
1

PRELIMINARIES

1.1 BASICS OF PROBABILITY

1.1.1 Introduction

In this chapter, we introduce some basics of probability that will be needed in the later chapters. We also take the liberty in stating some theorems without presenting proofs and emphasize that the contents of this chapter, by no means, represent all topics of probability that deserve a detailed discussion.

A *chance* or *random experiment* is an experiment whose outcomes or results of its performance are uncertain. A set of outcomes is called an *event*. The set of all possible outcomes is referred to as the *sample space*, denoted by Ω . Thus, an *event* is a subset of the sample space and an element of the sample space is a *sample point*.

Two events A_1 and A_2 are referred to as *mutually exclusive* if their intersection is empty. A set $\{A_1, A_2, \dots, A_n\}$ is called a *partition* of Ω if the events A_1, A_2, \dots, A_n are mutually exclusive, such that $A_1 \cup A_2 \cup \dots \cup A_n = \Omega$.

Probability of an event E , denoted by $P(E)$, indicates a number between 0 and 1 (inclusive), describing the likelihood of occurrence of the event E . There are two particular events: an event with probability 1 (referred to as *almost sure event*) and another with probability 0 (referred to as *null* or *impossible event*).

For a finite sample space with n elements, if all outcomes have the same chance to occur, each member is assigned a probability of $1/n$ and the sample space is called *equiprobable*. In case of an infinite sample space, elements are with *uniform measure*.

If a chance experiment is repeated, the chance of occurrence of an outcome is the ratio of the number of occurrences of the outcome to the total number of repetitions. Hence, for a sample space with n *equiprobable points*, the *probability of an event*

with k points is k/n , referred to as *relative frequency*, and it is an approximation of the probability of the event. In other words, for a sample space with equiprobable sample points, if E is an event with k points, the *probability of the event E* is given by

$$P(E) = \frac{\text{number of ways the event } E \text{ occurs}}{\text{total number of ways all outcomes could occur}}. \quad (1.1)$$

The number of elements of the event E is referred to as the “*size*” of E . Thus, *probability of the event E* may be defined as

$$P(E) = \frac{\text{size of } E}{\text{size of } \Omega}. \quad (1.2)$$

The triplet $(\Omega, \mathfrak{B}, P)$ is called the *probability space* associated with the random experiment, where

- (a) Ω is the sample space, that is, the set of all outcomes of the random experiment.
- (b) \mathfrak{B} is the set function containing all possible events drawn from Ω , which has the structure of the σ field. This means that \mathfrak{B} satisfies the following conditions:
 - (i) the empty set is in \mathfrak{B} ,
 - (ii) if $E \in \mathfrak{B}$, then the complement of E is also in \mathfrak{B} , and
 - (iii) if $E_1, E_2, \dots \in \mathfrak{B}$, then $\bigcup_{j=1}^{\infty} E_j \in \mathfrak{B}$.
- (c) P is the probability (measure) of an event. In fact, P is a function that associates a number $P(E)$ for each element E of \mathfrak{B} with the following properties (called *axioms of probability*):

Axiom 1 $0 \leq P(E) \leq 1$, for each event E in \mathfrak{B} .

Axiom 2 $P(\Omega) = 1$.

Axiom 3 For any sequence E_1, E_2, \dots of *mutually exclusive* events (disjoint sets, that is, $E_i \cap E_j = \phi$ if $i \neq j$) in \mathfrak{B} ,

$$P\left\{\bigcup_{i=1}^{\infty} E_i\right\} = \sum_{i=1}^{\infty} P\{E_i\}.$$

1.1.2 Conditional Probability

For the probability space $(\Omega, \mathfrak{B}, P)$, let B be an event (that is, $B \in \mathfrak{B}$) and $P(B) > 0$. Then, given the probability of an event A , the *conditional probability* of B , denoted by $P(A|B)$, defined on \mathfrak{B} , is given by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \text{ for any event } A \in \mathfrak{B}, \text{ and } P\{B\} \neq 0. \quad (1.3)$$

If $P(B) = 0$, then $P(A|B)$ is not defined.

It should be noted that conditional probability exhibits properties similar to ordinary probability, but restricted to a smaller space.

One of the concepts often needed is the “independence” of events. We offer the definition for two events that can be easily expanded. Hence, we have the following:

Two events A and B are **independent** if and only if

$$P(A \cap B) = P(A)P(B), \quad (1.4)$$

In other words, occurrence of one does not affect the chance of occurrence of the other. Relation (1.4) can be expanded for an arbitrary number of events. Hence, the arbitrary family of events $\{E_i, i \in \mathbb{N}\}$, where \mathbb{N} is the set of natural numbers, is independent if

$$P(E_{i_1} \cap \cdots \cap E_{i_n}) = P(E_{i_1}) \cdots P(E_{i_n}), \quad (1.5)$$

for every finite subset of indices $\{i_1, \dots, i_n\} \subset \mathbb{N}$.

As a consequence of Equation (1.4), it can be easily proved that if two events A and B are independent and $P(B) > 0$ then

$$P(A | B) = P(A); \quad (1.6)$$

and conversely, if $P(B) > 0$ and Equation (1.6) holds to be true, then A and B are independent.

As another application of independence, the following is called the **multiplicative law**. Although we offer the definition for only two events, it can be expanded for any finite number of events. Thus, for any two events A and B with conditional probability $P(B | A)$ or $P(A | B)P(A)$ and as long as $P(A)$ or $P(B)$ is nonnegative, we have

$$P(A \cap B) = P(B | A)P(A) = P(A | B)P(B). \quad (1.7)$$

Another property of conditional probability that can be easily verified called **the law of total probability** or **total probability theorem** is that if E_1, E_2, \dots, E_n is a partition of the sample space Ω , that is, n mutually exclusive events whose sum is unity, then for any given arbitrary event E , we have

$$P(E) = \sum_{i=1}^n P(E_i)P(E | E_i). \quad (1.8)$$

Using Equation (1.8) and the conditional probability, a very important theorem called the **Bayes' theorem** can be proved. It can be stated as follows: If E is an event and E_1, E_2, \dots, E_n a partition of the sample space Ω , then

$$P(E_i | E) = \frac{P(E_i)P(E | E_i)}{\sum_{j=1}^n P(E_j)P(E | E_j)}. \quad (1.9)$$

Relation (1.9) is referred to as **Bayes' formula**. The conditional probability $P(E_i | E)$ is called the **posterior probability**. The original probability of E_i is called the **prior probability** of E_i .

1.2 DISCRETE RANDOM VARIABLES AND DISTRIBUTIONS

Although sample points are numerical in many practical problems, there are nonnumerical in many others. For practical purposes, a numerical sample space is more desirable. The tool to quantify a sample space to a numerical one is the *random variable*.

Before defining a random variable, we define a countable set. A set is called *countable* if it has the same number of elements (cardinality) as some subset of the set of natural numbers $\mathbb{N} = \{1, 2, \dots\}$. In case of a finite subset of \mathbb{N} , the set is sometimes referred to as *finitely countable* or *at most countable*. For the case of the same cardinality as \mathbb{N} , it is sometimes referred to as *countably infinite*. A set that is not countable is called *uncountable* or *denumerable*. Throughout the book, we may use any of the terms as may be appropriate. Thus for us, the set of natural numbers $\mathbb{N} = \{1, 2, \dots\}$, the set of natural numbers including zero $\mathbb{N}_0 = \{0, 1, 2, \dots\}$, the set of integers $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, 3, \dots\}$, and the set of rational numbers (reactions) \mathbb{Q} are all countable or infinitely countable sets. However, $\{1, 2, 3\}$ is at most countable or finitely countable. We note that all infinitely countable sets are of the same size, infinite. However, the set of real numbers \mathbb{R} and an interval $[0, 1]$ are not countable. The latter was first proved by George Cantor using the *diagonal argument* method.

Thus, a random variable is defined as a function (mapping) that assigns a numerical value (or a set of values) to a sample point. Hence, If X is a random variable, it assigns a value $X(\omega)$ to each outcome ω in Ω . If the function X is on an at most countable sample space into the set of real numbers, \mathbb{R} , the random variable X is called a *discrete random variable*. Thus, a random variable is discrete if it takes at most countably many values. In other words, if there is a finitely countable set of real numbers, say A , such that $P\{X \in A\} = 1$.

For the sake of convenience, we may, allow a discrete random variable to assume positive infinity, as one of its values such as waiting time for an event to occur for the first time. This is because the event may never occur.

We leave it as an exercise for the reader to prove the following properties of discrete random variables. If X and Y are two discrete random variables, then $X \pm Y$, XY , and X/Y are also random variables, for the last one, provided Y is nonzero.

Probability distribution of a discrete random variable indicates the assignment of probabilities over the entire values of a random variable.

It is important to note that probabilities are nonnegative real numbers and total assignment of probabilities must be 1. Hence, these two simple properties establish two conditions for a function to be a probability distribution function.

Let the discrete random variable X be defined on a sample space Ω with a typical element x . Then, the *probability mass function (pmf) of X* , denoted by p_x , is defined as $p_x \equiv P(X = x)$. If no specific value is given, we denote it by p_X . The pmf p_X is sometimes described by a table or matrix. For instance, if X is a random variable with the general value x and specific values x_1, x_2, x_3, \dots , that are assigned probabilities p_1, p_2, p_3, \dots , then the pmf can be written as in Table 1.1.

It is important to note that according to the second axiom of probability, $\sum_x p_x = 1$, where x varies over all possible values of X .

When two random variables X and Y have the same distribution, say p_X and p_Y , we say X and Y as *equally distributed*.

TABLE 1.1 pmf

X	x_1	x_2	x_3	\dots
$P_x = P(X = x)$	p_1	p_2	p_3	\dots

Example 1.2.1 Bernoulli Distribution The simplest chance experiment (that is, an experiment whose outcomes are determined by chance, that is also called a *trial*) is an experiment with only two outcomes. Such an experiment is called a **Bernoulli Trial**. Independent repetitions of such trials are referred to as **Bernoulli Trials**. The random variable representing a Bernoulli trial is called a **Bernoulli random variable**. Thus, the sample space for a Bernoulli random variable has two sample points, referred to as **success** and **failure**. Denoting the probability of a success by p , $0 \leq p \leq 1$, the probability of failure will be $q = 1 - p$. Thus, if X is a Bernoulli random variable with values 1 and 0, corresponding to success and failure, respectively, then distribution of X is given by

$$P(X = 1) = p; \quad P(X = 0) = 1 - p, \tag{1.10}$$

with $p + (1 - p) = 1$.

