tail. Finally, it is also convenient to deal with the so called *cumulant generating function* $\kappa_X(s)$ defined as

$$\kappa_X(s) = \log M_X(s) = \log G_X(e^s).$$

Observe that

$$\begin{aligned} \mathbf{E}[X] &= \kappa'(0), \\ \mathbf{Var}[X] &= \kappa''(0). \end{aligned}$$

provided $\kappa(s)$ exists in the vicinity of $s = 0$. Then

$$\kappa(s) = \mu s + \sigma^2 \frac{s^2}{2} + O(s^3), \tag{7.42}$$

where $\mu = \mathbf{E}[X]$ and $\sigma^2 = \mathbf{Var}[X]$.

Example 7.17 *The Average Complexity of Quicksort*
Hoare's *Quicksort* algorithm is the most popular sorting algorithm due to its good performance in practice. The basic algorithm can be briefly described as follows [220, 249]: A partitioning key is selected at random from the unsorted list of keys, and used to partition the keys into two sublists to which the same algorithm is called recursively until the sublists have size one or zero. Let now X_n denote the number of comparisons needed to sort a random list of length n. It is known that after selecting randomly a key, the two sublists

are still "random," hence assuming any key is chosen with equal probability $1/n$ we obtain a stochastic equation

$$X_n = n - 1 + X_k + X_{n-1-k}.$$

It can be translated into the generating function $G_n(z) = \mathbf{E}[z^{X_n}]$ as

$$G_n(z) = z^{n-1} \sum_{k=0}^{n-1} \frac{1}{n} \mathbf{E}[z^{X_k}] \mathbf{E}[z^{X_{n-1-k}}] = \frac{z^{n-1}}{n} \sum_{k=0}^{n-1} G_k(z) G_{n-1-k}(z).$$

Then $c_n := \mathbf{E}[X_n] = G_n'(1)$ satisfies

$$c_n = n - 1 + \frac{2}{n} \sum_{i=0}^{n-1} c_i$$

with $c_0 = 0$. We solve this recurrence using the generating function $C(z) = \sum_{n \geq 0} c_n z^n$. By (7.6) and Item 1 in Table 7.2, we derive

$$C(z) = \frac{z^2}{(1-z)^2} + 2 \int_0^z \frac{C(x)}{1-x} dx,$$

which becomes, after differentiation,

$$C'(z) = \frac{2z}{(1-z)^2} + \frac{2z^2}{(1-z)^3} + \frac{2C(z)}{1-z}.$$

This is a linear differential equation that is simple to solve. Using, for example the method of variation of parameters, we arrive at

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

To find $C_n = [z^n]C(z)$ we proceed as in Example 1, and finally prove

$$C_n = 2(n+1)H_{n+1} - 4n - 2 = 2(n+1)H_n - 4n,$$

where H_n is the nth Harmonic number. □

7.5 Dirichlet Series

Finally, we study one more generating function of $\{a_n\}_1^\infty$ defined as

$$A(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s},$$

and known as *Dirichlet series*. When dealing with ordinary and exponential generating functions, we observed that the choice a particular generating function largely depends on the convolution formula. This is also the case here. We shall demonstrate that Dirichlet series is a very useful tool in number-theoretical problems.

7.5.1 Basic Properties

Let us start with analytic theory of Dirichlet series. For what complex s does the series

$$A(s) = \sum_{n=1}^{\infty} \frac{a_n}{n^s}$$

converge? Observe that

$$\left|\frac{a_n}{n^s}\right| = \frac{|a_n|}{n^{\Re(s)}}.$$

Therefore, if the series converges absolutely for $s = r + it$, then it must also converge for $\Re(s) \geq r$. Let r_a be a lower bound of such r. The series converges for $\Re(s) > r_a$ and diverges for $\Re(s) < r_a$. The real number r_a is called the *abscissa* of absolute convergence. This observation leads to the following known result (for a formal proof see [15, 158]).

Theorem 7.4 *Suppose the series $\sum_{n=1}^{\infty} |a_n/n^s|$ does not converge or diverge for all s. Then there exists a real number r_a called the abscissa of absolute convergence, such that the Dirichlet series $\sum_{n=1}^{\infty} \frac{a_n}{n^s}$ converges absolutely for $\Re(s) > r_a$ and does not converge absolutely for $\Re(s) < r_a$.*

As for an example, let us look at the Riemann zeta function (already discussed in Section 2.4.2) which is the Dirichlet series of $\{1\}_1^{\infty}$, that is,

$$\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}.$$

It converges absolutely for $\Re(s) > 1$, and for $s = 1$ it diverges, so it diverges for $\Re(s) < 1$. In fact, Dirichlet series converges absolutely for all $\Re(s) > 1$ as long as $|a_n|$ is uniformly bounded for all $n \geq 1$.

Let us now turn our attention to some important properties of Dirichlet series. How does the product of two Dirichlet series relate to the Dirichlet convolution? Observe that

$$A(s)B(s) = \sum_{n=1}^{\infty} a_n n^{-s} \sum_{m=1}^{\infty} b_m m^{-s} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} a_n b_m (nm)^{-s} = \sum_{k=1}^{\infty} k^{-s} \sum_{mn=k} a_n b_m.$$

Therefore, if $d|n$ reads "d divides n," then we just proved the following relationship

$$A(s)B(s) \quad \longleftrightarrow \quad \left\{ \sum_{d|n} a_d b_{\frac{n}{d}} \right\}_{n=1}^{\infty}. \tag{7.43}$$

In other words, Dirichlet convolution is a sum of products of the series coefficients whose product of the subscripts is fixed.

Example 7.18 *Some Dirichlet series*

Here are two applications of the Dirichlet convolution formula.

1. DIVISOR FUNCTION. Let $d(n)$ be the number of divisors of the integer n. Clearly

$$d(n) = \sum_{d|n} 1 \cdot 1,$$

hence

$$D(s) = \sum_{n=1}^{\infty} \frac{d(n)}{n^s} = \zeta^2(s).$$

2. FUNCTION $v_2(n)$. Let $v_2(n)$ be the exponent of 2 in the prime decomposition of n. Since

$$v_2(2n) = v_2(n) + 1,$$

we find, after easy algebra,

$$V(s) = \frac{\zeta(s)}{2^s - 1}$$

which is valid for $\Re(s) > 1$. □

7.5.2 Euler Products

For the purpose of this section, it is convenient to write $\{f(n)\}_1^\infty := \{f_n\}_1^\infty$. A function $f(n)$ whose domain is the set of positive integers is called *multiplicative* if

$$f(mn) = f(m)f(n)$$

for all pairs of *relatively prime* positive integers m and n (i.e., $\gcd(m, n) = 1$). Observe that since every integer n has a unique prime factorization

$$n = p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r} \tag{7.44}$$

where $p_1, \ldots, p_r$ are distinct primes and $e_1, \ldots, e_r$ are nonnegative integers, then for a multiplicative function f

$$f(n) = f(p_1^{e_1}) \cdots f(p_r^{e_r}).$$

Let us now consider the product

$$P(s) = \prod_p \left(1 + f(p)p^{-s} + f(p^2)p^{-2s} + \cdots\right)$$

where f is multiplicative and the product is over all prime numbers p. Let us for a moment look at $P(s)$ as a formal power series and multiply it out. What is the coefficient at n^{-s}? Assuming the factorization (7.44), we have

$$(p_1^{e_1} p_2^{e_2} \cdots p_r^{e_r})^{-s} f(p_1^{e_1}) f(p_2^{e_2}) \cdots f(p_r^{e_r}) = n^{-s} f(n).$$

Therefore, formally

$$\prod_{p>1}\left(1+f(p)p^{-s}+f(p^2)p^{-2s}+\cdots\right)=\sum_{n=1}^{\infty}\frac{f(n)}{n^s}.$$

For $s=0$ the above becomes the so called Euler product. In fact, it is not difficult to embrace analytic theory into our considerations. This leads to the following very important result (cf. [15] for a formal proof).

Theorem 7.5 (Euler 1737) *Let f be a multiplicative function such that the series $\sum_{n=1}^{\infty} f(n)n^{-s}$ is absolutely convergent. Then*

$$\sum_{n=1}^{\infty}\frac{f(n)}{n^s}=\prod_{p}\left(1+f(p)p^{-s}+f(p^2)p^{-2s}+\cdots\right) \tag{7.45}$$

and the infinite product is convergent, too.

Euler's theorem is a source of many interesting identities for the Riemann zeta function. Below, we consider some of them.

