Introduction to Probability and Stochastic Processes with Applications

Introduction to Probability and Stochastic Processes with Applications

Liliana Blanco Castañeda
National University of Colombia
Bogotá, Colombia

Viswanathan Arunachalam
Universidad de los Andes
Bogotá, Colombia

Delvamuthu Dharmaraja
Indian Institute of Technology Delhi
New Delhi, India

WILEY

A JOHN WILEY & SONS, INC., PUBLICATION

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
Published simultaneously in Canada

Library of Congress Cataloging-in-Publication Data:

Blanco Castaneda, Liliana.
Introduction to probability and stochastic processes with applications / Liliana Blanco Castaneda, Viswanathan Arunachalam, Selvamuthu Dharmaraja.
p. cm.
Includes bibliographical references and index.
ISBN 978-1-118-29440-6 (hardback)
1. Probabilities—Textbooks. 2. Stochastic processes—Textbooks. I. Arunachalam, Viswanathan, 1969- II. Dharmaraja, Selvamuthu, 1972- III. Title.
QA274.B53 2012
519.2—dc23 2012002024

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

To Sebastián and Paula
L.B.C.

To Akshaya and Abishek
V.A.

To Kathiravan and Madhuvanth
S.D.

CONTENTS IN BRIEF

CONTENTS

FOREWORD

Probability theory is the fulcrum around which the present-day mathematical modeling of random phenomena revolves. Given its broad and increasing application in everyday life-trade, manufacturing, reliability, or even biology and psychology, there is an ever-growing demand from researchers for strong textbooks expounding the theory and applications of probabilistic models.

This book is sure to be invaluable to students with varying levels of skill, as well as scholars who wish to pursue probability theory, whether pure or applied. It contains many different ideas and answers many questions frequently asked in classrooms. The extent of the exercises and examples chosen from a multitude of areas will be very helpful for students to understand the practical applications of probability theory.

The authors have extensively documented the origins of probability, giving the reader a clear idea of the needs and developments of the subject over many centuries. They have taken care to maintain an approach that is mathematically rigorous but at the same time simplistic and thus appealing to students.

Although a wide array of applications have been covered in various chapters, I must make particular mention of the chapters on queueing theory and financial mathematics. While the latter is an emerging topic, there is no limit on the applicability of queueing models to other diverse areas.

In all, the present book is the result of a long and distinguished teaching experience of probability, queueing theory, and financial mathematics, and this book is sure to advance the readers' knowledge of this field.

Professor Alagar Rangan
Eastern Mediterranean University
North Cyprus

PREFACE

This text is designed for a first course in the theory of probability and a subsequent course on stochastic processes or stochastic modeling for students in science, engineering, and economics, in particular for students who wish to specialize in probabilistic modeling. The idea of writing this book emerged several years ago, in response to students enrolled in courses that we were teaching who wished to refer to materials and problems covered in the lectures. Thus the edifice and the building blocks of the book have come mainly from our continuously updated and expanded lecture notes over several years.

The text is divided into twelve chapters supplemented by four appendices. The first chapter presents basic concepts of probability such as probability spaces, independent events, conditional probability, and Bayes' rule. The second chapter discusses the concepts of random variable, distribution function of a random variable, expected value, variance, probability generating functions, moment generating functions, and characteristic functions. In the third and fourth chapters, we present the distributions of discrete and continuous random variables, which are frequently used in the applications. The fifth chapter is devoted to the study of random vectors and their distributions. The sixth chapter presents the concepts of conditional probability and conditional expectation, and an introduction to the study of the multivariate normal distribution is discussed in seventh chapter. The law of large numbers and limit theorems are the goals of the eighth chapter, which studies four types of convergence for sequences of random variables, establishes relationships between them and discusses weak and strong laws of large numbers and the central limit theorem. The ninth chapter introduces stochastic processes with discrete and continuous-time Markov chains as the focus of study. The tenth chapter is devoted to queueing models and their applications. In eleventh chapter eleven we present an elementary introduction to stochastic calculus where martingales, Brownian motion, and Itô integrals are introduced. Finally, the last chapter is devoted to the introduction of mathematical finance. In this chapter, pricing methods such as risk-neutral valuation and Black-Scholes formula are discussed.

In the appendices, we summarize a few mathematical basics needed for the understanding of the material presented in the book. These cover ideas from set theory, combinatorial analysis, and linear algebra. Finally, the last appendix contains tables of standard distributions, which are used in applications. The bibliography is given at the end of the book, though it is not a complete list.

At the end of each chapter there is a list of exercises to facilitate understanding of the main body of each chapter, and in some cases, additional study material. Most of the examples and exercises are classroom tested in the courses that we taught over many years. We have also benefited from various books on probability and statistics for some of the examples and exercises in the text. To understand this text, the reader must have solid knowledge of differential and integral calculus and some linear algebra.

We do hope that this introductory book provides the foundation for students to learn other subjects in their careers. This book is comprehensible to students with diverse backgrounds. It is also well balanced, with lots of motivation to learn probability and stochastic processes and their applications. We hope that this book will serve as a valuable text for students and reference for researchers and practitioners who wish to consult probability and its applications.

L. BLANCO, V. ARUNACHALAM, S. DHARMARAJA

Bogotá, Colombia
December, 2011

ACKNOWLEDGMENTS

We are grateful to Professor Ignacio Mantilla for providing us with motivation, academic support, and advice for this book project. We are grateful to Professor Alagar Rangan for his encouragement and careful reading of the draft of this book and offering invaluable advice. This book has greatly benefited from his comments and suggestions. We thank Professor Diego Escobar for his useful suggestions. Our sincere thanks to Dr. Liliana Garrido for her careful reading as well as her suggestions. We record our appreciation to Laura Vielma, Christian Bravo, and Hugo Ramirez for their assistance in typing this book. We thank our students for their feedback, incisive questions and enthusiasm, and this has served as the platform for this project. We acknowledge National University of Colombia, Universidad de los Andes, and Indian Institute of Technology Delhi for the institutional support.

It is a pleasure to thank our Editor, Ms. Susanne Steitz-Filler, John Wiley & Sons, and her colleagues for providing advice and technical assistance.

Finally, last but foremost, we thank our family for their love and support. They were instrumental in bringing this book to fruition.

L.B.C, V.A. and S.D.

INTRODUCTION

Since its origin, probability theory has been linked to games of chance. In fact by the time of the first roman emperor, Augustus (63 B.C.–14 A.D.), random games were fairly common and mortality tables were being made. This was the origin of probability and statistics. Later on, these two disciplines started drifting apart due to their different objectives but always remained closely connected. In the sixteenth century philosophical discussions around probability were held and Italian philosopher Gerolamo Cardano (1501–1576) was among the first to make a mathematical approach to randomness. In the seventeenth and eighteenth centuries major advances in probability theory were made due in part to the development of infinitesimal calculus; some outstanding results from this period include: the law of large numbers due to James Bernoulli (1654–1705), a basic limit theorem in modern probability which can be stated as follows: if a random experiment with only two possible outcomes (success or failure) is carried out, then, as the number of trials increases the success ratio tends to a number between 0 and 1 (the success probability); and the DeMoivre-Laplace theorem (1733, 1785 and 1812), which established that for large values of n a binomial random variable with parameters n and p has approximately the same distribution of a normal random variable with mean np and variance $np(1-p)$. This result was proved by DeMoivre in 1733 for the case $p = \frac{1}{2}$ and then extended to arbitrary $0 < p < 1$ by Laplace in

1812. In spite of the utmost importance of the aforementioned theoretical results, it is important to mention that by the time they were stated there was no clarity on the basic concepts. Laplace's famous definition of probability as the quotient between cases in favor and total possible cases (under the assumption that all results of the underlying experiment were equally probable) was already known back then. But what exactly did it mean "equally probable"? In 1892 the German mathematician Karl Stumpf interpreted this expression saying that different events are equally probable when there is no knowledge whatsoever about the outcome of the particular experiment. In contrast to this point of view, the German philosopher Johannes von Kries (1853–1928) postulated that in order to determine equally probable events, an objective knowledge of the experiment was needed. Thereby, if all the information we possess is that a bowl contains black and white balls, then, according to Strumpf, it is equally probable to draw either color on the first attempt, while von Kries would admit this only when the number of black and white balls is the same. It is said that Markov himself had trouble regarding this: according to Krengel (2000) in Markov's textbook (1912) the following example can be found: "suppose that in an urn there are balls of four different colors $1, 2, 3$ and 4 each with unknown frequencies a, b, c and d, then the probability of drawing a ball with color 1 equals $\frac{1}{4}$ since all colors are equally probable". This shows the lack of clarity surrounding the mathematical modeling of random experiments at that time, even those with only a finite number of possible results.

The definition of probability based on the concept of equally probable led to certain paradoxes which were suggested by the French scientist Joseph Bertrand (1822–1900) in his book *Calcul des probabilites* (published in 1889). One of the paradoxes identified by Bertrand is the so-called paradox of the three jewelry boxes. In this problem, it is supposed that three jewelry boxes exist, A, B and C, each having two drawers. The first jewelry box contains one gold coin in each of the drawers, the second jewelry box contains one silver coin in each of the drawers and in the third one, one of the drawers contains a gold coin and the other a silver coin. Assuming Laplace's definition of probability, the probability of choosing the third jewelry box would be $\frac{1}{3}$. Let us suppose now that a jewelry box is randomly chosen and when one of the drawers is opened a gold coin is found. Then there are two options: either the other drawer contains a gold coin (in which case the chosen jewelry box would be A) or the other drawer contains a silver coin, which means the chosen jewelry box is C. If the coin originally found is silver, there would be two options: either the other drawer contains a gold coin, which means the chosen jewelry box is C, or the other drawer contains a silver coin, which would mean that the chosen jewelry box is B. Hence the probability of choosing C is $\frac{1}{2}$. Bertrand found it paradoxical that opening a drawer changed the probability of choosing jewelry box C.

The first mathematician able to solve the paradox of the three jewelry boxes, formulated by Bertrand, was Poincaré, who got the following solution as early as 1912. Let us assume that the drawers are labeled (in a place we are unable to see) as α and β and that the gold coin of jewelry box C is in drawer α. Then the following possibilities would arise:

1. Jewelry box A, drawer α: gold coin
2. Jewelry box A, drawer β: gold coin
3. Jewelry box B, drawer α: silver coin
4. Jewelry box B, drawer β: silver coin
5. Jewelry box C, drawer α: gold coin
6. Jewelry box C, drawer β: silver coin

If when opening a drawer a gold coin is found, there would be three possible cases: 1, 2 and 5. Of those cases the only one that favors is case 5. Hence $P(C) = \frac{1}{3}$.

At the beginning of the twentieth century and despite being the subject of works by famous mathematicians such as Cardano, Fermat, Bernoulli, Laplace, Poisson and Gauss, probability theory was not considered in the academic field as a mathematical discipline and it was questioned whether it was a rather empirical science. In the famous Second International Congress of Mathematicians held in Paris in 1900, David Hilbert, in his transcendental conference of August 8, proposed as part of his sixth problem the axiomatization of the calculus of probabilities. In 1901 G. Bohlmann formulated a first approach to the axiomatization of probability (Krengel, 2000): he defines the probability of an event E as a nonnegative number $p(E)$ for which the following hold:

i) If E is the sure event, then $p(E) = 1$.

ii) If E_1 and E_2 are two events such that they happen simultaneously with zero probability, then the probability of either E_1 or E_2 happening equals $p(E_1) + p(E_2)$.

By 1907 the Italian Ugo Broggi, under Hilbert's direction, wrote his doctoral dissertation titled "Die Axiome der Wahrscheinlichkeitsrechnung" (*The Axioms of the Calculus of Probabilities*). The definition of event is presented loosely and it is asserted that additivity and σ-additivity are equivalent (the proof of this false statement contains so many mistakes that it is to be assumed that Hilbert did not read it carefully). However, this work can be considered as the predecessor of Kolmogorov's.

At the International Congress of Mathematicians in Rome in 1908, Bohlmann defined the independence of events as it is currently known and showed the difference between this and 2×2 independence. It is worth noting that a precise definition of event was still missing.

According to Krengel (2000), in 1901 the Swedish mathematician Anders Wiman (1865–1959) used the concept of measure in his definition of geometric probability. In this regard, Borel in 1905 says: "When one uses the convention: the probability of a set is proportional to its length, area or volume, then one must be explicit and clarify that this is not a definition of probability but a mere convention".

Thanks to the works of Fréchet and Caratheodory, who "liberated" measure theory from its geometric interpretation, the path to the axiomatization of probability as it is currently known was opened. In the famed book *Grundbegriffe der Wahrscheinlichkeitsrechnung* (*Foundations of the Theory of Probability*), first published in 1933, the Russian mathematician Andrei Nikolaevich Kolmogorov (1903–1987) axiomatized the theory of probability by making use of measure theory, achieving rigorous definitions of concepts such as probability space, event, random variable, independence of events, and conditional probability, among others. While Kolmogorov's work established explicitly the axioms and definitions of probability calculus, it furthermore laid the ground for the theory of stochastic processes, in particular, major contributions to the development of Markov and ramification processes were made. One of the most important results presented by Kolmogorov is the consistency theorem, which is fundamental to guarantee the existence of stochastic processes as random elements of finite-dimensional spaces.

Probability theory is attractive not only for being a complex mathematical theory but also for its multiple applications to other fields of scientific interest. The wide spectrum of applications of probability ranges from physics, chemistry, genetics and ecology to communications, demographics and finance, among others. It is worth mentioning that Danish mathematician, statistician and engineer Agner Krarup Erlang (1878–1929) for his contribution to queueing theory.

At the beginning of the twentieth century, one of the most important scientific problems was the understanding of Brownian motion, named so after the English botanist Robert Brown (1773–1858), who observed that pollen particles suspended in a liquid, move in a constant and irregular fashion. Brown initially thought that the movement was due to the organic nature of pollen, but later on he would refute this after verifying with a simple experiment that the same behavior was observed with inorganic substances.

Since the work done by Brown and up to the end of the nineteenth century there is no record of other investigations on Brownian motion. In 1905 in his article "Über die von der molekularkinetischen Theorie der Warme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen" (On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat; (see Kahane, 1997) German theoretical physicist Albert Einstein (1879–1955) published the main characteristics of Brownian motion. He proved that the movement of the particle at instant t can be modeled by means of a normal distribution and concluded that this motion is a consequence of continuous collisions between the particle and the molecules of the liquid in which it is suspended. It is worth pointing out, however, that Einstein himself said he did not know Brown's works (Nelson, 1967). The first mathematical research regarding Brownian motion was carried out by French mathematician Louis Bachelier (1870–1946), whose 1900 doctoral dissertation "Theorie de la spéculation" (Speculation theory) suggested the Brownian motion as a model associated with speculative prices. One of the imperfections of such a model laid in the fact that it allowed prices to take negative values and therefore was forgotten for a long time. In 1960 the economist Samuelson (who received the Nobel Prize in Economics in 1970) suggested the exponential of the Brownian motion to model the behavior of prices subject to speculation.

The mathematical structure of Brownian motion, as it is known today, is due to the famed North American mathematician Norbert Wiener (1894–1964). For this reason Brownian motion is also called the Wiener process. The first articles about Brownian motion by Wiener are rather hard to follow and only the French mathematician Paul Lévy (1886–1971) was able to recognize its importance. Paul Lévy notably contributed to the development of probability by introducing the concept of the martingale, the Lévy processes among which we find the Brownian motion and the Poisson processes and the theorem of continuity of characteristic functions. Furthermore, Lévy deduced many of the most important properties of Brownian motion. It is said (see Gorostiza, 2001) that many times it has happened that major discoveries in probability theory believed to be new were actually somehow contained in Lévy's works.

During the 1970s, the Black-Scholes and Merton formula, which allows the pricing of put and call options for the European market, was written. For this work Scholes and Merton were awarded the 1997 Nobel Prize in Economics (Black's death in 1995 rendered him ineligible). Nevertheless, the research carried out by Black-Scholes and Merton would have been impossible without the previous works done by the Japanese mathematician Kiyoshi Itô (1915–2008), who in 1940 and 1946 published a series of articles introducing two of the most essential notions of modern probability theory: stochastic integrals and stochastic differential equations. These concepts have become an influential tool in many mathematical fields, e.g., the theory of partial differential equations, as well as in applications that go beyond financial mathematics and include theoretical physics, biology, and engineering, among others (see Korn and Korn, 2000).

CHAPTER 1

BASIC CONCEPTS

During the early development of probability theory, the evolution was based more on intuition rather than mathematical axioms. The axiomatic basis for probability theory was provided by A. N. Kolmogorov in 1933 and his approach conserved the theoretical ideas of all other approaches. This chapter is based on the axiomatic approach and starts with this notion.

1.1 PROBABILITY SPACE

In this section we develop the notion of probability measure and present its basic properties.

When an ordinary die is rolled once, the outcome cannot be accurately predicted; we know, however, that the set of all possible outcomes is $\{1, 2, 3, 4, 5, 6\}$. An experiment like this is called a *random experiment.*

Definition 1.1 (Random Experiment) *An experiment is said to be random if its result cannot be determined beforehand.*

It is assumed that the set of possible results of a random experiment is known. This set is called a *sample space.*

Introduction to Probability and Stochastic Processes with Applications, First Edition.
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

Definition 1.2 (Sample Space) *The set Ω of all possible results of a random experiment is called a sample space. An element $\omega \in \Omega$ is called an outcome or a sample point.*

■ **EXAMPLE 1.1**

Experiment: Flipping a fair coin. The possible results in this case are "head"=H and "tail"=T. That is, $\Omega = \{H, T\}$. ▲

■ **EXAMPLE 1.2**

Experiment: Rolling an ordinary die three consecutive times. In this case the possible results are triplets of the form (a, b, c) with $a, b, c \in \{1, 2, 3, 4, 5, 6\}$. That is:

$$\Omega = \{(a, b, c) : a, b, c \in \{1, 2, 3, 4, 5, 6\}\}. \quad \blacktriangle$$

■ **EXAMPLE 1.3**

Experiment: Items coming off a production line are marked defective (D) or nondefective (N). Items are observed and their condition noted. This is continued until two consecutive defectives are produced or four items have been checked, which ever occurs first. In this case:

$$\begin{aligned}\Omega = \{&DD, NDD, NDND, NNDD, NNDN, NNNN,\\ &NNND, NDNN, DNNN, DNDN, DNND, DNDD\}. \quad \blacktriangle\end{aligned}$$

■ **EXAMPLE 1.4**

Experiment: Observe the number of ongoing calls in a particular telephone exchange switch. In this case $\Omega = \{0, 1, 2, \cdots\}$. ▲

We notice that the elements of a sample space can be numbers, vectors, symbols, etc. and they are determined by the experiment being considered.

Definition 1.3 (Discrete Sample Space) *A sample space Ω is called discrete if it is either finite or countable. A random experiment is called finite (discrete) if its sample space is finite (discrete).*

Going back to Example 1.2, a question that arises naturally is: what's the "chance" of a given "event" such as "the sum of the results obtained is greater than or equal to 2"? In other words, what is the "chance" of

$$A := \{(a, b, c) \in \Omega : a + b + c \geq 2\}$$

happening?

Now, what is an event? Following the aforementioned idea, we can expect an event merely to be a subset of the sample space, but in this case, can we say that all subsets of the sample space are events? The answer is no. The class of subsets of the sample space for which the "chance" of happening is defined must have a σ-algebra structure, a concept we will further explain:

Definition 1.4 (σ-Algebra) *Let $\Omega \neq \emptyset$. A collection $\Im$ of subsets of Ω is called a σ-algebra (or a σ-field) over Ω:*

(i) If $\Omega \in \Im$.

(ii) If $A \in \Im$, then $A^c \in \Im$.

(iii) If $A_1, A_2, \cdots \in \Im$, then $\bigcup_{i=1}^{\infty} A_i \in \Im$.

The elements of $\Im$ are called events.

■ EXAMPLE 1.5

Consider Example 1.1. $\Omega = \{H, T\}$. Then $\mathcal{F} = \{\emptyset, \Omega\}$ is a trivial σ-algebra over Ω, whereas $\mathcal{G} = \{\emptyset, \{H\}\}$ is not a σ-algebra over Ω. ▲

■ EXAMPLE 1.6

Consider a random experiment of flipping two fair coins.
$\Omega = \{HH, HT, TH, TT\}$. Then $\mathcal{F} = \{\emptyset, \{HH, HT\}, \{TH, TT\}, \Omega\}$ is a σ-algebra over Ω. ▲

■ EXAMPLE 1.7

Consider Example 1.2. $\Omega = \{(a, b, c) : a, b, c \in \{1, 2, 3, 4, 5, 6\}\}$. Then
$\mathcal{F} = \{\emptyset, \{(1, 2, 3)\}, \Omega \setminus \{(1, 2, 3)\}, \Omega\}$ is a σ-algebra over Ω whereas
$\mathcal{G} = \{(1, 2, 3), (1, 1, 1)\}$ is not a σ-algebra over Ω. ▲

EXAMPLE 1.8

Let $\Omega \neq \emptyset$. Then $\Im_0 = \{\emptyset, \Omega\}$ and $\wp(\Omega) := \{A : A \subseteq \Omega\}$ are σ-algebras over Ω. $\Im_0$ is called the trivial σ-algebra over Ω while $\wp(\Omega)$ is known as the total σ-algebra over Ω . ▲

EXAMPLE 1.9

Let $\Omega = \{1, 2, 3\}$. Then $\Im = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ is a σ-algebra over Ω but the collection $\mathcal{G} = \{\emptyset, \{1\}, \{2\}, \{3\}, \Omega\}$ is not a σ-algebra over Ω. ▲

EXAMPLE 1.10

Let $\Omega \neq \emptyset$ be finite or countable, and let $\Im$ be a σ-algebra over Ω containing all subsets of the form $\{\omega\}$ with $\omega \in \Omega$. Then $\Im = \wp(\Omega)$. ▲

Theorem 1.1 *If $\Omega \neq \emptyset$ and $\Im_1, \Im_2, \cdots$ are σ-algebras over Ω, then $\bigcap\limits_{i=1}^{\infty} \Im_i$ is also a σ-algebra over Ω.*

Proof: Since $\Omega \in \Im_j$ for every j, this implies that $\Omega \in \bigcap_j \Im_j$. Let $A \in \bigcap_j \Im_j$, then $A \in \Im_j$ for all j, this means that $A^c \in \Im_j$ for all j. Hence $A^c \in \bigcap_j \Im_j$. Finally, let $A_1, A_2, \cdots \in \bigcap_j \Im_j$. Then $A_i \in \Im_j$, for all i and j, hence $\bigcup_i A_i \in \Im_j$ for all j. Thus we conclude that $\bigcup_{i=1}^{\infty} A_i \in \bigcap_j \Im_j$. ∎

Note that, in general, the union of σ-algebras over Ω is not a σ-algebra over Ω. For example, $\Omega = \{1, 2, 3\}, \Im_1 = \{\emptyset, \Omega, \{1\}, \{2, 3\}\}$ and $\Im_2 = \{\emptyset, \Omega, \{1, 2\}, \{3\}\}$. Clearly $\Im_1$ and $\Im_2$ are σ-algebras over Ω, but $\Im_1 \cup \Im_2$ is not a σ-algebra over Ω.

Definition 1.5 (Generated σ-Algebra) *Let $\Omega \neq \emptyset$ and let $\mathcal{A}$ be a collection of subsets of Ω. Let $\mathcal{M} = \{\Im : \Im$ is a σ-algebra over Ω containing $\mathcal{A}\}$. Then, the preceding example implies that $\sigma(\mathcal{A}) := \bigcap\limits_{\Im \in \mathcal{M}} \Im$ is the smallest σ-algebra over Ω containing $\mathcal{A}$. $\sigma(\mathcal{A})$ is called the σ-algebra generated by $\mathcal{A}$.*

EXAMPLE 1.11 Borel σ-Algebra

The smallest σ-algebra over $\mathbb{R}$ containing all intervals of the form $(-\infty, a]$ with $a \in \mathbb{R}$ is called the Borel σ-algebra and is usually written as $\mathcal{B}$. If $A \in \mathcal{B}$, then A is called a Borel subset of $\mathbb{R}$. Since $\mathcal{B}$ is a σ-algebra, if we take $a, b \in \mathbb{R}$ with $a < b$, then the following are Borel subsets of $\mathbb{R}$:

$$\begin{aligned}
(a, \infty) &= \mathbb{R} \setminus (-\infty, a] \\
(a, b] &= (-\infty, b] \cap (a, \infty) \\
(-\infty, a) &= \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n}\right] \\
[a, \infty) &= \mathbb{R} \setminus (-\infty, a)
\end{aligned}$$

$$\begin{aligned}
(a, b) &= (-\infty, b) \cap (a, \infty) \\
[a, b] &= \mathbb{R} \setminus ((-\infty, a) \cup (b, \infty)) \\
\{a\} &= [a, a] \\
\mathbb{N} &= \bigcup_{n=0}^{\infty} \{n\} \\
\mathbb{Q} &= \bigcup_{\substack{m, n \in \mathbb{Z} \\ n \neq 0}} \left\{\frac{m}{n}\right\} \\
\mathbb{Q}^c &= \mathbb{R} \setminus \mathbb{Q}. \qquad \blacktriangle
\end{aligned}$$

Can we say, then, that all subsets of $\mathbb{R}$ are Borel subsets? The answer to this question is no; see Royden (1968) for an example on this regard.

■ EXAMPLE 1.12 Borel σ-Algebra over $\mathbb{R}^n$

Let $a = (a_1, \cdots, a_n)$ and $b = (b_1, \cdots, b_n)$ be elements of $\mathbb{R}^n$ with $a \leq b$, that is, $a_i \leq b_i$ for all $i = 1, \cdots, n$. The σ-algebra, denoted by $\mathcal{B}_n$, generated by all intervals of the form

$$(a, b] := \{x = (x_1, \cdots, x_n) \in \mathbb{R}^n : a_i < x_i \leq b_i, i = 1, \cdots, n\}$$

is called the Borel σ-algebra over $\mathbb{R}^n$. $\blacktriangle$

Definition 1.6 (Measurable Space) *Let $\Omega \neq \emptyset$ and let $\Im$ be a σ-algebra over Ω. The couple $(\Omega, \Im)$ is called a measurable space.*

It is clear from the definition that both Ω and $\emptyset$ belong to any σ-algebra defined over Ω. $\emptyset$ is called the impossible event, Ω is called the sure event. An event of the form $\{\omega\}$ with $\omega \in \Omega$ is called a simple event.

We say that the event A happens if after carrying out the random experiment we obtain an outcome in A, that is, A happens if the result is a certain ω with $\omega \in A$. Therefore, if A and B are two events, then:

(i) The event $A \cup B$ happens if and only if either A or B or both happen.

(ii) The event $A \cap B$ happens if and only if both A and B happen.

(iii) The event A^c happens if and only if A doesn't happen.

(iv) The event $A \setminus B$ happens if and only if A happens but B doesn't.

■ **EXAMPLE 1.13**

If in Example 1.2 we consider the events: $A =$"the result of the first toss is a prime number" and $B =$"the sum of all results is less than or equal to 4". Then

$$A \cup B = \{(a, b, c) \in \Omega : a \in \{2, 3, 5\} \text{ or } (a + b + c) \leq 4\},$$

so $(2, 1, 1), (5, 3, 4), (1, 1, 1)$ are all elements of $A \cup B$. In addition:

$$A \cap B = \{(a, b, c) : a \in \{2, 3, 5\} \text{ and } (a + b + c) \leq 4\} = \{(2, 1, 1)\}.$$

The reader is advised to see what the events $A \setminus B$ and A^c are equal to.
▲

Definition 1.7 (Mutually Exclusive Events) *Two events A and B are said to be mutually exclusive if $A \cap B = \emptyset$.*

■ **EXAMPLE 1.14**

A coin is flipped once. Let $A =$"the result obtained is a head" and $B =$"the result obtained is a tail". Clearly the events A and B are mutually exclusive. ▲

■ **EXAMPLE 1.15**

A coin is flipped as many times as needed to obtain a head for the first time, and the number of tosses required is being counted. If

$A :=$ "no heads that are obtained before the third toss" $= \{3, 4, 5, \cdots\}$ and

$B :=$ "no heads that are obtained before the second toss" $= \{2, 3, 4, \cdots\}$

then A and B are not mutually exclusive. ▲

Our goal now is to assign to each event A a nonnegative real number indicating its "chance" of happening. Suppose that a random experiment is carried out n times keeping its conditions stable throughout the different repetitions.

Definition 1.8 (Relative Frequency) *For each event A, the number $fr(A) := \frac{n(A)}{n}$ is called the relative frequency of A, where $n(A)$ indicates the number of times the event A happened in the n repetitions of the experiment.*

■ EXAMPLE 1.16

Suppose a coin is flipped 100 times and 60 of the tosses produced a "head" as a result; then the relative frequencies of the events $A :=$"the result is head" and $B :=$"the result is tail" are respectively $\frac{3}{5}$ and $\frac{2}{5}$. ▲

■ EXAMPLE 1.17

A fair die is rolled 500 times and in 83 of those tosses the number 3 was obtained. In this case the relative frequency of the event

$$A := \text{"the result obtained is 3"}$$

equals $\frac{83}{500}$. ▲

Unfortunately for each fixed A , $fr(A)$ is not constant: its value depends on n; it has been observed, however, that when a random experiment is repeated under almost the same conditions for a large number of times, the relative frequency $fr(A)$ stabilizes around a specific value between 0 and 1.

■ EXAMPLE 1.18

Suppose a die is tossed n times and let:

$$A := \text{"the result obtained is 3"}.$$

The following table summarizes the values obtained:

n	frequency	relative frequency
100	14	0.14
200	29	0.145
300	51	0.17
400	65	0.1625
500	83	0.166

▲

The stabilization of the relative frequency is known as "statistic regularity" and this is what allows us to make predictions that eliminate, though partially, the uncertainty present in unforeseeable phenomena.

The value $P(A)$ around which the relative frequency of an event stabilizes indicates its "chance" of happening. We are interested now in describing the properties that such a number should have. First, we observe that since $n(A) \geq 0$ then $P(A)$ must be greater than or equal to zero, and because

$n(\Omega) = n$, $fr(\Omega) = 1$ and therefore $P(\Omega) = 1$. Furthermore, if A and B are mutually exclusive events, then $n(A \cup B) = n(A) + n(B)$ and therefore $fr(A \cup B) = fr(A) + fr(B)$, which in turn implies that whenever $A \cap B = \emptyset$ then $P(A \cup B) = P(A) + P(B)$. These considerations lead us to state the following definition:

Definition 1.9 (Probability Space) *Let $(\Omega, \Im)$ be a measurable space. A real-valued function P defined over $\Im$ satisfying the conditions*

(i) $P(A) \geq 0$ for all $A \in \Im$ (nonnegative property)

(ii) $P(\Omega) = 1$ (normed property)

(iii) if $A_1, A_2, \cdots$ are mutually exclusive events in $\Im$, that is,

$$A_i \cap A_j = \emptyset \text{ for all } i \neq j,$$

then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \qquad \textit{(countable additivity)}$$

is called a probability measure over $(\Omega, \Im)$. The triplet $(\Omega, \Im, P)$ is called a probability space.

■ EXAMPLE 1.19

Consider Example 1.9. Let $\Omega = \{1, 2, 3\}$, $\Im = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$ and P be the following map over $\Im$ for any $A \in \Im$:

$$P(A) = \begin{cases} 1 & \text{if } 3 \in A \\ 0 & \text{if } 3 \notin A. \end{cases}$$

It is easy to verify that P is indeed a probability measure over $(\Omega, \Im)$.
▲

■ EXAMPLE 1.20

Consider Example 1.4. Let $\Omega = \{0, 1, \cdots\}$, $\Im = \wp(\Omega)$ and P be defined on $\{i\}$:

$$P(\{i\}) = (1 - q)q^i, \quad i = 0, 1, \cdots, \quad 0 < q < 1.$$

Since all three properties of Definition 1.9 are satisfied, P is a probability measure over $(\Omega, \Im)$. ▲

■ **EXAMPLE 1.21**

Let $\Omega = \{1, 2\}$, $\Im = \wp(\Omega)$ and let P be the map over $\Im$ defined by:

$$P(A) = \begin{cases} 0 & \text{if } A = \emptyset \\ \frac{1}{3} & \text{if } A = \{1\} \\ \frac{2}{3} & \text{if } A = \{2\} \\ 1 & \text{if } A = \{1, 2\}. \end{cases}$$

P is a probability measure. ▲

Next we establish the most important properties of a probability measure P.

Theorem 1.2 *Let $(\Omega, \Im, P)$ be a probability space. Then:*

1. $P(\emptyset) = 0$.
2. *If $A, B \in \Im$ and $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$.*
3. *For any $A \in \Im$, $P(A^c) = 1 - P(A)$.*
4. *If $A \subseteq B$, then $P(A) \leq P(B)$ and $P(B \setminus A) = P(B) - P(A)$. In particular $P(A) \leq 1$ for all $A \in \Im$.*
5. *For any $A, B \in \Im$, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*
6. *Let $(A_n)_n$ be an increasing sequence of elements in $\Im$, that is, $A_n \in \Im$ and $A_n \subseteq A_{n+1}$ for all $n = 1, 2, \cdots$; then*

$$P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$$

where $\lim_{n \to \infty} A_n := \bigcup_{n=1}^{\infty} A_n$.

7. *Let $(A_n)_n$ be a decreasing sequence of elements in $\Im$, that is, $A_n \in \Im$ and $A_n \supseteq A_{n+1}$ for all $n = 1, 2, \cdots$; then*

$$P(\lim_{n \to \infty} A_n) = \lim_{n \to \infty} P(A_n)$$

where $\lim_{n \to \infty} A_n := \bigcap_{n=1}^{\infty} A_n$.

Proof:

1. $1 = P(\Omega) = P(\Omega \cup \emptyset \cup \emptyset \cup \cdots) = P(\Omega) + P(\emptyset) + P(\emptyset) + \cdots$. Then $0 \geq P(\emptyset) \geq 0$ and therefore $P(\emptyset) = 0$.
2. $A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \cdots$. Thus, the proof follows from property iii from the definition of probability measure and the previous result.

3. $P(A) + P(A^c) = P(A \cup A^c) = P(\Omega) = 1$.

4. $B = A \cup (B \setminus A)$. We obtain $P(B) = P(A) + P(B \setminus A) \geq P(A)$ by applying 2.

5. As an exercise for the reader.

6. Let $C_1 = A_1,\ C_2 = A_2 \setminus A_1, \cdots, C_n = A_n \setminus A_{n-1}$. It is clear that:

$$\bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} A_n \, .$$

Furthermore, since $C_i \cap C_j = \emptyset$ for all $i \neq j$, it follows from property iii of probability measures that:

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n\right) &= P\left(\bigcup_{n=1}^{\infty} C_n\right) \\ &= \sum_{n=1}^{\infty} P(C_n) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} P(C_k) \\ &= \lim_{n\to\infty} P\left(\bigcup_{k=1}^{n} C_k\right) \\ &= \lim_{n\to\infty} P(A_n) \, . \end{aligned}$$

7. Left as an exercise for the reader.

■

Note 1.1 *Let A, B and C be events. Applying the previous theorem:*

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\ &= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) \\ &\qquad - P(A \cap C) + P(A \cap B \cap C). \end{aligned}$$

An inductive argument can be used to see that if $A_1, A_2, \cdots, A_n$ are events, then

$$\begin{aligned} &P(A_1 \cup A_2 \cup \cdots \cup A_n) \\ &= \sum_{i=1}^{n} P(A_i) - \sum_{i_1 < i_2} P(A_{i_1} \cap A_{i_2}) + \cdots + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r}) \\ &\qquad + \cdots + (-1)^{n+1} P(A_1 \cap A_2 \cap \cdots \cap A_n) \end{aligned}$$

where the sum

$$\sum_{i_1<i_2<\cdots<i_r} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_r})$$

is taken over all possible subsets of size r *of the set* $\{1, 2, \cdots, n\}$.

Note 1.2 *Let* $(\Omega, \Im, P)$ *be a probability space with finite or countable* Ω *and* $\Im = \wp(\Omega)$. *Let* $\emptyset \neq A \in \Im$. *It is clear that*

$$A = \bigcup_{\omega \in A} \{\omega\}$$

and therefore

$$P(A) = \sum_{\omega \in A} P(\omega)$$

where $P(\omega) := P(\{\omega\})$. *That is,* P *is completely determined by* $p_j := P(\omega_j)$, *where* ω_j *with* $j = 1, 2, \cdots$ *denote the different elements of* Ω.

Clearly, the $|\Omega|$*-dimensional vector* $p := (p_1, p_2, \cdots)$ *(where* $|\Omega|$ *is the number of elements of* Ω*) satisfies the following conditions:*

(i) $p_j \geq 0$.

(ii) $\sum_{j=1}^{\infty} p_j = 1$.

A vector p *satisfying the above conditions is called a probability vector.*

Note 1.3 *Let* $\Omega = \{\omega_1, \omega_2, \cdots\}$ *be a (nonempty) finite or countable set,* $\wp(\Omega)$ *the total* σ*-algebra over* Ω *and* p *a* $|\Omega|$*-dimensional probability vector. It is easy to verify that the mapping* P *defined over* $\wp(\Omega)$ *by*

$$\begin{aligned} P(\emptyset) &= 0 \\ P(\omega_j) &= p_j \ , \ j = 1, 2, \cdots \\ P(A) &= \sum_{\{j \,:\, \omega_j \in A\}} p_j \quad \text{for } \emptyset \neq A \subseteq \Omega \end{aligned}$$

is a probability measure. The probability space $(\Omega, \wp(\Omega), P)$ *obtained in this fashion is called a discrete probability space.*

■ EXAMPLE 1.22

Let $(\Omega, \Im, P)$ be a probability space with:

$$\begin{aligned} \Omega &= \{1, 2, 3, 4\} \\ \Im &= \{\emptyset, \Omega, \{1\}, \{2, 3\}, \{4\}, \{1, 2, 3\}, \{2, 3, 4\}, \{1, 4\}\} \end{aligned}$$

$$P(\{1\}) = \frac{1}{4}, \quad P(\{2, 3\}) = \frac{1}{2}, \qquad P(\{4\}) = \frac{1}{4}.$$

Then:

$$P(\{1,2,3\}) = \frac{3}{4}$$
$$P(\{2,3,4\}) = \frac{3}{4}$$
$$P(\{1,4\}) = \frac{1}{2}. \qquad \blacktriangle$$

■ EXAMPLE 1.23

Let $(\Omega, \wp(\Omega), P)$ be a discrete probability space with $\Omega = \{a, b, c\}$ and P given by the probability vector $p = (\frac{1}{7}, \frac{4}{7}, \frac{2}{7})$. Then:

$$P(\{a,b\}) = \frac{5}{7}, \quad P(\{b,c\}) = \frac{6}{7}, \qquad P(\{a,c\}) = \frac{3}{7}. \qquad \blacktriangle$$

■ EXAMPLE 1.24

Let $(\Omega, \Im, P)$ be a probability space. If A and B are events such that $P(A) = p$, $P(B) = q$ and $P(A \cup B) = r$, then:

$$P(A \cap B) = p + q - r$$
$$P(A \setminus B) = r - q$$
$$P(A^c \cap B^c) = 1 - r$$
$$P(A \cup B^c) = p - r + 1\ . \qquad \blacktriangle$$

■ EXAMPLE 1.25

Consider three wireless service providers Vodafone, Aircel, and Reliance mobile in Delhi. For a randomly chosen location in this city, the probability of coverage for the Vodafone (V), Aircel(A), and Reliance mobile (R) are $P(V) = 0.52, P(A) = 0.51, P(R) = 0.36$, respectively. We also know that $P(V \cup A) = 0.84$, $P(A \cup R) = 0.76$ and $P(A \cap R \cap V) = 0.02$. What is the probability of not having coverage from Reliance mobile? Aircel claims it has better coverage than Vodafone. Can you verify this? If you own two cell phones, one from Vodafone and one from Aircel, what is your worst case coverage?

Solution: Given:

$$\begin{aligned} P(V) &= 0.52 \\ P(A) &= 0.51 \\ P(R) &= 0.36 \\ P(V \cup A) &= 0.84 \\ P(A \cup R) &= 0.84 \\ P(A \cap R \cap V) &= 0.02. \end{aligned}$$

Using the above information, the probability of not having coverage from Reliance mobile is:

$$P(\overline{R}) = 1 - P(R) = 1 - 0.36 = 0.64.$$

Aircel's claim is incorrect as $P(A) < P(V)$. The worst case coverage is:

$$\begin{aligned} P(\overline{A} \cap \overline{V}) &= 1 - P(A \cup V) \\ &= 1 - 0.84 \\ &= 0.16. \quad \blacktriangle \end{aligned}$$

EXAMPLE 1.26

Let $(\Omega, \Im, P)$ be a probability space, and let A and B be elements of $\Im$ with $P(A) = \frac{1}{2}$ and $P(B) = \frac{2}{3}$. Then

$$\frac{1}{6} \leq P(A \cap B) \leq \frac{1}{2}$$

since $P(A \cap B) \leq P(A) = \frac{1}{2}$ and $P(A \cup B) \leq 1$. ▲

EXAMPLE 1.27

A biased die is tossed once. Suppose that:

j	1	2	3	4	5	6
p_j	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{7}{32}$	$\frac{3}{32}$	$\frac{8}{32}$	$\frac{8}{32}$

Then, the probability of obtaining a number not divisible by 3 and whose square is smaller than 20 equals $\frac{9}{32}$, while the probability of getting a number i such that $|i - 5| \leq 3$ equals $\frac{31}{32}$. ▲

1.2 LAPLACE PROBABILITY SPACE

Among random experiments, the easiest to analyze are those with a finite number of possible results with each of them having the same likelihood. These experiments are called Laplace experiments. The tossing of a fair coin or a fair die a finite number of times is a classic example of Laplace experiments.

Definition 1.10 (Laplace Probability Space) *A probability space $(\Omega, \Im, P)$ with finite Ω, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{|\Omega|}$ for all $\omega \in \Omega$ is called a Laplace probability space. The probability measure P is called the uniform or classic distribution on Ω.*

Note 1.4 *If $(\Omega, \Im, P)$ is a Laplace probability space and $A \subseteq \Omega$, then:*

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} \frac{1}{|\Omega|} = \frac{|A|}{|\Omega|}.$$

In other words:

$$P(A) = \frac{\text{“number of cases favorable to } A\text{”}}{\text{“number of possible cases”}}.$$

This last expression is in no way a definition of probability, but only a consequence of assuming every outcome of the experiment to be equally likely and a finite number of possible results.

Thereby, in a Laplace probability space we have that probability calculus is reduced to counting the elements of a finite set, that is, we arrive to a combinatorial analysis problem. For readers not familiarized with this topic, Appendix B covers the basic concepts and results of this theory.

■ **EXAMPLE 1.28**

In a certain lottery six numbers are chosen from 1 to 49. The probability that the numbers chosen are $1, 2, 3, 4, 5$ and 6 equals:

$$\frac{1}{\binom{49}{6}} = 7.1511 \times 10^{-8}.$$

Observe that this is the same probability that the numbers $4, 23, 24, 35, 40$ and 45 have been chosen.

The probability p of 44 being one of the numbers chosen equals:

$$p = \frac{\binom{48}{5}}{\binom{49}{6}} = 0.12245. \qquad \blacktriangle$$

■ EXAMPLE 1.29

There are five couples sitting randomly at a round table. The probability p of two particular members of a couple sitting together equals:

$$p = \frac{2!8!}{9!} = \frac{2}{9}. \qquad \blacktriangle$$

■ EXAMPLE 1.30

In an electronics repair shop there are 10 TVs to be repaired, 3 of which are from brand A, 3 from brand B and 4 from brand C. The order in which the TVs are repaired is random. The probability p_1 that a TV from brand A will be the first one to be repaired equals:

$$p_1 = \frac{3 \cdot 9!}{10!} = 0.3.$$

The probability p_2 that all three TVs from the brand A will be repaired first equals:

$$p_2 = \frac{3 \cdot 2 \cdot 7!}{10!} = \frac{1}{120}.$$

The probability p_3 that the TVs will be repaired in the order $CABCABCABC$ equals:

$$p_3 = \frac{4 \cdot 3^3 \cdot 2^3}{10!} = \frac{1}{4200}. \qquad \blacktriangle$$

■ EXAMPLE 1.31

In a bridge game, the whole pack of 52 cards is dealt out to four players. We wish to find the probability that a player receives all 13 spades.

In this case, the total number of ways in which the pack can be dealt out is

$$\binom{52}{13, 13, 13, 13} = \binom{52}{13}\binom{39}{13}\binom{26}{13}\binom{13}{13}$$

and the total number of ways to divide the pack while giving a single player all spades equals:

$$\binom{4}{1}\binom{39}{13}\binom{26}{13}\binom{13}{13}.$$

Therefore, the probability p we look for is given by:

$$p = \frac{4}{\binom{52}{13}} = 6.2991 \times 10^{-12}. \qquad \blacktriangle$$

EXAMPLE 1.32

Suppose that all 365 days of the year are equally likely to be the day a person celebrates his or her birthday (we are ignoring leap years and the fact that birth rates aren't uniform throughout the year). The probability p that, in a group of 50 people, no two of them have the same birthday is:

$$\begin{aligned} p &= \frac{365 \times 364 \times \cdots \times (365 - 50 + 1)}{365^{50}} \\ &= \left(1 - \frac{1}{365}\right)\left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{49}{365}\right). \end{aligned} \quad \blacktriangle$$

EXAMPLE 1.33 Urn Models

An urn has N balls of the same type, R of them are red color and $N - R$ are white color. n balls are randomly drawn from the urn. We wish to find the probability that exactly $k \leq n$ of the balls drawn are red in color.

To simplify the argument, it will be assumed that the balls are numbered from 1 to N in such a way that the red balls are all numbered from 1 to R. We distinguish between two important cases: draw without replacement and draw with replacement. In the first case we must also consider two more alternatives: the balls are drawn one by one and the balls are drawn at the same time.

1. Draw without replacement (one by one): The n balls are extracted one by one from the urn and left outside of it. In this case the sample space is given by:

$$\Omega = \{(a_1, a_2, \cdots, a_n) : a_j \in \{1, 2, \cdots, N\}, a_i \neq a_j \text{ for all } i \neq j,\ j = 1, 2, \cdots, n\}.$$

Let:

$$A_k := \text{“exactly } k \leq n \text{ of the balls drawn are red”}.$$

Clearly A_k is made of all the n-tuples from Ω with exactly k components less than or equal to R. Therefore

$$|\Omega| = N \times (N - 1) \times \cdots \times (N - (n - 1)) =: (N)_n$$

and

$$|A_k| = \binom{n}{k} R \times (R-1) \times \cdots \times (R-k+1) \times (N-R) \times$$
$$\cdots \times (N-R-n+k+1)$$
$$= \binom{n}{k} (R)_k \times (N-R)_{(n-k)}.$$

Then:

$$P(A_k) = \frac{|A_k|}{|\Omega|} = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

2. Draw without replacement (at the same time): In this case the sample space is:
$$\Omega = \{T : T \subseteq \{1, 2, \cdots, N\} \text{ with } |T| = n\}.$$

   Here A_k consists of all the elements of Ω having exactly k elements less than or equal to R. Therefore:

$$|\Omega| = \binom{N}{n} \quad \text{and} \quad |A_k| = \binom{R}{k}\binom{N-R}{n-k}.$$

   Thus:

$$P(A_k) = \frac{|A_k|}{|\Omega|} = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

   As it can be seen, when the balls are drawn without replacement, it is irrelevant for the calculus of the probability whether the balls were extracted one by one or all at the same time.

3. Draw with replacement: In this case, each extracted ball is returned to the urn, and after mixing the balls, a new one is randomly drawn. The sample space is then given by:

$$\Omega = \{(a_1, a_2, \cdots, a_n) : a_j \in \{1, 2, \cdots, N\},\ j = 1, 2, \cdots, N\}.$$

   The event A_k consists of all n-tuples from Ω with k components less than or equal to R. Then,

$$|\Omega| = N^n \quad \text{and} \quad |A_k| = \binom{n}{k} R^k (N-R)^{n-k}$$

   and accordingly:

$$P(A_k) = \frac{|A_k|}{|\Omega|} = \binom{n}{k} p^k q^{n-k} \quad \text{where } p = \frac{R}{N} \text{ and } q = 1 - p. \qquad \blacktriangle$$

■ EXAMPLE 1.34

A rectangular box contains 4 Toblerones, 8 Cadburys and 5 Perks chocolates. A sample of size 6 is selected at random without replacement. Find the probability that the sample contains 2 Toblerones, 3 Cadbury and 1 Perk chocolates.
Solution: $|\Omega| = \binom{17}{6}$ where Ω is the set of possible outcomes. $|E| = \binom{4}{2}\binom{8}{3}\binom{5}{1}$ where E is the event of interest:

$$P|E| = \frac{\binom{4}{2}\binom{8}{3}\binom{5}{1}}{\binom{17}{6}}.$$

If in the above problem the sample is to be drawn with replacement, then the required probability is:

$$\frac{\frac{(4+2-1)!}{2!} \times \frac{(8+3-1)!}{3!} \times \frac{(5+1-1)!}{1!}}{\frac{(17+6-1)!}{6!}} = \frac{\frac{5!}{2!} \times \frac{10!}{3!} \times \frac{5!}{1!}}{\frac{22!}{6!}}. \quad ▲$$

■ EXAMPLE 1.35

(Hoel et al., 1971) Suppose that n balls are distributed in n urns in such a way that all the n^n possible arrangements are equally likely. Find the probability that only the first urn is empty.
Solution: Let A be the event of having only the first urn empty. This event happens only if the n balls are distributed in the $n-1$ remaining urns in such a way that none of them are empty. That means one of those $n-1$ urns must contain exactly two balls while the other $n-2$ must have one ball each. For $j = 2, \cdots, n$, let B_j be the event of having two balls in the urn j and exactly one ball in each of the other $n-2$ urns. Clearly the events B_j are mutually exclusive and their union yields the event A. To calculate $P(B_j)$, we observe that the two balls placed in the urn j can be chosen in $\binom{n}{2}$ ways and the remaining $n-2$ balls can be distributed in the remaining urns in $(n-2)!$ ways. Then,

$$P(B_j) = \frac{\binom{n}{2}(n-2)!}{n^n}$$

and therefore:

$$\begin{aligned} P(A) &= \sum_{j=2}^{n} P(B_j) \\ &= \frac{(n-1)\binom{n}{2}(n-2)!}{n^n} \\ &= \frac{(n-1)!\binom{n}{2}}{n^n}. \quad ▲ \end{aligned}$$

■ **EXAMPLE 1.36**

A wooden cube with painted faces is sawed up into 1000 little cubes all of the same size. The little cubes are then mixed up, and one is chosen at random. Find the probability that the selected cube has only two painted faces.

Solution: The probability that the selected cube has only two painted faces is:

$$\frac{12 \times 8}{1000} = \frac{96}{1000} = 0.096. \qquad \blacktriangle$$

1.3 CONDITIONAL PROBABILITY AND EVENT INDEPENDENCE

Many times, partial information about a random experiment can be obtained before its actual result is known. Resting upon this information, it is common to change the random experiment result's probabilistic structure. For example, a poker player can at some point peek at a rival's cards. Suppose he only managed to see that all cards had a red suit, that is, hearts or diamonds. Then our player knows his partner can't have all four kings, an event that previously had a positive probability. On the other hand, he suspects that the event "having all cards of the same suit" is more likely than before getting the extra information.

Following, we are going to analyze the situation from the relative frequencies perspective. Let B be an event whose chance of happening must be measured under the assumption that another event A has been observed. If the experiment is then repeated n times under the same circumstances, then the relative frequency of B under the condition A is defined as

$$rf(B \mid A) := \frac{n(A \cap B)}{n(A)} \quad \text{if } n(A) > 0$$

where $n(A \cap B)$ indicates the number of favorable cases to $A \cap B$.

It is clear that $rf(B \mid A)$ depends on n. However, when the experiment is performed for a large enough number of times, the relative frequencies tend to stabilize around a specific value between 0 and 1, known as the conditional probability of the event B under the condition A.

We observe that:

$$rf(B \mid A) = \frac{\frac{n(A\cap B)}{n}}{\frac{n(A)}{n}} = \frac{rf(A \cap B)}{rf(A)} \quad \text{if } n(A) > 0.$$

For large enough n, the numerator from the former expression tends to $P(A \cap B)$ while the denominator tends to $P(A)$. This motivates the following definition:

Definition 1.11 (Conditional Probability) *Let $(\Omega, \Im, P)$ be a probability space. If $A, B \in \Im$ with $P(A) > 0$, then the probability of the event B under the condition A is defined as follows:*

$$P(B \mid A) := \frac{P(A \cap B)}{P(A)}.$$

■ EXAMPLE 1.37

Two fair dice are rolled once. The probability that at least one of the results is 6 given that the results obtained are different equals $\frac{1}{3}$, as the following reasoning shows: Let A be the event "the results are different" and B the event "At least one of the results is 6". It is clear that:

$$\begin{aligned} A &= \{(a,b) : a, b \in \{1,2,\cdots,6\},\ a \neq b\} \quad \text{and} \\ B &= \{(a,6) : a \in \{1,2,\cdots,6\}\} \cup \{(6,b) : b \in \{1,2,\cdots,6\}\}. \end{aligned}$$

Then:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{10}{36}}{\frac{30}{36}} = \frac{1}{3}. \qquad \blacktriangle$$

■ EXAMPLE 1.38

An urn contains 12 balls, 8 of which are white color. A sample of size 4 is taken without replacement. Then, the probability that the first and third balls extracted are white given that our sample contains three white balls equals $\frac{1}{2}$. To this effect, let's assume the balls are numbered from 1 to 12; then:

$$\Omega = \{(a_1, a_2, a_3, a_4) : a_i \in \{1, 2, \ldots, 12\},\ \ a_i \neq a_j \ \text{ for all } \ i \neq j\}.$$

Let:

$$\begin{aligned} A &:= \text{“Exactly three of the balls extracted are white”}. \\ B &:= \text{“The first and the third balls removed are white”}. \end{aligned}$$

It is straightforward that:

$$P(B \mid A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{n(A \cap B)}{n(\Omega)}}{\frac{n(A)}{n(\Omega)}} = \frac{n(A \cap B)}{n(A)} = \frac{\binom{2}{1} 8.6.7.4}{\binom{4}{3} 4.8.7.6} = \frac{1}{2}. \qquad \blacktriangle$$

■ EXAMPLE 1.39

Let $(\Omega, \Im, P)$ be a probability space with

$$\Omega = \{a, b, c, d, e, f\}, \ \Im = \wp(\Omega)$$

and

ω	a	b	c	d	e	f
$P(\omega)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{3}{16}$	$\frac{1}{4}$	$\frac{5}{16}$

Let $A = \{a, c, e\}$, $B = \{c, d, e, f\}$ and $C = \{b, c, f\}$. Then:

$$\begin{aligned} P(A \mid B^c \cup C) &= \frac{P(A \cap (B^c \cup C))}{P(B^c \cup C)} \\ &= \frac{P(\{a, c\})}{P(\{a, b, c, f\})} \\ &= \frac{1}{3}. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 1.40

Consider the flights starting from Bogotá to Medellín. In these flights, 90% leave on time and arrive on time, 6% leave on time and arrive late, 1% leave late and arrive on time and 3% leave late and arrive late. What is the probability that, given a flight leaves late, it will arrive on time?
solution: Given:

$$\begin{aligned} P(\text{Flight leaves on time and arrives on time}) &= 0.9 \\ P(\text{Flight leaves on time and arrives late}) &= 0.06 \\ P(\text{Flight leaves late and arrives on time}) &= 0.01 \\ P(\text{Flight leaves late and arrives late}) &= 0.03. \end{aligned}$$

Now:

$$\begin{aligned} P(\text{Flight leaves late}) &= P(\text{Flight leaves late and arrives on time}) + \\ &\quad P(\text{Flight leaves late and arrives late}) \\ &= 0.01 + 0.03 \\ &= 0.04. \end{aligned}$$

Therefore:

$$\begin{aligned} P(\text{Flight arrives on time} \mid \text{it leaves late}) &= \frac{P(\text{Flight leaves late and arrives on time})}{P(\text{Flight leaves late})} \\ &= \frac{0.01}{0.04} \\ &= 0.25. \quad \blacktriangle \end{aligned}$$

The next theorem provides us with the main properties of conditional probability:

Theorem 1.3 (Conditional Probability Measure) *Let $(\Omega, \Im, P)$ be a probability space, and let $A \in \Im$ with $P(A) > 0$. Then:*

1. *$P(\,\cdot \mid A)$ is a probability measure over Ω centered on A, that is, $P(A \mid A) = 1$.*
2. *If $A \cap B = \emptyset$, then $P(B \mid A) = 0$.*
3. *$P(B \cap C \mid A) = P(B \mid A \cap C)P(C \mid A)$ if $P(A \cap C) > 0$.*
4. *If $A_1, A_2, \cdots, A_n \in \Im$ with $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$, then*

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 \cap A_2) \cdots \\ P(A_n \mid A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \end{aligned}$$

Proof:

1. The three properties of a probability measure must be verified:
   (i) Clearly $P(B \mid A) \geq 0$ for all $B \in \Im$.
   (ii) $P(\Omega \mid A) = \frac{P(A \cap \Omega)}{P(A)} = \frac{P(A)}{P(A)} = 1$. Therefore, we also have that $P(A \mid A) = 1$.
   (iii) Let $A_1, A_2, \cdots$ be a sequence of disjoint elements from $\Im$. Then:

$$\begin{aligned} P\left(\bigcup_{i=1}^{\infty} A_i \mid A\right) &= \frac{P\left(A \cap \left(\bigcup_{i=1}^{\infty} A_i\right)\right)}{P(A)} \\ &= \frac{P\left(\bigcup_{i=1}^{\infty} (A_i \cap A)\right)}{P(A)} \\ &= \frac{\sum_{i=1}^{\infty} P(A_i \cap A)}{P(A)} \\ &= \sum_{i=1}^{\infty} P(A_i \mid A)\,. \end{aligned}$$

2. Left as an exercise for the reader.

3.

$$\begin{aligned} P(B \cap C \mid A) &= \frac{P(B \cap C \cap A)}{P(A)} \\ &= \frac{P(B \cap C \cap A)}{P(C \cap A)} \times \frac{P(C \cap A)}{P(A)} \\ &= P(B \mid A \cap C) P(C \mid A). \end{aligned}$$

4. Left as an exercise for the reader.

■

■ EXAMPLE 1.41

An urn contains 12 balls, 4 of which are black while the remaining 8 are white. The following game is played: the first ball is randomly extracted and, after taking note of its color, it is then returned to the urn along with a new pair of balls of the same color. Find the probability that in the first three rounds of the game all the balls drawn are black.
Solution: For $i = 1, 2, 3$ we define:

$A_i :=$ "a black ball was extracted on the ith round of the game".

Clearly:

$$\begin{aligned} P(A_1 \cap A_2 \cap A_3) &= P(A_3 \mid A_2 \cap A_1) P(A_2 \mid A_1) P(A_1) \\ &= \frac{8}{16} \times \frac{6}{14} \times \frac{4}{12} \\ &= \frac{1}{14}. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 1.42

Three teenagers want to get into an R-rated movie. At the box office, they are asked to produce their IDs; after the clerk checks them and denies them the entrance, he returns the IDs randomly. Find the probability that none of the teenagers get their own ID.
Solution: Let:

$A :=$ "None of the teenagers get their own ID".

$B_i :=$ "The ith teenager gets his own ID".

Clearly, the probability we look for is:

$$\begin{aligned} P(A) &= P(B_1^c \cap B_2^c \cap B_3^c) \\ &= 1 - P(B_1 \cup B_2 \cup B_3) \\ &= 1 - P(B_1) - P(B_2) - P(B_3) + P(B_1 \cap B_2) \\ &\quad + P(B_2 \cap B_3) + P(B_1 \cap B_3) - P(B_1 \cap B_2 \cap B_3). \end{aligned}$$

Since there are three possible cases and only one is favorable:

$$P(B_i) = \frac{1}{3} \quad \text{for} \quad i = 1, 2, 3.$$

On the other hand, for any $i \neq j$,

$$P(B_i \cap B_j) = P(B_i)P(B_j \mid B_i) = \frac{1}{3} \times \frac{1}{2} = \frac{1}{6}$$

seeing that, after giving the ith teenager the right ID, for the jth teenager there is only one favorable option from two possible ones. In a similar fashion:

$$\begin{aligned} P(B_1 \cap B_2 \cap B_3) &= P(B_1)P(B_2 \mid B_1)P(B_3 \mid B_1 \cap B_2) \\ &= \frac{1}{3} \times \frac{1}{2} \times 1 = \frac{1}{6}. \end{aligned}$$

Therefore:

$$P(A) = \frac{1}{3}. \qquad \blacktriangle$$

The reliability of a device or its separate units is understood as the probability of their trouble-free operation without failure.

■ EXAMPLE 1.43

Let us consider a system composed of n units. It is assumed that unit fails independently of another unit. A series system is one in which all units must operate successfully. On the other had, a parallel system is one that will fail only if all its units fail. Let reliability of each unit be p.

Then, the reliability of series system R_1 is given by:

$$R_1 = p^n.$$

Similarly, the reliability of parallel system R_2 is given by:

$$R_2 = 1 - (1 - p)^n. \qquad \blacktriangle$$

In the following example, we illustrate the reliability of a non-series/parallel system.

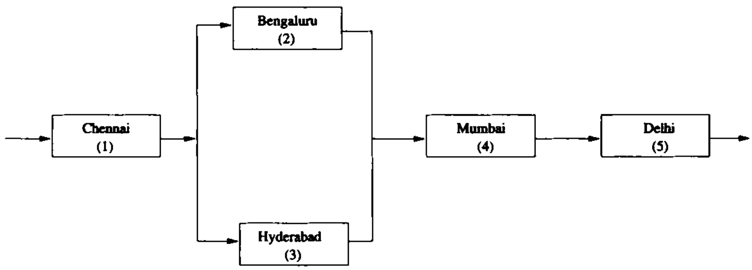


Figure 1.1 A communication network system

■ EXAMPLE 1.44

Consider a communication network system in India with five network switches placed as shown in Figure 1.1. Suppose that the probability that each network switch will perform a required function without failure, i.e., reliability of each network switch, is $R_1 = 0.98, R_2 = 0.99, R_3 = 0.99, R_4 = 0.96$ and $R_5 = 0.95$. Assume that each network switch is functioning independently. Find the reliability of the communication network system.

Solution: For $i = 1, 2, \cdots, 5$ let:

$$A_i := \text{"The } i\text{th network switch functioning"}.$$

We define:

$$R_i = P(A_i) \quad \text{for} \quad i = 1, 2, \cdots, 5.$$

The reliability of the communication network system is:

$$\begin{aligned} R &= P\left(A_1 \cap (A_2 \cup A_3) \cap A_4 \cap A_5\right) \\ &= P(A_1)P\left(A_2 \cup A_3\right) P(A_4)P(A_5) \\ &= R_1 \left(R_2 + R_3 - R_2R_3\right) R_4R_5 \\ &= 0.98 \times 0.9999 \times 0.96 \times 0.95 \\ &= 0.8937. \quad \blacktriangle \end{aligned}$$

The following results are vital for applications:

Theorem 1.4 (Total Probability Theorem) *Let $A_1, A_2, \cdots,$ be a finite or countable partition of Ω, that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $\bigcup_{i=1}^{\infty} A_i = \Omega$; such that $P(A_i) > 0$ for all $A_i \in \Im$. Then, for any $B \in \Im$:*

$$P(B) = \sum_i P(B \mid A_i)P(A_i).$$

Proof: We observe that

$$\begin{aligned} B &= B \cap \Omega \\ &= B \cap \cup_{i=1}^{\infty} A_i \\ &= \cup_{i=1}^{\infty} (B \cap A_i) \end{aligned}$$

and hence,

$$\begin{aligned} P(B) &= \sum_i P(B \cap A_i) \\ &= \sum_i P(B \mid A_i) P(A_i), \end{aligned}$$

which proves the theorem. ∎

As a corollary to the previous theorem we obtain a result known as Bayes' rule, which constitutes the base for an important statistical theory called Bayesian theory.

Corollary 1.1 (Bayes' Rule) *Let $A_1, A_2, \cdots$ be a finite or countable partition of Ω with $P(A_i) > 0$ for all i; then, for any $B \in \Im$ with $P(B) > 0$:*

$$P(A_i \mid B) = \frac{P(A_i)P(B \mid A_i)}{\sum\limits_j P(B \mid A_j)P(A_j)} \quad \textit{for all } i.$$

Proof:

$$\begin{aligned} P(A_i|B) &= \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{P(B)} \\ &= \frac{P(A_i)P(B|A_i)}{\sum\limits_j P(B|A_j)P(A_j)} \,. \end{aligned}$$

∎

To give an interpretation of Bayes' rule, suppose that the events $A_1, A_2, \cdots$ are all possible causes, mutually exclusive, of a certain event B. Under the assumption that we have indeed observed the event B, Bayes' formula allows us to know which of these causes is most likely to have produced the event B.

