Chapter 2

Probabilistic and Analytical Models

We first discuss several probabilistic models that are used throughout this book. Then, we briefly review some basic facts from probability theory (i.e., types of stochastic convergence) and complex analysis (e.g., Cauchy's residue theorem). We conclude with a brief discussion of certain special functions (e.g., Euler's gamma function, and Riemann's zeta function) that we shall use throughout the book.

Data structures and algorithms on strings are used in a variety of applications ranging from information theory, telecommunications, wireless communications, approximate pattern matching, molecular biology, game theory, coding theory, source coding, to stock market analysis. It is often reasonable, and sometimes *sine qua none*, to assume that strings are generated by a random source of known or unknown statistics. Applications often dictate what is a reasonable set of probabilistic assumptions. For example, it is inappropriate to postulate that suffixes of a (single) sequence are independent, but experiments support the claim that bits in hashing form an independent sequence, while bases in a DNA sequence are modeled by a Markov chain. We propose below a few basic (generic) probabilistic models that one often encounters in the analysis of problems on words that we used throughout the book. Finally, we review some facts from probability theory and complex analysis, and finish this chapter with a brief discussion of special functions that often arise in the analysis of algorithms.

2.1 Probabilistic Models of Strings

Throughout this book, we shall deal with sequences of discrete random variable. Thus, we need a good set of notations. We write $\{X_k\}_{k=1}^{\infty}$ for one-sided infinite sequence of random variables, however, we often abbreviate it to X provided it is clear from the context that we talk about a sequence not a single variable. We assume that the sequence $\{X_k\}$ is generated over a finite alphabet $\mathcal{A} = \{\omega_1, \ldots, \omega_V\}$ of size V. A partial sequence is denoted as $X_m^n = (X_m, \ldots, X_n)$ for $m < n$. When more than one sequence is analyzed (as in the digital trees) we use *either* an upper index to denote sequence (e.g., $X^1 = \{X_k^1\}_{k=1}^{\infty}, \ldots, X^m = \{X_k^m\}_{k=1}^{\infty}$ for m sequences) *or* we write $X(1) = \{X_k(1)\}_{k=1}^{\infty}, \ldots, X(m) = \{X_k(m)\}_{k=1}^{\infty}$. Finally, we shall always assume that a probability measure exists, and we write $P(X_1^n) = \Pr\{X_k = x_k,\ 1 \le k \le n,\ x_k \in \mathcal{A}\}$ for the probability mass, where we use lower-case letters for a realization of a stochastic process.

Sequences are generated by information sources, usually satisfying some constraints. We also call them probabilistic models. We start with the most elementary source, namely the **memoryless source**.

(B) Memoryless Source

Symbols of the alphabet $\mathcal{A} = \{\omega_1, \ldots, \omega_V\}$ occur independently of one another; thus, a string $X = X_1 X_2 X_3 \ldots$ can be described as the outcome of an infinite sequence of Bernoulli trials in which $\Pr\{X_j = \omega_i\} = p_i$ and $\sum_{i=1}^{V} p_i = 1$. If $p_1 = p_2 = \ldots = p_V = 1/V$, then the model is called *symmetric*; otherwise, it is *asymmetric*. In this book, we often consider *binary alphabet* $\mathcal{A} = \{0, 1\}$ with p being the probability of "0" and $q = 1 - p$ the probability of "1".

When one deals with many strings (e.g., when building a digital tree) additional assumption is made whether the strings are dependent or not.

In many cases, assumption (B) is not very realistic. For instance, if the strings are words from the English language, then there is certainly a dependence between consecutive letters. For example, h is much more likely to follow an s than a b. When this is the case, assumption (B) can be replaced by:

(M) Markov Source

There is a Markovian dependency between consecutive symbols in a string; that is, the probability $p_{ij} = \Pr\{X_{k+1} = \omega_j | X_k = \omega_i\}$ describes the conditional probability of sampling symbol ω_j immediately after symbol ω_i. We denote by $\mathsf{P} = \{p_{ij}\}_{i,j=1}^{V}$ the transition matrix, and by $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_V)$ the stationary vector satisfying $\boldsymbol{\pi}\mathsf{P} = \boldsymbol{\pi}$. In general, X_{k+1} may dependent on last r symbols, and then we have rth order Markov chains (however, in this book we mostly restraint ourselves to $r = 1$).

In information theory **Markov sources** are usually denoted as S_k (i.e., set of states). A generalization of Markov sources is the so–called **finite state source** (cf. [96]) or **hidden Markov source** (cf. [45]) in which the output string X_k is a function of a Markov source S_k, that is, $X_k = f(S_k)$ for some function $f(\cdot)$. Finally, if in a finite state source the current state, say S_k, is computable from the current source sample X_k and previous state S_{k-1} (i.e., $S_k = g(S_{k-1}, S_k)$ for some function g), then the source is called the **unifilar source** (cf. [96]).

There are two further generalizations of Markovian sources, namely **mixing source** and the **stationary ergodic source** that are very useful in practice, especially when dealing with problems of data compression or molecular biology when one expects long dependency among symbols of a string.

(MX) (Strongly) ψ-Mixing Source

Let $\mathbb{F}_m^n$ be a σ-field generated by $\{X_k\}_{k=m}^n$ for $m \le n$. The source is called *mixing*, if there exists a bounded function $\psi(g)$ such that for all $m, g \ge 1$ and any two events $A \in \mathbb{F}_1^m$ and $B \in \mathbb{F}_{m+g}^\infty$ the following holds

$$(1 - \psi(g))\mathsf{Pr}\{A\}\mathsf{Pr}\{B\} \le \mathsf{Pr}\{AB\} \le (1 + \psi(g))\mathsf{Pr}\{A\}\mathsf{Pr}\{B\}. \tag{2.1}$$

If, in addition, $\lim_{g\to\infty} \psi(g) = 0$, then the source is called *strongly* mixing.

In words, model (MX) postulates that the dependency between $\{X_k\}_{k=1}^m$ and $\{X_k\}_{k=m+g}^\infty$ is getting weaker and weaker as g becomes larger (note that when the sequence $\{X_k\}$ is i.i.d., then $\mathsf{Pr}\{AB\} = \mathsf{Pr}\{A\}\mathsf{Pr}\{B\}$). The "quantity" of dependency is characterized by $\psi(g)$ (cf. [30, 34, 67]). Occasionally, we write the ψ-mixing condition in an equivalent form as follows: There exists constants $c_1 \le c_2$ such that

$$c_1\mathsf{Pr}\{A\}\mathsf{Pr}\{B\} \le \mathsf{Pr}\{AB\} \le c_2\mathsf{Pr}\{A\}\mathsf{Pr}\{B\} \tag{2.2}$$

for all $m, g \ge 1$. In some derivations, we shall require that $c_1 > 0$ which will impose further restrictions on the process (cf. Theorem 2.1 below).