Example 7.19 *Another Representation for $\zeta(s)$*
Let us first assume $f(n)=1$ for all n. Then

$$\begin{aligned}\zeta(s) &= \prod_{p}\left(1+p^{-s}+p^{-2s}+\cdots\right)\\ &= \prod_{p}\frac{1}{1-p^{-s}}.\end{aligned} \tag{7.46}$$

Let us now recall the definition of the Möbius function $\mu(n)$ which we already discussed in Section 3.2. For prime p we define $\mu(p^e)$ as

$$\mu(p^e)=\begin{cases}+1 & \text{if} \quad e=0,\\ -1 & \text{if} \quad e=1,\\ 0 & \text{if} \quad e\geq 2.\end{cases}$$

Then, by Euler's theorem

$$\sum_{n=1}^{\infty}\frac{\mu(n)}{n^s}=\prod_{p}(1-p^{-s})=\frac{1}{\zeta(s)} \tag{7.47}$$

for $\Re(s)>1$. □

The last example shows how to invert relationships like

$$a_n = \sum_{d|n} b_d,$$

that is, to find b_n as a function of a_d. Let us consider again Example 4 of Chapter 3 where we counted the number of primitive sequences. If $f(n)$ is this number, then we already proved that

$$2^n = \sum_{d|n} f(d).$$

How can we recover $f(n)$? Let $A(s) = \sum_{n\geq 1} 2^n n^{-s}$. Then,

$$A(s) = F(s)\zeta(s)$$

where $F(s) = \sum_{n\geq 1} f(n)n^{-s}$. Hence,

$$F(s) = \frac{A(s)}{\zeta(s)} = A(s)\sum_{n\geq 1}\frac{\mu(n)}{n^s}.$$

This translates into

$$f(n) = \sum_{d|n} \mu\left(\frac{n}{d}\right) 2^d,$$

as already proved in Example 4. □

7.5.3 Perron–Mellin Formula

In this section, we present a surprising connection between *discrete* sums and complex integrals. We concentrate here on *exact* formulas for such sums. This opens a venue for a much deeper analysis presented in the next chapters (cf. Section 8.2.1, Section 8.5 and Section 10.1) where complex integrals are used to derive asymptotic expansions of discrete sums. Needless to say, evaluation of discrete quantities by complex integrals is at the core of an analytic approach to analysis of algorithms and analytic combinatorics. We shall return to this issue several times in the sequel.

Let $\{a_k\}_{k=1}^n$ be a sequence of reals. We are interested in the evaluation of the following discrete sum

$$S_n = \sum_{k=1}^n a_k.$$

From elementary calculus we know that such sums are well approximated by $\int_1^n a(x)dx$, where $a(x)|_{x=n} = a_n$ and $a(x)$ is a monotone function (cf. also Section 8.2.1). Here, we aim at *exact*

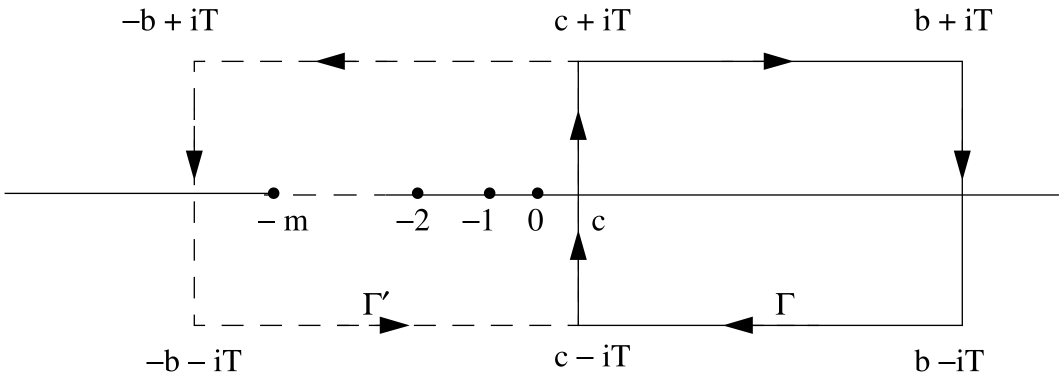

Figure 7.3: Illustration to the "closing-the-box" method.

evaluation of such sums through the Cauchy residue Theorem 2.5 (cf. Chapter 2.3). We recall that Cauchy's residue theorem states that

$$\frac{1}{2\pi i}\oint_{\mathcal{C}} f(z)dz = \sum_{j=1}^{N} \mathrm{Res}[f(z);\ z = a_j]\ ,$$

as long as $f(z)$ is analytic inside $\mathcal{C}$ except at N poles $a_1, \ldots, a_N$. Thus, if one selects a "good" function $f(z)$, then the right-hand side of the above can represent a discrete sum. In many applications, it is often convenient to take an infinite loop $\mathcal{C}$ as shown in Figure 7.3 where $T, b \to \infty$. In such a case, we write

$$\int_{c-i\infty}^{c+i\infty} f(s)ds := \lim_{T\to\infty}\int_{c-iT}^{c+iT} f(s)ds$$

where c is such that it assures the existence of the integral.

We illustrate the above idea on the so called Perron-Mellin formula. We shall follow the presentation of Apostol [15] and Flajolet *et al.* [119]. Let $\{a_n\}_1^\infty$ be a sequence whose Dirichlet series is $A(s)$. We prove that for $m \geq 1$

$$\frac{1}{m!}\sum_{k=1}^{n-1} a_k\left(1-\frac{k}{n}\right)^m = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} A(s)n^s \frac{ds}{s(s+1)\cdots(s+m)} \tag{7.48}$$

where $c > 0$ lies in the half-plane of absolute convergence of $A(s)$. When deriving the Perron-Mellin formula, we introduce the reader to a new technique, sometimes called the *closing-the-box* method, that allows to evaluate the integrals like the one above.

Consider the integral of the right-hand side of (7.48). We have

$$\frac{m!}{2\pi i}\int_{c-i\infty}^{c+i\infty} A(s)n^s \frac{ds}{s(s+1)\cdots(s+m)} \quad = \quad \frac{m!}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{n^s ds}{s(s+1)\cdots(s+m)}\sum_{k=1}^{\infty}\frac{a_k}{k^s}$$

$$= \sum_{k=1}^{\infty} a_k \frac{m!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{(n/k)^s}{s(s+1)\cdots(s+m)} ds,$$

where the interchange of the sum and the integral is justified since both converge absolutely. We need now to evaluate the integral

$$J_m(x) = \frac{m!}{2\pi i} \int_{c-i\infty}^{c+i\infty} \frac{x^s}{s(s+1)\cdots(s+m)} ds$$

for $x > 0$. We shall prove below that for all $m \geq 1$

$$J_m(x) = \begin{cases} 0 & \text{for } \ 0 < x < 1 \\ (1 - x^{-1})^m & \text{for } \ x > 1. \end{cases} \tag{7.49}$$

Provided the above is true, the proof of (7.48) follows (with the exception of $x = k/n = 1$ and $m = 0$ which we treat below).

In order to prove (7.49), we apply the "closing-the-box method." We consider two cases.

A. Case $0 < x < 1$

To evaluate the "infinite" integral in the definition of $J_m(x)$, we first consider a finite one, that is, $\int_{c-iT}^{c+iT}$ and later allow T to grow. The later integral is calculated by considering the rectangular contour Γ shown in Figure 7.3 (solid line). First, observe that there is no singularity inside the contour Γ since the integrand function appearing in the definition of $J_m(x)$ is analytic. Hence

$$\int_\Gamma \frac{x^s}{s(s+1)\cdots(s+m)} = 0,$$

and therefore

$$\int_{c-iT}^{c+iT} = \int_{b+iT}^{c+iT} + \int_{b-iT}^{b+iT} + \int_{c-iT}^{b-iT} .$$

We prove now that the right-hand side of the above tends to zero as $T, b \to \infty$ proving (7.49) for $x < 1$. Let $s = r + iy$. We have the following estimates (we use here the bound (2.21) from Section 2.3):

$$\begin{aligned} \left| \int_{c-iT}^{c+iT} \frac{x^s}{s(s+1)\cdots(s+m)} ds \right| &\leq 2\int_c^b \frac{x^{\Re(s)}}{T(T+1)\cdots(T+m)} ds + 2T\frac{x^b}{b^{m+1}} \\ &\leq \frac{2}{T^{m+1}} \int_c^\infty x^r \, dr + 2T\frac{x^b}{b^{m+1}} \\ &= \frac{2}{T^{m+1}} \left(\frac{-x^x}{\ln x} \right) + 2T\frac{x^b}{m^{m+1}} \\ &\to 0 \end{aligned}$$

since x^b decays exponentially to 0 as $b \to \infty$ for $0 < x < 1$.

B: CASE $x > 1$

In this case, we close the loop (rectangle) to the right of the line $(c - i\infty, c + i\infty)$ creating a new contour Γ' as shown in Figure 7.3 (dashed line). But, inside the loop Γ' there are $m+1$ poles at $a_i = -i$ for $i = 0, 1, \ldots, m$. By Cauchy

$$\frac{1}{2\pi i}\int_{\Gamma'} \frac{x^s}{s(s+1)\cdots(s+m)} ds = \sum_{i=0}^{m} \mathrm{Res}[f(s), s = -i],$$

where $f(s) = x^s/(s(s+1)\cdots(s+m))$. The above residues are easy to calculate. For example, using (2.23), we obtain

$$\mathrm{Res}[f(s), s = -i] = (-1)^i \frac{x^{-i}}{i!(m-i)!},$$

thus

$$\frac{m!}{2\pi i}\int_{\Gamma'} \frac{x^s}{s(s+1)\cdots(s+m)} ds = (1 - x^{-1})^m.$$

On the other hand, we can evaluate the integral over the contour Γ' as

$$\begin{aligned}\int_{\Gamma'} f(s)ds &= \int_{c-iT}^{c+iT} f(s)ds + \int_{-b-iT}^{c-iT} f(s)ds \\ &+ \int_{-b+iT}^{-b-iT} f(s)ds + \int_{c+iT}^{-b+iT} f(s)ds.\end{aligned}$$

We show now that the last three integrals vanish as $T, b \to \infty$ proving (7.49). Using the same argument as above, we have

$$\begin{aligned}\left|\int_{-b-iT}^{c-iT} \frac{x^s}{s(s+1)\cdots(s+m)} ds\right| &\le \frac{1}{T^{m+1}}\int_{-\infty}^{c} x^r dr = \frac{x^c}{T^{m+1}\ln x}, \\ \left|\int_{-b+iT}^{-b-iT} \frac{x^s}{s(s+1)\cdots(s+m)} ds\right| &\le 2\frac{x^{-b}}{T^m}, \\ \left|\int_{-b+iT}^{c+iT} \frac{x^s}{s(s+1)\cdots(s+m)} ds\right| &\le \frac{1}{T^{m+1}}\int_{-b}^{c} x^r dr \le \frac{1}{T^{m+1}}\frac{x^c}{\ln x}.\end{aligned}$$

Since $x > 1$, we see that all the three integrals tend to zero as $T, b \to \infty$.