■ **EXAMPLE 1.45**

Mr. Rodríguez knows that there is a chance of 40% that the company he works with will open a branch office in Montevideo (Uruguay). If that happens, the probability that he will be appointed as the manager

in that branch office is 80%. If not, the probability that Mr. Rodríguez will be promoted as a manager to another office is only 10%. Find the probability that Mr. Rodríguez will be appointed as the manager of a branch office from his company.
Solution: Let:

$M :=$ "Mr. Rodríguez is appointed as a manager".
$N :=$ "The company opens a new branch office in Montevideo".

Then:

$$\begin{aligned} P(M) &= P(M \mid N)P(N) + P(M \mid N^c)P(N^c) \\ &= 0.8 \times 0.4 + 0.10 \times 0.60 \\ &= 0.38. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 1.46

In the previous example, if we know that Mr. Rodríguez was indeed appointed manager of an office from the company he works for, what is the probability that the company opened a new office in Montevideo?
Solution: From Bayes' rule, is clear that:

$$\begin{aligned} P(N \mid M) &= \frac{P(N)P(M \mid N)}{P(M)} \\ &= \frac{0.4 \times 0.8}{0.38} \\ &= 0.84211. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 1.47

A signal can be green or red with probability $\frac{4}{5}$ or $\frac{1}{5}$, respectively. The probability that it is received correctly by a station is $\frac{3}{4}$. Of the two stations A and B, the signal is first received by A and then station A passes the signal to station B. If the signal received at station B is green, then find the probability that the original signal was green?
Solution: Let:

$B_G(B_R) :=$ "Signal received at station B is Green (Red)",
$A_G(A_R) :=$ "Signal received at station A is Green (Red)",

Then from Bayes' rule, it is clear that:

$$P(A_G \mid B_G) = \frac{P(B_G \cap A_G)}{P(B_G)}.$$

Now:

$$P(B_G \cap A_G) = \frac{4}{5} \times \frac{3}{4} \times \frac{3}{4} + \frac{4}{5} \times \frac{1}{4} \times \frac{1}{4}$$

$$P(B_G) = \frac{4}{5} \times \frac{3}{4} \times \frac{3}{4} + \frac{4}{5} \times \frac{1}{4} \times \frac{1}{4} + \frac{1}{5} \times \frac{3}{4} \times \frac{1}{4} + \frac{1}{5} \times \frac{1}{4} \times \frac{3}{4}.$$

Hence:

$$P(A_G \mid B_G) = \frac{40}{46}. \quad \blacktriangle$$

■ EXAMPLE 1.48

It is known that each of four people A, B, C, D tells the truth in a given instance with probability $\frac{1}{3}$. Suppose A makes a statement, and D says that C says that B says that A was telling the truth. What is the probability that A was actually telling the truth?

Solution: Let:

$T :=$ "A speaks the truth".

$E :=$ "D says the statement".

Then:

$$P(T) = \frac{1}{3}.$$

The required probability is $P(T|E)$. Now:

$P(E|T)$ = Probability that D speaks the truth to C and
C speaks truth to B and B speaks the truth,
$+D$ speaks truth to C, C lies to B and B lies,
$+D$ lies to, C, C speaks truth to B, B lies,
$+D$ lies to C and C lies to B and B speaks truth.

Hence:

$$\begin{aligned}
P(E|T) &= \left(\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{2}{3} \times \frac{2}{3}\right) + \left(\frac{2}{3} \times \frac{1}{3} \times \frac{2}{3}\right) + \left(\frac{2}{3} \times \frac{2}{3} \times \frac{1}{3}\right) \\
&= \frac{13}{27} \\
P(E|T^c) &= \left(\frac{2}{3} \times \frac{2}{3} \times \frac{1}{3}\right) + \left(\frac{2}{3} \times \frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{2}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{3} \times \frac{2}{3}\right) \\
&= \frac{10}{27}.
\end{aligned}$$

From Bayes' rule:

$$\begin{aligned} P(T|E) &= \frac{P(E|T)P(T)}{P(E|T)P(T)+P(E|T^c)P(T^c)} \\ &= \frac{\frac{13}{27}\times\frac{1}{3}}{\left(\frac{13}{27}\times\frac{1}{3}\right)+\left(\frac{14}{27}\times\frac{2}{3}\right)} \\ P(T|E) &= \frac{13}{13+28} \\ P(T|E) &= \frac{13}{41}. \quad \blacktriangle \end{aligned}$$

■ **EXAMPLE 1.49**

A quarter of a population is vaccinated against a certain contagious disease. During the course of an epidemic due to such disease, it is observed that from every 5 sick persons only 1 was vaccinated. It is also known that from every 12 vaccinated people, only 1 is sick. We wish to find the probability that a nonvaccinated person is sick.

Solution: Let:

$$\begin{aligned} V &:= \text{"The person is vaccinated"}. \\ S &:= \text{"The person is sick"}. \end{aligned}$$

From the information, we have:

$$\begin{aligned} P(V \mid S) &= \frac{1}{5} \\ P(S \mid V) &= \frac{1}{12} \\ P(V) &= \frac{1}{4}. \end{aligned}$$

Thus

$$\begin{aligned} x &:= P(S \mid V^c) \\ &= \frac{P(V^c \mid S)P(S)}{P(V^c)} \\ &= \frac{\frac{4}{5}\left(\frac{1}{12}\cdot\frac{1}{4}+\frac{3}{4}x\right)}{\frac{3}{4}}, \end{aligned}$$

and therefore

$$x = \frac{1}{9}. \quad \blacktriangle$$

EXAMPLE 1.50

It is known that the population of a certain city consists of 45% females and 55% males. Suppose that 70% of the males and 10% of the females smoke. Find the probability that a smoker is male.

Solution: Let S be the event that a person is a smoker, M be the event that a person is male and F be the event that a person is female:

$$\begin{aligned} P(M) &= \frac{55}{100} \\ P(F) &= \frac{45}{100} \\ P(S|M) &= \frac{70}{100}. \end{aligned}$$

The required probability is:

$$\begin{aligned} P(M|S) &= \frac{P(S|M)P(M)}{(P(S|M)P(M) + P(S|F)P(F)} \\ &= \frac{\left(\frac{70}{100}\right) \times \left(\frac{55}{100}\right)}{\left(\frac{70}{100}\right) \times \left(\frac{55}{100}\right) + \left(\frac{45}{100}\right) \times \left(\frac{10}{100}\right)} \\ &= 0.895. \quad \blacktriangle \end{aligned}$$

EXAMPLE 1.51

Dunlop tire Company produces tires which pass through an automatic testing machine. It is observed that 5% of the tires entering the testing machine are defective. However, the automatic testing machine is not entirely reliable. If a tire is defective, there is 0.04 probability that it is not to be rejected. If a tire is not defective there is 0.06 probability that it will be rejected. What is the probability that the tires rejected are actually not defective. Also, what fraction of those not rejected are defective?

Solution: Let D be the event that the tire is defective and R be the event that the tire is rejected:

$$P(D) = 0.05, \quad P(D^c) = 1 - P(D) = 0.95.$$

It is given that:

$$P(R^c \mid D) = 0.04 \text{ and } P(R \mid D^c) = 0.06.$$

Therefore:

$$\begin{aligned}
P(\text{Nondefective} \mid \text{Rejected}) &= \frac{P(D^c)\ P(R \mid D^c)}{P(D^c)\ P(R \mid D^c) + P(D)\ P(R \mid D)} \\
&= \frac{P(D^c)\ P(R \mid D^c)}{P(D^c)\ P(R \mid D^c) + P(D)\ (1 - P(R^c \mid D))} \\
&= \frac{0.95(0.06)}{0.95(0.06) + 0.05(1 - 0.04)} \\
&= 0.542.
\end{aligned}$$

Similarly:

$$\begin{aligned}
P(\text{Defective} \mid \text{Not Rejected}) &= \frac{P(D)\ P(R^c \mid D)}{P(D)\ P(R^c \mid D) + P(D^c)\ P(R^c \mid D^c)} \\
&= \frac{0.05(0.04)}{0.05(0.04) + 0.95(1 - 0.06)} \\
&= 0.002. \quad \blacktriangle
\end{aligned}$$

Definition 1.12 (A Priori and A Posteriori Distributions) *Let $A_1, A_2, \cdots$ be a finite or countable partition of Ω with $P(A_i) > 0$ for all i. If B is an element from $\Im$ with $P(B) > 0$, then $(P(A_n))_n$ is called the "a priori" distribution, that is, before B happens, and $(P(A_n \mid B))_n$ is called the "a posteriori" distribution, that is, after B has happened.*

■ EXAMPLE 1.52

In a city, tests are taken to detect a certain disease. Suppose that 1% of the healthy people are registered as sick, 0.1% of the population is actually ill and 90% of the sick are reported as such. We wish to calculate the probability that a randomly chosen person reported as ill is indeed sick.

If we define the events

$$S := \text{"the person is indeed sick"}$$
$$R := \text{"the person is reported as sick"}$$

from the information stated above, we know that:

$$\begin{aligned}
P(S) &= 0.001 \\
P(R \mid S^c) &= 0.01 \\
P(R \mid S) &= 0.9.
\end{aligned}$$

Therefore:

$$\begin{aligned} P(S \mid R) &= \frac{P(R \mid S)P(S)}{P(R \mid S^c)P(S^c) + P(R \mid S)P(S)} \\ &= \frac{0.9 \times 0.001}{0.01 \times 0.999 + 0.9 \times 0.001} \\ &= 8.2645 \times 10^{-2} \approx 0.083. \end{aligned}$$

In this case:

$$\begin{aligned} \text{A priori distribution} &= (P(S), P(S^c)) = (0.001, 0.999) \\ \text{A posteriori distribution} &= (P(S \mid R), P(S^c \mid R)) = (0.083, 0.917). \end{aligned}$$ ▲

Sometimes the occurrence of an event B does not affect the probability of an event A, that is:

$$P(A \mid B) = P(A). \tag{1.1}$$

In this case, we say the event A is "independent" from event B. The "definition" (1.1) requires the condition that $P(B) > 0$. To avoid this condition, we define independence as follows:

Definition 1.13 (Independent Events) *Two events A and B are said to be independent if and only if:*

$$P(A \cap B) = P(A)P(B) .$$

On the contrary, if the previous condition is not met, the events are said to be dependent.

■ EXAMPLE 1.53

Suppose a fair die is rolled two times. Let:

$A :=$ "The sum of the results obtained is an even number".
$B :=$ "The result from the second roll is even".

In this case:

$$P(A) = P(B) = \frac{1}{2}$$

Furthermore, $P(A \cap B) = \frac{1}{4}$. Accordingly, the events are independent.
▲

EXAMPLE 1.54

A die is biased in such a way that the probability of obtaining an even number equals $\frac{2}{5}$. Let A and B be defined as in the preceding example. Under these conditions we have:

$$P(A) = \frac{13}{25}$$

$$P(B) = \frac{2}{5}$$

$$P(A \cap B) = \frac{4}{25} \; .$$

Thus, A and B are not independent. ▲

Note 1.5 *A mistake that is commonly made is to assume that two events are independent if they are mutually exclusive. Note that this is not the case. For example, if a fair coin is flipped once and we consider the events*

$$A := \text{"the result obtained is head"}$$
$$B := \text{"the result obtained is tail"}$$

then, clearly, A and B are mutually exclusive. They are however not independent, since:

$$0 = P(A \cap B) \neq \frac{1}{4} = P(A)P(B) \; .$$

Theorem 1.5 *Let A and B be independent events. Then:*

1. *A and B^c are two independent events (and hence by symmetry A^c and B are two independent events).*

2. *A^c and B^c are two independent events.*

Proof:

1.

$$\begin{aligned} P(A) &= P(A \cap B^c) + P(A \cap B) \\ &= P(A \cap B^c) + P(A)P(B), \end{aligned}$$

Therefore:

$$P(A \cap B^c) = P(A)[1 - P(B)] = P(A)P(B^c).$$

2.

$$\begin{aligned} P(A^c \cap B^c) &= 1 - P(A \cup B) \\ &= 1 - P(A) - P(B) + P(A \cap B) \\ &= 1 - P(A) - P(B) + P(A)P(B) \\ &= [1 - P(A)][1 - P(B)] \\ &= P(A^c)P(B^c). \end{aligned}$$

■

In many cases it is necessary to analyze the independence of two or more events. In this context a broader definition of independence must be given.

Definition 1.14 (Independent Family) *A family of events $\{A_i : i \in I\}$ is said to be independent if*

$$P\left(\bigcap_{i \in J} A_i\right) = \prod_{i \in J} P(A_i)$$

for every finite subset $\emptyset \neq J$ of I. These events are mutually independent events.

Definition 1.15 (Pairwise Independent Events) *A family of events $\{A_i : i \in I\}$ is said to be pairwise (2×2) independent if:*

$$P(A_i \cap A_j) = P(A_i)P(A_j) \text{ for all } i \neq j.$$

Pairwise independence does not imply the independence of the family or mutually independent events, as the following example shows:

■ **EXAMPLE 1.55**

A fair die is rolled two consecutive times. Let A, B and C be the events defined as follows:

$$\begin{aligned} A &:= \text{"A 2 was obtained in the first toss"}. \\ B &:= \text{"A 5 was obtained in the second toss"}. \\ C &:= \text{"The sum of the results is 7"}. \end{aligned}$$

Clearly:

$$\begin{aligned} P(A) = P(B) = P(C) &= \frac{1}{6} \\ P(A \cap B) = P(A \cap C) = P(B \cap C) &= \frac{1}{36} \\ P(A \cap B \cap C) &= \frac{1}{36}. \end{aligned}$$

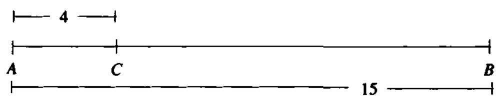


Figure 1.2 Sample and event space

The events are therefore pairwise independent, but they are however not mutually independent because:

$$P(A \cap B \cap C) \neq P(A)P(B)P(C) . \quad \blacktriangle$$

1.4 GEOMETRIC PROBABILITY

Let $(\Omega, \Im)$ be a measurable space, and assume that a geometric measure m, such as length, area or volume, is defined over $(\Omega, \Im)$. We define the geometric probability of an event A as follows:

$$P(A) := \frac{m(A)}{m(\Omega)}.$$

Next, we give some examples showing geometric probability calculus.

■ EXAMPLE 1.56

Find the probability that a point chosen at random lies on a line segment $\overline{AC}$ of a line $\overline{AB}$(see Figure 1.2).

Solution:

$$\begin{aligned} \text{Probability of the point is on } \overline{AC} &= \frac{\text{Length of } \overline{AC}}{\text{Length of } \overline{AB}} \\ &= \frac{4}{15}. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 1.57

María Victoria and Carlos agreed to meet downtown between 12 noon and 1 PM. They both get there at any moment in that time interval. Assuming their arrival times to be independent, find:

(i) The probability that Carlos and María Victoria will meet if both of them waits for the other 10 minutes at most.

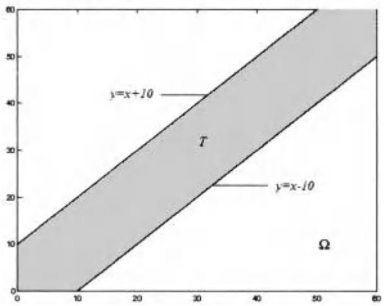


Figure 1.3 Event space for case (i)

(ii) The probability that Carlos and María Victoria will meet if María Victoria waits 5 minutes but Carlos waits 20.

Solution: (i) Let X and Y be events defined as follows:

$$X := \text{"María Victoria's arrival time"}$$
$$Y := \text{"Carlos' arrival time"}.$$

The sample space in this case is given by:

$$\Omega := \left\{(x, y) \in \mathbb{R}^2 : 0 \leq x \leq 60,\ 0 \leq y \leq 60\right\}.$$

We wish to measure the probability of the event:

$$T = \{(x, y) \in \Omega : |x - y| \leq 10\}.$$

Therefore (see Figure 1.3):

$$P(T) = \frac{\text{area of } T}{\text{area of } \Omega} = \frac{11}{36}.$$

(ii) The set of points T representing the arrival times of Carlos and María Victoria that allow them to meet is represented in Figure 1.4. Thus, the probability we wish to find equals:

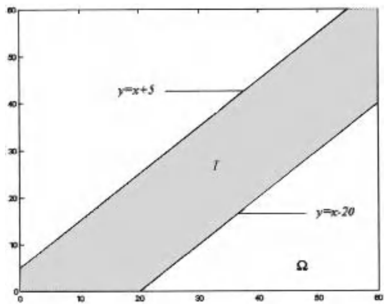


Figure 1.4 Event space for case (ii)

$$P(T) = \frac{103}{288}. \qquad \blacktriangle$$

EXERCISES

1.1 A fair coin is flipped three times in a row. Let:

$$A := \text{"The result of the first toss is head"}.$$
$$B := \text{"The result of the third toss is tail"}.$$

Describe in words the events $A \cap B$, $A \cup B$, A^c, $A^c \cap B^c$, $A \cap B^c$ and find its elements.

1.2 Let A, B and C be three arbitrary events. Give in terms of set operations of the following events:

a) A and B but not C.

b) All three of them.

c) Only A.

d) At least one of them.

e) At most one of them.

f) At most two of them.

1.3 A fair die is rolled twice in a row. Let A, B and C be the events given by:

$$A := \text{“The first result obtained is an even number”}.$$
$$B := \text{“The sum of the results is less than 7”}.$$
$$C := \text{“The second result obtained is a prime number”}.$$

List the elements belonging to the following events:

a) $A \cap B \cap C$.

b) $B \cup (A \cap C^c)$.

c) $(A \cap C) \cap [(A \cup B)^c]$.

1.4 A random experiment consists in extracting three light bulbs and classifying them as defective “D” or nondefective “N”. Consider the events:

$$A_i := \text{“The } i\text{th light bulb removed is defective”}, \quad i =.1, 2, 3.$$

a) Describe the sample space for this experiment.

b) List all the results in A_1, A_2, $A_1 \cup A_3$, $A_1^c \cap A_2^c \cap A_3$ and $(A_1 \cup A_2^c) \cap A_3$.

1.5 A worker makes n articles. The event “The ith article is defective” will be notated as A_i with $i = 1, 2, \cdots, n$. Describe the following events using the sets A_i and the usual operations between events:

a) $B :=$ “At least an article is defective”.

b) $C :=$ “None of the n articles is defective”.

c) $D :=$ “Exactly one article is defective”.

d) $E :=$ “At most one article is defective”.

1.6 Let A, B and C be arbitrary events. Depict the following events in terms of A, B and C:

a) $E_1 :=$ “At least one of the events A, B, C happens”.

b) $E_2 :=$ “Exactly two of the events A, B, C happen”.

c) $E_3 :=$ "At least two of the events A, B, C happen".

d) $E_4 :=$ "At most one of the events A, B, C happens".

1.7 Suppose that 35% of the students of a university are taking English, 7% are taking German and 2% are taking both English and German. What percentage of the student population is taking English but not German? What percentage of the students are taking neither English nor German?

1.8 Let $\Omega \neq \emptyset$ and let $\Im$ be a σ-algebra over Ω. Prove:

a) If $A_1, A_2, \cdots \in \Im$, then $\bigcap_{i=1}^{\infty} A_i \in \Im$.

b) If $A, B \in \Im$, then $A \cup B$, $A \cap B$, $A \setminus B$ and $A \triangle B$ all belong to $\Im$.

1.9 Let $\Omega = \{1,2,3,4\}$. Find four different σ-algebras $\{\Im_n\}$ for $n = 1,2,3,4$ such that $\Im_1 \subset \Im_2 \subset \Im_3 \subset \Im_4$.

1.10 Let Ω and $\widetilde{\Omega}$ be nonempty sets and $\widetilde{\Im}$ a σ-algebra over $\widetilde{\Omega}$. If $T : \Omega \to \widetilde{\Omega}$ is a function, then prove that the collection $T^{-1}(\widetilde{\Im}) = \{T^{-1}(A) : A \in \widetilde{\Im}\}$ is a σ-algebra over Ω.

1.11 Let $\Omega \neq \emptyset$ and $\overline{\Omega}$ be a nonempty subset of Ω. If $\Im$ is a σ-algebra over Ω, then prove that $\overline{\Im} = \{A \cap \overline{\Omega} : A \in \Im\}$ is a σ-algebra over $\overline{\Omega}$ called the trace of $\Im$ on $\overline{\Omega}$.

1.12 Let $A_1, A_2, \cdots, A_n$ be events in the probability space $(\Omega, \Im, P)$. Prove that:

$$P(\overset{n}{\underset{i=1}{\cup}} A_i) \leq \sum_{i=1}^{n} P(A_i).$$

1.13

a) Find the σ-algebra over $\Omega = \{1,2,3\}$ generated by $\{\{2\},\{3\}\}$.

b) Let $\mathcal{C}$ and $\mathcal{D}$ be two families of subsets from a nonempty set Ω with $\mathcal{C} \subseteq \mathcal{D}$. Is $\sigma(\mathcal{C}) \subseteq \sigma(\mathcal{D})$? Explain.

c) Let $\Omega = \{1,2\}$, $\Im = \{\emptyset, \Omega, \{1\}, \{2\}\}$ and μ defined over $\Im$ by:

$$\begin{aligned} \mu(\emptyset) &= 0 \\ \mu(\Omega) &= 1 \\ \mu(\{1\}) &= \frac{1}{3} \\ \mu(\{2\}) &= \frac{2}{3}. \end{aligned}$$

Is $(\Omega, \Im, \mu)$ a probability space? Explain.

1.14 Let $\Omega = \{a, b, c, d\}$, $\Im = \{\emptyset, \Omega, \{a\}, \{b, c\}, \{d\}, \{a, b, c\}, \{b, c, d\}, \{a, d\}\}$ and P a map of $\Im$ on $[0, 1]$ with $P(\{a\}) = \frac{2}{7}$, $P(\{b, c\}) = \frac{4}{9}$ and $P(\{d\}) = \alpha$.

a) Determine the value that α should take in order for P to be a probability measure over $(\Omega, \Im)$.

b) Find $P(\{a, b, c\})$, $P(\{b, c, d\})$ and $P(\{a, d\})$.

1.15 Let $(P_n)_{n\in\mathbb{N}}$ be a sequence of probability measures over a measurable space $(\Omega, \Im)$, $(\alpha_n)_{n\in\mathbb{N}}$ a sequence of nonnegative real numbers such that $\sum_{n=1}^{\infty} \alpha_n = 1$ and $P : \Im \to \mathbb{R}$ defined by:

$$P(A) := \sum_{n\in\mathbb{N}} \alpha_n P_n(A) \qquad \text{for all } A \in \Im.$$

Prove that P is a probability measure over $(\Omega, \Im)$.

1.16 Indicate whether the following statements are true or false. Give a brief account of your choice:

a) If $P(A) = 0$, then $A = \emptyset$.

b) If $P(A) = P(B) = 0$, then $P(A \cup B) = 0$.

c) If $P(A) = \frac{1}{2}$ and $P(B) = \frac{1}{3}$, then $\frac{1}{2} \leq P(A \cup B) \leq \frac{5}{6}$.

d) If $P(A) = P(B) = p$, then $P(A \cap B) \leq p^2$.

e) $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$.

f) If $P(A) = 0.5$, $P(B) = 0.4$ and $P(A \cup B) = 0.8$, then $P(A^c \cap B) = 0.1$

g) If A and B are independent events and $A \subset B$, then $P(A) = 0$ or $P(B) = 1$.

1.17 Prove that $P(A \cup B \cup C) = P(A) + P(A^c \cap B) + P(A^c \cap B^c \cap C)$.

1.18 Let A and B be two events with $P(A) = \frac{1}{2}$ and $P(B^c) = \frac{1}{4}$. Can A and B be mutually exclusive events? Justify.

1.19 A die is biased in such a way that the probability of getting an even number is twice that of an odd number. What is the probability of obtaining an even number? A prime number? An odd prime number?

1.20 From the 100 students majoring in philology and classic languages in the linguistics department of a certain University 28 take Latin classes, 26 take Greek, 16 Hebrew, 12 both Latin and Greek, 4 Latin and Hebrew and 6 Greek and Hebrew. Furthermore, 2 are taking all the aforementioned subjects.

a) If a philology and classic languages student is randomly chosen, what is the probability that he or she is taking only Hebrew?

b) If a philology and classic languages student is randomly chosen, what is the probability that he or she is taking Greek and Hebrew but not Latin?

c) If two philology and classic languages students are randomly chosen, what is the probability that at least one of them is attending one of the classes?

1.21 A company was hired to poll the 1000 subscribers of a magazine. The data presented on their report indicate that 550 subscribers are professional, 630 are married, 650 are over 35 years of age, 127 are professional and over 35, 218 are married and over 35, 152 are professional and married and 100 are married, professional and over 35. Is the data presented in the report right? Explain.

1.22 A fair coin is flipped n times. Let

$$A_k := \text{“The first head was obtained in the } k\text{th toss”}$$

where $k = 1, 2, \cdots, n$. What is $P(A_k)$ equal to?

1.23 Ten distinguishable balls are randomly distributed on 7 distinguishable urns. What is the probability that all urns have at least one ball? What is the probability that exactly two urns are left empty?

1.24 A group of 40 students is made from 20 men and 20 women. If this group is then divided in two equal groups, what is the probability that each group has the same number of men and women?

1.25 Let w be a complex cube root of unity with $w \neq 1$. A fair die is thrown three times. If x, y and z are the numbers obtained on the die. Find the probability that $w^x + w^y + w^z = 0$.

1.26 The coefficients a, b and c of the quadratic equation $ax^2+bx+c = 0$ are determined by rolling a fair die three times in a row. What is the probability that both roots of the equation are real? What is the probability that both roots of the equation are complex?

1.27 What is the chance that a leap year selected at random has 53 Sundays?

1.28 There are two urns A and B. Urn A contains 3 red and 4 black balls while urn B contains 5 red and 7 black balls. If a ball is randomly drawn from each urn, what is the probability that the balls have the same color?

1.29 An urn contains 3 red and 7 black balls. Players A and B consecutively extract a ball each until a red one is drawn. What is the probability that player A will remove the red ball from the urn? Assume the extraction is carried without replacement and player A starts the game.

1.30 There are 200 ornamental fishes in a lake; 50 of them are captured and tagged and then returned to the lake. A few days later 40 fishes are captured. What is the probability that 20 out of the 40 are tagged already?

1.31 Five men and 5 women are ordered according to their grades in a test. Suppose that no two grades are the same and that all 10! arrangements are equally likely. What is the probability that the best position achieved by a man is the fourth one?

1.32 In Bogotá, the father of a certain family decides to plan weekend activities with his children according to the result of the roll of a fair die. If the result is equal to or less than 3, he will take his children to their grandmother's house; if the result is 4 he will take them to the beach in the city of Cartagena; and if the result is 5 or 6, he will stay at home to watch movies with his children.
In order to have an idea of how his weekend activities are shaping up to be, the father decides to divide the year in 13 periods of 4 weeks each and is interested in the probability of the following events:

a) Go at least one trip to Cartagena.

b) Stay twice at home.

c) Go at least three times to grandmother's house.

d) Do each activity at least once.

Calculate the probability of these events?

1.33 In order to illuminate a stairway 7 lamps have been placed and labeled with letters from A to G. To guarantee a proper lighting, lamps A or B must work along with lamps F or G, or any of the lamps C, D or E must be working. The probability that any given lamp will be working is $\frac{1}{5}$.

a) What is the probability that the stairway will be well lit?

b) How does the probability from part i) change if the lamp D is never used?

1.34 What is the probability that among a group of 25 people at least 2 have the same birthday? Assume that the year has 365 days and that all of them are equally likely to be somebody's birthday.

1.35 There are n people at a Christmas party, each carrying a gift. All the presents are put in a bag and mixed, then each person randomly takes one. What is the probability that no one gets their own present?

1.36 In a bridge game, the whole pack of 52 cards is dealt between 4 players.

a) What is the probability that each player gets an ace?

b) What is the probability that one player gets 5 spades while another one receives the remaining 8?

1.37 An urn contains 15 balls, 9 of which are red and the other 6 are white. The following game is played: a ball is randomly extracted and, after taking note of its color, it is returned to the urn along with a new pair of balls of the same color. Find the probability that in the first three rounds of the game all the balls drawn are white.

1.38 Find the probability that in a group of 13 cards, from a pack of 52, there are exactly two kings and an ace. What is the probability that in such a group there is exactly one ace given that the group contains exactly two kings?

1.39 Let A and B be events such that $P(A) = 0.5$, $P(B) = 0.3$ and $P(A \cap B) = 0.1$. Find $P(A \mid B)$, $P(A \mid B^c)$, $P(A \mid A \cap B)$, $P(A^c \mid A \cup B)$ and $P(A \cap B \mid A \cup B)$.

1.40 A math student must take on the same day a probability and an algebra exam. Let:

$$A := \text{“The student fails the probability exam”}.$$
$$B := \text{“The student fails the algebra exam”}.$$

Let $P(A) = 0.4$, $P(B) = 0.3$ and $P(A \cap B) = 0.2$. What is the probability that the student passes the algebra exam given that he passed the probability one? What is the probability that the student passes the probability exam given that he failed the algebra one?

1.41 A survey was taken in a certain city producing the following results:

90%	of the families owns both a radio and a TV.
8%	of the families owns a radio but not a TV.
2%	of the families owns a TV but not a radio.
95%	of the families that owns a radio and a TV knows who the city mayor is.
80%	of the families that owns a radio but not a TV knows who the city mayor is.
1%	of the families that owns a TV but not a radio does not know who the city mayor is.

A family is randomly chosen in this city. Consider the events:

$$T := \text{“The family owns a TV”}$$
$$R := \text{“The family owns a radio”}$$
$$B := \text{“The family knows who is the city mayor”}.$$

Find the following probabilities:

a) $P(T \cup R)$.

b) $P(B \cap T)$.

c) $P(T \mid B)$.

1.42 Consider a population that develops according to the following rules: an initial individual constitutes the 0th generation and it can have 0, 1 or 2 descendants with probabilities of $\frac{1}{6}$, $\frac{2}{3}$ and $\frac{1}{6}$, respectively. After giving its offspring, the individual dies. Each descendant reproduces independently from one another and the family history, following the same rule as the original individual. The first generation will be made from the children of the first individual, the second generation will be made from its grandchildren, and so on. Given that there is only one individual in the second generation, what is the probability that the first generation had two individuals? What is the probability that there is at least one individual in the second generation?

1.43 Consider two urns A and B. Urn A contains 7 red balls and 5 white ones while urn B contains 2 red balls and 4 white ones. A fair die is rolled, if we obtain a 3 or a 6 a ball is taken from B and put into A and, after this, a ball is extracted from A. If the result is any other number, a ball is taken from A and put into B and then a ball is extracted from B. What is the probability that both balls extracted are red?

1.44 Suppose that you ask a classmate to sign you up for the class "Mathematics with no effort" that is being offered for the next term. If your classmate forgets to make the registration by the deadline set by the Mathematics Department, the probability that the class won't have its quota filled and you can therefore register is 2%; on the other hand, if your classmate registers you on time, the probability that the class won't have its quota filled is 80%. You are 95% sure that your classmate will register you on time. What is the probability that your classmate forgot to sign you up for the class if you could not register?

1.45 The probability that in a twin's birth both babies are males is 0.24, while the probability that they are both females is 0.36. What is the probability that in a twin's delivery the second baby born is a boy given that the first one was a boy? Suppose that it is equally likely for the first baby to be either male or female.

1.46 A particle starts at the origin and moves to and from on a straight line. At any move it jumps either 1 unit to the right or 1 unit to the left each with probability $\frac{1}{2}$. All successive moves are independent. Given that the particle is at the origin at the completion of the 6th move, find the probability that it never occupied a position to the left of the origin during previous moves.

1.47 An investor is considering buying a large number of shares of a company. The stock quote of the company in the past 6 months is of great interest to him. Based on this information, he observes that the share price is closely related to the gross national product (GNP): if the GNP goes up, the probability that the share price will rise as well is 0.7; if the GNP remains stable, the probability that the share price will increase is just 0.2; if the GNP falls, however, the probability that the price share will go up is only 0.1. If the probabilities that the GNP increases, remains the same or decreases are 0.5, 0.3 and 0.2, respectively, what is the probability that the shares will go up? If the shares rose their stock quote, what is the probability that the GNP had increased as well?

1.48 A person wrote n letters, sealed them in n envelopes and wrote the n different addresses randomly one on each of them. Find the probability that at least one of the letters reaches its correct destination.

1.49 Suppose that 15 power plants are distributed at random among 4 cities. What is the probability that exactly 2 cities will receive none.

1.50 Suppose that in answering a question on a multiple-choice test an examinee either knows the answer or he guesses. Let p be the probability that he will know the answer, and let $1-p$ be the probability that he will guess. Assume that the probability of answering a question correctly is unity for an examinee who knows the answer and $\frac{1}{m}$ for an examinee who guesses, where m is the number of multiple-choice alternatives. Find the conditional probability that an examinee knew the answer to a question given that he has correctly answered it.

1.51 There are eight coins in an urn. Two of them have two tails, three are fair coins and three are biased in such a way that the probability of getting a tail equals $\frac{3}{5}$. A coin is randomly drawn from the urn. If flipping the coin produced a head, what is the probability that the coin drawn was a common one?

1.52 Let $\Omega = \{a, b, c\}$, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{4}$ for all $\omega \in \Omega$. Let $A = \{b, c\}$. Find all the elements $B \in \Im$ such that A and B are independent.

1.53 Let $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{6}$ for all $\omega \in \Omega$. Prove that if A and B are independent elements in $\Im$ and A has 3 elements, then B must have an even number of elements.

1.54 Let $\Omega = \{1, 2, 3, 4, 5, 6\}, \Im = \wp(\Omega)$ and P be a probability measure with $P(\{1\}) = P(\{2\}) = \frac{1}{12}, P(\{3\}) = P(\{4\}) = \frac{1}{4}$. Find $P(\{5\})$ and $P(\{6\})$ if the events $\{1, 3, 4\}$ and $\{1, 2, 3, 5\}$ are independent.

1.55 Cards are taken out of a standard deck of 52 cards. Let event A be a card with a spade and event B a king. Are A and B independent events?

1.56 Prove that if $A = B$ and A and B are mutually independent (that is, A is independent of itself), then $P(A) = 0$ or $P(A) = 1$.

1.57 Let A be an event. Prove that the following conditions are equivalent:

a) A and B are independent for any event B.

b) $P(A) = 0$ or $P(A) = 1$.

1.58 Let A, B and C be independent events. Prove that A and $B \cup C$; A and $B \cap C$; A and $(B \setminus C)$ are independent.

1.59 Suppose that each of three men at a party throws his hat into the center of the room. The hats are first mixed up and then each man randomly selects a hat. What is the probability that none of the men selects his own hat?

1.60 Prove that events $A_1, A_2, \cdots, A_n$ are independent if and only if

$$P(B_1 \cap B_2 \cap \cdots \cap B_n) = P(B_1)P(B_2)\cdots P(B_n)$$

for all possible choices of $B_1, B_2, \cdots, B_n$ with $B_i = A_i$ or $B_i = A_i^c$ for all $i = 1, 2, \cdots, n$.

1.61 A fair coin is flipped three times in a row. Consider the following events:

$$A := \text{“The results of flips 1 and 2 are different”}.$$
$$B := \text{“The results of flips 2 and 3 are different”}.$$
$$C := \text{“The results of flips 1 and 3 are different”}.$$

a) Verify that $P(A) = P(A \mid B) = P(A \mid C)$ and that $P(A) \neq P(A \mid B \cap C)$.

b) Are A, B and C 2×2 independent? Are A, B and C independent? Explain your answers.

1.62 Let us pick one of the four points $(1,0,0), (0,1,0), (0,0,1)$ and $(1,1,1)$ at random with probability $\frac{1}{4}$ each. Define, for $k = 1, 2, 3$:

$$A_k = \{\text{the } k\text{th coordinate equals } 1\}.$$

Show that the events A_1, A_2 and A_3 are pairwise independent but not independent.

1.63 Let A, B and C be independent events with $P(A) = P(B) = P(C) = \frac{1}{3}$. Find the probability that:

a) At least one event happens.

b) At least two events happen.

c) Exactly two of the events happen.

1.64 Find the probability that among seven people:

a) No two of them were born the same day of the week (Sunday, Monday, Tuesday, etc.).

b) At least two of them were born the same day.

c) Two were born on Sunday and two on Tuesday.

1.65 In a town of $n + 1$ inhabitants, a person tells a rumor to a second person, who in turn repeats it to a third person, etc. At each step, the recipient of the rumor is chosen at random from the n people available. Find the probability that the rumor will be told r times without returning to the originator.

1.66 Suppose that an urn contains N balls numbered from 1 to N. A random sample of size n is taken without replacement, and the numbers obtained are taken down. After returning the balls to the urn, a second sample of size $m(\geq 1)$ is extracted without replacement. Find the probability that both samples have k balls in common.

1.67 An urn contains balls numbered from 1 to N. A ball is randomly drawn.

a) What is the probability that the number on the ball is divisible by 3 or 4?

b) What happens to the probability from the previous question when $n \to \infty$?

1.68 Pick a number x at random out of the integers 1 through 30. Let A be the event that x is even, B that x is divisible by 3 and C that x is divisible by 5. Are the events A, B and C independent?

1.69 Let $Q = (x, y)$ be a point chosen at random in a unit disc centered in $(0, 0)$ and with radius 1. Calculate the probabilities that Q is within 0.5 of the center; that $y > \frac{1}{\sqrt{2}}$; that both $\| x - y \| < 1$ and $\| x + y \| < 1$.

1.70 Suppose that a straight line is randomly subdivided into three parts. What is the probability that these parts can be assembled into a triangle?

1.71 An omnibus company always requires its drivers to wait for 10 minutes at a particular bus stop. The bus you hope to get arrives at this stop anywhere between noon and 1 PM. Assume that you arrive at the stop randomly between 12:30 PM and 1:30 PM and plan to spend at most 10 minutes waiting for the bus. What is the probability that you catch your bus on any day?

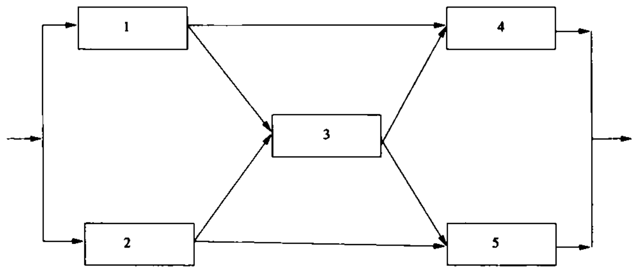


Figure 1.5 A system with five components

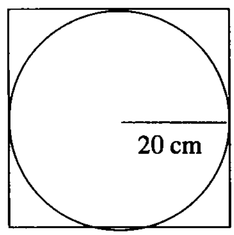


Figure 1.6 Sample space

1.72 Consider a system consisting of five independently functioning components as shown in Figure 1.5. Suppose that the reliability of the components is $R_i = 0.95$, $i = 1, 2, \cdots, 5$. Find the reliability of the system.

1.73 A point is randomly chosen from a disk of radius R. Calculate the probability that the point is closer to the circle than to the center.

1.74 Along a line segment $\overline{ab}$ two points l and m are randomly marked. Find the probability that l is closer to a than m.

1.75 There is a circular dartboard (see Figure 1.6). It costs \$5000 to throw a dart. You win \$10,000 if you hit the square outside of the circle. The radius of the circle is 20 cm. How long should the side of the square be made to make this game fair?

1.76 A point Q is selected at random in the square $ABCD$ and it is constructed a rectangle $AMPN$ (see Figure 1.7). Calculate the probability that the perimeter of the rectangle is less than the length of the square's side.

1.77 A communication system consists of n components, each of which will independently function with probability p. The total system will be able to operate effectively if at least half of its components function. For what values of p is a five-component system more likely to operate effectively than a three-component system?

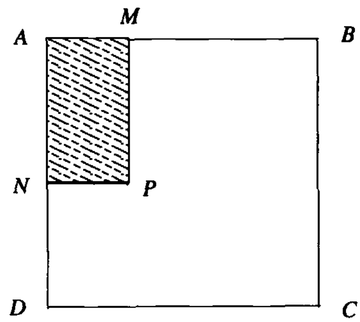


Figure 1.7 Sample space

1.78 Suppose a long-haul airplane has four engines and needs three or more engines to work in order to fly. Another airplane has two engines and needs one engine to fly. Assume that the engines are independent and suppose each has a constant probability p of staying functional during a flight. Find the reliability of both flights and conclude which one is safer.

1.79 A technological system consists of n units, the reliability of each unit being p. The failure of at least one unit results in the failure of the entire system. To increase the reliability of the system, it is duplicated by n similar units. Which way of duplication provides higher reliability: (a) the duplication of every unit (i.e., series of n units, each with duplication); (b) the duplication of the whole system (i.e., parallel system with n units in series).

CHAPTER 2

RANDOM VARIABLES AND THEIR DISTRIBUTIONS

In a random experiment, frequently there has been greater interest for certain numerical values that can be deduced from the results of the random experiment than the experiment itself. Suppose, for example, that a fair coin is tossed consecutively six times and that we want to know the number of heads obtained. In this case, the sample space is equal to:

$$\Omega = \{(a_1, a_2, a_3, a_4, a_5, a_6) : a_i \in \{H, T\}, i = 1, \cdots, 6\}.$$

If we define $X :=$ "number of heads obtained", then we have that X is a mapping of Ω to $\{1, 2, \cdots, 6\}$. Suppose, for example, we have $X((H, H, T, H, H, T)) = 4$. The mapping X is an example of a random variable. That is, a random variable is a function defined on a sample space. We will explain this concept in this chapter.

2.1 DEFINITIONS AND PROPERTIES

Definition 2.1 (Random Variable) *Let $(\Omega, \Im, P)$ be a probability space. A (real) random variable is a mapping $X : \Omega \longrightarrow \mathbb{R}$ such that, for all $A \in \mathcal{B}$, $X^{-1}(A) \in \Im$, where $\mathcal{B}$ is the Borel σ-algebra over $\mathbb{R}$.*

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

Note 2.1 *Let $(\Omega, \Im, P)$ be an arbitrary probability space. Given that the σ-algebra of Borel $\mathcal{B}$ in $\mathbb{R}$ is generated by the collection of all the intervals of the form $(-\infty, x]$ with $x \in \mathbb{R}$, it can be demonstrated that a function $X : \Omega \longrightarrow \mathbb{R}$ is a random variable if and only if $X^{-1}((-\infty, x]) \in \Im$ for all $x \in \mathbb{R}$.*

■ EXAMPLE 2.1

Let $\Omega = \{a, b, c\}$, $\Im = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$, P be an arbitrary probability measure defined over $\Im$. Assume $X : \Omega \longrightarrow \mathbb{R}$ given by:

$$X(\omega) = \begin{cases} 0 & \text{if} \quad \omega = a \\ 1 & \text{if} \quad \omega = b \text{ or } \omega = c \end{cases}$$

is a random variable since

$$X^{-1}((-\infty, x]) = \begin{cases} \emptyset & \text{if} \quad x < 0 \\ \{a\} & \text{if} \quad 0 \le x < 1 \\ \Omega & \text{if} \quad x \ge 1 \end{cases}$$

while the mapping $Y : \Omega \longrightarrow \mathbb{R}$ given by

$$Y(\omega) = \begin{cases} 0 & \text{if} \quad \omega = b \\ 1 & \text{if} \quad \omega = a \text{ or } \omega = c \end{cases}$$

is not a random variable because:

$$Y^{-1}((-\infty, x]) = \{b\} \notin \Im, \quad \text{if } 0 \le x < 1$$

If $\Im' = \{\emptyset, \{b\}, \{a, c\}, \Omega\}$ is taken as a σ-algebra over Ω, then Y is a random variable. ▲

■ EXAMPLE 2.2

Consider the rolling of two dice. $\Omega = \{(i, j), i, j = 1, 2, \cdots, 6\}$ is the set of all possible outcomes. Take $\Im = \wp(\Omega)$. Let X defined by

$$X(\omega) = i + j \text{ if } \omega = (i, j) \in \Omega$$

be a random variable, due to $X^{-1}((-\infty, x]) \in \Im$ for all $x \in \mathbb{R}$. ▲

■ EXAMPLE 2.3

A light bulb is manufactured. It is then tested for its lifetime X by inserting it into a socket and the time elapsed (in hours) until it burns out is recorded. In this case, $\Omega = \{t : t \geq 0\}$. Take σ-algebra $\Im = \mathcal{B}(\mathbb{R} \cap \Omega)$ (see Example 1.11 and Exercise 9.1). Define:

$$X : \Omega \to \mathbb{R} \text{ given by } X(w) = w, \ w \in \Omega.$$

Then:

$$X^{-1}\{(-\infty, x]\} = \begin{cases} \emptyset, & \text{if } x < 0 \\ (0, x] & \text{if } x \geq 0. \end{cases}$$

Hence, the lifetime X is a random variable. ▲

■ EXAMPLE 2.4

Consider Example 2.3. Assume that the light bulbs so manufactured are being sold in the market and from past experience it is known that there will be a profit of \$ 1.00 per bulb if the lifetime is less than 50 hours, a profit of \$ 2.00 per bulb if the lifetime is between 50 and 150 hours and a profit of US\$ 4.00 if the lifetime is more than 150 hours. Let Y be the profit function. Using the solution of Example 2.3, the profit function $Y : X \to \mathbb{R}$ is given by:

$$Y(t) = \begin{cases} 1 & \text{if } t < 50 \\ 2 & \text{if } 50 \leq t \leq 150 \\ 4 & \text{if } t > 150. \end{cases}$$

Since Y satisfies $Y^{-1}((-\infty, x]) \in \Im$ for all $x \in \mathbb{R}$, Y is a random variable. ▲

■ EXAMPLE 2.5

Let $(\Omega, \Im, P)$ be a probability space and $A \in \Im$ fixed. The function $\mathcal{X}_A : \Omega \longrightarrow \mathbb{R}$ given by

$$\mathcal{X}_A(\omega) = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}$$

is a real random variable. Indeed, if $\alpha \in \mathbb{R}$, then:

$$\{\omega \in \Omega : \mathcal{X}_A(\omega) \leq \alpha\} = \begin{cases} \emptyset & \text{if } \alpha < 0 \\ A^c & \text{if } 0 \leq \alpha < 1 \\ \Omega & \text{if } \alpha \geq 1 \end{cases}.$$

The function $\mathcal{X}_A$ is called an indicator function of A. Other notations frequently used for this function are I_A and 1_A. ▲

Notation 2.1 *Let X be a random variable defined over the probability space $(\Omega, \Im, P)$. It is also defined in the set notation form*

$$\{X \in B\} := \{\omega \in \Omega : X(\omega) \in B\} \quad \textit{with } B \in \mathcal{B}.$$

Theorem 2.1 *Suppose that X is a random variable defined over the probability space $(\Omega, \Im, P)$. The function P_X defined over the σ-algebra $\mathcal{B}$ through*

$$P_X(B) := P(\{X \in B\}) \textit{ for all } B \in \mathcal{B}$$

is a probability measure over $(\mathbb{R}, \mathcal{B})$ called the distribution of the random variable X.

Proof: It must be verified that P_X satisfies the three conditions that define a probability measure:

1. It is clear that, for all $B \in \mathcal{B}$, $P_X(B) \geq 0$.
2. $P_X(\mathbb{R}) = P(\{\omega \in \Omega : X(\omega) \in \mathbb{R}\} = P(\Omega) = 1$.
3. Let $A_1, A_2, \cdots$ be elements of $\mathcal{B}$ with $A_i \cap A_j = \emptyset$ for all $i \neq j$. Then:

$$\begin{aligned} P_X\left(\bigcup_{i=1}^{\infty} A_i\right) &= P\left(\left\{X \in \bigcup_{i=1}^{\infty} A_i\right\}\right) \\ &= P\left(\bigcup_{i=1}^{\infty} \{X \in A_i\}\right) \\ &= \sum_{i=1}^{\infty} P(\{X \in A_i\}) \\ &= \sum_{i=1}^{\infty} P_X(A_i). \end{aligned}$$

■

■ **EXAMPLE 2.6**

Let $\Omega = \{a, b, c\}, \Im = \{\emptyset, \{a\}, \{b, c\}, \Omega\}$ and P be given by:

$$\begin{aligned} P(\emptyset) &= 0 \\ P(\{a\}) &= \frac{1}{5} \\ P(\{b, c\}) &= \frac{4}{5} \\ P(\Omega) &= 1. \end{aligned}$$

Let $X : \Omega \longrightarrow \mathbb{R}$ be given by:

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = a \\ 2 & \text{if } \omega = b \text{ or } \omega = c. \end{cases}$$

Then in this case we have, for example, that:

$$\begin{aligned} P_X(\emptyset) &= 0 \\ P_X(\{1\}) &= P(\{a\}) &= \tfrac{1}{5} \\ P_X(\{2\}) &= P(\{b,c\}) &= \tfrac{4}{5} \\ P_X(\mathbb{R}) &= 1. \end{aligned} \qquad \blacktriangle$$

Definition 2.2 (Distribution Function) *Let X be a real random variable. The function F_X defined over $\mathbb{R}$ through*

$$\begin{aligned} F_X(x) &:= P_X((-\infty, x]) \\ &= P(X \leq x) \end{aligned}$$

is called the distribution function (or cumulative distribution function (cdf)) of the random variable X.

■ **EXAMPLE 2.7**

Consider Example 2.6. The distribution function of the random variable X is equal to:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{5} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2. \end{cases} \qquad \blacktriangle$$

■ **EXAMPLE 2.8**

Let $(\Omega, \Im, P)$ be a probability space and $A \in \Im$ be fixed. The distribution function F_X of the random variable $\mathcal{X}_A$ is given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ P(A^c) & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases} \qquad \blacktriangle$$

■ **EXAMPLE 2.9**

Consider the tossing of a fair coin three times and let X be a random variable defined by:

$$X := \text{“number of heads obtained”}.$$

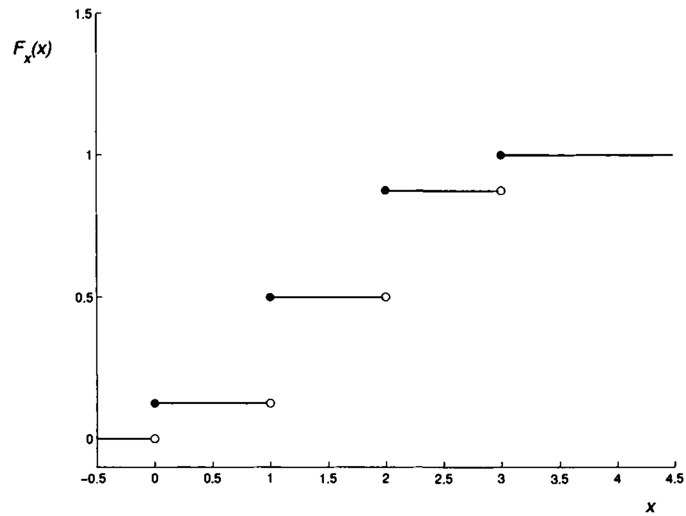


Figure 2.1 Distribution function for Example 2.9

In this case, the distribution function of X is equal to

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{1}{8} & \text{if} \quad 0 \le x < 1 \\ \frac{1}{2} & \text{if} \quad 1 \le x < 2 \\ \frac{7}{8} & \text{if} \quad 2 \le x < 3 \\ 1 & \text{if} \quad x \ge 3 \end{cases}$$

and its graph is shown in Figure 2.9. ▲

■ EXAMPLE 2.10

Suppose a fair die is thrown two consecutive times. Let Z be a random variable given by:

$$Z := \text{“absolute difference of the results obtained”}.$$

That is:

$$Z((x, y)) = |x - y| \qquad \text{with } x, y \in \{1, 2, 3, 4, 5, 6\}.$$

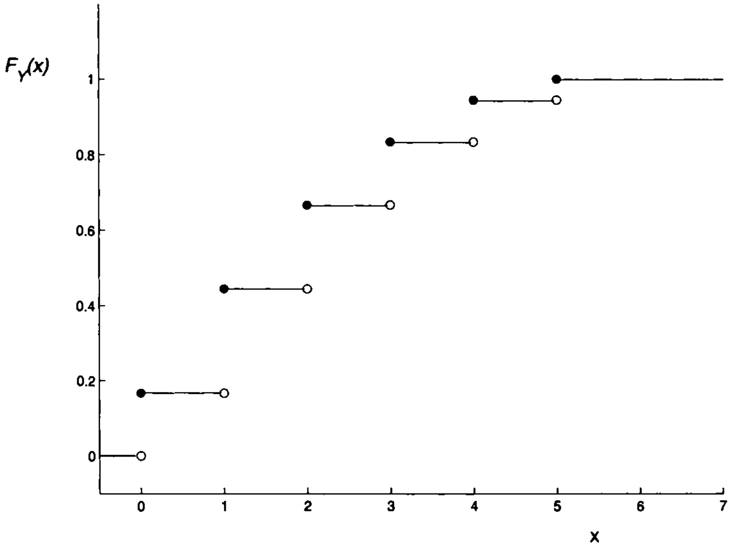


Figure 2.2 Distribution function for Example 2.10

In this case the distribution function of the random variable Z is given by

$$F_Z(z) = \begin{cases} 0 & \text{if} \quad z < 0 \\ \frac{6}{36} & \text{if} \quad 0 \le z < 1 \\ \frac{16}{36} & \text{if} \quad 1 \le z < 2 \\ \frac{24}{36} & \text{if} \quad 2 \le z < 3 \\ \frac{30}{36} & \text{if} \quad 3 \le z < 4 \\ \frac{34}{36} & \text{if} \quad 4 \le z < 5 \\ 1 & \text{if} \quad z \ge 5 \end{cases}$$

and its graph is presented in Figure 2.10. ▲

An important result of probability theory establishes that the distribution P_X of a real random variable X is completely determined by its distribution function F_X (see Muñoz and Blanco, 2002). The proof of this result is beyond the scope of the objectives of this book. Because of this, in the case of real random variables, it is common to identify the distribution of the variable as seen before as a probability measure over $(\mathbb{R}, \mathcal{B})$ with its distribution function.

It can be seen, in the previous examples, that the distribution functions of the random variables considered have certain common characteristics. For example, all of them are nondecreasing and right continuous and the limit when x tends to ∞ in all cases is 1; the limit when x tends to $-\infty$in all cases is equal to 0. These properties are characteristics of all distribution functions as it is stated in the next theorem.

Theorem 2.2 *Let X be a real random variable defined over $(\Omega, \Im, P)$. The distribution function F_X satisfies the following conditions:*

1. *If $x < y$, then $F_X(x) \leq F_X(y)$.*
2. $F_X(x^+) := \lim_{h \to 0^+} F_X(x+h) = F_X(x)$ *for all* $x \in \mathbb{R}$.
3. $\lim_{x \to \infty} F_X(x) = 1$.
4. $\lim_{x \to -\infty} F_X(x) = 0$.

Proof:

1. If $x < y$, then:
$$\{\omega \in \Omega : X(\omega) \leq x\} \subseteq \{\omega \in \Omega : X(\omega) \leq y\}$$
$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y).$$

2. Let $x \in \mathbb{R}$ be fixed. Suppose that $(x_n)_n$ is a decreasing sequence of real numbers with limit x. That is:
$$x_1 \geq x_2 \geq \cdots > x \quad \text{and} \quad \lim_{n \to \infty} x_n = x.$$
It can be seen that
$$\{X \leq x_1\} \supseteq \{X \leq x_2\} \supseteq \cdots$$
and that
$$\bigcap_{n=1}^{\infty} \{X \leq x_n\} = \{X \leq x\}.$$
Therefore, from the probability measure P, it follows that:
$$F_X(x^+) = \lim_{n \to \infty} F_X(x_n) = \lim_{n \to \infty} P(X \leq x_n) = P(X \leq x) = F_X(x).$$

3. It is evident that for all $n \in \mathbb{N}$ it is satisfied that
$$\{X \leq n\} \subseteq \{X \leq (n+1)\}$$
and
$$\Omega = \bigcup_{n=1}^{\infty} \{X \leq n\}.$$

So:

$$\lim_{x\to\infty} F_X(x) = \lim_{n\to\infty} F_X(n) = \lim_{n\to\infty} P(X \le n) = P(\Omega) = 1.$$

4. It is clear that, for all $n \in \mathbb{N}$, it is satisfied that

$$\{X \le -n\} \supseteq \{X \le -(n+1)\}$$

and in addition that

$$\emptyset = \bigcap_{n=1}^{\infty} \{X \le -n\}.$$

Concluding:

$$\lim_{x\to-\infty} F_X(x) = \lim_{n\to\infty} F_X(-n) = \lim_{n\to\infty} P(X \le -n) = P(\emptyset) = 0.$$

■

It can be shown that any function $F(x)$ satisfying conditions $\{1\} - \{4\}$ of Theorem 2.2 is the distribution function of some random variable X.

Corollary 2.1 *Let X be a real random variable defined over $(\Omega, \Im, P)$, F_X its distribution function and $a, b \in \mathbb{R}$ with $a < b$; then:*

1. $F_X(x^-) := \lim_{h\to 0^+} F_X(x-h) = P(X < x)$.
2. $P(a \le X \le b) = F_X(b) - F_X(a^-)$.
3. $P(a < X \le b) = F_X(b) - F_X(a)$.
4. $P(a \le X < b) = F_X(b^-) - F_X(a^-)$.
5. $P(a < X < b) = F_X(b^-) - F_X(a)$.
6. $P(X = a) = F_X(a) - F_X(a^-)$.
7. If $P(a < X < b) = 0$, *then* F_X *is constant in the interval* (a, b).

Proof: Proof of the items 1 and 2 will be elaborated while the rest are left as exercises.

1. Let $x \in \mathbb{R}$ be fixed. Suppose that $(x_n)_n$ is an increasing sequence of real numbers with limit x. That is:

$$x_1 \le x_2 \le \cdots < x \quad \text{and} \quad \lim_{n\to\infty} x_n = x.$$

It is clear that

$$\{X \le x_1\} \subseteq \{X \le x_2\} \subseteq \cdots$$

and

$$\bigcup_{n=1}^{\infty} \{X \le x_n\} = \{X < x\}.$$

Due to this:

$$F_X(x^-) = \lim_{n\to\infty} F_X(x_n) = \lim_{n\to\infty} P(X \le x_n) = P(X < x).$$

2. Given:

$$\Omega = \{X < a\} \cup \{a \le X \le b\} \cup \{X > b\}.$$

It is obtained:

$$1 = P(X < a) + P(a \le X \le b\} + P(X > b).$$

That is:

$$P(a \le X \le b\} = P(X \le b) - P(X < a) = F_X(b) - F_X(a^-).$$

■

■ EXAMPLE 2.11

Which of the following functions can represent the *cdf* of a random variable X?

(a)

$$F(x) = \begin{cases} 0 & \text{if } x < \frac{\pi}{4} \\ \sin x & \text{if } \frac{\pi}{4} \le x < \frac{3\pi}{4} \\ 1 & \text{if } x \ge \frac{3\pi}{4}. \end{cases}$$

(b)

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2\sin x & \text{if } 0 \le x < \frac{\pi}{4} \\ 1 & \text{if } x \ge \frac{\pi}{4}. \end{cases}$$

(c)

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \sqrt{2}\sin x & \text{if } 0 \le x < \frac{\pi}{4} \\ 1 & \text{if } x \ge \frac{\pi}{4}. \end{cases}$$

Solution: We will check all the four properties of $F(x)$. For all three functions:

$$\lim_{x\to-\infty} F(x) = 0 \quad \text{and} \quad \lim_{x\to+\infty} F(x) = 1.$$

(a) The function $F(x)$ is not a nondecreasing function.

(b) The function $F(x)$ is not a nondecreasing function.

(c) The function $F(x)$ satisfies all the four properties.

Hence, $F(x)$ in (c) represents a *cdf* but $F(x)$ in (a) and (b) does not. ▲

The random variables are classified according to their distribution function. If the distribution function F_X of the random variable X is a step function, then it is said that X is a *discrete* random variable. If F_X is an absolutely continuous function, then it is said that X is a *continuous* random variable. And if F_X can be expressed as a linear combination of a step function and a continuous function, then it is said that X is a *mixed* random variable.

■ EXAMPLE 2.12

Let X be a real random variable whose distribution function is given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < -\sqrt{2} \\ \frac{3}{5} & \text{if } -\sqrt{2} \leq x < \pi \\ 1 & \text{if } x \geq \pi. \end{cases}$$

Then X is a discrete random variable. Notice that X takes only the values $-\sqrt{2}$ and π with probability $\frac{3}{5}$ and $\frac{2}{5}$, respectively. ▲

■ EXAMPLE 2.13

Let X be a random variable whose distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{4} & \text{if } 0 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

given that F_X is a continuous function. Then X is a continuous random variable. ▲

■ EXAMPLE 2.14

Let X be a random variable whose distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ q + (1-q)(1-e^{-x}) & \text{if } x \geq 0 \end{cases}$$

where $0 < q < 1$. The distribution function has a jump at 0 since X takes the value 0 with probability q. Hence, X is a mixed random variable. ▲

Note 2.2 *If X is a real continuous random variable defined over the probability space $(\Omega, \Im, P)$, then $P(X = a) = 0$ for all $a \in \mathbb{R}$.*

2.2 DISCRETE RANDOM VARIABLES

The distribution of a discrete random variable has discontinuities that are called jumps. The corresponding points at which the jumps occur are called jump points. This concept will be established in a more precise way in the following definition.

Definition 2.3 *Let X be a real random variable and F_X its distribution function. It is said that F_X presents a jump at the point $a \in \mathbb{R}$ if:*

$$F_X(a) - F_X(a^-) \neq 0.$$

The difference $F_X(a) - F_X(a^-)$ is called the magnitude of the jump or the jump value and by the properties developed previously it is equal to $P(X = a)$.

■ **EXAMPLE 2.15**

In Example 2.10, it can be seen that the random variable Z has jump points $z = i$ with $i = 0, 1, \cdots, 5$. The jump values are, $\frac{1}{6}, \frac{5}{18}, \frac{2}{9}, \frac{1}{6}, \frac{1}{9}$ and $\frac{1}{18}$, respectively. ▲

Note 2.3 *If X is a real continuous random variable, then the collection of jump points of F_X is an empty set.*

The following result is very important since it guarantees that the number of jumps in a discrete real random variable is at most countable.

Theorem 2.3 *Let X be a discrete real random variable defined over the probability space $(\Omega, \Im, P)$ and F_X its distribution function. Then the number of jumps of F_X is at most countable.*

Proof: (Hernandez, 2003) Given that the magnitude of each jump is an element belonging to the interval $(0, 1]$ and the collection of intervals I_n with the form

$$I_n = \left(\frac{1}{2^{n+1}}, \frac{1}{2^n}\right] \quad \text{with } n = 0, 1, \cdots$$

forms a partition of $(0, 1]$, it is obtained that the magnitude of each jump must belong to one of the intervals I_n. Because the magnitudes of the jumps are probabilities, it is clear that at most there is a jump whose magnitude is in the interval I_0, there are at most three jumps whose magnitudes are in the interval I_1, there are at most seven jumps with magnitudes in the interval I_2 and in general, there are at most $2^{n+1} - 1$ jumps whose magnitudes are in the interval I_n. Therefore, due to the existence of a countable number of intervals I_n and at most $2^{n+1} - 1$ jumps in the interval I_n, it is concluded that the number of jumps is at most countable. ■

From the previous result it is concluded that the range of a discrete real random variable is at most a countable set.

Let X be a discrete real random variable and suppose that X takes the values $x_1, x_2, \cdots$ (all different). Let x be a real arbitrary number. Then:

$$\begin{aligned} F_X(x) &= P(X \leq x) \\ &= P\left(\bigcup_{x_i \leq x} (X = x_i) \right) \\ &= \sum_{x_i \leq x} P(X = x_i). \end{aligned}$$

That is, the distribution function of X is completely determined by the values of p_i with $i = 1, 2, \cdots$ where $p_i := P(X = x_i)$. This observation motivates the following definition:

Definition 2.4 (Probability Mass Function) *Let X be a discrete real random variable with values $x_1, x_2, \cdots$ (all different). The function p_X defined in $\mathbb{R}$ through*

$$p_X(x) = \begin{cases} P(X = x_i) & \text{if } x = x_1, x_2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

is called a probability mass function (pmf) of the discrete random variable X.

The following properties hold for the *pmf*:

1. $p(x_i) \geq 0$ for all i and
2. $\sum\limits_i p(x_i) = 1$.

■ EXAMPLE 2.16

Suppose that a fair die is tossed once and let X be a random variable that indicates the result obtained. In this case the possible values of X obtained are $1, \cdots, 6$. The probability mass function of X is given by:

$$p_X(x) = \begin{cases} \frac{1}{6} & \text{if } x = 1, 2, \cdots, 6 \\ 0 & \text{otherwise.} \end{cases}$$ ▲

It has been seen that the probability mass function of the discrete random variable X determines completely its distribution function. Conversely, it is obtained that for any $x \in \mathbb{R}$ it is satisfied that:

$$P(X = x) = F_X(x) - F_X(x^-).$$

That is, the distribution function of the discrete random variable X determines completely its probability mass function.