Relation (1.10) can also be expressed as

$$P(X = k) = p^k(1 - p)^{1-k}, \quad k = 0, 1. \tag{1.11}$$

In order to determine whether Equation (1.11) defines a distribution function, in this case, the **Bernoulli probability distribution function**, we note that

$$p^k(1 - p)^{1-k} > 0 \text{ and } \sum_{k=0}^1 p^k(1 - p)^{1-k} = p + (1 - p) = 1.$$

Example 1.2.2 The Indicator Function For an event A in the probability space, the indicator function of A is defined as the random variable

$$I_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A. \end{cases} \tag{1.12}$$

Hence, for every $\omega \in \Omega$, $I_\Omega(\omega) = 1$ and $I_\phi(\omega) = 0$. In other words, distribution function of an indicator function is a Bernoulli distribution with parameter $p = P(A)$. Because of this fact, sometimes, a Bernoulli random variable is called the **indicator random variable**. The random variable X , in this case, is an indicator of the event $\{X = 1\}$. This is because, when $X = 1$, ω is in the event A , the first part of Equation (1.12). For example, suppose two dice are rolled. Let X_1 and X_2 be random variables representing the numbers shown on the first and second rolling, respectively. Thus, the sample space Ω for each of these discrete random variables is $\Omega = \{1, 2, 3, 4, 5, 6\}$, and for both will be the cross-product $\Omega \times \Omega$, containing 36 elements, which are ordered pairs. In other words, $(k, l) \in \Omega \times \Omega$, $k, l = 1, 2, 3, 4, 5, 6$.

Suppose that we are interested in the sum of X_1 and X_2 be at least 7. Thus, we have $Y = X_1 + X_2$. Let the sum of interest be denoted by Z . That is, $Z = I_{Y \geq 7}$, or

$$Z = \begin{cases} 1, & \text{if } Y \geq 7, \\ 0, & \text{if } Y < 7. \end{cases} \quad (1.13)$$

Thus, we have to compute $P(Z = 1)$. Now, for the sum to be at least 7, we have to achieve the following ordered pairs: (1,6), (2,5), (3,4), (4,3), (5,2), and (6,1). That is, the sample space of Z , which is a subset of $\Omega \times \Omega$ contains (1,6), (2,5), (3,4), (4,3), (5,2) and (6,1). Probability of achieving each of these pairs is $1/36$. Thus, $P(Z = 1) = 1/6$.

Example 1.2.3 Binomial Distribution Consider Example 1.2.1. Let us assume that a Bernoulli trial is independently repeated n times and also that the random variable X represents the number of successes occurred in the n trials. Obviously, the sample space, in this case, will have 2^n sample points. The number of ways k success can occur in n trials is $\binom{n}{k}$. Requiring k success in n trials leaves $n - k$ failures. With the trials being independent, the probability of a succession of successes and failures will be the product $p^k(1 - p)^{n-k}$. Thus, the pmf of X , denoted by $B_k \equiv b(k; n, p)$, and called the *binomial random variable* is given by

$$B_k \equiv b(k; n, p) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, 2, \dots, n. \quad (1.14)$$

Relation (1.14) is, indeed, a probability distribution function, called the *binomial distribution function* with *parameters* n and p . We leave the proof as an exercise to the reader.

Example 1.2.4 Poisson Distribution The random variable X with probability distribution function (or pmf)

$$p_x = P(X = k) = \frac{e^{-\lambda} \lambda^k}{k!}, \quad k = 0, 1, 2, \dots, \quad (1.15)$$

where λ is a constant, is called a *Poisson random variable* and Equation (1.15) is called *Poisson distribution function (or Poisson pmf)* with *parameter* λ . We leave it as an exercise to show that Equation (1.15) is, actually, a distribution function.

The *discrete joint pmf* of the random variables X_1, \dots, X_n , with real values x_1, x_2, \dots, x_n , denoted by $p_{x_1 x_2 \dots x_n}$, is defined by

$$p_{x_1 x_2 \dots x_n} \equiv P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n), \quad (1.16)$$

where $p_{x_1 x_2 \dots x_n}$ is nonnegative and $\sum_{X_1} \sum_{X_2} \dots \sum_{X_n} p_{x_1 x_2 \dots x_n} = 1$.

From Equation (1.16), each individual probability, $p_{x_i} \equiv P\{X_i = x_i\}$, $i = 1, 2, \dots, n$, is called a **marginal mass function**, that is,

$$p_{x_i} = \sum_{X_1} \sum_{X_2} \cdots \sum_{X_{i-1}} \sum_{X_{i+1}} \cdots \sum_{X_n} p_{x_1 x_2 \cdots x_n}, \quad i = 1, 2, \dots, n, \tag{1.17}$$

where the summation is over all possible n -tuples with the i th component held fixed at a specific value, say x_i .

We list two properties of a random variable.

Property 1.2.1 We consider two random variables X and Y , with their respective pmfs p_X and p_Y and their joint mass function is denoted by $p_{X,Y}$. Then, the **conditional mass function of X given Y** is defined by

$$p_{X|Y} = P(X = x | Y = y) = \frac{P(X = x \text{ and } Y = y)}{P(Y = y)} = \frac{p_{X,Y}}{p_Y}, \tag{1.18}$$

provided that $p_Y > 0$.

As observed in defining the conditional probability function (1.18), random variables can be substituted for events. Thus, Equation (1.18) produces a different probability distribution for each value of y .

The pmf of the **conditional random variable $X | Y = y$** (read as “ X given $Y = y$ ”) is given by

$$P\{(X | Y = y) = x\} = P(X = x | Y = y), \tag{1.19}$$

for values x of X .

Example 1.2.5 Conditional pmf Let X and Y be jointly distributed as shown in Table 1.2.

TABLE 1.2 $p_{X,Y} = P(X = x \text{ and } Y = y)$

Y	X			$p_y = P(Y = y)$ Marginal pmf of Y
	2	4	6	
0	0.10	0.03	0.04	$p_0 = 0.17$
1	0.01	0.16	0.07	$p_1 = 0.24$
2	0.15	0.06	0.00	$p_2 = 0.21$
3	0.05	0.09	0.24	$p_3 = 0.38$
$p_x = P(X = x)$ Marginal pmf of X	$p_2 = 0.31$	$p_4 = 0.34$	$p_6 = 0.35$	1

From Table 1.1, $p_1 = P(Y = 1) = 0.24$. Then, the conditional mass function of X given $Y = 1$ is

$$p_{X|Y=1}(x) = P(X = x | Y = 1) = \begin{cases} \frac{0.01}{0.24}, & x = 2, \\ \frac{0.16}{0.24}, & x = 4, \\ \frac{0.07}{0.24}, & x = 6. \end{cases} \quad (1.20)$$

Keeping Equation (1.5) for events in mind, we leave it as an exercise to prove that two random variables X and Y , with their respective pmfs p_X and p_Y , and their joint mass function $p_{X,Y}$, are independent if and only if

$$p_{XY} = p_X p_Y. \quad (1.21)$$

Let X , Y , and Z be discrete random variables. Then, X and Y are said to be **conditionally independent**, given Z , if

$$P(X \leq x, Y \leq y | Z = z) = P(X \leq x | Z = z)P(Y \leq y | Z = z), \quad (1.22)$$

for all x , y , and z such that $P(Z = z) > 0$. As a consequence of Equation (1.22), if X and Y are conditionally independent, given Z , then

$$P(X = x | Y = y, Z = z) = P(X = x | Z = z), \quad (1.23)$$

for all x , y , and z such that $P(Z = z) > 0$.

Property 1.2.2 Functions of a Random Variable Let X be a discrete random variable with pmf, denoted by $F_X(x)$, where x is a real number representing the values of X . Let y be a real-valued function of the real variable x , denoted by $y = g(x)$. Hence, $Y = g(X)$ transforms the random variable X into the random variable Y , whose values are denoted by y . Thus, as X maps a sample point s to a real number, so does $Y = g(X(s))$. This indicates that cumulative distribution function (cdf) of Y depends on $g(x)$ and cdf of X .

We note that the domain of g should contain the range of X . Hence, pmf of Y can be expressed as

$$F_Y(y) = P(Y \leq y) = P[g(X) \leq y] = P[X \leq I_y], \quad (1.24)$$

where I_y is the **indicator function**.

1.3 DISCRETE MOMENTS

For k integers n_1, n_2, \dots, n_k , the **arithmetic average**, denoted by \bar{n} , is defined as the sum of integers divided by k . In other words, the arithmetic average of the k integers

is obtained by multiplying each number by $1/k$ and add, that is,

$$\bar{n} = n_1 \cdot \frac{1}{k} + n_2 \cdot \frac{1}{k} + \cdots + n_k \cdot \frac{1}{k}. \quad (1.25)$$

Similarly, let X be a random variable with values x_1, x_2, \dots , and x_k with probabilities p_1, p_2, \dots , and p_k , respectively. Then, the **mean** (or **weighted average** or **expected value** or **expectation** or **mathematical expectation**) of X , denoted by $E(X)$, is defined as

$$E(X) = \sum_{i=1}^k x_i p_i. \quad (1.26)$$

Note that each $p_i, i = 1, 2, \dots, k$, is the weight for each value x_1, x_2, \dots , and x_k , respectively.

Example 1.3.1 The Indicator Function We leave it as an exercise to prove that

$$E(I_A) = P(A). \quad (1.27)$$

Relation (1.26) may be expanded for a random variable with infinite values. In other words, let X be a random variable with values x_0, x_1, x_2, \dots , and respective probabilities p_0, p_1, p_2, \dots , that is $P(X = x_i) = p_i, i \in \mathbb{N}_0$. Then, the **mean** (or **weighted average** or **expected value** or **expectation** or **mathematical expectation**) of X is defined as

$$E(X) = \sum_{i=0}^{\infty} x_i p_i, \quad (1.28)$$

provided that the infinite series in Equation (1.28) converges absolutely; otherwise, the series does not exist, and hence, the expected value of X does not exist.