Finally, the case $x = 1$ and $m = 0$ must be considered separately. Direct calculations, that are left as an exercise for the reader, show that

$$\frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} \frac{ds}{s} = \frac{1}{2}$$

for $c > 0$.

In summary, we have just proved the Perron-Mellin formula.

Theorem 7.6 (Perron–Mellin Formula) *Let $A(s)$ be the Dirichlet' series of $\{a_n\}_1^\infty$. Let $c > 0$ lies inside the half-plane of absolute convergence of $A(s)$. Then*

$$\sum_{k=1}^{n-1} a_k + \frac{1}{2}a_n = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} A(s)n^s \frac{ds}{s}, \tag{7.50}$$

and for all $m \geq 1$

$$\frac{1}{m!}\sum_{k=1}^{n-1} a_k \left(1 - \frac{k}{n}\right)^m = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} A(s)n^s \frac{ds}{s(s+1)\cdots(s+m)}. \tag{7.51}$$

In Section 7.6.3 we shall use the Perron-Mellin formula to evaluate the total number of 1-digits in the binary representation of $1, 2, \ldots, n$. Here, we illustrate Theorem 7.6 on a simpler, but important example.

Example 7.20 *An Identity for $\zeta(s)$ Function*

Let us apply the Perron-Mellin formula for $m = 1$ and $a_k \equiv 1$. Then

$$\frac{n-1}{2} = \frac{1}{2\pi i}\int_{2-i\infty}^{2+i\infty} \zeta(s)n^s \frac{ds}{s(s+1)}. \tag{7.52}$$

Now, we evaluate the integral on the right-hand side by the closing-the-box approach. We consider a rectangle box with the line of integration shifted to the left up to $\Re(s) = -\frac{1}{4}$. Let us call the lines in the box as East (equal to the original line of integration), West (i.e., $\Re(s) = -\frac{1}{4}$), North (i.e., $\Im(s) = T$ as $T \to \infty$) and South (i.e., $\Im(s) = -T$). Call this contour (traversed counterclockwise) $\mathcal{C}$. Inside $\mathcal{C}$ there are two poles at $s = 0$ and $s = 1$ with residues $\mathrm{Res}[f(s); s = 0] = -1/2$ (since $\zeta(0) = -1/2$) and $\mathrm{Res}[f(s); s = 1] = n/2$, where $f(s) = \zeta(s)n^s/(s(s+1))$. Thus

$$\frac{1}{2\pi i}\int_{\mathcal{C}} \zeta(s)n^s \frac{ds}{s(s+1)} = \frac{n-1}{2},$$

that is, the same value as the integral (7.52). It easy to see (by a similar argument as above) that the integral over the South/North lines tend to zero as $T \to \infty$. More importantly, the integral over the West side converges, too, due to our choice $\Re(s) = -\frac{1}{4}$ and the estimate $\zeta(s) = O(y^{\frac{1}{2}-r})$ (cf. (2.45) in Section 2.4.2)), where $s = r + iy$. Indeed

$$\left|\int_{-\frac{1}{4}-iT}^{-\frac{1}{4}+iT} \zeta(s)n^s \frac{ds}{s(s+1)}\right| = O(T^{-1/4}).$$

In view of this, we prove

$$\frac{1}{2\pi i}\int_{-\frac{1}{4}-i\infty}^{-\frac{1}{4}+i\infty} \zeta(s)n^s \frac{ds}{s(s+1)} = 0 \tag{7.53}$$

which is a source of several *exact* instead of *asymptotic* formulas (e.g., see Section 7.6.3). □

7.6 Applications

We discuss three different applications of generating functions. In the first one, we consider a class of recurrences arising in the analysis of digital trees, and show how to solve them exactly. Then we study an old problem of counting the number of pattern occurrences in a random string. We use the combinatorial structure approach (i.e., language approach) to derive the generating function for the frequency of pattern occurrence when the text string is generated by a Markovian source. Finally, we use the Perron–Mellin formula to derive Delange's formula for the total number of 1 digits in the binary representation of $1, 2, \ldots, n$.

7.6.1 Recurrences Arising in the Analysis of Digital Trees

We consider certain recurrences that arise in problems on words; in particular, in digital trees discussed in Section 1.1. Let x_n be a generic notation for a quantity of interest (e.g., depth, size or path length in a digital tree built over n strings). Given x_0 and x_1, the following three recurrences originate from problems on tries, PATRICIA tries and digital search trees, respectively,

$$x_n = a_n + \beta \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (x_k + x_{n-k}) , \qquad n \geq 2, \tag{7.54}$$

$$x_n = a_n + \beta \sum_{k=1}^{n-1} \binom{n}{k} p^k q^{n-k} (x_k + x_{n-k}) - \alpha (p^n + q^n) x_n , \qquad n \geq 2, \tag{7.55}$$

$$x_{n+1} = a_n + \beta \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (x_k + x_{n-k}) \qquad n \geq 0, \tag{7.56}$$

where a_n is a known sequence (also called the additive term), α and β are some constants, and finally $p + q = 1$.

To solve these recurrences we apply exponential generating functions. We start with (7.54). Multiplying it by z^n, summing up, using the convolution formula for exponential generating function (7.16) from Table 7.3, and taking into account the initial conditions, we arrive at

$$x(z) = a(z) + \beta x(zp) e^{zq} + \beta x(zq) e^{zp} + d(z) \tag{7.57}$$

where $d(z) = d_0 + d_1 z$ and d_0 and d_1 depend on the initial condition for $n = 0, 1$. This functional equation is still hard to solve, hence we introduce the so called Poisson transform

$$\widetilde{X}(z) = x(z) e^{-z} = e^{-z} \sum_{n=0}^{\infty} x_n \frac{z^n}{n!}.$$

The Poisson transform is a very useful tool in analysis of algorithms. We devote to it the whole Chapter 10. Observe now that (7.57) reduces to following functional equation

$$\widetilde{X}(z) = \widetilde{A}(z) + \beta \widetilde{X}(zp) + \beta \widetilde{X}(z) + d(z) e^{-z} . \tag{7.58}$$

where $\widetilde{A}(z) = a(z)e^{-z} = e^{-z}\sum_{n=0}^{\infty} a_n \frac{z^n}{n!}$. (Throughout this book we write $\widetilde{A}(z)$ for the Poisson transform of the sequence $\{a_n\}_{n=0}^{\infty}$). Since $\widetilde{x}_n = n![z^n]\widetilde{X}(z)$ the quantities $\widetilde{x}_n$ and x_n are related by

$$x_n = \sum_{k=0}^{n} \binom{n}{k} \widetilde{x}_k.$$

After comparing coefficients of $\widetilde{X}(z)$ at z^n, we finally obtain

$$x_n = x_0 + n(x_1 - x_0) + \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{\hat{a}_k + kd_1 - d_0}{1 - \beta(p^k + q^k)} , \tag{7.59}$$

where $n![z^n]\widetilde{A}(z) = \widetilde{a}_n := (-1)^n \hat{a}_n$. In passing, we point out that $\hat{a}_n$ and a_n form the so called *binomial inverse relations*, and

$$\hat{a}_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k a_k , \qquad a_n = \sum_{k=0}^{n} \binom{n}{k} (-1)^k \hat{a}_k , \tag{7.60}$$

that is, $\hat{\hat{a}}_n = a_n$.

Example 7.21 *The Average External Path Length in a Trie*
To illustrate our discussion so far, let us consider a trie with a memoryless source and estimate the average ℓ_n of the external path length, i.e., $\ell_n = \mathbf{E}[L_n]$ (cf. Section 1.1). Clearly, $\ell_0 = \ell_1 = 0$ and for $n \geq 2$

$$\ell_n = n + \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (\ell_k + \ell_{n-k}) .$$

Thus by (7.59)

$$\ell_n = \sum_{k=2}^{n} (-1)^k \binom{n}{k} \frac{k}{1 - p^k - q^k}$$

for $n \geq 2$. □

Let us now consider recurrence (7.55) which is much more intricate. It has an exact solution only for some special cases that we discuss below. We only consider a simplified version of (7.55), namely

$$x_n(2^n - 2) = 2^n a_n + \sum_{k=1}^{n-1} \binom{n}{k} x_k$$

with $x_0 = x_1 = 0$ (for a more general recurrence of this type see [317]). The EGF $x(z)$ of x_n is

$$x(z) = (e^{z/2} + 1)x(z/2) + a(z) - a_0 \tag{7.61}$$

where $a(z)$ is the exponential generating function of a_n. To solve this recurrence we observe that after multiplying both sides by $z/(e^z-1)$ and defining

$$\check{X}(z) = x(z)\frac{z}{e^z-1}$$

we obtain a new simpler functional equation, namely:

$$\check{X}(z) = \check{X}(z/2) + \check{A}(z) \tag{7.62}$$

where in the above we assume for simplicity $a_0 = 0$. This is due to the identity $e^z - 1 = (e^{z/2}-1)(e^{z/2}+1)$ that luckily translates (7.61) into (7.62). But (7.62) is of the same type as the functional equation (7.58), and this allows to extract the coefficient $\check{x}_n = n![z^n]\check{X}(z)$. One must, however, translate the coefficients $\check{x}_n$ into the original sequence x_n. We use the Bernoulli polynomials $B_n(x)$ and the Bernoulli numbers $B_n = B_n(0)$ (cf. Table 7.2 for detailed discussion) defined as

$$\frac{ze^{tz}}{e^z-1} = \sum_{k=0}^{\infty} B_k(t)\frac{z^k}{k!} .$$

Furthermore, we introduce the *Bernoulli inverse relations* for a sequence a_n as

$$\check{a}_n = \sum_{k=0}^{n}\binom{n}{k}B_k a_{n-k} \qquad \longleftrightarrow \qquad a_n = \sum_{k=0}^{n}\binom{n}{k}\frac{\check{a}_k}{k+1} . \tag{7.63}$$