■ EXAMPLE 2.17

Let X be a random variable whose distribution function is given by:

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -2 \\ \frac{1}{7} & \text{if} \quad -2 \leq x < \frac{1}{2} \\ \frac{4}{7} & \text{if} \quad \frac{1}{2} \leq x < \sqrt{2} \\ 1 & \text{if} \quad x \geq \sqrt{2}. \end{cases}$$

In this case, the probability mass function of the random variable X is given by:

$$p_X(x) = \begin{cases} \frac{1}{7} & \text{if } x = -2 \\ \frac{3}{7} & \text{if } x = \frac{1}{2} \\ \frac{3}{7} & \text{if } x = \sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$ ▲

■ EXAMPLE 2.18

Let X be a discrete random variable with probability mass function given by:

$$p_X(x) = \begin{cases} \frac{5}{9} & \text{if } x = -1 \\ \frac{4}{9} & \text{if } x = \frac{3}{2} \\ 0 & \text{otherwise.} \end{cases}$$

In this case, the distribution function of the random variable X is given by:

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < -1 \\ \frac{5}{9} & \text{if} \quad -1 \leq x < \frac{3}{2} \\ 1 & \text{if} \quad x \geq \frac{3}{2}. \end{cases}$$ ▲

EXAMPLE 2.19

Let X be a discrete random variable with values $\{0, \pm 1, \pm 2\}$. Suppose that $P(X = -2) = P(X = -1)$ and $P(X = 1) = P(X = 2)$ with the information that $P(X > 0) = P(X < 0) = P(X = 0)$. Find the probability mass function and distribution function of the random variable X.

Solution: Let

$$P(X = -2) = P(X = -1) = \alpha \text{ and } P(X = 1) = P(X = 2) = \beta.$$

$$\begin{aligned} P(X > 0) &= P(X = 1) + P(X = 2) = 2\beta \\ P(X < 0) &= P(X = -1) + P(X = -2) = 2\alpha. \end{aligned}$$

Given:

$$P(X > 0) = P(X < 0) = P(X = 0) .$$

Hence, $2\alpha = 2\beta$. Therefore:

$$P(X > 0) = 2\alpha, P(X < 0) = 2\alpha \text{ and } P(X = 0) = 2\alpha.$$

Assuming total probability is 1, $\alpha = \frac{1}{6}$. The probability mass function of the random variable X is given by:

$$P(X = i) = \begin{cases} \frac{1}{6} & \text{if } i = \pm 1, \pm 2 \\ \frac{1}{3} & \text{if } i = 0 \\ 0 & \text{otherwise.} \end{cases}$$

The distribution function of the random variable X is given by:

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{6} & -2 \leq x < -1 \\ \frac{1}{3} & -1 \leq x < 0 \\ \frac{2}{3} & 0 \leq x < 1 \\ \frac{5}{6} & 1 \leq x < 2 \\ 1 & x \geq 2. \end{cases}$$ ▲

EXAMPLE 2.20

A fair coin is tossed two times. Let X be the number of heads obtained.

1. Find the probability mass function of the random variable X.
2. Determine $P(0.5 < X \leq 4), P(-1.5 \leq X < 1)$ and $P(X \leq 2)$.

Solution:

1. The random variable X takes values 0, 1 and 2:

$$\begin{aligned} P(X=0) &= P(\{T,T\}) = \frac{1}{4} \\ P(X=1) &= P(\{T,H\}) + P(\{H,T\}) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \\ P(X=2) &= P(\{H,H\}) = \frac{1}{4}. \end{aligned}$$

The probability mass function of the random variable X is given by

x	0	1	2
$p_X(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

2.

$$\begin{aligned} P(0.5 < X \leq 4) &= P(X=1) + P(X=2) = \frac{3}{4} \\ P(-1.5 \leq X < 1) &= P(X=0) = \frac{1}{4} \\ P(X \leq 2) &= P(X=0) + P(X=1) + P(X=2) = 1. \end{aligned}$$ ▲

EXAMPLE 2.21

A random variable X can take all nonnegative integer values and $P(X = m)$ is proportional to $\alpha^m (0 < \alpha < 1)$. Find $P(X = 1)$.
Solution: Let $p_m = P(X = m) = k\alpha^m$ where k is a constant. Also since p_m is a pmf, we have:

$$\sum_{m=0}^{\infty} p_m = 1$$

$$\sum_{m=0}^{\infty} k\alpha^m = 1$$

$$k \cdot \frac{1}{1-\alpha} = 1$$

$$p_1 = k \cdot \alpha = (1-\alpha)\alpha.$$ ▲

2.3 CONTINUOUS RANDOM VARIABLES

In the study of continuous random variables, special emphasis is paid to the absolutely continuous random variables defined as follows:

Definition 2.5 (Absolute Continuous Random Variables)
Let X be a real random variable defined over the probability space $(\Omega, \Im, P)$. It is said that X is absolutely continuous if and only if there exists a nonnegative and integrable real function f_X such that for all $x \in \mathbb{R}$ it is satisfied that:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt. \tag{2.1}$$

The function f_X receives the name probability density function (pdf) or simply density function of the continuous random variable X.

Note 2.4 *A probability density function f satisfies the following properties:*

(a) $f(x) \geq 0$ for all possible values of x.

(b) $\int_{-\infty}^{\infty} f(x)dx = 1$.

Property (a) follows from the fact that $F(x)$ is nondecreasing and hence its derivative $f(x) \geq 0$, while (b) follows from the condition that $\lim\limits_{x\to\infty} F(x) = 1$. It is clear that any real-valued function satisfying the above two properties will be a probability density function of some continuous random variable.

■ **EXAMPLE 2.22**

Let X be a random variable with distribution function given by:

$$F_X(x) = \begin{cases} 0 & \text{if} \quad x < 0 \\ x & \text{if} \quad 0 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1. \end{cases}$$

The function f_X defined by

$$f_X(x) = \begin{cases} 1 & \text{if } 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases}$$

is a *pdf* of the random variable X. ▲

■ EXAMPLE 2.23

Let X be a random variable with distribution function given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{x^2}{4} & \text{if } 0 \le x \le 2 \\ 1 & \text{if } x > 2. \end{cases}$$

It is easy to verify that the function f_X given by

$$f_X(x) = \begin{cases} \frac{x}{2} & \text{if } 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

is a *pdf* of the random variable X. ▲

Suppose that X is an absolutely continuous random variable with probability density function f_X. Some properties that satisfy this function will be deduced shortly.

It is known that:

$$\begin{aligned} \int_{-\infty}^{\infty} f_X(t)dt &= \lim_{x\to\infty} \int_{-\infty}^{x} f_X(t)dt \\ &= \lim_{x\to\infty} F_X(x) = 1. \end{aligned}$$

If $a, b \in \mathbb{R}$, then:

$$\begin{aligned} P(a < X \le b) &= F_X(b) - F_X(a) \\ &= \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx \\ &= \int_{a}^{b} f_X(x)dx. \end{aligned}$$

Notice that because F_X is a continuous function, the previous integral is also equal to $P(a \le X \le b)$, $P(a < X < b)$ and $P(a \le X < b)$. Moreover, it is obtained that

$$P(X \in B) = \int_B f_X(x)dx \tag{2.2}$$

for all Borel sets B. The proof of this result is beyond the scope of this text. It is clear that the integral given in (2.2) must be interpreted as a Lebesgue integral given that the Riemannn integral is not defined for all Boolean sets. Nevertheless, if B is an interval or a union of intervals, it makes sense that, for practical effects, the Riemann integral is sufficient.

Since $F_X(x)$ is absolutely continuous, it is differentiable at all x except perhaps at a countable number of points. Using the fundamental theorem of integrals, we have $f_X(x) = \frac{dF_X(x)}{dx}$ for all x where $F_X(x)$ is differentiable. For the points $a_1, a_2, \cdots, a_n, \cdots$, where $F_X(x)$ is not differentiable, we define:

$$\frac{dF_X(x)}{dx} = 0.$$

With this we can conclude that:

$$f_X(x) = \frac{dF_X(x)}{dx} \qquad \text{for all } x.$$

This last property implies that, for $\Delta x \approx 0$, it is obtained that:

$$\begin{aligned} P(x - \Delta x < X \leq x + \Delta x) &= F_X(x + \Delta x) - F_X(x - \Delta x) \\ &\approx 2\Delta x f_X(x). \end{aligned}$$

That is, the probability that X belongs to an interval of small length around x is the same as the probability density function of X evaluated in x times the interval's length.

When the context in which the random variable is referenced is clear, we will eliminate the subscript X in the distribution function as well as in the density function.

■ EXAMPLE 2.24

Let X be a random variable with density function given by:

$$f(x) = \begin{cases} kx(1-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

Determine:

1. The value of k.
2. The distribution function of the random variable X.
3. $P(-1 \leq X \leq \frac{1}{2})$.

Solution:

1. Given that

$$1 = \int_{-\infty}^{\infty} f(x)dx = k \int_0^1 (x - x^2)dx = k\frac{1}{6}.$$

It is obtained that $k = 6$.

2.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)dt \\ &= \begin{cases} 0 & \text{if} \quad x < 0 \\ 3x^2 - 2x^3 & \text{if} \quad 0 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1. \end{cases} \end{aligned}$$

3.

$$P\left(-1 \leq X \leq \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F(-1)$$
$$= \frac{1}{2}. \qquad \blacktriangle$$

■ EXAMPLE 2.25

Consider Example 2.3. Let X be the random variable representing the lifetime of the light bulb. The *cdf* of X is given by:

$$F(x) = \begin{cases} 0 & x \leq 0 \\ kx & 0 < x \leq 100 \\ 1 & x > 100. \end{cases}$$

1. Determine k.

2. Find the probability that the light bulb lasts more than 20 hours but no more than 70 hours.

Solution:

1. Since $F(x)$ is a cumulative distribution function of the continuous random variable X, it is a continuous function. Applying right continuity at $x = 100$, we get:

$$\begin{aligned} 100k &= 1 \\ k &= \frac{1}{100}. \end{aligned}$$

2. The probability that the light bulb lasts more than 20 hours but no more than 70 hours is given by:

$$\begin{aligned} P(20 \leq X \leq 70) &= F(70) - F(20) \\ &= 0.5. \qquad \blacktriangle \end{aligned}$$

■ EXAMPLE 2.26

Let X be a random variable whose distribution function is given by:

$$F(x) = \begin{cases} 1 - (1+x)e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

1. Determine the density function of X.
2. Calculate $P(X \leq \frac{1}{3})$.

Solution:

1.

$$f(x) = \begin{cases} xe^{-x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

2.

$$\begin{aligned} P\left(X \leq \frac{1}{3}\right) &= F\left(\frac{1}{3}\right) \\ &= 1 - \frac{4}{3}e^{-\frac{1}{3}} \\ &= 4.4625 \times 10^{-2}. \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 2.27

A continuous random variable X has density function given by

$$f(x) = \begin{cases} kxe^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$.

1. Determine k.
2. Find the distribution function of X.

Solution:

1. Given that

$$1 = \int_{-\infty}^{\infty} f(x)dx = k\int_{0}^{\infty} xe^{-\lambda x}dx = \frac{k}{\lambda^2}.$$

It is obtained that $k = \lambda^2$.

2.

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(u)du \\ &= \begin{cases} -\lambda(xe^{-\lambda x}) + 1 - e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise.} \end{cases} \quad \blacktriangle \end{aligned}$$

2.4 DISTRIBUTION OF A FUNCTION OF A RANDOM VARIABLE

Suppose that X is a real random variable defined over the probability space $(\Omega, \Im, P)$ and let g be a function such that $Y = g(X)$ is a random variable defined over $(\Omega, \Im, P)$. We are interested in determining (if possible) the distribution function of the random variable Y in terms of the distribution function of the random variable X.

■ EXAMPLE 2.28

Let X be a random variable and Y be defined as $Y = \mid X \mid$. We know that the modulus function is continuous and hence Y is also a random variable. Let F_X be the *cdf* of X. The *cdf* of Y is given by:

$$\begin{aligned} P(Y \leq y) &= P(|X| \leq y) \\ &= P(-y \leq X \leq y) \\ &= P(X \leq y) - P(X \leq -y) + P(X = -y) \\ &= F_X(y) - F_X(-y) + P(X = -y). \end{aligned}$$

If X is a continuous random variable, then $P(X = -y) = 0$. When X is a continous random variable, the *cdf* of Y is given by:

$$F_Y(y) = \begin{cases} F_X(y) - F_X(-y) & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

When X is a discrete random variable, the *cdf* of Y is given by:

$$F_Y(y) = \begin{cases} F_X(y) - F_X(-y) + P(X = -y) & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$ ▲

■ EXAMPLE 2.29

Let X be a random variable with *cdf* $F_X(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt$. Find the *cdf* of $Y = X^+$.

Solution: We know that:

$$Y = X^+ = \begin{cases} X, & X \geq 0 \\ 0, & X < 0. \end{cases}$$

The *cdf* of Y is given by:

$$\begin{aligned}
F_Y(y) &= P(Y \le y) = P(X^+ \le y) \\
&= \begin{cases} 0 & \text{if } \ y < 0 \\ P(X \le 0) & \text{if } \ y = 0 \\ P(X < 0) + P(0 \le X \le y) & \text{if } \ y > 0 \end{cases} \\
F_Y(y) &= \begin{cases} 0 & \text{if } \ y < 0 \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{0} e^{-\frac{t^2}{2}} dt & \text{if } \ y = 0 \\ \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{-\frac{t^2}{2}} dt & \text{if } \ y > 0. \end{cases}
\end{aligned}$$

The distribution function of Y has a jump at 0. Hence, Y is a mixed random variable. ▲

■ EXAMPLE 2.30

Let X be a discrete or continuous real random variable. Find the *cdf* of $Y := aX + b$ where $a, b \in \mathbb{R}$ and $a \neq 0$.

Solution: It is clear that:

$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P\left(aX + b \le y\right) \\
&= \begin{cases} P\left(X \le \frac{y-b}{a}\right) & \text{if } \ a > 0 \\ P\left(X \ge \frac{y-b}{a}\right) & \text{if } \ a < 0 \end{cases} \\
&= \begin{cases} F_X\left(\frac{y-b}{a}\right) & \text{if } \ a > 0 \\ 1 - F_X\left(\left(\frac{y-b}{a}\right)^-\right) & \text{if } \ a < 0. \end{cases}
\end{aligned}$$

▲

■ EXAMPLE 2.31

Let X be a real random variable with density function given by:

$$f_X(x) = \begin{cases} \frac{1}{2} & \text{if } \ -1 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Let $Y = |X|$. Then:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= \begin{cases} P(|X| \le y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0 \end{cases} \\
&= \begin{cases} P(-y \le X \le y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0 \end{cases} \\
&= \begin{cases} F_X(y) - F_X(-y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0. \end{cases}
\end{aligned}
$$

Therefore, the density function of the random variable Y is given by:

$$
\begin{aligned}
f_Y(y) &= \begin{cases} f_X(y) + f_X(-y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0 \end{cases} \\
&= \begin{cases} 1 & \text{if } \; 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases} \qquad \blacktriangle
\end{aligned}
$$

■ **EXAMPLE 2.32**

Let X be a real random variable with density function given by:

$$
f_X(x) = \begin{cases} 1 & \text{if } \; 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}
$$

Let $Y = e^X$. In this case, it is obtained that:

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= \begin{cases} P(X \le \ln y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0 \end{cases} \\
&= \begin{cases} F_X(\ln y) & \text{if } \; y > 0 \\ 0 & \text{if } \; y \le 0. \end{cases}
\end{aligned}
$$

Therefore, the density function of the random variable Y is given by:

$$
\begin{aligned}
f_Y(y) &= \begin{cases} \frac{1}{y} f_X(\ln y) & \text{if } \; y > 0 \\ 0 & \text{otherwise} \end{cases} \\
&= \begin{cases} \frac{1}{y} & \text{if } \; 0 < \ln y < 1 \\ 0 & \text{otherwise.} \end{cases} \qquad \blacktriangle
\end{aligned}
$$

EXAMPLE 2.33

The distribution of a random variable X is given by:

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{6} + \frac{1}{3}x & 0 \leq x \leq 1 \\ 1 & x \geq 1. \end{cases}$$

Find the distribution of $Y = \alpha X + \beta, \alpha > 0$.

Solution: Since the distribution function of X has jumps at 0 and 1, X is a mixed random variable. Using $Y = \alpha X + \beta$, the distribution function of Y is given by:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\alpha X + \beta \leq y) \\ &= P\left(X \leq \frac{y-\beta}{\alpha}\right) \\ &= \begin{cases} 0 & y < \beta \\ \frac{1}{6} + \frac{1}{3}\frac{y-\beta}{\alpha} & \beta \leq y < \alpha + \beta \\ 1 & \alpha + \beta \leq y. \end{cases} \end{aligned}$$

Note that the distribution function of Y has jumps at β and $\alpha + \beta$. Hence, Y is also a mixed random variable. ▲

EXAMPLE 2.34

Let X be a discrete random variable with probability mass function given by:

x	-1	0	1	2	3
$P_X(x)$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

Let $Y = X^2$. It can be seen that the possible values of Y are $0, 1, 4$ and 9. Additionally:

y	0	1	4	9
$P(Y = y)$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

The corresponding distribution functions of X and of Y are:

$$F_X(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{1}{7} & \text{if } -1 \le x < 0 \\ \frac{3}{7} & \text{if } 0 \le x < 1 \\ \frac{4}{7} & \text{if } 1 \le x < 2 \\ \frac{6}{7} & \text{if } 2 \le x < 3 \\ 1 & \text{if } x \ge 3 \end{cases}$$

$$F_Y(y) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{2}{7} & \text{if } 0 \le x < 1 \\ \frac{4}{7} & \text{if } 1 \le x < 4 \\ \frac{6}{7} & \text{if } 4 \le x < 9 \\ 1 & \text{if } x \ge 9. \end{cases} \qquad \blacktriangle$$

■ EXAMPLE 2.35

Let X be a random variable with probability mass function $P(X = x) = \binom{n}{x} p^x(1-p)^{n-x}$, $x = 0, 1, \cdots, n$. Find the distribution of $Y = n - X$.

Solution: The *pmf* of Y is given by:

$$\begin{aligned} P(Y = y) &= P(n - X = y) \\ &= P(X = n - y) \\ &= \binom{n}{n-y} p^{n-y}(1-p)^y \\ &= \binom{n}{y} p^{n-y}(1-p)^y \\ &= \binom{n}{y} (1-p)^y p^{n-y}, \quad y = 0, 1, \cdots, n. \end{aligned} \qquad \blacktriangle$$

■ EXAMPLE 2.36

Let X be a random variable with distribution over the set of integers $\{-n, -(n-1), \cdots, -1, 0, 1, \cdots, (n-1), n\}$ given by:

$$P(X = x) = \begin{cases} \frac{1}{2n+1} & \text{if } x = 0, \pm 1, \pm 2, \cdots, \pm n \\ 0 & \text{otherwise.} \end{cases}$$

Find the distribution of (i) $|X|$ and (ii) X^2.
Solution: (i) Let $Y = |X|$. For values of X, the corresponding values of Y are given in the following table:

X	$-n, n$	$-(n-1), n-1$	$\cdots$	-1, 1	0
Y	n	$n-1$	$\cdots$	1	0

For $y = 0$:

$$P(Y=0) = P(X=0) = \frac{1}{2n+1}.$$

For $y = 1, 2, \cdots, n$:

$$P(Y=y) = P(X=-y) + P(X=y) = \frac{1}{2n+1} + \frac{1}{2n+1} = \frac{2}{2n+1}.$$

Hence, the probability mass function of Y is given by:

$$P(Y=y) = \begin{cases} \frac{1}{2n+1} & \text{if } y=0 \\ \frac{2}{2n+1} & \text{if } y=1,2,\cdots, n \\ 0 & \text{otherwise.} \end{cases}$$

(ii) Let $Y = X^2$. Using a similar approach, the probability mass function of Y is given by

$$P(Y=y) = \begin{cases} \frac{1}{2n+1} & \text{if } y=0 \\ \frac{2}{2n+1} & \text{if } y=1^2, 2^2, 3^2, \cdots, n^2 \\ 0 & \text{otherwise.} \end{cases}$$ ▲

■ EXAMPLE 2.37

Suppose that X is a continuous random variable. Define:

$$Y = \begin{cases} 1 & \text{if } X \geq 0 \\ -1 & \text{if } X < 0. \end{cases}$$

Find the *cdf* of Y.
Solution:

$$\begin{aligned} P(Y=1) &= P(X \geq 0) \\ P(Y=-1) &= P(X < 0). \end{aligned}$$

Therefore:

$$P(Y \leq y) = \begin{cases} 0 & y < -1 \\ P(X<0) & -1 \leq y < 1 \\ 1 & 1 \leq y < \infty. \end{cases}$$

In this example, note that X is a continuous random variable while Y is a discrete random variable. ▲

Theorem 2.4 *Let X be an absolutely continuous real random variable with density function f_X. If h is a strictly monotonous and differentiable function, then the probability density function of the random variable $Y = h(X)$ is given by*

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y)) \left|\frac{d}{dy} h^{-1}(y)\right| & \text{if } \; y = h(x) \; \text{ for some } x \\ 0 & \text{if } \; y \neq h(x) \; \text{ for all } x \end{cases}$$

where $h^{-1}(\cdot)$ is the unique inverse function of $h(\cdot)$.

Proof: Suppose that h is a strictly increasing function and let $y \in \mathbb{R}$ such that $y = h(x)$ for some x.

Then:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(h(X) \leq y) \\ &= P(X \leq h^{-1}(y)) \\ &= F_X(h^{-1}(y)). \end{aligned}$$

Differentiating yields

$$\begin{aligned} f_Y(y) &= f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y) \\ &= f_X(h^{-1}(y)) \left|\frac{d}{dy} h^{-1}(y)\right| \end{aligned}$$

given that the derivative of h is positive.

Let h be a strictly decreasing function and $y \in \mathbb{R}$ such that $y = h(x)$ for some x. Then:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(h(X) \leq y) \\ &= P(X \geq h^{-1}(y)) \\ &= 1 - F_X(h^{-1}(y)). \end{aligned}$$

Differentiating yields

$$\begin{aligned} f_Y(y) &= -f_X(h^{-1}(y)) \frac{d}{dy} h^{-1}(y) \\ &= f_X(h^{-1}(y)) \left|\frac{d}{dy} h^{-1}(y)\right| \end{aligned}$$

because in this case the derivative of h is negative.

If $y \in \mathbb{R}$ is such that $y \neq h(x)$ for all x, then, $F_Y(y) = 0$ or $F_Y(y) = 1$, that is, $f_Y(y) = 0$. ■

Corollary 2.2 *Let h be piecewise strictly monotone and continuously differentiable, that is, there exist intervals $I_1, I_2, \cdots, I_n$ which partition $\mathbb{R}$ such that h is strictly monotone and continuously differentiable on the interior of each I_i. Let $Y = h(X)$. Then the density of Y exists and is given by*

$$f_Y(y) = \sum_{k=1}^{n} f_X(h_k^{-1}(y)) \left| \frac{dh_k^{-1}(y)}{dy} \right|$$

where h_k^{-1} is the inverse of h in I_k.

■ EXAMPLE 2.38

Let X denote the measurement error in a certain physical experiment and let Y denote the square of the defined random variable X. Given that the *pdf* of X is known, find the *pdf* of Y.

Solution: Given that X is a continuous random variable and it denotes the error in the physical experiment. In order to find the *pdf* of the newly defined random variable Y, we shall first obtain the *cdf* of Y and then by differentiating the *cdf* we will obtain the corresponding *pdf*.

Let $F_X(x)$ denote the *cdf* of the given random variable X. The *cdf* of Y is given by:

$$\begin{aligned} P(Y \leq y) &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= P(X \leq \sqrt{y}) - P(X \leq -\sqrt{y}) \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \, . \end{aligned}$$

Thus, we have obtained the *cdf* of the newly defined random variable Y. Now, by differentiating the *cdf* of Y, we get the *pdf* of Y:

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} \left[f_X(\sqrt{y}) + f_X(-\sqrt{y}) \right] & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$ ▲

■ EXAMPLE 2.39

Let X be a random variable with distribution

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

Define $Y = e^X$. Find the *pdf* of Y.

Solution: Consider $Y = h(X) = e^X$ is a strictly increasing function of x and also $h(X)$ is differentiable. Now:

$$y = h(x) = e^x.$$

Then:

$$h^{-1}(y) = x = \ln y$$
$$\frac{dh^{-1}(y)}{dy} = \frac{1}{y}.$$

Since $Y = e^X$ satisfies all the conditions of Theorem 2.4, we get:

$$f_Y(y) = f_X(h^{-1}(y)) \left| \frac{dh^{-1}(y)}{dy} \right| = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\ln y)^2} \frac{1}{y}.$$

Hence, the *pdf* of Y is given by:

$$f_Y(y) = \begin{cases} \frac{1}{y\sqrt{2\pi}} e^{-\frac{1}{2}(\ln y)^2} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

▲

2.5 EXPECTED VALUE AND VARIANCE OF A RANDOM VARIABLE

Let X be a real random variable. It is known that its probabilistic-type properties are determined by its distribution function $F_X(\cdot)$. Nevertheless, it is important to know a group of values that "summarize" in a certain way this information. For example, we want to define an "average" of the values taken by X and a measurement that "quantifies" by how much the values of X vary with respect to that value. This information will be given by two real values called expected value (average) and variance of the random variable, respectively. These concepts are stated precisely in the next definitions.

Definition 2.6 (Expected Value) *Let X be a real random variable defined over the probability space $(\Omega, \Im, P)$.*

1. *If X is a discrete random variable with values $x_1, x_2, \cdots$, it is said that X has an expected value if:*

$$\sum_{k=1}^{\infty} |x_k| P(X = x_k) < \infty.$$

In such a way, the expected value $E(X)$ (mathematical expectation, average) of X is determined as:

$$E(X) = \sum_{k=1}^{\infty} x_k P(X = x_k).$$

2. *If X is a continuous random variable with density function f_X, it is said that X has an expected value if:*

$$\int_{-\infty}^{\infty} |x| \, f_X(x)dx < \infty.$$

In this case, the expected value $E(X)$ (mathematical expectation, average) of X is determined as:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx.$$

Note 2.5 *If X is a real random variable that takes only a finite number of values, then $E(X)$ always exists.*

■ **EXAMPLE 2.40**

Suppose that a normal die is thrown once and let X be a random variable that represents the result obtained. It is clear that:

$$E(X) = \sum_{k=1}^{6} \frac{k}{6} = \frac{21}{6}. \qquad \blacktriangle$$

■ **EXAMPLE 2.41**

Let $(\Omega, \Im, P)$ be an arbitrary probability space and $A \in \Im$ fixed. It is known that $X := \mathcal{X}_A$ is a discrete random variable that only takes the values 0 and 1 with probability $P(A^c)$ and $P(A)$, respectively. Therefore, $E(X)$ exists and it is equal to $P(A)$. ▲

■ **EXAMPLE 2.42**

Let X be a discrete random variable with pmf given by:

$$p(x) = \begin{cases} e^{-3}\frac{3^x}{x!} & \text{if } x = 0, 1, 2, \cdots \\ 0 & \text{otherwise.} \end{cases}$$

Then:

$$
\begin{aligned}
E(X) &= \sum_{k=1}^{\infty} k e^{-3} \frac{3^k}{k!} \\
&= e^{-3} \sum_{k=1}^{\infty} \frac{3^k}{(k-1)!} \\
&= e^{-3} \sum_{j=0}^{\infty} \frac{3^{j+1}}{j!} \\
&= 3e^{-3} \sum_{j=0}^{\infty} \frac{3^j}{j!} \\
&= 3e^{-3} e^3 = 3. \qquad \blacktriangle
\end{aligned}
$$

■ EXAMPLE 2.43

Let X be a continuous random variable with *pdf* given by:

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Given that:

$$\int_{-\infty}^{\infty} |x| f(x) dx = \int_0^{\infty} 2xe^{-2x} = \frac{1}{2}.$$

It is concluded that the expected value of X exists and it is equal to $\frac{1}{2}$. ▲

■ EXAMPLE 2.44

Let X be a random variable with values in $\mathbb{Z}$. Suppose that

$$P(X = j) = \begin{cases} \frac{c}{j^2} & \text{if } j \neq 0 \\ 0 & \text{if } j = 0 \end{cases}$$

where $c > 0$ is a constant such that

$$\sum_j \frac{c}{j^2} = 1.$$

Given that

$$\sum_j |j| P(X = j) = \infty,$$

$E(X)$ does not exist. ▲

■ EXAMPLE 2.45

Let X be an absolutely continuous random variable with density function given by

$$f(x) = \frac{\alpha}{\pi(\alpha^2 + x^2)} \quad \text{if } x \in \mathbb{R}$$

where $\alpha > 0$ is a constant.

Then

$$\int_{-\infty}^{\infty} |x|\, f(x)dx = \frac{2\alpha}{\pi} \int_{0}^{\infty} \frac{x}{\alpha^2 + x^2} dx = \infty,$$

which means that $E(X)$ does not exist. ▲

The following result allows us to find the expected value of an absolutely continuous random variable from its distribution function.

Theorem 2.5 *Let Y be an absolutely continuous random variable with density function f. If $E(Y)$ exists, then:*

$$E(Y) = \int_0^{\infty} [1 - F_Y(y)]dy - \int_0^{\infty} F_Y(-y)dy\ .$$

Proof:

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{\infty} x f(x)dx \\
&= \int_0^{\infty} x f(x)dx + \int_{-\infty}^{0} x f(x)dx \\
&= \int_0^{\infty} \left(\int_0^x dy \right) f(x)dx - \int_{-\infty}^{0} \left(\int_0^{-x} dy \right) f(x)dx \\
&= \int_0^{\infty} \int_y^{\infty} f(x)dxdy - \int_0^{\infty} \int_{-\infty}^{-y} f(x)dxdy \\
&= \int_0^{\infty} P(Y > y)dy - \int_0^{\infty} P(Y \le -y)dy \\
&= \int_0^{\infty} [1 - F_Y(y)]dy - \int_0^{\infty} F_Y(-y)dy\ .
\end{aligned}$$

This is illustrated in Figure 2.3. ■

Sometimes we are faced with a situation where we must deal not with the random variable whose distribution is known but rather with some function of the random variable as discussed in the previous section.

Suppose that X is a random variable. Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be a function such that $Y = g(X)$ is also a random variable and whose expected value exists. It is clear that, to calculate $E(Y)$, we need to find the density function of the random variable Y. However, fortunately there is a method that allows us to

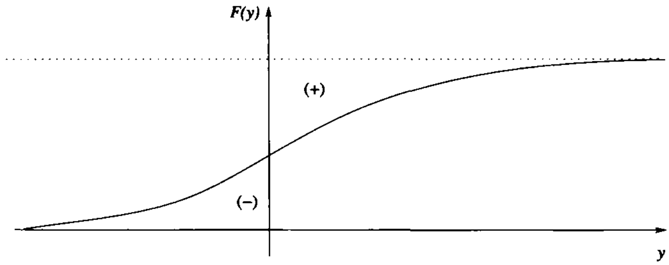


Figure 2.3 Illustration of expectation value for continuous random variable

do the calculations in a simpler way, known as the "*law of the unconscious statistician*". This method has been stated in the following theorem.

Theorem 2.6 *Let X be a real random variable with density function f_X and $g : \mathbb{R} \longrightarrow \mathbb{R}$ a function such that $Y = g(X)$ is a random variable. Then*

$$E(g(X)) = \begin{cases} \sum_x g(x)p_X(x) & \text{if } X \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} g(x)f_X(x)dx & \text{if } X \text{ is an absolutely continuous random variable} \end{cases}$$

whenever the sum, in the discrete case, or the integral in the continuous case converges absolutely.

Proof:

1. Suppose that X is a discrete random variable that takes the values $x_1, x_2, \cdots$. In this case the probability mass function of X is given by:

$$p_X(x) = \begin{cases} P(X = x) & \text{if } x = x_1, x_2, \cdots \\ 0 & \text{otherwise.} \end{cases}$$

The random variable $Y = g(X)$ takes the values $g(x_1), g(x_2), \cdots$. It is clear that some of these values can be the same. Let us suppose that y_j with $j \geq 1$ represents different values of $g(x_i)$. Then, grouping all the

$g(x_i)$ that have the same value, it can be obtained that:

$$\begin{aligned}\sum_i g(x_i)p_X(x_i) &= \sum_j \sum_{i:g(x_i)=y_j} g(x_i)p_X(x_i)\\ &= \sum_j y_j \sum_{i:g(x_i)=y_j} p_X(x_i)\\ &= \sum_j y_j P(g(X)=y_j)\\ &= \sum_j y_j P(Y=y_j)\\ &= E(Y) = E(g(X)).\end{aligned}$$

2. Let X be an absolutely continuous random variable with density function f_X. Suppose that g is a nonnegative function. Then

$$\begin{aligned}E(g(X)) &= \int_0^\infty P(g(X)>y)dy - \int_0^\infty P(g(X)\le -y)dy\\ &= \int_0^\infty P(g(X)>y)dy\\ &= \int_0^\infty \left(\int_B f_X(x)dx\right)dy\end{aligned}$$

where $B := \{x : g(x) > y\}$. Therefore:

$$\begin{aligned}E(g(X)) &= \int_0^\infty \int_0^{g(x)} f_X(x)dydx\\ &= \int_0^\infty g(x)f_X(x)dx.\end{aligned}$$

The proof for the general case requires results that are beyond the scope of this text. The interested reader may find them in the text of Ash (1972).

■

■ EXAMPLE 2.46

Let X be a continuous random variable with density function given by:

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1\\ 0 & \text{otherwise.}\end{cases}$$

Then:

$$\begin{aligned} E(X^3) &= \int_{-\infty}^{\infty} x^3 f(x)dx \\ &= \int_0^1 2x^4 dx \\ &= \frac{2}{5}. \quad \blacktriangle \end{aligned}$$

Next, some important properties of the expected value of a random variable are presented.

Theorem 2.7 *Let X be a real random variable.*

1. *If $P(X \geq 0) = 1$ and $E(X)$ exists, then, $E(X) \geq 0$.*
2. *$E(\alpha) = \alpha$ for every constant α.*
3. *If X is bounded, that is, there exists a real constant $M > 0$ such that $P(|X| \leq M) = 1$, then $E(X)$ exists.*
4. *If α and β are constants and if g and h are functions such that $g(X)$ and $h(X)$ are random variables whose expected values exist, then the expected value of $(\alpha g(X) + \beta h(X))$ exists and:*

$$E(\alpha g(X) + \beta h(X)) = \alpha E(g(X)) + \beta E(h(X)).$$

5. *If g and h are functions such that $g(X)$ and $h(X)$ are random variables whose expected values exist and if $g(x) \leq h(x)$ for all x, then:*

$$E(g(X)) \leq E(h(X)).$$

In particular:

$$|E(X)| \leq E(|X|) \ .$$

Proof:

1. Suppose that X is a discrete random variable that takes the values $x_1, x_2, \cdots$. Given that $P(X < 0) = 0$, it is obtained that $P(X = x_j) = 0$ for all $x_j < 0$. Therefore:

$$\begin{aligned} E(X) &= \sum_i x_i P(X = x_i) \\ &= \sum_{i:x_i \geq 0} x_i P(X = x_i) \geq 0 \ . \end{aligned}$$

If X is an absolutely continuous random variable with density function f, then:

$$
\begin{aligned}
E(X) &= \int_0^\infty [1 - F_X(x)]dx - \int_0^\infty F_X(-x)dx \\
&= \int_0^\infty [1 - F_X(x)]dx \\
&= \int_0^\infty P(X > x)dx \geq 0 .
\end{aligned}
$$

2.

$$
\begin{aligned}
E(\alpha) &= \alpha P(X = \alpha) \\
&= \alpha .
\end{aligned}
$$

3. If X is a discrete random variable that takes values $x_1, x_2, \cdots$, then, because $P(|X| > M) = 0$, it can be supposed that $\{x_1, x_2, \cdots\} \subseteq [-M, M]$. In conclusion:

$$
\sum_i |x_i| \, P(X = x_i) \leq M \sum_i P(X = x_i) = M < \infty.
$$

If X is an absolutely continuous random variable with density function f, then, given that $P(|X| > M) = 0$, it can be supposed that $f(x) = 0$ for all $x \notin [-M, M]$. Therefore:

$$
\begin{aligned}
\int_{-\infty}^{\infty} |x| \, f(x)dx &= \int_{-M}^{M} |x| \, f(x)dx \\
&\leq M \int_{-M}^{M} f(x)dx = M < \infty.
\end{aligned}
$$

4. The proof for the continuous case is elaborated here while the discrete case is left as an exercise. If X is an absolutely continuous random variable with density function f, then:

$$
\begin{aligned}
\int_{-\infty}^{\infty} |\alpha g(x) + \beta h(x)| \, f(x)dx &\leq \int_{-\infty}^{\infty} |\alpha g(x)| \, f(x)dx \\
&\quad + \int_{-\infty}^{\infty} |\beta h(x)| \, f(x)dx \\
&= |\alpha| \int_{-\infty}^{\infty} |g(x)| \, f(x)dx \\
&\quad + |\beta| \int_{-\infty}^{\infty} |h(x)| \, f(x)dx < \infty.
\end{aligned}
$$

The expected value of $(\alpha g(X) + \beta h(X))$ exists and it is clear that:

$$\int_{-\infty}^{\infty} [\alpha g(x) + \beta h(x)] f(x) dx = \int_{-\infty}^{\infty} [a g(x)] f(x) dx + \int_{-\infty}^{\infty} [\beta h(x)] f(x) dx$$
$$= \alpha E(g(X)) + \beta E(h(X)).$$

5. The proof for the continuous case is elaborated here while the discrete case is left as an exercise. If X is an absolutely continuous random variable with density function f, then:

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx \leq \int_{-\infty}^{\infty} h(x) f(x) dx = E(h(X)).$$

Given that
$$-|X| \leq X \leq |X|$$
it is obtained that:
$$E(-|X|) \leq E(X) \leq E|X|$$
$$-E(|X|) \leq E(X) \leq E(|X|).$$

That is:
$$|E(X)| \leq E(|X|).$$

■

The expected value of the random variable is the "first central moment" of a variable around zero. In general, central moments around zero of random variables are the expected values of the power of the variable. More precisely:

Definition 2.7 (Central Moment Around Zero)
Let X be a real random variable. The rth central moment of X around zero, denoted by μ_r', is defined as

$$\mu_r' := E(X^r)$$

whenever the expected value exists.

The following result allows to confirm that if $s < r$ and μ_r' exists, then μ_s' exists.

Theorem 2.8 *If X is a random variable such that μ_r' exists, then μ_s' exists for all $s < r$.*

Proof: Proof for the continuous case is elaborated here while the discrete case is left for the reader.

Given that
$$|x^s| < 1 + |x^r| \quad \text{for } x \in \mathbb{R}$$

it is obtained that:

$$\begin{aligned}\int_{-\infty}^{\infty} |x^s| f_X(x)dx &\leq \int_{-\infty}^{\infty} [1 + |x^r|] f_X(x)dx \\ &= 1 + \int_{-\infty}^{\infty} |x^r| f_X(x)dx < \infty.\end{aligned}$$

■

Definition 2.8 (Central Moment Around the Average) *Let X be a real random variable whose expected value exists. The rth central moment of X around $E(X)$ is defined as*

$$\mu_r := E([X - E(X)]^r)$$

whenever the expected value exists.

The following result allows to relate the rth order moments around the mean with the rth order moments around the origin.

Theorem 2.9 *Let X be a random variable whose expected value exists. If μ_r exists, then:*

$$\mu_r = \sum_{k=0}^{r} \binom{r}{k} \mu'_k (-\mu'_1)^{r-k}.$$

Proof:

$$\begin{aligned}\mu_r &= E\left(X - E(X)\right)^r \\ &= E\left(X - \mu'_1\right)^r \\ &= E\left(\sum_{k=0}^{r} \binom{r}{k} X^k (-\mu'_1)^{r-k}\right) \\ &= \sum_{k=0}^{r} \binom{r}{k} E\left(X^k\right) \left(-\mu'_1\right)^{r-k} \\ &= \sum_{k=0}^{r} \binom{r}{k} \mu'_k (-\mu'_1)^{r-k}.\end{aligned}$$

■

Note that for any random variable whose expected value exists it is satisfied that $\mu_0 = 1$ and $\mu_1 = 0$. The second central moment of X, around the mean, receives the name of *variance* of the random variable X and it is generally denoted by σ_X^2; the square root of the variance is called *standard deviation* of X and it is denoted usually by σ_X. Other notations with frequent use for the variance of X are $Var(X)$ and $V(X)$. In this text, any of these notations will be used indistinctively.

The variance measures the dispersion of the values of the variable around its mean. The term $(X - E(X))^2$ is the square of the distance from X to $E(X)$ and therefore, $E(X - E(X))^2$ represents the average of the squares of the distances of each value from X to $E(X)$. Then, if a random variable has a small variance, the possible values of X will be very close to the mean, while if X has large variance, then the values of X tend to be far away from the mean.

In the applications (see Ospina, 2001) a measure of relative dispersion is commonly used. It is called variation coefficient and is defined as:

$$CV(X) := \frac{\sigma_X}{E(X)} \text{ if } E(X) \neq 0.$$

When $|E(X)|$ is not near zero, $CV(X)$ is used as an indicator of how large the variance is. Empirically, it has been seen that when $|CV(X)| < 0.1$ the variance generally is small.

Some important properties of the variance of a random variable are presented next.

Theorem 2.10 *Let X be a random variable whose expected value exists and $\alpha, \beta \in \mathbb{R}$ are constants. Then:*

1. $Var(X) \geq 0$.

2. $Var(\alpha) = 0$.

3. $Var(\alpha X) = \alpha^2 Var(X)$.

4. $Var(X + \beta) = Var(X)$.

5. $Var(X) = 0$ *if and only if* $P(X = E(X)) = 1$.

Proof:

1. It is clear from the definition of variance and from the properties of the expected value.

2.
$$Var(\alpha) = E(\alpha - E(\alpha))^2 = E(0) = 0.$$

3.
$$\begin{aligned} Var(\alpha X) &= E[\alpha X - E(\alpha X)]^2 \\ &= E(\alpha X - \alpha E(X))^2 \\ &= \alpha^2 E(X - E(X))^2 \\ &= \alpha^2 Var(X). \end{aligned}$$

4.

$$\begin{aligned} Var(X+\beta) &= E[(X+\beta)-E(X+\beta)]^2 \\ &= E[X+\beta-E(X)-\beta]^2 \\ &= Var(X). \end{aligned}$$

5. (a) If $X = E(X)$ with probability 1, it is clear that $Var(X) = 0$.

   (b) Suppose that $Var(X) = 0$, and let $a := E(X)$.
   If $P(X = a) < 1$, then $c > 0$ exists such that

$$P((X-a)^2 > c) > 0$$

   given that

$$(x-a)^2 \geq c\mathcal{X}_{\{(x-a)^2>c\}}.$$

   Then

$$\begin{aligned} E(X-a)^2 &\geq E(c\mathcal{X}_{\{(X-a)^2>c\}}) \\ Var(X) &\geq cE(\mathcal{X}_{\{(X-a)^2>c\}}) \\ Var(X) &\geq cP((X-a)^2>c) > 0, \end{aligned}$$

   which is a contradiction. Therefore $P(X = E(X)) = 1$.

■

To calculate the variance of a random variable, the following result will be very useful.

Theorem 2.11 *Let X be a random variable whose $E(X^2)$ exists. Then:*

$$Var(X) = E(X^2) - (E(X))^2.$$

Proof:

$$\begin{aligned} Var(X) &= E(X-E(X))^2 \\ &= E(X^2 - 2XE(X) + (E(X))^2) \\ &= E(X^2) - 2E(X)E(X) + E(E(X))^2 \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 \\ &= E(X^2) - (E(X))^2. \end{aligned}$$

■

EXAMPLE 2.47

Suppose that a fair die is thrown once and let X be a random variable that represents the result obtained. It is known that $E(X) = \frac{21}{6}$ and:

$$E(X^2) = \sum_{k=1}^{6} \frac{1}{6} k^2 = \frac{91}{6}.$$

Then:

$$Var(X) = \frac{35}{12} \approx 2.92 \,. \qquad \blacktriangle$$

EXAMPLE 2.48

Let X be defined as in Example 2.41. Then:

$$E(X^2) = P(X = 1) = P(A).$$

Therefore:

$$\begin{aligned} Var(X) &= P(A) - [P(A)]^2 \\ &= P(A)P(A^c). \qquad \blacktriangle \end{aligned}$$

EXAMPLE 2.49

Let X be defined as in Example 2.42. Then in this case:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 e^{-3} \frac{3^k}{k!} \\ &= e^{-3} \sum_{k=1}^{\infty} k \frac{3^k}{(k-1)!} \\ &= e^{-3} \sum_{j=0}^{\infty} (j+1) \frac{3^{j+1}}{j!} \\ &= e^{-3} [9e^3 + 3e^3] \\ &= 12. \end{aligned}$$

Therefore:

$$Var(X) = 12 - 9 = 3. \qquad \blacktriangle$$

■ EXAMPLE 2.50

Let X be defined as in Example 2.43. Then:

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$$
$$= \int_0^{\infty} 2x^2 e^{-2x} dx$$
$$= \frac{1}{2}.$$

Therefore:

$$Var(X) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}. \quad \blacktriangle$$

The probability generating function (pgf) plays an important role in many areas of applied probability, in particular, in stochastic processes. In this section we will discuss pgf's, which are defined only for nonnegative integer valued random variables.

Definition 2.9 (Probability Generating Functions) *Let X be a nonnegative integer-valued random variable and let $p_k = P(X = k)$, $k = 0, 1, 2, \cdots$ with $\sum_{k=0}^{\infty} p_k = 1$. The pgf of X is defined as:*

$$G_X(s) = E\left(s^X\right) = \sum_{k=0}^{\infty} p_k s^k, \quad |s| < 1.$$

Since $G_X(1) = 1$, the series converges for $|s| \leq 1$.

■ EXAMPLE 2.51

For $\lambda > 0$, let X be a discrete random variable with probability mass function given by:

$$p(k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad k = 0, 1, 2, \cdots.$$

Then:

$$G_X(s) = E\left(s^X\right) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} s^k$$
$$= e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} s^k$$
$$= e^{\lambda(s-1)}. \quad \blacktriangle$$

EXAMPLE 2.52

Let X be a discrete random variable with probability mass function given by

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and $p \in (0,1)$. Then:

$$\begin{aligned} G_X(s) &= E[s^X] \\ &= \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} s^k \\ &= \sum_{k=0}^{n} \binom{n}{k} (ps)^k q^{n-k} \\ &= (ps+q)^n. \quad \blacktriangle \end{aligned}$$

Note 2.6 *If X and Y have pgf's $G_X(s)$ and $G_Y(s)$, respectively, then $G_X(s) = G_Y(s)$ if and only if $X \stackrel{d}{=} Y$.*

Theorem 2.12 *Let X be a nonnegative integer-valued random variable with $E(|X|^k) < \infty$ for all $k = 1, 2, \cdots$. Then, $G_X^{(r)}(1) = E[X(X-1)\cdots(X-r+1)]$ where $G_X^{(r)}(1)$ is the rth derivative of its pgf $G_X(s)$ at $s = 1$.*

Proof:

$$G_X(s) = \sum_{n=0}^{\infty} p_n s^n \quad (|s| < 1)$$

with $p_n = P(X = n), n = 0, 1, \cdots$. Differentiate with respect to s, and it is assumed that the order of summation and differentiation can be interchanged. We have:

$$G_X'(s) = \sum_{n=1}^{\infty} n p_n s^{n-1}.$$

When $s = 1$:

$$G_X'(1) = \sum_{n=1}^{\infty} n p_n = E(X) = G_X'(1).$$

In general:

$$\begin{aligned} G_X^{(r)}(s) &= \frac{d^r}{ds^r}[G_X(s)] \\ &= \frac{d^r}{ds^r}\left[\sum_{n=0}^{\infty} p_n s^n\right] \\ &= \sum_{n=0}^{\infty} p_n n(n-1)\cdots(n-r+1)s^{n-r}. \end{aligned}$$

Since the series is convergent for $|s| \leq 1$ and using Abel's lemma:

$$G_X^{(r)}(1) = E[X(X-1)\cdots(X-r+1)], \quad r \geq 1.$$

■

■ EXAMPLE 2.53

If $G_X(s) = e^{\lambda(s-1)}$, $|s| < 1$, then:

$$\begin{aligned} G'_X(s) &= \lambda e^{\lambda(s-1)} \\ E(X) &= G'_X(1) = \lambda \\ G_X^{(2)}(s) &= \lambda^2 e^{\lambda(s-1)} \\ Var(X) &= \lambda^2 - \lambda^2 + \lambda = \lambda. \end{aligned} \quad \blacktriangle$$

The moments of a random variable X have a very important role in statistics not only for the theory but also for applied statistics. Due to this, it is very convenient to have mechanisms that allow easy calculations for the moments of the random variable. This mechanism is provided by the so-called moment generating function which will be as follows.

Definition 2.10 (Moment Generating Function) *Let X be a random variable such that $E(e^{tX})$ is finite for all $t \in (-\alpha, \alpha)$ with real positive α. The moment generating function (mgf) of X, denoted by $m_X(\cdot)$, is defined as:*

$$m_X(t) = E(e^{tX}) \quad \textit{with } t \in (-\alpha, \alpha).$$

That is:

$$m_X(t) = \begin{cases} \sum_k e^{tx_k} p_X(x_k) & \textit{if } X \textit{ is a discrete random variable} \\ & \quad \textit{with pmf } p_X(x) \\ \int_{-\infty}^{\infty} e^{tx} f(x)dx & \textit{if } X \textit{ is a continuous random variable} \\ & \quad \textit{with pdf } f(x). \end{cases}$$

It is important to notice that not all probability distributions have moment generating functions associated with them. Later on some examples will be given that ratify this statement.

Before giving the important properties of the mgf of a random variable X, some examples are presented.

EXAMPLE 2.54

Let X be a discrete random variable with probability mass function given by

$$p_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and $p \in (0,1)$.

In this case:

$$\begin{aligned} m_X(t) &= \sum_{k=0}^{n} e^{tk} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{k=0}^{n} \binom{n}{k} (pe^t)^k (1-p)^{n-k} \\ &= (pe^t + q)^n \quad \text{where } q := 1-p. \qquad \blacktriangle \end{aligned}$$

Note 2.7 *The distribution shown in Example 2.54 receives the name of binomial distribution with parameters n and p. In the next chapter, this distribution will be studied in detail.*

EXAMPLE 2.55

Let X be the random variable as in Example 2.42. In this case:

$$\begin{aligned} m_X(t) &= \sum_{k=0}^{\infty} e^{tk} e^{-3} \frac{3^k}{k!} \\ &= e^{-3} \sum_{k=0}^{\infty} \frac{(3e^t)^k}{k!} \\ &= e^{-3} \exp(3e^t) \\ &= \exp(3(e^t - 1)). \qquad \blacktriangle \end{aligned}$$

EXAMPLE 2.56

Let X be a random variable with density function given by:

$$f(x) = \begin{cases} 2e^{-2x} & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that:

$$\begin{aligned} m_X(t) &= \int_0^\infty e^{tx} 2e^{-2x}\,dx \\ &= \left(1 - \frac{t}{2}\right)^{-1} \quad \text{with } t < 2. \qquad \blacktriangle \end{aligned}$$

■ EXAMPLE 2.57

Let X be a random variable with density function given by

$$f(x) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi x}} \exp\left\{-\frac{1}{2}\left[\frac{\ln(x)-\mu}{\sigma}\right]^2\right\} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ are constants.

Then:

$$E(e^{tX}) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^\infty \frac{1}{x} \exp\left\{tx - \frac{1}{2}\left[\frac{\ln(x) - \mu}{\sigma}\right]^2\right\} dx.$$

With a change of variable

$$u = \frac{\ln(x) - \mu}{\sigma}$$

it is obtained that:

$$\begin{aligned} E(e^{tX}) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^\infty \exp\left\{te^{\mu+\sigma u} - \frac{u^2}{2}\right\} du \\ &\geq \frac{1}{\sqrt{2\pi}} \int_0^\infty \exp\left\{te^{\mu+\sigma u} - \frac{u^2}{2}\right\} du. \end{aligned}$$

Given that $\exp(\sigma u) > \frac{(\sigma u)^3}{3!}$ it is deduced that for $t > 0$:

$$te^{\mu+\sigma u} - \frac{u^2}{2} > \frac{u^2}{2}\left[\frac{te^{\mu}\sigma^3}{3}u - 1\right].$$

Therefore, if $u > \frac{6}{te^{\mu}\sigma^3}$ is taken, it is obtained that

$$\left[\frac{te^{\mu}\sigma^3}{3}u - 1\right] > 1\ .$$

Thus by taking

$$a = \frac{6}{te^{\mu}\sigma^3}$$

we obtain:

$$E(e^{tX}) \geq \frac{1}{\sqrt{2\pi}} \int_a^\infty \exp(\frac{u^2}{2})du = \infty.$$

This means that there does not exist a neighborhood of the origin in which $E(e^{tX})$ is finite and in consequence the moment generating function of X does not exist. ▲

The existence of the moment generating function of a random variable X guarantees the existence of all the rth order moments around zero of X. If the moment generating function exists, then it is differentiable in a neighborhood of the origin and satisfies

$$\left.\frac{d^r}{dt^r} m_X(t)\right|_{t=0} = E(X^r).$$

More precisely, the next two theorems state this and the proofs can be found in Hernandez (2003).

Theorem 2.13 *If X is a random variable whose moment generating function exists, then $E(X^r)$ exists for all $r \in \mathbb{N}$.*

It is important to clarify that the converse of the previous theorem is not valid: the fact that all the rth order moments of a random variable X exist does not guarantee the existence of the mgf of the variable. For example, for the random variable in Example 2.57, it is known that for all $r \in \mathbb{N}$:

$$E(X^r) = \exp\left\{r\mu + \frac{(r\sigma)^2}{2}\right\}.$$

Theorem 2.14 *If X is a random variable whose moment generating function $m_X(.)$ exists, then, $h \in (0, \infty)$ exists such that:*

$$m_X(t) = \sum_{k=0}^{\infty} E(X^k)\frac{t^k}{k!}; \quad \textit{for all } t \in (-h, h).$$

Therefore:

$$E(X^r) = \left.\frac{d^r}{dt^r} m_X(t)\right|_{t=0}.$$

An important property of the mgf of a random variable is that, when it exists, it characterizes the distribution of the variable. More precisely, the following theorem is obtained. Its proof is beyond the scope of this text. The interested reader may refer to Ash (1972).

Theorem 2.15 *Let X and Y be random variables whose moment generating functions exist. If*

$$m_X(t) = m_Y(t) \textit{ for all } t,$$

then X and Y have the same distribution.

As it can be seen, the moment generating function of a random variable, when it exists, is a very helpful tool to calculate the rth order moments around zero of a variable. Unfortunately, this function does not always exist, and thus it is necessary to introduce a new class of functions that are equally useful and always exist.

Definition 2.11 (Characteristic Function) *Let X be a random variable. The characteristic function of X is the function $\varphi_X : \mathbb{R} \longrightarrow \mathbb{C}$ defined by:*

$$\varphi_X(t) := E(e^{itX}) = E(\cos tX) + iE(\sin tX) \quad \textit{where } i = \sqrt{-1}.$$

■ **EXAMPLE 2.58**

Let X be a random variable with:

$$P(X = 1) = P(X = -1) = \frac{1}{2}.$$

Then:

$$\begin{aligned}\varphi_X(t) &= \frac{1}{2}(\cos t + \cos(-t)) + \frac{i}{2}(\sin t + \sin(-t)) \\ &= \cos t. \qquad \blacktriangle\end{aligned}$$

■ **EXAMPLE 2.59**

Let X be a discrete random variable with probability mass function given in Example 2.42. Then:

$$\begin{aligned}\varphi_X(t) &= E(e^{itX}) \\ &= \sum_{k=0}^{\infty} e^{itk} e^{-3} \frac{3^k}{k!} \\ &= e^{-3} \sum_{k=0}^{\infty} \frac{(3e^{it})^k}{k!} \\ &= e^{-3} \exp(3e^{it}) \\ &= \exp[3(e^{it} - 1)]. \qquad \blacktriangle\end{aligned}$$

■ EXAMPLE 2.60

Let X be a continuous random variable with density function f given by:

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise.} \end{cases}$$

Then, for $t \neq 0$

$$\begin{aligned} \varphi_X(t) &= \int_0^1 \cos(tx)dx + i \int_0^1 \sin(tx)dx \\ &= \frac{1}{t}(\sin t - i\cos t + i) \\ &= \frac{1}{it}(e^{it} - 1) \end{aligned}$$

and for $t = 0$

$$\varphi_X(0) = E(e^0) = 1. \quad \blacktriangle$$

Next, some principal properties of the characteristic function, without proof, will be presented. The interested reader may consult these proofs in Grimmett and Stirzaker (2001) and Hernandez (2003).

Theorem 2.16 *If X is a discrete or absolutely continuous random variable, then $E(e^{itX})$ exists for all $t \in \mathbb{R}$.*

Theorem 2.17 *Let X be a random variable. The characteristic function $\varphi_X(.)$ of X satisfies:*

1. $\varphi_X(0) = 1$.
2. $|\varphi_X(t)| \leq 1$ *for all t.*
3. *If $E(X^k)$ exists, then:*

$$\frac{d^k}{dt^k}\varphi_X(t) \mid_{t=0} = i^k E(X^k).$$

Finally, it is important to notice that the characteristic function of a random variable, as in the case of mgf (when it exists), determines the distribution of the variable. That is, it satisfies:

Theorem 2.18 *If X and Y are random variables and*

$$\varphi_X(t) = \varphi_Y(t) \text{ for all } t,$$

then X and Y have the same distribution.

EXERCISES

2.1 Let $\Omega = \{1, 2, 3\}$, $\Im = \{\emptyset, \{1\}, \{2, 3\}, \Omega\}$. Let us define, for $A \in \Im$:

$$P(A) := \begin{cases} 1 & \text{if } 1 \in A \\ 0 & \text{if } 1 \notin A. \end{cases}$$

Is $X : \Omega \longrightarrow \mathbb{R}$ defined by $X(\omega) = \omega^2$ a real random variable? Justify.

2.2 Let $\Omega = \{1, 2, 3, 4\}$. Determine the least σ-algebra over Ω so that $X(\omega) := \omega + 1$ is a real random variable.

2.3 A coin is biased in such a way that $P(H) = \frac{3}{7}$ and $P(T) = \frac{4}{7}$. Suppose that the coin is tossed three consecutive times and let X be the random variable that indicates the number of heads obtained. Find the distribution function of the random variable X and calculate $E(X)$.

2.4 The sample space of a random experiment is $\Omega = \{a, b, c, d, e, f\}$ and each result is equally probable. Let us define the random variable X as follows:

ω	a	b	c	d	e	f
$X(\omega)$	0	0	1	1	-1	2

Calculate the following probabilities:

a) $P(X = 1)$.

b) $P(|X - 1| \leq 1)$.

c) $P(X \geq 0 \text{ or } X < 2)$.

2.5 Prove that if all values of a real random variable X are in the interval $[a, b]$ with $a < b$, then $F_X(x) = 0$ for all $x < a$ and $F_X(x) = 1$ for all $x \geq b$.

2.6 A fair coin is tossed four consecutive times. Let X be the random variable that denotes the number of heads obtained. Find the distribution function of the random variable $Y := X - 2$ and graph it.

2.7 Let $(\Omega, \Im, P)$ be a probability space defined as follows:

$$\Omega := \{(1, 2, 3), (2, 3, 1), (3, 1, 2)\}, \ \Im := \wp(\Omega) \text{ and } P(\omega) := \frac{1}{3}$$

for all $\omega \in \Omega$. Consider the random variables X_n defined over Ω, as:

$$X_n(\omega) = \omega_n \text{ for } n = 1, 2, 3 \text{ and } \omega = (\omega_1, \omega_2, \omega_3) \in \Omega.$$

a) Determine the set of values S taken by the random variable X_n.

b) Verify that:

$$P(X_3 = 3 \mid X_2 \in \{1, 2\}, \ X_1 = 3) \neq P(X_3 = 3 \mid X_2 \in \{1, 2\}).$$

2.8 A box contains 5 white balls and 10 black ones. A die is thrown. A number of balls equal to the result obtained in the die are taken out of the box. What is the probability that all balls taken out of the box are white balls? What is the probability that the result obtained when throwing the die is 3 if all the balls taken out of the box are white?

2.9 Let X be a random variable with density function given by:

$$f(x) = \begin{cases} 0.2 & \text{if } -1 < x \leq 0 \\ 0.2 + cx & \text{if } 0 < x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

a) Determine the value of c.

b) Obtain the distribution function of the random variable X.

c) Calculate $P(0 \leq X < 0.5)$.

d) Determine $P(X > 0.5 \mid X > 0.1)$.

e) Calculate the distribution function and the density function of the random variable $Y := 2X + 3$.

2.10 Let X be a random variable with density function given by:

$$f(x) = \begin{cases} c(2-x) & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

a) Calculate the value of c.

b) Determine the distribution function of the random variable X.

c) Calculate $P(|X| \geq 0.2)$.

2.11 The cumulative distribution function of a random variable X is given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 2x - x^2 & \text{if } 0 \leq x \leq 1 \\ 1 & \text{if } x > 1. \end{cases}$$

a) Calculate $P\left(X \geq \frac{3}{2}\right)$ and $P\left(-2 \leq X \leq \frac{3}{4}\right)$.

b) Determine the density function $f_X(\cdot)$.

2.12 Let X, Y and Z be random variables whose distribution functions are respectively:

$$F_X(x) = \begin{cases} 0 & \text{if } -\infty < x < -1 \\ 0.2 & \text{if } -1 \leq x < 0 \\ 0.7 & \text{if } 0 \leq x < 1 \\ 0.8 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

$$F_Y(x) = \begin{cases} 0 & \text{if } -\infty < x < 0 \\ \frac{1}{2} + \frac{1}{2}x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

$$F_Z(x) = \begin{cases} 0 & \text{if } -\infty < x < -1 \\ \frac{1}{2} + \frac{1}{2}x & \text{if } -1 \leq x < 1 \\ 1 & \text{if } x \geq 1 \ . \end{cases}$$

a) Which variables are discrete variables? Which variables are continuous variables? Explain.

b) Calculate $P(X = 0)$, $P(\frac{1}{2} < X \leq 2)$ and $P(X > 1.5)$.

c) Calculate $P(Y = 0)$, $P(\frac{1}{2} < Y \leq 2)$ and $P(Y > 0)$.

d) Calculate $P(Z = 0)$, $P(-\frac{1}{2} < Z \leq \frac{1}{2})$ and $P(Z \geq 2)$.

2.13 A circular board with radius 1 is sectioned in n concentric discs with radii $\frac{1}{n}, \frac{2}{n}, \cdots, 1$. A dart is thrown randomly inside the circle. If it hits the ring between the circles with radii $\frac{i}{n}$ and $\frac{i+1}{n}$ for $i = 0, \cdots, n-1$, $n-i$ monetary units are won. Let X be the random variable that denotes the amount of money won. Find the probability mass function of the random variable X.

2.14 (Grimmett and Stirzaker, 2001) In each of the following exercises determine the value of the constant C so that the functions given are probability mass functions over the positive integers:

a) $p(x) = C2^{-x}$.

b) $p(x) = \frac{C2^{-x}}{x}$.

c) $p(x) = Cx^{-2}$.

d) $p(x) = \frac{C2^{x}}{x!}$.

2.15 (Grimmett and Stirzaker, 2001) In each of the following exercises determine the value of the constant C so that the functions given are probability mass functions:

a) $p(x) = C\{x(1-x)\}^{-\frac{1}{2}}, \quad 0 < x < 1.$

b) $p(x) = C\exp(-x - e^{-x}),\ x \in \mathbb{R}$.

2.16 An absolutely continuous random variable X takes values in the interval $[0, 4]$ and its density function is given by:

$$f(x) = \frac{1}{2} - cx.$$

a) Determine the value of c.

b) Calculate $P(\frac{1}{2} \leq X < 3)$.

2.17 Let X be an absolutely continuous random variable with density function f. Prove that the random variables X and $-X$ have the same distribution function if and only if $f(x) = f(-x)$ for all $x \in \mathbb{R}$.

2.18 In the following cases, determine the distribution function of the discrete random variable X whose probability mass function is given by:

a) $P(X = k) = pq^{k-1}$, $k = 1, 2, \cdots$ with $p \in (0,1)$ fixed and $q := 1 - p$.

b)

$$P(X = k) = \frac{6k^2}{n(n+1)(2n+1)}, \quad k = 1, \cdots, n.$$

2.19 A person asks for a key ring that has seven keys but he does not know which is the key that will open the lock. Therefore, he tries with each one of the keys until he opens the lock. Let X be the random variable that indicates the number of tries needed to achieve the goal of opening the lock.

a) Determine the density function of the random variable X.

b) Calculate $P(X \leq 2)$ and $P(X = 5)$.

2.20 Four balls are taken out randomly and without replacement from each box that contains 25 balls numbered from 1 to 25. If you bet that at least 1 of the 4 balls taken out has a number less than or equal to 5, what is the probability that you win the bet?