The following are some of the properties of Equation (1.28):

1. The only way Equation (1.28) would not exist is when

$$\lim_{n \rightarrow \infty} \sum_{i=0}^n i p_i = +\infty. \quad (1.29)$$

2. Let X and Y be random variables and a, b , and c constants, then

$$E(aX + bY + c) = aE(X) + bE(Y) + c. \quad (1.30)$$

3. Let X and Y be two independent random variables with marginal pmf p_X and p_Y , respectively. Assume $E(X)$ and $E(Y)$ exist. Then,

$$E(XY) = E(X)E(Y). \quad (1.31)$$

Relation (1.31) can be expanded for a finite number of independent random variables.

4. For random variables X and Y ,

$$E(X | Y = y) = \sum_x xP(X = x | Y = y), \quad (1.32)$$

where x and y are values of X and Y , respectively.

It is important to note that Equation (1.32) is a function of y , which implies that $E(X | Y)$ is a random variable in its own right.

Example 1.3.2 Conditional Expected Value Consider Example 1.2.5 and Table 1.2. In Example 1.2.5, we found the conditional pmf of X given $Y = 1$. Now we want to find the expected value of this conditional variable. From Table 1.2 and Equation (1.20), we have

$$E(X | Y = 1) = 2 \cdot \frac{0.01}{0.24} + 4 \cdot \frac{0.16}{0.24} + 6 \cdot \frac{0.07}{0.24} = \frac{1.08}{0.24} = 4.50.$$

5. For two random variables X and Y , we have

$$E[E(X | Y)] = E(X). \quad (1.33)$$

6. For a random variable X , with x representing its values, the *n th moment*, denoted by M^n , is defined as

$$E_n = E(X^n) = \sum_x x^n p_X, \quad (1.34)$$

where n is a nonnegative integer.

- (a) If $n = 1$, then Equation (1.34) reduces to the expected value or the first moment of X . With this note in mind, the variance of the random variable X , denoted by $\text{Var}(X)$, can be defined as

$$\text{Var}(X) = \sum_x [x - E(X)]^2 p_X, \quad (1.35)$$

where x represents values of X . The following are some properties of variance (relation (1.35)):

- (b) Let X and Y be random variables and a , b , and c constants, then

$$\text{Var}(aX \pm bY + c) = a^2 \text{Var}(X) + b^2 \text{Var}(Y). \quad (1.36)$$

- (c) For $n = 2$, then Equation (1.34) reduces to the second moment and that in turn leads to the variance of X , that is

$$\text{Var}(X) = E_2 - E_1^2. \quad (1.37)$$

7. **Central Moment.** The expected value of the random variable of X is denoted by μ , that is, $\mu = E(X)$. The n th moment of the random variable $X - \mu$, that is,

$E[(X - \mu)^n]$, is called the **central moment** of X . The random variable $X - \mu$ measures the **deviation of X from its expectation**. This deviation can be positive or negative depending on the values of $X - \mu$. Hence, its absolute value gives an **absolute measure of deviation** of the random variable X from its mean μ . Yet a better measure, the **mean square deviation**, that is, $E[(X - \mu)^2]$, called the **second central moment of X** , is, indeed, the variance of X . Thus, the variance measures the **average deviation or dispersion** of the random variable from its mean. As the variance measures the deviation from the mean by squares, in order to adjust for this squaring process, the square root is used. The positive square root of the variance of a random variable X is called the **standard deviation**.

8. **Generating Function.** Let X be a discrete random variable with nonnegative integer values with p_X as the pmf of X . The **probability generating function (pgf)** of X , denoted by $G(z)$, where z is a complex number, that was defined in Chapter 1, can be redefined as

$$G(z) \equiv E(z^X) = \sum_{x=0}^{\infty} p_X z^x, \quad (1.38)$$

where the power series in Equation (1.38) converges absolutely at least for all z , such that $|z| \leq 1$. The idea of applying generating function to probability is intended to encapsulate all the information about the random variable.

- When X takes k as its value, z^X takes z^k as its value.
- The power series in Equation (1.38) follows all the rules of convergence of power series with nonnegative coefficients.
- The product of two generating functions $G(z)$ and $H(z)$ is given by

$$G(z)H(z) = \left(\sum_{i=0}^{\infty} a_i z^i \right) \left(\sum_{j=0}^{\infty} b_j z^j \right) = \sum_{k=0}^{\infty} c_k z^k, \quad (1.39)$$

where

$$c_k = \sum_{i=0}^k a_i b_{k-i}. \quad (1.40)$$

The sequence $\{c_k\}$ defined in Equation (1.40) is called the **convolution** of the two sequences $\{a_i\}$ and $\{b_j\}$. In general, the convolution initiates two independent random variables X and Y with pmfs p_X and p_Y , respectively. Here it is defined as follows. Let X and Y be two independent discrete random variables with pmfs of p_X and p_Y , respectively. Then, the convolution of p_X and p_Y is a pmf $p_Z = p_X * p_Y$ such that

$$p_l = \sum_k p_k \cdot p_{l-k}, \quad l = \dots, -2, -1, 0, 1, 2, \dots \quad (1.41)$$

The pmf defined in Equation (1.41) is distribution function of the random variable $Z = X + Y$.

- (d) For a random variable X that takes $k, k = 0, 1, \dots$, as its values, pmf of X is recovered by the derivatives of the generating function as follows:

$$p_k = P(X = k) = \frac{G^{(k)}(0)}{k!}, k = 0, 1, 2, \dots \quad (1.42)$$

- (e) If p_X is the pmf of a random variable X , then

$$E(1) = G(1^-) = \lim_{z \rightarrow 1} G(z) = \sum_{x=0}^{\infty} p_x = 1. \quad (1.43)$$

- (f) If $z = 1$, then

$$G(1) = \sum_{n=0}^{\infty} p_n = 1. \quad (1.44)$$

- (g) For a random variable X ,

$$E(X) = G'(1^-) \quad (1.45)$$

$$\text{Var}(X) = G''(1^-) + G'(1^-) - [G'(1^-)]^2. \quad (1.46)$$

- (h) The *n th factorial moment*, denoted by $E[(X)_n]$, is defined as

$$E[(X)_n] = E[X(X-1) \cdots (X-n+1)], \quad (1.47)$$

and for a random variable X , it is given by

$$E[(X)_n] = E\left(\frac{X!}{(X-n)!}\right) = G^{(n)}(1^-), \quad (1.48)$$

where

$$(X)_n = X(X-1) \cdots (X-n+1), \quad (1.49)$$

All the moments of X can be obtained from Equation (1.48). For instance, if $n = 1$, then we obtain Equation (1.43). If $n = 2$, we obtain $E(X(X-1)) = E(X^2) - E(X)$, which leads to the variance given in Equation (1.44).

9. **The Moment generating function** of a random variable X , denoted by $M_X(t)$, with the pmf p_X is defined as

$$M_X(t) = E(e^{tX}) = G_X(e^t) = \sum_x e^{tx} p_x, \quad (1.50)$$

where $G_X(t)$ is the generating function of X and t is a nonnegative real number.

It is important to note the following points:

- It is possible that t is a complex number with nonnegative real part.
- The moment generating function generates all the moments of X , as did the generating function.

(c) If $t = 0$, then

$$M_X(0) = \sum_{x=0}^{\infty} p_x = 1. \quad (1.51)$$

Recall that for generating function, we had relation (1.42), that is, if $z = 1$, then $G(1) = \sum_{x=0}^{\infty} p_x = 1$.

(d) The n th moment of the random variable X can be obtained from the moment generating function as

$$E(X^n) = \sum_x x^n p_x = \lim_{t \rightarrow 0} \frac{d^n M_X(t)}{dt^n}. \quad (1.52)$$

Recall that for generating function, we had a similar relation (1.49), that is, to obtain the n th factorial moment of a random variable X , we need to evaluate the n th derivative of the generating function at 1. Hence, for higher moments than the second, the moment generation yields direct results.

(e) For computational purposes, it is more convenient to use the following relations:

$$E(X) = \lim_{t \rightarrow 0} \frac{d[\ln M_X(t)]}{dt}. \quad (1.53)$$

and

$$\text{Var}(X) = \lim_{t \rightarrow 0} \frac{d^2[\ln M_X(t)]}{dt^2}. \quad (1.54)$$

1.4 CONTINUOUS RANDOM VARIABLES, DENSITY, AND CUMULATIVE DISTRIBUTION FUNCTIONS

A sample space is called *continuous* if the outcomes of a chance experiment can assume real values, that is, Ω is the entire real number line \mathbb{R} or a part of it. However, in practice, a smaller set of subsets can be taken to contain all events of our interest. The smallest such set, denoted by \mathcal{B} , is referred to as the **Borel set**. For instance, the set of all rational numbers in the interval $[0, 1]$ is a Borel set for $[0, 1]$. The Borel set leads to a new **probability space** (Ω, \mathcal{B}, P) , where P is the probability of events in the Borel set.

We define the function $f_X(x)$ on the set of real numbers, \mathbb{R} , such that $f_X(x) \geq 0$, for all real x , and $\int_{-\infty}^{\infty} f_X(x) dx = 1$. Then, $f_X(x)$ is called a **continuous probability density function (pdf) on \mathbb{R}** . Let Ω be a continuous sample space. A random variable X , which takes its values for such Ω , is called a **continuous random variable**. In other words, a continuous random variable maps Ω to a subset of the real line, that is, X is a real-valued function defined on Ω such that $X : \Omega \rightarrow \mathbb{R} \subset \mathbb{R}$. In terms of pdf, $f_X(x)$, we have the following properties:

$$P(a \leq X \leq b) = \int_a^b f_X(x) dx, \quad \text{for any interval } [a, b], \quad (1.55)$$

$$P(X = a) = \int_a^a f_X(x)dx = 0, \quad \text{for any fixed real } a, \quad (1.56)$$

$$P(X < b) = \int_{-\infty}^b f_X(x)dx. \quad (1.57)$$

Suppose X is a continuous random variable defined on a sample space Ω with a probability density function $f_X(x)$. Then, the **cdf** of X , denoted by $F_X(x)$, is defined as

$$F_X(x) = P(X \leq x) = \int_{-\infty}^x f_X(t)dt. \quad (1.58)$$

As it can be seen, if the density function $f_X(x)$ is continuous, then $f_X(x) = \frac{dF_X(x)}{dx}$.