In Exercise 9 we asked the reader to prove

$$a_n = \binom{n}{r}q^n \qquad \longleftrightarrow \qquad \check{a}_n = \binom{n}{r}q^r B_{n-r}(q) \tag{7.64}$$

for $0 < q < 1$. Using now (7.62), and comparing coefficients we show that for $a_n = \binom{n}{r}q^n$ the above recurrence has a particularly simply solution, namely:

$$x_n = \sum_{k=1}^{n}(-1)^k\binom{n}{k}\frac{B_{k+1}(1-q)}{k+1}\frac{1}{2^{k+1}-1} .$$

A general solution to the above recurrence can be found in [317] (cf. also [135]), and it involves $\check{a}_n$.

Example 7.22 *The Average Unsuccessful Search*

As an example, consider the number of trials u_n in an unsuccessful search of a string in a PATRICIA trie in which string are generated by an unbiased memoryless source (i.e., $p = q = 1/2$). Knuth [220] proved that

$$u_n(2^n-2) = 2^n(1-2^{1-n}) + \sum_{k=1}^{n-1}\binom{n}{k}u_k$$

and $u_0 = u_1 = 0$. A simple application of the above leads, after some algebra, to

$$u_n = 2 - \frac{4}{n+1} + 2\delta_{n0} + \frac{2}{n+1}\sum_{k=2}^{n}\binom{n+1}{k}\frac{B_k}{2^{k-1}-1}$$

where $\delta_{n,k}$ is the Kronecker delta, that is, $\delta_{n,k} = 1$ for $n = k$ and zero otherwise. □

In summary, so far we were able to solve exactly functional equations (7.54) and (7.55) since they can be reduced to a simple functional equation of the form (7.58). In particular, equation (7.55) became (7.58) since luckily $e^z - 1 = (e^{z/2} - 1)(e^{z/2} + 1)$, as already pointed out by Knuth [220], but one cannot expect that much luck with other functional equations. Nevertheless, there is a large class of recurrences that can be reduced to the following general functional equation

$$F(z) = a(z) + b(z)F(\sigma(z)) \tag{7.65}$$

where $a(z), b(z), \sigma(z)$ are known function. Formally, iterating this equation we obtain its solution as

$$F(z) = \sum_{k=0}^{\infty} a(\sigma^{(k)}(z)) \prod_{j=0}^{k-1} b(\sigma^{(j)}(z)) \tag{7.66}$$

where $\sigma^{(k)}(z)$ is the kth iterate of $\sigma(\cdot)$. When applying the above to solve real problems, one must assure the existence of the infinite series involved. In some cases we can provide asymptotic solutions to such complicated formulas by appealing to the Mellin transform that we discuss in Chapter 9.

Example 7.23 *A simple Functional Equation*
Consider the following functional equation

$$f(z) = \beta a(z/2) f(z/2) + b(z),$$

where $|\beta| \le 1$. By (7.66) we obtain

$$f(z) = \sum_{n=0}^{\infty} \beta^n b(z2^{-n}) \prod_{k=1}^{n} a(z2^{-k})$$

provided the series converges. Define

$$\varphi(z) = \prod_{j=0}^{\infty} a(z2^j)\ ,$$

provided the infinite product converges. Then, for $\varphi(z,u) \neq 0$

$$f(z)\varphi(z) = \sum_{n=0}^{\infty} \beta^n b(z2^{-n}) \varphi(z2^{-n}). \tag{7.67}$$

This form is quite useful in some computations, as we shall see in the next chapters. □

Finally, we deal with recurrence (7.56). The EGF $x(z)$ of x_n satisfies

$$x'(z) = a(z) + x(zp)e^{zq} + x(zq)e^{zp} ,$$

which becomes after the substitution $\widetilde{X}(z) = x(z)e^{-z}$

$$\widetilde{X}'(z) + \widetilde{X}(z) = \widetilde{A}(z) + \widetilde{X}(zp) + \widetilde{X}(zq) . \tag{7.68}$$

The above is a *differential*-functional equation that we did not discuss so far. It can be solved since a direct translation of coefficients gives: $\widetilde{x}_{n+1} + \widetilde{x}_n = \widetilde{a}_n + \widetilde{x}_n(p^n + q^n)$. Fortunately, this is a simple linear recurrence that has an explicit solution. Since

$$x_n = \sum_{k=0}^{n} \binom{n}{k} \widetilde{x}_k,$$

we finally obtain

$$x_n = x_0 - \sum_{k=1}^{n} (-1)^k \binom{n}{k} \sum_{i=1}^{k-1} \hat{a}_i \prod_{j=i+1}^{k-1} (1 - p^j - q^j) = x_0 - \sum_{k=1}^{n} (-1)^k \binom{n}{k} \sum_{i=1}^{k-1} \hat{a}_i \frac{Q_k}{Q_i} , \tag{7.69}$$

where $Q_k = \prod_{j=2}^{k} (1 - p^j - q^j)$, and $\hat{a}_n$ is the binomial inverse of a_n as defined in (7.60).

Example 7.24 The Average External Path Length in a Digital Search Tree

To illustrate this, let ℓ_n be the expected path length in a digital search tree. Then for all $n \geq 0$

$$\ell_{n+1} = n + \sum_{k=1}^{n} \binom{n}{k} p^k q^{n-k} (\ell_k + \ell_{n-k})$$

with $\ell_0 = 0$. By (7.56) it has the following solution

$$\ell_n = \sum_{k=2}^{n} (-1)^k \binom{n}{k} Q_{k-1}$$

where Q_k is defined above. □

In passing, we should observe that solutions of the recurrences (7.54)-(7.56) have a form of an alternating sum, that is,

$$x_n = \sum_{k=1}^{n} (-1)^k \binom{n}{k} f_k$$

where f_n has an explicit formula. In Section 8.5 we shall discuss in depth such sums and provide a tool to find asymptotic expansions for these sums.

We were quite lucky so far since we dealt mostly with linear recurrences of *first* order. However, this is not any longer true when we consider b-digital search trees (discussed in Chapter 1.1) in which one assumes that a node of such a tree can store up to b strings. Then, the general recurrence (7.56) becomes

$$x_{n+b} = a_n + \beta \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} (x_k + x_{n-k}) \qquad n \geq 0 \tag{7.70}$$

provided $x_0, \ldots, x_{b-1}$ are given. Our previous approach would lead to a linear recurrence of order b that does not possess a nice explicit solution. The "culprit" lies in the fact that the exponential generating function of $\{x_{n+b}\}_{n=0}^{\infty}$ is the b-th derivative of the exponential generating function of $\{x_n\}_{n=0}^{\infty}$. We know, however, that the *ordinary* generating function of $\{x_{n+b}\}_{n=0}^{\infty}$ translates into $z^{-b}(X(z)-x_0-\cdots-x_{b-1}z^{b-1})$, where $X(z)$ is OGF of $\{x_n\}_0^{\infty}$. This simple observation led Flajolet and Richmond [113] to reconsider the standard approach to the above binomial recurrences, and to introduce ordinary generating function into the play. A careful reader observes, however, that then one must translate into ordinary generating functions sequences such as $s_n = \sum_{k=0}^{n} \binom{n}{k} a_k$ (which were easy under exponential generating functions since they become $a(z)e^z$). In Example 11 we prove that

$$s_n = \sum_{k=0}^{n} \binom{n}{k} a_k \qquad \longleftrightarrow \qquad S(z) = \frac{1}{1-z} A\left(\frac{z}{1-z}\right) .$$

We shall return to this problem in Section 9.4.1 and Section 10.5.2.

Example 7.25 *A Simple Differential-Functional Equation*
For example, recurrence (7.70) for $p = q = 1/2$ and any $b \geq 1$ can be translated into ordinary generating functions as

$$\begin{aligned} X(z) &= \frac{1}{1-z} G\left(\frac{z}{1-z}\right) \\ G(z)(1+z)^b &= 2z^b G(z/2) + P(z) \end{aligned}$$

where $P(z)$ is a function of a_n and initial conditions. The latter functional equation falls under (7.65), thus can be solved by the method presented above. However, for $p \neq 1/2$ an explicit solution is harder to find, and we must restore to an asymptotic solution. We shall discuss it again in Section 10.5.2. □

7.6.2 Pattern Occurrences in a Random Text

Let us consider two strings, a pattern string $H = h_1 h_2 \ldots h_m$ and a text string $T = t_1 t_2 \ldots t_n$ of respective lengths equal to m and n over an alphabet $\mathcal{S}$ of size V. (In this section, we drop our standard notation $\mathcal{A}$ for the alphabet since it is traditionally reserved in the pattern matching problems for the autocorrelation language defined below.) We shall write $\mathcal{S} = \{1, 2, \ldots, V\}$ to simplify the presentation. We assume that the pattern string is *fixed* and given, while the text string is random and generated by a stationary Markovian source. We also present results for memoryless sources. For the stationary Markov model, as always, we denote the transition matrix as $\mathsf{P} = \{p_{i,j}\}_{i,j\in\mathcal{S}}$, the stationary distribution as $\boldsymbol{\pi}$, and we write Π for the stationary matrix that consists of V identical rows equal to $\boldsymbol{\pi}$.