A weaker mixing condition, namely, *ϕ-mixing* is defined as above except that ψ is replaced by ϕ function and (2.1) is read as

$$-\phi(g)) \le \mathsf{Pr}\{B|A\} - \mathsf{Pr}\{B\} \le \phi(g), \qquad \mathsf{Pr}\{A\} > 0, \tag{2.3}$$

provided $\phi(g) \to 0$ as $g \to \infty$. In general, ψ-mixing implies ϕ-mixing condition but not *vice versa* (cf. [34]). In this book, we shall mostly work with ψ-mixing condition. Observe, however, that one can define $\phi(g) = \psi(g)\mathsf{Pr}\{B\}$ to translate ϕ-mixing into ψ-mixing *provided* $\phi(g) \to 0$ implies $\psi(g) \to 0$ as $g \to \infty$ (e.g., this is true if for all $B \in \mathbb{F}_{m+g}^\infty$ the probability $\mathsf{Pr}\{B\} > 0$ is away from zero).

The following result shows that Markov sources are special cases of strongly mixing sources.

Theorem 2.1 *If X_k is a finite-state irreducible stationary aperiodic Markov chain, then X_k is strongly ϕ-mixing process, hence also strongly ψ-mixing.*
If, in addition, the transition probabilities $p_{ij} > 0$ for all $i, j \in \mathcal{A}$, then the constant c_1 in (2.2) is positive, i.e., $c_1 > 0$.

Proof. Define two events

$$\begin{aligned} A &= \{X_{t_1} = i_1, X_{t_2} = i_2, \ldots, X_{t_m} = i_m\} & t_1 < t_2 < \cdots < t_m, \\ B_g &= \{X_{s_1+g} = j_1, X_{s_2+g} = j_2, \ldots, X_{s_r+g} = j_r\} & t_m \le s_1 < s_2 < \cdots < s_r \end{aligned}$$

for any $g > 0$. By Markov property

$$\mathsf{Pr}\{A \cap B_g\} = \mathsf{Pr}\{B_g | X_{s_1+g} = j_1\}\mathsf{Pr}\{X_{s_1+g} = j_1 | X_{t_m} = i_m\}\mathsf{Pr}\{A\}. \tag{2.4}$$

Since X_k is irreducible aperiodic Markov chain over a finite alphabet, hence it is ergodic, and since it is also a stationary chain its stationary distribution is $\Pr\{X_t = i\}$ for every $t \geq 0$. We also know from Perron-Frobenius theorem (cf. Chapter 4) that the transition probability $\Pr\{X_{s_1+g} = j_1 | X_{t_m} = i_m\}$ converges exponentially fast to $\Pr\{X_{s_1+g} = j_1\}$. That is, there exists $\rho < 1$ (which is related to the second largest eigenvalue of the transition matrix of the Markov chain) such that

$$|\Pr\{X_{s_1+g} = j_1 | X_{t_m} = i_m\} - \Pr\{X_{s_1+g} = j_1\}| \leq O(\rho^g)$$

as $g \to \infty$. In view of this, and the above we arrive at

$$-O(\rho^g) \leq \Pr\{B_g | A\} - \Pr\{B_g\} \leq O(\rho^g)$$

that proves the first part of the theorem.

To show that $c_1 > 0$ when $p_{ij} > 0$ for all $i, j \in \mathcal{A}$ observe that by (2.4)

$$\Pr\{A \cap B_g\} = C_{ij}(g) \Pr\{A\} \Pr\{B_g\}$$

where

$$C_{ij}(g) = \frac{\Pr\{X_{s_1+g} = j | X_{t_m} = i\}}{\Pr\{X_{s_1+g} = j\}}.$$

Observe that $\Pr\{X_{s_1+g} = j\} > 0$ for all $j \in \mathcal{A}$. Furthermore, if $p_{ij} > 0$ for all $i, j \in A$, then $C_{ij}(1) > 0$ and hence $C_{ij}(g) > 0$ for all $g > 1$. Since $c_1 = \min_{i,j \in \mathcal{A}, g \geq 1}\{C_{ij}(g)\}$, we establish the announced result. ■

The most general probabilistic model that can provide some useful results, is the stationary ergodic source in which we only assume that the sequence is stationary and ergodic, but otherwise dependent in any way.

(S) Stationary and Ergodic Source

The sequence $\{X_k\}_{k=1}^{\infty}$ of letters from a finite alphabet is a *stationary and ergodic* sequence of random variables.

In the stationary model, the probability mass $\Pr\{X_1^n = x_1^n\}$ does not depend on time-shift, that is, if m is an integer , then for every n and m the following holds $\Pr\{X_{1+m}^{n+m} = x_1^n\} = \Pr\{X_1^n = x_1^n\}$ It turns out that that a stationary distribution may be approximated by a k order Markov distribution (cf. Section 4.2.1). Such an approximation becomes asymptotically accurate as $k \to \infty$ (cf. Theorem 6.33 of Chapter 6). In passing we should point out that Bernoulli model and Markov model may be stationary or not. In this book, we shall always work with stationary models, unless stated otherwise.

2.2 Review from Probability

Probability theory is a basic tool in the "average case" analysis of algorithms. We shall learn in this book a considerable number of probabilistic "tricks" that help to gain insights into algorithms behavior. In this section, we briefly review some elementary definitions and results from probability theory. The reader is referred to [67, 70, 71, 214] for more detailed discussions.

In this book, we mostly deal with discrete random variables. Let X_n denote the value of a parameter of interest depending on n (e.g., depth in a suffix tree and/or trie built over n strings). We write $\Pr\{X_n = k\}$ for the probability mass of X_n. The expected value $\mathbf{E}[X_n]$ or mean and the variance $\mathbf{Var}[X_n]$ are computed in a standard way as:

$$\begin{aligned}\mathbf{E}[X_n] &= \sum_{k=0}^{\infty} k\Pr\{X_n = k\},\\ \mathbf{Var}[X_n] &= \sum_{k=0}^{\infty}(k - \mathbf{E}[X_n])^2\Pr\{X_n = k\}.\end{aligned}$$

We also write $\mathbf{E}[X^r] = \sum_{k=0}^{\infty} k^r \Pr\{X_n = k\}$ for the rth moment of X_n. Finally, we write $I(A)$ for the *indicator function*, that is, $I(A) = 1$ when the event A occurs, and zero otherwise.