We note that the distribution $F_X(x)$ is nondecreasing, right continuous, and exhibits the following properties:

$$\lim_{x \rightarrow -\infty} F_X(x) = 0 \text{ and } \lim_{x \rightarrow \infty} F_X(x) = 1. \quad (1.59)$$

If there is no danger of confusion, we will suppress the subscript X from $f_X(x)$ and $F_X(x)$. Equation (1.55) can be rewritten using Equation (1.55) as

$$P(a \leq X \leq b) = F(b) - F(a). \quad (1.60)$$

It is important to note that $0 \leq F(x) \leq 1$. Thus, for an event, the longer the elapse time, the greater would be the probability of occurrence of the event. If $1 - F(x)$ is denoted by $S(x)$, that is, $S(x) = P(X > x)$, then $S(x)$ is called the **survival probability**. In other words, if x represents the time, then $S(x)$ is the probability that the event will not happen until the time x . The ratio of density function and survival probability is referred to as **hazard or failure rate**. Denoting the hazard rate by $h(x)$ the hazard rate will be defined as

$$h(x) = \frac{f(x)}{S(x)}. \quad (1.61)$$

Similar to the discrete random variable, we define the convolution of two density functions. Let X and Y be two continuous random variables with density functions $f(x)$ and $g(y)$, respectively, for all real numbers x and y . Then, the **convolution** of functions f and g , denoted by $h \equiv f * g$, is defined by

$$\begin{aligned} h(z) &= (f * g)(z) = \int_{-\infty}^{\infty} f(z-y)g(y)dy \\ &= \int_{-\infty}^{\infty} g(z-x)f(x)dx. \end{aligned} \quad (1.62)$$

Furthermore, we note that Equation (1.62) is the pdf of sum of two independent random variables, $Z = X + Y$.

Example 1.4.1 Uniform Random Variable Let X be a continuous random variable with pdf as follows:

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{otherwise.} \end{cases} \quad (1.63)$$

Relation (1.63) is the pdf of the *uniform distribution* over the interval $[a, b]$. The random variable with this pdf is called a *uniform random variable*.

Example 1.4.2 Exponential Random Variable A continuous random variable X with pdf

$$f_X(x) = \begin{cases} \mu e^{-\mu x}, & x \geq 0, \\ 0, & \text{otherwise,} \end{cases} \quad (1.64)$$

is referred to as *negative exponential* (or *exponential*) *random variable*, where μ is a parameter of the *pdf*.

The following is an important property of exponential distribution:

Theorem 1.4.1 Memoryless Property A random variable $X : \Omega \rightarrow (0, \infty)$ has an exponential distribution if and only if it exhibits the following property, referred to as the *memoryless* (or *forgetful*) *property*:

$$P\{X > s + t | X > s\} = P\{X > t\}, \quad \text{for all } s, t \geq 0. \quad (1.65)$$

Proof: Let X have an exponential distribution with parameter $\mu > 0$. Then, we have

$$\begin{aligned} P\{X > s + t | X > s\} &= \frac{P\{X > s + t \text{ and } T > t\}}{P\{X > s\}} = \frac{P\{X > s + t\}}{P\{X > s\}} \\ &= \frac{e^{-\mu(s+t)}}{e^{-\mu s}} = e^{-\mu t} = P\{X > t\}. \end{aligned}$$

The converse follows because the function $g(t) = P\{X > t\}$ satisfies

$$g(s + t) = g(s)g(t), \quad \text{for all } s, t \geq 0 \text{ and } g(0) = 1.$$

Example 1.4.3 Hyperexponential Random Variable Let $Y_i, i = 1, 2, \dots, n$, be n independent exponentially distributed random variables with pdf $f_{Y_i}, i = 1, 2, \dots, n$, and parameters $\mu_i, i = 1, 2, \dots, n$, with $\mu_i \neq \mu_j$, for $i \neq j$. Furthermore, let $p_i, i = 1, 2, \dots, n$, be positive constants such that $\sum_{i=1}^n p_i = 1$. Then, a continuous random variable X is called a *hyperexponential random variable* if its mixed pdf has the following form:

$$f_X(x) = \sum_{i=1}^n p_i f_{Y_i}(x), \quad (1.66)$$

with $2n$ parameters μ_i and $p_i, i = 1, 2, \dots, n$.

Example 1.4.4 Gamma Random Variable As a generalization of Equation (1.63), we consider a continuous random variable with two-parameter pdf as follows:

$$f_X(x; \mu, t) = \begin{cases} \frac{\mu^t x^{t-1}}{\Gamma(t)} e^{-\mu x}, & x > 0, \mu, t > 0 \\ 0, & \text{otherwise,} \end{cases} \quad (1.67)$$

where

$$\Gamma(t) = \int_0^{\infty} x^{t-1} e^{-x} dx \quad (1.68)$$

is the **gamma function** and t is a positive real number. In Equation (1.67) if $t = 1$, we will obtain the exponential density function with parameter μ defined by Equation (1.64).

Example 1.4.5 Erlang Random Variable In Equation (1.67) if t is a nonnegative integer, say n , it reduces to pdf of another random variable, called the **Erlang pdf of order r** . In other words, a random variable X has **Erlang** pdf with shape parameter r and rate parameter μ , denoted by $e(x; r, \mu)$, which is defined as

$$e(x; r, \mu) = \frac{\mu^r x^{r-1} e^{-\mu x}}{(r-1)!}, \quad \text{for } x, \mu \geq 0. \quad (1.69)$$

In Equation (1.69), if $r = 1$, then we will have the exponential pdf. In addition to the shape parameter r ; if rather than the rate parameter μ , a scale parameter, $\beta = 1/\mu$, is of interest, then an alternative, but equivalent, pdf is defined as follows:

$$e(x; r, \beta) = \frac{x^{r-1} e^{-x/\beta}}{\beta^r (r-1)!}, \quad \text{for } x, \beta \geq 0. \quad (1.70)$$

We note that when $\beta = 2$, the Erlang distribution simplifies to the chi-squared distribution with $2r$ degrees of freedom. Thus, Equation (1.70) can be regarded as pdf of a generalized chi-squared distribution for even numbers of degrees of freedom. We also note that because of the factorial function in the denominator of Eqs. (1.69) and (1.70), the Erlang distribution is only defined when the shape parameter r is a positive integer.

Example 1.4.6 Normal Distribution Function A continuous random variable X with mean μ , variance σ^2 , and pdf

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad (1.71)$$

is called a **normal** random variable, denoted by $X \sim N(\mu, \sigma^2)$ or $X \sim \Phi(\mu, \sigma^2)$. If $\mu = 0$ and $\sigma^2 = 1$, then X is called a **standard normal** random variable.

1.5 CONTINUOUS RANDOM VECTOR

The idea of random vector is a multidimensional generalization of the random variable. Thus, let Ω be the sample space. A **random vector** X is an n -dimensional vector, whose components are random variables, that is $X = (X_1, X_2, \dots, X_n)$. Continuous random vectors are described by the **joint pdf**. For instance, consider two continuous random variables X and Y with pdf $f_X(x)$ and $f_Y(x)$, respectively. The **joint cumulative bivariate cdf of X and Y** , denoted by $F_{X,Y}(x, y)$ or simply $F(x, y)$, is defined as

$$F(x, y) = P(X \leq x, Y \leq y). \quad (1.72)$$

The joint pdf for X and Y , denoted by $f_{X,Y}(x, y)$, is obtained as

$$f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x, y)}{\partial x \partial y}. \quad (1.73)$$

Relations (1.72) and (1.73) can be finitely extended for many random variables as in the discrete case. We note that $f_{X,Y}(x, y) \geq 0$, for every $(x, y) \in \mathbb{R}^2$. We also note that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1.$$

Finally, we note that for any region D in the xy plane, the following condition holds:

$$P\{(X, Y) \in D\} = \int \int_D f_{X,Y}(x, y) dx dy.$$

From the joint pdf, individual pdfs, called **marginal pdf**,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy \text{ and } f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx. \quad (1.74)$$

The joint bivariate cdf for X and Y is

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y). \quad (1.75)$$

Similar to the discrete case, from the joint pdf, we can define the conditional pdf. Let us consider two continuous random variables X and Y , with their respective pdfs f_X and f_Y and joint mass function $f_{X,Y}$. Then, for any y such that $f_Y(y) > 0$, the **conditional pdf of X given $Y=y$** , denoted by $f_{X|Y=y}(x)$, is defined by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad (1.76)$$

for all $x \in \mathbb{R}$. The quantity $f_{X|Y=y}(x)$ is not defined when $f_Y(y) = 0$.

Example 1.5.1 Marginal pdf An industry produces two types of productions. There are defective items in each production. Let us denote the proportion of defective items of Type 1 production by X and Type 2 production by Y . The joint pdf of X and Y , denoted by $f_{X,Y}(x, y)$ is given by

$$f_{X,Y}(x, y) = \begin{cases} 1 - x^2, & 0 \leq x \leq 1, 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

We need to determine

1. marginal probability $P(0 \leq X \leq 0.3)$,
2. marginal probability $P(0.1 \leq Y \leq 0.7)$,
3. joint probability $P(0 \leq X \leq 0.5, 0.4 \leq Y \leq 0.9)$,
4. marginal pdf f_X ,
5. marginal pdf f_Y , and
6. conditional probability $P(X \leq 0.3 | Y = 0.6)$.

The answers to the aforementioned questions are as follows:

$$1. P(0 \leq X \leq 0.3) = \int_0^{0.3} (1 - x^2) dx = 0.291.$$

$$2. P(0.1 \leq Y \leq 0.7) = \int_{0.1}^{0.7} (1 - x^2) dy = 0.6(1 - x^2).$$

$$3. P(0 \leq X \leq 0.5, 0.5 \leq Y \leq 0.9) = \int_0^{0.5} \int_{0.4}^{0.9} (1 - x^2) dy dx = 0.2292.$$

$$\begin{aligned} 4. \quad f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy = \int_{-\infty}^{\infty} (1 - x^2) I_{\{0 \leq x \leq 1, 0 \leq y \leq 1\}} dy \\ &= \int_0^1 (1 - x^2) I_{\{0 \leq x \leq 1\}} dy = (1 - x^2) I_{\{0 \leq x \leq 1\}} \\ &= \begin{cases} (1 - x^2), & \text{if } 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

$$\begin{aligned} 5. \quad f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx = \int_{-\infty}^{\infty} (1 - x^2) I_{\{0 \leq x \leq 1, 0 \leq y \leq 1\}} dx \\ &= \int_0^1 (1 - x^2) I_{\{0 \leq y \leq 1\}} dx = \left[x - \frac{x^3}{3} \right]_0^1 I_{\{0 \leq y \leq 1\}} \\ &= \begin{cases} \frac{2}{3}, & \text{if } 0 \leq y \leq 1, \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

We note that the indicator functions are used in answers (4) and (5) to cover various cases at once and, thus, make the calculations easier.