A word about notation: To extract a particular element, say with index (i,j), from a matrix, say P, we shall write $[\mathsf{P}]_{i,j} = p_{i,j}$. Finally, we recall that $(\mathsf{I} - \mathsf{P})^{-1} = \sum_{k\geq 0} \mathsf{P}^k$ provided the inverse matrix exists (i.e., $\det(\mathsf{I} - \mathsf{P}) \neq 0$ or $||\mathsf{P}|| < 1$ for any matrix norm $||\cdot||$). Below, we write $P(H_i^j) = \Pr\{T_{i+k}^{j+k} = H_i^j\}$ for the probability of the substring $H_i^j = h_i \ldots h_j$ occurring in the random text T_{i+k}^{j+k} between symbols $i+k$ and $j+k$ for any k (in particular, $P(H)$ denotes the probability of H appearing in the text).

Our goal is to estimate the frequency of pattern occurrences in a text generated by a Markov source. We allow pattern to overlap when counting occurrences (e.g., if $H = abab$, then it occurs twice in $T = abababb$ when overlapping is allowed; it occurs only once if overlapping is not allowed). We write O_n or $O_n(H)$ for the number of pattern H occurrences in the text of size n. We derive formulas on two generating functions, namely,

$$\begin{aligned} T^{(r)}(z) &= \sum_{n\geq 0} \Pr\{O_n(H) = r\} z^n \\ T(z,u) &= \sum_{r=1}^{\infty} T^{(r)}(z) u^r = \sum_{r=1}^{\infty}\sum_{n=0}^{\infty} \Pr\{O_n(H) = r\} z^n u^r \end{aligned}$$

that are defined for $|z| \leq 1$ and $|u| \leq 1$. We adopt here the presentation of Régnier and Szpankowski [281].

We use unlabeled combinatorial structures approach to find $T(z,u)$. In the case of strings, such structures are known as *languages*, and we use this term throughout this section. We extend, however, the definition of languages to include languages defined on a Markovian source. Therefore, we adopt the following definition.

Definition 7.7 *For any language $\mathcal{L}$ we define its generating function $L(z)$ as*

$$L(z) = \sum_{w\in\mathcal{L}} P(w) z^{|w|}$$

where $P(w)$ is the stationary probability of word w occurrence, $|w|$ is the length of w, and we assume that $P(\epsilon) = 1$, where ϵ is the empty word. In addition, we define H-conditional generating function of $\mathcal{L}$ as

$$L_H(z) = \sum_{w\in\mathcal{L}} P(w|w_{-m} = h_1 \cdots w_{-1} = h_m) z^{|w|} = \sum_{w\in\mathcal{L}} P(w|w_{-m}^{-1} = H) z^{|w|}$$

where w_{-i} stands for a symbol preceding the first character of w at distance i.

Since we allow overlaps, the structure of the pattern has a profound impact on the number of occurrences. To capture this, we introduce the autocorrelation language and the autocorrelation polynomial.

Definition 7.8 *Given a string H, we define the* autocorrelation set $\mathcal{A}$ *as:*

$$\mathcal{A} = \{H_{k+1}^m : \; H_1^k = H_{m-k+1}^m\} \; , \tag{7.71}$$

and by HH we denote the set of positions k satisfying $H_1^k = H_{m-k+1}^m$. The generating function of language $\mathcal{A}$ is denoted as $A(z)$ and we call it the autocorrelation polynomial. *Its H-conditional generating function is denoted $A_H(z)$. In particular,*

$$A_H(z) = \sum_{k \in HH} P(H_{k+1}^m | H_k^k) z^{m-k} \tag{7.72}$$

for a Markov source.

Before we proceed, we present a simple example illustrating the definitions we introduced so far. Let us assume that $H = 101$ over a binary alphabet $\mathcal{S} = \{0, 1\}$. Observe that $HH = \{1, 3\}$ and $\mathcal{A} = \{\epsilon, 01\}$, where ϵ is the empty word. Thus, for the unbiased memoryless source we have $A(z) = 1 + \frac{z^2}{4}$, while for the Markovian model of order one, we obtain $A_{101}(z) = 1 + p_{10}p_{01}z^2$.

As mentioned above, we want to estimate the number of pattern occurrences in a text. Alternatively, we can seek the generating function of the language $\mathcal{T}$ that consists of words containing at least one occurrence of H. We will follow this approach, and introduce some other languages to describe $\mathcal{T}$.

Definition 7.9 *Given a pattern H:*

(i) Let $\mathcal{T}_r$ be a language that consists of all words containing exactly $r \geq 1$ occurrences of H.

(ii) Let $\mathcal{R}$ be a set of words containing only one occurrence of H, located at the right end. We also define $\mathcal{U}$ as

$$\mathcal{U} = \{u : \; H \cdot u \in \mathcal{T}_1\} \tag{7.73}$$

where the operation $\cdot$ means concatenation of words. In other words, a word $u \in \mathcal{U}$ if $H \cdot u$ has exactly one occurrence of H at the left end of $H \cdot u$.

(iii) Let $\mathcal{M}$ be the language:

$$\mathcal{M} = \{w : \; H \cdot w \in \mathcal{T}_2 \text{ and } H \text{ occurs at the right end of } H \cdot w\},$$

that is, $\mathcal{M}$ is a language such that $H \cdot \mathcal{M}$ has exactly two occurrences of H at the left and right end of a word from $\mathcal{M}$.

We now can describe languages $\mathcal{T}$ and $\mathcal{T}_r$ in terms of $\mathcal{R}$, $\mathcal{M}$ and $\mathcal{U}$. This will further lead to a simple formula for the generating function of $O_n(H)$.

Theorem 7.10 *Language $\mathcal{T}$ satisfies the fundamental equation*

$$\mathcal{T} = \mathcal{R} \cdot \mathcal{M}^* \cdot \mathcal{U} \ . \tag{7.74}$$

Notably, language $\mathcal{T}_r$ can be represented for any $r \geq 1$ as follows

$$\mathcal{T}_r = \mathcal{R} \cdot \mathcal{M}^{r-1} \cdot \mathcal{U} \ . \tag{7.75}$$

Here, by definition $\mathcal{M}^0 := \{\epsilon\}$ and $\mathcal{M}^* := \bigcup_{r=0}^{\infty} \mathcal{M}^r$.

Proof. We first prove (7.75) and obtain our decomposition of $\mathcal{T}_r$ as follows: The first occurrence of H in a word belonging to $\mathcal{T}_r$ determines a prefix p that is in $\mathcal{R}$. Then, one concatenates a non-empty word w that creates the second occurrence of H. Hence, w is in $\mathcal{M}$. This process is repeated $r-1$ times. Then, one adds after the last H occurrence a suffix u that does not create a new occurrence of H. Equivalently, Hu is such that u is in $\mathcal{U}$, and w is a proper subword of Hu. Finally, a word belongs to $\mathcal{T}$ if for some $1 \leq r < \infty$ it belongs to $\mathcal{T}_r$. The set union $\bigcup_{r=1}^{\infty} \mathcal{M}^{r-1}$ yields precisely $\mathcal{M}^*$. ■

We now prove the following result that summarizes relationships between the languages $\mathcal{R}$, $\mathcal{M}$ and $\mathcal{U}$.

Theorem 7.11 *The languages $\mathcal{M}$, $\mathcal{U}$ and $\mathcal{R}$ satisfy*

$$\bigcup_{k\geq 1} \mathcal{M}^k = \mathcal{W} \cdot H + \mathcal{A} - \{\epsilon\} \ , \tag{7.76}$$

$$\mathcal{U} \cdot \mathcal{S} = \mathcal{M} + \mathcal{U} - \{\epsilon\} \ , \tag{7.77}$$

$$H \cdot \mathcal{M} = \mathcal{S} \cdot \mathcal{R} - (\mathcal{R} - H) \ , \tag{7.78}$$

where $\mathcal{W}$ is the set of all words, $\mathcal{S}$ is the alphabet set and $+$ and $-$ are disjoint union and subtraction of languages. In particular, a combination of (7.77) and (7.78) gives

$$H \cdot \mathcal{U} \cdot \mathcal{S} - H \cdot \mathcal{U} = (\mathcal{S} - \epsilon)\mathcal{R} \ . \tag{7.79}$$

Additionally, we have:

$$\mathcal{T}_0 \cdot H = \mathcal{R} \cdot \mathcal{A}. \tag{7.80}$$

Proof. All the relations above are proved in a similar fashion. We first deal with (7.76). Let k be the number of H occurrences in $\mathcal{W} \cdot H$. By definition, $k \geq 1$ and the last occurrence is on the right: this implies that $\mathcal{W} \cdot H \subseteq \bigcup_{k\geq 1} \mathcal{M}^k$. Furthermore, a word w in $\bigcup_{k\geq 1} \mathcal{M}^k$ is not in $\mathcal{W} \cdot H$ iff its size $|w|$ is smaller than $|H|$. Then, the second H occurrence in Hw overlaps with H, which means that w is in $\mathcal{A} - \epsilon$.