2.21 A player takes out, simultaneously and randomly, 2 balls from a box that contains 8 white balls, 5 black balls and 3 blue balls. Suppose that the player wins 5000 pesos for each black ball selected and loses 3000 pesos for every white ball selected. Let X be the random variable that denotes the player's fortune. Find the density function of the random variable X.

2.22 A salesman has two different stores where he sells computers. The probability that he sells, in one day, a computer in the first store is 0.4 and independently, the probability that he sells, in one day, a computer in the second store is 0.7. Additionally, suppose that it is equally probable that he sells a computer of type 1 or type 2. A type 1 computer costs \$1800 while the type 2 computer, with the same specifications, costs \$1000. Let X be the amount, in dollars, that the salesman sells in one day. Find the distribution of the random variable X.

2.23 Prove that

$$\int_{-\infty}^{\infty} \exp(-x^2)dx = \sqrt{\pi}$$

and use the above result to show that

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad x \in \mathbb{R},$$

is a density function if $\sigma > 0$.

2.24 Suppose that f and g are density functions and that $0 < \lambda < 1$ is a constant. Is $\lambda f + (1-\lambda)\, g$ a density function? Is fg a density function? Explain.

2.25 Let X be a random variable with cumulative distribution function given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{1}{4}x + \frac{1}{4} & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

$$F_c(x) = \begin{cases} 0 & \text{if } x < 1 \\ x - 1 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2. \end{cases}$$

Determine a cumulative discrete distribution function $F_d(\cdot)$ and one continuous $F_c(\cdot)$ and the constants α and β with $\alpha + \beta = 1$ such that:

$$F_X(x) = \alpha F_d(x) + \beta F_c(x).$$

2.26 Let X be a random variable with cumulative distribution function given by:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \\ \frac{1}{25} & \text{if } 1 \leq x < 2 \\ \\ x^2 & \text{if } 2 \leq x < \frac{3}{5} \\ \\ 1 & \text{if } x \geq \frac{3}{5}. \end{cases}$$

Determine the cumulative discrete distribution function $F_d(\cdot)$ and one continuous $F_c(\cdot)$ and the constants α and β with $\alpha + \beta = 1$ such that:

$$F_X(x) = \alpha F_d(x) + \beta F_c(x).$$

2.27 We say that a discrete real random variable X has Fisher logarithmic distribution with parameter θ if its density function is given by

$$p(x) = \begin{cases} \frac{1}{|\ln(1-\theta)|} \frac{\theta^x}{x} & \text{if } x = 1, 2, \cdots \\ \\ 0 & \text{otherwise} \end{cases}$$

with $\theta \in (0, 1)$. Verify that p is a probability mass function.

2.28 Let X be a random variable with distribution function given by:

$$F(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{x^2-2x+2}{2} & \text{if } 1 \leq x \leq 2 \\ 1 & \text{if } x > 2. \end{cases}$$

Calculate $Var(X)$.

2.29 Let X and Y be two nonnegative continuous random variables having respective *cdf's* F_X and F_Y. Suppose that for some constants $a > 0$ and $b > 0$:

$$F_X(x) = F_Y\left(\frac{x-a}{b}\right).$$

Determine $E(X)$ in terms of $E(Y)$.

2.30 Let X be a continuous random variable with strictly increasing distribution function F. What type of distribution have the random variable $Y := -\ln(F_X(X))$?

2.31 Let X be a random variable which can only take values $-1, 0, 1$ and 2 with $P(X = 0) = 0.1$, $P(X = 1) = 0.4$ and $P(X = 2) = \frac{1}{2}P(X = -1)$. Find $E(X)$.

2.32 An exercise on a test for small children required to show the correspondence between each of three animal pictures with the word that identifies the animal. Let us define the random variable Y as the number of correct answers if a student assigns randomly the three words to the three pictures.

a) Find the probability distribution of Y.

b) Find $E(Y)$ and $Var(Y)$.

2.33 Let X be a random variable with density function given by:

$$f(x) = C\frac{1}{1+x^2} \text{ if } x \in \mathbb{R}.$$

a) Determine the value of C.

b) Calculate $P(X \geq 0)$.

c) Find (if they exist) $E(X)$ and $Var(X)$.

d) Find the distribution function of X.

2.34 A die is tossed two times. Let X and Y be the random variables defined by:

$$X := \text{“result of the first throw”}.$$
$$Y := \text{“result of the second throw”}.$$

Calculate $E(\max\{X,Y\})$ and $E(\min\{X,Y\})$.

2.35 Let X be a random variable with density function given by:

$$f(x) = \begin{cases} x^3 & \text{if } 0 < x \le 1 \\ (2-x)^3 & \text{if } 1 < x \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

a) Calculate $\mu := E(X)$ and $\sigma^2 := Var(X)$.

b) Find $P(\mu - 2\sigma < X < \mu + 2\sigma)$.

2.36 Let X be a random variable with density function given by:

$$f(x) = \begin{cases} 12x^3 - 21x^2 + 10x & \text{if } 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

a) Calculate $\mu := E(X)$ and $\sigma^2 := Var(X)$.

b) Determine the value of c so that $P(X > c) = \frac{7}{16}$.

2.37 Let X be a random variable with density function given by:

$$f(x) = \begin{cases} C\sin\left(\frac{1}{5}\pi x\right) & \text{if } 0 \le x \le 5 \\ 0 & \text{otherwise.} \end{cases}$$

a) Determine the value of C.

b) Find (if they exist) the mean and variance of X.

2.38 Suppose that X is a continuous random variable with *pdf* $f_X(x) = e^{-x}$ for $x > 0$. Find the *pdf* for the random variable Y given by:

$$Y = \begin{cases} X & \text{if } X \le 1 \\ \frac{1}{X} & \text{if } X > 1. \end{cases}$$

2.39 Let X be a random variable with continuous and strictly increasing cumulative distribution function.

a) Determine the density function of the random variable $Y := |X|$.

b) Find the cumulative distribution function of the random variable $Y := X^3$.

2.40 There are three boxes A, B and C. Box A contains 5 red balls and 4 white balls, box B contains 8 red balls and 5 white balls and box C contains 2 red balls and 6 white ones. A ball is taken out randomly from each box. Let X be the number of white balls taken out. Calculate the probability mass function of X.

2.41 Let X be a random variable with $P(X = 0) = 0$ and such that $E(X)$ exits and is different from zero.

a) Is it valid to say, in general, that:

$$E\left(\frac{1}{X}\right) = \frac{1}{E(X)} \ ? \tag{2.3}$$

b) Is there a random variable X such that equation (2.3) is satisfied? Explain.

2.42 Let X be a discrete random variable with values in the nonnegative integers and such that $E(X)$ and $E(X^2)$ exist.

a) Prove that:

$$E(X) = \sum_{k=0}^{\infty} P(X \geq k).$$

b) Verify that:

$$\sum_{k=0}^{\infty} kP(X > k) = \frac{1}{2}\left(E(X^2) - E(X)\right) .$$

2.43 Determine if the following propositions are true or false. Justify your answer.

a) If $P(X > Y) = 1$, then $E(X) > E(Y)$.

b) If $E(X) > E(Y)$, then $P(X > Y) = 1$.

c) If $Y = X + 1$, then $F_X(x) = F_Y(x+1)$ for all x.

2.44 Let X be a random variable such that:

$$P(X = 1) = p = 1 - P(X = -1) .$$

Find a constant $c \neq 1$ such that $E\left(c^X\right) = 1$.

2.45 Let X be a random variable with values in $\mathbb{Z}^+$with $P(X = k) = \frac{C}{3^k}$.

a) Determine the value of C.

b) Find (if it exists) $E(X)$.

2.46 Let X be a random variable with values $\frac{3^k}{2^k}$, $k = 0, 1, \cdots$, and such that $P\left(X = \frac{3^k}{2^k}\right) = \frac{1}{2^{k+1}}$. Does $E(X)$ exist? Does $Var(X)$ exists? Explain.

2.47 **(Markov's Inequality)** Let X be a real random variable with $X \geq 0$ and such that $E(X)$ exists. Prove that, for all $\alpha > 0$, it is satisfied that:

$$P(X \geq \alpha) \leq \frac{E(X)}{\alpha}.$$

2.48 **(Chebyschev's Inequality)** Let X be a random variable with mean μ and variance σ^2.

a) Prove that for all $\epsilon > 0$ it is satisfied that:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}.$$

b) If $\mu = \sigma^2 = 20$, what can be said about $P(0 \leq X \leq 40)$?

2.49 Let X be a random variable with mean 11 and variance 9. Use Chebyschev's inequality to find (if possible):

a) A lower bound for $P(6 < X < 16)$.

b) The value of k such that $P(|X - 11| \geq k) \leq 0.09$.

2.50 (Meyer, 1970) Let X be a random variable with mean μ and variance σ^2. Suppose that H is a function two times differentiable in $x = \mu$ and that $Y := H(X)$ is a random variable such that $E(Y)$ and $E(Y^2)$ exist.

a) Prove the following approximations for $E(Y)$ and $Var(Y)$:

$$E(Y) \approx H(\mu) + \frac{H''(\mu)}{2}\sigma^2$$

$$Var(Y) \approx (H'(\mu))^2 \sigma^2 .$$

b) Using the result in part (a), calculate (in an approximate way) the mean and variance of the random variable

$$Y := 2(1 - 0.005X)^{1.2}$$

where X is a random variable whose density function is given by:

$$f_X(x) = 3000x^{-4}\mathcal{X}_{[10,\infty)}(x) .$$

2.51 Find the characteristic function of the random variable X with probability function $P(X = 1) = \frac{2}{5}$ and $P(X = 0) = \frac{3}{5}$.

2.52 Determine the characteristic function of the random variable X with density function given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

where $a, b \in \mathbb{R}$ and $a < b$.

2.53 Determine the characteristic function of the random variable X with density function given by:

$$f(x) = \begin{cases} 1 - |x| & \text{if } -1 \leq x \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

2.54 **(Quartile of order q)** For any random variable X, a quartile of order q with $0 < q < 1$, is any number denoted by x_q that satisfies simultaneously the conditions

(i) $P(X \leq x_q) \geq q$

(ii) $P(X \geq x_q) \leq 1 - q$.

The most frequent quartiles are $x_{0.5}$, $x_{0.25}$ and $x_{0.75}$, named, respectively, *median, lower quartile* and *upper quartile.*

A quartile x_q is not necessarily unique. When a quartile is not unique, there exists an interval in which every point satisfies the conditions (i) and (ii). In this case, some authors suggest to consider the lower value of the interval and others suggest the middle point of the interval (see Hernández, 2003). Taking into account this information, solve the following exercises:

a) Let X be a discrete random variable with density function given by:

$$f_X(x) = \begin{cases} \binom{5}{x} \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x} & \text{if } x = 0, \cdots, 5 \\ 0 & \text{otherwise.} \end{cases}$$

Determine the lower quartile, the upper quartile and the median of X.

b) Let X be a random variable with density function given by:

$$f_X(x) = \begin{cases} \frac{1}{800} \exp\left(-\frac{x}{800}\right) & \text{if } x > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Determine the mean and median of X.

c) Let X be a random variable with density function given by:

$$f_X(x) = \begin{cases} \frac{10}{x^2} & \text{if } x > 10 \\ 0 & \text{otherwise.} \end{cases}$$

Determine (if they exist) the mean and the median of X.

2.55 (**Mode**) Let X be a random variable with density function $f(.)$. A mode of X (if it exists) is a real number ζ such that:

$$f(\zeta) \geq f(x) \text{ for all } x \in \mathbb{R}.$$

a) Let X be a discrete random variable with density function given by

x	1	2	3	4	5	6
$P(X = x)$	$\frac{9}{40}$	$\frac{1}{5}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{3}{20}$	$\frac{7}{40}$

Find (if it exists) the mode of X.

b) Suppose that the random variable X with density function given by

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} \exp(-\lambda x) & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $r > 0$ and $\lambda > 0$ are constants and Γ is the function defined by:

$$\Gamma(r) := \int_0^\infty t^{r-1} \exp(-t)\, dt .$$

If $E(X) = 28$ and the mode is 24, determine the values of r and λ.

c) Verify that if X is a random variable with density function given by

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } 0 < x < b \\ 0 & \text{otherwise} \end{cases}$$

with $a, b \in \mathbb{R}$, $a < b$, then any $\zeta \in (a, b)$ is a mode for X.

d) Verify that if X is a random variable with density function given by

$$f(x) = \begin{cases} \frac{1}{2} x^{-\frac{3}{2}} \exp\left(-x^{\frac{1}{2}}\right) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

then the mode of X does not exist.

2.56 Suppose that the time (in minutes) that a phone call lasts is a random variable with density function given by:

$$f(t) = \begin{cases} \frac{1}{5} \exp\left(\frac{-t}{5}\right) & \text{if } t > 0 \\ 0 & \text{otherwise .} \end{cases}$$

a) Determine the probability that the phone call:

i) Takes longer than 5 minutes.

ii) Takes between 5 and 6 minutes.

iii) Takes less than 3 minutes.

iv) Takes less than 6 minutes given that it took at least 3 minutes.

b) Let $C(t)$ be the amount (in pesos) that must be paid by the user for every call that lasts t minutes. Assume:

$$C(t) = \begin{cases} 500 & \text{if } 0 < t \leq 5 \\ 750 & \text{if } 5 < t \leq 10 \\ 100t & \text{if } t > 10 . \end{cases}$$

Calculate the mean cost of a call.

2.57 Let X be a nonnegative integer-valued random variable with distribution

$$\begin{aligned} P(X = 2n) &= \frac{1}{2} \cdot P(X = 2n-1) \\ &= \frac{2}{3} \cdot P(X = 2n+1) \\ \text{with } P(X = 0) &= \frac{2}{3} \cdot P(X = 1). \end{aligned}$$

Find the pgf of X.

2.58 Find the pgf, if it exists, for the random variable with pmf

a) $p(n) = \frac{1}{n!(e-1)}, \quad n = 1, 2, \ldots$

b) $p(n) = \frac{1}{n(n+1)}, \quad n = 1, 2, \ldots$.

2.59 Let X be a continuous random variable with *pdf*

$$f_X(x) = \begin{cases} \frac{e^{-x}x^3}{6} & \text{if } 0 < x < \infty \\ 0 & \text{otherwise} . \end{cases}$$

Find the moment generating function of X.

2.60 Find the characteristic function of X if X is a random variable with probability mass function given by

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and $0 < p < 1$.

2.61 Find the characteristic function of X if X is a random variable with *pdf* given by

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\lambda > 0$.

2.62 Let X be a random variable with characteristic function given by:

$$\varphi_X(t) = \frac{1}{7}\left(2 + e^{-it} + e^{it} + 3e^{2it}\right).$$

Determine:

a) $P\left(-1 \leq X \leq \frac{1}{2}\right)$.

b) $E(X)$.

2.63 Let X be a random variable. Show that

$$\overline{\varphi_X(t)} = \varphi_X(-t)$$

where $\overline{z}$ denote the complex conjugate of z.

2.64 Let X be a random variable. Show that $\varphi_X(t)$ is a real function if and only if X and $-X$ have the same distribution.

2.65 Let X be a random variable with characteristic function given by:

$$\varphi_X(t) = e^{2(e^{it}-1)}.$$

Determine $E(2X^2 - 5X + 1)$.

2.66 Let X be a random with characteristic function $\varphi_X(t)$. Find the characteristic function of $Y := 2X - 5$.

CHAPTER 3

SOME DISCRETE DISTRIBUTIONS

In this chapter we present some frequently used discrete distributions.

3.1 DISCRETE UNIFORM, BINOMIAL AND BERNOULLI DISTRIBUTIONS

Definition 3.1 (Discrete Uniform Distribution) *A random variable X has a discrete uniform distribution with N points, where N is a positive integer with possible distinct values x_i, $i = 1, 2, \cdots, N$, if its probability mass function is given by:*

$$p(x) = \begin{cases} \frac{1}{N} & \text{if } x = x_1, x_2, \cdots, x_N \\ 0 & \text{otherwise} . \end{cases}$$

If in particular $x_i = i$, $i = 1, 2, \cdots, N$, the probability mass function is shown in Figure 3.1.

Theorem 3.1 (Properties of a Discrete Uniform Random Variable) *If X is a random variable having a discrete uniform distribution with N points, then:*

Introduction to Probability and Stochastic Processes with Applications, First Edition.
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

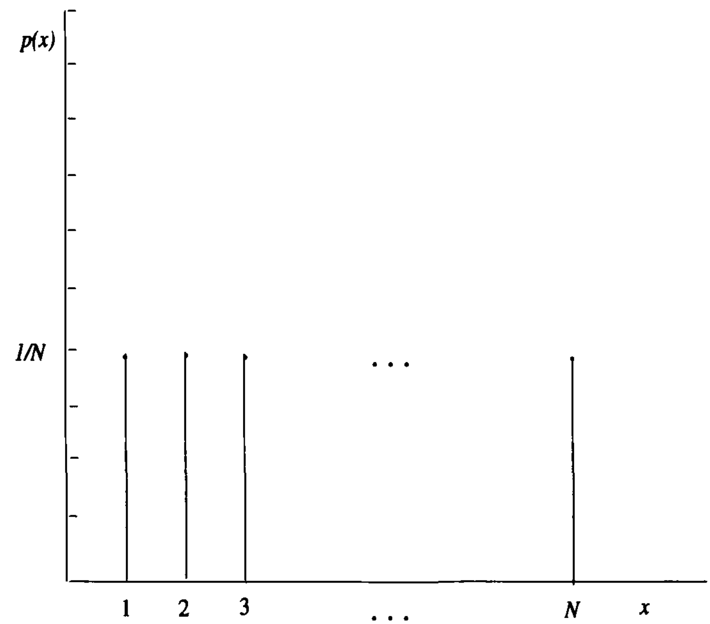


Figure 3.1 Probability mass function for a discrete uniform distribution

1. $E(X) = \frac{1}{N}\sum_{i=1}^{N} x_i.$

2. $E(X^r) = \frac{1}{N}\sum_{i=1}^{N} x_i^r.$

3. $m_X(t) = \sum_{i=1}^{N} \frac{1}{N} e^{tx_i}.$

Corollary 3.1 *If* $x_k = k, \quad k = 1, 2, \cdots, N,$ *then:*

1. $E(X) = \frac{N+1}{2}.$

2. $Var(X) = \frac{N^2-1}{12}.$

3. $m_X(t) = \sum_{k=1}^{N} \frac{1}{N} e^{kt}.$

Proof:

1.

$$E(X) = \sum_{k=1}^{N} \frac{k}{N} = \frac{1}{N} \times \frac{N(N+1)}{2} = \frac{N+1}{2}.$$

2. Left as an exercise for the reader.

3. Follows from the definition of the mgf.

■

■ EXAMPLE 3.1

A fair die is tossed. The sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ and each event occurs with probability $\frac{1}{6}$. Therefore we have a uniform distribution with:

$$\begin{aligned} p(n) &= \frac{1}{6}, \quad n = 1, 2, 3, 4, 5, 6 \\ E(X) &= \frac{1+2+3+4+5+6}{6} = 3.5 \\ Var(X) &= \frac{35}{12}. \quad \blacktriangle \end{aligned}$$

Definition 3.2 (Binomial and Bernoulli Distributions) *A random variable X is said to have a binomial distribution with parameters n and p if its probability mass function is given by*

$$p(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{if } x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

where n is a positive integer and $0 < p < 1$.

If $n = 1$, the binomial distribution is called a Bernoulli distribution with parameter p.

The probability mass function of a binomial distribution with parameters $n = 6$, $p = 0.5$ is shown in Figure 3.2.

Notation 3.1 *We shall write $X \stackrel{d}{=} \mathcal{B}(n, p)$ to indicate that the random variable X has a binomial distribution with parameters n and p.*

The binomial distribution is frequently used to describe experiments whose outcome can be regarded in terms of the occurrence or not of a certain phenomenon. If the random variable X denotes the number of successes achieved after carrying out n independent repetitions of the experiment, then X has a binomial distribution with parameters n and p, where p is the probability of success, that is, the probability that the desired phenomenon is observed; the probability of failure $1 - p$ is usually referred to with the letter q.

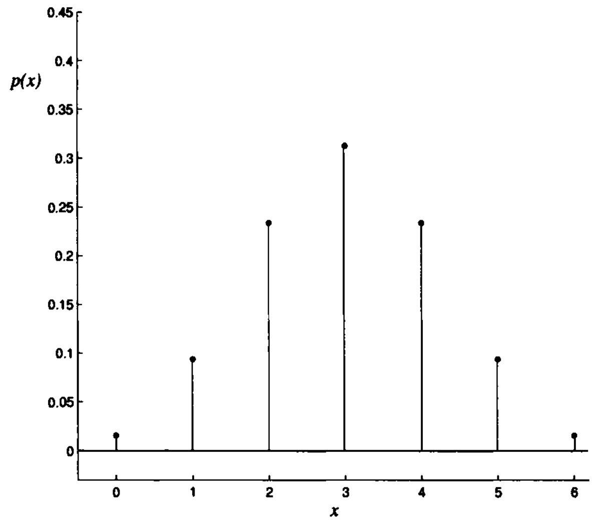


Figure 3.2 Probability mass function of a binomial distribution

■ EXAMPLE 3.2

A fair die is rolled five consecutive times. Let X be the random variable representing the number of times that the number 5 was obtained. Find the probability mass function of X.

Solution: The random variable X has a binomial distribution with parameters 5 and $\frac{1}{6}$. Therefore:

$$P(X=0) = \binom{5}{0} \times \left(\frac{1}{6}\right)^0 \times \left(\frac{5}{6}\right)^5 = 0.40188$$

$$P(X=1) = \binom{5}{1} \times \left(\frac{1}{6}\right)^1 \times \left(\frac{5}{6}\right)^4 = 0.40188$$

$$P(X=2) = \binom{5}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^3 = 0.16075$$

$$P(X=3) = \binom{5}{3} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^2 = 0.03215$$
$$P(X=4) = \binom{5}{4} \times \left(\frac{1}{6}\right)^4 \times \left(\frac{5}{6}\right)^1 = 3.215 \times 10^{-3}$$
$$P(X=5) = \binom{5}{5} \times \left(\frac{1}{6}\right)^5 \times \left(\frac{5}{6}\right)^0 = 1.286 \times 10^{-4}. \quad \blacktriangle$$

EXAMPLE 3.3

A salesman from a travel agency knows from experience that the opportunity to sell a travel package is greater when he makes more contact with potential buyers. He has established that the probability that a costumer will buy a package after a visit is constant and equal to 0.01. If the set of visits done by the salesman constitute an independent set of trials, how many potential buyers must he visit in order to guarantee that the probability of selling at least a travel package will equal 0.85?
Solution: Let:

$$X := \text{“Number of people that buy a travel package after the salesman's visit”}.$$

We have that $X \stackrel{d}{=} \mathcal{B}(n, 0.01)$ and wish to find n such that

$$P(X \geq 1) = 0.85$$

or equivalently:

$$\begin{aligned} 0.85 &= 1 - P(X=0) \\ &= 1 - \binom{n}{0}(0.01)^0(0.99)^n. \end{aligned}$$

Therefore:

$$\begin{aligned} (0.99)^n &= 0.15 \\ n\ln(0.99) &= \ln(0.15) \\ n &= \frac{\ln(0.15)}{\ln(0.99)} \approx 189. \end{aligned}$$

Consequently, the salesman must visit at least 189 people to achieve his goal. $\blacktriangle$

■ **EXAMPLE 3.4**

Paula's best friend invited her to a party. Since Paula is still very young, her parents conditioned their consent to Paula's brother going along with her. Paula's brother proposes the following deal to her: "You pick a number, whichever you like, from 1 to 6; you then roll a fair die four times, and if the number you picked appears at least twice, then I will come with you. Otherwise, I won't". What is the probability that Paula is going to the party?
Solution: Let:

$X :=$ "Number of times the number chosen by Paula appears".

Clearly $X \stackrel{d}{=} \mathcal{B}(4, \frac{1}{6})$. Then, the required probability equals:

$$\begin{aligned}
P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\
&= 1 - \binom{4}{0} \times \left(\frac{1}{6}\right)^0 \times \left(\frac{5}{6}\right)^4 - \binom{4}{1} \times \left(\frac{1}{6}\right)^1 \times \left(\frac{5}{6}\right)^3 \\
&= 0.13194. \quad \blacktriangle
\end{aligned}$$

■ **EXAMPLE 3.5**

A player bets on a number from 1 to 6. Once he has bet, three fair dice are rolled. If the number chosen by the player appears i times, with $i = 1, 2, 3$, then the player is rewarded with $2i$ money tokens. If the number bet does not appear on any die, the player loses 3 tokens. Is this game fair to the player? Explain your answer.
Solution: Let X be the random variable representing the fortune of the player. The values X can take are $-3, 2, 4$ and 6. We have that:

$$\begin{aligned}
P(X = -3) &= \binom{3}{0} \times \left(\frac{1}{6}\right)^0 \times \left(\frac{5}{6}\right)^3 = \frac{125}{216} \\
P(X = 2) &= \binom{3}{1} \times \left(\frac{1}{6}\right)^1 \times \left(\frac{5}{6}\right)^2 = \frac{75}{216} \\
P(X = 4) &= \binom{3}{2} \times \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^1 = \frac{15}{216} \\
P(X = 6) &= \binom{3}{3} \times \left(\frac{1}{6}\right)^3 \times \left(\frac{5}{6}\right)^0 = \frac{1}{216}.
\end{aligned}$$

Accordingly:

$$E(X) = \frac{(-3) \times 125 + 2 \times 75 + 4 \times 15 + 6}{216} = \frac{-159}{216}.$$

This indicates that in the long run the player loses 159 tokens for every 216 games played. Therefore, the game does not favor him. ▲

■ EXAMPLE 3.6

(Hoel et al., 1971) Suppose that n balls are randomly distributed in r urns. Find the probability that exactly k balls were put in the first r_1 urns.

Solution: Let $X :=$ "number of balls in the first r_1 urns". Since $X \stackrel{d}{=} \mathcal{B}(n,p)$ with $p = \frac{r_1}{r}$, then:

$$P(X=k) = \binom{n}{k}\left(\frac{r_1}{r}\right)^k\left(1-\frac{r_1}{r}\right)^{n-k}. \quad ▲$$

■ EXAMPLE 3.7

Consider the k-out-of-n structure which is a special case of parallel redundant system. This type of configuration requires that at least k components succeed out of the total n parallel components for the system to succeed. The problem is to compute the system reliability given the component reliabilities p_i. The reliability of a k-out-of-n structure of independent components, which all have the same reliability p, equals:

$$\text{Reliability} = \sum_{i=k}^{n}\binom{n}{i}p^i(1-p)^{n-i}.$$

This formula holds since the sum of n random variables has a Bernoulli distribution with parameters n and p under given assumptions. ▲

Next, we present some properties of the binomial distribution.

Theorem 3.2 (Properties of the Binomial Distribution) *Let X be a random variable having a binomial distribution with parameters n and p. Then:*

1. $E(X) = np$.
2. $Var(X) = npq$, *where* $q := 1-p$.
3. $m_X(t) = (pe^t + q)^n$.

Proof: In Example 2.54 we verified that the mgf of a random variable with binomial distribution of parameters n and p is given by $m_X(t) = (pe^t+q)^n$. Thus:

$$E(X) = \left.\frac{d}{dt}m_X(t)\right|_{t=0} = npe^t(pe^t+q)^{n-1}\,|_{t=0} = np.$$

Furthermore, we have:

$$\begin{aligned}E(X^2) &= \left.\frac{d^2}{dt^2}m_X(t)\right|_{t=0}\\ &= \left.[n(n-1)(pe^t)^2(pe^t+q)^{n-2} + npe^t(pe^t+q)^{n-1}]\right|_{t=0}\\ &= n(n-1)p^2 + np.\end{aligned}$$

Hence, we obtain:

$$\begin{aligned}Var(X) &= n^2p^2 - np^2 + np - n^2p^2\\ &= np(1-p)\\ &= npq.\end{aligned}$$

■

Note 3.1 *Suppose that X is a random variable with a binomial distribution of parameters n and p. Let:*

$$B(k) := \binom{n}{k}p^kq^{n-k} .$$

Seeing that

$$\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$$

then for $k = 1, \cdots, n$:

$$\begin{aligned}B(k) &= \frac{n-k+1}{k}\binom{n}{k-1}p \times p^{k-1} \times q^{n-k+1} \times \frac{1}{q}\\ &= \frac{n-k+1}{k} \times \frac{p}{q} \times B(k-1).\end{aligned}$$

Hence, starting with $B(0) = q^n$ the values of $B(k)$ for $k = 1, \cdots, n$ can be recursively obtained.

Algorithm 3.1

Input: p, n, where n is the number of terms.

Output: $B(k)$ for $k = 0(1)n$.

Initialization: $q := (1-p)$,

$B(0) := q^n$.

Iteration: For $k = 0(1)n-1$ do:

$B(k+1) = \frac{n-k}{k+1} \times \frac{p}{q} \times B(k)$ ▲

Using the last algorithm we obtain, for example, for $n = 5$ and $p = 0.3$:

k	$B(k)$
0	0.16807
1	0.36015
2	0.30870
3	0.13230
4	0.02835
5	0.00243

Note 3.2 *From the previous remark, we have that:*

$$B(k) > B(k-1) \text{ if and only if } \frac{n-k+1}{k} \times \frac{p}{q} > 1 \; .$$

That is:

$$B(k) > B(k-1) \text{ if and only if } (n+1)p > k.$$

Thus, if $X \stackrel{d}{=} \mathcal{B}(n,p)$, then $P(X = k)$ increases monotonically until it reaches a maximum for $k = \lfloor (n+1)p \rfloor$ (where $\lfloor a \rfloor$ denotes the integer part or floor of a) and, after this value, it decreases monotonically.

3.2 HYPERGEOMETRIC AND POISSON DISTRIBUTIONS

In Chapter 1, we saw that if an urn contains N balls, R of which are red and $N - R$ white, and a sample of size n is extracted without substitution, then the probability $P(A_k)$ that exactly k of the extracted balls are red equals:

$$P(k) := P(A_k) = \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}}.$$

If we are only interested in the number k of red balls among the n balls extracted, then $P(k) := P(A_k)$ defines a probability measure over the set $\{0, 1, \cdots, n\}$ called the hypergeometric distribution with parameters n, R and N. More precisely, we have:

Definition 3.3 (Hypergeometric Distribution) *A random variable X is said to have a hypergeometric distribution with parameters n, R and N if its probability mass function is given by*

$$p_X(x) = \begin{cases} \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & \text{if } x = 0, 1, \cdots, n \\ 0 & \text{otherwise} \end{cases}$$

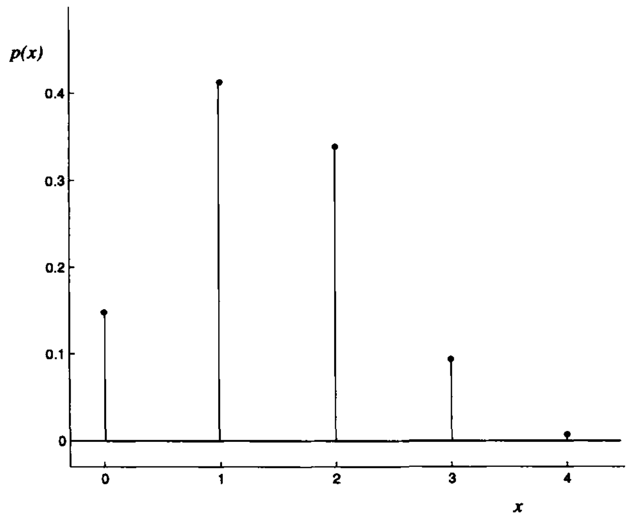


Figure 3.3 Probability mass function of a hypergeometric distribution

where N is a positive integer, R is a nonnegative integer less than or equal to N and n is a positive integer less than or equal to N.

The probability mass function of a hypergeometric distribution with parameters $n = 4$, $R = 7$ and $N = 20$ is shown in Figure 3.3.

Notation 3.2 *The expression $X \stackrel{d}{=} Hg(n, R, N)$ means that the random variable X has a hypergeometric distribution with parameters n, R and N.*

■ EXAMPLE 3.8

The surveillance division of a university has acquired 50 communication devices in order to optimize the security in the campus. Eight of them are randomly selected and tested to identify any possible flaws. If 3 of the 50 devices are defective, what is the probability that the sample contains at most two defective devices?

Solution: Let X :="number of defective devices found in the sample". It follows that $X \stackrel{d}{=} Hg(8, 3, 50)$. Therefore:

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \frac{\binom{3}{0}\binom{47}{8}}{\binom{50}{8}} + \frac{\binom{3}{1}\binom{47}{7}}{\binom{50}{8}} + \frac{\binom{3}{2}\binom{47}{6}}{\binom{50}{8}} \\ &= 0.99714. \quad \blacktriangle \end{aligned}$$

EXAMPLE 3.9

A work team established by the Department of the Environment programmed inspection of 25 factories in order to investigate possible violations to the environmental contamination act. Nevertheless, cuts in budget have substantially reduced the work team size, so that only 5 out of 25 factories are going to be inspected. If it is known that 10 of the factories are working outside the environmental contamination act, find the probability that at least one of the factories sampled is working against the regulations.

Solution: Let $X :=$ "number of factories operating in contravention of the environmental contamination act". It is straightforward that $X \stackrel{d}{=} Hg(5, 10, 25)$, whereupon we have:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - \frac{\binom{10}{0} \times \binom{15}{5}}{\binom{25}{5}} \\ &= 0.94348. \quad \blacktriangle \end{aligned}$$

Note 3.3 *Suppose that the size N of the population is unknown, but we wish to find it without counting one by one each individual. A way of doing this is the so-called capture-recapture method, which consists in capturing R individuals from the population, tagging them and then returning them to the population. Once the tagged and untagged individuals have mixed, a sample of size n is taken. Let X be the random variable representing the number of tagged individuals in the sample. Clearly $X \stackrel{d}{=} Hg(n, R, N)$. Suppose that the observed value of X is k. Then $P_k(N) := P(X = k)$ represents the probability that the sample contains k tagged individuals when the population has size N. So we will estimate N as the value $\widehat{N}$ for which the probability that X equals k is maximum. Such a value is known as the maximum likelihood estimate of N. To find it, we observe that:*

$$P_k(N) \geq P_k(N-1) \text{ if and only if } \frac{(N-R) \times (N-n)}{N \times (N-R-n+k)} \geq 1 .$$

That is,

$$P_k(N) \geq P_k(N-1) \text{ if and only if } \frac{R \times n}{k} \geq N$$

and thereby:

$$\widehat{N} = \left[\frac{R \times n}{k}\right].$$

■ EXAMPLE 3.10

To establish the number of fishes in a lake the following procedure is followed: 2000 fishes are captured, tagged and then returned to the lake. Days after, 350 fishes are captured and it is observed that 50 of them are tagged. Then, according to the previous remark, the maximum likelihood estimate of the size N of the population is:

$$\widehat{N} = \left[\frac{2000 \times 350}{50}\right] = 12,727. \quad \blacktriangle$$

Below we present some of the properties of the hypergeometric distribution.

Theorem 3.3 (Properties of the Hypergeometric Distribution) *Let $X \stackrel{d}{=} Hg(n, R, N)$. Then:*

1. $E(X) = \frac{nR}{N}$.

2. $Var(X) = n \times \frac{R}{N} \times \frac{N-R}{N} \times \frac{N-n}{N-1}$.

Proof:

1.

$$\begin{aligned} E(X) &= \sum_{x=0}^{n} x \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} \\ &= \sum_{x=1}^{n} n \frac{R}{N} \frac{\binom{R-1}{x-1}\binom{N-R}{n-x}}{\binom{N-1}{n-1}} \\ &= n \frac{R}{N} \sum_{k=0}^{n-1} \frac{\binom{R-1}{k}\binom{N-R}{n-k-1}}{\binom{N-1}{n-1}} \\ &= n \frac{R}{N}, \end{aligned}$$

where we used the following identity:

$$\sum_{k=0}^{n-1} \binom{R-1}{k}\binom{N-R}{n-k-1} = \binom{N-1}{n-1}.$$

2. Seeing that

$$Var(X) = E(X(X-1)) + E(X) - (E(X))^2$$

it suffices to find $E(X(X-1))$. Thus:

$$
\begin{aligned}
E(X(X-1)) &= \sum_{x=0}^{n} x(x-1)P(X=x) \\
&= \sum_{x=2}^{n} \left[x(x-1)\frac{R(R-1)(R-2)!}{(R-x)!x(x-1)(x-2)!} \times \right. \\
&\qquad \left. \frac{(N-n)!n(n-1)(n-2)!}{N(N-1)(N-2)!} \times \binom{N-R}{n-x} \right] \\
&= n(n-1)\frac{R(R-1)}{N(N-1)} \sum_{x=2}^{n} \left[\frac{\binom{R-2}{x-2}\binom{N-R}{n-x}}{\binom{N-2}{n-2}} \right] \\
&= n(n-1)\frac{R(R-1)}{N(N-1)} \sum_{k=0}^{n-2} \left[\frac{\binom{R-2}{k}\binom{N-R}{n-k-2}}{\binom{N-2}{n-2}} \right] \\
&= n(n-1)\frac{R(R-1)}{N(N-1)}
\end{aligned}
$$

Using the relation

$$
\sum_{k=0}^{n-2} \binom{R-2}{k}\binom{N-R}{n-k-2} = \binom{N-2}{n-2}.
$$

We obtain

$$
\begin{aligned}
Var(X) &= n(n-1)\frac{R(R-1)}{N(N-1)} + \frac{nR}{N} - \frac{n^2R^2}{N^2} \\
&= n\frac{R}{N}\left[\frac{(N-R)(N-n)}{N(N-1)}\right].
\end{aligned}
$$

■

Next we will see that if the population size N is large enough in comparison to the sample size n, then the hypergeometric distribution can be approximated by a binomial distribution. More precisely, we have the following result:

Theorem 3.4 *Let $0 < p < 1$. If $N, R \longrightarrow \infty$ in such a way that $\frac{R}{N} \longrightarrow p$, then:*

$$
Hg(n,R,N)(k) := \frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \longrightarrow \mathcal{B}(n,p)(k) := \binom{n}{k}p^k(1-p)^{n-k}.
$$

Proof: Observing that $\frac{N-R}{N} \longrightarrow (1-p) = q > 0$ when $N, R \longrightarrow \infty$, we see that $(N-R) \longrightarrow \infty$ when $N \longrightarrow \infty$. Therefore:

$$
\begin{aligned}
&\frac{\binom{R}{k}\binom{N-R}{n-k}}{\binom{N}{n}} \\
&= \binom{n}{k} \frac{R(R-1)\cdots(R-k+1)(N-R)\cdots(N-R-(n-k)+1)}{N(N-1)\cdots(N-n+1)} \\
&= \binom{n}{k} \left(\frac{R}{N}\right)^k \left(\frac{N-R}{N}\right)^{n-k} \frac{R(R-1)\cdots(R-k+1)}{R^k} \\
&\quad \times \frac{(N-R)\cdots(N-R-(n-k)+1)}{(N-R)^{n-k}} \times \frac{N^n}{N(N-1)\cdots(N-n+1)} \\
&\xrightarrow[N,R\longrightarrow\infty]{} \binom{n}{k} p^k (1-p)^{n-k}.
\end{aligned}
$$

■

■ **EXAMPLE 3.11**

In a city with 2 million inhabitants, 60% belong to political party A. One hundred people are randomly chosen from the inhabitants. The distribution of the number of people, among the 100 drawn who belong to party A is hypergeometric with parameters 100, 1200000 and 2000000. Applying the previous result we can approximate this distribution by a binomial with parameters $n = 100$ and $p = 0.6$. That way, for example, the probability that among the 100 people chosen exactly 40 belong to party A equals:

$$\binom{100}{40} (0.6)^{40} (0.4)^{60} = 2.4425 \times 10^{-5}. \quad \blacktriangle$$

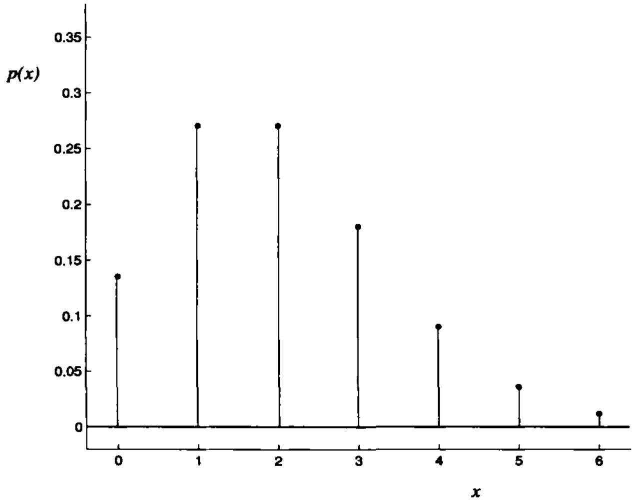


Figure 3.4 Probability mass function of a Poisson distribution

The following table compares the binomial and hypergeometric distributions:

k	$Hg(4,60,100)$	$Hg(4,600,1000)$	$Hg(4,6000,10,000)$	$\mathcal{B}\left(4,\frac{3}{5}\right)$
0	0.02331	0.02537	0.02558	0.0256
1	0.15118	0.15337	0.15358	0.1536
2	0.35208	0.34624	0.34566	0.3456
3	0.34907	0.34595	0.34563	0.3456
4	0.12436	0.12908	0.12955	0.1296

Definition 3.4 (Poisson Distribution) *A random variable X is said to have a Poisson distribution with parameter $\lambda > 0$ if its probability mass function is given by:*

$$p(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!} & \textit{if} \quad x = 0, 1, \cdots \\ 0 & \textit{otherwise} \end{cases} .$$

The probability mass function of a Poisson distribution with parameter $\lambda = 2$ is shown in Figure 3.4.

Notation 3.3 *Let X be a random variable. We write $X \stackrel{d}{=} \mathcal{P}(\lambda)$ to indicate that X has a Poisson distribution with parameter λ.*

Note 3.4 *Let* $X \stackrel{d}{=} \mathcal{P}(\lambda)$ *and* $\mathcal{P}(\lambda)(k) := P(X = k)$. *It is straightforward that, for* $k = 1, 2, \cdots$:

$$\mathcal{P}(\lambda)(k) = \frac{\lambda}{k}\mathcal{P}(\lambda)(k-1) .$$

Hence, starting with $\mathcal{P}(\lambda)(0) = e^{-\lambda}$ *the values of* $\mathcal{P}(\lambda)(k)$ *can be recursively obtained for* $k = 1, 2, \cdots$.

A simple algorithm to calculate the values of the Poisson distribution making use of the previous expression is given by:
Algorithm 3.2[Calculus of $\mathcal{P}(\lambda)(k)$]

Input: λ, n, with n being the number of terms.

Output: $\mathcal{P}(\lambda)(k)$ for $k = 0(1)n$.

Initialization: $\mathcal{P}(\lambda)(0) = e^{-\lambda}$.

Iteration: For $k = 0(1)n - 1$ do:
$\mathcal{P}(\lambda)(k+1) = \frac{\lambda}{k+1}\mathcal{P}(\lambda)(k)$. ▲

Analyzing the graph of the probability mass function from a random variable with Poisson distribution, we observe that the individual probabilities become smaller as the variable takes larger values. This is precisely one of the general characteristics of the Poisson distribution. Other properties of the Poisson distribution are given in the following theorem.

Theorem 3.5 *Let* X *be a random variable with Poisson distribution of parameter* λ. *Then:*

1. $E(X) = \lambda$.

2. $Var(X) = \lambda$.

3. $m_X(t) = \exp(\lambda(e^t - 1))$.

Proof: We start by proving 3 and by applying the properties of the mgf we deduce 1 and 2. Accordingly:

$$\begin{aligned} m_X(t) &= \sum_{k=0}^{\infty} e^{tk} e^{-\lambda} \frac{\lambda^k}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= \exp\left(\lambda(e^t - 1)\right) . \end{aligned}$$

Thus

$$\begin{aligned} E(X) &= \frac{d}{dt} m_X(t) \mid_{t=0} \\ &= \lambda e^t \exp\left(\lambda(e^t - 1)\right) \mid_{t=0} \\ &= \lambda \,. \end{aligned}$$

Also

$$\begin{aligned} E(X^2) &= \frac{d^2}{dt^2} m_X(t) \mid_{t=0} \\ &= \left[\lambda e^t \exp\left(\lambda(e^t - 1)\right) + \lambda^2 e^{2t} \exp\left(\lambda(e^t - 1)\right)\right] \mid_{t=0} \\ &= \lambda(\lambda + 1), \end{aligned}$$

which yields

$$Var(X) = \lambda.$$

The property $E(X) = Var(X)$ holds only for Poisson distribution and is a characterizing property of the distribution. ■

■ EXAMPLE 3.12

The number of patients who come daily to the emergency room (E.R.) of a certain hospital has a Poisson distribution with mean 10. What is the probability that, during a normal day, the number of patients admitted in the emergency room of the hospital will be less than or equal to 3?

Solution: Let X :="number of patients who come daily to the E.R.". From the data supplied by the statement of the problem, we know that $X \stackrel{d}{=} \mathcal{P}(10)$. Therefore:

$$\begin{aligned} P(X \leq 3) &= P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) \\ &= e^{-10} + 10e^{-10} + 50e^{-10} + \frac{1000}{6} e^{-10} \\ &= 1.0336 \times 10^{-2}. \quad \blacktriangle \end{aligned}$$

It will be proved below that for large enough n and suitably small p a binomial distribution with parameters n and p can be approximated by a Poisson distribution of parameter $\lambda = np$; that is, under these circumstances, the Poisson distribution is a limit of the binomial distribution.

Theorem 3.6 *If $p(n)$ is a sequence satisfying $0 < p(n) < 1$ and $n \times (p(n)) \xrightarrow[n\to\infty]{} \lambda$, then:*

$$\mathcal{B}_{n,p(n)}(k) := \binom{n}{k} (p(n))^k (1 - p(n))^{n-k} \xrightarrow[n\to\infty]{} e^{-\lambda} \frac{\lambda^k}{k!} =: \mathcal{P}(\lambda)(k).$$

Proof: Let $\lambda_n = n \times (p(n))$. Then:

$$\mathcal{B}_{n,p(n)}(k) = \frac{1}{k!} \times \frac{n}{n} \times \cdots \times \frac{n-k+1}{n} \times \lambda_n^k \times \left(1 - \frac{\lambda_n}{n}\right)^n \times \left(1 - \frac{\lambda_n}{n}\right)^{-k}.$$

Taking the limit when $n \to \infty$ we see that the quotients $\frac{n}{n}, \cdots, \frac{n-k+1}{n}$ as well as the factor $\left(1 - \frac{\lambda_n}{n}\right)^{-k}$ tend to 1, while the expression $\left(1 - \frac{\lambda_n}{n}\right)^n$ tends to $e^{-\lambda}$. As a result:

$$\mathcal{B}_{n,p(n)}(k) \xrightarrow[n\to\infty]{} \mathcal{P}(\lambda)(k).$$

■

The following table shows how good the approximation is when $\lambda = np = 1$:

k	$\mathcal{P}(\lambda)(k)$	$\mathcal{B}_{100,\frac{1}{100}}(k)$	$\mathcal{B}_{10,\frac{1}{10}}(k)$
0	0.3678	0.3660	0.3487
1	0.3679	0.3697	0.3874
2	0.1839	0.1848	0.1937
3	0.0613	0.0610	0.0574
4	0.0153	0.0149	0.0112

Assume $p(n) = p$, i.e., the probability that the success in any trial is p and all the trials are independent for finite n. For large n, the sum of Bernoulli random variables with parameters n and p tends to Poisson random variable with parameter $\lambda = np$.

The preceding theorem implies that the Poisson distribution offers a good probabilistic model for those random experiments where there are independent repetitions and only two possible outcomes, success or failure, with small probability of success, and where the main interest lies on knowing the number of successes obtained after a large enough number of repetitions of the experiment. Empirically, it has been determined that the approximation can be safely used if $n \geq 100$, $p \leq 0.01$ and $np \leq 20$.

■ EXAMPLE 3.13

There are 135 students inside a conference hall. The probability that one of the students celebrates his or her birthday today equals $\frac{1}{365}$. What is the probability that two or more students from the conference hall are celebrating their birthdays today?

Solution: Let X := "number of students celebrating their birthdays today". It is known that $X \stackrel{d}{=} \mathcal{B}(135, \frac{1}{365})$; however, this distribution can be approximated by means of a Poisson distribution with parameter

$\lambda = \frac{27}{73}$. Thereby:

$$\begin{aligned} P(X \geq 2) &= 1 - P(X = 0) - P(X = 1) \\ &= 1 - e^{-\frac{27}{73}} - \frac{27}{73} e^{-\frac{27}{73}} \\ &= 5.3659 \times 10^{-2}. \quad \blacktriangle \end{aligned}$$

3.3 GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

When working with a binomial distribution we proceeded as follows: a random experiment was repeated n times and the probability of obtaining exactly k successes was found. In this case, the number of repetitions remains constant while the number of successes is random. Instead suppose the question is the following: What is the probability that the experiment has to be repeated n times to obtain exactly k successes? That is, now the number of successes remains constant while the number of repetitions is a random variable X.

What is the probability that X takes the values $j = k, k+1, \cdots$? If $X = j$, then the jth result was necessarily a success and therefore the other $k - 1$ successes are obtained in the remaining $j - 1$ repetitions of the experiment. That is,

$$P(X = j) = \binom{j-1}{k-1} p^k (1-p)^{j-k}, \quad j = k, k+1, \cdots,$$

where $0 < p < 1$ is the success probability.

Definition 3.5 (Geometric and Negative Binomial Distributions) *A random variable X is said to have a negative binomial distribution with parameters k and p if its probability mass function is given by:*

$$p(x) = \begin{cases} \binom{x-1}{k-1} p^k (1-p)^{x-k} & \textit{if} \quad x = k, k+1, \cdots \\ 0 & \textit{otherwise} . \end{cases}$$

In the special case where $k = 1$ it is said that the random variable has a geometric distribution with parameter p.

The probability mass function of a negative binomial distribution with parameters $k = 2$ and $p = \frac{1}{2}$ is shown in Figure 3.5.

Notation 3.4 *The expressions $X \stackrel{d}{=} \mathcal{B}_N(k,p)$ and $X \stackrel{d}{=} \mathcal{G}(p)$ mean that X has a negative binomial distribution with parameters k and p and X has a geometric distribution of parameter p.*

Note 3.5 *Suppose that we are interested not in the number of repetitions required to obtain k successes but in the number of failures Y that happened*

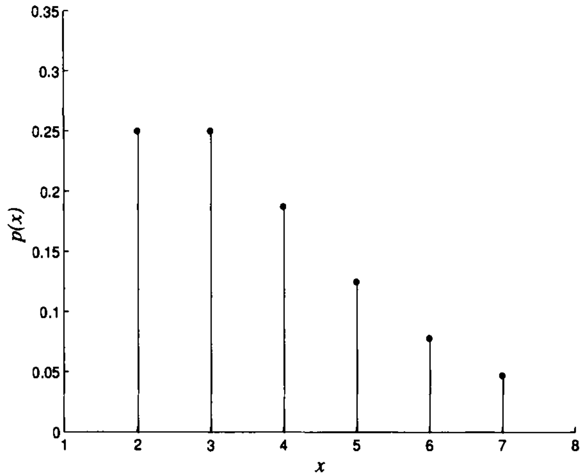


Figure 3.5 Probability mass function of a negative binomial distribution

before we obtained exactly k successes. In this case we have that $X = k + Y$ *and therefore:*

$$P(Y = j) = \binom{k+j-1}{k-1} p^k (1-p)^j \quad \text{with } j = 0, 1, 2, \cdots .$$

Some authors call this last distribution the negative binomial and the one defined before as the Pascal distribution.

■ EXAMPLE 3.14

In a quality control department, units coming from an assembly line are inspected. If the proportion of defective units is 0.03, what is the probability that the twentieth unit inspected is the third one found defective?

Solution: Let $X :=$ "number of units necessary to inspect in order to obtain exactly three defective ones". Clearly $X \stackrel{d}{=} \mathcal{B}_N(3, 0.03)$. Hence:

$$\begin{aligned} P(X = 20) &= \binom{19}{2} (0.03)^3 (1 - 0.03)^{17} \\ &= 2.7509 \times 10^{-3}. \quad \blacktriangle \end{aligned}$$

Below, we find the mean and variance of a random variable having a negative binomial distribution.

Theorem 3.7 *Let X be a random variable having a negative binomial distribution with parameters k and p. Then:*

1. $E(X) = \frac{k}{p}$.

2. $Var(X) = \frac{k(1-p)}{p^2}$.

3. $m_X(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]^k$.

Proof: The rth moment of X around the origin is given by

$$
\begin{aligned}
E(X^r) &= \sum_{j=k}^{\infty} j^r \binom{j-1}{k-1} p^k (1-p)^{j-k} \\
&= \frac{k}{p} \sum_{j=k}^{\infty} j^{r-1} \binom{j}{k} p^{k+1} (1-p)^{j-k} \\
&= \frac{k}{p} \sum_{n=k+1}^{\infty} (n-1)^{r-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} \\
&= \frac{k}{p} E\left((Y-1)^{r-1}\right),
\end{aligned}
$$

where $Y \stackrel{d}{=} \mathcal{B}_N(k+1, p)$. Therefore

$$E(X) = \frac{k}{p}$$

and

$$
\begin{aligned}
E(X^2) &= \frac{k}{p} E(Y-1) \\
&= \frac{k}{p} \times \left[\frac{k+1}{p} - 1\right] \\
&= \frac{k}{p} \times \frac{k+1-p}{p},
\end{aligned}
$$

which yields:

$$
\begin{aligned}
Var(X) &= \frac{k^2+k-kp}{p^2} - \frac{k^2}{p^2} \\
&= \frac{k(1-p)}{p^2}.
\end{aligned}
$$

On the other hand,

$$\begin{aligned} m_X(t) &= \sum_{j=k}^{\infty} e^{tj} \binom{j-1}{k-1} p^k (1-p)^{j-k} \\ &= e^{tk} \sum_{l=0}^{\infty} e^{tl} \binom{l+k-1}{k-1} p^k (1-p)^l \\ &= e^{tk} \sum_{l=0}^{\infty} (-1)^l \binom{-k}{l} p^k \left((1-p)\, e^t\right)^l , \end{aligned}$$

where:

$$\begin{aligned} \binom{-k}{l} &:= (-1)^l \frac{k(k+1)\cdots(k+l-1)}{l!} \\ &= (-1)^l \binom{l+k-1}{k-1} . \end{aligned}$$

Using the Taylor series expansion of the function $g(x) := (1-x)^{-k}$ about the origin, we obtain:

$$(1-x)^{-k} = \sum_{l=0}^{\infty} \binom{-k}{l} (-x)^l.$$

Accordingly:

$$\begin{aligned} m_X(t) &= \left(pe^t\right)^k \left(1 - (1-p)\, e^t\right)^{-k} \\ &= \left[\frac{pe^t}{1-(1-p)e^t}\right]^k . \end{aligned}$$

■

Corollary 3.2 *If X is a random variable having a geometric distribution with parameter p, then:*

1. $E(X) = \frac{1}{p}$.
2. $Var(X) = \frac{1-p}{p^2}$.
3. $m_X(t) = \left[\frac{pe^t}{1-(1-p)e^t}\right]$.

Note 3.6 *For the random variable $Y = X - k$ we have:*

$$\begin{aligned} E(Y) &= \frac{k}{p} - k = \frac{k(1-p)}{p} \\ Var(Y) &= Var(X) = \frac{k(1-p)}{p^2} \\ m_Y(t) &= \left[\frac{p}{1-(1-p)e^t}\right]^k . \end{aligned}$$

Note 3.7 *Suppose that the size N of a population is unknown. We wish to determine N without counting each of the individuals. One way of doing this is the so-called inverse capture-recapture method, which consists in capturing R individuals from the population and returning them to it after they have been properly tagged. Once the tagged and nontagged individuals are homogeneously mixed, a new sample is taken from the population drawing one individual at a time until a predetermined number of tagged individuals k is reached. Let X be the random variable representing the number of necessary extractions to obtain k tagged individuals. Clearly $X \stackrel{d}{=} \mathcal{B}_N(k,p)$, where $p = \frac{R}{N}$. Suppose that the observed value X equals j. Then $P_k(N) := P(X = j)$ represents the probability that j extractions were required in order to obtain exactly k tagged individuals when the population size is N. Therefore, we can take as an estimator $\widehat{N}$ of N the value for which the probability that X equals j is maximum. Such an estimator is called a maximum likelihood estimate of N and equals:*

$$\widehat{N} = \left[\frac{R \times j}{k}\right].$$

■ EXAMPLE 3.15

To establish the number of fishes in a lake, it is proceeded as follows: 1000 fishes are captured and tagged, then returned to the lake. Days after, fishes are captured until 15 of them are tagged. If it was necessary to make 120 extractions to obtain the desired number of marked fishes, then according to the previous remark, the maximum likelihood estimate of the population size is:

$$\widehat{N} = \left[\frac{1000 \times 120}{15}\right] = 8000. \quad \blacktriangle$$

Note 3.8 *Let X be a random variable having a negative binomial distribution with parameters k and p, and let Y be a random variable having a binomial distribution with parameters n and p. Then:*

$$\begin{aligned} P(X > n) &= \sum_{j=n+1}^{\infty} P(X = j) \\ &= \sum_{j=n+1}^{\infty} \binom{j-1}{k-1} p^k (1-p)^{j-k} \\ &= \sum_{j=0}^{k-1} \binom{n}{j} p^j (1-p)^{n-j} \\ &= P(Y < k). \end{aligned}$$

EXERCISES

3.1 A fair coin is flipped an even number of times. What is the probability that half the times the result obtained is a "head"?

3.2 In a national automobile club a telephonic campaign is started in order to increase the number of its members. From previous experience, it is known that one out of every 10 people called actually join the club. If in a day 20 people are called, what is the probability that at least two of them join the club? What is the expected number?

3.3 A fair die is rolled 10 consecutive times. Find the probability that the number 5 is obtained at least 3 times.

3.4 In the computer room of an educational center there are 20 computers. The probability that any given computer is being used in peak hours is 0.9. What is the probability that at least one computer is available in peak hours? What is the probability that all the computers are being used?

3.5 An airline knows that 5% of the people making reservations on a certain flight will not show up. Consequently, their policy is to sell 52 tickets for a flight that can hold only 50 passengers. Assume that passengers who come to the airport are independent with each other. What is the probability that there will be a seat available for every passenger who shows up?

3.6 Find the probability distribution of a binomial random variable X with parameters n and p truncated to the right at $X = r, r > 0$.

3.7 Let X be a random variable having a binomial distribution with parameters n and p. Suppose that $E(X) = 12$ and $Var(X) = 4$. Find n and p.

3.8 A coin is flipped as many times as needed to obtain a head for the first time.

a) Describe the sample space of this experiment.

b) Let X be the random variable representing the required number of tosses. Find $P(X > 1)$ and $P(X \leq n)$.

3.9 Assume that the number of clients X arriving in an hour at a bank is a Poisson random variable and $P(X = 0) = 0.02$. Find the mean and variance of X.

3.10 Let X be a random variable having a Poisson distribution with parameter λ. If $P(X = 0) = 0.4$, find $P(X \leq 3)$.

3.11 Let X be a Poisson random variable with parameter λ. Assume $P(X = 2) = 2P(X = 1)$. Find $P(X = 0)$.

3.12 The quality control department of a factory inspects the units finished by the assembly line. It is believed that the proportion of defective units equals 0.01. What is the probability that the fifth unit inspected is the second defective one?

3.13 Casinos use roulettes with 38 pockets, of which 18 are black, 18 are red and 2 are green. Let X be the random variable representing the number of times it is necessary to spin the wheel to obtain a red number for the first time. Find the probability mass function of X.

3.14 A number from the interval $(0, 1)$ is randomly chosen. What is the probability that:

a) The first decimal digit is 1.

b) The second decimal digit is 5.

c) The first decimal digit of its square root is 3.

3.15 A multiple-choice test contains 30 questions, each with four possible options. Suppose that a student only guesses the answers:

a) What is the probability that the student answers right more than 20 questions?

b) What is the probability that the student answers less than 3 questions right?

c) What is the probability that the student answers all the questions wrong?

3.16 Recent studies determined that the probability of dying due to a certain flu vaccine is 0.00001. If 200,000 people are vaccinated, and we assume that each person can be regarded as an independent trial, what is the probability that no one dies because of the vaccine?

3.17 A professor wishes to set up the final examination with multiple-choice test. The examination must have 15 questions, and on account of internal regulations, a student approves if he or she answers at least 10 questions right. To minimize the "risk" of a student approving the examination by guessing the answers, the teacher wishes to put k possible choices to each question with only one being correct. What value must k have so that a student randomly answering the test has a probability of passing equal to 0.001?

3.18 Fifteen daltonic people are required for a medical experiment. If it is known that only 0.1% of the population has this condition, what is the expected number of interviews that must be carried out to find the 15 people required?

3.19 There are N fishes in a lake from which R $(\leq N)$ are tagged. Assume that n $(\leq R)$ fishes are caught one by one without replacement and let $T_i :=$ "the ith fish captured is tagged" for $i = 1, 2, \cdots, n$. What is $p(T_i)$ equal to? Are T_1 and T_3 independent?

3.20 At a certain governmental institution, the probability that a call is answered in less than 30 seconds is 0.25. Assume the calls to be independent.

a) If someone calls 10 times, what is the probability that 9 calls are answered in less than 30 seconds?

b) If someone calls 20 times, what is the average number of calls that are going to be answered in less than 30 seconds?

c) What is the average number of calls that have to be made in order to get an answer in less than 30 seconds?

d) What is the probability that 6 calls have to be made to get 2 of them answered in less than 30 seconds?

3.21 Suppose that 5% of the articles produced in a factory are defective. If 15 articles are randomly chosen and inspected, what is the probability that there are at most 3 defective articles in the sample?

3.22 Suppose that X is a random variable having a geometric distribution with parameter p. Find the probability mass function of X^2 and $X + 3$.

3.23 Three copies of the book "Get rich in less than 24 hours" in a library are lent to users for a day. The number of daily requests for a copy is a Poisson random variable with mean 5. Find:

a) The proportion of days wherein the demand of a copy of the book is zero.

b) The proportion of the days wherein the demand of a copy of the book surpasses the supply.

3.24 A certain system of a spacecraft must work correctly in order for the ship to make a safe entry to the atmosphere. A component of such a system operates without complications only 85% of the time. With the purpose of increasing the reliability of the system, four of these components are installed in such a way that the system will function properly if at least one of the components operates without complications.

a) What is the probability that the system will malfunction? Assume that the components operate independently from each other.

b) If the system fails, what can be inferred from the alleged 85% probability of success of a single component?

3.25 The records of an insurance company show that only 0.1% of the population suffers from a certain type of accident each year. If 10,000 people are randomly selected to be insured, what is the probability that no more than 5 of those clients have that type of accident next year?

3.26 There are 20 spectacled bears in a forest, 5 of which are captured, tagged and then released. Weeks later 4 of the 20 bears are captured again. Find the probability that at most 2 of the captured bears are tagged.

3.27 A company analyzes the shipments from its suppliers to detect products that do not comply with the minimum quality specifications required. It is known that 3% of such products do not meet the quality standards of the company. What size must a sample have in order to have at least a 0.90 probability that at least one article selected does not comply with the quality requirements? Suppose that in this case the hypergeometric distribution can be approximated by a binomial distribution.

3.28 A 300-page book has 253 typographical errors. Assuming that each page has 5000 characters, what is the probability that there are no typos on the first page? What is the probability that there is at least one typo on the first page?

3.29 A factory makes ink cartridges for stylographic pens. One out of every 30 cartridges made by the factory turns out to be defective. The cartridges are packed in six-unit boxes. Find the expected number of boxes that contain, respectively, no defective cartridges, 1 defective cartridge, 2 or more defective cartridges, in a shipment of 1000 boxes.

3.30 The police suspect that in a truck loaded with 40 rice bundles there might be cocaine packages camouflaged. To confirm their suspicion, the police randomly pick 5 bundles to be inspected. If indeed, 10 from the 40 bundles contain cocaine camouflaged, what is the probability that at least 1 of the inspected bundles has cocaine?

3.31 At a TV contest the following game is played: the contestant must simultaneously extract 3 ballots from an urn containing 5 ballots marked with a prize and 9 unmarked ballots. If the 3 ballots drawn are marked, the contestant has two choices: he or she can either pick one of the marked prizes and retire or repeat the extraction three more times, and if in those repetitions all three extracted ballots are marked, then the contestant will win, in addition to all prizes marked in the ballots, a brand new car. If a contestant chooses this last alternative, what is the probability that he or she has of winning?

3.32 (Ross, 1998) The number of times a person suffers from a cold in a year is a random variable having a Poisson distribution with parameter $\lambda = 3/\text{year}$. Suppose that there is a new medicine (based on large quantities of vitamin C) that reduces the Poisson parameter to $\lambda = 2/\text{year}$ on 85% of the population and has no major effects on preventing colds for the remaining 15%. If a given person takes the medicine throughout a year and during that time has 2 colds, what is the probability that the medicine did not have an effect on that person?