6.

$$\begin{aligned}
 P(X \leq 0.3 | Y = .6) &= \int_{-\infty}^{0.3} f_{X|Y=0.6}(x) dx \\
 &= \int_0^{0.3} \frac{1-x^2}{\frac{2}{3}} dx = \frac{3}{2} \left[x - \frac{x^3}{3} \right]_0^{0.3} = 0.4485.
 \end{aligned}$$

Referring to Equation (1.61), a distribution function $F(x)$ is called **heavy-tailed**, if $1 - F(x) = S(x) > 0, x \geq 0$ and

$$\lim_{x \rightarrow \infty} P\{X > x + y | X > x\} = \lim_{x \rightarrow \infty} \frac{S(x + y)}{S(x)} = 1, \quad \forall y \geq 0. \tag{1.77}$$

In other words, if X ever exceeds a given larger value, then it is likely to exceed any larger value than that value. The Pareto distribution function with $S(x) = X^{-\alpha}, x \geq 1$, where $\alpha > 0$ is a parameter, is a standard example. On the contrary, an exponential distribution function with parameter μ is not a heavy-tailed, because

$$\lim_{x \rightarrow \infty} \frac{S(x + y)}{S(x)} = \frac{e^{-\mu(x+y)}}{e^{-\mu x}} = e^{-\mu y}, \quad \forall x, y \geq 0, \tag{1.78}$$

which becomes 1 only for $y = 0$ and not for all values of y .

1.6 FUNCTIONS OF RANDOM VARIABLES

Let us consider a sample space Ω with a general element ω . Let X be a random variable, that is, a function relating the real number $X(\omega)$ to the outcome ω , with pdf $f_X(x)$ and cdf $F_X(x)$, if X is continuous, and p_X if X is discrete. Let $y = g(x)$ be a real-valued function of the real variable x , and $Y = g(X)$ be a transformation of the random variable X into the random variable Y . Thus, as $X(\omega)$ is a random variable, $g(X(\omega))$ is called **function of the random variable X** . Then, the question is how to find pdf $f_Y(y)$ and cdf $F_Y(y)$ of $g(X(\omega))$. We note that the domain of function g should contain the range of X .

The idea of a function of a random variable may be expanded for a finite number of variables. Thus, if $\mathbf{X} = (X_1, X_2, \dots, X_n)$ is a random vector of n random variable that associates the sample space Ω to the space \mathbb{R}^n of real n -tuples, then the joint distribution function of X , for the continuous case, is defined as

$$F_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}, \tag{1.79}$$

and pmf, for the discrete case, is defined as

$$p_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = P\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}. \tag{1.80}$$

For the continuous case, the pdf is defined as

$$f_{X_1, X_2, \dots, X_n}(x) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F_{X_1, \dots, X_n}(x_1, \dots, x_n). \quad (1.81)$$

As in the case of one variable, the marginal pdf for each random variable is obtained as follows:

$$f_{X_1}(x_1) = \int \cdots \int f_{X_1, \dots, X_n}(x_1, \dots, x_n) dx_2 \cdots dx_n. \quad (1.82)$$

For the discrete case, summation will replace the integrals.

Example 1.6.1 Sum of Random Variables Consider the random variable Z as the sum of two random variables X and Y . Hence,

$$P\{Z \leq z\} = P\{X + Y \leq z\} \quad (1.83)$$

and

$$F_Z(z) = \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_{X,Y}(x, y) dx dy. \quad (1.84)$$

Thus,

$$f_Z(z) = \frac{dF_Z(z)}{dz} = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) dx. \quad (1.85)$$

If X and Y are independent random variables, then the pdf of Z is the convolution of the two pdfs. That is,

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx. \quad (1.86)$$

If g is a function of a discrete random variable X , $g = g(X)$, then expected value of $g(X)$ can be computed as

$$E[g(X)] = \sum_k g(x_k) P(X = x_k). \quad (1.87)$$

We will discuss the continuous case in the next section.

We now define, without proof, three types of convergence in the theory of probability. The first has two versions. It is important to note that in a classical sense, when we say a sequence $\{X_n\}$ converges to X as n approaches infinity, we mean that the difference of X_n and X , in absolute value, gets closer to 0 as n becomes larger and larger.

Theorem 1.6.1 Convergence in Probability Version a: The sequence of random variables $\{X_n, n \in \mathbb{N}\}$, not necessarily all defined on the same sample space Ω , **converges in probability** to a real number c , is denoted by $X_n \xrightarrow{\text{prob}} c$, if and only if

$$\lim_{n \rightarrow \infty} P(|X_n - c| \geq \epsilon) = 0, \tag{1.88}$$

for any arbitrary positive number ϵ .

Version b: The sequence of random variables $\{X_n, n \in \mathbb{N}\}$ defined on a sample space Ω **converges in probability** to a random variable X , defined on a sample space Ω , is denoted by $X_n \xrightarrow{\text{prob}} X$, only if

$$\lim_{n \rightarrow \infty} P(|X_n - X| \geq \epsilon) = 0, \tag{1.89}$$

for any arbitrary positive number ϵ . In other words, as n increases, $P(|X_n - X| > \epsilon)$ decreases, and X is called the **probability limit** of the sequence.

It is important to note that if X is a constant c rather than a random variable, then both versions are the same. Furthermore, if each of the two sequences $\{X_n, n \in \mathbb{N}\}$ and $\{Y_n, n \in \mathbb{N}\}$ converges in probability to X and Y , respectively, with all random variables defined on the same sample space Ω , that is, $X_n \xrightarrow{\text{prob}} X$, and $Y_n \xrightarrow{\text{prob}} Y$, then, we leave it as an exercise to the readers to prove that $(X_n + Y_n) \xrightarrow{\text{prob}} (X + Y)$.

We note that the idea of convergence in probability can be extended to a sequence of random vectors.

Theorem 1.6.2 Almost Surely Convergence The sequence of random variables $\{X_n, n \in \mathbb{N}\}$, defined on the sample space Ω , **converges almost surely** (or **with probability one**) to a random variable X , denoted by $X_n \xrightarrow{\text{a.s.}} X$ (or $X_n \rightarrow X$ with probability one) if there is a set $A \subset \Omega$ such that

$$\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega), \text{ for all } \omega \in A, \tag{1.90}$$

and $P(A) = 1$, or

$$P(\{\lim_{n \rightarrow +\infty} X_n(\omega) = X(\omega), \forall \omega \in \Omega\}) = 1. \tag{1.91}$$

We leave it as an exercise that the **almost sure convergence** implies **convergence in probability**, but the converse is not true. We also note that the almost sure convergence reduces the idea of convergence of deterministic sequences. We further note that convergence almost surely is the probabilistic version of pointwise convergence in real analysis. Finally, we note that the random variables $X_n, n \geq 1$, are generally highly dependent.

Example 1.6.2 Convergence in Probability A random experiment (per day income) runs over time n . Let the income on day n be denoted by R_n . Let the income on the first n days be denoted by S_n , that is,

$$S_n = \sum_{k=1}^n R_k. \quad (1.92)$$

Finally, we denote the lifetime income by S , that is,

$$S = \sum_{k=1}^{\infty} R_k. \quad (1.93)$$

Assuming S_k is positive for all k , we have $S_n \xrightarrow{\text{a.s.}} S$.

Theorem 1.6.3 Convergence in Distribution The sequence of random variables $\{X_n, n \geq 1\}$ *converges in distribution* to a random variable X , denoted by $X_n \xrightarrow{\text{dist}} X$, if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x), \quad (1.94)$$

for every $x \in \mathbb{R}$ at which $F_X(x)$ is continuous.

We also state the following two main limit theorems in the theory of probability, without proof.

Theorem 1.6.4 Law of Large Numbers Let $\{X_n, n \geq 1\}$ be a sequence of independent, identically distributed random variables (iid), sometimes referred to as a *random sample*, such that $\mu \equiv E[|(X_1)|] = E[|(X_2)|] = \dots$ is finite. Then,

$$\bar{X}_n \equiv \frac{X_1 + \dots + X_n}{n} \xrightarrow{\text{a.s.}} \mu. \quad (1.95)$$

Theorem 1.6.4 states that the mean of the infinite sequence of iid random variables $\{X_n, n \geq 1\}$ converges to the expected value as n approaches to infinity. If the convergence is almost possible, the law is called the *strong law of large numbers*; if it is in distribution, it is called the *weak law of large numbers*. The strong and weak laws of large numbers, respectively, state that

$$P\{\lim_{n \rightarrow \infty} |\bar{X}_n - \mu| = 0\} = 1 \quad (1.96)$$

and

$$\lim_{n \rightarrow \infty} P\{|\bar{X}_n - \mu| > \epsilon\} = 0, \text{ for any } \epsilon > 0. \quad (1.97)$$

For the proof of the strong law of large number, one may refer to Grimmet and Stirzaker (2004).

We note that Equation (1.95) can be generalized replacing random variables by functions of random variables. Thus, let g be a function from reals to reals, that is, $g : \mathbb{R} \rightarrow \mathbb{R}$ such that $\mu = E[g(X_1)] = E[g(X_2)] = \dots$ exists and

$$\frac{g(X_1) + \dots + g(X_n)}{n} \rightarrow \mu = \int_{-\infty}^{\infty} g(x)f_{X_1}(x)dx, \text{ as } n \rightarrow \infty. \tag{1.98}$$

Theorem 1.6.5 Central Limit Theorem Let $\{X_n, n \geq 1\}$ be a sequence of iid random variables, such that $E[|X_1^2|] (E[|X_1^2|] = E[|X_2^2|] = \dots)$ is finite. Let $\mu \equiv E(X_1) = E(X_2) \dots$, and $\sigma^2 \equiv \text{Var}(X_1) = \text{Var}(X_2) = \dots$. Then,

$$\frac{X_1 + \dots + X_n - n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{dist}} \Phi(0, 1). \tag{1.99}$$

In other words, the left-hand side of Equation (1.99) approaches standard normal distribution when n approaches infinity.