Below, we first discuss some standard inequalities for expectations, and then review types of stochastic convergence.

2.2.1 Some Useful Inequalities

We review here some inequalities that play a considerable role in probability theory and they are often used in the probabilistic analysis of algorithms.

Markov's Inequality: *For a nonnegative function $g(\cdot)$ and a random variable X the following holds*

$$\Pr\{g(X) \geq t\} \leq \frac{\mathbf{E}[g(X)]}{t} \tag{2.5}$$

for any $t > 0$. Indeed, we have the following chain of obvious inequalities

$$\mathbf{E}[g(X)] \geq \mathbf{E}[g(X)I(g(X) \geq t)] \geq t\mathbf{E}[I(g(X) \geq t)] = t\Pr\{g(X) \geq t\},$$

where we recall that $I(A)$ is the indicator function of the event A.

Chebyshev's Inequality: If one replaces X by $|X - \mathbf{E}[X]|$ in the Markov inequality, then

$$\Pr\{|X - \mathbf{E}[X]| > t\} \leq \frac{\mathbf{Var}[X]}{t^2} \tag{2.6}$$

which is known as Chebyshev's inequality.

Schwarz's Inequality (also called Cauchy-Schwarz): *Let X and Y be such that $\mathbf{E}[X^2] < \infty$ and $\mathbf{E}[Y^2] < \infty$. Then*

$$\mathbf{E}[|XY|]^2 \le \mathbf{E}[X^2]\mathbf{E}[Y^2] \ , \tag{2.7}$$

where $\mathbf{E}[X]^2 := (\mathbf{E}[X])^2$.

Jensen's Inequality: *Let $f(\cdot)$ be a downward convex function, that is, for $\lambda \in (0,1)$*

$$\lambda f(x) + (1-\lambda)f(y) \ge f(\lambda x + (1-\lambda)y) \ .$$

Then

$$f(\mathbf{E}[X]) \le \mathbf{E}[f(X)], \tag{2.8}$$

with the equality when $f(\cdot)$ is a linear function.

Minkowski's Inequality: *If $\mathbf{E}[|X|^p] < \infty$, and $\mathbf{E}[|Y|^p] < \infty$, then $\mathbf{E}[|X+Y|]^p < \infty$ and*

$$\mathbf{E}[|X+Y|^p]^{1/p} \le E[|X|^p]^{1/p} + E[|Y|^p]^{1/p} \tag{2.9}$$

for $1 \le p < \infty$.

Hölder's Inequality: *Let $0 \le \theta \le 1$, and $\mathbf{E}[X] < \infty$, $\mathbf{E}[Y] < \infty$. Then*

$$\mathbf{E}[X^\theta Y^{1-\theta}] \le \mathbf{E}[X]^\theta E[Y]^{1-\theta}. \tag{2.10}$$

Inequality on Means: Let $(p_1, p_2, \ldots, p_n)$ be a probability vector (i.e., $\sum_{i=1}^n p_i = 1$) and $(a_1, \ldots, a_n)$ any vector of positive numbers. The *mean of order* $b \ne 0$ $(-\infty \le b \le \infty)$ is defined as

$$M_n(b) := \left(\sum_{i=1}^n p_i a_i^b \right)^{\frac{1}{b}} .$$

The *inequality on means* asserts that $M_n(b)$ is a increasing function of b, that is,

$$r < s \qquad \Rightarrow \qquad M_n(r) < M_n(s). \tag{2.11}$$

Furthermore, **Hilbert's inequality on means** states

$$\lim_{b \to -\infty} M_n(b) = \min\{a_1, \ldots, a_n\}, \tag{2.12}$$

$$\lim_{b \to \infty} M_n(b) = \max\{a_1, \ldots, a_n\}. \tag{2.13}$$

We shall see in Chapter 4 that Markov's inequality and Chebyshev's inequality constitute a foundation for the so called "first and second moment methods". These two simple methods are probably the most often used probabilistic tools in the analysis of algorithms.

2.2.2 Types of Stochastic Convergence

It is important to know different ways in which random variables can converge. To examine various methods of convergence, let $X_n = \{X_n, n \geq 1\}$ be a sequence of random variables, and let their distribution functions be $F_n(x)$, respectively. A good and easy to read account on various types of convergence can be found in Shiryayev [214].

The first notion of convergence of a sequence of random variables is known as **convergence in probability**. The sequence X_n converges to a random variable X *in probability*, denoted $X_n \to X$ (pr.) or $X_n \stackrel{pr}{\to} X$, if for any $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr\{|X_n - X| < \epsilon\} = 1 .$$

Note that this does not say that the difference between X_n and X becomes very small. What converges here is the *probability* that the difference between X_n and X becomes very small. It is, therefore, possible, although unlikely, for X_n and X to differ by a significant amount and for such differences to occur infinitely often. It is knowing that (cf. [30, 67] *if* $X_n \stackrel{pr}{\to} X$, *then* $f(X_n) \stackrel{pr}{\to} f(X)$ *provided* f *is a continuous function.* The reader is asked to prove it in Exercise 6.

A stronger kind of convergence which does not allow such behavior is called **almost sure convergence** or **strong convergence**. A sequence of random variables X_n converges to a random variable X *almost surely*, denoted $X_n \to X$ (a.s.) or $X_n \stackrel{(a.s.)}{\to} X$, if for any $\epsilon > 0$,

$$\lim_{N\to\infty} \Pr\{\sup_{n\geq N} |X_n - X| < \epsilon\} = 1 .$$

¿From this formulation of almost sure convergence, it is clear that if $X_n \to X$ (a.s.), the probability of infinitely many large differences between X_n and X is zero. As the term strong implies, almost sure convergence implies convergence in probability. An equivalent definition of this convergence says that *the set of sample points for which X_n does not converge to X has probability zero.*

A simple sufficient condition for almost sure convergence can be inferred from the *Borel-Cantelli lemma.* We give it in the following corollary that we invoke frequently throughout the book.