3.33 Determine the expected value and the variance of the number of times that is necessary to roll a fair die until the result "1" happens 4 consecutive times.

3.34 A player has the following strategy in the roulette: He bets two tokens to the red color. If on the first spin of the wheel appears a red number, he takes the money won and retires; if on the first spin of the wheel appears a black or green number, he spins the wheel two more times betting two tokens to the red color each time and then retires. Let X be the random variable representing the fortune of the player. Find $E(X)$ (see Exercise 3.13).

3.35 To pay for his college fees, a young man has decided to sell cheese and ham sandwiches. The money needed to make each sandwich is \$0.50 and he expects to sell them at \$1.50 each. However, the sandwiches that are not sold on any day cannot be sold on the next day. If the daily demand of sandwiches is a random variable having a binomial distribution with parameters $n = 20$ and $p = \frac{2}{3}$, how many sandwiches must he make to maximize his expected daily income?

3.36 A fair coin is flipped as many times as necessary to obtain "head" for the first time. Let X be the number of required flips.

a) Find $E(X)$.

b) Find $E(2^X)$ (if it exists).

3.37 A coin is biased in such a way that the probability of obtaining a "head" equals 0.4. Use Chebyschev's inequality to determine how many times must the coin be flipped in order to have at least a 0.9 probability that the quotient between the number of heads and the number of total flips lies between 0.3 and 0.5.

3.38 Let X be a random variable having a Poisson distribution with parameter λ. Find the value of λ for which $P(X = k)$ is maximum.

3.39 A company rents out time on a computer for periods of t hours, for which it receives \$400 an hour. The number of times the computer breaks down during t hours is a random variable having the Poisson distribution with $\lambda = (0.8)t$, and if the computer breaks down x times it costs $50x$ dollars to fix it. How should the company select t in order to maximize its expected profit?

3.40 A film supplier produces 10 rolls of a specifically sensitized film each year. If the film is not sold within a year it must be discarded. Past experience indicates that D, the small demand for the film, is a Poisson-distributed random variable with parameter 8. If a profit of \$7 is made on every roll which is sold, while a loss of \$2 is incurred on every roll which must be discarded, compute the expected profit which the supplier may realize on the 10 rolls which he produces.

3.41 Let X be a random variable having a Poisson distribution with parameter λ. Prove the following:

$$P(X \leq n) = \frac{1}{n!} \int_{\lambda}^{\infty} e^{-x} x^n dx, \qquad n = 0, 1, \cdots .$$

3.42 Let X be a random variable having a binomial distribution with parameters n and p. Prove that:

$$E\left(\frac{1}{X+1}\right) = \frac{1-(1-p)^{n+1}}{(n+1)p}.$$

3.43 Let X be a random variable having a binomial distribution with parameters n and p. Find the value of p that maximizes $P(X = k)$ for $k = 0, 1, \cdots, n$.

3.44 Let X be a random variable having a hypergeometric distribution with parameters n, R and N.

a) Prove that:

$$P(X = j+1) = \frac{(R-j)(n-j)}{(j+1)((N-R)-n+j+1)} P(X = j).$$

b) Verify

$$P(X = j+1) > P(X = j)$$

if and only if

$$j < \frac{(n+1)(R+1)}{N+2} - 1.$$

3.45 Let X be a random variable having a Poisson distribution with parameter λ. Prove that:

$$E(X^n) = \lambda E\left((X+1)^{n-1}\right).$$

3.46 Let X be a random variable having a geometric distribution with parameter p. Prove that:

$$P(X = n+k \mid X > n) = P(X = k) .$$

3.47 Let X be a random variable having a geometric distribution with parameter p. Find $E\left(\frac{1}{X}\right)$.

3.48 Let X be a random variable having a Poisson distribution with parameter λ. Prove:

a) $E(X^2) = \lambda E(X+1)$.

b) If $\lambda = 1$, then $E(|X-1|) = \frac{2}{e}$.

3.49 Find the characteristic function of a random variable having a binomial distribution with parameters n and p.

3.50 Find the characteristic function of a random variable having a Poisson distribution with parameter λ.

3.51 How many children must a couple have so that, with a 0.95 probability, she gives birth to at least one boy and one girl?

3.52 A reputed publisher claims that in the handbooks published by them misprints occur at the rate of 0.0024 per page. What is the probability that in a randomly chosen handbook of 300 pages, the third misprint will occur after examining 100 pages?

CHAPTER 4

SOME CONTINUOUS DISTRIBUTIONS

In this chapter we will study some of the absolute continuous-type distributions most frequently used.

4.1 UNIFORM DISTRIBUTION

Suppose that a school bus arrives always at a certain bus stop between 6 AM and 6:10 AM and that the probability that the bus arrives in any of the time subintervals, in the interval $[0, 10]$, is proportional to the length of the subinterval. This means it is equally probable that the bus arrives between 6:00 AM and 6:02 AM as it is that it arrives between 6:07 AM and 6:09 AM. Let X be the time, measured in minutes, that a student must wait in the bus stop if he or she arrived exactly at 6:00 AM. If throughout several mornings the time of the bus arrival is measured carefully, with the data obtained, it is possible to construct a histogram of relative frequencies. From the previous description it can be noticed that the relative frequencies observed of X between 6:00 and 6:02 AM and between 6:07 and 6:09 AM are practically the same. The variable X is an example of a random variable with uniform distribution. More precisely it can be defined in the following way:

Introduction to Probability and Stochastic Processes with Applications, First Edition. **145**
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

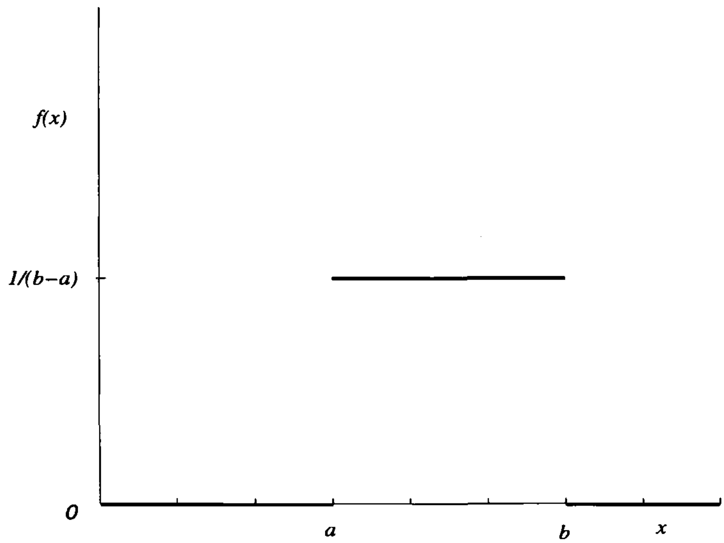


Figure 4.1 Density function of a uniform distribution

Definition 4.1 (Uniform Distribution) *It is said that a random variable X is uniformly distributed over the interval $[a, b]$, with $a < b$ real numbers, if its density function is given by:*

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a \leq x \leq b \\ 0 & \text{otherwise .} \end{cases}$$

The probability density function of a uniform distribution over the interval $[a, b]$ is shown in Figure 4.1.

Notation 4.1 *The expression $X \stackrel{d}{=} \mathcal{U}[a, b]$ means that the random variable X has a uniform distribution over the interval $[a, b]$.*

It is easy to verify that if $X \stackrel{d}{=} \mathcal{U}[a, b]$, then the cumulative distribution function of X is given by:

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{b-a} & \text{if } a \leq x < b \\ 1 & \text{if } x \geq b \text{ .} \end{cases}$$

The distribution function of a random variable with uniform distribution over the interval $[a, b]$ is shown in Figure 4.2.

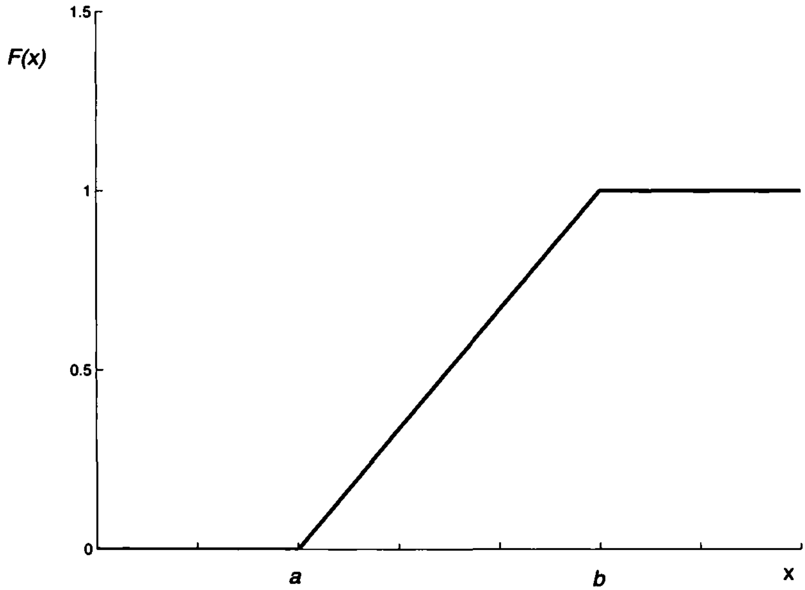


Figure 4.2 Distribution function of a random variable with uniform distribution over the interval $[a, b]$

■ EXAMPLE 4.1

Let $X \stackrel{d}{=} \mathcal{U}[-3, 2]$. Calculate:

1. $P(X \geq 0)$.
2. $P(-5 \leq X \leq \frac{1}{2})$.

Solution: In this case the density function of the random variable X is given by:

$$f(x) = \begin{cases} \frac{1}{5} & \text{if } -3 \leq x \leq 2 \\ 0 & \text{otherwise .} \end{cases}$$

Therefore,

$$P(X \geq 0) = \int_0^2 \frac{1}{5} dx = \frac{2}{5}$$

and

$$P\left(-5 \leq X \leq \frac{1}{2}\right) = \int_{-3}^{\frac{1}{2}} \frac{1}{5} dx = \frac{1}{5}\left(\frac{1}{2} + 3\right) = \frac{7}{10}. \quad \blacktriangle$$

EXAMPLE 4.2

Let $a, b \in \mathbb{R}$ be fixed, with $a < b$. A number X is chosen randomly in the interval $[a, b]$. This means that any subinterval of $[a, b]$ with length τ has the same probability of containing X. Therefore, for any $a \leq x \leq y \leq b$, we have that $P(x \leq X \leq y)$ depends only on $(y - x)$. If f is the density function of the random variable X, then:

$$kdx = P(x < X \leq x + dx) = f(x)dx \; .$$

This means $f(x) = k$ being k an appropriate constant. Given that

$$1 = \int_{-\infty}^{\infty} f(x)dx = \int_{a}^{b} kdx$$

it can be deduced that $k = \frac{1}{b-a}$. This is, $X \stackrel{d}{=} \mathcal{U}[a, b]$. ▲

EXAMPLE 4.3

A number is randomly chosen in the interval $[1, 3]$. What is the probability that the first digit to the right side of the decimal point is 5? What is the probability that the second digit to the right of the decimal point is 2?

Solution: Let $X :=$ "number randomly chosen in the interval $[1, 3]$". The density function of the random variable X according to the previous example is:

$$f(x) = \begin{cases} \frac{1}{2} & \text{if } 1 \leq x \leq 3 \\ 0 & \text{otherwise} \; . \end{cases}$$

Therefore,

$$\begin{aligned} & P(\text{"first digit to the right side of the decimal point of } X \text{ is 5"}) \\ & = P(1.5 \leq X < 1.6) + P(2.5 \leq X < 2.6) \\ & = \int_{1.5}^{1.6} \frac{1}{2} dx + \int_{2.5}^{2.6} \frac{1}{2} dx = 0.1 \end{aligned}$$

and

$$P(\text{"second digit to the right of the decimal point of } X \text{ is 2"})$$

$$\begin{aligned} & = P\left(X \in \bigcup_{k=0}^{9} \{[1.k2, 1.k3) \cup [2.k2, 2.k3)\}\right) \\ & = 20 \times \frac{1}{2} \times 0.01 = 0.1 \; . \end{aligned}$$ ▲

EXAMPLE 4.4

A point X is chosen at random in the interval $[-1,3]$. Find the *pdf* of $Y = X^2$.

Solution: For $y < 0$, $F_Y(y) = 0$. For $y \in [0,1)$:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(X^2 \leq y) \\ &= P(-\sqrt{y} \leq X \leq \sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{4} dx \\ &= \frac{\sqrt{y}}{2}. \end{aligned}$$

For $y \in [1,9)$:

$$\begin{aligned} F_Y(y) &= \int_{-1}^{1} \frac{1}{4} dx + \int_{1}^{\sqrt{y}} \frac{1}{4} dx \\ &= \frac{1}{2} + \frac{\sqrt{y}-1}{4}. \end{aligned}$$

For $y \in [9,\infty)$, $F_Y(y) = 1$. Hence, the *pdf* of Y is:

$$f_Y(y) = \begin{cases} \frac{1}{4\sqrt{y}} & \text{if } 0 < y < 1 \\ \frac{1}{8\sqrt{y}} & \text{if } 1 < y < 9 \\ 0 & \text{otherwise .} \end{cases} \quad \blacktriangle$$

EXAMPLE 4.5

An angle θ is chosen randomly on the interval $(0, \frac{\pi}{2})$. What is the probability distribution of $X = \tan\theta$? What would be the distribution of X if θ was to be chosen from $(-\pi/2, \pi/2)$?

Solution: Given that $\theta \stackrel{d}{=} \mathcal{U}[0, \frac{\pi}{2}]$:

$$f_\theta(\theta) = \begin{cases} \frac{2}{\pi} & \text{if } 0 < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise .} \end{cases}$$

Let $X = \tan\theta$. Since $\tan\theta$ is a strict monotonous and differentiable function in $\theta \in (0, \frac{\pi}{2})$, by applying Theorem 2.4:

$$f_X(x) = f_\theta(\tan^{-1}(x)) \times \left| \frac{d\tan^{-1}(x)}{dx} \right| .$$

After simplifications, we get:

$$f_X(x) = \begin{cases} \frac{2}{\pi(1+x^2)} & \text{if} \quad 0 < x < \infty \\ 0 & \text{otherwise} . \end{cases}$$

When θ is randomly chosen on the interval $(-\pi/2, \pi/2)$,

$$f_\theta(\theta) = \begin{cases} \frac{4}{\pi} & \text{if} \quad -\frac{\pi}{2} < \theta < \frac{\pi}{2} \\ 0 & \text{otherwise} . \end{cases}$$

$\tan\theta$ is again a strict monotonous and differentiable function. By applying Theorem 2.4, we get:

$$f_X(x) = \frac{4}{\pi(1+x^2)}, \quad -\infty < x < \infty. \qquad \blacktriangle$$

Theorem 4.1 *If X is a random variable with uniform distribution over the interval $[a, b]$, then:*

1. $E(X) = \frac{a+b}{2}$.
2. $Var(X) = \frac{(b-a)^2}{12}$.
3. $m_X(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$.

Proof:

1.

$$\begin{aligned} E(X) &= \int_a^b x \frac{1}{b-a} dx \\ &= \frac{a+b}{2}. \end{aligned}$$

2.

$$\begin{aligned} E(X^2) &= \int_a^b x^2 \frac{1}{b-a} dx \\ &= \frac{b^2+ab+a^2}{3} . \end{aligned}$$

Therefore:

$$\begin{aligned} Var(X) &= \frac{b^2-2ab+a^2}{12} \\ &= \frac{(b-a)^2}{12}. \end{aligned}$$

3. Follows from the definition of the mgf.

$\blacksquare$

■ **EXAMPLE 4.6**

Suppose that $X \stackrel{d}{=} \mathcal{U}[a,b]$ and that $E(X) = 2$ and $Var(X) = \frac{3}{4}$. Calculate $P(X \leq 1)$.

Solution: We have that $\frac{a+b}{2} = 2$ and $\frac{(b-a)^2}{12} = \frac{3}{4}$. Due to this, $a = \frac{1}{2}$ and $b = \frac{7}{2}$. Then:

$$P(X \leq 1) = \int_{\frac{1}{2}}^{1} \frac{1}{3} dx = \frac{1}{6}. \quad \blacktriangle$$

Note 4.1 *Suppose that X is a continuous random variable with increasing distribution function $F_X(x)$. Let Y be a random variable with uniform distribution in the interval $(0,1)$ and let Z be a random variable defined as $Z := F_X^{-1}(Y)$. The distribution function of the random variable Z is given by:*

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P\left(F_X^{-1}(Y) \leq z\right) \\ &= P(Y \leq F_X(z)) \\ &= F_X(z). \end{aligned}$$

That is, the random variables X and Z have the same probability distribution.

4.2 NORMAL DISTRIBUTION

The normal distribution is one of the most important and mainly used not only in probability theory but also in statistics. Some authors name it Gaussian distribution in honor of Gauss, who is considered the "father" of this distribution. The importance of the normal distribution is due to the famous central limit theorem, which will be discussed in Chapter 8.

Definition 4.2 *It is said that a random variable X has normal distribution with parameters μ and σ, where μ is a real number and σ is a positive real number, if its density function is given by:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right], \quad x \in \mathbb{R}.$$

It is left as an exercise for the reader to verify that f is effectively a density function. That is, f is nonnegative and:

$$\int_{-\infty}^{\infty} f(x)dx = 1.$$

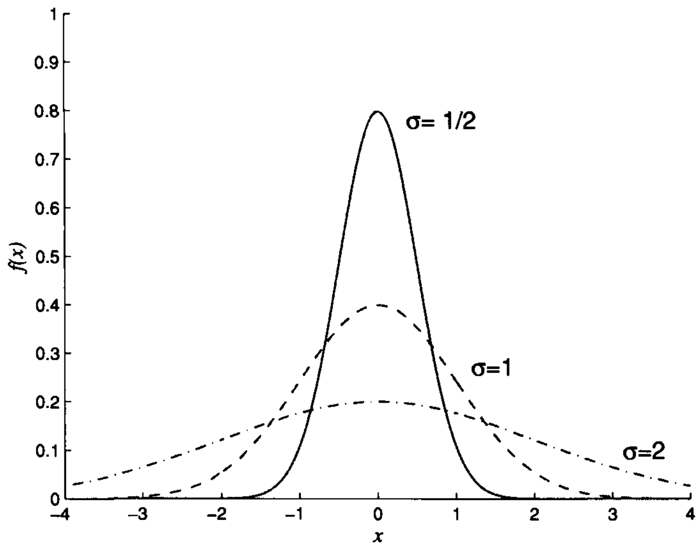


Figure 4.3 Probability density function of normal distribution with $\mu = 0$ and different values of σ

The parameters μ and σ are called location parameter and scale parameter, respectively. We precisely give the concepts to follow.

Definition 4.3 (Location and Scale Parameters) *Let Y be a random variable. It is said that θ_1 is a location parameter if for all $c \in \mathbb{R}$ we have that a random variable $Z := Y + c$ has parameter $\theta_1 + c$. That is, if $f_Y(\cdot; \theta_1, \theta_2)$ is the density function of Y, then the density function of Z is $f_Z(\cdot; \theta_1 + c, \cdot)$. It is said that θ_2 is a scale parameter if $\theta_2 > 0$ and for all $c \in \mathbb{R}$ the random variable $W := cY$ has parameter $|c|\,\theta_2$. That is, if $f_Y(\cdot; \theta_1, \theta_2)$ is the density function of Y, then the density function of W is $f_W(\cdot; \cdot, |c|\,\theta_2)$.*

Notation 4.2 *We write $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ to indicate that X is a random variable with normal distribution with parameters μ and σ.*

Figure 4.3 shows the density function of the random variable X with normal distribution with $\mu = 0$ and different values of σ.

In Figure 4.4, it can be seen the density function of the random variable $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ for $\sigma = 1.41$ and different values of μ.

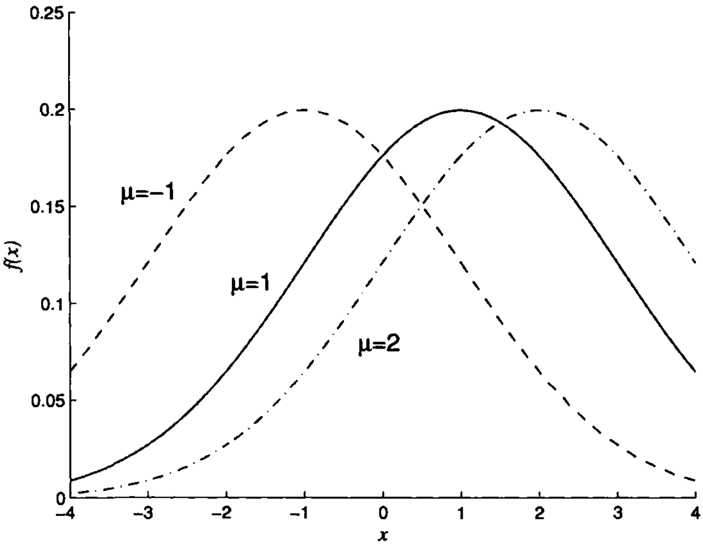


Figure 4.4 Probability density function of normal distribution with $\sigma = 1.41$ and different values of μ

The distribution function of the random variable $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ is given by:

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left[-\frac{1}{2}\left(\frac{u-\mu}{\sigma}\right)^2\right] du.$$

The graph of F, with $\mu = 0$ and $\sigma = 1$, is given in Figure 4.5.

Definition 4.4 (Standard Normal Distribution) *If $X \stackrel{d}{=} \mathcal{N}(0, 1)$, then it is said that X has a standard normal distribution. The density function and the distribution function of the random variable are denoted by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively.*

Note 4.2 *The density function of a standard normal random variable is symmetric with respect to the y axis. Therefore, for all $z < 0$ it is satisfied that:*

$$\Phi(z) = 1 - \Phi(-z).$$

Note 4.3 *Let $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$ and let $Y := aX + b$ where a and b are real constants with $a \neq 0$. As seen in Chapter 2, it is known that the density*

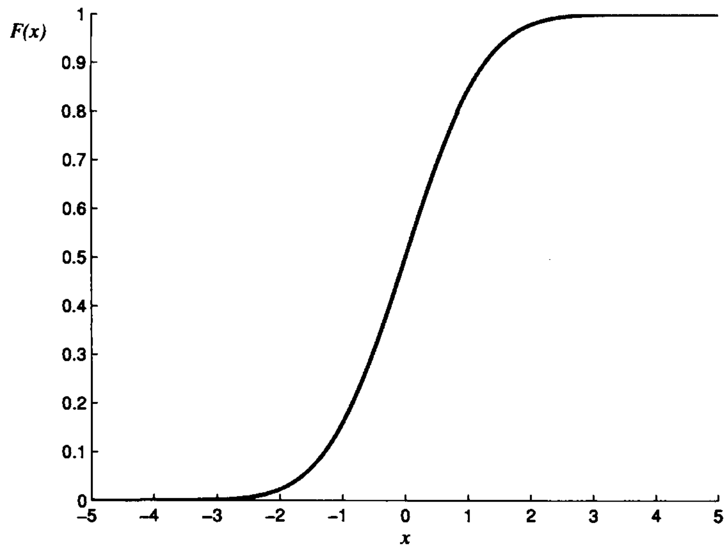


Figure 4.5 Standard normal distribution function with $\mu = 0$ and $\sigma = 1$

function of the random variable Y is given by:

$$\begin{aligned} f_Y(x) &= \frac{1}{|a|} f_X\left(\frac{x-b}{a}\right) \\ &= \frac{1}{|a|}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2}\left(\frac{x-(a\mu+b)}{a\sigma}\right)^2\right]. \end{aligned}$$

This means Y has a normal distribution with location parameter $a\mu + b$ and scale parameter $|a|\,\sigma$. Particularly, if $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$, then $Y = \frac{X-\mu}{\sigma}$ has a standard normal distribution. That is, in order to know the values of the random variable's distribution function with arbitrary normal distribution, it is sufficient to know the values of the random variable with standard normal distribution. In Appendix D.3 the table of values of the standard normal distribution are provided.

■ EXAMPLE 4.7

Let $X \stackrel{d}{=} \mathcal{N}(1, 4)$. Calculate:

1. $P(0 \leq X < 1)$.
2. $P(X^2 > 4)$.

Solution: It is known that

$$\begin{aligned} P(0 \leq X < 1) &= P\left(-\frac{1}{2} \leq \frac{X-1}{2} < 0\right) \\ &= \Phi(0) - \Phi\left(-\frac{1}{2}\right) \\ &= 0.5 - 0.30854 \\ &= 0.19146 \end{aligned}$$

and

$$\begin{aligned} P\left(X^2 > 4\right) &= 1 - P(|X| \leq 2) \\ &= 1 - P(-2 \leq X \leq 2) \\ &= 1 - P\left(-\frac{3}{2} \leq \frac{X-1}{2} \leq \frac{1}{2}\right) \\ &= 1 - \left[\Phi\left(\frac{1}{2}\right) - \Phi\left(-\frac{3}{2}\right)\right] \\ &= 1 - [0.69146 - 0.06681] \\ &= 0.37535\ . \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 4.8

Let $X \stackrel{d}{=} \mathcal{N}(12, 4)$. Find the value of c such that $P(X > c) = 0.10$.
Solution:

$$\begin{aligned} P(X > c) &= 1 - P(X \leq c) \\ &= 1 - P\left(\frac{X-12}{2} \leq \frac{c-12}{2}\right) \\ &= 1 - \Phi\left(\frac{c-12}{2}\right). \end{aligned}$$

That is:

$$\Phi\left(\frac{c-12}{2}\right) = 0.9.$$

So that the values given in the table we have that:

$$\frac{c-12}{2} = 1.285.$$

and $c = 14.57$. ▲

■ EXAMPLE 4.9

Suppose that the life lengths of two electronic devices say, D_1 and D_2, have normal distributions $\mathcal{N}(40, 36)$ and $\mathcal{N}(45, 9)$, respectively. If a device is to be used for 45 hours, which device would be preferred? If it is to be used for 42 hours, which one should be preferred?

Solution: Given that $D_1 \stackrel{d}{=} \mathcal{N}(40, 36)$ and $D_2 \stackrel{d}{=} \mathcal{N}(45, 9)$. We will find which device has greater probability of lifetime more than 45 hours:

$$\begin{aligned} P(D_1 > 45) &= P\left(\frac{D_1 - 40}{6} > \frac{45 - 40}{6}\right) \\ &= 1 - \Phi\left(\frac{5}{6}\right) \\ &= 1 - 0.7995 \\ &= 0.2005. \end{aligned}$$

Also:

$$\begin{aligned} P(D_2 > 45) &= P\left(\frac{D_2 - 45}{3} > \frac{45 - 45}{3}\right) \\ &= 1 - \Phi(0) \\ &= 1 - 0.5 \\ &= 0.5. \end{aligned}$$

Hence, in this case, the device D_2 will be preferred. Now, we will find which device has greater probability of lifetime more than 42 hours. Similar calculation yields:

$$\begin{aligned} P(D_1 > 42) &= 1 - \Phi\left(\frac{1}{3}\right) \\ &= 0.3707 \\ P(D_2 > 42) &= 1 - \Phi(-1) \\ &= 0.8413. \end{aligned}$$

In this case also, the device D_2 will be preferred. ▲

■ EXAMPLE 4.10

Let X denote the length of time (in minutes) an automobile battery will continue to crank an engine. Assume that $X \sim \mathcal{N}(10, 4)$. What is the probability that the battery will crank the engine longer than $10 + x$ minutes given that it is still cranking at 10 minutes?

Solution: We want to find:

$$P(X > 10 + x \mid X > 10) \quad = \quad \frac{P(X > 10 + x)}{P(X > 10)} = \frac{P(Z > x/2)}{1/2} \ .$$

For a specific choice $x = 2$, we get:

$$P(X > 10 + x \mid X > 10) = 2[1 - \phi(1)] = 0.3174. \qquad \blacktriangle$$

Note 4.4 *Suppose that* $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$. *Then*

$$\begin{aligned} P\left(\mu - 3\sigma \leq X \leq \mu + 3\sigma\right) &= P\left(\frac{\mu - 3\sigma - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{\mu + 3\sigma - \mu}{\sigma}\right) \\ &= \Phi(3) - \Phi(-3) \\ &= 0.99865 - 0.00135 \\ &= 0.9973 \end{aligned}$$

or equivalently $P\left(|X - \mu| > 3\sigma\right) = 0.0027$.

Next, we will find the expected value, the variance and the moment generating function of a random variable with normal distribution.

Theorem 4.2 *Let* $X \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$. *Then:*

1. $E(X) = \mu$.

2. $Var(X) = \sigma^2$.

3. $m_X(t) = \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]$.

Proof: We will calculate the moment generating function. From it, we can easily find the expected value and the variance:

$$\begin{aligned} m_X\left(t\right) &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[tx - \frac{\left(x - \mu\right)^2}{2\sigma^2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\infty} \exp\left[-\frac{(x - \mu - \sigma^2 t)^2}{2\sigma^2} + \mu t + \frac{\sigma^2 t^2}{2}\right] dx \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{\left(x - \{\mu + \sigma^2 t\}\right)^2}{2\sigma^2}\right] dx \\ &= \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]. \end{aligned}$$

Hence

$$E(X) = [\mu + \sigma^2 t] \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]\bigg|_{t=0} = \mu$$

and

$$E(X^2) = \left(\sigma^2 + [\mu + \sigma^2 t]^2\right) \exp\left[\mu t + \frac{\sigma^2 t^2}{2}\right]\bigg|_{t=0} = \sigma^2 + \mu^2$$

we get:

$$Var(X) = \sigma^2.$$

■

Note 4.5 *It can be verified that the characteristic function of the random variable X with normal distribution with parameters μ and σ is given by:*

$$\Phi_X(t) = \exp\left[i\mu t - \frac{\sigma^2 t^2}{2}\right].$$

Note 4.6 *The normal distribution is another limit form of the binomial distribution only if the following conditions over the parameters n and p are satisfied in the binomial distribution: $n \to \infty$ and if neither p nor $q = 1 - p$ is very small.*

Suppose that $X \stackrel{d}{=} \mathcal{B}(n,p)$. Then:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \cdots, n \ .$$

When $n \to \infty$, we have that $x \to \infty$ and additionally:

$$n! \approx \sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}} \qquad \textit{(Stirling's formula)} \ .$$

Therefore:

$$\begin{aligned}
\lim_{n\to\infty} \binom{n}{x} p^x (1-p)^{n-x} &= \lim_{n\to\infty} \frac{\sqrt{2\pi} e^{-n} n^{n+\frac{1}{2}} p^x (1-p)^{n-x}}{\sqrt{2\pi} e^{-x} x^{x+\frac{1}{2}} \sqrt{2\pi} e^{-(n-x)} (n-x)^{(n-x)+\frac{1}{2}}} \\
&= \lim_{n\to\infty} \frac{n^{n+\frac{1}{2}} p^x (1-p)^{n-x} \sqrt{np(1-p)}}{\sqrt{2\pi} x^{x+\frac{1}{2}} (n-x)^{(n-x)+\frac{1}{2}} \sqrt{np(1-p)}} \\
&= \lim_{n\to\infty} \frac{(np)^{x+\frac{1}{2}} \left(n\left(1-p\right)\right)^{n-x+\frac{1}{2}}}{\sqrt{2\pi} x^{x+\frac{1}{2}} (n-x)^{(n-x)+\frac{1}{2}} \sqrt{np(1-p)}} \\
&= \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}\sqrt{np(1-p)}} \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{n(1-p)}{n-x}\right)^{(n-x)+\frac{1}{2}} .
\end{aligned}$$

Let $\frac{1}{N} := \left(\frac{np}{x}\right)^{x+\frac{1}{2}} \left(\frac{n(1-p)}{n-x}\right)^{(n-x)+\frac{1}{2}}$. *It is clear that:*

$$\ln N = \left(x + \frac{1}{2}\right) \ln\left(\frac{x}{np}\right) + \left(n - x + \frac{1}{2}\right) \ln\left(\frac{n-x}{n(1-p)}\right) . \tag{4.1}$$

If we take $Z := \frac{X-np}{\sqrt{np(1-p)}}$, *we have that* Z *takes the values* $z = \frac{x-np}{\sqrt{np(1-p)}}$. *When considering the limit when* $n \to \infty$, *we have that* Z *takes all of the values between* $-\infty$ *and* ∞. *Isolating* x *in the previous equation it is obtained that* $x = z\sqrt{np(1-p)} + np$. *Replacing in (4.1) we get:*

$$\begin{aligned} \ln N &= \left(z\sqrt{np(1-p)} + np + \frac{1}{2}\right) \ln\left(\frac{z\sqrt{np(1-p)} + np}{np}\right) \\ &\quad + \left(n - \left(z\sqrt{np(1-p)} + np\right) + \frac{1}{2}\right) \ln\left(\frac{n - \left(z\sqrt{np(1-p)} + np\right)}{n(1-p)}\right) \\ &= \left(np + z\sqrt{np(1-p)} + \frac{1}{2}\right) \ln\left(1 + z\sqrt{\frac{1-p}{np}}\right) \\ &\quad + \left(n(1-p) - z\sqrt{np(1-p)} + \frac{1}{2}\right) \ln\left(1 - z\sqrt{\frac{p}{n(1-p)}}\right). \end{aligned}$$

Developing the function $h(x) = \ln(1+x)$, *it is obtained:*

$$\begin{aligned} \ln N &= \left(z\sqrt{np(1-p)} + np + \frac{1}{2}\right) \left[z\sqrt{\frac{1-p}{np}} - \frac{1}{2}z^2\left(\frac{1-p}{np}\right) + \cdots\right] \\ &\quad + \left(n(1-p) - z\sqrt{np(1-p)} + \frac{1}{2}\right) \\ &\qquad \cdot \left[-z\sqrt{\frac{p}{n(1-p)}} - \frac{1}{2}z^2\left(\frac{p}{n(1-p)}\right) - \cdots\right] \\ &= \left[z^2(1-p) - \frac{1}{2}z^3\sqrt{\frac{(1-p)^3}{np}} + z\sqrt{np(1-p)} - \frac{1}{2}z^2(1-p)\right. \\ &\qquad \left. + \frac{1}{2}z\sqrt{\frac{1-p}{np}} - \frac{1}{4}z^2\left(\frac{1-p}{np}\right) + \cdots\right] \\ &\quad + \left[-z\sqrt{np(1-p)} - \frac{1}{2}z^2p + z^2p + \frac{1}{2}z^3\sqrt{\frac{p^3}{n(1-p)}} - \frac{1}{2}z\sqrt{\frac{p}{n(1-p)}}\right. \\ &\qquad \left. - \frac{1}{4}z^2\left(\frac{p}{n(1-p)}\right) - \cdots\right]. \end{aligned}$$

That is:

$$\ln N = -\frac{1}{2}z^2 + z^2 + \frac{z}{2\sqrt{n}}\left(\sqrt{\frac{1-p}{p}} + \sqrt{\frac{p}{1-p}}\right) + o(n^{-\frac{1}{2}}).$$

Therefore,

$$\lim_{n\to\infty} \ln N = \frac{1}{2}z^2$$

and hence:

$$\lim_{n\to\infty} N = e^{\frac{1}{2}z^2}.$$

Given that

$$\begin{aligned}
P(X = x) &= P(x < X \leq x + dx) \\
&= P\left(\frac{x - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{x + dx - np}{\sqrt{np(1-p)}}\right) \\
&= P(z < Z \leq z + dz) \\
&\approx g(z)dz
\end{aligned}$$

where $g(\cdot)$ is the density function of the random variable Z, and

$$\begin{aligned}
g(z) &= \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \times \frac{1}{N} \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right).
\end{aligned}$$

That is, $Z \stackrel{d}{=} \mathcal{N}(0,1)$. In other words, if n is sufficiently large $\mathcal{B}(n,p) \approx \mathcal{N}(np, np(1-p))$. In practice, the approximation is generally acceptable when $p \in \left(0, \frac{1}{2}\right)$ and $np(1-p) > 9$ or $np > 5$, or if $p \in \left(\frac{1}{2}, 1\right)$ and $n(1-p) > 5$. In the case of $p = \frac{1}{2}$ it is obtained that the approximation is quite good even in the case in which n is "small" (see Hernandez, 2003).

The result that has recently been deduced is known as the Moivre-Laplace theorem and as it will be seen later on that it is a particular case of the central limit theorem.

■ EXAMPLE 4.11

A normal die is tossed 1000 consecutive times. Calculate the probability that the number 6 shows up between 150 and 200 times. What is the probability that the number 6 appears exactly 150 times?

Solution: Let $X :=$"Number of times the number 6 is obtained as a result". It is clear that $X \stackrel{d}{=} \mathcal{B}(1000, \frac{1}{6})$. According to the previous result, it can be supposed that X has a normal distribution with parameters

$\mu = \frac{500}{3}$ and $\sigma^2 = \frac{1250}{9}$. Therefore:

$$P(150 \le X \le 200) = P\left(\frac{150 - \frac{500}{3}}{\sqrt{\frac{1250}{9}}} \le \frac{X - \frac{500}{3}}{\sqrt{\frac{1250}{9}}} \le \frac{200 - \frac{500}{3}}{\sqrt{\frac{1250}{9}}}\right)$$

$$\begin{aligned}
&= P(-1.14142 \le Z \le 2.8284) \quad \text{where } Z \stackrel{d}{=} \mathcal{N}(0,1) \\
&= \Phi(2.8284) - \Phi(-1.14142) \\
&= 0.9976 - 0.07927 \\
&= 0.91833. \quad \blacktriangle
\end{aligned}$$

To answer the second part of the question, it may be seen that because the binomial distribution is discrete and the normal distribution is a continuous one, an appropriate approximation is obtained as below:

$$\begin{aligned}
P(X = 150) &= P(149.5 \le X \le 150.5) \\
&= P\left(\frac{149.5 - \frac{500}{3}}{\sqrt{\frac{1250}{9}}} \le \frac{X - \frac{500}{3}}{\sqrt{\frac{1250}{9}}} \le \frac{150.5 - \frac{500}{3}}{\sqrt{\frac{1250}{9}}}\right) \\
&= P(-1.4566 \le Z \le -1.3718) \\
&= \Phi(-1.3718) - \Phi(-1.4566) \\
&= 0.08534 - 0.07215 \\
&= 0.01319\ . \quad \blacktriangle
\end{aligned}$$

4.3 FAMILY OF GAMMA DISTRIBUTIONS

Some random variables are always nonnegative and they have distributions that are biased to the right, that is, the greater part of the area below the graph of the density function is close to the origin and the values of the density function decrease gradually when x increases. An example of such distributions is the gamma distribution whose density function is shown in Figure 4.6.

The gamma distribution is used in an extensive way in a variety of areas as, for example, to describe the intervals of time between two consecutive failures of an airplane's motor or the intervals of time between arrivals of clients to a queue in a supermarket's cashier point.

The gamma distribution is the generalization of three particular cases that, historically, came first: the exponential function, the Erlang function and the chi-square distribution.

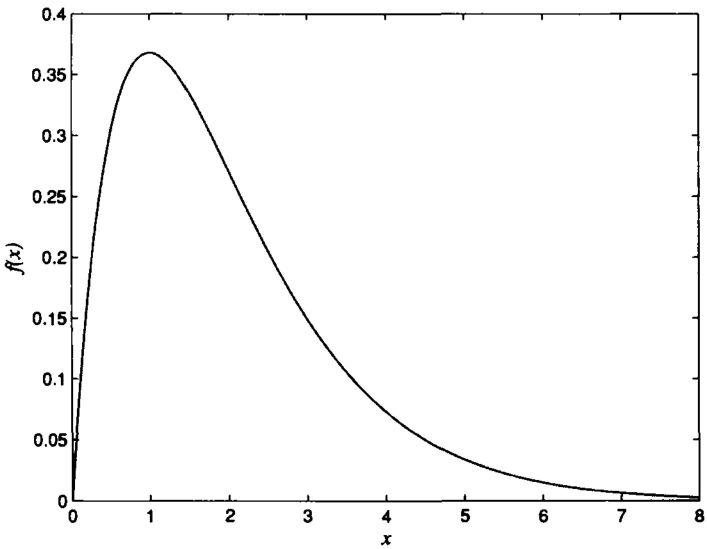


Figure 4.6 Probability density function of a gamma distribution with parameters $r = 2$ and $\lambda = 1$

Definition 4.5 (Gamma Distribution) *It is said that the random variable X has gamma distribution with parameters $r > 0$ and $\lambda > 0$ if its density function is given by*

$$f(x) = \begin{cases} \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} \exp(-\lambda x) & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \tag{4.2}$$

where $\Gamma(\cdot)$ is the gamma function, that is:

$$\Gamma(r) := \int_0^\infty t^{r-1} \exp(-t) dt.$$

The order of the parameters is important due to the fact that r is the shape parameter while λ is the scale parameter.

The verification that f is a density function is left as an exercise for the reader.

Figure 4.7 shows the gamma density function for $\lambda = 1$ and different values of r.

Figure 4.8 shows the form of the gamma density function for $r = 1.5$ and different values of λ.

Notation 4.3 *The expression $X \stackrel{d}{=} \Gamma(r, \lambda)$ means that X has a gamma distribution with parameters r and λ.*

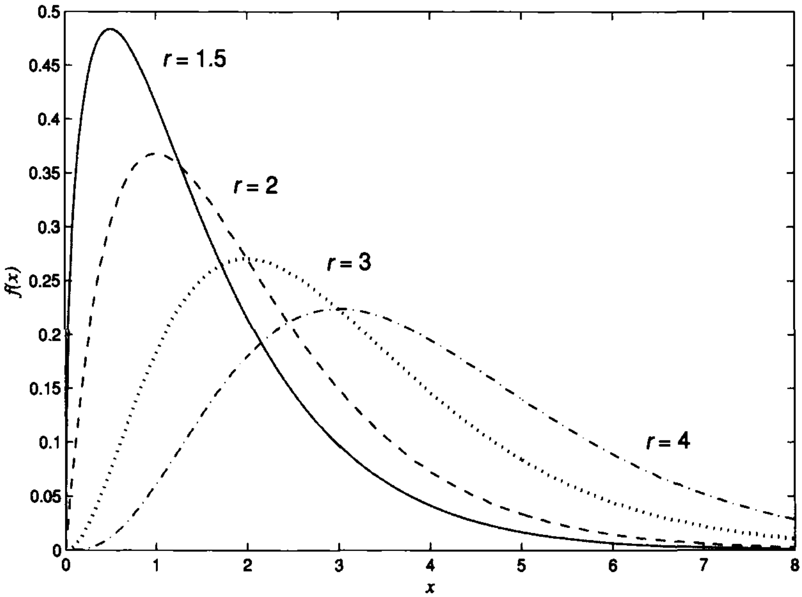


Figure 4.7 Probability density function of a gamma density function for $\lambda = 1$ and different values of r

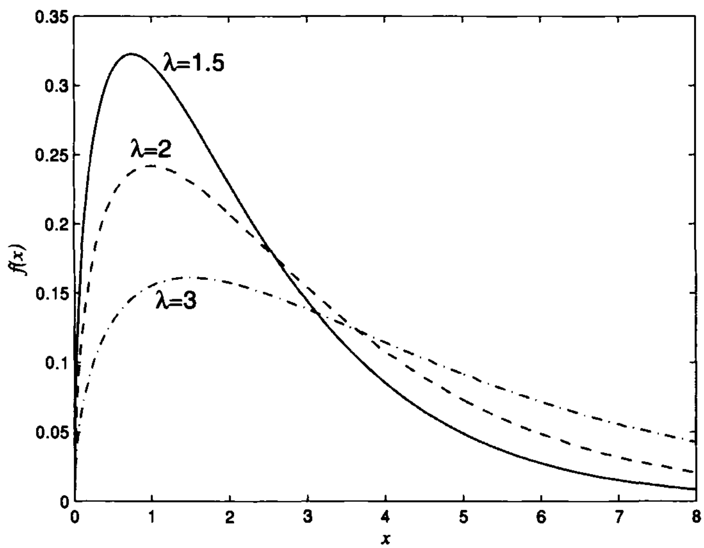


Figure 4.8 Probability density function of a gamma density function for $r = 1.5$ and different values of λ

The distribution function of the random variable X with gamma distribution with parameters r and λ is given by:

$$F(x) = \int_0^x \frac{\lambda}{\Gamma(r)} (\lambda t)^{r-1} \exp(-\lambda t)\, dt$$
$$= \frac{1}{\Gamma(r)} \int_0^{\lambda x} u^{r-1} \exp(-u) du.$$

When r is a positive integer, it is known that $\Gamma(r) = (r-1)!$ using which we get:

$$F(x) = 1 - \exp(-\lambda x) \sum_{k=0}^{r-1} \frac{(\lambda x)^k}{k!}.$$

It can be seen that the right side of the above equation corresponds to $P(Y \geq r)$ where $Y \overset{d}{=} \mathcal{P}(\lambda x)$. In Chapter 9 we will see that there is a relationship between the Poisson distribution and the gamma distribution.

In the following theorem we will determine the expected value, the variance and the moment generating function of a random variable with gamma distribution.

Theorem 4.3 *If $X \overset{d}{=} \Gamma(r, \lambda)$, then:*

1. $E(X) = \frac{r}{\lambda}$.
2. $Var(X) = \frac{r}{\lambda^2}$.
3. $m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r$ *if* $t < \lambda$.

Proof: We will calculate the moment generating function of X and then, from it, we will find $E(X)$ and $Var(X)$. It is known that:

$$\begin{aligned} m_X(t) &= \int_0^\infty \exp(tx) \frac{\lambda}{\Gamma(r)} (\lambda x)^{r-1} \exp(-\lambda x)\, dx \\ &= \left(\frac{\lambda}{\lambda - t}\right)^r \int_0^\infty \frac{(\lambda - t)^r}{\Gamma(r)} x^{r-1} \exp(-(\lambda - t)x) dx. \end{aligned}$$

If $(\lambda - t) > 0$, then

$$g(x) := \frac{(\lambda - t)^r}{\Gamma(r)} x^{r-1} \exp(-(\lambda - t)x) \mathcal{X}_{(0,\infty)}(x)$$

is gamma density function, and therefore:

$$\int_0^\infty \frac{(\lambda - t)^r}{\Gamma(r)} x^{r-1} \exp(-(\lambda - t)x) dx = 1.$$

Then:

$$m_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r \quad \textit{if } t < \lambda.$$

Further:

$$\begin{aligned} E(X) &= \frac{d}{dt} m_X(t)\bigg|_{t=0} = \frac{r}{\lambda} \\ E(X^2) &= \frac{d^2}{dt^2} m_X(t)\bigg|_{t=0} = \frac{r^2 + r}{\lambda^2}. \end{aligned}$$

■

In particular cases, $r = 1$ and $\lambda > 0$, $\lambda = \frac{1}{2}$ and $r = \frac{k}{2}$ with k positive integer, and $r > 1$ and $\lambda > 0$, we get,respectively, the exponential distribution, the chi-square distribution with k degrees of freedom and the Erlang distribution.

Notation 4.4 *The expression $X \stackrel{d}{=} Exp(\lambda)$ indicates that X has an exponential distribution with parameter λ.*

The expression $X \stackrel{d}{=} \mathcal{X}^2_{(k)}$ indicates that X has a chi-square distribution with k degrees of freedom.

The expression $X \stackrel{d}{=} Erlang(r, \lambda)$ indicates that X has an Erlang distribution with parameters r and λ.

■ EXAMPLE 4.12

The time (in minutes) required to obtain a response in a human exposed to tear gas A has a gamma distribution with parameter $r = 2$ and $\lambda = \frac{1}{2}$. The distribution for a second tear gas B is also gamma but has parameters $r = 1$ and $\lambda = \frac{1}{4}$.

1. Calculate the mean time required to get a response in a human exposed to each tear gas formula.

2. Calculate the variance for both distributions.

3. Which tear gas is more likely to cause a human response in less than 1 minute?

Solution: Let X_1 X_2 be response times from tear gas A and B, respectively.

1. Mean:

$$E(X_1) = 4; \quad E(X_2) = 4.$$

2. Variance:
$$Var(X_1) = 8; \quad Var(X_2) = 16.$$

3. We need to evaluate $P(X_i < 1),\ i = 1, 2$.
$$\begin{aligned} P(X_1 < 1) &= 1 - \frac{3}{2\sqrt{e}} = 0.0902 \\ P(X_2 < 1) &= 1 - e^{-4} = 0.9817 \ . \end{aligned}$$

   Hence the second tear gas is more likely to cause a human response. ▲

In the next example, we will prove that if $Z \stackrel{d}{=} \mathcal{N}(0,1)$ then $Z^2 \stackrel{d}{=} \chi_1^2$.

■ **EXAMPLE 4.13**

If Z is a standard normal random variable, find the *pdf* of $Y = Z^2$.
Solution: The *cdf* of Y is given by:

$$\begin{aligned} F_Y(y) &= P[Y \le y] = P[Z^2 \le y] = P[-\sqrt{y} \le Z \le \sqrt{y}] \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = \Phi(\sqrt{y}) - (1 - \Phi(\sqrt{y})) \ . \end{aligned}$$

Hence,
$$F_Y(y) = 2\Phi(\sqrt{y}) - 1$$
where $\Phi(z)$ is the *cdf* of Z. By differentiating the above equation, we obtain:

$$\begin{aligned} f_Y(y) &= 2\frac{1}{2\sqrt{y}} f_Z(\sqrt{y}) \\ &= \frac{1}{\sqrt{y}\sqrt{2\pi}} e^{-y/2} \ . \end{aligned}$$

Hence, the *pdf* of Y is given by:
$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, from equation (4.2) with $\lambda = \frac{1}{2}$ and $r = \frac{1}{2}$, we conclude that Y has a chi-square distribution with 1 degree of freedom. This problem can also be solved using Corollary 2.2. Since $y = z^2$, then $z_1 = \sqrt{y}, z_2 = -\sqrt{y}$.

$$\begin{aligned} f_Y(y) &= f_Z(\sqrt{y}) \left| \frac{d(\sqrt{y})}{dy} \right| + f_Z(-\sqrt{y}) \left| \frac{d(-\sqrt{y})}{dy} \right| \\ &= \begin{cases} \frac{1}{\sqrt{2\pi y}} e^{-y/2} & \text{if } y > 0 \\ 0 & \text{otherwise.} \end{cases} \quad ▲ \end{aligned}$$

Note 4.7 *If* $X \stackrel{d}{=} Exp(\lambda)$, *then* $E(X) = \frac{1}{\lambda}$; $Var(X) = \frac{1}{\lambda^2}$ *and* $m_X(t) = \frac{\lambda}{\lambda - t}$ *for* $t < \lambda$.

If $X \stackrel{d}{=} \mathcal{X}^2_{(k)}$, *then* $E(X) = k$; $Var(X) = 2k$ *and* $m_X(t) = (1-2t)^{-\frac{k}{2}}$ *for* $t < \frac{1}{2}$

The exponential distribution is frequently used as a model to describe the distribution of the time elapsed between successive occurrences of events, as in the case of the clients who arrive at a bank, calls that enter a call-center, etc. It is also used to model the distribution of the lifetime of components that do not deteriorate or get better through time, that is, those components whose distribution of the remaining lifetime are independent of the actual age. Therefore, this model adjusts to reality only if the distribution of the component's remaining lifetime does not depend on its age. More precisely, we have the following result:

Theorem 4.4 *Let* X *be a random variable such that* $P(X > 0) > 0$. *Then*

$$X \stackrel{d}{=} Exp(\lambda) if \text{ and only } if \; P(X > x + t \mid X > t) = P(X > x)$$

for all $x, t \in [0, \infty)$.

Proof: $\Longrightarrow$) *Suppose that* $X \stackrel{d}{=} Exp(\lambda)$. *Then:*

$$\begin{aligned} P(X > x + t \mid X > t) &= \frac{P(X > x + t)}{P(X > t)} \\ &= \frac{\exp(-\lambda(x+t))}{\exp(-\lambda t)} \\ &= \exp(-\lambda x) \\ &= P(X > x). \end{aligned}$$

$\Longleftarrow$) *Let* $G(x) = P(X > x)$. *Then, by hypothesis,*

$$G(x+t) = G(x)G(t)$$

which implies that $G(x) = \exp(-\lambda x)$ *with* λ *being a constant greater than* 0. *Indeed:*

$$G\underbrace{\left(\frac{1}{n} + \frac{1}{n} + \cdots + \frac{1}{n}\right)}_{n \ times} = \underbrace{G\left(\frac{1}{n}\right) G\left(\frac{1}{n}\right) \cdots G\left(\frac{1}{n}\right)}_{n \ times}.$$

Thus,

$$G(1) = \left[G\left(\frac{1}{n}\right)\right]^n$$

or equivalently:

$$G\left(\frac{1}{n}\right) = [G(1)]^{\frac{1}{n}}.$$

In the same way, we can obtain for $m, n \in \mathbb{N}$ the following result:

$$G\left(\frac{m}{n}\right) = \left[G\left(\frac{1}{n}\right)\right]^m$$
$$= [G(1)]^{\frac{m}{n}}.$$

As G is a continuous function to the right, it can be concluded that:

$$G(x) = [G(1)]^x .$$

On the other hand, we have that $0 < G(1) < 1$. Indeed, if $G(1) = 1$, then $G(x) = 1$, which contradicts that $G(\infty) = 0$. If $G(1) = 0$ then $G\left(\frac{1}{m}\right) = 0$ and due to the continuity to the right it can be concluded that $G(0) = 0$, which contradicts the hypothesis. Therefore, we can take $\lambda := -\ln[G(1)]$ in order to obtain the result. ■

The exponential distribution is used in some cases to describe the lifetime of a component. Let T be the random variable that denotes the lifetime of a given component and let f be its density function. It is clear that T is nonnegative. The reliability function of the device is defined by

$$R(t) = P(T > t) = 1 - F_T(t)$$

where $F_T(t)$ is the distribution function of the random variable T. The mean time to failure (MTTF) is defined to be the expected lifetime of the device:

$$E[T] = \int_0^\infty P(T > t) = \int_0^\infty R(t)dt.$$

Suppose that we want to know the probability that the component fails during the next Δt units of time given that it is working correctly until time t. If F is the distribution function of the random variable T and if $F(t) < 1$, then:

$$P(t \leq T \leq t + \Delta t \mid T > t) = \frac{P(t \leq T \leq t + \Delta t)}{1 - P(T \leq t)}$$
$$\approx \frac{f(t)\Delta t}{1 - F(t)} =: \lambda(t)\Delta t.$$

The function $\lambda(t)$ is known as the *risk function or failure rate* associated with the random variable T. The function $R(t) := 1 - F(t)$ is also known as the *confiability function*. The previous expression indicates that if we know the density function of the lifetime of the component, then we know its failure rate. Next, we will see that the converse is also valid. Writing $\lambda(t)$ as

$$\lambda(t) = \frac{\frac{d}{dt}F(t)}{1 - F(t)},$$

and integrating on both sides, we obtain

$$\ln(1 - F(t)) = -\int_0^t \lambda(s)ds + C.$$

That is:

$$F(t) = 1 - \exp(C)\exp\left(-\int_0^t \lambda(s)ds\right).$$

It is reasonable to suppose that $F(0) = 0$, that is, that the probability of instant failure of the component is zero. In such cases, we have that $C = 0$ and therefore

$$F(t) = \begin{cases} 1 - \exp\left(-\int_0^t \lambda(s)ds\right) & \text{if} \quad t \geq 0. \end{cases}$$

which let us know the distribution function of the random variable T from the risk function. If the failure rate is assumed to be a constant and equal to $\lambda > 0$, it can be seen that for $t \geq 0$:

$$F(t) = 1 - \exp(-\lambda t).$$

This indicates that the random variable T has an exponential distribution with parameter λ.

■ **EXAMPLE 4.14**

The length of lifetime T, in hours, of a certain device has an exponential distribution with mean 100 hours. Calculate the reliability at time $t = 200$ hours.

Solution: The density function of the random variable T is given by:

$$f(t) = \frac{1}{100}\exp(-\frac{1}{100}t)\mathcal{X}_{(0,\infty)}(t) .$$

Therefore:

$$\begin{aligned} P(T \geq 200) &= 1 - P(T < 200) \\ &= 1 - \int_0^{200} \frac{1}{100}\exp\left(-\frac{1}{100}t\right)dt \\ &= 1 - [-\exp(-2) + 1] \\ &= \exp(-2) = 0.13534 . \quad \blacktriangle \end{aligned}$$

EXAMPLE 4.15

Let X be the lifetime of an electron tube and suppose that X may be represented as a continuous random variable which is exponentially distributed with parameter λ. Let $p_j = P(j \leq X < j+1)$. Prove that p_j is of the form $(1-a)a^j$ and determine a.

Solution: For $j = 0, 1, \ldots$

$$\begin{aligned} p_j &= P(j \leq X < j+1) \\ &= \int_j^{j+1} \lambda e^{-\lambda x} dx \\ &= e^{-j\lambda} - e^{-(j+1)\lambda} \\ &= e^{-j\lambda}(1 - e^{-\lambda}) \\ &= a^j(1-a) \end{aligned}$$

where $a = e^{-\lambda}$. Hence we have the result. ▲

EXAMPLE 4.16

Let X be a uniformly distributed random variable on the interval $(0,1)$. Show that $Y = -\lambda^{-1} ln(1-X)$ has an exponential distribution with parameter $\lambda > 0$.

Solution: We observe that Y is a nonnegative random variable implying $F_Y(y) = 0$ for $y \leq 0$. For $y > 0$, we have:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] = P[-\lambda^{-1} ln(1-X) \leq y] \\ &= P[ln(1-X) \geq -\lambda y] \\ &= P[(1-X) \geq e^{-\lambda y}] \quad \text{(since } e^x \text{ is an increasing function of } x\text{)} \\ &= P[X \leq 1 - e^{-\lambda y}] \\ &= F_X\left(1 - e^{-\lambda y}\right) . \end{aligned}$$

Since $X \stackrel{d}{=} \mathcal{U}[0,1]$, $F_X(x) = x$, $0 \leq x \leq 1$. Thus:

$$F_Y(y) = 1 - e^{-\lambda y},$$

so that Y is exponentially distributed with parameter λ. ▲

4.4 WEIBULL DISTRIBUTION

The Weibull distribution is widely used in engineering as a model to describe the lifetime of a component. This distribution was introduced by a Swedish scientist Weibull who proved that the effort to which the materials are subject to may be modeled through this distribution.

Suppose now that the risk function of a random variable T is given by

$$\lambda(t) = \alpha\beta t^{\beta-1}$$

where α and β are positive constants. In such a case we have that:

$$\begin{aligned} F(t) &= \begin{cases} 1 - \exp\left(-\int_0^t \alpha\beta s^{\beta-1} ds\right) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 \end{cases} \\ &= \begin{cases} 1 - \exp(-\alpha t^{\beta}) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 . \end{cases} \end{aligned}$$

The density function of T is given by:

$$f(t) = \begin{cases} \alpha\beta t^{\beta-1} \exp(-\alpha t^{\beta}) & \text{if} \quad t \geq 0 \\ 0 & \text{if} \quad t < 0 . \end{cases}$$

A random variable with the above density function receives a special name:

Definition 4.6 (Weibull Distribution) *It is said that a random variable X has Weibull distribution with parameters α and β if its density function is given by:*

$$f(x) = \begin{cases} \alpha\beta x^{\beta-1} \exp(-\alpha x^{\beta}) & \textit{if} \quad x \geq 0 \\ 0 & \textit{if} \quad x < 0 . \end{cases}$$

Notation 4.5 *The expression $X \stackrel{d}{=} W(\alpha, \beta)$ indicates that the random variable X has a Weibull distribution with parameters α and β.*

Figure 4.9 shows the graph of the Weibull distribution for $\alpha = 1$ and different values of β.

Theorem 4.5 *Let $X \stackrel{d}{=} W(\alpha, \beta)$. Then:*

1. $E(X) = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}} \Gamma(1 + \frac{1}{\beta})$.

2. $Var(X) = \left(\frac{1}{\alpha}\right)^{\frac{2}{\beta}} \left[\Gamma(1 + \frac{2}{\beta}) - \Gamma^2(1 + \frac{1}{\beta})\right]$.

Proof: *Left as an exercise.* ■

Note 4.8 *Some authors, for example, Ross (1998) and Hernández (2003), define the density function of a Weibull distribution considering three parameters, a location parameter c, a scale parameter a and a form parameter b, and they say that the random variable X has a Weibull distribution with parameters a, b and c if its density function is given by:*

$$f(x) = \begin{cases} \frac{b}{a} \left(\frac{x-c}{a}\right)^{b-1} \exp\left[-\left(\frac{x-c}{a}\right)^{b}\right] & \textit{if} \quad x > c \\ 0 & \textit{if} \quad x \leq c . \end{cases}$$

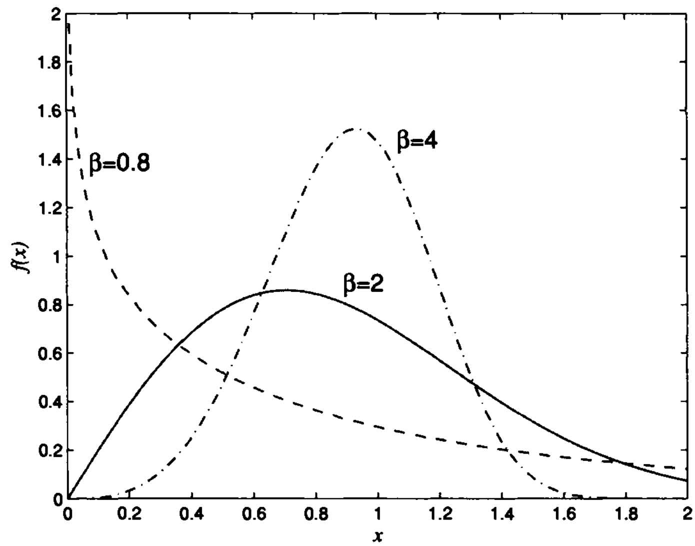


Figure 4.9 Probability density function of a Weibull distribution for $\alpha = 1$ and different values of β

However, in the majority of the applications it is a common practice to make $c = 0$, having as a result the density function considered initially by taking $\alpha = \left[\frac{1}{a}\right]^b$ and $\beta = b$.

4.5 BETA DISTRIBUTION

The distribution that will be presented here is used frequently as a mathematical model that represents physical variables whose values are restricted to an interval of finite length or as a model for fractions such as purity proportions of a chemical product or the fraction of time that takes to repair a machine.

Definition 4.7 (Beta Distribution) *It is said that the random variable X has a beta distribution with parameters $a > 0$ and $b > 0$ if its density function is given by*

$$f(x) = \frac{1}{B(a,b)} x^{a-1}(1-x)^{b-1}\mathcal{X}_{(0,1)}(x)$$

where $B(a,b)$ is the beta function. That is:

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx \ .$$

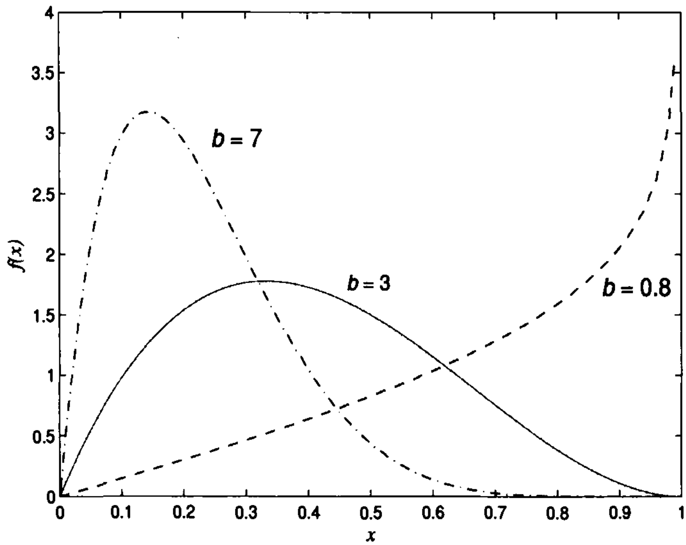


Figure 4.10 Probability density function of a beta distribution for $a = 2$ and different values of b

Notation 4.6 *The expression* $X \stackrel{d}{=} \beta(a, b)$ *means that* X *has a beta distribution with parameters* a *and* b*.*

The beta and gamma functions are related through the following expression:

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

so that the density function can be expressed in the form

$$f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1}(1-x)^{b-1}\mathcal{X}_{(0,1)}(x).$$

If a and b are positive integers, then:

$$f(x) = \frac{(a+b-1)!}{(a-1)!(b-1)!} x^{a-1}(1-x)^{b-1}\mathcal{X}_{(0,1)}(x).$$

Figure 4.10 shows the graphs of the beta density function for $a = 2$ and different values of b.