Let us turn now to (7.77). When one adds a character s right after a word u from $\mathcal{U}$, two cases may occur. Either Hus still does not contain a second occurrence of H, which means that us is a non-empty word of $\mathcal{U}$. Or a new H appears, clearly at the right end. Then, us is in $\mathcal{M}$. Furthermore, the whole set $\mathcal{M} + (\mathcal{U} - \epsilon)$ is attained, i.e., a strict prefix of $\mathcal{M}$ cannot contain a new H occurrence. Hence, it is in $\mathcal{U}$, and a strict prefix of a $\mathcal{U}$-word is in $\mathcal{U}$.

We now prove (7.78). Let $x = sw$ be a word in $H \cdot \mathcal{M}$ where s is a symbol from $\mathcal{S}$. As x contains exactly two occurrences of H located at its left and right ends, w is in $\mathcal{R}$ and x is in $\mathcal{A} \cdot \mathcal{R} - \mathcal{R}$. Reciprocally, if a word swH from $\mathcal{S} \cdot \mathcal{R}$ is not in $\mathcal{R}$, then swH contains a second H occurrence starting in sw. As wH is in $\mathcal{R}$, the only possible position is at the left end, and then x is in $H \cdot \mathcal{M}$. We now rewrite:

$$\mathcal{S} \cdot \mathcal{R} - \mathcal{R} = \mathcal{S} \cdot \mathcal{R} - (\mathcal{R} \cap \mathcal{S} \cdot \mathcal{R}) = \mathcal{S} \cdot \mathcal{R} - (\mathcal{R} - H)$$

which yields $H \cdot \mathcal{M} - H = (\mathcal{S} - \epsilon) \cdot \mathcal{R}$.

Finally, we leave the derivation of (7.80) to the reader as an exercise. ∎

Now we translate the language relationships into the associated generating functions. We should point out, however, that one must be careful with products of two languages (i.e., concatenation of two languages). For the memoryless source, the situation is quite simple. Indeed, let $\mathcal{L}_1$ and $\mathcal{L}_2$ be two arbitrary languages with generating functions $L_1(z)$ and $L_2(z)$, respectively. Then, for Bernoulli sources the language $\mathcal{L} = \mathcal{L}_1 \cdot \mathcal{L}_2$ is transferred into the generating function $L(z)$ such that

$$L(z) = L_1(z)L_2(z) \tag{7.81}$$

since $P(wv) = P(w)P(v)$ for $w \in \mathcal{L}_1$ and $v \in \mathcal{L}_2$. In particular, the generating function $L(z)$ of $\mathcal{L} = \mathcal{S} \cdot \mathcal{L}_1$ is $L(z) = zL_1(z)$, where $\mathcal{S}$ is the alphabet set, since $S(z) = \sum_{s \in \mathcal{S}} P(s)z = z$.

For Markovian sources $P(wv) \neq P(w)P(v)$, thus property (7.81) is not any longer true. We have to replace it by a more sophisticated one. We have to condition $\mathcal{L}_2$ on symbols preceding a word from $\mathcal{L}_2$ (i.e., belonging to $\mathcal{L}_1$). In general, for a K order Markov chain, one must distinguish V^K ending states for $\mathcal{L}_1$ and V^K initial states for $\mathcal{L}_2$. For simplicity of presentation, we only consider first-order Markov chains (i.e., $K = 1$), and we write $\ell(w)$ for the last symbol of a word w. In particular, to re-write property (7.81) we must introduce the following conditional generating function for a language $\mathcal{L}$:

$$L_i^j(z) = \sum_{w \in \mathcal{L}} P(w, \ell(w) = j | w_1 = i) z^{|w|}$$

Let now $\mathcal{L} = \mathcal{W} \cdot \mathcal{V}$. Then

$$L_k^l(z) = \sum_{i,j \in \mathcal{S}} p_{ji} W_k^j(z) V_i^l(z) \tag{7.82}$$

where $W_k^j(z)$ and $V_i^l(z)$ are conditional generating functions for $\mathcal{W}$ and $\mathcal{V}$ respectively. To prove this, let $w \in \mathcal{W}$ and $v \in \mathcal{V}$ and observe

$$\begin{aligned} P(wv) &= \sum_{j\in\mathcal{S}} P(wv,\ \ell(w)=j) \\ &= \sum_{j\in\mathcal{S}} P(w,\ \ell(w)=j)P(v|\ell(w)=j) \\ &= \sum_{j\in\mathcal{S}}\sum_{i\in\mathcal{S}} P(w,\ \ell(w)=j)p_{ji}P(v|v_1=i)\ . \end{aligned}$$

After conditioning on the first symbol of $\mathcal{W}$ and the last symbol of $\mathcal{V}$, we prove (7.82).

The lemma below translates (7.76)–(7.78) into generating functions.

Lemma 7.12 *The generating functions associated with languages $\mathcal{M}$,$\mathcal{U}$ and $\mathcal{R}$ satisfy*

$$\frac{1}{1-M_H(z)} = A_H(z) + P(H)z^m\left(\frac{1}{1-z} + F(z)\right) \quad (7.83)$$

$$U_H(z) = \frac{M_H(z)-1}{z-1}\ , \quad (7.84)$$

$$R(z) = P(H)z^m \cdot U_H(z) \quad (7.85)$$

provided the underlying Markov chain is stationary.

Proof. We first prove (7.84). Interestingly, it does not need the stationarity assumption. Let us consider the language relationship (7.77) from Theorem 7.11 which we re-write as $\mathcal{U}\cdot\mathcal{S} - \mathcal{U} = \mathcal{M} - \epsilon$. Observe that $\sum_{j\in\mathcal{S}} p_{i,j}z = z$. Hence, set $\mathcal{U}\cdot\mathcal{S}$ yields (conditioning on the left occurrence of H):

$$\sum_{w\in\mathcal{U}}\sum_{j\in\mathcal{S}} P(wj|H)z^{|wj|} = \sum_{i\in\mathcal{S}}\sum_{w\in\mathcal{U},\ell(w)=i} P(w|H)z^{|w|}\sum_{j\in\mathcal{S}} p_{i,j}z = U_H(z)\cdot z\ .$$

Of course, $\mathcal{M}-\epsilon$ and $\mathcal{U}$ translate into $M_H(z)-1$ and $U_H(z)$, and (7.84) is proved.

We now turn our attention to (7.85), and we use relationship (7.78) of Theorem 7.11. Observe that $\mathcal{S}\cdot\mathcal{R}$ can be rewritten as

$$\sum_{j,i\in\mathcal{S}^2}\sum_{iw\in\mathcal{R}} P(jiw)z^{|jiw|} = z^2\sum_{j\in\mathcal{S}}\sum_{i\in\mathcal{S}}\pi_j p_{j,i}\sum_{iw\in\mathcal{R}} P(w|w_{-1}=i)z^{|w|}\ .$$

But, due to the stationarity of the underlying Markov chain $\sum_j \pi_j p_{j,i} = \pi_i$. As $\pi_i P(w|w_{-1} = i) = P(iw)$, we get $zR(z)$. Furthermore, in (7.78) $H\cdot\mathcal{M} - H$ translates into $P(H)z^m\cdot(M_H(z)-1)$. But, by (7.84), this becomes $P(H)z^m\cdot U_H(z)(z-1)$, and after a simplification, we prove (7.85).

Finally, we deal with (7.83), and prove it using (7.76) from Theorem 7.11. Left-hand side of (7.76) involves language $\mathcal{M}$, hence we must condition on the left occurrence of H. In particular, $\bigcup_{r\geq 1}\mathcal{M}^r+\epsilon$ of (7.76) translates into $\frac{1}{1-M_H(z)}$. Now we deal with $\mathcal{W}\cdot H$ of the right-hand side of (7.76). *Conditioning* on the left occurrence of H, the generating function $W(z)H(z)$ of $\mathcal{W}\cdot H$ becomes

$$\begin{aligned}W_H(z)H(z) &= \sum_{n\geq 0}\sum_{|w|=n} z^{n+m}P(wH|w_{-1}=\ell(H))\\ &= \sum_{n\geq 0}\sum_{|w|=n} z^{n}P(wh_1|w_{-1}=\ell(H))P(v=h_2\dots h_m|v_{-1}=h_1)z^m\ .\end{aligned}$$

We have $P(v=h_2\dots h_m|v_{-1}=h_1)z^m=\frac{1}{\pi_{h_1}}z^mP(H)$, and for $n\geq 0$:

$$\sum_{|w|=n} P(wh_1|w_{-1}=\ell(H))=[\mathsf{P}^{n+1}]_{\ell(H),h_1}$$

where, we recall, $\ell(H)=h_m$ is the last character of H. In summary: language $\mathcal{W}\cdot H$ contributes $P(H)z^m\left[\frac{1}{\pi_{h_1}}\sum_{n\geq 0}\mathsf{P}^{n+1}z^n\right]_{\ell(H),h_1}$, while language $\mathcal{A}-\{\epsilon\}$ introduces $A_H(z)-1$. We now observe that for any symbols i and j

$$\left[\frac{1}{\pi_j}\sum_{n\geq 0}\Pi z^n\right]_{i,j}=\sum_{n\geq 0}z^n=\frac{1}{1-z}.$$

Using the equality $\mathsf{P}^{n+1}-\Pi=(\mathsf{P}-\Pi)^{n+1}$ (which follows from a consecutive application of the identity: $\Pi\mathsf{P}=\Pi$), we finally obtain the sum in (7.83). This completes the proof of the theorem. ■

The lemma above together with Theorem 7.10 suffice to establish the following main result.