1.7 CONTINUOUS MOMENTS

We offer definition of moments for a continuous random variable similarly to discrete random variables (Eqs. (1.26) and (1.27)). Thus, let X be a continuous random variable with pdf $f_X(x)$ and a real-valued function $g(x)$. Then, the expected value of $g(X)$ is defined as

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f_X(x)dx, \tag{1.100}$$

provided the integral exists, that is, $E[g|X]$ is finite. In particular, if $g(X) = X$, the mean μ and variance σ^2 of X are obtained as follows:

$$\mu = E(X) = \int_{-\infty}^{\infty} xf_X(x)dx, \tag{1.101}$$

and

$$\sigma^2 = \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x)dx. \tag{1.102}$$

Example 1.7.1 Expected Value of a Function of a Variable Let X be a normal random variable with $\mu = 0$, variance σ^2 , and $g(x) = e^{\tau x}$, where τ is a constant. Then, $E[g(x)]$ can be computed as follows:

$$\begin{aligned} E[g(x)] &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} e^{\tau x} e^{-\frac{x^2}{2\sigma^2}} dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{\sigma^2\tau^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(x-\sigma^2\tau)^2}{2\sigma^2}} dx \\ &= e^{\frac{\sigma^2\tau^2}{2}}. \end{aligned}$$

Example 1.7.2 Characteristic Function The characteristic function of a real-valued random variable X , denoted by $\phi_X(t)$, is defined by

$$\phi_X(t) = E[e^{itx}], \quad (1.103)$$

where i is the imaginary unit and t is a real number. If $F_X(x)$ is the cumulative distribution of X , then

$$\phi_X(t) = \int_{\mathbb{R}} e^{itx} dF_X(x), \quad (1.104)$$

where \mathbb{R} is the set of real numbers, and the integral in Equation (1.104) is of the Riemann–Stieltjes kind. All moments of X can be found using derivatives of Equation (1.103) with respect to t , evaluated at $t = 0$. For instance, we leave it to the reader as an exercise to show that for the Poisson random variable with parameter λ , its characteristic function is

$$\phi(t) = e^{\lambda(e^{it} - 1)}, \quad (1.105)$$

where λ is the mean.

Now, let $f_X(x)$ be a real-valued continuous function of a real variable about a number c . Then, the n th moment of $f_X(x)$, denoted by μ_n , is defined as

$$\mu_n = \int_{-\infty}^{\infty} (x - c)^n f_X(x) dx. \quad (1.106)$$

If c is the mean, then Equation (1.106) is referred to as the **central moments**. The n th moment of the pdf $f_X(x)$ when $c = 0$ in Equation (1.106) is the expected value of X^n and is called a **raw** or **crude** moment. In this case, the n th **moment** of the cdf of X is defined by the Riemann–Stieltjes integral as

$$\mu_n = E(X^n) = \int_{-\infty}^{\infty} x^n dF_X(x). \quad (1.107)$$

If the value of the integral in Equation (1.107) for the absolute value of X^n is infinite, then the moment does not exist. We note that if $n = 0$, then the moment of the pdf is 1. When $n = 1$ and $n = 2$, we obtain the mean μ and variance σ^2 of X using Eqs. (1.101) and (1.102), respectively.

Example 1.7.3 Hyperexponential We leave it as an exercise to the reader to show that the first two moments and variance of a hyperexponential distribution function are respectively

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \sum_{i=1}^n p_i \int_0^{\infty} x \mu_i e^{-\mu_i x} dx = \sum_{i=1}^n \frac{p_i}{\mu_i}, \quad (1.108)$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f_X(x) dx = \sum_{i=1}^n p_i \int_0^{\infty} x^2 \mu_i e^{-\mu_i x} dx = \sum_{i=1}^n \frac{2p_i}{\mu_i}, \quad (1.109)$$

and

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \left[\sum_{i=1}^n \frac{p_i}{\mu_i} \right]^2 + \sum_{i=1}^n \sum_{j=1}^n p_i p_j \left(\frac{1}{\mu_i} - \frac{1}{\mu_j} \right). \quad (1.110)$$

1.8 DIFFERENCE EQUATIONS

1.8.1 Introduction

Recently, there has been an increasing interest in the calculus of difference and differential-difference equations, particularly for solving some Markovian queueing models. Some other particular interests in the theory of difference equations are: because of the emerge of development and advancement of high-speed computers; numerous applications of difference equations to engineering, sciences (such as physics, chemistry, biology, probability, and statistics), economics, and psychology; and the mathematical theory in view of the analogy of the theory to differential equations. One main interest in difference equations versus differential equations is that the former involve discrete changes of an unknown function while the latter involve instantaneous rates of changes of an unknown function. In other words, difference equations are the discrete analogs of differential equations.

The theory and solutions of difference equations in many ways are parallel to the theory and solutions of differential equations. In fact, derivative of a function of one variable defined as a limit of a difference quotient is the main reason for many analogies between the differential calculus and the calculus of finite differences. In other words, as we may recall from differential calculus, the derivative of a given function $f(x)$ is defined as

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{\Delta[f(x)]}{h},$$

if the limit exists, where Δ denotes the ***difference operator***, that is, $\Delta f(x) = f(x+h) - f(x)$. For higher order differences, we may continue the Δ operator.

1.8.2 Basic Definitions and Properties

Definition 1.8.1 This definition consists of the following six parts:

1. Given a function F and a set S , for k in S and some positive integer n , a ***difference equation*** over a given set S is an equation of the form

$$F(k; y_k, y_{k+1}, \dots, y_{k+n}) = 0. \quad (1.111)$$

A rewording of this statement is as follows: A difference equation is the one that involves differences. We write y_k for $y(x)$ to indicate the value of y at k . The

range of values of the integer k must be stated. It should be noted that shifting indices has no bearing in the order of the difference equation. Hence,

$$F(k+r; y_{k+r}, y_{k+r+1}, \dots, y_{k+n+r}) = 0 \quad (1.112)$$

still is of order n , which is equivalent to Equation (1.111).

2. The **order** of a difference equation is the difference between the highest and lowest indices that appear in the equation.
3. A **solution** of the difference equation (Equation (1.111)) indicates a sequence $\{y_k\}$, which satisfies the difference equation Equation (1.111) for all values of k in S . The solution is referred to a **general solution** of Equation (1.111) if it involves exactly n arbitrary constants. A **particular solution** of Equation (1.111) is obtained from the general solution by assigning values to the arbitrary constants.
4. A difference equation of the form

$$f_0(k)y_{k+n} + f_1(k)y_{k+n-1} + \dots + f_{n-1}(k)y_{k+1} + f_n(k)y_k = f(k), \quad (1.113)$$

where each of f and $f_i, i = 0, 1, 2, \dots, n$, is a function of k defined for all values of $k \in S$, is called **linear** over the set S . If a difference equation is not linear, it is called **nonlinear**. The linear difference equation (1.113) is called of **order n** , if both $f_0(k)$ and $f_n(k)$ are different from zero at each point of S . In other words, the order of the difference equation (1.113) is the difference between the highest and lowest indices that appear in the equation, if and only if

$$f_0(k)f_n(k) \neq 0, \forall k \in S. \quad (1.114)$$

If $f(k)$ on the right-hand side of Equation (1.113) is zero, then the equation is called a **homogeneous linear difference equation of order n** ; otherwise, it is called **nonhomogeneous** or **inhomogeneous**. In general, an **n th order linear homogeneous difference equations with constant coefficients** is an equation of the form

$$y_{k+n} + a_1y_{k+n-1} + \dots + a_{n-1}y_{k+1} + a_ny_k = 0, \quad (1.115)$$

where a_1, a_2, \dots, a_n are n given constants with $a_n \neq 0$.

5. In addition to differences in an equation, there may be derivatives or integrals. In such cases, we refer to the equation as **differential-difference equation** or **integral-difference equation**, respectively.
6. Consider $z = f(x, y)$ as a function with two variables x and y . We write differences of z with respect to x and y . That is, we consider y as a constant and denote the difference as

$$\Delta_x f(x, y) = f(x+h, y) - f(x, y). \quad (1.116)$$

Similarly, we consider x as a constant and denote the difference as

$$\Delta_y f(x, y) = f(x, y + k) - f(x, y). \quad (1.117)$$

Relations (1.116) and (1.117) are called **partial differences** of the function $z = f(x, y)$. An equation that contains partial differences is referred to as a **partial difference equation**.

Theorem 1.8.1 Existence and Uniqueness Theorem If there are n initial conditions y_0, y_1, \dots, y_{n-1} for Equation (1.113), then corresponding to each of these conditions, Equation (1.113) has a unique solution.

Proof: The proof for this is left as an exercise for the reader.

1.9 METHODS OF SOLVING LINEAR DIFFERENCE EQUATIONS WITH CONSTANT COEFFICIENTS

In general, not every method mentioned below could be applied to solve every equation. Perhaps hence different methods have been developed.

1.9.1 Characteristic Equation Method

Homogeneous Case For the n th order linear homogeneous difference equation with constant coefficient given by Equation (1.115), the n th degree polynomial equation

$$r^n + a_1 r^{n-1} + \dots + a_{n-1} r + a_n = 0 \quad (1.118)$$

is called **characteristic equation** (or **auxiliary equation**).

Theorem 1.9.1 If $r_i, i = 1, 2, \dots, n$, is a root of the characteristic equation (Equation 1.118), then

$$(a) \quad y_k^{(i)} = r_i^k \quad (1.119)$$

is a solution of Equation (1.118).

(b) If all n roots of the characteristic equation (1.118) are distinct real numbers, then the general solution of the equation is

$$y_k = c_1 y_k^{(1)} + c_2 y_k^{(2)} + \dots + c_n y_k^{(n)}, \quad (1.120)$$

where $c_i, i = 1, 2, \dots, n$, are n arbitrary constants.