Figure 4.11 shows the graph of the beta density function with $b = 3$ and different values of a. It is clear that if $a = b = 1$, then the beta distribution coincides with the uniform distribution over the interval $(0, 1)$. In addition to this, we have that:

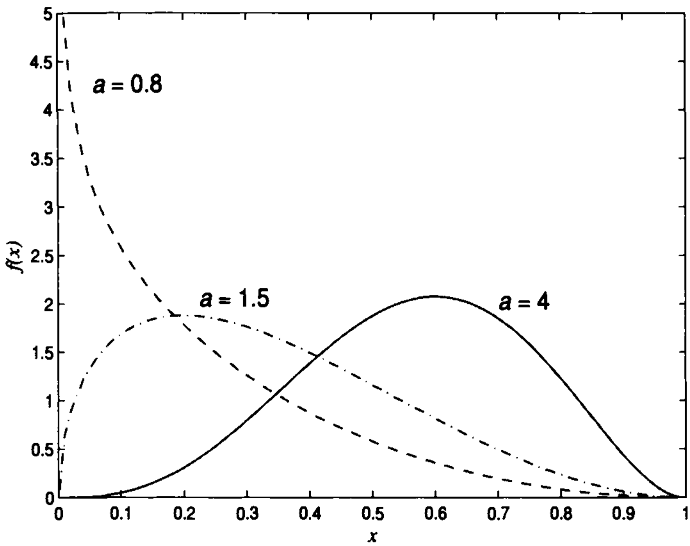


Figure 4.11 Probability density function of beta density function with $b = 3$ and different values of a

1. If $a > 1$ and $b > 1$, the function f has a global maximum.
2. If $a > 1$ and $b < 1$, the function f is an increasing function.
3. If $a < 1$ and $b > 1$, the function f is a decreasing function.
4. If $a < 1$ and $b < 1$, the graph of f has a U form.

The distribution function of the random variable with beta distribution is given by:

$$F(x) = \left[\int_0^x \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1} dt\right] \mathcal{X}_{(0,1)}(x) + \mathcal{X}_{[1,\infty)}(x) \ .$$

The moment generating function of a random variable with beta distribution does not have a simple form. Due to this, it is convenient to find its moments from the definition.

Theorem 4.6 *Let* $X \stackrel{d}{=} \beta(a,b)$. *Then:*

1. $E(X) = \frac{a}{a+b}$.

2. $Var(X) = \frac{ab}{(a+b+1)(a+b)^2}$.

Proof:

$$\begin{aligned} E(X^k) &= \frac{1}{B(a,b)} \int_0^1 x^{k+a-1}(1-x)^{b-1}dx \\ &= \frac{B(a+k,b)}{B(a,b)} \\ &= \frac{\Gamma(a+k)\Gamma(a+b)}{\Gamma(a+k+b)\Gamma(a)}. \end{aligned}$$

Therefore:

$$\begin{aligned} E(X) &= \frac{\Gamma(a+1)\Gamma(a+b)}{\Gamma(a+1+b)\Gamma(a)} \\ &= \frac{a\Gamma(a)\Gamma(a+b)}{(a+b)\Gamma(a+b)\Gamma(a)} \\ &= \frac{a}{a+b} \\ E(X^2) &= \frac{\Gamma(a+2)\Gamma(a+b)}{\Gamma(a+2+b)\Gamma(a)} \\ &= \frac{a(a+1)}{(a+b)\,(a+b+1)}. \end{aligned}$$

■

■ **EXAMPLE 4.17**

(Wackerly et al., 2008) A gas distributor has storage tanks that hold a fixed quantity of gas and that are refilled every Monday. The proportion of the storage sold during the week is very important for the distributor. Through observations done during several weeks, it was found that an appropriate model to represent the required proportion was a beta distribution with parameters $a = 4$ and $b = 2$. Find the probability that the distributor sells at least 90% of his stored gas during a given week.
Solution: Let X :="proportion of the stored gas that is sold during the week". Given that $X \stackrel{d}{=} \beta(4,2)$ we have that:

$$\begin{aligned} P(X \geq 0.9) &= 1 - P(X < 0.9) \\ &= 1 - \int_0^{0.9} 20x^3(1-x)dx \\ &= 0.08146 \ . \quad ▲ \end{aligned}$$

4.6 OTHER CONTINUOUS DISTRIBUTIONS

Definition 4.8 (Cauchy Distribution) *It is said that a random variable X has a Cauchy distribution with parameters θ and β, $\theta \in \mathbb{R}$ and $\beta \in \mathbb{R}^+$, if*

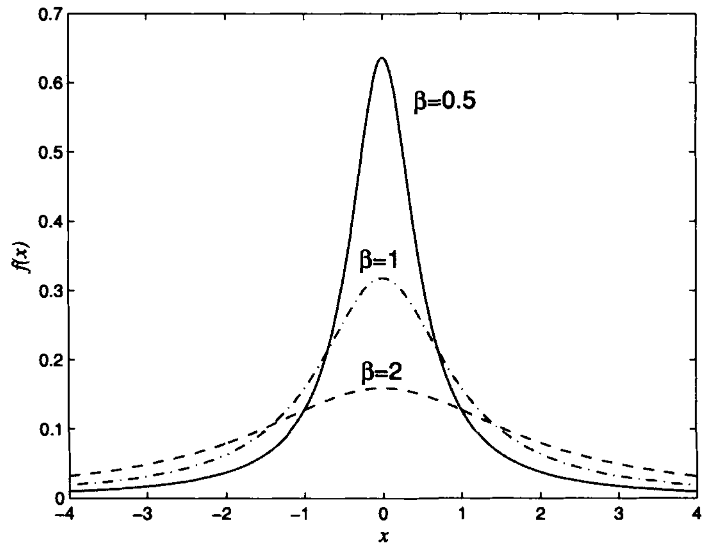


Figure 4.12 Probability density function of a Cauchy distribution for $\theta = 0$ and some values of β

its density function is given by:

$$f(x) = \frac{1}{\pi\beta} \frac{1}{1 + \left(\frac{x-\theta}{\beta}\right)^2}, \quad x \in \mathbb{R}.$$

When $\theta = 0$ and $\beta = 1$, it is obtained that

$$f(x) = \frac{1}{\pi(1 + x^2)},$$

which is known as the standardized Cauchy density function.

Figure 4.12 shows the graphs of f for $\theta = 0$ and some values of β.
Figure 4.13 shows the graphs of f for $\beta = 1$ and some values of θ.
The distribution function of the random variable with Cauchy distribution is given by:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x - \theta}{\beta}\right).$$

The Cauchy distribution has the characteristic of heavy tails. This means that the values that are farthest away from θ have high probabilities of occurrence. That is why this distribution presents atypical behavior in several ways and is

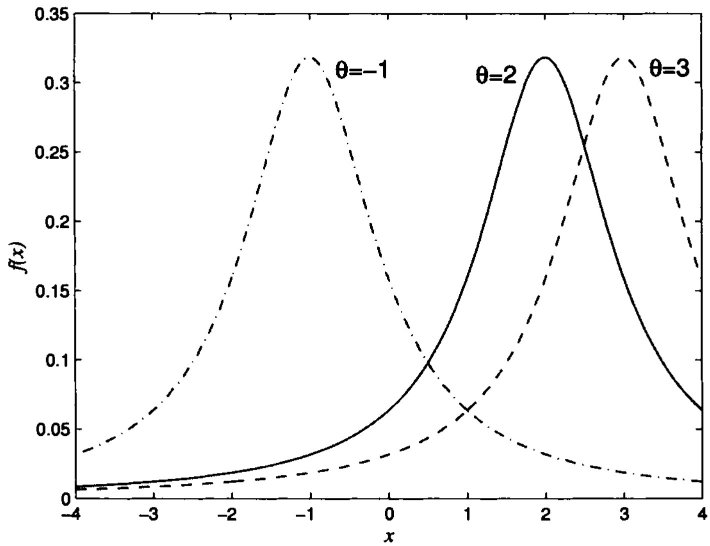


Figure 4.13 Probability density function of a Cauchy distribution for $\beta = 1$ and some values of θ

an excellent counterexample for various assertions that in the beginning might seem reasonable. Remember, for example, that in Chapter 2 it was proven that the expected value of the random variable with Cauchy distribution does not exist.

Definition 4.9 (Laplace Distribution) *It is said that a random variable X has a Laplace distribution or double exponential with parameters α, β if its density function is given by*

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\alpha|}{\beta}\right),$$

where $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^+$.

Figure 4.14 shows the graphs of a Laplace density function with $\alpha = 0$ and different values of β.

Definition 4.10 (Exponential Power) *It is said that a random variable X distributes with an exponential power with parameters α, β and γ with $\alpha \in \mathbb{R}$ and $\beta, \gamma \in \mathbb{R}^+$ if its density function is given by:*

$$f(x) = \frac{1}{2\beta\Gamma\left(1+\frac{1}{\gamma}\right)} \exp\left[-\left|\frac{x-\alpha}{\beta}\right|^{\gamma}\right].$$

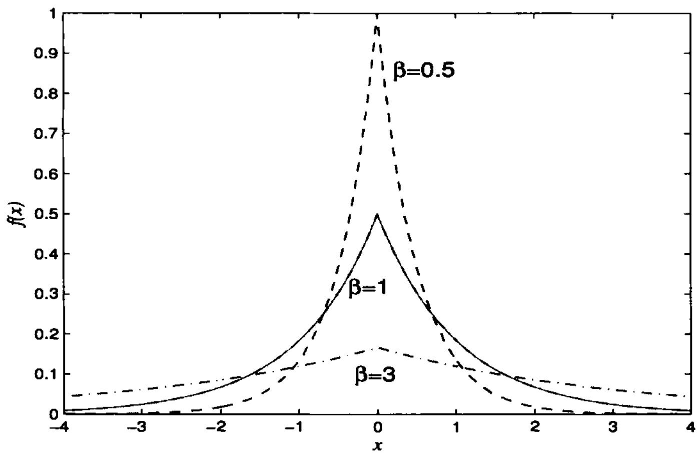


Figure 4.14 Probability density function of a Laplace density function with $\alpha = 0$ and different values of β

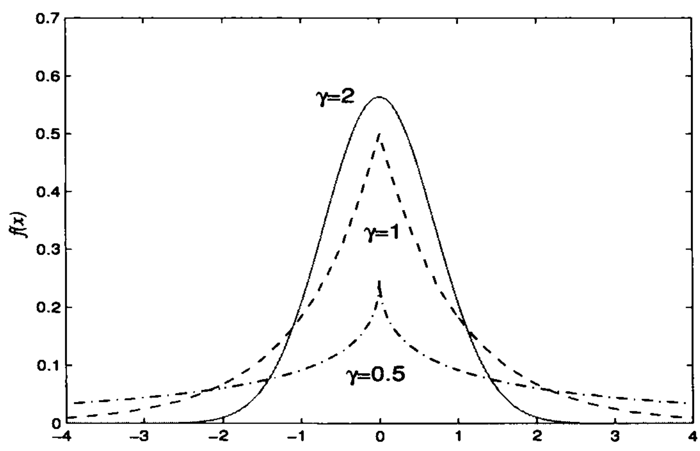


Figure 4.15 Probability density function of an exponential power distribution for $\alpha = 0$, $\beta = 1$ and different values of γ

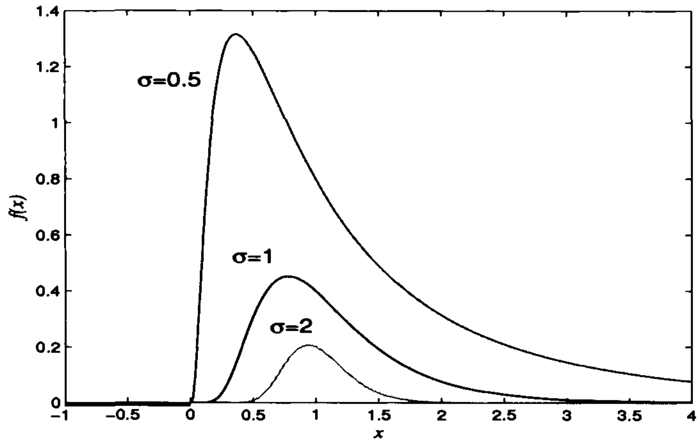


Figure 4.16 Probability density function of a lognormal distribution for $\mu = 0$ and different values of σ

Figure 4.15 shows the graphs of a density function of a random variable with exponential power distribution for $\alpha = 0$, $\beta = 1$ and different values of γ. If $\gamma = 2$, then

$$f(x) = \frac{1}{\sqrt{\pi}\beta} \exp\left[-\left(\frac{x-\alpha}{\beta}\right)^2\right],$$

that is, we obtain the density function of a random variable with normal distribution with parameters $\mu = \alpha$ and $\sigma = \frac{\beta}{\sqrt{2}}$. If $\gamma = 1$ we have that

$$f(x) = \frac{1}{2\beta} \exp\left(-\frac{|x-\alpha|}{\beta}\right),$$

which is the Laplace density function with parameters α and β.

Definition 4.11 (Lognormal Distribution) *Let X be a nonnegative variable and $Y := \ln X$. If the random variable Y has a normal distribution with parameters μ and σ, then it is said that X has a lognormal distribution with parameters μ and σ.*

It is clear that if X has lognormal distribution, its density function is given by:

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma x} \exp\left[-\left(\frac{\ln x - \mu}{\sqrt{2}\sigma}\right)^2\right] & \text{if} \quad x > 0 \\ 0 & \text{otherwise} . \end{cases}$$

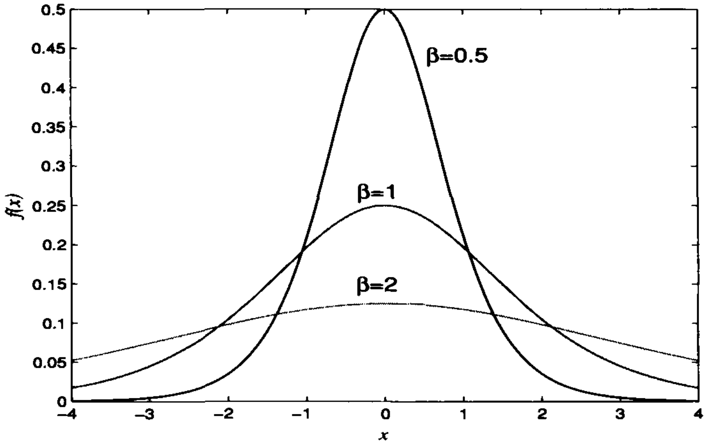


Figure 4.17 Probability density function of a logistic distribution for $\alpha = 0$ and different values of β

Figure 4.16 shows the graph of a lognormal density function for $\mu = 0$ and different values of σ.

Definition 4.12 (Logistic Distribution) *It is said that a random variable X has a logistic distribution with parameters α and β with $\alpha \in \mathbb{R}$ and $\beta \in \mathbb{R}^+$ if its density function is given by:*

$$f(x) = \frac{1}{\beta} \frac{\exp\left[-\left(\frac{x-\alpha}{\beta}\right)\right]}{\left[1 + \exp\left[-\left(\frac{x-\alpha}{\beta}\right)\right]\right]^2}, \quad x \in \mathbb{R}.$$

The distribution function of the random variable with logistic distribution of parameters α and β is given by:

$$F(x) = \frac{1}{1 + \exp\left[-\left(\frac{x-\alpha}{\beta}\right)\right]}.$$

Figure 4.17 shows the graph of the logistic density function for $\alpha = 0$ and different values of β.

In this chapter we have seen some important distributions of continuous random variables which are frequently used in applications. To conclude this chapter, we now present the relation between various continuous distributions which is illustrated in Figure 4.18.

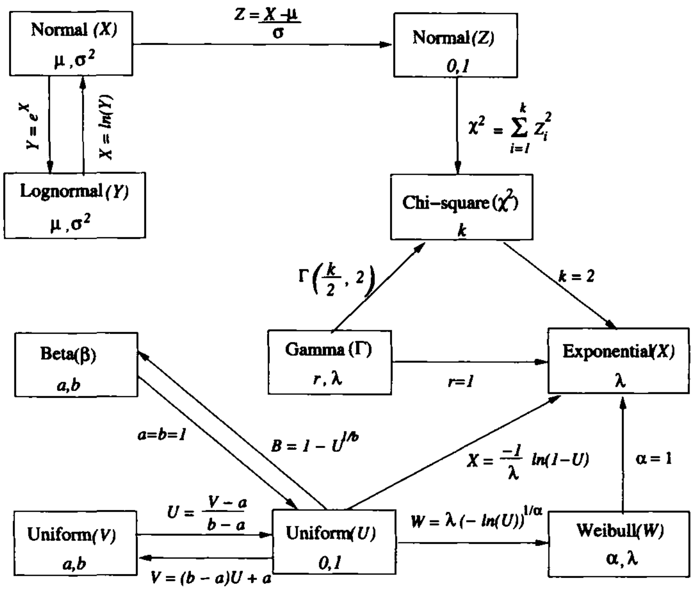


Figure 4.18 Relationship between distributions

EXERCISES

4.1 Let X be a random variable with continuous uniform distribution in the interval $\left[-\frac{1}{2}, \frac{3}{2}\right]$.

a) Calculate: mean, variance and standard deviation of X.

b) Determine the value of x so that $P(|X| < x) = 0.9$.

4.2 A number is randomly chosen in the interval $(0, 1)$. Calculate:

a) The probability that the first digit to the right of the decimal point is 6.

b) The probability that the second digit to the right of the decimal point is 1.

c) The probability that the second digit to the right of the decimal point is 8 given that the first digit was 3.

4.3 Let $X \stackrel{d}{=} \mathcal{U}(a, b)$. If $E(X) = 2$ and $Var(X) = \frac{3}{4}$, which are the values of the parameters a and b?

4.4 A student arrives at the bus station at 6:00 AM sharp knowing that the bus will arrive any moment, uniformly distributed between 6:00 AM and

6:20 AM. What is the probability that the student must wait more than 5 minutes? If at 6:10 AM the bus has not arrived yet, what is the probability that the student has to wait at least 5 more minutes?

4.5 A bus on line A arrives at a bus station every 4 minutes and a bus on line B every 6 minutes. The time interval between an arrival of a bus for line A and a bus for line B is uniformly distributed between 0 and 4 minutes. Find the probability:

a) That the first bus that arrives will be for line A.

b) That a bus will arrive within 2 minutes (for line A or B).

4.6 Assume that N stars are randomly scattered, independently of each other, in a sphere of radius R (measured in parsecs).

a) What is the probability that the star nearest to the center is at a distance at least r?

b) Find the limit of the probability in (a) if $R \to \infty$ and $N/R^3 \to 4\pi\lambda/3$ where $\lambda \simeq 0.0063$.

4.7 Consider a random experiment of choosing a point in an annular disc of inner radius r_1 and outer radius r_2 ($r_1 < r_2$). Let X be the distance of a chosen point from the annular disc to the center of the annular disc. Find the *pdf* of X.

4.8 Let X be an exponentially distributed random variable with parameter b. Let Y be defined by $Y = i$ for $i = 0, 1, 2, \cdots$ whenever $i \leq X < i + 1$. Find the distribution of Y. Also obtain the variance of Y if it exists.

4.9 Let X be a uniformly distributed random variable on the interval $(0, 1)$. Define:

$$Y = a + (b - a)X, \quad a < b.$$

Find the distribution of Y.

4.10 Suppose that X is a random variable with uniform distribution over the interval $(0, 4)$. Calculate the probability that the roots of the equation $2x^2 + 2xX + X + 1 = 0$ are both complex.

4.11 Let X be a random variable with uniform distribution over $(-2, 2)$. Find:

a) $P\left(|X| < \frac{1}{3}\right)$.

b) A density function of the random variable $Y = |X|$.

4.12 Let X be a random variable with uniform distribution over $(0, 1)$. Find the density functions for the following random variables:

a) $Y := \ln X$.

b) $Z := X^3 + 2$.

c) $W := \frac{1}{X}$.

4.13 A player throws a dart at a dartboard. Suppose that the player receives 10 points for his throw if it hits 2 cm from the center, 5 points if it lands between 2 and 6 cm from the center and 3 points if it hits between 6 and 10 cm from the center. Find the expected number of points obtained by the player knowing that the distance from the place where the dart hits and the center of the board is a random variable with uniform distribution in $(0, 10)$.

4.14 Let $X \stackrel{d}{=} \mathcal{N}(0,1)$. Calculate:

a) $P(-1 < X \leq 1.2)$.

b) $P(-0.34 < X < 0)$.

c) $P(-2.32 < X < 2.4)$.

d) $P(X \geq 1.43)$.

e) $P(|X - 1| \leq 0.5)$.

4.15 Let $X \stackrel{d}{=} \mathcal{N}(3, 16)$. In each of the following exercises, obtain the value of x that solves the equation:

a) $P(X > x) = 0.5$.

b) $P(x < X < 5) = 0.1$.

c) $P(X \geq x) = 0.01$.

4.16 In the following exercises, find the value of c that will satisfy the equalities.

a) $P(W \leq c) = 0.95$ if $W \stackrel{d}{=} \mathcal{X}^2_{10}$.

b) $P(|X - 1| \leq 3) = c$ if $X \stackrel{d}{=} \mathcal{N}(1, 9)$.

c) $P(W > c) = 0.25$ if $W \stackrel{d}{=} \mathcal{X}^2_5$.

d) $P(X < c) = 0.990$ if $X \stackrel{d}{=} \mathcal{X}^2_{25}$.

e) $P\left(\mathcal{X}^2_k \leq c\right) = 0.005$ for $k = 3, 7, 27$.

4.17 Suppose that the marks on an examination are distributed normally with mean 76 and standard deviation 15. Of the best students 15% obtained A as grade and of the worst students 10% lost the course and obtained P.

a) Find the minimum mark to obtain A as a grade.

b) Find the minimum mark to pass the test.

4.18 The lifetime of a printer is a normal random variable with mean 5.2 years and standard deviation 1.4 years. What percentage of the printers have a lifetime less than 7 years? Less than 3 years? Between 3 and 7 years?

4.19 Let X be a random variable with normal distribution with mean μ and variance σ^2. Find the distribution of the random variable $Y := 5X - 1$.

4.20 Suppose that the lifespan of a certain type of lamp is a random variable having a normal distribution with mean 180 hours and standard deviation 20 hours. A random sample of four lamps is taken.

a) What is the probability that all four lamps have a lifespan greater than 200 hours?

b) All four lamps of the random sample are placed inside an urn. If one lamp is then randomly selected, what is the probability that the extracted lamp has a lifespan greater than 200 hours?

4.21 Let X be a random variable with binomial distribution with parameters $n = 30$ and $p = 0.3$. Is it reasonable to approximate this distribution to a normal with parameter $\mu = 9$ and variance $\sigma^2 = 6.3$? Explain.

4.22 A fair die is tossed 1000 consecutive times. Calculate the probability that the number 3 is obtained less than 500 times given that the number 1 was obtained exactly 200 times.

4.23 Let $X \stackrel{d}{=} \mathcal{N}(3, 4)$. Find the number α so that:

$$P(X > \alpha) = \frac{1}{2} P(X \le \alpha).$$

4.24 Determine the tenths of the standard normal distribution, that is, the values $x_{0.1}, x_{0.2}, \cdots, x_{0.9}$, so that $\Phi(x_{0.i}) = 0.i$ for $i = 1, \cdots, 9$.

4.25 Consider a nonlinear amplifier whose input X and output Y are related by its transfer characteristic:

$$Y = \begin{cases} X^{\frac{1}{2}} & \text{if } X > 0 \\ -|X|^{\frac{1}{2}} & \text{if } X < 0 . \end{cases}$$

Find the *pdf* of Y if X has $\mathcal{N}(0,1)$ distribution.

4.26 Let X be a continuous random variable with distribution function F and *pdf* $f(x)$. The truncated distribution of X to the left at $X = \alpha$ and to the right at $X = \beta$ is defined as:

$$g(x) = \begin{cases} \frac{f(x)}{F(\beta)-F(\alpha)} & \text{if } \alpha < x \leq \beta \\ 0 & \text{otherwise}\ . \end{cases}$$

Find the *pdf* of a truncated normal $\mathcal{N}(\mu, \sigma^2)$ random variable truncated to the left at $X = \alpha$ and to the right at $X = \beta$.

4.27 A marketing study determined that the daily demand for a recognized newspaper is a random variable with normal distribution with mean $\mu = 50,000$ and standard deviation $\sigma = 12,500$. Each newspaper sold leaves 500 Colombian pesos as revenue, while each paper not sold gives 300 Colombian pesos as loss. How many newspapers are required in order to produce a maximum expected revenue?

4.28 Let X be a random variable with:

a) Uniform distribution over $[-1, 1]$.

b) Exponential distribution with parameter λ.

c) Normal distribution with parameters μ and σ.

Calculate the distribution function F_Y and the density function f_Y of the random variable $Y = aX + b$, where a and b are real numbers and $a \neq 0$.

4.29 In the claim office of a public service enterprise, it is known that the time (in minutes) that the employee takes to take a claim from a user is a random variable with exponential distribution with mean 15 minutes. If you arrive at 12 sharp to the claim office and in that moment there is no queue but the employee is taking a claim from a client, what is the probability that you must wait for less than 5 minutes to talk to the employee?

4.30 Suppose that the number of kilometers that an automobile travels before its battery runs out is distributed exponentially with a mean value of 10,000 km. If a person wants to travel 5,000km, what is the probability that the person finishes his trip without having to change the battery?

4.31 The time that has elapsed between the calls to an office has an exponential distribution with mean time between calls of 15 minutes.

a) What is the probability that no calls have been received in a 30-minute period of time?

b) What is the probability of receiving at least one call in the interval of 10 minutes?

c) What is the probability of receiving the first call between 5 and 10 minutes after opening the office?

4.32 The length of time T of an electronic component is a random variable with exponential distribution with parameter λ.

a) Determine the probability that the electronic component works for at least until $t = 3\lambda^{-1}$.

b) What is the probability that the electronic component works for at least until $t = k\lambda^{-1}$ if it work's until time $t = (k-1)\lambda^{-1}$?

4.33 Let X be a random variable having an exponential distribution with parameter $\lambda = \frac{1}{3}$. Compute:

a) $P(X > 3)$.

b) $P(X > 6 \mid X > 3)$.

c) $P(X > t + 3 \mid X > t)$.

4.34 Let X be a random variable with exponential distribution with parameter λ. Find the density functions for the following random variables:

a) $Y := \ln X$.

b) $Z := X^2 + 1$.

c) $W := \frac{1}{X}$.

4.35

a) In 1825 Gompertz proposed the risk function $\lambda(t)$ given by

$$\lambda(t) = \alpha\beta^t \mathcal{X}_{(0,\infty)}(t) \text{ with } \alpha > 0 \text{ and } \beta > 1$$

to model the lifetime of human beings. Determine the distribution function F corresponding to the failure function $\lambda(t)$.

b) In the year 1860 Makeham modified the risk function proposed by Gompertz and suggested the following:

$$\lambda(t) = \left(\gamma + \alpha\beta^t\right) \mathcal{X}_{(0,\infty)}(t) \quad \text{with } \alpha, \beta \in \mathbb{R}^+ \text{ and } \beta > 1.$$

Determine the distribution function F corresponding to the failure function $\lambda(t)$ (this distribution is also known as the Makeham distribution).

4.36 Calculate the failure function of the exponential distribution with parameter μ.

4.37 Determine the distribution of F whose failure distribution is given by

$$\lambda(t) = \alpha + \beta t, t \geq 0,$$

where α and β are constants.

4.38 Suppose that T denotes the lifetime of a certain component and that the risk function associated with T is given by:

$$\lambda(t) = \frac{1}{(t+1)}\ .$$

Find the distribution of T.

4.39 Suppose that T denotes the lifetime of a certain component and that the risk function associated with T is given by

$$\lambda(t) = \alpha t^{\beta}$$

where $\alpha > 0$ and $\beta > 0$ are constants. Find the distribution of T.

4.40 Let X be a Weibull distribution. Prove that $Y = bX$ for some $b > 0$ has the exponential distribution.

4.41 The lifetime of a certain electronic component has a Weibull distribution with $\beta = 0.5$ and a mean life of 600 hours. Calculate the probability that the component lasts at least 500 hours.

4.42 The time elapsed (in months after maintenance) before a fail in a vigilance equipment with closed circuit in a beauty shop has a Weibull distribution with $\alpha = \frac{1}{75}$ and $\beta = 2.2$. If the shop wants to have a probability of damage before the next programmed maintenance of 0.04, then what is the time elapsed of the equipment to receive maintenance?

4.43 A certain device has the Weibull failure rate

$$r(t) = \lambda p t^{p-1}, \quad t > 0.$$

a) Find the reliability $R(t)$.

b) Find the mean time to failure.

c) Find the density function $f_T(t)$.

4.44 A continuous random variable X is said to have Pareto distribution if it has density function

$$f(x;\alpha) = \frac{\theta}{x_0}\left(\frac{x_0}{x}\right)^{\theta+1}, \quad x > x_0, \quad \theta > 0.$$

The Pareto distribution is commonly used in economics.

a) Find the mean and variance of the distribution.

b) Determine the density function of $Z = lnX$.

4.45 A certain electrical component of a mobile phone has the Pareto failure rate

$$r(t) = \begin{cases} \frac{a}{t} & \text{if } t \geq t_0 \\ 0 & \text{otherwise.} \end{cases} \tag{4.3}$$

a) Find the reliability $R(t)$ for $t > 0$.

b) Sketch $R(t)$ for $t_0 = 1$ and $a = 2$.

c) Find the mean time to failure if $a > 0$.

4.46 Let X be a random variable with standard normal distribution. Find $E(|X|)$.

4.47 Consider Note 4.1. Let X be a continuous random variable with strictly increasing distribution function F. Let $Y = F_X(X)$. Prove that Y has uniform distribution over $(0, 1)$.

4.48 Let X be a random variable with uniform distribution over $(0, 1)$. Find the function $g : \mathbb{R} \longrightarrow \mathbb{R}$, so that $Y = g(X)$ has a standard normal distribution.

4.49 Prove that if X is a random variable with Erlang distribution with parameters r and λ, then:

$$F_X(x) = 1 - \sum_{j=0}^{r-1} \frac{(\lambda x)^j \exp(-\lambda x)}{j!} .$$

4.50 Prove that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$.

4.51 Prove that:

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

4.52 Let X be a random variable with standard Cauchy distribution. Prove that $E(X)$ does not exist.

4.53 Let X be a random variable with standard Cauchy distribution. What type of distribution has the random variable $Y := \frac{1}{X}$?

4.54 Let X be a normal random variable with parameters 0 and σ^2. Find a density function for:

a) $Y = |X|$.

b) $Y = \sqrt{|X|}$.

4.55 Let X be a continuous random variable with *pdf*

$$f_X(x) = \frac{1}{\pi(1+x^2)}, \quad -\infty < x < \infty .$$

Find the distribution of:

$$Y = \begin{cases} X, & |X| \geq 2 \\ 0, & |X| < 2 \,. \end{cases}$$

4.56 Let X be a random variable with standard normal distribution. Prove that

$$\left(x^{-1} - x^{-3}\right) \exp\left(-\frac{1}{2}x^2\right) < \sqrt{2\pi}\,[P(X > x)] < x^{-1} \exp\left(-\frac{1}{2}x^2\right)$$

for $x > 0$.

CHAPTER 5

RANDOM VECTORS

In the previous chapters, we have been concentrating on a random variable defined in an experiment, e.g., the number of heads obtained when three coins are tossed. In other words, we have been discussing the random variables but one at a time. In this chapter, we learn how to treat two random variables, of both the discrete as well as the continuous types, simultaneously in order to understand the interdependence between them. For instance, in a telephone exchange, the time of a call arrival X and a call duration Y or the shock arrivals X and the consequent damage Y are of interest. One of the questions which also comes to mind is, can we relate two random variables defined on the same sample space, or in other words, can two random variables of the same spaces be correlated? Some of these questions will be discussed and answered in this chapter.

5.1 JOINT DISTRIBUTION OF RANDOM VARIABLES

In many cases it is necessary to consider the joint behavior of two or more random variables. Suppose, for example, that a fair coin is flipped three

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

consecutive times and we wish to analyze the joint behavior of the random variables X and Y defined as follows:

X := "Number of heads obtained in the first two flips".

Y := "Number of heads obtained in the last two flips".

Clearly:

$$P(X=0,Y=0) = P((T,T,T)) = \frac{1}{8}$$
$$P(X=0,Y=1) = P((T,T,H)) = \frac{1}{8}$$
$$P(X=1,Y=0) = P((H,T,T)) = \frac{1}{8}$$
$$P(X=1,Y=1) = P(\{(H,T,H),(T,H,T)\}) = \frac{1}{4}$$
$$P(X=1,Y=2) = P((T,H,H)) = \frac{1}{8}$$
$$P(X=2,Y=1) = P((H,H,T)) = \frac{1}{8}$$
$$P(X=2,Y=2) = P((H,H,H)) = \frac{1}{8}.$$

This information can be summarized in the following table:

$X\backslash Y$	0	1	2
0	$\frac{1}{8}$	$\frac{1}{8}$	0
1	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
2	0	$\frac{1}{8}$	$\frac{1}{8}$

Definition 5.1 (n-Dimensional Random Vector) *Let $X_1, X_2, \cdots, X_n$ be n real random variables defined over the same probability space $(\Omega, \Im, P)$. The function $\mathbf{X} : \Omega \longrightarrow \mathbb{R}^n$ defined by*

$$\mathbf{X}(\omega) := (X_1(\omega), \cdots, X_n(\omega))$$

is called an n-dimensional random vector.

Definition 5.2 (Distribution of a Random Vector) *Let $\mathbf{X}$ be an n-dimensional random vector. The probability measure defined by*

$$P_{\mathbf{X}}(B) := P(\mathbf{X} \in B) \ ; \ B \in \mathcal{B}_n$$

is called the distribution of the random vector $\mathbf{X}$.

Definition 5.3 (Joint Probability Mass Function) *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be an n-dimensional random vector. If the random variables X_i, with $i = 1, \cdots, n$, are all discrete, it is said that the random vector $\mathbf{X}$ is discrete. In this case, the probability mass function of $\mathbf{X}$, also called the joint distribution function of the random variables $X_1, X_2, \cdots, X_n$, is defined by:*

$$p_{\mathbf{X}}(\mathbf{x}) := \begin{cases} P(\mathbf{X} = \mathbf{x}) & \textit{if } \ \mathbf{x} \textit{ belongs to the image of } \mathbf{X} \\ 0 & \textit{otherwise} . \end{cases}$$

Note 5.1 *Let X_1 and X_2 be discrete random variables. Then:*

$$\begin{aligned} P(X_1 = x) &= P\left((X_1 = x) \cap \bigcup_y (X_2 = y) \right) \\ &= P\left(\bigcup_y (X_1 = x, X_2 = y) \right) \\ &= \sum_y P(X_1 = x, X_2 = y). \end{aligned}$$

In general, we have:

Theorem 5.1 *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a discrete n-dimensional random vector. Then, for all $j = 1, \cdots, n$ we have:*

$$P(X_j = x) = \sum_{x_1} \cdots \sum_{x_{j-1}} \sum_{x_{j+1}} \cdots \sum_{x_n} P(X_1 = x_1, \cdots, X_{j-1} = x_{j-1}, X_j = x, X_{j+1} = x_{j+1}, \cdots, X_n = x_n).$$

The function

$$p_{X_j}(x) := \begin{cases} P(X_j = x) & \textit{if } \ x \textit{ belongs to the image of } X_j \\ 0 & \textit{otherwise} \end{cases}$$

is called the marginal distribution of the random variable X_j.

■ EXAMPLE 5.1

Let X and Y be discrete random variables with joint distribution given by:

$X \backslash Y$	0	1	2	3
-1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{1}{16}$	$\frac{5}{16}$
1	$\frac{1}{16}$	$\frac{3}{16}$	$\frac{2}{16}$	0

The marginal distributions of X and Y are given, respectively, by:

x	-1	1
$P(X=x)$	$\frac{10}{16}$	$\frac{6}{16}$

y	0	1	2	3
$P(Y=y)$	$\frac{2}{16}$	$\frac{6}{16}$	$\frac{3}{16}$	$\frac{5}{16}$

▲

■ EXAMPLE 5.2

Suppose that a fair coin is flipped three consecutive times and let X and Y be the random variables defined as follows:

$X :=$ "Number of heads obtained".

$Y :=$ "Flip number where a head was first obtained" (if there are none, we define $Y = 0$).

1. Find the joint distribution of X and Y.
2. Calculate the marginal distributions of X and Y.
3. Calculate $P(X \leq 2, Y = 1)$, $P(X \leq 2, Y \leq 1)$ and $P(X \leq 2 \text{ or } Y \leq 1)$.

Solution:

1. The joint distribution of X and Y is given by:

$X\backslash Y$	0	1	2	3
0	$\frac{1}{8}$	0	0	0
1	0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
2	0	$\frac{2}{8}$	$\frac{1}{8}$	0
3	0	$\frac{1}{8}$	0	0

2. The marginal distributions of X and Y are presented in the following tables:

x	0	1	2	3
$P(X=x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

y	0	1	2	3
$P(Y=y)$	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{8}$

3. It follows from the previous items that:

$$\begin{aligned}
P(X \le 2, Y = 1) &= P(X = 0, Y = 1) + P(X = 1, Y = 1) \\
&\quad + P(X = 2, Y = 1) \\
&= \frac{3}{8} \\
P(X \le 2, Y \le 1) &= P(X \le 2, Y = 1) + P(X \le 2, Y = 0) \\
&= \frac{3}{8} + P(X = 0, Y = 0) + P(X = 1, Y = 0) \\
&\quad + P(X = 2, Y = 0) \\
&= \frac{1}{2} \\
P(X \le 2 \text{ or } Y \le 1) &= P(X \le 2) + P(Y \le 1) - P(X \le 2, Y \le 1) \\
&= P(X = 0) + P(X = 1) + P(X = 2) \\
&\quad + P(Y = 0) + P(Y = 1) - \frac{1}{2} \\
&= \frac{7}{8} + \frac{5}{8} - \frac{1}{2} = 1. \quad \blacktriangle
\end{aligned}$$

■ EXAMPLE 5.3

A box contains three nails, four drawing-pins and two screws. Three objects are randomly extracted without replacement. Let X and Y be the number of drawing-pins and nails, respectively, in the sample. Find the joint distribution of X and Y.

Solution: Since

$$P(X = x, Y = y) = \frac{\binom{4}{x}\binom{3}{y}\binom{2}{3-x-y}}{\binom{9}{3}} \quad \text{for } x, y = 0, 1, 2, 3,$$

the joint distribution of the variables X and Y is given by:

$X\backslash Y$	0	1	2	3
0	0	$\frac{3}{84}$	$\frac{6}{84}$	$\frac{1}{84}$
1	$\frac{4}{84}$	$\frac{24}{84}$	$\frac{12}{84}$	0
2	$\frac{12}{84}$	$\frac{18}{84}$	0	0
3	$\frac{4}{84}$	0	0	0

The marginal distributions of X and Y are, respectively:

x	0	1	2	3
$P(X = x)$	$\frac{10}{84}$	$\frac{40}{84}$	$\frac{30}{84}$	$\frac{4}{84}$

y	0	1	2	3
$P(Y = y)$	$\frac{20}{84}$	$\frac{45}{84}$	$\frac{18}{84}$	$\frac{1}{84}$

▲

■ EXAMPLE 5.4

A fair dice is rolled twice in a row. Let:

$X_1 :=$ "Greatest value obtained".
$X_2 :=$ "Sum of the results obtained".

The density function of the random vector $\mathbf{X} = (X_1, X_2)$ is given by:

$X_1\backslash X_2$	2	3	4	5	6	7	8	9	10	11	12
1	$\frac{1}{36}$	0	0	0	0	0	0	0	0	0	0
2	0	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0	0	0
3	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0	0	0
4	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0	0	0
5	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$	0	0
6	0	0	0	0	0	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

The marginal distributions of X_1 and X_2 are, respectively:

x_1	1	2	3	4	5	6
$P(X_1 = x_1)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

x_2	2	3	4	5	6	7	8	9	10	11	12
$P(X_2 = x_2)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

Furthermore, we have that

$$\begin{aligned}
P\left(X_1 \le \frac{3}{2}, X_2 \le \frac{11}{3}\right) &= P\left(X_1 = 1, X_2 = 2\right) + P\left(X_1 = 1, X_2 = 3\right) \\
&= \frac{1}{36} \\
P\left(X_1 \le \pi, X_2 \le 2\right) &= P\left(X_1 = 1, X_2 = 2\right) + P\left(X_1 = 2, X_2 = 2\right) \\
&\quad + P\left(X_1 = 3, X_2 = 2\right) \\
&= \frac{1}{36}
\end{aligned}$$

and in general:

$X_2 \backslash X_1$	$(-\infty, 1)$	$[1, 2)$	$[2,3)$	$[3,4)$	$[4,5)$	$[5,6)$	$[6, \infty)$
$(-\infty, 2)$	0	0	0	0	0	0	0
$[2, 3)$	0	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
$[3, 4)$	0	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{3}{36}$	$\frac{3}{36}$
$[4, 5)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$	$\frac{6}{36}$
$[5, 6)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{8}{36}$	$\frac{10}{36}$	$\frac{10}{36}$	$\frac{10}{36}$
$[6,7)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{13}{36}$	$\frac{15}{36}$	$\frac{15}{36}$
$[7, 8)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{15}{36}$	$\frac{19}{36}$	$\frac{21}{36}$
$[8,9)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{22}{36}$	$\frac{26}{36}$
$[9,10)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{24}{36}$	$\frac{30}{36}$
$[10, 11)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{33}{36}$
$[11, 12)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	$\frac{35}{36}$
$[12, \infty)$	0	$\frac{1}{36}$	$\frac{4}{36}$	$\frac{9}{36}$	$\frac{16}{36}$	$\frac{25}{36}$	1

Each entry in the table represents the probability that

$$P(X_1 \leq x_1, X_2 \leq x_2)$$

for the values x_1 and x_2 indicated in the first row and column, respectively. ▲

The previous example leads us to the following definition:

Definition 5.4 (Joint Cumulative Distribution Function) *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ *be an n-dimensional random vector. The function defined by*

$$F(x_1, x_2, \cdots, x_n) := P(X_1 \leq x_1, X_2 \leq x_2, \cdots, X_n \leq x_n),$$

for all $(x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$ *is called the joint cumulative distribution function of the random variables* $X_1, X_2, \cdots, X_n$*, or simply the distribution function of the n-dimensional random vector* $\mathbf{X}$.

Note 5.2 *Just like the one-dimensional case, we have that the distribution of the random vector* $\mathbf{X}$ *is completely determined by its distribution function.*

Note 5.3 *Let* X_1 *and* X_2 *be random variables with joint cumulative distribution function* F*. Then:*

$$\begin{aligned} F_{X_1}(x) &= P(X_1 \leq x) \\ &= P\left((X_1 \leq x) \cap \bigcup_y (X_2 \leq y)\right) \\ &= P\left(\bigcup_y (X_1 \leq x, X_2 \leq y)\right) \\ &= \lim_{y \to \infty} P(X_1 \leq x, X_2 \leq y) \\ &= \lim_{y \to \infty} F(x, y). \end{aligned}$$

Likewise, we have that $F_{X_2}(y) = \lim\limits_{x \to \infty} F(x, y)$*. This can be generalized in the following theorem:*

Theorem 5.2 *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ *be an n-dimensional random vector with joint cumulative distribution function* F*. For each* $j = 1, \cdots, n$*, the cumulative distribution function of the random variable* X_j *is given by:*

$$F_{X_j}(x) = \lim_{x_1 \to \infty} \cdots \lim_{x_{j-1} \to \infty} \lim_{x_{j+1} \to \infty} \cdots \lim_{x_n \to \infty} F(x_1, \cdots, x_n) \ .$$

The distribution function F_{X_j} *is called the marginal cumulative distribution function of the random variable* X_j.

The previous theorem shows that if the cumulative distribution function of the random variables $X_1, \cdots, X_n$, is known, then the marginal distributions are also known. The converse however does not always hold.

Next we present some of the properties of the joint distribution function.

Theorem 5.3 *Let* $\mathbf{X} = (X, Y)$ *be a two-dimensional random vector. The joint cumulative distribution function* F *of the random variables* X, Y *has the following properties:*

1.

$$\Delta_a^b F := F(b_1, b_2) + F(a_1, a_2) - F(a_1, b_2) - F(b_1, a_2) \geq 0$$

where $a = (a_1, a_2)$, $b = (b_1, b_2) \in \mathbb{R}^2$ *with* $a_1 \leq b_1$ *and* $a_2 \leq b_2$.

2.

$$\lim_{x \searrow x_0} F(x,y) = F(x_0,y) \tag{5.1}$$

$$\lim_{y \searrow y_0} F(x,y) = F(x,y_0). \tag{5.2}$$

3.

$$\lim_{x \to -\infty} F(x,y) = 0 \quad \textit{and} \quad \lim_{y \to -\infty} F(x,y) = 0.$$

4.

$$\lim_{(x,y) \to (\infty,\infty)} F(x,y) = 1.$$

Proof:

1. Let $a = (a_1, a_2)$, $b = (b_1, b_2)$ with $a_1 < b_1$, $a_2 < b_2$ and:

$$\begin{aligned}
A &:= \{(x,y) \in \mathbb{R}^2 : x \leq b_1, y \leq b_2\} \\
B &:= \{(x,y) \in \mathbb{R}^2 : x \leq a_1, y \leq a_2\} \\
C &:= \{(x,y) \in \mathbb{R}^2 : x \leq a_1, y \leq b_2\} \\
D &:= \{(x,y) \in \mathbb{R}^2 : x \leq b_1, y \leq a_2\}.
\end{aligned}$$

If $I := (A - C) - (D - B)$, then, clearly:

$$\begin{aligned}
0 \leq P_{\mathbf{X}}(I) &= (P_{\mathbf{X}}(A) - P_{\mathbf{X}}(C)) - (P_{\mathbf{X}}(D) - P_{\mathbf{X}}(B)) \\
&= P_{\mathbf{X}}(A) + P_{\mathbf{X}}(B) - P_{\mathbf{X}}(C) - P_{\mathbf{X}}(D) \\
&= F(b_1,b_2) + F(a_1,a_2) - F(a_1,b_2) - F(b_1,a_2) \\
&= \Delta_a^b F.
\end{aligned}$$

2. We prove (5.1) and leave (5.2) as an exercise for the reader:

$$\begin{aligned}
\lim_{x \searrow x_0} F(x,y) &= \lim_{x \searrow x_0} P\left(X \leq x, Y \leq y\right) \\
&= P\left(\lim_{x \searrow x_0} (X \leq x, Y \leq y)\right) \\
&= P(X \leq x_0, Y \leq y) \\
&= F(x_0, y).
\end{aligned}$$

3. Since

$$\begin{aligned}
\lim_{x \searrow -\infty} [X \leq x, Y \leq y] &= \lim_{x \searrow -\infty} ([X \leq x] \cap [Y \leq y]) \\
&= \emptyset \cap [Y \leq y] = \emptyset,
\end{aligned}$$

we have:

$$\lim_{x \searrow -\infty} P\left[X \le x, Y \le y\right] = P\left(\lim_{x \searrow -\infty}\left[X \le x, Y \le y\right]\right)$$
$$= P(\emptyset) = 0.$$

Analogously, we can verify that:

$$\lim_{y \searrow -\infty} P\left[X \le x, Y \le y\right] = 0.$$

4. It is straightforward that:

$$\lim_{(x,y)\to(\infty,\infty)} F(x,y) = \lim_{x\to\infty}\left(\lim_{y\to\infty} P(X \le x, Y \le y)\right)$$
$$= \lim_{x\to\infty} P(X \le x) = 1.$$

∎

The following theorem is the general case for n-dimensional random vectors of the above theorem which can be proved in a similar fashion.

Theorem 5.4 *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be an n-dimensional random vector. The joint cumulative distribution function F of the random variables $X_1, X_2, \cdots, X_n$ has the following properties:*

1. *$\Delta_a^b F \ge 0$ for all $a = (a_1, \cdots, a_n)$, $b = (b_1, \cdots, b_n) \in \mathbb{R}^n$ with $a \le b$, where:*

$$\Delta_a^b F := \sum_{(\epsilon_1,\cdots,\epsilon_n)\in\{0,1\}^n} (-1)^{\left(\sum_{i=1}^{n}\epsilon_i\right)} F\left(\epsilon_1 a_1 + (1-\epsilon_1)b_1, \cdots, \epsilon_n a_n + (1-\epsilon_n)b_n\right).$$

2. *F is right continuous on each component.*

3. *For all $a_1, \cdots, a_{i-1}, a_{i+1}, \cdots a_n \in \mathbb{R}$ with $i = 1, \cdots, n$, we have:*

$$\lim_{x \searrow -\infty} F\left(a_1, \cdots, a_{i-1}, \underset{\text{ith position}}{\underset{\downarrow}{x}}, a_{i+1}, \cdots, a_n\right) = 0.$$

4.

$$\lim_{(x_1,\cdots,x_n)\to(\infty,\cdots,\infty)} F(x_1, \cdots, x_n) = 1.$$

■ EXAMPLE 5.5

Check whether the following functions are joint cumulative distribution functions:

1.
$$F(x,y) = \begin{cases} e^{-(x+y)} & \text{if } 0 < x < \infty, 0 < y < \infty \\ 0 & \text{otherwise.} \end{cases}$$

2.
$$F(x,y) = \begin{cases} 1 & \text{if } x + 2y \geq 1 \\ 0 & \text{if } x + 2y < 1. \end{cases}$$

3.
$$F(x,y) = \begin{cases} 0 & \text{if } \quad x + y < 0 \\ 1 & \text{if } \quad x + y \geq 0. \end{cases}$$

4.
$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } \quad x > 0, y > 0 \\ 0 & \text{otherwise.} \end{cases}$$

Solution:

1.
$$\begin{aligned} \lim_{x\to\infty, y\to\infty} F(x,y) &= \lim_{x\to\infty, y\to\infty} e^{-(x+y)} \\ &= \lim_{x\to\infty, y\to\infty} e^{-x}.e^{-y} \\ &= \lim_{x\to\infty} e^{-x}. \lim_{y\to\infty} e^{-y} \\ &= 0 \times 0 = 0 \neq 1. \end{aligned}$$

$F(x,y)$ is not a joint cumulative distribution function.

2. For instance, take $x_1 = \frac{1}{3}$, $x_2 = 1$, $y_1 = \frac{1}{4}$, $y_2 = 1$.

$$\begin{aligned} P\left(\frac{1}{3} < X \leq 1, \frac{1}{4} < y \leq 1\right) &= F(1,1) - F\left(1, \frac{1}{4}\right) \\ &\quad - F\left(\frac{1}{3}, 1\right) + F\left(\frac{1}{3}, \frac{1}{4}\right) \\ &= 1 - 1 - 1 + 0 = -1 < 0. \end{aligned}$$

$F(x,y)$ is not a joint cumulative distribution function.

3. For instance, take $x_1 = -\frac{1}{2}$, $x_2 = 0$, $y_1 = 0$, $y_2 = \frac{1}{2}$

$$\begin{aligned} P\left(-\frac{1}{2} < X \le 0,\ 0 < y \le \frac{1}{2}\right) &= F\left(0, \frac{1}{2}\right) - F\left(-\frac{1}{2}, \frac{1}{2}\right) \\ &\quad - F(0,0) + F\left(-\frac{1}{2}, 0\right) \\ &= 1 - 1 - 1 + 0 = -1 < 0. \end{aligned}$$

$F(x, y)$ is not a joint cumulative distribution function.

4. Since all the four properties of Theorem 5.3 are satisfied, $F(x, y)$ is a joint cumulative distribution function. ▲

Definition 5.5 (Jointly Continuous Random Variables) *Let $X_1, X_2, \cdots, X_n$ be n real-valued random variables defined over the same probability space. It is said that the random variables are jointly continuous, if there is an integrable function $f : \mathbb{R}^n \longrightarrow [0, +\infty)$ such that for every Borel set C of $\mathbb{R}^n$:*

$$P((X_1, X_2, \cdots, X_n) \in C) = \int \cdots \int_C f(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n .$$

The function f is called the joint probability density function of the random variables $X_1, X_2, \cdots, X_n$.

Note 5.4 *From the definition above, we have, in particular:*

1. $\displaystyle\int \cdots \int_{\mathbb{R}^n} f(x_1, x_2, \cdots, x_n)\, dx_1\, dx_2 \cdots dx_n = 1.$

2.

$$\begin{aligned} & P(X_1 \le x_1, X_2 \le x_2, \cdots, X_n \le x_n) \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \cdots, t_n)\, dt_1\, dt_2 \cdots dt_n . \end{aligned} \tag{5.3}$$

The previous remark shows that if the joint probability density function f of the random variables X_1, $X_2, \cdots, X_n$ is known, then the joint distribution function F is also known. This raises the question: Does the converse hold as well? That is, is it possible, starting with the joint distribution function F, to find the joint probability density function f? The answer is given in the next theorem:

Theorem 5.5 *Let X and Y be continuous random variables having a joint distribution function F. Then, the joint probability density function f is*

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x\, \partial y} = \frac{\partial^2 F(x, y)}{\partial y \partial x}$$

for all the points (x, y) where $f(x, y)$ is continuous.

Proof: By applying the fundamental theorem of calculus to (5.3), we obtain

$$\frac{\partial F(x,y)}{\partial x} = \int_{-\infty}^{y} f(x,v)dv$$

and therefore:

$$\frac{\partial}{\partial y}\left(\frac{\partial F(x,y)}{\partial x}\right) = f(x,y) \ .$$

Since $\frac{\partial^2 F(x,y)}{\partial x \partial y}$ and $\frac{\partial^2 F(x,y)}{\partial y \partial x}$ exist and are both continuous, then

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{\partial^2 F(x,y)}{\partial y \partial x},$$

which completes the proof of the theorem. ■

Furthermore, if $X_1, \cdots, X_n$ are n continuous random variables with joint distribution function F, then the function $g(\cdot, \cdots, \cdot)$ defined over $\mathbb{R}^n$ by

$$g(u_1, \cdots, u_n) := \begin{cases} \left.\frac{\partial^n F(x_1,x_2,\cdots,x_n)}{\partial x_1 \partial x_2 \cdots \partial x_n}\right|_{(x_1,x_2,\cdots,x_n)=(u_1,\cdots,u_n)} & \text{if the partial derivative exits} \\ & \text{derivative exists} \\ 0 & \text{otherwise} \end{cases}$$

is a joint probability density function of the random variables $X_1, \cdots, X_n$.

Suppose now that X and Y are continuous random variables having joint probability density function f, and let g be the function defined by:

$$g(x) := \int_{-\infty}^{\infty} f(x,y)dy.$$

Clearly:

$$\begin{aligned} \int_{-\infty}^{x} g(u)\,du &= \int_{-\infty}^{x}\int_{-\infty}^{\infty} f(u,y)\,dy\,du \\ &= \lim_{t\to\infty}\int_{-\infty}^{x}\int_{-\infty}^{t} f(u,y)\,dy\,du \\ &= \lim_{t\to\infty} F(x,t) = F_X(x). \end{aligned}$$

Moreover, since

$$\int_{-\infty}^{\infty} g(x)\,dx = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1 \ ,$$

g is the density function of the random variable X called the *marginal density function* of X and is commonly notated as $f_X(x)$.

In similar fashion,

$$f_Y(y) := \int_{-\infty}^{\infty} f(x,y)dx$$

is the density function of the random variable Y.

In general, we have the following result:

Theorem 5.6 *If $X_1, X_2, \cdots, X_n$ are n real-valued random variables, having joint pdf f, then*

$$f_{X_j}(x) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \cdots, x_{j-1}, x, x_{j+1}, \cdots, x_n)\, dx_1 \cdots dx_{j-1}\, dx_{j+1} \cdots dx_n$$

is the density function of the random variable X_j for $j = 1, 2, \cdots, n$.

EXAMPLE 5.6

Let X and Y be random variables with joint *pdf* given by:

$$f(x,y) = \begin{cases} 6xy^2 & \text{if} \quad 0 < x < 1, \quad 0 < y < 1 \\ 0 & \text{otherwise} . \end{cases}$$

1. Calculate $P\left(\frac{1}{2} < X \leq \frac{3}{4}, 0 < Y \leq \frac{1}{3}\right)$.

2. Compute $P(\frac{1}{2} < X \leq \frac{3}{4})$.

Solution:

1.

$$\begin{aligned} P\left(\frac{1}{2} < X \leq \frac{3}{4},\ 0 < Y \leq \frac{1}{3}\right) &= \int_0^{\frac{1}{3}} \int_{\frac{1}{2}}^{\frac{3}{4}} 6xy^2\, dx\, dy \\ &= 1.1574 \times 10^{-2} . \end{aligned}$$

2.

$$\begin{aligned} P\left(\frac{1}{2} < X \leq \frac{3}{4}\right) &= \int_{\frac{1}{2}}^{\frac{3}{4}} \int_0^1 6xy^2\, dy\, dx \\ &= 0.3125 . \quad \blacktriangle \end{aligned}$$

■ EXAMPLE 5.7

Let X and Y be random variables with joint *pdf* given by:

$$f(x,y) = \begin{cases} \frac{24x^2}{y^3} & \text{if} \quad 0 < x < 1;\ y > 2 \\ 0 & \text{otherwise} . \end{cases}$$

1. Calculate $P\left(X < \frac{1}{2} \mid Y > 6\right)$.

2. Find the marginal density functions of X and Y.

Solution:

1.

$$\begin{aligned} P\left(X < \frac{1}{2} \mid Y > 6\right) &= \frac{P(X < \frac{1}{2}, Y > 6)}{P(Y > 6)} \\ &= \frac{\int_0^{\frac{1}{2}} \int_6^{\infty} \frac{24x^2}{y^3} dydx}{\int_6^{\infty} \int_0^1 \frac{24x^2}{y^3} dxdy} = \frac{1.3889 \times 10^{-2}}{0.1111} = 0.125. \end{aligned}$$

2.

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\, dy \\ &= \begin{cases} \displaystyle\int_2^{\infty} \frac{24x^2}{y^3}\, dy & \text{if} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} 3x^2 & \text{if} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases} \\ f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)\, dx \\ &= \begin{cases} \displaystyle\int_0^1 \frac{24x^2}{y^3}\, dx & \text{if} \quad y > 2 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} \frac{8}{y^3} & \text{if} \quad y > 2 \\ 0 & \text{otherwise} . \end{cases} \end{aligned}$$

▲

■ EXAMPLE 5.8

Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} (x+y) & \text{if} \quad 0 < x < 1;\ 0 < y < 1 \\ 0 & \text{otherwise .} \end{cases}$$

Find:

1. The joint cumulative distribution function of X and Y.
2. The marginal density functions of X and Y.

Solution:

1.

$$\begin{aligned} F(x,y) &= \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,t)\, dt\, du \\ &= \begin{cases} \frac{x^2y+xy^2}{2} & \text{if} \quad 0 < x < 1;\ 0 < y < 1 \\ \frac{x^2+x}{2} & \text{if} \quad 0 < x < 1;\ y \geq 1 \\ \frac{y+y^2}{2} & \text{if} \quad x \geq 1;\ 0 < y < 1 \\ 1 & \text{if} \quad x \geq 1;\ y \geq 1 \\ 0 & \text{otherwise .} \end{cases} \end{aligned}$$

2.

$$f_X(x) = \begin{cases} x + \frac{1}{2} & \text{if} \quad 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} y + \frac{1}{2} & \text{if} \quad 0 < y < 1 \\ 0 & \text{otherwise.} \end{cases} \qquad \blacktriangle$$

■ EXAMPLE 5.9

Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} k & \text{if} \ \ 0 \leq x \leq 1;\ 0 \leq y \leq 1 \ \text{ and } \ 3y \leq x \\ 0 & \text{otherwise .} \end{cases}$$

Find:

1. The value of the constant k.
2. The joint cumulative distribution function of X and Y.
3. The marginal density functions of X and Y.
4. $P(2Y \leq X \leq 5Y)$.

Solution:

1.

$$1 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy = \int_0^1\int_0^{\frac{x}{3}} k\,dy\,dx = \frac{k}{6}$$

so that $k = 6$.

2.

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(u,t)\,dt\,du$$
$$= \begin{cases} 0 & \text{if } x < 0 \text{ or } y < 0 \\ 6xy - 8y^2 & \text{if } 0 \leq x \leq 1,\ 0 \leq y \leq \frac{1}{3},\ 3y \leq x \\ x^2 & \text{if } 0 \leq x \leq 1,\ \frac{x}{3} \leq y \leq \frac{1}{3} \\ 6y - 9y^2 & \text{if } x > 1,\ 0 \leq y \leq \frac{1}{3} \\ x^2 & \text{if } 0 \leq x \leq 1,\ y \geq \frac{1}{3} \\ 1 & \text{if } x \geq 1,\ y \geq \frac{1}{3}\ . \end{cases}$$

3.

$$f_X(x) = \begin{cases} \displaystyle\int_0^{\frac{x}{3}} 6dy & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
$$= \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$
$$f_Y(y) = \begin{cases} \displaystyle\int_{3y}^{1} 6dx & \text{if } 0 \leq y \leq \frac{1}{3} \\ 0 & \text{otherwise} \end{cases}$$
$$= \begin{cases} 6(1-3y) & \text{if } 0 \leq y \leq \frac{1}{3} \\ 0 & \text{otherwise} \end{cases}.$$

4.

$$P(2Y \leq X \leq 5Y) = \int_0^1 \left(\int_{\frac{x}{5}}^{\frac{x}{3}} 6\,dy\right) dx = 0.4. \qquad \blacktriangle$$

EXAMPLE 5.10

The system analyst at an email server in a university is interested in the joint behavior of the random variable X, defined as the total time between an email's arrival at the server and departure from the service window, and Y, the time an email waits in the buffer before reaching the service window. Because X includes the time an email waits in the buffer, we must have $X > Y$. The relative frequency distribution of observed values of X and Y can be modeled by the joint *pdf*

$$f(x,y) = \begin{cases} e^{-x} & \text{if } 0 \leq y < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

with time measured in seconds.

1. Find $P(X < 2, Y > 1)$.

2. Find $P(X > 2Y)$.

3. Find $P(X - Y \geq 1)$.

Solution:

1.

$$P(X < 2, Y > 1) = \int\int_{\{(x,y):x<2,y>1\}} f(x,y)dxdy = 0.$$

2.

$$\begin{aligned} P(X > 2Y) &= \int\int_{\{(x,y):x>2y\}} f(x,y)dxdy \\ &= \int_0^\infty \int_0^{x/2} e^{-x} dydx \\ &= \frac{1}{2}. \end{aligned}$$

3.

$$\begin{aligned} P(X - Y \geq 1) &= \int\int_{\{(x,y):x-y\geq 1\}} f(x,y)dxdy \\ &= \int_0^\infty \int_{y+1}^\infty e^{-x} dxdy \\ &= e^{-1}. \quad \blacktriangle \end{aligned}$$

5.2 INDEPENDENT RANDOM VARIABLES

Definition 5.6 (Independent Random Variables) *Let X and Y be two real-valued random variables defined over the same probability space. If for any pair of Borel sets A and B of $\mathbb{R}$ we have*

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B),$$

then X and Y are said to be independent.

We say that X and Y are independent and identically distributed (i.i.d.) random variablesif X and Y are independent random variables and have the same distributions.

Note 5.5 (Independent Random Vectors) *The previous definition can be generalized to random vectors as follows: Two n-dimensional random vectors $\mathbf{X}$ and $\mathbf{Y}$ defined over the same probability space $(\Omega, \Im, P)$ are said to be independent, if for any A and B Borel subsets of $\mathbb{R}^n$ they satisfy:*

$$P(\mathbf{X} \in A, \ \mathbf{Y} \in B) = P(\mathbf{X} \in A)\, P(\mathbf{Y} \in B) \ .$$

Assume that X and Y are independent random variables. Then it follows from the definition above that:

$$F(x, y) = P(X \le x, Y \le y) = P(X \le x)P(Y \le y) \ \text{ for all } x, y \in \mathbb{R}.$$

That is:

$$F(x, y) = F_X(x)F_Y(y) \ \text{ for all } x, y \in \mathbb{R}. \tag{5.4}$$

Conversely, if the condition (5.4) is met, then the random variables are independent.

Suppose now that X and Y are independent discrete random variables. Then

$$P(X = x, Y = y) = P(X = x)P(Y = y) \tag{5.5}$$

for all x in the image of X and all y in the image of Y. Conversely, if the condition (5.5) is met, then, the random variables are independent.

If X and Y are independent random variables with joint density function $f(x, y)$, then:

$$P(x < X \le x + dx, \ y < Y \le y + dy) = P(x < X \le x + dx)P(y < Y \le y + dy).$$

That is:

$$f(x, y) = f_X(x)f_Y(y) \ \text{ for all } x, y \in \mathbb{R} \ . \tag{5.6}$$

Conversely, if the condition (5.6) is satisfied, then the random variables are independent. In conclusion, we have that the random variables X and Y are independent if and only if its joint density function $f(x, y)$ can be factorized as the product of their marginal density functions $f_X(x)$ and $f_Y(y)$.

■ EXAMPLE 5.11

There are nine colored balls inside an urn, three of which are red while the remaining six are blue. A random sample of size 2 (with replacement) is extracted. Let X and Y be the random variables defined by:

$$X := \begin{cases} 1 & \text{if the first extracted ball is red} \\ 0 & \text{if the first extracted ball is blue} \end{cases}$$

$$Y := \begin{cases} 1 & \text{if the second extracted ball is red} \\ 0 & \text{if the second extracted ball is blue} . \end{cases}$$

Are X and Y independent? Explain.