Lemma 2.2 (Borel-Cantelli) *Let $\epsilon > 0$. If $\sum_{n=0}^{\infty} \Pr\{|X_n - X| > \epsilon\} < \infty$, then $X_n \stackrel{a.s.}{\to} X$.*

Proof. It follows directly from the following chain of inequalities

$$\Pr\{\sup_{n\geq N} |X_n - X| \geq \epsilon\} = \Pr\{\bigcup_{n\geq N} (|X_n - X| \geq \epsilon)\} \leq \sum_{n\geq N} \Pr\{|X_n - X| \geq \epsilon\} \to 0 .$$

The inequality above is a consequence of the fact that the probability of a sum of events is smaller than the sum of the probability of the events. The last convergence is a consequence of our assumption that $\sum_{n=0}^{\infty} \Pr\{|X_n - X| > \epsilon\} < \infty$. ■

A third type of convergence is defined on the distribution functions $F_n(x)$. The sequence of random variables X_n **converges in distribution** or **converges in law** to the random variable X, denoted $X_n \xrightarrow{d} X$ if

$$\lim_{n\to\infty} F_n(x) = F(x) \tag{2.14}$$

for each point of continuity of $F(x)$. One can prove that the above definition is equivalent to the following: $X_n \xrightarrow{d} X$ *if*

$$\lim_{n\to\infty} \mathbf{E}[f(X_n)] = \mathbf{E}[f(X)] \tag{2.15}$$

for all bounded continuous functions f. We shall return to this type of convergence in Section 8.6.

The next type of convergence is the **convergence in mean of order** p or **convergence in** L^p which postulates that $\mathbf{E}[|X_n - X|^p] \to 0$ as $n \to \infty$. We write is as $X_n \xrightarrow{L^p} X$. Finally, we introduce **convergence in moments** for which $\lim_{n\to\infty} \mathbf{E}[X_n^p] = \mathbf{E}[X^p]$ for any $p \geq 1$.

We now describe relationships (implications) between various type of convergence. The reader is referred to [30, 67, 214] for a proof.

Theorem 2.3 *We have the following implications:*

$$X_n \xrightarrow{a.s.} X \quad \Rightarrow \quad X_n \xrightarrow{pr} X, \tag{2.16}$$

$$X_n \xrightarrow{L^p} X \quad \Rightarrow \quad X_n \xrightarrow{pr} X, \tag{2.17}$$

$$X_n \xrightarrow{pr} X \quad \Rightarrow \quad X_n \xrightarrow{d} X \tag{2.18}$$

$$X_n \xrightarrow{L^p} X \quad \Rightarrow \quad \mathbf{E}[X_n^p] \to \mathbf{E}[X^p]. \tag{2.19}$$

No other implications hold in general.

It is easy to devise an example showing that convergence in probability does not imply the convergence in mean (e.g., take $X_n = n$ with probability $1/n$ and zero otherwise). To obtain such an implication one needs somewhat stronger conditions. For example, if $|X_n| \leq Y$ and $\mathbf{E}[Y] < \infty$, then by the *dominated convergence theorem* we know that convergence in probability implies convergence in mean. To generalize it, probabilist introduced the so called **uniform integrability**. *It is said that a sequence* $\{X_n,\ n \geq 1\}$ *is uniformly integrable if*

$$\sup_{n\geq 1} \mathbf{E}[|X_n| I(|X_n| > a)] \to 0 \tag{2.20}$$

when $a \to \infty$. The above is equivalent to

$$\lim_{a\to\infty} \sup_{n\geq 1} \int_{|x|>a} x dF_n(x) = 0.$$

Then the following is true: *if* X_n *is uniformly integrable, then* $X_n \xrightarrow{pr} X$ *implies* $X_n \xrightarrow{L^p} X$, *and* $X_n \xrightarrow{L^p} X$ *implies convergence in moments, that is,* $\mathbf{E}[X_n^p] \to \mathbf{E}[X^p]$.

2.3 Review from Complex Analysis

Much of the necessary complex analysis involves the use of Cauchy's integral formula and Cauchy's residue theorem. We informally recall a few facts from analytical functions, and then discuss the above two theorems. We shall return to them in Chapter 7. For precise definitions and formulations the reader is referred to many excellent books such as Henrici [114], Hille [115], Remmert [204] or Titchmarsh [238]. We shall follow here Flajolet and Sedgewick [85].

The main topic of this section is **analytic function**, its definitions, properties and a few applications. Analytic functions can be characterized in a three equivalent manner, namely: by convergent series, by differentiability property, and by integrals vanishing on cycles. While discussing these three definitions, we introduce the necessary concepts from complex analysis.

Convergent Series. A function $f(z)$ of complex variable z is analytical at point $z = a$ if it has a convergent series representation in the neighborhood of $z = a$, that is,

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

¿From this definition, one immediately conclude that if $f(z)$ and $g(z)$ are analytic, then $f(z) + g(z)$, $f(z)g(z)$, and $f(z)/g(z)$ with $g(z) \neq 0$ are analytic, too. Furthermore, the above implies that if $f(z)$ is analytic at $z = a$, then there is a disk called the *disk of convergence* such that the series representing $f(z)$ is convergent inside this disk and divergent outside the disk. The radius of this disk is called *radius of convergence*. In Chapter 7 we discuss it in depth; in particular, we prove Hadamard's theorem (cf. Theorem 7.1) that gives the radius of convergence of a series in terms of its coefficients.

Holomorphic Function. Another equivalent name for analytic functions is holomorphic. The holomorphic function $f(z)$ it has a derivative at a point $z = a$ defined as

$$\frac{df(z)}{dz} = \lim_{\Delta z \to 0} \frac{f(z + \Delta z) - f(z)}{\Delta z}$$

that does not depend on the way Δz goes to zero. Riemann proved that a holomorphic function is analytic, that is, it has a local convergent series representation.

Integrals. Finally, we discuss a very important concept of complex integrals. We should first define a *simply connected* domain as an open connected set having the property that any simple closed path can be continuously deformed to a point inside this set. It is a basic fact of complex analysis that *if $f(z)$ is an analytic function on an open simply connected set, then*

$$\oint f(z) = 0$$

along any closed path inside this set. This fact actually is equivalent to analyticity of $f(z)$. It is worth knowing that one can bound the integrals as follows:

$$\left|\int_{\mathcal{C}} f(z)\right| \leq |\mathcal{C}| \max_{\mathcal{C}}\{|f(z)|\} \tag{2.21}$$

where $|\mathcal{C}|$ is the path length of the curve $\mathcal{C}$. Finally, we cite one more result that plays a significant role in the analysis.