- (c) If roots r_i of the characteristic equation (1.118) are with multiplicity m_i , $i = 1, 2, \dots, l$, such that

$$\sum_{i=1}^l m_i = n, \quad (1.121)$$

then the general solution of the equation is

$$\begin{aligned} y_k = & r_1^k (c_1^{(1)} + c_2^{(1)}k + \dots + c_{m_1}^{(1)}k^{m_1-1}) + r_2^k (c_1^{(2)} + c_2^{(2)}k + \dots + c_{m_2}^{(2)}k^{m_2-1}) \\ & + \dots + r_{m_l}^k (c_1^{(l)} + c_2^{(l)}k + \dots + c_{m_l}^{(l)}k^{m_l-1}), \end{aligned} \quad (1.122)$$

where $c_i^{(j)}$, $i = 1, \dots, l$; $j = 1, 2, \dots, m_i$, are arbitrary constants.

Proof: See Mickens (1990), Second Edition, p. 124.

Theorem 1.9.2 For arbitrary constants $c_{i,j+1}$ and c_j , let

$$y(x) = \sum_{i=1}^m \left(\sum_{j=0}^{n_i-1} c_{i,j+1} x^j \right) e^{r_i x} + \sum_{j=(n_1+\dots+n_m)+1}^n c_j e^{r_j x}, \quad (1.123)$$

where $n_i \geq 1$, $i = 1, 2, \dots, m$, with $n_1 + n_2 + \dots + n_m \leq n$, the general solution of the n th order linear differential equation

$$\frac{d^n y(x)}{dx^n} + a_1 \frac{d^{n-1} y(x)}{dx^{n-1}} + \dots + a_{n-1} \frac{dy(x)}{dx} + a_n y(x). \quad (1.124)$$

In the above equation, a_i , $i = 1, 2, \dots, n$, $a_n \neq 0$, are constants. In addition, let the characteristic equation (1.118) has r_j simple roots and r_i roots with multiplicity n_i , $i = 1, 2, \dots, m$. Furthermore, let y_k be the general solution of the difference equation (1.115). Then,

$$y_k = \left. \frac{d^k y(x)}{dx^k} \right|_{x=0}, \quad (1.125)$$

and

$$y_k = \sum_{i=1}^m \left(c_{i1} + \sum_{l=1}^{n_i-1} \gamma_{i,l} k^l \right) r_i^k + \sum_{j=(n_1+\dots+n_m)+1}^n c_j r_j^k, \quad (1.126)$$

Proof: See Mickens (1990, p. 139).

Nonhomogeneous Case Consider the first-order linear nonhomogeneous difference equation

$$y_{k+1} - ay_k = b, \quad k = 0, 1, 2, \dots, \quad (1.127)$$

with initial condition $y_0 = c$, where a , b , and c are constants and $a \neq 0$. The characteristic equation of the homogeneous part, when $b = 0$,

$$y_{k+1} - ay_k = 0, \quad k = 0, 1, 2, \dots, \quad (1.128)$$

of Equation (1.127) is $r - a = 0$, with only one root $r = a$. Hence, the general solution of Equation (1.128) is

$$Y_k = Ca^k, \quad k = 0, 1, 2, \dots, \quad (1.129)$$

where C is an arbitrary constant. In order to find the particular solution of Equation (1.127), different cases for values of a and b may be considered.

1.9.2 Recursive Method

Considering the nonhomogeneous case, we rewrite Equation (1.127) as

$$y_{k+1} = ay_k + b, \quad k = 0, 1, 2, \dots \quad (1.130)$$

Now we consider different values of k , starting with 0.

$k = 0$: This value of k gives the initial value $y_0 = c$, substituting into Equation (1.130) gives

$$y_1 = ay_0 + b = ac + b. \quad (1.131)$$

$k = 1$: Substituting this value of k into Eqs. (1.130) and (1.131), we obtain

$$y_2 = ay_1 + b = a(ac + b) + b = a^2c + b(1 + a). \quad (1.132)$$

$k = 2$: Substituting $k = 2$ into Eqs. (1.130–132) results in

$$y_3 = ay_2 + b = a[a^2c + b(1 + a) + b] + b = a^3c + b(1 + a + a^2). \quad (1.133)$$

From Equation (1.131) to Equation (1.133), we can obtain a pattern that would conjecture the general case and prove it by mathematical induction. Thus, we choose the solution for the initial problem (Equation 1.127) as

$$y_k = a^k c + b(1 + a + a^2 + \dots + a^{k-1}). \quad (1.134)$$

We leave it as an exercise for the reader to prove Equation (1.134) that would be the proof of the following theorem:

Theorem 1.9.3 Existence and Uniqueness Theorem The linear first-order initial-valued differential equation given by Equation (1.127) has a unique solution given by Equation (1.134).

The following theorem will provide a special case of Equation (1.127), that is, linear first-order nonhomogeneous case. We leave its proof as an exercise to the reader.

Theorem 1.9.4 If the sequence $\{y_k, k = 0, 1, \dots\}$ is the solution of Equation (1.127) when $a = 1$, that is, the linear first-order nonhomogeneous equation

$$y_{k+1} = y_k + b, k = 0, 1, 2, \dots, y_0 = c, \quad (1.135)$$

then,

- (a) if $b < 0$, it will diverge to $-\infty$,
- (b) if $b = 0$, $\{y_k\}$ is a constant sequence, and
- (c) if $b > 0$, the solution sequence will diverge to $+\infty$.

We will provide an example later in this chapter after the Laplace transform method to show how the recursive method works.

1.9.3 Generating Function Method

Let $\{a_n\}$ be a sequence of real numbers such that $a_n = 0$, for $n < 0$. If the infinite series

$$G(z) \equiv \sum_{n=0}^{\infty} a_n z^n, \quad (1.136)$$

where z is a dummy variable, converges for $|z| < R$, where R is the radius of convergence, then $G(z)$ is called the **generating function** of the sequence $\{a_n\}$. If the sequence $\{a_n\}$ is bounded, then the series (Equation 1.136) converges for at least $|z| < 1$.

If $\{a_n\}$ is the probability function $P_n(t)$, $n = 0, 1, 2, \dots$, of a nonnegative discrete random variable $X(t)$, Equation (1.136) is called the **pgf of $X(t)$** , denoted by $G(z, t)$, and is defined as

$$G(z, t) \equiv E(z^{X(t)}) = \sum_{n=0}^{\infty} P_n(t) z^n, |z| < 1. \quad (1.137)$$

The essence of the method of pgf is to encapsulate all the information about the random variable. The following are some properties of pgf:

- (a)

$$G(1, t) = \sum_{k=0}^{\infty} P_n(t) = 1. \quad (1.138)$$

- (b) If X is a random variable, all its moments can be found via Equation (1.137) because

$$\frac{\partial^k G(z, t)}{\partial z^k} = \sum_{k=n}^{\infty} k! P_k(t) \frac{z^{k-n}}{(n-k)!} \quad (1.139)$$

and, hence,

$$\begin{aligned} \left. \frac{\partial^k G(z, t)}{\partial z^k} \right|_{z=1} &= \sum_{k=n}^{\infty} k! P_k(t) \frac{z^{k-n}}{(n-k)!} \\ &= E[X(X-1) \cdots (X-n+1)]. \end{aligned} \quad (1.140)$$

For instance, if $n = 1$, from Equation (1.140) we have $E(X)$ and if $k = 2$, we have $E[X(X-1)] = E(X^2) - E(X)$. Thus, variance can be obtained as

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = E[X(X-1)] + E(X) - [E(X)]^2. \quad (1.141)$$

- (c) Applying Taylor expansion in Equation (1.136), we obtain

$$\left. \frac{\partial^k G(z, t)}{\partial z^k} \right|_{z=0} = n! P_k(t). \quad (1.142)$$

Theorem 1.9.5 The generating function of sum of n independent random variables is the product of the generating function of each variable.

Proof: The proof of this is left as an exercise for the reader.

Example 1.9.1 An Application of pgf This example is to show how the pgf method works, particularly on a system of differential difference equation. This example is extracted from Example 4.5.2 of Haghghi and Mishev (2013).

Let us solve the following system of differential difference equations using the pgf method:

$$\begin{cases} P'_0(t) = -\lambda P_0(t), \\ P'_n(t) = -\lambda P_n(t) + \lambda P_{n-1}(t), \quad n = 1, 2, \dots, \end{cases} \quad (1.143)$$

with initial condition $P_0(0) = \delta_{i0}$, where δ_{i0} is **Kronecker's delta function** defined as

$$\delta_{i0} = \begin{cases} 1, & \text{if } i = 0, \\ 0, & \text{if } i \neq 0, \end{cases} \quad (1.144)$$

and normalizing equation $\sum_{n=0}^{\infty} P_n(t) = 1$. We need to find $P_n(t)$, $n = 0, 1, 2, \dots$

Using Equation (1.137), we first find the pgf of Equation (1.144) that gives

$$G(z, 0) = z^i. \quad (1.145)$$

Differentiating Equation (1.137) term by term, within the region of convergence, with respect to t , gives

$$\frac{\partial}{\partial t} G(z, t) = \sum_{n=0}^{\infty} \frac{\partial}{\partial t} [P_n(t)] z^n = P'_0(t) + \sum_{n=1}^{\infty} P'_n(t) z^n. \quad (1.146)$$

Now, multiplying the first equation of system (Equation 1.137) by z^n and summing over $n = 1, 2, \dots$, we have

$$\frac{\partial}{\partial t} G(z, t) - P'_0(t) = -\lambda[G(z, t) - P_0(t) - zG(z, t)]. \quad (1.147)$$

Using the second equation of system (Equation 1.137), we have

$$\frac{\partial}{\partial t} G(z, t) = G(z, t)[\lambda(z - 1)]. \quad (1.148)$$

Solving the system of two equations (Eqs. 1.147 and 1.148) together with Equation (1.145), we have

$$G(z, t) = z^i e^{\lambda(z-1)t}. \quad (1.149)$$

Thus, $P_n(t)$, $n = 0, 1, 2, \dots$, are the coefficients of the Taylor expansion of Equation (1.149), which are obtained as

$$P_n(t) = \begin{cases} 0, & n = 0, 1, 2, \dots, i - 1, \\ \frac{(\lambda t)^{n-1}}{(n-1)!} e^{-\lambda t} & n = i, i + 1, i + 2, \dots \end{cases} \quad (1.150)$$

1.9.4 Laplace Transform Method



Portrait of Pierre-Simon Laplace

Before we start the discussion of Laplace transform, it would be instructive to give a brief history of this great mathematician. **Pierre-Simon, marquis de Laplace** (23 March 1749–5 March 1827) was a French mathematician and astronomer, who was born in Beaumont-en-Auge, Normandy. He formulated Laplace's equation and pioneered the Laplace transform, which appears in many branches of mathematical physics, a field that he took a leading role in forming. The Laplacian differential operator, widely used in applied mathematics, is also named after him. He is remembered as one of the greatest scientists of all time, sometimes referred to as a French Newton or Newton of France, with a phenomenal natural mathematical faculty superior to any of his contemporaries (http://en.wikipedia.org/wiki/Pierre-Simon_Laplace).