Theorem 7.13 (Régnier and Szpankowski 1998) *Let H be a given pattern of size m, and T be a random text of length n generated according to a stationary Markov chain over a V-ary alphabet $\mathcal{S}$. The generating functions $T^{(r)}(z)$ and $T(z,u)$ satisfy*

$$T^{(r)}(z) = R(z)M^{r-1}(z)U_H(z)\ ,r\geq 1 \tag{7.86}$$

$$T(z,u) = R(z)\frac{u}{1-uM_H(z)}U_H(z)\ , \tag{7.87}$$

where

$$M_H(z) = 1+\frac{z-1}{D_H(z)}\ , \tag{7.88}$$

$$
\begin{aligned}
U_H(z) &= \frac{1}{D_H(z)} \ , & (7.89)\\
R(z) &= z^m P(H) \frac{1}{D_H(z)} \ . & (7.90)
\end{aligned}
$$

with

$$D_H(z) = (1-z)A_H(z) + z^m P(H)(1 + (1-z)F(z)). \tag{7.91}$$

The function $F(z)$ *is defined for* $|z| \le R$ *where* $R = \frac{1}{||\mathsf{P}-\mathsf{\Pi}||}$ *as follows*

$$F(z) = \frac{1}{\pi_{h_1}}[(\mathsf{P}-\mathsf{\Pi})(\mathsf{I}-(\mathsf{P}-\mathsf{\Pi})z)^{-1}]_{h_m,h_1} \ . \tag{7.92}$$

where h_1 *and* h_m *are the first and the last symbols of* H*, respectively.*

In the Bernoulli model, $F(z) = 0$*, and hence*

$$D(z) = (1-z)A_H(z) + z^m P(H)$$

with the other formulas as above.

Theorem 7.13 is a starting point of the next findings concerning the average and the variance of O_n. The limiting distribution of O_n is established in Example 8 of Chapter 8.

Theorem 7.14 (Régnier and Szpankowski 1998) *Let the hypotheses of Theorem 7.13 be fulfilled and* $nP(H) \to \infty$*.*

(i) MOMENTS. *The expectation* $\mathbf{E}[O_n(H)]$ *satisfies for* $n \ge m$*:*

$$\mathbf{E}[O_n(H)] = P(H)(n-m+1) \ , \tag{7.93}$$

while the variance becomes for some $r > 1$

$$\mathbf{Var}[O_n(H)] = nc_1 + c_2 + O(r^{-n}) \ , \tag{7.94}$$

where

$$
\begin{aligned}
c_1 &= P(H)(2A_H(1) - 1 - (2m-1)P(H) + 2P(H)E_1)) \ , & (7.95)\\
c_2 &= P(H)((m-1)(3m-1)P(H) - (m-1)(2A_H(1)-1) - 2A'_H(1)) & \\
&- 2(2m-1)P(H)^2E_1 + 2E_2P(H)^2, & (7.96)
\end{aligned}
$$

and the constants E_1*,* E_2 *are*

$$
\begin{aligned}
E_1 &= \frac{1}{\pi_{h_1}}[(\mathsf{P}-\mathsf{\Pi})\mathsf{Z}]_{h_m,h_1}, & (7.97)\\
E_2 &= \frac{1}{\pi_{h_1}}[(\mathsf{P}^2-\mathsf{\Pi})\mathsf{Z}^2]_{h_m,h_1} & (7.98)
\end{aligned}
$$

where $\mathsf{Z} = (\mathsf{I} - (\mathsf{P} - \Pi))^{-1}$ *is the fundamental matrix of the underlying Markov chain.*

For the memoryless source, $E_1 = E_2 = 0$ *since* $\mathsf{P} = \Pi$*, and (7.94) reduces to an equality for* $n \geq 2m - 1$*. Thus*

$$\mathbf{Var}[O_n(H)] = nc_1 + c_2 \ . \tag{7.99}$$

with

$$\begin{aligned} c_1 &= P(H)(2A(1) - 1 - (2m-1)P(H)) \ , \\ c_2 &= P(H)((m-1)(3m-1)P(H) - (m-1)(2A(1)-1) - 2A'(1)). \end{aligned}$$

Proof. We first prove part (i). We evaluate the first two moments of $T(z,u)$ at $u = 1$, that is, $\mathbf{E}[O_n] = [z^n]T'(z,1)$ and $\mathbf{E}[O_n(O_n - 1)] = T''(z,1)$. From Theorem 7.13 we conclude that

$$\begin{aligned} T_u(z,1) &= \frac{z^m P(H)}{(1-z)^2} \ , \\ T_{uu}(z,1) &= \frac{2z^m P(H) M_H(z) D_H(z)}{(1-z)^3} \end{aligned}$$

where $T_u(z,1)$ and $T_{uu}(z,1)$ are the first and the second derivatives of $T(z,u)$ at $u = 1$. Now, we observe that both expressions admit as a numerator a function that is analytic beyond the unit circle. This allows for a very simple computation of the expectation and variance based on the following basic formula

$$[z^n](1-z)^{-p} = \frac{\Gamma(n+p)}{\Gamma(p)\Gamma(n+1)}. \tag{7.100}$$

To obtain $\mathbf{E}[O_n]$ we proceed as follows, for $n \geq m$:

$$\mathbf{E}[O_n] = [z^n]T_u(z,1) = P(H)[z^{n-m}](1-z)^{-2} = (n-m+1)P(H)$$

Let

$$\Phi(z) = 2z^m P(H) M_H(z) D_H(z),$$

which is a polynomial for the memoryless source. Observe that

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

where $f(z)$ is a polynomial of degree $2m-2$. It follows that $[z^n](z-1)f(z)$ is 0 for $n \geq 2m-1$ and, using formula (7.100), we get

$$\mathbf{E}[O_n(O_n-1)] = [z^n]T_{uu}(z,1) = \Phi(1)\frac{(n+2)(n+1)}{2} - \Phi'(1)(n+1) + \frac{1}{2}\Phi''(1).$$

Observing that $M_H(z)D_H(z) = D_H(z) + (1-z)$, we obtain (7.94). In the Markov case, we have to compute the additional term

$$[z^n]\frac{2(z^{2m}P(H)^2F(z))}{(1-z)^2}$$

where $F(z)$ is analytic beyond the unit circle for $|z| \le R$, with $R > 1$. The Taylor expansion of $F(z)$ is $E_1 + (1-z)E_2$ and, applying (7.100) again yields the result. ■

In passing, we add that Nicodéme, Salvy and Flajolet [260] extended these results to patterns that are regular expressions.

7.6.3 Delange's Formula for a Digital Sum

Let $\nu(k)$ be the number of 1-digits in the binary representation of the integer k. Define

$$S(n) = \sum_{k=1}^{n-1} \nu(k),$$

that is, the total number of 1-digits in the binary representations of the integers $k = 1, 2, \ldots, n-1$. We want to derive an exact formula for $S(n)$. We adopt here the presentation of Flajolet *et al.* [119].

Let $v_2(k)$ be the exponent of 2 in the prime decomposition of k. We first prove that for all $k \ge 1$

$$\nu(k) - \nu(k-1) = 1 - v_2(k). \tag{7.101}$$

Indeed, let $k = (d_m, d_{m-1}, \ldots, 1, 0, 0, \ldots, 0)_2$ be the binary representations of k with digits $d_i \in \{0, 1\}$. Observe that 1 is at position $v_2(k)$. Then $k - 1 = (d_m, d_{m-1}, \ldots, 0, 1, 1, \ldots, 1)_2$, where 0 is at position $v_2(k)$. Let M be the number of 1's among the digits $d_m, \ldots, d_{v_2(k)-1}$. Then $\nu(k) = M + 1$ and $\nu(k) = M + v_2(k)$, and (7.101) follows.

Now, we proceed as follows. Summing up (7.101) from $k = 1$ to some $m - 1$, and then again summing from $m = 1$ to $n - 1$ we obtain

$$S(n) = \sum_{m=1}^{n-1}(S(m) - S(m-1)) = \frac{n(n-1)}{2} - \sum_{k=1}^{n-1}(n-k)v_2(k). \tag{7.102}$$

But the sum on the right-hand side can be handled by the Perron–Mellin formula with $m = 1$. Having this in mind, we shall prove the following result.