Theorem 2.4 (i) *Let $\mathcal{C}$ be a piecewise continuous differentiable path and Ω a region. If $g(w,z)$ is a continuous on $\mathcal{C} \times \Omega$ and for every $w \in \mathcal{C}$ the function $g(w,z)$ is analytic in Ω, then the function*

$$h(z) = \int_{\mathcal{C}} g(\xi, z)d\xi$$

is analytic in Ω.

(ii) *The function*

$$h(z) = \int_a^\infty f(t,z)dt$$

is analytic function of z at all points z if: (I) the integral converges; (ii) $f(z,t)$ is an analytic function of z and t; (iii) $\frac{\partial f(z,t)}{\partial t}$ is a continuous function of both variables; (iv) the integral $\int_a^\infty \frac{\partial f(z,t)}{\partial z}dt$ converges uniformly.

Example 2.1: *A Simple Integral*

In complex analysis a circular integral over z^{-n} for $n = 0, \pm 1, \dots$ plays a very special role. Observe that this function is not analytic at $z = 0$. It is easy to see the following (we substitute $z = e^{2\pi i t}$)

$$\oint z^{-n}dz = \int_0^1 e^{-2\pi i n}dt = \begin{cases} 2\pi i & \text{for } n = 1 \\ 0 & \text{otherwise.} \end{cases} \tag{2.22}$$

We shall often use this simple fact in this book. □

Meromorphic Functions and Residues. A quotient of two analytical functions gives a *meromorphic functions* that are analytical with an exception of a finite number of points called poles where the denominator vanishes. More formally, a meromorphic function $f(z)$ can be represented in a neighborhood of $z = a$ with $z \neq a$ by the Laurent series as:

$$f(z) = \sum_{n \geq -M} f_n(z-a)^n \ ,$$

for some integer M. If the above holds with $f_{-M} \neq 0$, then it is said that $f(z)$ has a *pole* of order M at $z = a$. An important tool frequently used in the complex analysis is the *residue.*

The residue of $f(z)$ at a point a is the coefficient at $(z-a)^{-1}$ in the expansion of $f(z)$ around a, and it is denoted as

$$\mathrm{Res}[f(z);\ z=a] := f_{-1} = \lim_{z\to a}(z-a)f(z).$$

There are many simple rules to evaluate residues and the reader can find them in any standard book on complex analysis (e.g., [114, 204]). For example, if $f(z)$ and $g(z)$ are analytic around $z=a$ and a is a simple pole, then

$$\mathrm{Res}[\frac{f(z)}{g(z)};\ z=a] = \frac{f(a)}{g'(a)}, \qquad g(a)=0, \quad g'(a)\neq 0 \tag{2.23}$$

$$\mathrm{Res}[f(z)g(z);\ z=a] = g(a)\mathrm{Res}[f(z);\ z=a] \tag{2.24}$$

Evaluating residues of multiply poles is much more computationally involved. Actually, the easiest way is to use the `series` commend in MAPLE that produces a series development of a function. The residue is simply the coefficient at $(z-a)^{-1}$. For example, the following session of MAPLE computes series of $f(z) = 1/(1-2^z)^2$ at $z=0$:

```
> series(1/(1-2^z)^2, z=0, 4);
```

$$\frac{1}{\ln(2)^2}z^{-2} - \frac{1}{\ln(2)}z^{-1} + \frac{5}{12} + \mathrm{O}(z)$$

¿From the above we see that $\mathrm{Res}[f(z);\ z=0] = \frac{1}{\ln 2}$.

Residues are very important in evaluating contour integrals. In fact,

Theorem 2.5 (Cauchy Residue Theorem) *If $f(z)$ is analytic within and on the boundary of a simple closed curve C except at a finite number of poles $a_1, a_2, \ldots, a_N$ inside of C having residues* $\mathrm{Res}[f(z);\ z=a_1], \ldots, \mathrm{Res}[f(z);\ z=a_N]$, *then*

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

where the curve $\mathcal{C}$ is traversed counterclockwise.

Sketch of Proof: Let us assume there is only one pole at $z=a$. Since $f(z)$ is meromorphic, it has the Laurent expansion which after integration over $\mathcal{C}$ leads to

$$\int_C f(z)dz = \sum_{n\geq 0} f_n \int_C (z-a)^n + \int_C \frac{dz}{z-a} = 2\pi i \mathrm{Res}[f(z), z=a],$$

since the first integral is zero by the calculation of Example 1. ■

The Cauchy residue theorem can be used to prove the next *very* important result, namely **Cauchy's Integral Theorem**. In Part III of this book, it will be the most often used paradigm.

Theorem 2.6 (Cauchy Coefficient Formula) *Let $f(z)$ be analytic inside a simply connected region with $\mathcal{C}$ being a closed curve oriented counter-clockwise that encircles the origin $z = 0$. Then*

$$f_n := [z^n]f(z) = \frac{1}{2\pi i}\oint f(z)\frac{dz}{z^{n+1}}. \tag{2.25}$$

Also, the following holds

$$f^{\langle k\rangle}(z) = \frac{k!}{2\pi i}\oint \frac{f(w)dw}{(w-z)^{k+1}} \tag{2.26}$$

where $f^{\langle k\rangle}(z)$ is the kth derivative of $f(z)$.

Proof: These formulæ are direct consequences of Cauchy's Residue Theorem. For example,

$$[z^n]f(z) = \mathrm{Res}[f(z)z^{-n-1}: \ z = 0].$$

where $f_n = [z^n]f(z)$. ■

We finish this section with two examples that illustrate what we have learned so far. The second example is particularly important since we shall do this type of calculation quite often in Part III of this book.