Solution: The joint distribution function of X and Y is given by:

$X\backslash Y$	0	1
0	$\frac{4}{9}$	$\frac{2}{9}$
1	$\frac{2}{9}$	$\frac{1}{9}$

It can be easily verified that for any choice of $x, y \in \{0, 1\}$:

$$P(X = x, Y = y) = P(X = x)\ P(Y = y) .$$

That is, X and Y are independent random variables, which was to be expected since the composition of the urn is the same for each extraction.
▲

■ EXAMPLE 5.12

Solve the previous exercise, now under the assumption that the extraction takes place without replacement.

Solution: In this case, the joint distribution function of X and Y is given by:

$X\backslash Y$	0	1
0	$\frac{5}{12}$	$\frac{3}{12}$
1	$\frac{3}{12}$	$\frac{1}{12}$

Given:

$$P(X = 0, Y = 0) = \frac{5}{12} \neq \frac{2}{3} \times \frac{2}{3} = P(X = 0)P(Y = 0).$$

Then X and Y are *not* independent. ▲

■ **EXAMPLE 5.13**

Let X and Y be independent random variables with density functions given by:

$$f(x) = \frac{x^2}{9}\mathcal{X}_{(0,3)}(x)$$
$$g(y) = y^{-2}\mathcal{X}_{(1,\infty)}(y).$$

Find $P(XY > 1)$.
Solution: Since X and Y are independent, their joint density function is given by:

$$h(x,y) = \begin{cases} \frac{x^2}{9y^2} & \text{if } 0 < x < 3;\ y > 1 \\ 0 & \text{otherwise }. \end{cases}$$

Accordingly:

$$P(XY > 1) = \int_1^\infty \int_{\frac{1}{y}}^3 \frac{x^2}{9y^2}\, dx\, dy = 0.99074\ . \qquad ▲$$

Note 5.6 *Let X and Y be independent discrete random variables with values in $\mathbb{N}$. Clearly:*

$$\begin{aligned} P(X+Y=z) &= \sum_{x=0}^{z} P(X=x,\, Y=z-x) \\ &= \sum_{x=0}^{z} P(X=x)\, P(Y=z-x), \quad z \in \mathbb{N}\ . \end{aligned}$$

■ **EXAMPLE 5.14**

Suppose that X and Y are independent random variables with $X \stackrel{d}{=} \mathcal{P}(\lambda)$ and $Y \stackrel{d}{=} \mathcal{P}(\mu)$. Find the distribution of $Z = X + Y$.
Solution: The random variables X and Y take the values $0, 1, \cdots$, and therefore the random variable Z also takes the values $0, 1, \cdots$. Let

$z \in \{0, 1, \cdots\}$. Then:

$$\begin{aligned} P(Z = z) &= \sum_{x=0}^{z} P(X = x)P(Y = z - x) \\ &= \sum_{x=0}^{z} e^{-\lambda} \frac{\lambda^x}{x!} e^{-\mu} \frac{\mu^{z-x}}{(z-x)!} \\ &= \frac{1}{z!} e^{-(\lambda+\mu)} \sum_{x=0}^{z} \binom{z}{x} \lambda^x \mu^{z-x} \\ &= \frac{1}{z!} e^{-(\lambda+\mu)} (\lambda + \mu)^z . \end{aligned}$$

That is, $Z \overset{d}{=} \mathcal{P}(\lambda + \mu)$. ▲

Note 5.7 *Suppose that X and Y are independent random variables having joint probability density function $f(x, y)$ and marginal density functions f_X and f_Y, respectively. The distribution function of the random variable $Z = X + Y$ can be obtained as follows:*

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= \iint_{\{x+y \leq z\}} f(x, y)\, dx\, dy \\ &= \iint_{\{x+y \leq z\}} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{z-y} f_X(x) f_Y(y)\, dx\, dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{z-y} f_X(x)\, dx \right) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} F_X(z - y) f_Y(y)\, dy . \end{aligned} \tag{5.7}$$

Upon derivation, equation (5.7) yields the density function of the random variable Z as follows:

$$\begin{aligned} f_Z(z) &= \frac{d}{dz} \int_{-\infty}^{\infty} F_X(z - y) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} \frac{d}{dz} F_X(z - y) f_Y(y)\, dy \\ &= \int_{-\infty}^{\infty} f_X(z - y) f_Y(y)\, dy . \end{aligned} \tag{5.8}$$

The density function of the random variable Z is called the convolution of the density functions f_X and f_Y and is notated $f_X * f_Y$.

■ EXAMPLE 5.15

Let X and Y be i.i.d. random variables having an exponential distribution of parameter $\lambda > 0$. The density function of $Z = X + Y$ is given by:

$$\begin{aligned} f_Z(z) &= (f_X * f_Y)(z) \\ &= \int_{-\infty}^{\infty} \lambda e^{-\lambda(z-u)} \mathcal{X}_{(0,\infty)}(z-u)\, \lambda e^{-\lambda u} \mathcal{X}_{(0,\infty)}(u)\, du \\ &= \int_{-\infty}^{\infty} \lambda^2 e^{-\lambda z} \mathcal{X}_{(0,\infty)}(z-u) \mathcal{X}_{(0,\infty)}(u) du. \end{aligned}$$

Given:

$$\mathcal{X}_{(0,\infty)}(z-u)\mathcal{X}_{(0,\infty)}(u) = \begin{cases} 1 & \text{if } z > u > 0 \\ 0 & \text{otherwise .} \end{cases}$$

Then :

$$\begin{aligned} (f_X * f_Y)(z) &= \lambda^2 z e^{-\lambda z} \mathcal{X}_{(0,\infty)}(z) \\ &= \frac{\lambda}{\Gamma(2)} (\lambda z) e^{-\lambda z} \mathcal{X}_{(0,\infty)}(z). \end{aligned}$$

That is, $Z \stackrel{d}{=} \Gamma(2, \lambda)$. ▲

■ EXAMPLE 5.16

Let X and Y be independent random variables such that $X \stackrel{d}{=} \mathcal{U}[-1, 2]$ and $Y \stackrel{d}{=} Exp(1)$. Calculate $P(X + Y \geq 1)$.

Solution: The joint probability density function of X and Y is given by

$$f(x,y) = \begin{cases} \frac{1}{3} e^{-y} & \text{if } -1 \leq x \leq 2;\ y > 0 \\ 0 & \text{otherwise} \end{cases}$$

so that:

$$\begin{aligned} P(X + Y \geq 1) &= \iint_{\{(x,y): x+y \geq 1\}} f(x,y)\, dx\, dy \\ &= \int_{-1}^{1} \int_{1-x}^{\infty} \frac{1}{3} e^{-y}\, dy\, dx + \int_{1}^{2} \int_{0}^{\infty} \frac{1}{3} e^{-y}\, dy\, dx \\ &= \int_{-1}^{1} \frac{1}{3} e^{-(1-x)}\, dx + \int_{1}^{2} \frac{1}{3} dx \\ &= 0.62155\ . \end{aligned}$$

▲

Below, the notion of independence is generalized to n random variables:

Definition 5.7 (Independence of n Random Variables) *n real-valued random variables $X_1, X_2, \cdots, X_n$ defined over the same probability space are called independent if and only if for any collection of Borel subsets $A_1, A_2, \cdots, A_n$ of $\mathbb{R}$ the following condition holds:*

$$P(X_1 \in A_1, \cdots, X_n \in A_n) = P(X_1 \in A_1) \cdots P(X_n \in A_n).$$

Note 5.8 (Independence of n Random Vectors) *The definition above can be extended to random vectors in the following way: n k-dimensional random vectors $\mathbf{X}_1, \mathbf{X}_2, \cdots, \mathbf{X}_n$, defined over the same probability space $(\Omega, \Im, P)$, are said to be independent if and only if they satisfy*

$$P(\mathbf{X}_1 \in A_1, \cdots, \mathbf{X}_n \in A_n) = P(\mathbf{X}_1 \in A_1) \cdots P(\mathbf{X}_n \in A_n)$$

for any collection of Borel subsets $A_1, A_2, \cdots, A_n$ of $\mathbb{R}^k$.

■ EXAMPLE 5.17 Distribution of the Maximum and Minimum

Let $X_1, \cdots, X_n$ be n real-valued random variables defined over $(\Omega, \Im, P)$ and consider the random variables Y and Z defined as follows:

$$\begin{array}{rccc} Y := \max(X_1, \cdots, X_n): & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & \max(X_1(\omega), \cdots, X_n(\omega)) \end{array}$$

and

$$\begin{array}{rccc} Z := \min(X_1, \cdots, X_n): & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & \min(X_1(\omega), \cdots, X_n(\omega)) \end{array}.$$

Clearly,

$$F_Y(y) := P(Y \leq y) = P(X_1 \leq y, \cdots, X_n \leq y)$$

and:

$$F_Z(z) := P(Z \leq z) = 1 - P(X_1 > z, \cdots, X_n > z).$$

Therefore, if the random variables $X_1, \cdots, X_n$ are independent, then

$$F_Y(y) = \prod_{k=1}^{n} P(X_k \leq y) = \prod_{k=1}^{n} F_{X_k}(y)$$

$$\begin{aligned} F_Z(z) &= 1 - \prod_{k=1}^{n} [1 - P(X_k \leq z)] \\ &= 1 - \prod_{k=1}^{n} [1 - F_{X_k}(z)] \end{aligned}$$

where $F_{X_k}(\cdot)$ is the distribution function of the random variable X_k for $k = 1, \cdots, n$. ▲

EXAMPLE 5.18

Let X and Y be independent and identically distributed random variables having an exponential distribution of parameter a. Determine the density function of the random variable $Z := \max\{X, Y^3\}$.

Solution: The density function of the random variable $U = Y^3$ is given by:

$$f_U(u) = \begin{cases} \frac{a}{3\sqrt[3]{u^2}} e^{-a\sqrt[3]{u}} & \text{if } u > 0 \\ 0 & \text{otherwise .} \end{cases}$$

Since the random variables X and Y^3 are independent, it follows from the previous example that $Z = \max\{X, Y^3\}$ has the following cumulative distribution function:

$$F_Z(z) = F_X(z) F_{Y^3}(z).$$

Therefore, the density function of Z is given by:

$$f_Z(z) = f_X(z) F_{Y^3}(z) + F_X(z) f_{Y^3}(z).$$

That is:

$$f_Z(z) = \begin{cases} ae^{-az} + \frac{a}{3\sqrt[3]{z^2}} e^{-a\sqrt[3]{z}} - a(1 + \frac{1}{3\sqrt[3]{z^2}}) e^{-a\sqrt[3]{z} - az} & \text{if } z > 0 \\ 0 & \text{otherwise .} \end{cases}$$ ▲

Note 5.9 *If $X_1, X_2, \cdots, X_n$ are n independent random variables, then $X_1, X_2, \cdots, X_k$ with $k \le n$ are also independent. To this effect, let $A_1, A_2, \cdots, A_k$ be Borel subsets of $\mathbb{R}$. Then:*

$$\begin{aligned} P(X_1 \in A_1, \cdots, X_k \in A_k) &= P(X_1 \in A_1, \cdots, X_k \in A_k, X_{k+1} \in \mathbb{R}, \\ &\qquad \cdots, X_n \in \mathbb{R}) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k) P(X_{k+1} \in \mathbb{R}) \\ &\qquad \cdots P(X_n \in \mathbb{R}) \\ &= P(X_1 \in A_1) \cdots P(X_k \in A_k). \end{aligned}$$

Suppose that X_1, X_2 and X_3 are independent discrete random variables and let $Y_1 := X_1 + X_2$ and $Y_2 := X_3^2$. Then:

$$\begin{aligned} P(Y_1 = z, Y_2 = w) &= \sum_x P(X_1 = x, X_2 = z - x, X_3^2 = w) \\ &= \sum_x P(X_1 = x) P(X_2 = z - x) P(X_3 = \pm\sqrt{w}) \\ &= P(X_3 = \pm\sqrt{w}) \sum_x P(X_1 = x) P(X_2 = z - x) \\ &= P(X_1 + X_2 = z) P(X_3^2 = w). \end{aligned}$$

That is, $Y_1 := X_1 + X_2$ and $Y_2 := X_3^2$ are independent random variables. Does this result hold in general? The answer to this question is given by the following theorem whose demonstration is omitted since it relies on measure-theoretic results.

Theorem 5.7 *Let $X_1, \cdots, X_n$, be n independent random variables. Let Y be a random variable defined in terms of $X_1, \cdots, X_k$ and let Z be a random variable defined in terms of $X_{k+1}, \cdots, X_n$, where $1 \leq k < n$. Then Y and Z are independent.*

■ **EXAMPLE 5.19**

Let $X_1, \cdots, X_5$ be independent random variables. By Theorem 5.7, it is obvious that $Y = X_1X_2 + X_3$ and $Z = e^{X_5} \sin X_4$ are independent random variables. ▲

5.3 DISTRIBUTION OF FUNCTIONS OF A RANDOM VECTOR

Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a random vector and let $g_1(\cdot, \cdots, \cdot), \cdots, g_k(\cdot, \cdots, \cdot)$ be real-valued functions defined over $A \subseteq \mathbb{R}^n$. Consider the following random variables: $Y_1 := g_1(X_1, \cdots, X_n), \cdots, Y_k := g_k(X_1, \cdots, X_n)$. We wish to determine the joint distribution of $Y_1, \cdots, Y_k$ in terms of the joint distribution of the random variables $X_1, \cdots, X_n$.

Suppose that the random variables $X_1, \cdots, X_n$ are discrete and that their joint distribution is known. Clearly:

$$\begin{aligned} P(Y_1 = y_1, \cdots, Y_k = y_k) &= P\big(g_1(X_1, \cdots, X_n) = y_1, \cdots, g_k(X_1, \cdots, X_n) = y_k\big) \\ &= \sum_{\substack{(x_1, \cdots, x_n) \in \mathbb{R}^n : \\ g_1(x_1, \cdots, x_n) = y_1 \\ \vdots \\ g_k(x_1, \cdots, x_n) = y_k}} P(X_1 = x_1, \cdots, X_n = x_n). \end{aligned}$$

■ **EXAMPLE 5.20**

Let X_1 and X_2 be random variables with joint distribution given by:

$X_1 \backslash X_2$	0	1
-1	$\frac{1}{7}$	$\frac{1}{7}$
0	$\frac{2}{7}$	$\frac{1}{7}$
1	$\frac{1}{7}$	$\frac{1}{7}$

Let $g_1(x_1, x_2) := x_1 + x_2$ and $g_2(x_1, x_2) = x_1 x_2$. Clearly, the random variables

$$Y_1 := g_1(X_1, X_2) = X_1 + X_2 \quad \text{and} \quad Y_2 := g_2(X_1, X_2) = X_1 X_2$$

take, respectively, the values $-1, 0, 1, 2$ and $-1, 0, 1$. The joint distribution of Y_1 and Y_2 is given by:

$Y_1 \backslash Y_2$	-1	0	1
-1	0	$\frac{1}{7}$	0
0	$\frac{1}{7}$	$\frac{2}{7}$	0
1	0	$\frac{2}{7}$	0
2	0	0	$\frac{1}{7}$

▲

■ EXAMPLE 5.21

Let X_1, X_2 and X_3 be random variables having joint distribution function given by:

$\mathbf{x} = (x_1, x_2, x_3)$	$(0,0,0)$	$(0,0,1)$	$(0,1,1)$	$(1,0,1)$	$(1,1,0)$	$(1,1,1)$
$P((X_1, X_2, X_3) = \mathbf{x})$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$

Let $g_1(x_1, x_2, x_3) := x_1 + x_2 + x_3$, $g_2(x_1, x_2, x_3) = |x_3 - x_2|$. The joint distribution of $Y_1 := g_1(X_1, X_2, X_3)$ and $Y_2 := g_2(X_1, X_2, X_3)$ is given by:

$Y_2 \backslash Y_1$	0	1	2	3
0	$\frac{1}{8}$	0	$\frac{1}{8}$	$\frac{1}{8}$
1	0	$\frac{3}{8}$	$\frac{2}{8}$	0

▲

In the case of absolutely continuous random variables, we have the following result. The interested reader can consult its proof in Jacod and Protter (2004).

Theorem 5.8 (Transformation Theorem) *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ *be a random vector with joint density function* $f_{\mathbf{X}}$. *Let* $g : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ *be an* *injective map. Suppose that both g and its inverse $h : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ are continuous. If the partial derivatives of h exist and are continuous and if their Jacobian J is different from zero, then the random vector $\mathbf{Y} := g(\mathbf{X})$ has joint density function $f_{\mathbf{Y}}$ given by:*

$$f_{\mathbf{Y}}(\mathbf{y}) = \begin{cases} |J(\mathbf{y})| \, f_{\mathbf{X}}(h(\mathbf{y})) & \text{if } \mathbf{y} \text{ is on } g\text{'s range} \\ 0 & \text{otherwise} . \end{cases}$$

■ EXAMPLE 5.22

Let $\mathbf{X} = (X_1, X_2)$ be a random vector having joint probability density function given by:

$$f_{\mathbf{X}}(x_1, x_2) = \begin{cases} 1 & \text{if } 0 < x_1 < 1, \ 0 < x_2 < 1 \\ 0 & \text{otherwise} . \end{cases}$$

Find the joint probability density function of $\mathbf{Y} = (Y_1, Y_2)$, where $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.
Solution: In this case we have

$$g(x_1, x_2) = (g_1(x_1, x_2), g_2(x_1, x_2)) = (x_1 + x_2, x_1 - x_2)$$

and the inverse transformation is given by:

$$h(x_1, x_2) = \left(\frac{x_1 + x_2}{2}, \frac{x_1 - x_2}{2} \right) .$$

The Jacobian J of the inverse transformation would then equal:

$$J = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2} .$$

Therefore, the joint probability density function of $\mathbf{Y}$ is:

$$f_{\mathbf{Y}}(y_1, y_2) = \begin{cases} \frac{1}{2} & \text{if } 0 < y_1 + y_2 < 2 \ , \ 0 < y_1 - y_2 < 2 \\ 0 & \text{otherwise} . \quad \blacktriangle \end{cases}$$

In general we have the following result for distribution of the sum and the difference of random variables.

Theorem 5.9 *Let X and Y be random variables having joint probability density function f. Let $Z := X + Y$ and $W := X - Y$. Then the probability density functions of Z and W are given by*

$$f_Z(z) = \int_{-\infty}^{\infty} f(z - u, u) \, du$$

and

$$f_W(w) = \int_{-\infty}^{\infty} f(u + w, u)\, du,$$

respectively.

Proof: Just like in the previous example, we have that

$$g(x, y) := (x + y, x - y)$$

and the inverse transformation h is given by:

$$h(x, y) = \left(\frac{x+y}{2}, \frac{x-y}{2}\right).$$

Therefore, the joint probability density function of Z and W equals

$$f_{(Z,W)}(z, w) = \frac{1}{2} f\left(\frac{z+w}{2}, \frac{z-w}{2}\right),$$

from which we obtain that:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{2} f\left(\frac{z+w}{2}, \frac{z-w}{2}\right) dw \\ &= \int_{-\infty}^{\infty} f(u, z - u)\, du \\ &= \int_{-\infty}^{\infty} f(z - u, u)\, du. \end{aligned}$$

In a similar fashion:

$$\begin{aligned} f_W(w) &= \int_{-\infty}^{\infty} \frac{1}{2} f\left(\frac{z+w}{2}, \frac{z-w}{2}\right) dz \\ &= \int_{-\infty}^{\infty} f(u, u - w)\, du \\ &= \int_{-\infty}^{\infty} f(u + w, u)\, du\ . \end{aligned}$$

▲

■

Note 5.10 *In the particular case where the random variables X and Y are independent, the density function of the random variable $Z := X + Y$ is given by*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z - u)\, du = \int_{-\infty}^{\infty} f_X(z - u) f_Y(u)\, du, \tag{5.9}$$

where $f_X(\cdot)$ and $f_Y(\cdot)$ represent the density functions of X and Y respectively. The expression given in (5.9) is the convolution of f_X and f_Y, notated as $f_X * f_Y$ [compare with (5.8)].

■ EXAMPLE 5.23

Let X and Y be independent random variables having density functions given by:

$$f_X(x) = \begin{cases} 1 & \text{if } 2 < x < 3 \\ 0 & \text{otherwise} \end{cases}$$

$$f_Y(y) = \begin{cases} \frac{1}{2} & \text{if } 7 < y < 9 \\ 0 & \text{otherwise} \end{cases} .$$

Determine the density function of $Z = X + Y$.
Solution: We know that:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(u) f_Y(z-u)\, du.$$

Since

$$f_X(u) f_Y(z-u) = \begin{cases} \frac{1}{2} & \text{if } 2 < u < 3 \text{ and } 7 + u < z < 9 + u \\ 0 & \text{otherwise} \end{cases}$$

we obtain:

$$f_Z(z) = \begin{cases} \frac{1}{2}(z-9) & \text{if } 9 < z \le 10 \\ \frac{1}{2} & \text{if } 10 < z < 11 \\ \frac{1}{2}(12-z) & \text{if } 11 \le z < 12 \\ 0 & \text{otherwise .} \end{cases} \quad \blacktriangle$$

■ EXAMPLE 5.24

Let X and Y be random variables having the joint probability density function given by:

$$f(x,y) = \begin{cases} 3x & \text{if } 0 \le y \le x \le 1 \\ 0 & \text{otherwise .} \end{cases}$$

Find the density function of $W := X - Y$.
Solution: According to Theorem 5.9:

$$f_W(w) = \int_{-\infty}^{\infty} f(u, u-w)\, du.$$

Here:

$$f(u, u-w) = \begin{cases} 3u & \text{if } 0 \le u - w \le u \le 1 \\ 0 & \text{otherwise .} \end{cases}$$

Then:

$$f_W(w) = \int_w^1 3u\,\mathcal{X}_{[0,1]}(w)\,du = \frac{3}{2}(1-w^2)\mathcal{X}_{[0,1]}(w)\,. \qquad \blacktriangle$$

Theorem 5.10 (Distribution of the Product of Random Variables) *Let X and Y be random variables having joint probability density function f. Let $Z := XY$. Then the density function of Z is given by:*

$$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|u|} f\left(u, \frac{z}{u}\right)\,du.$$

Proof: We introduce $W := Y$. Let g be the function defined by:

$$g(x,y) := (g_1(x,y), g_2(x,y)) = (xy, y).$$

The inverse transformation equals:

$$h(x,y) = \left(\frac{x}{y}, y\right).$$

The Jacobian of the inverse transformation is:

$$J(x,y) = \begin{vmatrix} \frac{1}{y} & -\frac{x}{y^2} \\ 0 & 1 \end{vmatrix} = \frac{1}{y}.$$

Therefore, the joint probability density function of Z and W is given by:

$$f_{ZW}(z,w) = \left|\frac{1}{w}\right| f\left(\frac{z}{w}, w\right).$$

Now, we obtain the following expression for the density function of $Z = XY$:

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{|w|} f\left(\frac{z}{w}, w\right)\,dw \\ &= \int_{-\infty}^{\infty} \frac{1}{|u|} f\left(u, \frac{z}{u}\right)\,du. \end{aligned}$$

■

■ EXAMPLE 5.25

Let X and Y be two independent and identically distributed random variables, having uniform distribution on the interval $(0,1)$. According to the previous example, the density function of $Z := XY$ is given by

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} \frac{1}{|u|} f\left(u, \frac{z}{u}\right)\,du \\ &= \int_{-\infty}^{\infty} \frac{1}{|u|} f_X(u) f_Y\left(\frac{z}{u}\right)\,du, \end{aligned}$$

where f_X and f_Y represent the density functions of X and Y respectively. Since

$$f_X(u) f_Y\left(\frac{z}{u}\right) = \begin{cases} 1 & \text{if } 0 < u < 1 \text{ and } 0 < z < u \\ 0 & \text{otherwise} \end{cases}$$

then:

$$\begin{aligned} f_Z(z) &= \begin{cases} \int_z^1 \frac{1}{u} du & \text{if } 0 < z < 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} -\ln z & \text{if } 0 < z < 1 \\ 0 & \text{otherwise} . \end{cases} \quad \blacktriangle \end{aligned}$$

Theorem 5.11 (Distribution of the Quotient of Random Variables) *Let X and Y be random variables having joint probability density function f. Let $Z := \frac{X}{Y}$ [which is well defined if $P(Y = 0) = 0$)]. Then the density function of Z is given by:*

$$f_Z(z) = \int_{-\infty}^{\infty} |w| f(zw, w)\, dw.$$

Proof: We introduce $W := Y$. Now, consider the function

$$g(x, y) := (g_1(x, y), g_2(x, y)) = \left(\frac{x}{y}, y\right).$$

The inverse transformation is then given by:

$$h(x, y) = (xy, y).$$

Its Jacobian equals:

$$J(x, y) = \begin{vmatrix} y & x \\ 0 & 1 \end{vmatrix} = y.$$

Then the joint density function of Z and W is given by:

$$f_{ZW}(z, w) = |w| f(zw, w).$$

Therefore, a density function of Z is:

$$f_Z(z) = \int_{-\infty}^{\infty} |w| f(zw, w)\, dw.$$

■

EXAMPLE 5.26

Let X and Y be two independent and identically distributed random variables having uniform distribution on the interval $(0, 1)$. According to the previous example, the density function of $Z := \frac{X}{Y}$ is given by

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} |w| \, f(zw, w) \, dw \\ &= \int_{-\infty}^{\infty} |w| \, f_X(zw) f_Y(w) \, dw \end{aligned}$$

where f_X and f_Y represent the density functions of X and Y, respectively. Since

$$f_X(zw) f_Y(w) = \begin{cases} 1 & \text{if } 0 < zw < 1 \text{ and } 0 < w < 1 \\ 0 & \text{otherwise} \end{cases}$$

then:

$$\begin{aligned} f_Z(z) &= \begin{cases} \int_0^1 w \, dw & \text{if } 0 < z < 1 \\ \int_0^{\frac{1}{z}} w \, dw & \text{if } z \geq 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \frac{1}{2} \mathcal{X}_{(0,1)}(z) + \frac{1}{2z^2} \mathcal{X}_{[1,\infty)}(z) \,. \end{aligned}$$ ▲

EXAMPLE 5.27

Assume that the lifetime X of an electric device is a continuous random variable having a probability density function given by:

$$f(x) = \begin{cases} \frac{1000}{x^2} & \text{if } x > 1000 \\ 0 & \text{otherwise} \,. \end{cases}$$

Let X_1 and X_2 be the life spans of two independent devices. Find the density function of the random variable $Z = \frac{X_1}{X_2}$.

Solution: It is known that:

$$f_Z(z) = \int_{-\infty}^{\infty} |v| \, f(vz) f(v) \, dv.$$

Since

$$f(vz) = \begin{cases} \frac{1000}{v^2 z^2} & \text{if } vz > 1000 \\ 0 & \text{otherwise} \end{cases}$$

then:

(i) If $z \geq 1$, then $f(vz)f(v)$ does not equal zero if and only if $v > 1000$. Thus:

$$f_Z(z) = \int_{1000}^{\infty} \frac{1000}{v^2 z^2} \cdot \frac{1000}{v^2} v\, dv = \frac{(1000)^2}{z^2} \int_{1000}^{\infty} \frac{dv}{v^3} = \frac{1}{2z^2} \,.$$

(ii) If $0 < z < 1$, then $f(vz)f(v)$ does not equal zero if and only if $vz > 1000$ if and only if $v > \frac{1000}{z}$. Therefore:

$$f_Z(z) = \int_{\frac{1000}{z}}^{\infty} \frac{1000}{v^2 z^2} \cdot \frac{1000}{v^2} v\, dv = \frac{1}{2} \,.$$

That is:

$$f_Z(z) = \begin{cases} \frac{1}{2z^2} & \text{if } z > 1 \\ \frac{1}{2} & \text{if } 0 < z < 1 \\ 0 & \text{otherwise} \,. \end{cases} \quad \blacktriangle$$

■ EXAMPLE 5.28 *t*-Student Distribution

Let X and Y be independent random variables such that $X \stackrel{d}{=} \mathcal{N}(0,1)$ and $Y \stackrel{d}{=} \mathcal{X}^2_{(k)}$. Consider the transformation

$$g(x,y) := \left(\frac{x}{\sqrt{\frac{y}{k}}}, y \right) .$$

The inverse transformation is given by

$$h(x,y) = \left(x\sqrt{\frac{y}{k}}, y \right) ,$$

whose Jacobian is:

$$J(x,y) = \sqrt{\frac{y}{k}} \,.$$

Therefore, the joint probability density function of $Z := \frac{X}{\sqrt{\frac{Y}{k}}}$ and $W := Y$ is given by

$$f_{ZW}(z,w) = \sqrt{\frac{w}{k}} f(z\sqrt{\frac{w}{k}}, w), \qquad w > 0,$$

where:

$$f(x,y) = \frac{1}{\sqrt{2\pi}} e^{\left(-\frac{1}{2}x^2\right)} \left[\Gamma\left(\frac{k}{2}\right) \right]^{-1} \left(\frac{1}{2}\right)^{\frac{k}{2}} y^{\frac{k}{2}-1} e^{\left(-\frac{1}{2}y\right)} \;;\; x \in R, y > 0.$$

Integration of $f_{ZW}(z,w)$ with respect to w yields the density function of the random variable Z which is then given by:

$$f_Z(z) = \frac{1}{\sqrt{k\pi}}\Gamma\left(\frac{k+1}{2}\right)\left[\Gamma\left(\frac{k}{2}\right)\right]^{-1}\frac{1}{\left[1+\frac{z^2}{k}\right]^{\frac{k+1}{2}}};\ z \in \mathbb{R}. \qquad (5.10)$$

A real-valued random variable Z is said to have a t-Student distribution with k degrees of freedom, notated as $Z \stackrel{d}{=} t_{(k)}$, if its density function is given by (5.10). ▲

■ EXAMPLE 5.29 *F* Distribution

Let X and Y be independent random variables such that $X \stackrel{d}{=} \mathcal{X}^2_{(m)}$ and $Y \stackrel{d}{=} \mathcal{X}^2_{(n)}$. Consider the map

$$g(x,y) := \left(\frac{\frac{x}{m}}{\frac{y}{n}}, y\right).$$

The inverse of g is given by

$$h(x,y) = \left(\frac{m}{n}xy, y\right)$$

and has a Jacobian equal to $J(x,y) = \frac{m}{n}y$. Therefore, the joint probability density function of $Z := \frac{nX}{mY}$ and $W := Y$ is given by:

$$f_{ZW}(z,w) =$$
$$\frac{1}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}\left(\frac{1}{2}\right)^{\frac{m}{2}}\left(\frac{1}{2}\right)^{\frac{n}{2}}\left(\frac{m}{n}zw\right)^{\frac{m}{2}-1}w^{\frac{n}{2}-1}e^{\left[-\frac{1}{2}w\left(\frac{m}{n}z+1\right)\right]},\ z>0, w>0.$$

Integrating with respect to w we find that the density function of Z is given by the following expression:

$$f_Z(z) = \Gamma\left(\frac{m+n}{2}\right)\left[\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)\right]^{-1}\left(\frac{m}{n}\right)^{\frac{m}{2}}\frac{z^{\frac{m}{2}-1}}{\left(1+\frac{m}{n}z\right)^{\frac{m+n}{2}}},\ z>0. \qquad (5.11)$$

A random variable Z is said to have an F distribution, with m degrees of freedom on the numerator and n degrees of freedom in the denominator, notated as $Z \stackrel{d}{=} F^m_n$, if its density function is given by (5.11). ▲

The transformation theorem can be generalized in the following way:

Theorem 5.12 *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ *be a random vector with joint probability density function* $f_{\mathbf{X}}$. *Let* g *be a map of* $\mathbb{R}^n$ *into itself. Suppose* *that $\mathbb{R}^n$ can be partitioned in k disjoint sets $A_1, \cdots, A_k$ in such a way that the map g restricted to A_i for $i = 1, \cdots, k$ is a one-to-one map with inverse h_i. If the partial derivatives of h_i exist and are continuous and if the Jacobians J_i are not zero on the range of the transformation for $i = 1, \cdots, k$, then the random vector $\mathbf{Y} := g(\mathbf{X})$ has joint probability density function given by:*

$$f_{\mathbf{Y}}(\mathbf{y}) = \sum_{i=1}^{k} |J_i(\mathbf{y})| \, f_{\mathbf{X}}(h_i(\mathbf{y})) \, .$$

Proof: Interested reader may refer to Jacod and Protter (2004) for the above theorem. ■

■ EXAMPLE 5.30

Let X be a random variable with density function f_X. Let $g : \mathbb{R} \longrightarrow \mathbb{R}$ be given by $g(x) = x^4$. Clearly $\mathbb{R} = (-\infty, 0) \cup [0, \infty)$ and the maps

$$\begin{array}{rccc} g_1 : & [0,\infty) & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & x^4 \end{array}$$

$$\begin{array}{rccc} g_2 : & (-\infty,0) & \longrightarrow & \mathbb{R} \\ & x & \longmapsto & x^4 \end{array}$$

have respectively the inverses

$$\begin{array}{rccc} h_1 : & [0,\infty) & \longrightarrow & [0,\infty) \\ & x & \longmapsto & \sqrt[4]{x} \end{array}$$

$$\begin{array}{rccc} h_2 : & (0,\infty) & \longrightarrow & (-\infty,0) \\ & x & \longmapsto & -\sqrt[4]{x} \, . \end{array}$$

The Jacobians of the inverse transforms are given, respectively, by:

$$J_1(x) = \frac{1}{4} x^{-\frac{3}{4}} \qquad \text{and} \qquad J_2(x) = -\frac{1}{4} x^{-\frac{3}{4}}.$$

Therefore, the density function of the random variable $Y = X^4$ is given by:

$$\begin{aligned} f_Y(y) &= \sum_{i=1}^{2} |J_i(y)| \, f_X(h_i(y)) \\ &= \frac{1}{4} y^{-\frac{3}{4}} f_X(\sqrt[4]{y}) + \frac{1}{4} y^{-\frac{3}{4}} f_X(-\sqrt[4]{y}). \end{aligned} \qquad \blacktriangle$$

5.4 COVARIANCE AND CORRELATION COEFFICIENT

Next, the expected value of a function of an n-dimensional random variable is defined.

Definition 5.8 (Expected Value of a Function of a Random Vector) *Let $(X_1, X_2, \cdots, X_n)$ be an n-dimensional random vector and let $g(\cdot, \cdots, \cdot)$ be a real-valued function defined over $\mathbb{R}^n$. The expected value of the function $g(X_1, \cdots, X_n)$, notated as $E(g(X_1, \cdots, X_n))$, is defined as*

$$E(g(X_1, \cdots, X_n)) :=$$
$$\begin{cases} \sum_{x_1} \cdots \sum_{x_n} g(x_1, \cdots, x_n) P(X_1 = x_1, \cdots, X_n = x_n), \\ \qquad\qquad\qquad\qquad \textit{discrete random variables} \\ \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \cdots, x_n) f(x_1, \cdots, x_n)\, dx_1 \cdots dx_n, \\ \qquad\qquad\qquad\qquad \textit{continuous random variables} \end{cases}$$

provided that the multiple summation in the discrete case or the multiple integral in the continuous case converges absolutely.

■ EXAMPLE 5.31

Suppose that a fair dice is rolled twice in a row. Let $X :=$"the maximum value obtained" and $Y :=$"the sum of the results obtained". In this case, $E(XY) = \frac{1232}{36}$. ▲

■ EXAMPLE 5.32

Let (X, Y, Z) be a three-dimensional random vector with joint probability density function given by:

$$f(x, y, z) = \begin{cases} 8\,x\,y\,z & \text{if } 0 < x < 1;\ 0 < y < 1 \text{ and } 0 < z < 1 \\ 0 & \text{otherwise}. \end{cases}$$

Then $E(5X - 2Y + Z) = \frac{8}{3}$ and $E(XY) = \frac{4}{9}$. ▲

Theorem 5.13 *If X and Y are random variables whose expected values exist, then the expected value of $X + Y$ also exists and equals the sum of the expected values of X and Y.*

Proof: We prove the theorem for the continuous case. The discrete case can be treated analogously.

Suppose that f is the joint probability density function of X and Y. Then:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |x + y|\, f(x, y)\, dx\, dy \le \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (|x| + |y|)\, f(x, y)\, dx\, dy$$

$$\begin{aligned}
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x|\, f(x,y)\, dx\, dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |y|\, f(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} |x| \left(\int_{-\infty}^{\infty} f(x,y)\, dy\right) dx + \int_{-\infty}^{\infty} |y| \left(\int_{-\infty}^{\infty} f(x,y)\, dx\right) dy \\
&= \int_{-\infty}^{\infty} |x|\, f_X(x)\, dx + \int_{-\infty}^{\infty} |y|\, f_Y(y)\, dy < \infty\ .
\end{aligned}$$

Therefore, $E(X+Y)$ exists.

Furthermore:

$$\begin{aligned}
E(X+Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (x+y) f(x,y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} x \left(\int_{-\infty}^{\infty} f(x,y)\, dy\right) dx + \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f(x,y)\, dx\right) dy \\
&= \int_{-\infty}^{\infty} x f_X(x)\, dx + \int_{-\infty}^{\infty} y f_Y(y)\, dy \\
&= E(X) + E(Y)\ .
\end{aligned}$$

■

Note 5.11 *If X is a discrete random variable and Y is a continuous random variable, the previous result still holds.*

In general we have the following:

Theorem 5.14 *If $X_1, X_2, \cdots, X_n$ are n random variables whose expected values exist, then the expected value of the sum of the random variables exists as well and equals the sum of the expected values.*

Proof: Left as an exercise for the reader. ■

■ EXAMPLE 5.33

An urn contains N balls, R of which are red-colored while the remaining $N-R$ are white-colored. A sample of size n is extracted without replacement. Let X be the number of red-colored balls in the sample. From the theory described in Section 3.2, it is known that X has a hypergeometric distribution of parameters n, R and N; therefore $E(X) = \frac{nR}{N}$. We are now going to deduce the same result by writing X as the sum of random variables and then applying the last theorem.

Let X_i, with $i = 1, \cdots, n$, be the random variables defined as follows:

$$X_i := \begin{cases} 1 & \text{if the } i\text{th ball extracted is red} \\ 0 & \text{if the } i\text{th ball extracted is white.} \end{cases}$$

Clearly:

$$X = \sum_{i=1}^{n} X_i.$$

Accordingly:

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} \frac{R}{N} = \frac{nR}{N}. \qquad \blacktriangle$$

Theorem 5.15 *If X and Y are independent random variables whose expected values exist, then the expected value of XY also exists and equals the product of the expected values.*

Proof: We prove the theorem for the continuous case. The discrete case can be treated analogously.

Suppose that f is the joint probability density function of X and Y. Then:

$$\begin{aligned}
\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |xy|\, f(x,y)\, dx\, dy &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x|\,|y|\, f_X(x) f_Y(y)\, dx\, dy \\
&= \int_{-\infty}^{\infty} |x| \left(\int_{-\infty}^{\infty} |y|\, f_Y(y)\, dy \right) f_X(x)\, dx \\
&= \left(\int_{-\infty}^{\infty} |x|\, f_X(x)\, dx \right) \left(\int_{-\infty}^{\infty} |y|\, f_Y(y)\, dy \right) < \infty .
\end{aligned}$$

That is, $E(XY)$ exists.

Suppressing the absolute-value bars on the previous proof, we obtain:

$$E(XY) = E(X)E(Y).$$

$\blacksquare$

Two random variables can be independent or be closely related to each other. It is possible, for example, that the random variable Y will increase as a result of an increment of the random variable X or that Y will increase as X decreases. The following quantities, known as the covariance and the correlation coefficient, allow us to determine if there is a linear relationship of this type between the random variables under consideration.

Definition 5.9 (Covariance) *Let X and Y be random variables defined over the same probability space and such that $E(X^2) < \infty$ and $E(Y^2) < \infty$. The covariance between X and Y is defined by:*

$$Cov(X,Y) := E\left((X - E(X))(Y - E(Y))\right).$$

Note 5.12 *Let X and Y be random variables defined over the same probability space and such that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Since $|X| \leq 1+X^2$, then* *it follows that $E(X)$ exists. On the other hand, $|XY| \leq X^2 + Y^2$ implies the existence of the expected value of the random variable $(X - E(Y))(Y - E(Y))$.*

Theorem 5.16 *Let X and Y be random variables defined over the same probability space and such that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then:*

(i) $Cov(X,Y) = E(XY) - E(X)E(Y)$.

(ii) $Cov(X,Y) = Cov(Y,X)$.

(iii) $VarX = Cov(X,X)$.

(iv) $Cov(aX+b,Y) = a\,Cov(X,Y)$ *for any* $a, b \in \mathbb{R}$.

Proof: We prove (iv) and leave the others as exercises for the reader:

$$\begin{aligned} Cov(aX+b,Y) &= E\left(\left((aX+b) - E(aX+b)\right)\left(Y - E(Y)\right)\right) \\ &= E\big((aX + b - aE(X) - b)(Y - E(Y))\big) \\ &= a\,Cov(X,Y). \end{aligned}$$

■

Note 5.13 *From the first property above, it follows that if X and Y are independent, then $Cov(X,Y) = 0$. The converse, however, does not always hold, as the next example illustrates.*

■ EXAMPLE 5.34

Let $\Omega = \{1,2,3,4\}$, $\Im = \wp(\Omega)$ and let $P : \Im \to [0,1]$ be given by $P(1) = P(2) := \frac{2}{5}$, $P(3) = P(4) := \frac{1}{10}$. The random variables X and Y defined by $X(1) = Y(2) := 1$, $X(2) = Y(1) := -1$, $X(3) = Y(3) := 2$ and $X(4) = Y(4) := -2$ are not independent in spite of the fact that $Cov(X,Y) = 0$. ▲

Theorem 5.17 (Cauchy-Schwarz Inequality) *Let X and Y be random variables such that $E(X^2) < \infty$ and $E(Y^2) < \infty$. Then $|E(XY)|^2 \leq (E(X^2))(E(Y^2))$. Furthermore, the equality holds if and only if there are real constants a and b, not both simultaneously zero, such that $P(aX + bY = 0) = 1$.*

Proof: Let $\alpha = E(Y^2)$ and $\beta = -E(XY)$. Clearly $\alpha \geq 0$. Since the result is trivially true when $\alpha = 0$, let us consider the case when $\alpha > 0$. We have:

$$\begin{aligned} 0 \leq E((\alpha X + \beta Y)^2) &= E(\alpha^2 X^2 + 2\alpha\beta XY + \beta^2 Y^2) \\ &= \alpha(E(X^2)E(Y^2) - E(XY)E(XY)). \end{aligned}$$

Since $\alpha > 0$, the result follows.

If $(EXY)^2 = E(X^2)E(Y^2)$, then $E(\alpha X + \beta Y)^2 = 0$. Therefore, with probability 1, we have that $(\alpha X + \beta Y) = 0$. If $\alpha > 0$, we can take $a = \alpha$ and $b = \beta$. If $\alpha = 0$, then we can take $a = 0$ and $b = 1$.

Conversely, if there are real numbers a and b not both of them zero such that with probability 1 $(aX + bY) = 0$, then, $aX = -bY$ with probability 1, and in that case it can be easily verified that $|E(XY)|^2 = E(X^2)E(Y^2)$. ∎

Note 5.14 *Taking $|X|$ and $|Y|$ instead of X and Y in the previous theorem yields:*

$$E|XY| \leq \sqrt{E(X^2)}\sqrt{E(Y^2)}\ .$$

Applying this last result to the random variables $X - E(X)$ and $Y - E(Y)$ we obtain $|Cov(X,Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}$.

Theorem 5.18 *If X and Y are real-valued random variables having finite variances, then $Var(X+Y) < \infty$ and we have:*

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y). \tag{5.12}$$

Proof: To see that $Var(X+Y) < \infty$ it suffices to verify that:

$$E(X+Y)^2 < \infty\ .$$

To this effect:

$$\begin{aligned} E(X+Y)^2 &= E(X^2) + 2E(XY) + E(Y^2) \\ &\leq E(X^2) + 2E|XY| + E(Y^2) \\ &\leq 2(E(X^2) + E(Y^2)) < \infty. \end{aligned}$$

Applying the properties of the expected value, we arrive at (5.12). ∎

In general we have:

Theorem 5.19 *If $X_1, X_2, \cdots, X_n$ are n random variables having finite variances, then $Var\left(\sum_{i=1}^{n} X_i\right) < \infty$ and:*

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + \sum\sum_{i \neq j} Cov(X_i, X_j).$$

Proof: Left as an exercise for the reader. ∎

Note 5.15 *Since each pair of indices i, j with $i \neq j$ appears twice in the previous summation, it is equivalent to:*

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum\sum_{i<j} Cov(X_i, X_j).$$

If $X_1, X_2, \cdots, X_n$ are n independent random variables with finite variances, then:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) \ .$$

■ EXAMPLE 5.35

A person is shown n pictures of babies, all corresponding to well-known celebrities, and then asked to whom they belong. Let X be the random variable representing the number of correct answers. Find $E(X)$ and $Var(X)$.

Solution: Let:

$$X_i := \begin{cases} 1 & \text{if the person correctly identifies the } i\text{th picture} \\ 0 & \text{otherwise.} \end{cases}$$

Then:

$$X = \sum_{i=1}^{n} X_i.$$

Hence:

$$E(X) = \sum_{i=1}^{n} E(X_i) = \sum_{i=1}^{n} P(X_i = 1) = n\frac{1}{n} = 1 \ .$$

On the other hand, we have that:

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i<j}\sum Cov(X_i, X_j).$$

Noting that

$$Var(X_i) = E(X_i^2) - (E(X_i))^2 = \frac{1}{n} - \left(\frac{1}{n}\right)^2 = \frac{1}{n}\left(\frac{n-1}{n}\right)$$

and

$$\begin{aligned} E(X_iX_j) &= P(X_i = 1, X_j = 1) \\ &= P(X_j = 1 \mid X_i = 1)P(X_i = 1) \\ &= \frac{1}{n-1} \times \frac{1}{n}, \end{aligned}$$

it follows that

$$\begin{aligned} Cov(X_i, X_j) &= E(X_iX_j) - (E(X_i))(E(X_j)) \\ &= \frac{1}{n-1} \times \frac{1}{n} - \frac{1}{n^2} \\ &= \frac{1}{n}\left[\frac{1}{n(n-1)}\right] . \end{aligned}$$

Thus:

$$Var(X) = \frac{n(n-1)}{n^2} + 2\binom{n}{2}\frac{1}{n^2(n-1)} = \frac{n-1}{n} + \frac{1}{n} = 1 . \qquad \blacktriangle$$

The covariance is a measure of the linear association between two random variables. A "high" covariance means that with probability 1 there is a linear relationship between the two variables. But what exactly does it mean that the covariance is "high"? How can we qualify the magnitude of the covariance? In fact, property (iv) of the covariance shows that its magnitude depends on the measure scale used. For this reason, it is difficult in concrete cases to determine by inspection if the covariance between two random variables is "high" or not. In order to get rid of this complication, the English mathematician Karl Pearson, who developed most of the modern statistical techniques, introduced the following concept:

Definition 5.10 (Correlation Coefficient) *Let X and Y be real-valued random variables with $0 < Var(X) < \infty$ and $0 < Var(Y) < \infty$. The correlation coefficient between X and Y is defined as follows:*

$$\rho(X,Y) := \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} .$$

Theorem 5.20 *Let X and Y be real-valued random variables with $0 < Var(X) < \infty$ and $0 < Var(Y) < \infty$.*

(i) $\rho(X,Y) = \rho(Y,X)$.

(ii) $|\rho(X,Y)| \leq 1$.

(iii) $\rho(X,X) = 1$ *and* $\rho(X,-X) = -1$.

(iv) $\rho(aX+b,Y) = \rho(X,Y)$ *for any $a,b \in \mathbb{R}$ with $a > 0$.*

(v) $|\rho(X;Y)| = 1$ *if and only if there are constants $a,b \in \mathbb{R}$ not both of them zero and $c \in \mathbb{R}$ such that $P(aX+bY=c) = 1$.*

Proof: We prove (ii) and (v) and leave the others as exercise for the reader.

(ii) Let:

$$X^* := \frac{X - E(X)}{\sqrt{Var(X)}}$$

$$Y^* := \frac{Y - E(Y)}{\sqrt{Var(Y)}} .$$

Clearly $E(X^*) = E(Y^*) = 0$ and $Var(X^*) = Var(Y^*) = 1$. Therefore

$$\begin{aligned}\rho(X^*, Y^*) &= E(X^*Y^*)\\ &= \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}}\\ &= \rho(X, Y)\end{aligned}$$

and

$$0 \le Var(X^* \pm Y^*) = Var(X^*) \pm 2Cov(X^*, Y^*) + Var(Y^*) = 2(1 \pm \rho(X, Y)).$$

Hence, it follows that:

$$|\rho(X, Y)| \le 1 .$$

(v) Let X^* and Y^* be defined as in (ii). Clearly

$$\begin{aligned}\rho(X, Y) = 1 &\iff \rho(X^*, Y^*) = 1\\ &\iff (E(X^*Y^*))^2 = E(X^*)^2\, E(Y^*)^2\\ &\iff \text{There exist } \alpha, \beta \text{ where } \alpha \neq 0 \text{ or } \beta \neq 0\\ &\qquad \text{such that } P(\alpha X^* + \beta Y^* = 0) = 1\\ &\iff P\left(\alpha\left(\frac{X - E(X)}{\sqrt{Var(X)}}\right) + \beta\left(\frac{Y - E(Y)}{\sqrt{Var(Y)}}\right) = 0\right) = 1\\ &\iff P(aX + bY = c) = 1,\end{aligned}$$

where:

$$a = \frac{\alpha}{\sqrt{Var(X)}}, \quad b = \frac{\beta}{\sqrt{Var(Y)}} \quad \text{and } c = \frac{\alpha E(X)}{\sqrt{Var(X)}} + \frac{\beta E(Y)}{\sqrt{Var(Y)}} .$$

The case $\rho(X, Y) = -1$ can be handled analogously.

Part (v) above indicates that if $|\rho(X, Y)| \approx 1$, then $Y(\omega) \approx aX(\omega) + b$ for all $\omega \in \Omega$. In practice, a "high" absolute-value correlation coefficient indicates that Y can be predicted from X and vice versa.

∎

5.5 EXPECTED VALUE OF A RANDOM VECTOR AND VARIANCE-COVARIANCE MATRIX

In this section we generalize the concepts of expected value and variance of a random variable to a random vector.

Definition 5.11 (Expected Value of a Random Vector) *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ *be a random vector. The expected value (or expectation) of* $\mathbf{X}$*, notated as* $E(\mathbf{X})$*, is defined as*

$$E(\mathbf{X}) := (E(X_1), E(X_2), \cdots, E(X_n))$$

subject to the existence of $E(X_j)$ for all $j = 1, \cdots, n$.

This definition can be extended even further as follows:

Definition 5.12 (Expected Value of a Function of a Random Vector) *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a random vector and let $h : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the function given by*

$$h(x_1, \cdots, x_n) = (h_1(x_1, \cdots, x_n), \cdots, h_m(x_1, \cdots, x_n))$$

where h_i, for $i = 1, \cdots, m$, are real-valued functions defined over $\mathbb{R}^n$. The expected value of $h(\mathbf{X})$ is given by

$$E\big(h(\mathbf{X})\big) := \big(E\big(h_1(\mathbf{X})\big), \cdots, E\big(h_m(\mathbf{X})\big)\big)$$

subject to the existence of $E(h_j(\mathbf{X}))$ for all $j = 1, \cdots, m$.

Definition 5.13 (Expected Value of a Random Matrix) *If X_{ij}, with $i = 1, \cdots, m$ and $j = 1, \cdots, n$, are real-valued random variables defined over the same probability space, then the matrix $\mathbf{A} = (X_{ij})_{m\times n}$ is called a random matrix, and its expected value is defined as the matrix whose entries correspond to the expectations of the random variables X_{ij}, that is*

$$E(\mathbf{A}) := (E(X_{ij}))_{m\times n}$$

subject to the existence of $E(X_{ij})$ for all $i = 1, \cdots, m$ and all $j = 1, \cdots, n$.

Definition 5.14 (Variance-Covariance Matrix) *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a random vector such that $E(X_j^2) < \infty$ for all $j = 1, \cdots, n$. The variance-covariance matrix, notated $\sum$, of $\mathbf{X}$ is defined as follows:*

$$\sum := \begin{pmatrix} Var(X_1) & Cov(X_1, X_2) & \cdots & Cov(X_1, X_n) \\ Cov(X_2, X_1) & Var(X_2) & \cdots & Cov(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & Cov(X_n, X_2) & \cdots & Var(X_n) \end{pmatrix}.$$

Note 5.16 *Observe that* $\sum = E\big([\mathbf{X}-E(\mathbf{X})]^T\,[\mathbf{X}-E(\mathbf{X})]\big)$.

■ EXAMPLE 5.36

Let X_1 and X_2 be discrete random variables having the joint distribution given by:

$X_1 \backslash X_2$	-1	0	1
1	$\frac{1}{6}$	$\frac{2}{6}$	0
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Clearly, the expected value of $\mathbf{X} = (X_1, X_2)$ equals

$$E(\mathbf{X}) = \left(\frac{3}{2}, -\frac{1}{6}\right)$$

and the variance-covariance matrix is given by:

$$\sum = \begin{pmatrix} \frac{1}{4} & \frac{1}{12} \\ \frac{1}{12} & \frac{17}{36} \end{pmatrix} .$$

For any $a = (a_1, a_2)$ we see that:

$$\begin{aligned} a \sum a^T &= (a_1, a_2) \begin{pmatrix} \frac{1}{4} & \frac{1}{12} \\ \frac{1}{12} & \frac{17}{36} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\ &= \frac{1}{4} a_1^2 + \frac{1}{6} a_1 a_2 + \frac{17}{36} a_2^2 \\ &= \left(\frac{1}{2} a_1 + \frac{1}{6} a_2\right)^2 + \frac{4}{9} a_2^2 \geq 0 . \end{aligned}$$

That is, the matrix $\sum$ is positive semidefinite. ▲

■ EXAMPLE 5.37

Let $\mathbf{X} = (X, Y)$ be a random vector having the joint probability density function given by:

$$f(x, y) = \begin{cases} \frac{1}{x} & \text{if } 0 < x \leq 1,\ 0 < y \leq x \\ 0 & \text{otherwise} . \end{cases}$$

In this case we have:

$$\begin{aligned} E(XY) &= \int_0^1 \int_0^x xy \frac{1}{x} \, dy \, dx = \frac{1}{6} \\ E(X) &= \int_0^1 \int_y^1 x \frac{1}{x} \, dx \, dy = \frac{1}{2} \\ E(Y) &= \int_0^1 \int_0^x y \frac{1}{x} \, dy \, dx = \frac{1}{4} \\ E(X^2) &= \int_0^1 \int_0^x x^2 \frac{1}{x} \, dy \, dx = \frac{1}{3} \\ E(Y^2) &= \int_0^1 \int_0^x y^2 \frac{1}{x} \, dy \, dx = \frac{1}{9} . \end{aligned}$$

Therefore, the expected value of $\mathbf{X}$ is given by

$$E(\mathbf{X}) = \left(\frac{1}{2}, \frac{1}{4}\right)$$

and its variance-covariance matrix equals:

$$\sum = \begin{pmatrix} \frac{1}{12} & \frac{1}{24} \\ \frac{1}{24} & \frac{7}{144} \end{pmatrix} .$$

And for any $a = (a_1, a_2) \in \mathbb{R}^2$ we have:

$$\begin{aligned} a \sum a^T &= (a_1, a_2) \begin{pmatrix} \frac{1}{12} & \frac{1}{24} \\ \frac{1}{24} & \frac{7}{144} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} \\ &= \frac{1}{12}\left(a_1^2 + 2\frac{1}{2}a_1a_2 + \frac{1}{4}a_2^2 - \frac{1}{4}a_2^2 + \frac{7}{12}a_2^2\right) \\ &= \frac{1}{12}\left(a_1 + \frac{1}{2}a_2\right)^2 + \frac{1}{36}a_2^2 \geq 0 . \end{aligned}$$

That is, the matrix $\sum$ is positive semidefinite. ▲

In general, we have the following result:

Theorem 5.21 *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector. If $E(X_j^2) < \infty$ for all $j = 1, \cdots, n$, then the variance-covariance matrix $\sum$ of $\mathbf{X}$ is positive semidefinite.*

Proof: Let $a = (a_1, \cdots, a_n)$ be any vector in $\mathbb{R}^n$. Consider the random variable Y defined as follows:

$$\begin{aligned} Y &:= (X_1, \cdots, X_n)(a_1, \cdots, a_n)^T \\ &= \sum_{i=1}^{n} a_i X_i . \end{aligned}$$

Since $Var(Y) \geq 0$, it suffices to verify that $Var(Y) = a \sum a^T$.
Indeed:

$$Var(Y) = E\left([Y - E(Y)]^2\right) .$$

Seeing that, $E(Y) = \sum_{i=1}^{n} a_i E(X_i) = \mu a^T$, where $\mu := E(\mathbf{X})$, we have:

$$\begin{aligned} Var(Y) &= E\left([\mathbf{X}a^T - \mu a^T]\,[\mathbf{X}a^T - \mu a^T]\right) \\ &= E\left([a\mathbf{X}^T - a\mu^T]\,[\mathbf{X}a^T - \mu a^T]\right) \\ &= E\left(a[\mathbf{X} - \mu]^T[X - \mu]a^T\right) \\ &= a\,E\left([\mathbf{X} - \mu]^T[X - \mu]\right)a^T \\ &= a \sum a^T . \end{aligned}$$

■

Note 5.17 *Clearly, from the definition of the variance-covariance matrix $\sum$, if the random variables $X_1, X_2, \cdots, X_n$ are independent, then $\sum$ is a matrix whose diagonal elements correspond to the variances of the random variables X_j for $j = 1, \cdots, n$.*

We end this section by introducing the correlation matrix, which plays an important role in the development of the theory of multivariate statistics.

Definition 5.15 (Correlation Matrix) *Let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a random vector with $0 < Var(X_j) < \infty$ for all $j = 1, \cdots, n$. The correlation matrix of $\mathbf{X}$, notated as $\mathbf{R}$, is defined as follows:*

$$\mathbf{R} := \begin{pmatrix} 1 & \rho(X_1, X_2) & \cdots & \rho(X_1, X_n) \\ \rho(X_2, X_1) & 1 & \cdots & \rho(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \rho(X_n, X_1) & \rho(X_n, X_2) & \cdots & 1 \end{pmatrix} .$$

■ **EXAMPLE 5.38**

Let $\mathbf{X} = (X, Y)$ be a two-dimensional random vector with joint probability density function given by:

$$f(x, y) = \begin{cases} 2 & \text{if } 0 < x < y, \ 0 < y < 1 \\ 0 & \text{otherwise} . \end{cases}$$

We have

$$E(X) = \int_0^1 \int_0^y 2x \, dx \, dy = \frac{1}{3}$$
$$E(Y) = \int_0^1 \int_x^1 2y \, dy \, dx = \frac{2}{3}$$

and

$$Var(X) = \int_0^1 \int_0^y 2(x - \frac{1}{3})^2 \, dx \, dy = \frac{1}{18}$$
$$Var(Y) = \int_0^1 \int_x^1 2(y - \frac{2}{3})^2 \, dy \, dx = \frac{1}{18}$$
$$Cov(X, Y) = \int_0^1 \int_0^y 2(x - \frac{1}{3})(y - \frac{2}{3}) \, dx \, dy = \frac{1}{36} .$$

Therefore the correlation matrix is given by:

$$\mathbf{R} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}. \quad \blacktriangle$$

Note 5.18 *The correlation matrix* $\mathbf{R}$ *inherits all the properties of the variance-covariance matrix* $\sum$ *because*

$$\mathbf{R} = diag\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_n}\right) \cdot \sum \cdot diag\left(\frac{1}{\sigma_1}, \cdots, \frac{1}{\sigma_n}\right)$$

where $\sigma_j := \sqrt{VarX_j}$ *for* $j = 1, \cdots, n$.

Therefore, $\mathbf{R}$ *is symmetric and positive semidefinite. Furthermore, if* $\sigma_j > 0$ *for all* $j = 1, \cdots, n$*, then* $\mathbf{R}$ *is nonsingular if and only if* $\sum$ *is nonsingular.*

5.6 JOINT PROBABILITY GENERATING, MOMENT GENERATING AND CHARACTERISTIC FUNCTIONS

This section is devoted to the generalization of the concepts of probability generating, moment generating and characteristic functions, introduced in Chapter 2 for the one-dimensional case, to n-dimensional random vectors.

Now, we define the joint probability generating function for n-dimensional random vectors.

Definition 5.16 *Let* $X = (X_1, X_2, \cdots, X_n)$ *be an* n*-dimensional random vector with joint probability mass function* $P_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n)$*. Then the* pgf *of a random vector is defined as*

$$G_X(s_1, s_2, \cdots, s_n) = \sum_{x_1, x_2, \cdots, x_n} P_{X_1,X_2,\cdots,X_n}(x_1, x_2, \cdots, x_n) s_1^{x_1} \cdots s_n^{x_n}$$

for $|s_1| \le 1, |s_2| \le 1, \cdots, |s_n| \le 1$ *provided that the series is convergent.*

Theorem 5.22 *The* pgf *of the marginal distribution of* X_i *is given by*

$$G_{X_i} = G_X(1, 1, \cdots, s_i, \cdots, 1)$$

and the pgf *of* $X_1 + X_2 + \cdots + X_n$ *is given by*

$$H(s) = G_X(s, s, \cdots, s).$$

Proof: Left as an exercise. ∎

Now, we present the pgf for the sum of independent random variables.

Theorem 5.23 *Let X and Y be independent nonnegative integer-valued random variables with pgf's $P_X(s)$ and $P_Y(s)$, respectively, and let $Z = X + Y$. Then:*

$$G_Z(s) = G_{X+Y}(s) = G_X(s) \cdot G_Y(s).$$

Proof:

$$\begin{aligned} G_Z(s) &= E[s^Z] \\ &= E[s^{X+Y}] \\ &= E[s^X] \cdot E[s^Y] \quad \text{given that } X \text{ and } Y \text{ are independent} \\ &= G_X(s) \cdot G_Y(s) \ . \end{aligned}$$

■

Corollary 5.1 *If $X_1, X_2, \cdots, X_n$ are independent nonnegative integer-valued random variables with pgf's $G_{X_1}(s), \cdots, G_{X_n}(s)$ respectively, then:*

$$\begin{aligned} G_{X_1+\cdots+X_n}(s) &= G_{X_1}(s), \cdots, G_{X_n}(s) \\ &= \prod_{i=1}^{n} G_{X_i}(s) \ . \end{aligned}$$

■ **EXAMPLE 5.39**

Find the distribution of the sum of n independent random variables X_i, $i = 1, 2, \cdots, n$, where $X_i \stackrel{d}{=} \text{Poisson}(\lambda)$.
Solution: $G_{X_i}(s) = e^{\lambda_i(s-1)}$. So

$$\begin{aligned} G_{X_1+\cdots+X_n}(s) &= \prod_{i=1}^{n} e^{\lambda_i(s-1)} \\ &= e^{(\lambda_1+\cdots+\lambda_n)(s-1)} \ . \end{aligned}$$

This means that:

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}(\lambda_1 + \cdots + \lambda_n)$$

$$\sum_{i=1}^{n} X_i \sim \text{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right) \ . \qquad \blacktriangle$$

If $X_1, X_2, \cdots, X_n$ are i.i.d. random variables, then:

$$G_{X_1+X_2+\cdots+X_n}(s) = (G_{X_i}(s))^n \ .$$

Theorem 5.24 *Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables with pgf $P(s)$. We consider the sum $S_N = X_1 + X_2 + \cdots + X_N$ where N is a discrete random variable independent of the X_i's with distribution given by $P(N = n) = g_n$. The pgf of N is $G(s) = \sum_n g_n s^n \quad |s| < 1$.*

We prove that the pgf of S_N is $H(s) = G(P(s))$.

Proof:

$$\begin{aligned}
H(s) &= E[s^{S_N}] \\
&= \sum_{n=0}^{\infty} E[s^{S_N} | N = n] P(N = n) \\
&= \sum_{n=0}^{\infty} E[s^{S_n}] P(N = n) \\
&= \sum_{n=0}^{\infty} (P(s))^n \, P(N = n) \\
&= \sum_{n=0}^{\infty} g_n \, (P(s))^n \\
&= G\left(P(s)\right) \ .
\end{aligned}$$

■

From the above result, we immediately obtain:

1. $E(S_N) = E(N)E(X)$.
2. $Var(S_N) = E(N)Var(X) + Var(N)[E(X)]^2$.

Now we generalize the concept of mgf for random vectors.