Theorem 7.15 (Delange 1975) *The sum-of-digits function $S(n)$ satisfies*

$$S(n) = \frac{1}{2}n\log_2 n + n\left(\frac{\log_2 \pi}{2} - \frac{1}{2\ln 2} - \frac{3}{4}\right) + nF(\log_2 n) \tag{7.103}$$

where $F(x)$ is representable by the Fourier series

$$F(x) = \sum_{k\in\mathbb{Z}\setminus\{0\}}^{\infty} f_k e^{2\pi ikx} = \sum_{k\neq 0} f_k e^{2\pi ikx}$$

with

$$f_k = -\frac{1}{\ln 2}\frac{\zeta(\chi_k)}{\chi_k(\chi_k+1)}, \qquad \chi_k = \frac{2\pi ik}{\ln 2}, \qquad k\neq 0.$$

Proof. Expression (7.102) suggests to apply the Perron–Mellin formula with $a_k = v_2(k)$ and $m = 1$. In Example 18 we proved that the Dirichlet series $V(s)$ of $v_2(k)$ is

$$V(s) = \sum_{k=1}^{\infty}\frac{v_2(k)}{k^s} = \frac{\zeta(s)}{2^s-1},$$

hence by the Perron–Mellin formula (7.51) we obtain

$$S(n) = \frac{n(n-1)}{2} - \frac{n}{2\pi i}\int_{2-i\infty}^{2+i\infty}\frac{n^s\zeta(s)}{2^s-1}\frac{ds}{s(s+1)}. \tag{7.104}$$

We now evaluate the integral in (7.104) by the closing-the-box method. We shift the line of integration to the left such that the new contour $\mathcal{C}$ consists of vertical lines at $\Re(s) = 2$ and $\Re(s) = -\frac{1}{4}$ and closed on the North and South at high $\pm T$. This contour will enclose residues of the integrand at $s = 1$, $s = 0$ (a double pole) and simple poles at $s = \chi_k$ for integers $k \neq 0$. To compute the residue at the double pole $s = 0$, we use the following expansions

$$\begin{aligned}
\zeta(s) &= -\frac{1}{2} - s\frac{1}{2}\ln(2\pi) + O(s^2)\\
n^s &= 1 + s\ln(n) + O(s^2)\\
\frac{1}{2^s-1} &= \frac{1}{s\ln 2} - \frac{1}{2} + O(s)\\
\frac{1}{s(s+1)} &= \frac{1}{s} - 1 + O(s).
\end{aligned}$$

Multiplying the above and finding the coefficient at s^{-1} we compute that the residue is equal to

$$-\frac{1}{2}\log_2 n - \log_2\sqrt{\pi} + \frac{1}{2\ln 2} - \frac{1}{4}.$$

The residue at χ_k for $k \neq 0$ gives $F(\log_2 n)$, that is,

$$F(\log_2 n) = \frac{1}{\ln 2}\sum_{k\in\mathbb{Z}\setminus\{0\}}\frac{\zeta(\chi_k)}{\chi_k(\chi_k+1)}n^{\chi_k}.$$

Summing up, we obtain

$$\frac{1}{2\pi i}\int_C \frac{n^s\zeta(s)}{2^s-1}\frac{ds}{s(s+1)} = \frac{1}{2}n\log_2 n + n\left(\frac{\log_2 \pi}{2} - \frac{1}{2\ln 2} - \frac{3}{4}\right) nF(\log_2 n) - nR(n)$$

It remains to prove that the integral over the South/North line vanishes at $T \to \infty$, and that $R(n) \equiv 0$.

Let us first estimate the integral over the horizontal lines that at positions $\pm T$. Due to the bound (2.45) for the $\zeta(s)$ function, we have

$$\left|\int_{-\frac{1}{4}\pm iT}^{2\pm iT} \frac{n^s\zeta(s)}{2^s-1}\frac{ds}{s(s+1)}\right| \le c|\zeta(s)|\left(\frac{n}{2}\right)^2\frac{1}{T^2} = O(T^{-5/4}).$$

Thus, the integral vanishes at $T \to \infty$.

To prove $R(n) \equiv 0$ we use (7.53). We have

$$\begin{aligned} R(n) &= \frac{1}{2\pi i}\int_{-1/4-i\infty}^{-1/4+i\infty} \frac{n^s\zeta(s)}{2^s-1}\frac{ds}{s(s+1)} = \frac{1}{2\pi i}\int_{-1/4-i\infty}^{-1/4+i\infty} \frac{n^s\zeta(s)}{s(s+1)}ds\sum_{k\ge 0}\left(-2^{ks}\right) \\ &= \sum_{k\ge 0}\frac{1}{2\pi i}\int_{-1/4-i\infty}^{-1/4+i\infty} \zeta(s)(2^k n)^s\frac{ds}{s(s+1)} \\ &\overset{(7.53)}{=} 0 \end{aligned}$$

where the second line is justifiable for $\Re(s) < 0$ since both the series and the integral converge absolutely. This proves the theorem. ∎

7.7 Extensions and Exercises

7.1 Prove (7.15).

7.2 Derive a closed-form formula for the Bell numbers.

7.3 Derive (7.34) from the Lagrange formula.

7.4 Let

$$S_n = \sum_{k=0}^{n}\binom{n}{k}a_k p^k q^{n-k}$$

where $p + q = 1$, and a_k is a given sequence. Let $S(z)$ and $A(z)$ be ordinary generating functions of S_n and a_n, respectively. Prove that

$$S(z) = \frac{1}{1-qz}A\left(\frac{pz}{1-qz}\right).$$

7.5 Consider the following sum

$$S_{n,k} = \sum_{i=0}^{n} \binom{n-k}{i} \left(\frac{i}{n}\right)^i \left(1 - \frac{i}{n}\right)^{n-i}.$$

Define

$$s_k(z) = \sum_{n \geq 0} S_{n,k} \frac{n^n}{n!} z^n.$$

Prove that

$$s_k(z) = z^k B(z) B^{(k)}(z)$$

where $B(z) = 1/(1 - T(z))$ is defined in (7.34), and $B^{(k)}(z)$ denotes the kth derivative of $B(z)$.

7.6 Let for $|z| < 1$

$$M(z) = \sum_{n \geq 1} z^n \log n.$$

Prove the following

$$\begin{aligned} M_1(z) &= \sum_{n \geq 1} z^n \log n! = \frac{M(z)}{1-z}, \\ M_2(z) &= \sum_{n \geq 1} z^n \log(2n)! = \frac{1}{2} \left(\frac{M(\sqrt{z})}{1 - \sqrt{z}} + \frac{M(-\sqrt{z})}{1 + \sqrt{z}} \right). \end{aligned}$$

7.7 ⚠ Let $0 < p < 1$. For fixed integer M and all $0 \leq \ell < M$, defined

$$\pi(\ell) = \sum_{k \geq 0} \binom{n}{\ell + kM} p^{\ell + kM} (1-p)^{n - \ell - kM}.$$

Prove that for all $0 \leq \ell \leq M$ we have

$$\pi(\ell) = \frac{1}{M} + O(\rho^n)$$

where $\rho < 1$.

7.8 ⚠ Prove that the variance $\mathbf{Var}[X_n]$ of the quicksort complexity (discussed in Example 17) is

$$\mathbf{Var}[X_n] = 7n^2 - 2(n+1)H_n + 13n - 4(n+1)^2 H_n^{(2)}$$

where $H_n^{(2)} = \sum_{k=1}^{n} \frac{1}{k^2}$.

7.9 Prove (7.60) and (7.64).

7.10 Under what conditions on the function $a(s)$ the following is true

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

7.11 ⚠ Let $\mathcal{H}$ be a set of patterns (e.g., set of all strings that are within given distance from a given pattern H). For such a set we define the correlation language $\mathcal{A}$ and the correlation matrix $\mathsf{A}(z)$ as follows: Given two strings H and F, let $\mathcal{A}$ be the set of words:

$$\mathcal{A} = \{F_{k+1}^m : H_{m-k+1}^m = F_1^k\}$$

and the set of positions k satisfying $F_1^k = H_{m-k+1}^m$ is denoted as HF. Let $\mathsf{A}(z) = \{A_{H_iH_j}(z)\}_{i,j=1,M}$ be the *correlation matrix* where $A_{H_iH_j}(z)$ is the correlation polynomial for H_i and H_j. Let $O_n(\mathcal{H})$ be the number of patters from $\mathcal{H}$ occurring in a random text of length n generating by a memoryless source. As in Section 7.6.2 we write

$$\begin{aligned} T^{(r)}(z) &= \sum_{n\geq 0}\mathsf{Pr}\{O_n(\mathcal{H}) = r\}z^n, \\ T(z,u) &= \sum_{r=1}^{\infty}T^{(r)}(z)u^r = \sum_{r=1}^{\infty}\sum_{n=0}^{\infty}\mathsf{Pr}\{O_n(\mathcal{H}) = r\}z^n u^r. \end{aligned}$$

Prove the following theorem.

Theorem 7.16 (Régnier and Szpankowski 1997) *Let $\mathcal{H}$ be a given set of patterns of length m, and T be a random text of length n generated by a memoryless source. The generating functions $T^{(r)}(z)$ and $T(z,u)$ can be computed as follows:*

$$\begin{aligned} T^{(r)}(z) &= \mathbf{R}^t(z)\mathsf{M}^{r-1}(z0\mathbf{U}(z), \\ T(z,u) &= \mathbf{R}^t(z)u(\mathsf{I} - u\mathsf{M}(z))^{-1}\mathbf{U}(z), \end{aligned}$$

where

$$\begin{aligned} \mathsf{M}(z) &= (\mathsf{D}(z) + (z-1)\mathsf{I})[\mathsf{D}(z)]^{-1}, \\ (\mathsf{I} - \mathsf{M}(z))^{-1} &= \mathsf{A}(z) + \frac{z^m}{1-z}\mathbf{1}\cdot\mathbf{H}^t, \\ \mathbf{U}(z) &= \frac{1}{1-z}(\mathsf{I} - \mathsf{M}(z))\cdot\mathbf{1}, \\ \mathbf{R}^t(z) &= \frac{z^m}{1-z}\mathbf{H}^t\cdot(\mathsf{I} - \mathsf{M}(z)), \end{aligned}$$

and

$$\mathsf{D}(z) = (1 - z)\mathsf{A}(z) + z^m \mathbf{1} \cdot \mathbf{H}^t,$$

In the above, $\mathbf{H} = (P(H_1), \dots, P(H_M))^t$*, upper index* t *means transpose, and* $\mathsf{A}(z) = \{A_{H_i,H_j}(z)\}_{i,j=1.M}$ *is the matrix of the correlation polynomials of patterns from the set* $\mathcal{H}$*.*

Derive from the above the average and the variance of $O_n(\mathcal{H})$.