Example 2.2: *Liouville's Theorem*

We prove the following theorem of Liouville: *If a function $f(z)$ is analytic in the whole complex plane $\mathbb{C}$ and $f(z) \le B$ for some constant B, then $f(z)$ must be constant.* Let us first assume that $f(z) = B$. Then by Example 1 we show that $f_n = 0$ for $n \ge 1$. Conversely, let $|f(z)| \le B$. Then $f_n = O(R^{-n})$ for *any* R (it doesn't matter how large) by (2.25), so f_n must be zero except for $n = 0$. □

Example 2.3: *Expansion of Meromorphic Functions*

Let $f(z)$ be a meromorphic for $|z| < R$ with poles at $a_1, \ldots, a_m$. Then

$$f_n := [z^n]f(z) = -\sum_{j=1}^{m}\mathrm{Res}[f(z), z = a_j] + O(R^{-n}).$$

Indeed, let us compute the following integral

$$I_n = \frac{1}{2\pi i}\oint_R f(z)\frac{dz}{z^{n+1}}$$

where the integration is along a circle of radius R and center at the origin. Observe that the circle of radius R contains all the poles of the function in addition to the pole at $z = 0$. On one hand, by our estimate (2.21) this integral is $|I_n| = O(R^{-n})$. On the other hand, inside the circle of radius R there are poles at $z = 0$ and $z = a_j$. The pole at $z = 0$ contributes f_n by Cauchy's Coefficient Formula while the poles at $z = a_j$ contribute $\mathrm{Res}[f(z), z = a_j]$which leads to our result. $\square$

2.4 Special Functions

In this subsection, we review properties of two special functions, namely: **Euler's gamma function** and **Riemann's zeta function**. Both are used very frequently in the average case analysis of algorithms. We shall meet them quite often in Part III of this book. There are many excellent books treating these functions in depth. The reader might inspect a new book of Temme [236], an excellent account on special functions and asymptotics by Olver [186], an old traditional (but still in use) book of Whittaker and Watson [250]. The new book [9] by Andrews, Askey and Roy is also excellent. We will follow here Temme. A full list of properties for these special functions can be found in the handbook of Abramowitz and Stegun [1] and the three volume opus on transcendental functions [22] by H. Bateman.

2.4.1 Euler's Gamma Function

A desire to generalize $n!$ to the complex plane, led Euler to introduce one of the most useful special function, namely: *gamma function*. It is defined as

$$\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt, \qquad \Re(z) > 0. \tag{2.27}$$

To see that the above integral generalized $n!$, let us integrate it by parts. We obtain

$$\Gamma(z+1) = -\int_0^\infty t^z d\left(e^{-t}\right) = z\Gamma(z). \tag{2.28}$$

Observe now that $\Gamma(1) = 1$, and then $\Gamma(n) = n!$ for n natural, as desired.

Analytical Continuation. We now analytically continue the gamma function to the whole complex plane. We first extend the definition to $-1 < \Re(z) < 0$ by considering (2.28) and writing

$$\Gamma(z) = \frac{\Gamma(z+1)}{z}, \qquad -1 < \Re(z) < 0.$$

Since $\Gamma(z+1)$ is well defined for $-1 < \Re(z) < 0$ (indeed, $\Re(z+1) > 0$), we can enlarged the region of definition to the strip $-1 < \Re(z) < 0$. However, at $z = 0$ there is a pole whose residue is easily to evaluate, that is,

$$\mathrm{Res}[\Gamma(z); z = 0] = \lim_{z \to 0} z\Gamma(z) = 1.$$

Now, we can further extend to $-2 < \Re(z) < -1$ by applying (2.28) twice to get

$$\Gamma(z) = \frac{\Gamma(z+2)}{z(z+1)}, \qquad -2 < \Re(z) < -1.$$

Observe that

$$\mathrm{Res}[\Gamma(z); z = -1] = \lim_{z \to -1} (z+1)\Gamma(z) = -1.$$

In general, let us assume we have already defined $\Gamma(z)$ up to the strip $-n < \Re(z) < -n+1$. Then, extension to $-n-1 < \Re(z) < -n$ is obtained by

$$\Gamma(z) = \frac{\Gamma(z+n+1)}{z(z+1)\cdots(z+n)}.$$

The residue at $z = -n$ becomes

$$\mathrm{Res}[\Gamma(z); z = -n] = \lim_{z \to -n} (z+n)\Gamma(z) = \frac{(-1)^n}{n!} \tag{2.29}$$

for all $n = 0, -1, \ldots$. The above formula is the most useful, and certainly responsible for many applications of the gamma function to discrete mathematics and the analysis of algorithms. The reader should take a good look at it.

Beta Function. Closely related to the gamma function is the *beta function* defined as

$$B(w, z) = \int_0^1 t^{w-1}(1-t)^{z-1} dt, \qquad \Re(w) > 0 \text{ and } \Re(z) > 0. \tag{2.30}$$

The reader is asked to prove (cf. Ex. 11) the following formula that relates the gamma and the beta functions

$$B(w, z) = \frac{\Gamma(w)\Gamma(z)}{\Gamma(w+z)}. \tag{2.31}$$

Beta function is often used to prove properties of the gamma function (see exercises at the end of this chapter).

Infinite Products. Weierstrass in 1856 defined the gamma function as

$$\frac{1}{\Gamma(z)} = z e^{\gamma z} \prod_{n=1}^{\infty} \left[\left(1 + \frac{z}{n}\right) e^{-z/n} \right]$$

where the constant γ is known as Euler's constant and is defined as

$$\gamma = \lim_{n \to \infty} \left(\sum_{k=1}^{n} \frac{1}{k} - \ln(n+1) \right) = 0.5772157\ldots \ .$$

Using this definition of γ, and some algebraic manipulation the reader is asked in Ex. 12 to derive the following Euler's product representation

$$\Gamma(z) = \lim_{n\to\infty} \frac{n!n^z}{z(z+1)\cdots(z+n)}.$$

In Ex. 2.23 we suggest to show that the above representation is equivalent to the integral formula (2.28) for $\Re(z) > 0$.

Gamma Function on Imaginary Line. Several applications of the gamma function follow from its desirable behavior for purely imaginary variable, that is, for $z = iy$ where $y \in \mathbb{R}$. Let us start with a well–know *reflection formula* of the gamma function (see Ex. 14)

$$\Gamma(z)\Gamma(1-z) = \frac{\pi}{\sin \pi z}, \qquad z \notin \mathbb{Z}. \tag{2.32}$$

Set now $z = iy$ for $y \in \mathbb{R}$ and use (2.28) to get

$$\Gamma(iy)\Gamma(-iy) = \frac{\pi}{iy \sin \pi iy}.$$

But $i \sin iy = \sinh y = (e^y - e^{-y})/2$, we finally derive

$$\Gamma(iy)\Gamma(-iy) = |\Gamma(iy)|^2 = \frac{\pi}{y \sinh \pi y},$$

and even more importantly

$$|\Gamma(iy)| \sim \sqrt{\frac{2\pi}{|y|}} e^{-\pi|y|/2} \quad \text{as} \quad |y| \to \pm\infty. \tag{2.33}$$

The above shows that the gamma function decays exponentially fast over imaginary lines. In general, using (2.28) one can show that

$$|\Gamma(x+iy)| \sim \sqrt{2\pi}|y|^{x-\frac{1}{2}} e^{-\pi|y|/2} \quad \text{as} \quad |y| \to \pm\infty \tag{2.34}$$

for any $x \in \mathbb{R}$.