Laplace transform has been a powerful method for solving linear ordinary differential equations, their corresponding initial-valued problems, and systems of linear differential equations. In order to obtain functions from their Laplace transforms, the inverse Laplace transform was introduced at the same time. Therefore, we start with the definition of Laplace transform.

The **Laplace transform** (or **unilateral Laplace transform**) of a function $f(t)$, $t \geq 0$, is an integral operator to this function on $[0, \infty)$, denoted by $F(s)$, $\mathcal{L}(f)(s)$, $\mathcal{L}\{f(t)\}$ or $F * (s)$ (and many other notations used by different authors). The parameter s is arbitrary and is taken as in the definition (Equation 1.151) below:

$$F(s) = \mathcal{L}(f)(s) = \mathcal{L}\{f(t)\} = \int_0^{\infty} e^{-st}f(t)dt. \quad (1.151)$$

The **bilateral Laplace transform** of a function $f(t)$, $t \geq 0$, (although not often used) is defined as

$$F^{(2)}(s) = \mathcal{L}^{(2)}(f)(s) = \mathcal{L}^{(2)}\{f(t)\}(s) = \int_{-\infty}^{\infty} e^{-st}f(t)dt. \quad (1.152)$$

The Laplace–Stieltjes transform of a distribution function $F(x)$ of a random variable X , denoted by $F^*(s)$, is defined as

$$F^*(s) = E[e^{-sX}] = \int_0^{\infty} e^{-sx}dF(x), \quad (1.153)$$

where s is a complex number.

The **inverse Laplace transform** of $F(s)$, denoted by $f(t)$, $\mathcal{L}^{-1}(F)(t)$ or $\mathcal{L}^{-1}\{F(s)\}$, is defined as

$$f(t) = \mathcal{L}^{-1}(F)(t) = \mathcal{L}^{-1}\{F(s)\}, \quad (1.154)$$

if the function $F(s)$ is determined from the function $f(t)$ in (1.151). See Lerch (1903).

One of the basic properties of Laplace transform is that it is a **linear operator**, that is

$$\mathcal{L}\{af(t) + bg(t)\} = a\mathcal{L}\{f(t)\} + b\mathcal{L}\{g(t)\}, \quad (1.155)$$

for any constant a and b , provided that both $\mathcal{L}\{f(t)\}$ and $\mathcal{L}\{g(t)\}$ exist.

There are various tables for Laplace transforms and their inverses; Table 1.3 is one such table.

TABLE 1.3 Laplace Transforms of Some Elementary Functions

Formula #	$f(t)$	$F(s) \equiv \mathcal{L}\{f(t)\}$
1	$f(at)$	$\frac{1}{s}F\left(\frac{s}{a}\right), \quad s > a,$
2	$f'(t)$	$s\mathcal{L}\{f(t)\} - f(0)$
3	$f''(t)$	$s^2\mathcal{L}\{f(t)\} - sf(0) - f'(0)$
4	1	$\frac{1}{s}$
5	$\frac{t^n}{n!}, \quad n = 0, 1, 2, \dots$	$\frac{1}{s^{n+1}}, \quad n = 0, 1, 2, \dots$
6	$\frac{t^{a-1}}{\Gamma(a)}, \quad a \geq 0$	$\frac{1}{s^a}, \quad a > 0$
7	$u(t-a)$ Heaviside function	$\frac{1}{s}e^{-as}$
8	$u(t-a)f(t-a)$	$e^{-as}F(s)$
9	$\delta(t)$, unit impulse at $t = 0$.	1
10	$\delta(t-a)$ Dirac delta function	e^{-as}
11	$e^{at}, \quad a \geq 0$	$\frac{1}{s-a}, \quad s > a$
12	$e^{at}f(t), \quad a \geq 0$	$F(s-a), \quad s > a,$
13	$\frac{(ac+d)e^{at} - (bc+d)e^{bt}}{a-b}$	$\frac{cs+d}{(s-a)(s-b)}, \quad a \neq b$
14	$\frac{1}{t}(e^{at} - e^{bt})$	$\ln\left(\frac{s-b}{s-a}\right)$
15	$\sin(at)$	$\frac{a}{s^2+a^2}$
16	$\cos(at)$	$\frac{s}{s^2+a^2}$

Theorem 1.9.6 Existence of Laplace Transform Let $f(t)$ be a function that (1) it is defined on $[0, \infty)$; (2) it is piecewise continuous on $[0, \infty)$; and (3) it satisfies

$$|f(t)| \leq Me^{kt}, \quad t \in [0, \infty], \quad (1.156)$$

for some constants M and k , with $M > 0$. Then, $f(t)$'s Laplace transform $\mathcal{L}(f)$ exists for $s > k$.

Proof: Proof of this is left as an exercise for the reader.

The following example is extracted from Example 4.5.2 of Haghghi and Mishev (2013):

Example 1.9.2 Let us consider the system (Equation 1.143) discussed in Example 1.9.1. We now solve the same system using recursive and Laplace methods combined and find $P_n(t)$, $n = 0, 1, 2, \dots$

Let us denote by $P_0^*(s)$ and $P_n^*(s)$, $n = 1, 2, \dots$, the Laplace transforms of $P_0(t)$ and $P_n(t)$, $n = 1, 2, \dots$, respectively. Then, using formula # 2 in Table 1.3 we will have:

$$\mathcal{L}[P'_0(t)] = sP_0^*(s) - P_0(0) = sP_0^*(s) - 1 \tag{1.157}$$

and

$$\mathcal{L}[P'_n(t)] = sP_n^*(s) - P_n(0) = sP_n^*(s). \tag{1.158}$$

Substituting Eqs. (1.157) and (1.158) into Equation (1.137), we have

$$\begin{cases} sP_0^*(s) - 1 = -\lambda P_0^*(s), \\ sP_n^*(s) = -\lambda P_n^*(s) + \lambda P_{n-1}^*(s), \quad n = 1, 2, \dots, \end{cases} \tag{1.159}$$

or

$$\begin{cases} (s + \lambda)P_0^*(s) = 1 \\ (s + \lambda)P_n^*(s) = \lambda P_{n-1}^*(s), \quad n = 1, 2, \dots \end{cases} \tag{1.160}$$

Recursively, we will have the following:

$$\left. \begin{aligned} P_0^*(s) &= \frac{1}{(s + \lambda)}, \\ P_1^*(s) &= \frac{\lambda}{(s + \lambda)} P_0^*(s) = \frac{\lambda}{(s + \lambda)} \frac{\lambda}{(s + \lambda)} \\ P_2^*(s) &= \frac{\lambda}{(s + \lambda)^2}, \\ &\vdots \\ P_n^*(s) &= \frac{\lambda^n}{(s + \lambda)^{n+1}}. \end{aligned} \right\} \tag{1.161}$$

Then, inverting Equation (1.161), we have

$$P_n(t) = \mathcal{L}^{-1}[P_n^*(s)] = \mathcal{L}^{-1} \left[\frac{\lambda^n}{(s + \lambda)^{n+1}} \right] = \lambda^n \frac{t^n e^{-\lambda t}}{n!}, \quad n = 1, 2, \dots, \tag{1.162}$$

and

$$P_0(t) = \mathcal{L}^{-1}[P_0^*(s)] = \mathcal{L}^{-1} \left[\frac{1}{(s + \lambda)} \right] = e^{-\lambda t}. \tag{1.163}$$

Combining Eqs. (1.162) and (1.163), we will have

$$P_n(t) = \frac{(\lambda t)^n}{n!} e^{-\lambda t}, \quad n = 0, 1, 2, \dots, \quad (1.164)$$

which is the Poisson probability distribution function.

EXERCISES

- 1.1 If X and Y are two discrete random variables, then prove that $X \pm Y$, XY , and X/Y are also random variables, for the last one, provided Y is nonzero.
- 1.2 Prove that relation (1.14) is, indeed, a probability distribution function.
- 1.3 Prove that relation (1.15) is, indeed, a probability distribution function.
- 1.4 Prove that two random variables X and Y , with their respective pmfs p_X and p_Y and joint mass function, denoted by $p_{X,Y}$, are independent if and only if $p_{XY} = p_X p_Y$.
- 1.5 Prove that: $E(I_A)P(A)$, where I_A is the indicator function of A .
- 1.6 If each of the two sequences $\{X_n, n \in \mathbb{N}\}$ and $\{Y_n, n \notin \mathbb{N}\}$ converges in probability to X and Y , respectively, with all random variables defined on the same sample space Ω , that is, $X_n \xrightarrow{\text{prob}} X$, and $Y_n \xrightarrow{\text{prob}} Y$, then prove that $(X_n + Y_n) \xrightarrow{\text{prob}} (X + Y)$.
- 1.7 Show that almost sure convergence implies convergence in probability, but the converse is not true.
- 1.8 Give the detail that the first two moments and variance of a hyperexponential distribution function are, respectively, Eqs. (1.79), (1.109), and (1.110).
- 1.9 Prove Theorem 1.8.1.
- 1.10 Prove relation (1.134).
- 1.11 Prove Theorem 1.9.4.
- 1.12 Prove Theorem 1.9.5.
- 1.13 Prove Theorem 1.9.6.
- 1.14 Let the discrete time random process $\{X_n\}$ be defined by $X_n = s^n$, $n = 0, 1, \dots$, where s is selected randomly from $(0,1)$.
 - a. Find the cumulative distribution of X_n .
 - b. Find the joint cumulative distribution of X_n and X_{n+1} .

- 1.15** Let $X(t) = At + B$, where A and B are independent random variables.
- Find pdf of $X(t)$, $t > 0$.
 - Find the mean of $X(t)$.
- 1.16** Let X_n consist of an iid sequence of random variables with Poisson distribution with parameter λ , $\lambda > 0$, and $S_n = X_1 + X_2 + \cdots + X_n$.
- Find the pmf of S_n .
 - Find the joint pmf of S_n and S_{n+k} .
 - Find the covariance of S_n and S_{n+k} .