Asymptotic Expansions. In many applications Stirling's asymptotic formula for $n!$ proves to be extremely useful. Not surprisingly then, the same is true for asymptotic expansion of the gamma function. It can be proved [236, 250] that (cf. Example 3 in Chapter 8)

$$\Gamma(z) = \sqrt{2\pi} z^{z-\frac{1}{2}} e^{-z} \left(1 + \frac{1}{12z} + \frac{1}{288z^2} + \cdots\right)$$

for $z \to \infty$ when $|\arg(z)| < \pi$.

The above asymptotics is helpful in deriving another approximation for large z. Let us consider the ratio of $\Gamma(z+a)$ and $\Gamma(z+b)$ for large z. To see how it behaves consider

$$\Gamma^*(z) = \frac{\Gamma(z)}{\sqrt{2\pi}z^{z-\frac{1}{2}}e^{-z}}$$

which tends to one as $z \to \infty$. Then,

$$\frac{\Gamma(z+a)}{\Gamma(z+b)} = z^{a-b}\frac{\Gamma^*(z+b)}{\Gamma^*(z+b)}Q(z,a,b)$$

where $Q(z,a,b) = 1 + O(1/z)$ as $z \to \infty$. Hence, the above ratio is approximately z^{a-b}. More precisely

$$z^{b-a}\frac{\Gamma(z+a)}{\Gamma(z+b)} = 1 + \frac{(a-b)(a+b-1)}{2z} + O(1/z^2) \tag{2.35}$$

as $z \to \infty$ along any curve joining $z = 0$ and $z = \infty$ providing $z \neq -a, -a-1, \ldots$, and $z \neq -b, -b-1, \ldots$. A full asymptotic expansion of this ratio can be found in Exercise 16 and in Temme [236].

Psi Function. The derivative of the logarithm of the gamma function plays an important role in the theory and applications of special functions. It is known under the name of *psi function* and is defined as:

$$\psi(z) = \frac{d}{dz}\ln\Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}.$$

Using the Weierstrass's product form of the gamma function, one can derive the following

$$\psi(z) = -\gamma + \sum_{n=0}^{\infty}\left(\frac{1}{n+1} - \frac{1}{z+n}\right), \qquad z \neq 0, -1, -2, \ldots. \tag{2.36}$$

¿From the above, we conclude that the psi function possesses simple poles at all non-positive integers, and

$$\mathrm{Res}[\psi(z); z = -n] = \lim_{z \to -n}(z+n)\psi(z) = -1, \qquad n \in \mathbb{N}. \tag{2.37}$$

Laurent's Expansions. As we observed above, the gamma function and the psi function do have simple poles at all non-positive integers. Thus, one can expand these functions around $z = -n$ using the Laurent's series. We obtain the following results:

$$\begin{aligned}
\Gamma(z) &= \frac{(-1)^n}{n!}\frac{1}{(z+n)} + \psi(n+1) \qquad\qquad (2.38)\\
&+ \frac{1}{2}(z+n)\left(\pi^2/3 + \psi^2(n+1) - \psi'(n+1)\right) + O((z+n)^2),\\
\psi(z) &= \frac{-1}{(z+n)} + \psi(m+1) + \sum_{k=2}^{\infty}\left((-1)^n\zeta(n) + \sum_{i=1}^{k} i^{-k}\right)(z+n)^{k-1}, \qquad (2.39)
\end{aligned}$$

where $\zeta(z)$ is the Riemann zeta function defined below in Section 2.4.2. In particular,

$$\Gamma(z) = \frac{1}{z} - \gamma + O(z), \tag{2.40}$$

$$\Gamma(z) = \frac{-1}{z+1} + \gamma - 1 + O(z+1). \tag{2.41}$$

We shall use the above formulæ quite often in Part III of this book (cf. Section 7.5.3).

2.4.2 Riemann's Zeta Function

We discuss here the Riemman zeta function $\zeta(z)$. This is the most fascinating special function that still hides from us its beautiful properties (e.g., the Riemann conjecture concerning zeros of $\zeta(z)$). We only uncover a tip of the iceberg. The reader is referred to [239] for more in depth discussion. The Riemann zeta function is defined as

$$\zeta(z) = \sum_{n=1}^{\infty} \frac{1}{n^z}, \qquad \Re(z) > 1. \tag{2.42}$$

The *generalized* zeta function $\zeta(z, a)$ is defined as

$$\zeta(z, a) = \sum_{n=0}^{\infty} \frac{1}{(n+a)^z}, \qquad \Re(z) > 1,$$

where $a \neq 0, -1, -2, \ldots$ is a constant. It is evident that $\zeta(z, 1) = \zeta(z)$.

Integral Representation and Analytical Continuation. To continue it analytically to the whole complex plane we use an integral representation. We start with a simple observation that

$$\frac{1}{n^z} = \frac{1}{\Gamma(z)} \int_0^{\infty} t^{z-1} e^{-nt} dt$$

which follows from the definition of the gamma function and a substitution $w = nt$. Summing up the above over all $n \geq 1$, and interchanging summation and integration (which is allowed due to uniform convergence) we obtain

$$\zeta(z) = \frac{1}{\Gamma(z)} \int_0^{\infty} \frac{t^{z-1}}{e^t - 1} dt, \qquad \Re(z) > 1. \tag{2.43}$$

A modification of this integral (cf. [236]) can be used to analytically continue the zeta function to the whole plane except $z = 1$. Furthermore, it can be used to prove Riemman's functional equation (cf. [250])

$$\zeta(s)\Gamma\left(\frac{s}{2}\right)\pi^{-s/2} = \zeta(1-s)\Gamma\left(\frac{1-s}{2}\right)\pi^{-(1-s)/2} \tag{2.44}$$

that is true for all s where $\zeta(s)$ and $\Gamma(s)$ functions are defined.

Zeta Function on Imaginary Line. The zeta function $\zeta(z)$ behaves in a different way than the gamma function on imaginary lines, which has important consequences for certain asymptotics that we shall learn in Chapter 8 of this book. Let now $z = x+iy$, where $x, y \in \mathbb{R}$, and we inspect the behavior of $\zeta(z)$ for fixed x and $y \to \pm\infty$. The following result is proved in Whittaker and Watson [250]

$$\zeta(z) = \begin{cases} O(|y|^{\frac{1}{2}-x}) & \text{for } x < 0 \\ O(|y|^{\frac{1-x}{2}}) & \text{for } 0 < x < 1 \\ O(1) & \text{for } x > 1. \end{cases} \tag{2.45}$$

as $y \to \pm\infty$. It is worth remembering the above relationships.

Laurent's Expansion. Using Hankel's type integral representation of the zeta function (cf. Part III), one can prove that $\zeta(z)$ has only one pole at $z = 1$. Its Laurent's series around $z = 1$ is

$$\zeta(z) = \frac{1}{z-1} + \sum_{k=0}^{\infty} \frac{(-1)^k \gamma_k}{k!}(z-1)^k \tag{2.46}$$

where γ_k are the so–called Stieltjes constants for $k \geq 0$ defined as

$$\gamma_k = \lim_{m\to\infty} \left(\sum_{i=1}^{m} \frac{\ln^k i}{i} - \frac{\ln^{k+1} m}{k+1} \right).$$

In particular, $\gamma_0 = \gamma = 0.577215\ldots$ is the Euler's constant, and $\gamma_1 = -0.072815\ldots$. From the above, we conclude that

$$\mathrm{Res}[\zeta(z); z = 1] = 1.$$

We shall use $\zeta(z)$ function quite extensively in Part III of this book. We learn more about this fascinating function. Among others, we relate the value of the zeta function on nonnegative integers to the so called Bernoulli numbers that we also encounter of the pages of this book (cf. Table 8.2.1). Stay tune!

2.5 Extensions and Exercises

2.1 Prove the following formula for the $(k+1)$st moment of a discrete nonnegative random variable X:

$$\mathbf{E}[X^{k+1}] = \sum_{m\geq 0} \Pr\{X > m\} \sum_{i=0}^{k} (m+1)^{k-i} m^i.$$

In particular,

$$\mathbf{E}[X] = \sum_{m\geq 0} \Pr\{X > m\}.$$

2.2 Prove that (2.14) and (2.15) are equivalent.

2.3 Construct counter-examples to every reverse implication from Theorem 2.3.

2.4 Prove Theorem 2.3.

2.5 Prove the following result: *If X_n is a sequence of nonnegative random variables such that $X_n \stackrel{a.s.}{\to} X$ and $\mathbf{E}[X_n] \to \mathbf{E}[X]$, then $X_n \stackrel{L^1}{\to} X$.*

2.6 Prove that if $X_n \stackrel{pr}{\to} X$, then $f(X_n) \stackrel{pr}{\to} f(X)$ provided f is a continuous function.

2.7 Prove (2.23)–(2.24).

2.8 Compute the residue of $f(z) = \Gamma(z)(1-2^z)^{-1}$ at $z = 0$.

2.9 Extend Liouville's Theorem to polynomial functions, that is, prove that *if $f(z)$ is of at most polynomial growth, $|f(z)| \le B|z|^r$, then it is a polynomial.*

2.10 Estimate the growth of the coefficients $f_n = [z^n]f(z)$ of

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

using our arguments from Example 3.

2.11 Prove formula (2.31).
Hint. Compute the integral

$$I(w,z) = \int_0^\infty \int_0^\infty x^{2w-1} y^{2z-1} e^{-(x^2+y^2)} dx dy$$

in two different ways (cf. [236])

2.12 Show that the Weierstrass product formula for the gamma function implies the Euler product form representation for the gamma function.
Hint. Use the definition the Euler's constant.

2.13 Prove that the product formulas for the gamma function are equivalent to the integral formula (2.27).
Hint. Compute

$$\int_0^\infty \left(1 - \frac{t}{n}\right)^n t^{z-1} dt$$

in two different manners (cf. [236]).

2.14 Prove the *reflection formula* (2.32) for the gamma function.
Hint. Compute

$$B(z,1-z)=\int_0^\infty \frac{s^{z-1}}{1+s}ds$$

in two ways (cf. [236]).

2.15 Prove the growth (2.45) of $\zeta(z)$ function on imaginary lines for $0<x<1$. In general, prove that: *if* $\sum_{n\geq 0} a_n<\infty$, *then the series*

$$f(z)=\sum_{n=1}^{\infty}\frac{a_n}{n^z}$$

grows like $|f(x+iy)|=O(|y|^{1-x})$ *for* $0<x<1$ *as* $y\to\pm\infty$ (cf. [238]).

2.16 In this exercise we extend the asymptotic formula (2.35) for $\Gamma(z+a)/\Gamma(z+b)$. Prove the following asymptotic expansion for $\Re(b-a)>0$ (cf. [236])

$$\frac{\Gamma(z+a)}{\Gamma(z+b)}\sim z^{a-b}\sum_{n=0}^{\infty}c_n\frac{\Gamma(b-a+n)}{\Gamma(b-a)}\frac{1}{z^n}\qquad \text{as}\qquad z\to\infty \tag{2.47}$$

where

$$c_n=(-1)^n\frac{B_n^{a-b+1}(a)}{n!}$$

and $B_n^{(w)}(x)$ are the so-called *generalized Bernoulli polynomials* defined as

$$e^{xz}\left(\frac{z}{e^z-1}\right)^w=\sum_{n=0}^{\infty}\frac{B_n^{(w)}(x)}{n!}z^n\qquad |z|<2\pi.$$

Hint. Express the ratio $\Gamma(z+a)/\Gamma(z+b)$ in terms of the beta function, that is,

$$\frac{\Gamma(z+a)}{\Gamma(z+b)}=\frac{B(z+a,b-a)}{\Gamma(b-a)}.$$

Then, use the integral representation of the beta function to show that

$$\frac{\Gamma(z+a)}{\Gamma(z+b)}=\frac{1}{\Gamma(b-a)}\int_0^\infty u^{b-a-1}e^{-zu}f(u)du$$

where

$$f(u)=e^{-au}\left(\frac{1-e^{-u}}{u}\right)^{b-a-1}.$$