Definition 5.17 (Joint Moment Generating Function) *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector. If there exist $M > 0$ such that $E\big(\exp(\mathbf{X}\mathbf{t}^T)\big)$ is finite for all $\mathbf{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n$ with*

$$\|\mathbf{t}\| := \sqrt{t_1^2 + \cdots + t_n^2} < M,$$

then the joint moment generating function of $\mathbf{X}$, notated as $m_{\mathbf{X}}(\mathbf{t})$, is defined as:

$$m_{\mathbf{X}}(\mathbf{t}) := E\big(\exp(\mathbf{X}\mathbf{t}^T)\big) \quad \textit{for} \;\; \|\mathbf{t}\| < M \ .$$

■ **EXAMPLE 5.40**

Let X_1 and X_2 be discrete random variables with joint distribution given by:

$X_1 \backslash X_2$	-1	0	1
1	$\frac{1}{6}$	$\frac{2}{6}$	0
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

The joint moment generating function of $\mathbf{X} = (X_1, X_2)$ is given by:

$$\begin{aligned} m_{\mathbf{X}}(\mathbf{t}) &= E\left(\exp((X_1, X_2)(t_1, t_2)^T)\right) \\ &= E\big(\exp(X_1 t_1 + X_2 t_2)\big) \\ &= \sum_{x_1}\sum_{x_2} \exp\big(t_1 x_1 + t_2 x_2\big) P(X_1 = x_1, X_2 = x_2) \\ &= \frac{1}{6}\exp(t_1 - t_2) + \frac{1}{6}\exp(2t_1 - t_2) + \frac{1}{3}\exp(t_1) \\ &\qquad + \frac{1}{6}\exp(2t_1) + \frac{1}{6}\exp(2t_1 + t_2)\ . \end{aligned}$$

Notice that the moment generating functions of X_1 and X_2 also exist and are given, respectively, by

$$\begin{aligned} m_{X_1}(t_1) &= E(\exp(t_1 X_1)) \\ &= \sum_{x_1} \exp(t_1 x_1) P(X_1 = x_1) \\ &= \frac{1}{2}\exp(t_1) + \frac{1}{2}\exp(2t_1) \end{aligned}$$

and

$$\begin{aligned} m_{X_2}(t_2) &= E(\exp(t_2 X_2)) \\ &= \sum_{x_2} \exp(t_2 x_2) P(X_2 = x_2) \\ &= \frac{1}{3}\exp(-t_2) + \frac{1}{6}\exp(t_2) + \frac{1}{2}\ . \qquad \blacktriangle \end{aligned}$$

In general we have the following result:

Theorem 5.25 *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector. The joint moment generating function of $\mathbf{X}$ exists if and only if the marginal moment generating functions of the random variables X_i for $i = 1, \cdots, n$ exist.*

Proof:

$\Longrightarrow$) Suppose initially that the joint moment generating function of $\mathbf{X}$ exists. In that case, there is a constant $M > 0$ such that $m_{\mathbf{X}}(\mathbf{t}) =$

$E(\exp(\mathbf{X}\mathbf{t}^T)) < \infty$ for all $\mathbf{t}$ with $\|\mathbf{t}\| < M$. Then, for all $i = 1, \cdots, n$ we have:

$$\begin{aligned} m_{X_i}(t_i) &= E(\exp(t_i X_i)) \\ &= E\Big(\exp\big(\mathbf{X} \cdots (0, 0, \cdots, \underset{\substack{\uparrow \\ i\text{th} \\ \text{position}}}{t_i}, 0, \cdots, 0)^T\big)\Big) \\ &= m_{\mathbf{X}}((0, \cdots, t_i, \cdots, 0)^T) \\ &< \infty \quad \text{whenever } |t_i| < M . \end{aligned}$$

That is, the moment generating function of X_i for $i = 1, \cdots, n$ exist.

$\Longleftarrow$) Suppose now that the moment generating functions of the random variables X_i for $i = 1, \cdots, n$ exist. Then, for each $i = 1, \cdots, n$ there exist $M_i > 0$ such that $m_{X_i}(t_i) = E(\exp(t_i X_i)) < \infty$ whenever $|t_i| < M_i$. Let $\mathbf{h} : \mathbb{R}^n \longrightarrow \mathbb{R}$ be defined by:

$$\mathbf{h}(\mathbf{t}) := \exp\left(\mathbf{t} \cdot \mathbf{a}^T\right) \quad \text{for } \mathbf{a} \in \mathbb{R}^n$$

The function $\mathbf{h}$ so defined is convex, and consequently, if $\mathbf{x}_i \in \mathbb{R}^n$, for $i = 1, \cdots, m$, and we choose $\alpha_i \in (0, 1)$ for $i = 1, \cdots, m$ such that $\sum\limits_{i=1}^{m} \alpha_i = 1$, then, we must have that:

$$\mathbf{h}\left(\sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i\right) \le \sum_{i=1}^{m} \alpha_i \, \mathbf{h}(\mathbf{x}_i) .$$

Therefore:

$$\exp\left\{\left(\sum_{i=1}^{m} \alpha_i \, \mathbf{x}_i\right) \mathbf{a}^T\right\} \le \sum_{i=1}^{m} \alpha_i \exp\left(\mathbf{x}_i \, \mathbf{a}^T\right) .$$

In particular, for $\mathbf{a} = \mathbf{X} = (X_1, \cdots, X_n)$, $\mathbf{x}_i = (0, \cdots, 0, \underset{\substack{\uparrow \\ i\text{th} \\ \text{position}}}{t_i}, 0, \cdots, 0)$ and $n = m$, the preceding expression yields:

$$\exp\left\{\sum_{i=1}^{n} \alpha_i \, t_i \, X_i\right\} \le \sum_{i=1}^{n} \alpha_i \exp(t_i \, X_i) .$$

Taking expectations we get:

$$E\left(\exp\left\{\sum_{i=1}^{n} \alpha_i \, t_i \, X_i\right\}\right) \le \sum_{i=1}^{n} \alpha_i \, m_{X_i}(t_i) .$$

Therefore, if for a fixed choice of $\alpha_i \in (0, 1)$, $i = 1, \cdots, n$, satisfying $\sum_{i=1}^{n} \alpha_i = 1$, we define $\mathcal{R}$ by

$$\mathcal{R} := \{\mathbf{u} = (u_1, \cdots, u_n) \in \mathbb{R}^n : u_i = \alpha_i \, t_i, \text{ with } |t_i| < M_i\}$$

then for all $\mathbf{u} \in \mathcal{R}$ we have

$$E(\exp(\mathbf{X}\mathbf{u}^T)) < \infty$$

and furthermore, by taking $M := \min\{\alpha_1 M_1, \cdots, \alpha_n M_n\}$, for all $\mathbf{t} \in \mathbb{R}^n$ with $\|\mathbf{t}\| < M$ we can guarantee that:

$$E(\exp(\mathbf{X}\mathbf{t}^T)) < \infty .$$

That is, the joint moment generating function of $\mathbf{X}$ exists. ∎

The previous theorem establishes that the joint moment generating function of the random variables $X_1, \cdots, X_n$, exists if and only if the marginal moment generating functions also exist; nevertheless, it does not say that the joint moment generating function can be found from the marginal distributions. That is possible if the random variables are independent, as stated in the following theorem:

Theorem 5.26 *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector. Suppose that for all $i = 1, \cdots, n$ there exists $M_i > 0$ such that:*

$$m_{X_i}(t) := E\big(\exp(tX_i)\big) < \infty \quad \text{if } |t| < M_i.$$

If the random variables $X_1, \cdots, X_n$ are independent, then $m_{\mathbf{X}}(\mathbf{t}) < \infty$ for all $\mathbf{t} = (t_1, \cdots, t_n)$ with $\|\mathbf{t}\| < M$, where $M := \min\{M_1, \cdots, M_n\}$. Moreover:

$$m_{\mathbf{X}}(\mathbf{t}) = \prod_{i=1}^{n} m_{X_i}(t_i) .$$

Proof: We have that:

$$\begin{aligned} m_{\mathbf{X}}(\mathbf{t}) &= E\big(\exp(\mathbf{X}\mathbf{t}^T)\big) \\ &= E\left\{\exp\left(\sum_{i=1}^{n} t_i X_i\right)\right\} \\ &= E\left\{\prod_{i=1}^{n} \exp(t_i X_i)\right\} \\ &= \prod_{i=1}^{n} E\big(\exp(t_i X_i)\big) \quad \text{(since, } X_i\text{'s are independent random variables)} \\ &= \prod_{i=1}^{n} m_{X_i}(t_i) . \end{aligned}$$

∎

Just like the one-dimensional case, the joint moment generating function allows us, when it exists, to find the *joint moment of the random vector* $\mathbf{X}$ *around the origin* in the following sense:

Definition 5.18 (Joint Moment) *Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector. The joint moment of order $k_1, \cdots, k_n$, with $k_j \in \mathbb{N}$, of $\mathbf{X}$ around the point $\mathbf{a} = (a_1, \cdots, a_n)$ is defined by*

$$\mu_{k_1 \cdots k_n} := E\left(\prod_{j=1}^{n} (X_j - a_j)^{k_j}\right)$$

provided that the expected value above does exist.

The joint moment of order $k_1, \cdots, k_n$ of $\mathbf{X}$ around the origin is written $\mu'_{k_1 \cdots k_n}$.

Note 5.19 *If X and Y are random variables whose expected values exist, then the joint moment of order $1, 1$ of the random vector $\mathbf{X} = (X, Y)$ around $\mathbf{a} = (EX, EY)$ is:*

$$\mu_{11} = E((X - EX)(Y - EY)) = Cov(X, Y) .$$

■ EXAMPLE 5.41

Let X_1 and X_2 be the random vectors with joint distribution given by:

$X_1 \backslash X_2$	-1	0	1
1	$\frac{1}{6}$	$\frac{2}{6}$	0
2	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

We have:

$$\begin{aligned} \mu'_{12} &= E(X_1 X_2^2) \\ &= \sum_{x_1} \sum_{x_2} x_1 \, x_2^2 \, P(X_1 = x_1, X_2 = x_2) \\ &= \frac{5}{6} . \end{aligned}$$

On the other hand:

$$\begin{aligned} m_X(\mathbf{t}) = \frac{1}{6} \exp(t_1 - t_2) + \frac{1}{6} \exp(2t_1 - t_2) + \frac{1}{3} \exp(t_1) \\ + \frac{1}{6} \exp(2t_1) + \frac{1}{6} \exp(2t_1 + t_2) . \end{aligned}$$

Furthermore, it can be easily verified that:

$$\left.\frac{\partial^3 m_{\mathbf{X}}(t_1,t_2)}{\partial t_1 \partial t_2 \partial t_2}\right|_{(t_1,t_2)=(0,0)}$$

$$= \left.\left[\frac{1}{6}\exp(t_1-t_2)+\frac{1}{3}\exp(2t_1-t_2)+\frac{1}{3}\exp(2t_1+t_2)\right]\right|_{(t_1,t_2)=(0,0)}$$

$$= \frac{5}{6} = \mu'_{12}. \qquad \blacktriangle$$

The property observed in the last example holds in more general situations, as the following theorem (given without proof) states.

Theorem 5.27 *Let* $\mathbf{X} = (X_1, \cdots, X_n)$ *be an n-dimensional random vector. Suppose that the joint moment generating function* $m_{\mathbf{X}}(\mathbf{t})$ *of* $\mathbf{X}$ *exists. Then the joint moments of all orders, around the origin, are finite and satisfy:*

$$\mu'_{k_1\cdots k_n} = \left.\frac{\partial^{k_1+\cdots+k_n}}{\partial t_1^{k_1}\cdots \partial t_n^{k_n}} m_{\mathbf{X}}(\mathbf{t})\right|_{(t_1,\cdots,t_n)=(0,\cdots,0)}.$$

We end this section by presenting the definition of the joint characteristic function of a random vector.

Definition 5.19 (Joint Characteristic Function) *Let* $\mathbf{X} = (X_1, \cdots, X_n)$ *be an n-dimensional random vector and* $\mathbf{t} = (t_1, \cdots, t_n) \in \mathbb{R}^n$. *The joint characteristic function of a random vector* $\mathbf{X}$*, notated as* $\varphi_{\mathbf{X}}(\mathbf{t})$*, is defined by*

$$\varphi_{\mathbf{X}}(\mathbf{t}) := E\left[\exp\left(i\mathbf{X}\mathbf{t}^T\right)\right] \quad ,$$

where $i = \sqrt{-1}$.

Just like the univariate case, the joint characteristic function of a random vector always exists. Another property carried over from the one-dimensional case is that the joint characteristic function of a random vector completely characterizes the distribution of the vector. That is two random vectors $\mathbf{X}$ and $\mathbf{Y}$ will have the same joint distribution function if and only if they have the same joint characteristic function. In addition, it can be proved that successive differentiation of the characteristic function followed by the evaluation at the origin of the derivatives thus obtained yield the presented below expression for the joint moments, around the origin,

$$\mu'_{k_1\cdots k_n} = \frac{1}{i^{k_1+\cdots+k_n}}\left[\left.\frac{\partial^{k_1+\cdots+k_n}}{\partial t_1^{k_1}\cdots \partial t_n^{k_n}} \varphi_{\mathbf{X}}(\mathbf{t})\right|_{(t_1,\cdots,t_n)=(0,\cdots,0)}\right]$$

whenever the moment is finite.

The proofs of these results given are beyond the scope of this text and the reader may refer to Hernandez (2003).

Suppose now that X and Y are independent random variables whose moment generating functions do exist. Let $Z := X + Y$. Then, we have that:

$$\begin{aligned} m_Z(t) &= E(\exp tZ) \\ &= E(\exp(tX + tY)) \\ &= E(\exp(tX)\exp(tY)) \\ &= E(\exp(tX))E(\exp(tY)) \\ &= m_X(t)m_Y(t). \end{aligned}$$

That is, the moment generating function of Z exists and it is equal to the product of the moment generating functions of X and Y.

In general we have that:

Note 5.20 *If $X_1, \cdots, X_n$ are n independent random variables whose moment generating functions do exist, then the moment generating function of the random variable $Z := X_1 + \cdots + X_n$ also exists and equals the product of the moment generating functions of $X_1, \cdots, X_n$. A similar result holds for the characteristic functions and for the probability generating functions.*

■ EXAMPLE 5.42

Let $X_1, X_2, \cdots, X_n$ be n independent and identically distributed random variables having an exponential distribution of parameter $\lambda > 0$. Then $Z := X_1 + \cdots + X_n$ has a gamma distribution of parameters n and λ. Indeed, the moment generating function of $Z := X_1 + \cdots + X_n$ is given by

$$\begin{aligned} m_Z(t) &= E(\exp tZ) \\ &= \prod_{k=1}^{n} E(\exp tX_k) \\ &= \prod_{k=1}^{n} \left(\frac{\lambda}{\lambda - t}\right) \quad \text{if } t < \lambda \\ &= \left(\frac{\lambda}{\lambda - t}\right)^n \quad \text{if } t < \lambda, \end{aligned}$$

which corresponds to the moment generating function of a gamma distributed random variable of parameters n and λ. ▲

■ EXAMPLE 5.43

Suppose you participate in a chess tournament in which you play n games. Since you are an average player, each game is equally likely to be a win, a loss or a tie. You collect 2 points for each win, 1 point for each tie and 0 points for each loss. The outcome of each game is independent of the outcome of every other game. Let X_i be the number of points you earn for game i and let Y equal the total number of points earned over the n games. Find the moment generating function $m_{X_i}(s)$ and $m_Y(s)$. Also, find $E(Y)$ and $Var(Y)$.

Solution: The probability distribution of X_i is:

X_i	$P_{X_i}(x)$
0	$\frac{1}{3}$
1	$\frac{1}{3}$
2	$\frac{1}{3}$

So the moment generating function of X_i is:

$$m_{X_i}(s) = \frac{1}{3}(1 + e^s + e^{2s}) .$$

Since it is identical for all i, we refer to it as $m_X(s)$. The mgf of Y is:

$$m_Y(s) = (M_X(s))^n = \frac{1}{3^n}(1 + e^s + e^{2s})^n .$$

Further:

$$E(Y) = n, \ Var(Y) = \frac{2}{3}n . \quad \blacktriangle$$

■ EXAMPLE 5.44

Let $X_1, X_2, \cdots, X_n$ be n independent and identically distributed random variables having a standard normal distribution and let:

$$Z := X_1^2 + \cdots + X_n^2 .$$

It is known that if a random variable has a standard normal distribution, then its square has a chi-squared distribution with one degree of freedom.

Therefore, the moment generating function of Z is given by

$$\begin{aligned} m_Z(t) &= \prod_{k=1}^{n} m_{X_k}(t) \\ &= \prod_{k=1}^{n} \left(\frac{1}{1-2t}\right)^{\frac{1}{2}} \quad \text{if } t < \frac{1}{2} \\ &= \left(\frac{1}{1-2t}\right)^{\frac{n}{2}} \quad \text{if } t < \frac{1}{2} \end{aligned}$$

that is, $Z \stackrel{d}{=} \mathcal{X}^2_{(n)}$. ▲

■ EXAMPLE 5.45

Let $X_1, X_2, \cdots, X_n$ be n independent random variables. Suppose that $X_k \stackrel{d}{=} \mathcal{N}(\mu_k, \sigma_k^2)$ for each $k = 1, \cdots, n$. Let $\alpha_1, \cdots, \alpha_n$ be n real constants. Then the moment generating function of the random variable $Z := \alpha_1 X_1 + \cdots + \alpha_n X_n$ is given by:

$$\begin{aligned} m_Z(t) &= E\left(\exp t\left(\sum_{k=1}^{n} \alpha_k X_k\right)\right) \\ &= E\left(\prod_{k=1}^{n} \exp(t\alpha_k X_k)\right) \\ &= \prod_{k=1}^{n} E(\exp(t\alpha_k X_k)) \\ &= \prod_{k=1}^{n} m_{X_k}(t\alpha_k) \\ &= \prod_{k=1}^{n} \exp\left(t\alpha_k\mu_k + \frac{(t\alpha_k)^2}{2}\sigma_k^2\right) \\ &= \exp\left(t\sum_{k=1}^{n}(\alpha_k\mu_k) + \frac{t^2}{2}\sum_{k=1}^{n}(\alpha_k\sigma_k)^2\right). \end{aligned}$$

Hence, $Z \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$, where $\mu := \sum_{k=1}^{n} (\alpha_k\mu_k)$ and $\sigma^2 := \sum_{k=1}^{n} (\alpha_k\sigma_k)^2$. ▲

■ **EXAMPLE 5.46**

Let X_1, $X_2, \cdots, X_n$ be n independent and identically distributed random variables having a normal distribution of parameters μ and σ^2. The random variable $\overline{X}$ defined by

$$\overline{X} := \frac{1}{n}\sum_{i=1}^{n} X_i$$

has a normal distribution with parameters μ and $\frac{\sigma^2}{n}$. Consequently:

$$\frac{\sum\limits_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \stackrel{d}{=} \mathcal{N}(0,1). \quad \blacktriangle$$

EXERCISES

5.1 Determine the constant h for which the following gives a joint probability mass function for (X, Y):

$$P(X = x, Y = y) = \begin{cases} (x+1)(y+1)h & \text{if } x = 0,1,2; y = 0,1,2,3 \\ 0 & \text{otherwise .} \end{cases}$$

Also evaluate the following probabilities:
a) $P(X \leq 1, Y \leq 1)$ b) $P(X + Y \leq 1)$ c) $P(XY > 0)$.

5.2 Suppose that in a central electrical system there are 15 devices of which 5 are new, 4 are between 1 and 2 years of use and 6 are in poor condition and should be replaced. Three devices are chosen randomly without replacement. If X represents the number of new devices and Y represents the number of devices with 1 to 2 years of use in the chosen devices, then find the joint probability distribution of X and Y. Also, find the marginal distributions of X and Y.

5.3 Prove that the function

$$F(x,y) = \begin{cases} 0 & x < 0 \text{ or } y < 0 \\ x + y & x + y < 1, x \geq 0, y \geq 0 \\ 1 & x + y \geq 1, x \geq 0, y \geq 0 \end{cases}$$

has the properties 1, 2, 3 and 4 of Theorem 5.3 but it is not a *cdf*.

5.4 Let (X, Y) a random vector with joint *pdf*

$$f(x,y) = \begin{cases} \frac{3}{2}(x^2 + y^2) & \text{if } 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise .} \end{cases}$$

Show that the *cdf* is:

$$F(x,y)=\begin{cases} 0 & x\le 0 \text{ or } y\le 0\\ \frac{1}{2}(x^3y+xy^3) & 0<x<1, 0<y<1\\ \frac{1}{2}(x^3+x) & 0<x<1, y\ge 1\\ \frac{1}{2}(y+y^3) & 0<y<1, x\ge 1\\ 1 & x\ge 1, y\ge 1\,. \end{cases}$$

5.5 Let X_1 and X_2 be two continuous random variables with joint probability density function

$$f_{(X_1,X_2)}(x_1,x_2)=\begin{cases} 4x_1x_2 & \text{if } 0<x_1<1, 0<x_2<1\\ 0 & \text{otherwise}\,. \end{cases}$$

Find the joint probability density function of $Y_1 = X_1^2$ and $Y_2 = X_1X_2$.

5.6 (**Buffon Needle Problem**) Parallel straight lines are drawn on the plane $\mathbb{R}^2$ at a distance d from each other. A needle of length L is dropped at random on the plane. What is the probability that the needle shall meet at least one of the lines?

5.7 Let A, B and C be independent random variables each uniformly distributed on $(0,1)$. What is the probability that $Ax^2 + Bx + C = 0$ has real roots?

5.8 Suppose that a two-dimensional random variable (X,Y) has joint probability density function

$$f_{X,Y}(x,y)=\begin{cases} \frac{e^{-(x+y)}x^8y^4}{8!4!} & \text{if } 0<x<\infty, 0<y<\infty\\ 0 & \text{otherwise}\,. \end{cases}$$

a) Find the *pdf* of $U = \frac{X}{Y}$.

b) Find $E(U)$.

5.9 Suppose that the joint distribution of the random variables X and Y is given by:

$X\backslash Y$	1	2	3	4
0	0.1	0	0	0
-1	0.1	0.1	0	0
-2	0.1	0.1	0.1	0
-3	0.1	0.1	0.1	0.1

Calculate:

a) $P(X \geq -2, Y \geq 2)$.

b) $P(X \geq -2 \text{ or } Y \geq 2)$.

c) The marginal distributions of X and Y.

d) The distribution of $Z := X + Y$.

5.10 Suppose that two cards are drawn at random from a deck of cards. Let X be the number of aces obtained and Y be the number of queens obtained. Find the joint probability mass function of (X, Y). Also find their joint distribution function.

5.11 Assume that X and Y are random variables with the following joint distribution:

$X \backslash Y$	0	1	2
1	0.2	α	β
2	γ	0.1	δ
3	η	κ	0.3

Find the values of $\alpha, \beta, \gamma, \delta, \eta$ and κ, so that the conditions below will hold:

$$P(X = 1) = 0.51,\ P(X = 2) = 0.22,\ P(Y = 0) = 0.33 \text{ and } P(Y = 2) = 0.62\,.$$

5.12 An urn contains 11 balls, 4 of them are red-colored, 5 are black-colored and 2 are blue-colored. Two balls are randomly extracted from the urn without replacement. Let X and Y be the random variables representing the number of red-colored and black-colored balls, respectively, in the sample. Find:

a) The joint distribution of X and Y.

b) $E(X)$ and $E(Y)$.

5.13 Solve the previous exercise under the assumption that the extraction takes place with replacement.

5.14 Let X, Y and Z be the random variables with joint distribution given by:

$\mathbf{x} = (x, y, z)$	$(1,2,3)$	$(2,1,3)$	$(3,2,1)$	$(1,3,2)$	$(3,1,2)$	$(2,3,1)$
$P(\mathbf{X} = \mathbf{x})$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Find:

a) $E(X+Y+Z)$.

b) $E(XYZ)$.

5.15 **(Multinomial Distribution)** Suppose that there are $k+1$ different results of a random experiment and that p_i is the probability of obtaining the ith result for $i = 1, \cdots, k+1$ (note that $\sum_{i=1}^{k+1} p_i = 1$). Let X_i be the number of times the ith result is obtained after n independent repetitions of the experiment. Verify that the joint density function of the random variables $X_1, \cdots, X_{k+1}$ is given by

$$f(x_1, \cdots, x_{k+1}) = \frac{n!}{\prod_{i=1}^{k+1} x_i!} \prod_{i=1}^{k+1} p_i^{x_i}$$

where $x_i = 0, \cdots, n$ and $\sum_{i=1}^{k+1} x_i = n$.

5.16 Surgeries scheduled in a hospital are classified into 4 categories according to their priority as follows: "urgent", "top priority", "low priority" and "on hold". The hospital board estimates that 10% of the surgeries belong to the first category, 50% to the second one, 30% to the third one and the remaining 10% to the fourth one. Suppose that 30 surgeries are programmed in a month. Find:

a) The probability that 5 of the surgeries are classified in the first category, 15 in the second, 7 in the third and 3 in the fourth.

b) The expected number of surgeries of the third category.

5.17 Let X and Y be independent random variables having the joint distribution given by the following table:

$X \backslash Y$	-1	0	1
0	$\frac{1}{6}$	d	$\frac{1}{6}$
1	a	e	k
2	b	f	h

If $P(X=1) = P(X=2) = \frac{1}{5}$ find the values missing from the table. Calculate $E(XY)$.

5.18 A bakery sells an average of 1.3 doughnuts per hour, 0.6 bagels per hour and 2.8 cupcakes per hour. Assume that the quantities sold of each product are independent and that they follow a Poisson distribution. Find:

a) The distribution of the total number of doughnuts, bagels and cupcakes sold in 2 hours.

b) The probability that at least two of the products are sold in a 15-minutes period.

5.19 Let X and Y be random variables with joint distribution given by:

$X\backslash Y$	-1	0	1
0	$\frac{1}{16}$	$\frac{1}{4}$	0
1	$\frac{3}{16}$	$\frac{1}{8}$	$\frac{3}{16}$
2	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

Find the joint distribution of $Z := X + Y$ and $W := X - Y$.

5.20 Let X and Y be discrete random variables having the following joint distribution:

$X\backslash Y$	-1	0	2
1	$\frac{1}{18}$	$\frac{3}{18}$	$\frac{1}{9}$
2	$\frac{1}{9}$	0	$\frac{1}{6}$
3	0	$\frac{2}{9}$	$\frac{1}{6}$

Compute $Cov(X, Y)$ and $\rho(X, Y)$.

5.21 Suppose X has a uniform distribution on the interval $(-\pi, \pi)$. Define $Y = \cos X$. Show that $Cov(X, Y) = 0$ though X, Y are dependent.

5.22 An urn contains 3 red-colored and 2 black-colored balls. A sample of size 2 is drawn without replacement. Let X and Y be the numbers of red-colored and black-colored balls, respectively, in the sample. Find $\rho(X, Y)$.

5.23 Let X and Y be independent random variables with $X \stackrel{d}{=} \mathcal{B}\left(3, \frac{1}{3}\right)$ and $Y \stackrel{d}{=} \mathcal{B}\left(2, \frac{1}{2}\right)$. Calculate $P(X = Y)$.

5.24 Let $X_1, X_2, \cdots, X_5$ be i.i.d. random variables with uniform distribution on the interval $(0, 1)$.

a) Find the probability that $\min(X_1, X_2, \cdots, X_5)$ lies between $(\frac{1}{4}, \frac{3}{4})$.

b) Find the probability that X_1 is the minimum and X_5 is the maximum among these random variables.

5.25 Let X and Y be independent random variables having Bernoulli distributions with parameter $\frac{1}{2}$. Are $Z := X + Y$ and $W := |X - Y|$ independent? Explain.

5.26 Consider an "experiment" of placing three balls into three cells. Describe the sample space of the experiment. Define the random variables

$$\begin{aligned} N &= \text{number of occupied cells} \\ X_i &= \text{number of balls in cell number } i \ (i = 1, 2, 3). \end{aligned}$$

a) Find the joint probability distribution of (N, X_1).

b) Find the joint probability distribution of (X_1, X_2).

c) What is $Cov(N, X_1)$?

d) What is $Cov(X_1, X_2)$?

5.27 Let X_1, X_2 and X_3 be independent random variables with finite positive variances σ_1^2, σ_2^2 and σ_3^2, respectively. Calculate the correlation coefficient between $X_1 - X_2$ and $X_2 + X_3$.

5.28 Let X and Y be two random variables such that $\rho(X, Y) = \frac{1}{2}$, $Var(X) = 1$ and $Var(Y) = 2$. Compute $Var(X - 2Y)$.

5.29 Given the uncorrelated random variable X_1, X_2, X_3 whose means are 2, 1 and 4 and whose variances are 9, 20 and 12:

a) Find the mean and the variance of $X_1 - 2X_2 + 5X_3$.

b) Find the covariance between $X_1 + 5X_2$ and $2X_2 - X_3 + 5$.

5.30 Let X, Y, Z be i.i.d. random variables each having uniform distribution in the interval $(1, 2)$. Find $Var\left(\frac{4X}{3Y} + \frac{3Y}{2Z}\right)$.

5.31 Let X and Y be independent random variables. Assume that both variables take the values 1 and -1 each with probability $\frac{1}{2}$. Let $Z := XY$. Are X, Y and Z pair-wise independent? Are X, Y and Z independent? Explain.

5.32 A certain lottery prints $n \geq 2$ lottery tickets m of which are sold. Suppose that the tickets are numbered from 1 to n and that they all have the same "chance" of being sold. Calculate the expected value and the variance of the random variable representing the sum of the numbers of the lottery tickets sold.

5.33 What is the expected number of days of the year for which exactly k of r people celebrate their birthday on that day? Suppose that each of the 365 days is just as likely to be the birthday of someone and ignore leap years.

5.34 Under the same assumptions given in problem 5.33, what is the expected number of days of the year for which there is more than one birthday? Verify with a calculator that this expected number is, for all $r \geq 29$, greater than 1.

5.35 Let X and Y be random variables with mean 0, variance 1 and correlation coefficient ρ. Prove that:

$$E\left(\max\left\{X^2, Y^2\right\}\right) \leq 1 + \sqrt{1-\rho^2}\ .$$

Hint: $\max\{u, v\} = \frac{1}{2}(u+v) + \frac{1}{2}|u-v|$. Use Cauchy-Schwarz inequality.

5.36 Let X and Y be random variables with mean 0, variance 1 and correlation coefficient ρ.

a) Show that the random variables $Z := X - \rho Y$ and Y are not correlated.

b) Compute $E(Z)$ and $Var(Z)$.

5.37 Let X, Y and Z be random variables with mean 0 and variance 1. Let $\rho_1 := \rho(X,Y)$, $\rho_2 := \rho(Y,Z)$ and $\rho_3 := \rho(X,Z)$. Prove that:

$$\rho_3 \geq \rho_1\rho_2 - \sqrt{1-\rho_1^2}\sqrt{1-\rho_2^2}\ .$$

Hint:

$$XZ = [\rho_1 Y + (X - \rho_1 Y)]\,[\rho_2 Y + (Z - \rho_2 Y)]\,.$$

5.38 Let $X \stackrel{d}{=} \mathcal{U}(0,1)$ and $Y = X^2$.

a) Find $\rho(X,Y)$.

b) Are X and Y independent? Explain.

5.39 Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} c & \text{if } 0 < x < 4,\ 0 < y \text{ and } (x-1) < y < (x+1) \\ 0 & \text{otherwise}\ . \end{cases}$$

a) Find the value of the constant c.

b) Compute $P\left(X < \frac{1}{2}, Y < \frac{1}{2}\right)$ and $P\left(X < \frac{1}{2}\right)$.

c) Calculate $E(X)$ and $E(Y)$.

5.40 Ten costumers, among them John and Amanda, arrive at a store between 8:00 AM and noon. Assuming that the clients arrive independently and that the arrival time of each of them is a random variable with uniform distribution on the interval $[0, 4]$. Find:

a) The probability that John arrives before 11:00 AM.

b) The probability that John and Amanda both arrive before 11:00 AM.

c) The expected number of clients arriving before 11:00 AM.

5.41 Let X and Y be random variables with joint cumulative distribution function given by:

$$F(x, y) = \begin{cases} (1 - \exp(-x)) \left(\frac{1}{2} + \frac{1}{\pi} \tan^{-1} y\right) & \text{if } x > 0 \\ 0 & \text{otherwise} . \end{cases}$$

Does a joint probability density function of X and Y exist? Explain.

5.42 Let X and Y be random variables having the following joint probability density function:

$$f(x, y) = \begin{cases} c \sin(x + y) & \text{if } 0 \leq x, y \leq \frac{\pi}{2} \\ 0 & \text{otherwise} . \end{cases}$$

Compute:

a) The value of the constant c.

b) The marginal density functions of X and Y.

5.43 Let X and Y be random variables with joint probability density function given by:

$$f(x, y) = \begin{cases} \frac{4x}{y^3} & \text{if } 0 < x < 1,\ 1 < y \\ 0 & \text{otherwise} . \end{cases}$$

Find:

a) $P\left(\frac{1}{2} < X \leq \frac{3}{4},\ 0 < Y \leq \frac{1}{3}\right)$.

b) $P(Y > 5)$.

5.44 Andrew and Sandra agreed to meet between 7:00 PM and 8:00 PM in a restaurant. Let X be the random variable representing the arrival time (in minutes) of Andrew and let Y be the random variable representing the arrival time (in minutes) of Sandra. Suppose that X and Y are independent and identically distributed with a uniform distribution over $[7, 8]$.

a) Find the joint probability density function of X and Y.

b) Calculate the probability that both Andrew and Sandra arrive at the restaurant between 7:15 PM and 8:15 PM.

c) If the first one waits only 10 minutes before leaving and eating elsewhere, what is the probability that both Sandra and Andrew eat at the restaurant initially chosen?

5.45 The two-dimensional random variable (X,Y) has the following joint probability density function:

$$f(t_1,t_2) = \frac{1}{16\pi} e^{-\frac{1}{2}\left(\frac{(t_1-3)^2}{4}+\frac{(t_2+2)^2}{16}\right)}, \quad -\infty < t_1, t_2 < \infty.$$

Find:

a) The marginal density functions f_X and f_Y.

b) $E(X)$ and $E(Y)$.

c) $Var(X)$ and $Var(Y)$.

d) $Cov(X,Y)$ and $\rho(X,Y)$.

5.46 Let (X,Y) be the coordinates of a point randomly chosen from the unit disk. That is, X and Y are random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} \frac{1}{\pi} & \text{if } 0 < x^2+y^2 \le 1 \\ 0 & \text{otherwise} . \end{cases}$$

Compute $P(X<Y)$.

5.47 A point Q with coordinates (X,Y) is randomly chosen from the square $[0,1]\times[0,1]$. Find the probability that $(0,0)$ is closer to Q than $\left(\frac{1}{2},\frac{1}{2}\right)$.

5.48 Let X and Y be independent random variables with $X \stackrel{d}{=} \mathcal{U}(0,1)$ and $Y \stackrel{d}{=} \mathcal{U}(0,2)$. Calculate:

a) $P([X+Y]>1)$.

b) $P\left(Y \le X^2+1\right)$.

5.49 Let X and Y be independent and identically distributed random variables having a uniform distribution over the interval $(0,1)$. Compute:

a) $P\left(|X-Y| \le \frac{1}{3}\right)$.

b) $P\left(\left|\frac{X}{Y}-1\right|\leq\frac{1}{3}\right)$.

c) $P\left(Y\leq X\mid X<\frac{1}{3}\right)$.

5.50 Let X and Y be random variables having the following joint probability density function:

$$f(x,y)=\begin{cases}\exp\left[-(x+y)\right] & \text{if } x>0,\ y>0\\ 0 & \text{otherwise}.\end{cases}$$

Calculate:

a) $P\left(\frac{1}{2}<X+Y<5\right)$.

b) $P\left(X<Y+3\mid X>2Y\right)$.

c) $P\left(Y>\frac{1}{2}\right)$.

5.51 A certain device has two components and if one of them fails the device will stop working. The joint *pdf* of the lifetime of the two components (measured in hours) is given by:

$$f(x,y)=\begin{cases}\frac{x+y}{2} & \text{if } 0<x<3, 0<y<3\\ 0 & \text{otherwise}.\end{cases}$$

Compute the probability that the device fails during the first operation hour.

5.52 Let X and Y be independent random variables. Assume that X has an exponential distribution with parameter λ and that Y has the density function given by $f(x)=2x\mathcal{X}_{(0,1)}(x)$.

a) Find a density function for $Z:=X+Y$.

b) Calculate $Cov(X,Z)$.

5.53 Prove or disprove: If X and Y are random variables such that the characteristic function of $X+Y$ equals the product of the characteristic functions of X and Y, then X and Y are independent. Justify your answer.

5.54 Let X_1 and X_2 be two identical random variables which are binomial distributed with parameters n and p. Find the $Cov(X_1-X_2,X_1+X_2)$.

5.55 Let X and Y be independent and identically distributed random variables having an exponential distribution with parameter λ. Find the density functions of the following random variables:

a) $Z:=|X-Y|$.

b) $W:=\min\left\{X,Y^3\right\}$.

5.56 Let X, Y and Z be independent and identically distributed random variables having a uniform distribution over the interval $[0, 1]$.

a) Find the joint density function of $W := XY$ and $V := Z^2$.

b) Calculate $P(V \leq W)$.

5.57 Let X and Y be independent and identically distributed random variables having a standard normal distribution. Let $Y_1 = X + Y$ and $Y_2 = X/Y$. Find the joint probability density function of the random variables Y_1 and Y_2. What kind of distribution does Y_2 have?

5.58 Let X and Y be random variables having the following joint probability density function:

$$f(x,y) = \begin{cases} \frac{3}{8}xy & \text{if } x \geq 0 \;\; y \geq 0 \text{ and } x + y \leq 2 \\ 0 & \text{otherwise .} \end{cases}$$

Find the joint probability density function of X^2 and Y^2.

5.59 Let X and Y be random variables having the joint probability density function given below:

$$f(x,y) = \begin{cases} cx^2y^2 & \text{if } 0 < x < 1 \text{ and } 0 < y < 1 \\ 0 & \text{otherwise .} \end{cases}$$

Calculate the constant c. Find the joint probability density function of X^3 and Y^3.

5.60 Let $X_1 < X_2 < X_3$ be ordered observations of a random sample of size 3 from a distribution with *pdf*

$$f(x) = \begin{cases} 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise .} \end{cases}$$

Show that $Y_1 = X_1/X_2, Y_2 = X_2/X_3$ and $Y_3 = X_3$ are independent.

5.61 Let X, Y and Z be random variables having the following joint probability density function:

$$f(x,y,z) = \begin{cases} e^{-(x+y+z)} & \text{if } x > 0, \; y > 0, \; z > 0 \\ 0 & \text{otherwise .} \end{cases}$$

Find f_U, where $U := \frac{X+Y+Z}{3}$.

5.62 Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} \frac{3x}{2} & \text{if } 0 < x < 1 \text{ and } -x < y < x \\ 0 & \text{otherwise .} \end{cases}$$

Find f_Z, where $Z = X - Y$.

5.63 Suppose that $X \stackrel{d}{=} F_n^m$. Find the value x for which:

a) $P(X \leq x) = 0.99$ with $m = 7, n = 3$.

b) $P(X \leq x) = 0.005$ with $m = 20$, $n = 30$.

c) $P(X \leq x) = 0.95$ with $m = 2$, $n = 9$.

5.64 If $X \stackrel{d}{=} t_n$, what kind of distribution does X^2 have?

5.65 Prove that, if $X \stackrel{d}{=} F_n^m$, then $E(X) = \frac{n}{n-2}$ for $n > 2$ and $Var(X) = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)}$.

Hint: Suppose that $X = \frac{U/m}{V/n}$, where U and V are independent, $U \stackrel{d}{=} \mathcal{X}_m^2$ and $V \stackrel{d}{=} \mathcal{X}_n^2$.

5.66 Let X be a random variable having a t-Student distribution with k degrees of freedom. Compute $E(X)$ and $Var(X)$.

5.67 Let X be a random variable having a standard normal distribution. Are X and $|X|$ independent? Are they not correlated? Explain.

5.68 Let $\mathbf{X}$ be an n-dimensional random vector with variance-covariance matrix $\sum$. Let A be a nonsingular square matrix of order n and let $\mathbf{Y} := \mathbf{X}A$.

a) Prove that $E(\mathbf{Y}) = E(\mathbf{X})A$.

b) Compute the variance-covariance matrix of $\mathbf{Y}$.

5.69 If the Y_i's, $i = 1, \cdots, n$, are independent and identically distributed continuous random variables with probability density $f(y_i)$, then prove that the joint density of the order statistics $Y_{(1)}, Y_{(2)}, \cdots, Y_{(n)}$ is:

$$f(y_1, y_2, \cdots, y_n) = n! \prod_{i=1}^{n} f(y_i), \quad y_1 < y_2 < \cdots < y_n.$$

Also, if the Y_i, $i = 1, \cdots, n$, are uniformly distributed over $(0, t)$, then prove that the joint density of the order statistics is:

$$f(y_1, y_2, \cdots, y_n) = \frac{n!}{t^n}, \quad y_l < y_2 < \cdots < y_n.$$

5.70 Let X and Y be independent random variables with moment generating functions given by:

$$m_X(t) = \exp\left(2e^t - 2\right) \quad \text{and} \quad m_Y(t) = \frac{1}{5}\left(1 + 2e^{-t} + 2e^t\right).$$

Find:

a) $P([X+Y]=2)$.

b) $E(XY)$.

5.71 Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with a $\mathcal{N}(\mu, \sigma^2)$ distribution. What kind of distribution do the random variables defined by $Z_i := \frac{X_i - \mu}{\sigma}$, for $i = 1, 2, \cdots, n$, have? What is the distribution of $Z_1^2 + \cdots + Z_n^2$?

5.72 Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with a $\mathcal{N}(\mu, \sigma^2)$ distribution. Let

$$S^2 := \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$$

where:

$$\overline{X} := \frac{1}{n} \sum_{i=1}^{n} X_i \ .$$

What kind of distribution does $\frac{(n-1)S^2}{\sigma^2}$ have? Explain.

5.73 The tensile strength for a certain kind of wire has a normal distribution with unknown mean μ and unknown variance σ^2. Six wire sections were randomly cut from an industrial roll. Consider the random variables $Y_i :=$"tensile strength of the ith segment" for $i = 1, 2, \cdots, 6$. The population mean μ and variance σ^2 can be estimated by $\overline{Y}$ and S^2, respectively. Find the probability that $\overline{Y}$ is at most at a $\frac{2S}{\sqrt{n}}$ distance from the real population mean μ.

5.74 Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with a $\mathcal{N}(\mu, \sigma^2)$ distribution. What kind of distribution does the following random variable have:

$$\frac{\sqrt{n(n-1)}\left(\overline{X} - \mu\right)}{\sqrt{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}} \ ?$$

Explain.

CHAPTER 6

CONDITIONAL EXPECTATION

One of the most important and useful concepts of probability theory is the conditional expected value. The reason for it is twofold: in the first place, in practice usually it is interesting to calculate probabilities and expected values when some partial information is already known. On the other hand, when one wants to find a probability or an expected value, many times it is convenient to condition first with respect to an appropriate random variable.

6.1 CONDITIONAL DISTRIBUTION

The relationship between two random variables can be seen by finding the conditional distribution of one of them given the value of the other. In Chapter 1, we defined the conditional probability of an event A given another event B as:

$$P(A \mid B) := \frac{P(A \cap B)}{P(B)}, \qquad P(B) > 0 \ .$$

It is natural, then, to have the following definition:

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

Definition 6.1 (Conditional Probability Mass Function) *Let X and Y be two discrete random variables. The conditional probability mass function of X given $Y = y$ is defined as*

$$p_{X|Y}(x \mid y) := P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

for all y for which $P(Y = y) > 0$.

Definition 6.2 (Conditional Distribution Function) *The conditional distribution function of X given $Y = y$ is defined as*

$$F_{X|Y}(x \mid y) := P(X \leq x \mid Y = y) = \sum_{k \leq x} p_{X|Y}(k \mid y)$$

for all y for which $P(Y = y) > 0$.

Definition 6.3 (Conditional Expectation) *The conditional expectation of X given $Y = y$ is defined as:*

$$E(X \mid Y = y) := \sum_x x p_{X|Y}(x \mid y) \ .$$

The quantity $E(X \mid Y = y)$ is called the regression of X on $Y = y$.

■ EXAMPLE 6.1

A box contains five red balls and three green ones. A random sample of size 2 (without replacement) is drawn from the box. Let:

$$X := \begin{cases} 1 & \text{if the first ball taken out is red} \\ 0 & \text{if the first ball taken out is green} \end{cases}$$

$$Y := \begin{cases} 1 & \text{if the second ball taken out is red} \\ 0 & \text{if the second ball taken out is green .} \end{cases}$$

The joint probability distribution of the random variables X and Y is given by:

$X\backslash Y$	0	1
0	$\frac{6}{56}$	$\frac{15}{56}$
1	$\frac{15}{56}$	$\frac{20}{56}$

Hence:

$$p_{X|Y}(x \mid 0) = \begin{cases} \frac{6}{21} & \text{if} \quad x = 0 \\ \frac{15}{21} & \text{if} \quad x = 1 \end{cases}$$

$$p_{X|Y}(x \mid 1) = \begin{cases} \frac{3}{7} & \text{if} \quad x = 0 \\ \frac{4}{7} & \text{if} \quad x = 1 \end{cases}$$

$$F_{X|Y}(x \mid 0) = \begin{cases} 0 & \text{if} \quad x < 0 \\ \frac{6}{21} & \text{if} \quad 0 \leq x < 1 \\ 1 & \text{if} \quad x \geq 1 \end{cases}$$

$$E(X \mid Y = y) = \begin{cases} \frac{15}{21} & \text{if} \quad y = 0 \\ \frac{4}{7} & \text{if} \quad y = 1\,. \end{cases}$$ ▲

■ **EXAMPLE 6.2**

Let X and Y be independent Poisson random variables with parameters λ_1 and λ_2, respectively. Calculate the expected value of X under the condition that $X + Y = n$, where n is a nonnegative fixed integer.
Solution: Let:

$$p_{X|X+Y}(x \mid n) = P(X = x \mid X + Y = n) = \frac{P(X = x, Y = n - x)}{P(X + Y = n)}$$

$$= \binom{n}{x} \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^x \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^{n-x}.$$

That is, X has, under the condition $X + Y = n$, a binomial distribution with parameters n and $p := \frac{\lambda_1}{\lambda_1+\lambda_2}$. Therefore:

$$E(X \mid X + Y = n) = n\left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right).$$ ▲

Definition 6.4 (Conditional Probability Density Function) *Let X and Y be continuous random variables with joint probability density function f. The conditional probability density function of X given $Y = y$ is defined as*

$$f_{X|Y}(x \mid y) := \frac{f(x, y)}{f_Y(y)}$$

for all y with $f_Y(y) > 0$.

Definition 6.5 (Conditional Distribution Function) *The conditional distribution function of X given $Y = y$ is defined as*

$$F_{X|Y}(x \mid y) := P(X \leq x \mid Y = y) = \int_{-\infty}^{x} f_{X|Y}(t \mid y)\, dt$$

for all y with $f_Y(y) > 0$.

Definition 6.6 (Conditional Expectation) *The conditional expectation of X given $Y = y$ is defined as*

$$E(X \mid Y = y) := \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx$$

for all y with $f_Y(y) > 0$.

■ EXAMPLE 6.3

Let X and Y be random variables with joint probability density function given by:

$$f(x, y) = \begin{cases} x + y & \text{if } 0 < x < 1,\ 0 < y < 1 \\ 0 & \text{other cases.} \end{cases}$$

For $0 < y < 1$ we can obtain that:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{x+y}{0.5+y} & \text{if } 0 < x < 1 \\ 0 & \text{other cases} \end{cases}$$

$$F_{X|Y}(x \mid y) = \begin{cases} 0 & \text{if } x \leq 0 \\ \frac{1}{0.5+y}\left(\frac{x^2}{2} + xy\right) & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

$$E(X \mid Y = y) = \frac{1}{0.5 + y}\left(\frac{1}{3} + \frac{y}{2}\right). \quad \blacktriangle$$

■ EXAMPLE 6.4

Let X and Y be random variables with joint probability density function given by

$$f(x, y) = \begin{cases} \lambda^2 e^{-\lambda y} & \text{if } 0 < x < y \\ 0 & \text{other cases} \end{cases}$$

where $\lambda > 0$. For $y > 0$ we obtain:

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{1}{y} & \text{if } 0 < x < y \\ 0 & \text{other cases} \end{cases}$$

$$F_{X|Y}(x \mid y) = \begin{cases} \frac{x}{y} & \text{if } 0 < x < y \\ 1 & \text{if } x > 0,\ x > y \\ 0 & \text{other cases} \end{cases}$$

$$E(X \mid Y = y) = \frac{y}{2}\,. \qquad \blacktriangle$$

■ EXAMPLE 6.5

Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} 6 - x - y & \text{if } \ 0 < x < 2 \text{ and } 2 < y < 4 \\ 0 & \text{other cases.} \end{cases}$$

Calculate $f_{X|Y}(x \mid y)$ and $E(X \mid Y = y)$.
Solution: The marginal density function of Y is equal to:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x,y)dx \\ &= \int_0^2 (6 - x - y)\mathcal{X}_{(2,4)}(y)dx \\ &= (10 - 2y)\,\mathcal{X}_{(2,4)}(y)\,. \end{aligned}$$

Then, for $2 < y < 4$, we obtain that:

$$\begin{aligned} f_{X|Y}(x \mid y) &= \left(\frac{6 - x - y}{10 - 2y}\right)\mathcal{X}_{(0,2)}(x) \\ E(X \mid Y = y) &= \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)dx \\ &= \int_0^2 x\left(\frac{6 - x - y}{10 - 2y}\right)dx \\ &= \frac{28 - 6y}{3\,(10 - 2y)}\,. \qquad \blacktriangle \end{aligned}$$

Note 6.1 *For all y, with $f_Y(y) > 0$ and all Borel set A in $\mathbb{R}$, it can be said that:*

$$P(X \in A \mid Y = y) = \int_A f_{X|Y}(x \mid y)dx \; .$$

■ EXAMPLE 6.6

If X and Y are random variables with joint probability density function given by

$$f(x, y) = \frac{e^{-\frac{x}{y}} e^{-y}}{y} \mathcal{X}_{(0,\infty)}(x) \mathcal{X}_{(0,\infty)}(y),$$

then $P(X > 1 \mid Y = \frac{1}{2}) = e^{-2}$. ▲

Note 6.2 *If X and Y are independent random variables, then the conditional density of X given $Y = y$ is equal to the density of X.*

Note 6.3 (Bayes' Rule)

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)} \quad \textit{if } f_Y(y) > 0 \; .$$

As $f(x,y) = f_X(x) f_{Y|X}(y \mid x)$ and $f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$, then

$$f_{X|Y}(x \mid y) = \frac{f_X(x) f_{Y|X}(y \mid x)}{\int_{-\infty}^{\infty} f_X(x) f_{Y|X}(y \mid x)dx} \; . \quad ▲$$

So far we have defined the conditional distributions when both the random variables under consideration are either discrete or continuous. Suppose now that X is an absolutely continuous random variable and that N is a discrete random variable. In this case:

$$\begin{aligned} f_{X|N}(x \mid n) &= \lim_{\Delta x \to 0} \frac{P(x < X \le x + \Delta x \mid N = n)}{\Delta x} \\ &= \lim_{\Delta x \to 0} \frac{P(N = n \mid x < X \le x + \Delta x)}{P(N = n)} \frac{P(x < X \le x + \Delta x)}{\Delta x} \\ &= \frac{P(N = n \mid X = x)}{P(N = n)} f_X(x) \; . \end{aligned}$$

■ EXAMPLE 6.7

Let X be a random variable with uniform distribution over the interval $(0,1)$ and N, a binomial random variable with parameters $n + m$ and

X. Then, for $0 < x < 1$, we have that

$$
\begin{aligned}
f_{X|N}(x \mid n) &= \frac{P(N = n \mid X = x)}{P(N = n)} f_X(x) \\
&= \frac{\binom{n+m}{n}}{P(N = n)} x^n (1 - x)^m \\
&= C x^n (1 - x)^m
\end{aligned}
$$

where $C = \frac{1}{P(N=n)} \binom{n+m}{n}$. That is, under the condition $N = n$, X has a beta distribution with parameters $n + 1$ and $m + 1$. ▲

■ EXAMPLE 6.8

Let Y be a Poisson random variable with parameter Λ, where the parameter Λ itself is distributed as $\Gamma(\alpha, \beta)$. Calculate $f_{\Lambda|Y}(\lambda \mid y)$.
Solution: It is known that:

$$
p_{Y|\Lambda}(y \mid \lambda) = \begin{cases} \frac{\lambda^y e^{-\lambda}}{y!} & \text{if } y = 0, 1, \cdots \\ 0 & \text{other cases .} \end{cases}
$$

Then:

$$
\begin{aligned}
f(\lambda, y) &= p_{Y|\Lambda}(y \mid \lambda) f_\Lambda(\lambda) \\
&= \begin{cases} f_\Lambda(\lambda) \frac{\lambda^y e^{-\lambda}}{y!} & \text{if } y = 0, 1, \cdots \\ 0 & \text{other cases .} \end{cases}
\end{aligned}
$$

Given that:

$$
f_\Lambda(\lambda) = \frac{\beta^\alpha \lambda^{\alpha-1} \exp(-\lambda\beta)}{\Gamma(\alpha)} \mathcal{X}_{(0,\infty)}(\lambda)
$$

it is obtained that:

$$
\begin{aligned}
p_Y(y) &= \int_0^\infty \frac{\beta^\alpha \lambda^{\alpha-1} \exp(-\lambda\beta)}{\Gamma(\alpha)} \frac{\lambda^y e^{-\lambda}}{y!} d\lambda \\
&= \frac{\beta^\alpha}{y! \Gamma(\alpha)} \int_0^\infty \lambda^{\alpha+y-1} \exp(-\lambda(\beta + 1)) d\lambda \\
&= \frac{\beta^\alpha \Gamma(\alpha + y)}{y! \, (\beta + 1)^{\alpha+y} \, \Gamma(\alpha)} \, .
\end{aligned}
$$

Consequently, for $\lambda > 0$ and y a nonnegative integer:

$$\begin{aligned} f_{\Lambda|Y}(\lambda \mid y) &= \frac{f(\lambda, y)}{p_Y(y)} \\ &= \frac{\beta^\alpha \lambda^{\alpha+y-1} y!(\beta+1)^{\alpha+y}\Gamma(\alpha)\exp(-\lambda(\beta+1))}{y!\beta^\alpha\Gamma(\alpha)\Gamma(\alpha+y)} \\ &= \frac{(\beta+1)^{\alpha+y}\,\lambda^{\alpha+y-1}\exp(-\lambda(\beta+1))}{\Gamma(\alpha+y)}\,. \end{aligned}$$

That is, under the condition $Y = y$, Λ has a gamma distribution with parameters $\alpha + y$ and $\beta + 1$. ▲

Definition 6.7 *Let X and Y be real random variables and h a real function such that $h(X)$ is a random variable. Define*

$$E(h(X) \mid Y = y) := \begin{cases} \sum\limits_x h(x)P(X = x \mid Y = y) & \text{if } X \text{ and } Y \text{ are discrete} \\ & \quad\text{random variables} \\ \int_{-\infty}^{\infty} h(x) f_{X|Y}(x \mid y)dx & \text{if } X \text{ and } Y \text{ are continuous} \\ & \quad\text{random variables} \end{cases}$$

for all values y of Y for which $P(Y = y) > 0$ in the discrete case and $f_Y(y) > 0$ in the continuous case.

■ EXAMPLE 6.9

Let X and Y be random variables with joint probability density function given by:

$$f(x, y) = \begin{cases} \exp(-y) & \text{if } x > 0;\ y > x \\ 0 & \text{other cases}\,. \end{cases}$$

We have that:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)dx \\ &= \int_0^{\infty} \exp(-y)\mathcal{X}_{(0,\infty)}(y)dx \\ &= ye^{-y}\mathcal{X}_{(0,\infty)}(y)\,. \end{aligned}$$

Therefore, for $y > 0$, we have:

$$f_{X|Y}(x \mid y) = \frac{1}{y}\mathcal{X}_{(0,y)}(x)\,.$$

Now, it can be deduced that:

$$\begin{aligned}E\left(\exp\left(\frac{X}{2}\right) \mid Y=1\right) &= \int_{-\infty}^{\infty} \exp\left(\frac{x}{2}\right) f_{X|Y}(x \mid 1)dx \\ &= \int_0^1 \exp\left(\frac{x}{2}\right) dx \\ &= 2\ . \quad \blacktriangle\end{aligned}$$

From the previous definition, a new random variable can be defined as follows:

Definition 6.8 (Conditional Expectation) *Let X and Y be real random variables defined over $(\Omega, \Im, P)$ and h a real-valued function such that $h(X)$ is a random variable. The random variable $E(h(X) \mid Y)$ defined by*

$$\begin{aligned} E(h(X) \mid Y):\ \Omega &\longrightarrow \mathbb{R} \\ \omega &\longmapsto E(h(X) \mid Y = Y(\omega)) \end{aligned}$$

is called the conditional expected value of $h(X)$ given Y.

■ EXAMPLE 6.10

Let $\Omega = \{a, b, c\}$, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{3}$ for all $\omega \in \Omega$.
Consider the random variables X and Y defined as follows:

$$X(\omega) = \begin{cases} 1 & \text{if} \quad \omega = a, b \\ 0 & \text{if} \quad \omega = c \end{cases}$$

$$Y(\omega) = \begin{cases} \pi & \text{if} \quad \omega = a \\ \frac{1}{2} & \text{if} \quad \omega = b \\ -1 & \text{if} \quad \omega = c\ . \end{cases}$$

Then:

$$E(X \mid Y) = \begin{cases} E(X \mid Y = \pi) & \text{if} \quad \omega = a \\ E(X \mid Y = \frac{1}{2}) & \text{if} \quad \omega = b \\ E(X \mid Y = -1) & \text{if} \quad \omega = c\ . \end{cases}$$

It is obtained that:

$$\begin{aligned} E(X \mid Y = \pi) &= \sum_x x P(X = x \mid Y = \pi) \\ &= P(X = 1 \mid Y = \pi) \\ &= \frac{P(X = 1, Y = \pi)}{P(Y = \pi)} \\ &= \frac{P(a)}{P(a)} = 1\ . \end{aligned}$$

In the same way, it can be verified that:

$$E\left(X \mid Y = \frac{1}{2}\right) = \frac{P(b)}{P(b)} = 1$$

$$E(X \mid Y = -1) = \frac{P(\emptyset)}{P(c)} = 0 .$$

Then:

$$E(X \mid Y) = \begin{cases} 1 & \text{if} \quad \omega = a, b \\ 0 & \text{if} \quad \omega = c . \end{cases}$$

It may be observed, additionally, that:

$$\begin{aligned} E\left(E(X \mid Y)\right) &= \sum_y E(X \mid Y = y) P(Y = y) \\ &= \frac{2}{3} = E(X) . \quad \blacktriangle \end{aligned}$$

The above result of the example is proved in the following theorem in a general setup.

Theorem 6.1 *Let X, Y be real random variables defined over $(\Omega, \Im, P)$ and h a real-valued function such that $h(X)$ is a random variable. If $E(h(X))$ exists, then:*

$$E(h(X)) = E(E(h(X)) \mid Y) .$$

Proof: Suppose that X and Y are discrete random variables. Then:

$$\begin{aligned} E(E(h(X)) \mid Y) &= \sum_y E(h(X) \mid Y = y) P(Y = y) \\ &= \sum_y \sum_x h(x) P(X = x \mid Y = y) P(Y = y) \\ &= \sum_y \sum_x h(x) P(X = x, Y = y) \\ &= \sum_x h(x) \left(\sum_y P(X = x, Y = y) \right) \\ &= \sum_x h(x) P(X = x) \\ &= E(h(X)) . \end{aligned}$$

If X and Y are random variables with joint probability density function f, then:

$$\begin{aligned} E(E(h(X) \mid Y)) &= \int_{-\infty}^{\infty} E(h(X) \mid Y = y) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \left(\int_{-\infty}^{\infty} h(x) f_{X|Y}(x \mid y) dx \right) f_Y(y) dy \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x) f(x, y) dx dy \\ &= \int_{-\infty}^{\infty} h(x) \left(\int_{-\infty}^{\infty} f(x, y) dy \right) dx \\ &= \int_{-\infty}^{\infty} h(x) f_X(x) dx \\ &= E(h(X)) . \end{aligned}$$

■

■ EXAMPLE 6.11

The number of clients who arrive at a store in a day is a Poisson random variable with mean $\lambda = 10$. The amount of money (in thousands of pesos) spent by each client is a random variable with uniform distribution over the interval $(0, 100]$. Determine the amount of money that the store is expecting to collect in a day.

Solution: Let X and M be random variables defined by:

$X :=$ "Number of clients who arrive at the store in a day".

$M :=$ "Amount of money that the store collects in a day".

It is clear that

$$M = \sum_{i=1}^{X} M_i ,$$

where:

$M_i :=$ "Amount of money spent by the ith client".

According to the previous theorem, it can be obtained that:

$$E(M) = E(E(M \mid X)) .$$

Given:

$$\begin{aligned} E\left(M \mid X=k\right) &= E\left(\sum_{i=1}^{k} M_i\right) \\ &= \sum_{i=1}^{k} E(M_i) \\ &= 50k. \end{aligned}$$

Then:

$$E(M) = E(50X) = 50E(X) = 500{,}000 \text{ pesos.} \qquad \blacktriangle$$

Note 6.4 *In particular it is said that, if $(\Omega, \Im, P)$ is an arbitrary probability space and if $A \in \Im$ is fixed, then:*

$$\begin{aligned} P(A) &= E\left(\mathcal{X}_A\right) \\ &= E\left[E\left(\mathcal{X}_A \mid Y\right)\right] \\ &= \begin{cases} \sum_y P\left(A \mid Y=y\right) P\left(Y=y\right) & \text{if } Y \text{ is a discrete random variable} \\ \int_{-\infty}^{\infty} P\left(A \mid Y=y\right) f_Y(y)dy & \text{if } Y \text{ is a random variable with pdf } f_Y. \end{cases} \end{aligned}$$

■ EXAMPLE 6.12

Let X and Y be independent random variables with densities f_X and f_Y, respectively. Calculate $P(X < Y)$.
Solution: Let $A := \{X < Y\}$. Then

$$\begin{aligned} P(A) &= E\left(\mathcal{X}_A\right) \\ &= E(E(\mathcal{X}_A \mid Y)) \\ &= \int_{-\infty}^{\infty} E(\mathcal{X}_A \mid Y=y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} P(X<Y \mid Y=y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} P(X<y) f_Y(y)dy \\ &= \int_{-\infty}^{\infty} F_X(y) f_Y(y)dy \end{aligned}$$

where $F_X(.)$ is the distribution function of X. ▲

Theorem 6.2 *If X, Y and Z are real random variables defined over $(\Omega, \Im, P)$ and if h is a real function such that $h(Y)$ is a random variable, then the conditional expected value satisfies the following conditions:*

1. *$E(X \mid Y) \geq 0$ if $X \geq 0$ a.s.*

2. *$E(1 \mid Y) = 1$.*

3. *If X and Y are independent, then $E(X \mid Y) = E(X)$.*

4. *$E(Xh(Y) \mid Y) = h(Y)E(X \mid Y)$.*

5. *$E(\alpha X + \beta Y \mid Z) = \alpha E(X \mid Z) + \beta E(Y \mid Z)$ for $\alpha, \beta \in \mathbb{R}$.*

Proof: We present the proof for the discrete case. Proof for the continuous case can be obtained in a similar way.

1. Suppose that X takes the values $x_1, x_2, \cdots$. Given that $P(X < 0) = 0$, then $P(X = x_j) = 0$ for $x_j < 0$. Therefore,

$$\begin{aligned} E(X \mid Y = y) &= \sum_j x_j P(X = x_j \mid Y = y) \\ &= \sum_{j:x_j \geq 0} x_j P(X = x_j \mid Y = y) \geq 0 \end{aligned}$$

and in consequence $E(X \mid Y) \geq 0$.

2. Let $X := 1$. Then:

$$E(X \mid Y = y) = 1P(X = 1 \mid Y = y) = 1.$$

3. As X and Y are independent, it is obtained that

$$P(X = x \mid Y = y) = P(X = x)$$

for all y with $P(Y = y) > 0$. Therefore:

$$\begin{aligned} E\left(X \mid Y = y\right) &= \sum_x x P(X = x \mid Y = y) \\ &= \sum_x x P(X = x) \\ &= E(X). \end{aligned}$$

4.

$$\begin{aligned} E\left(Xh(Y) \mid Y = y\right) &= \sum_x x h(y) P(X = x \mid Y = y) \\ &= h(y) E(X \mid Y = y) \end{aligned}$$

and we have:

$$E(Xh(Y) \mid Y) = h(Y)E(X \mid Y).$$

5.

$$\begin{aligned}
E(\alpha X+\beta Y \mid Z=z) &= \sum_{x,y}(\alpha x+\beta y)\,P(X=x,Y=y \mid Z=z) \\
&= \alpha\sum_{x,y} xP(X=x,Y=y \mid Z=z) \\
&\qquad + \beta\sum_{x,y} yP(X=x,Y=y \mid Z=z) \\
&= \alpha\sum_{x} x\sum_{y} P(X=x,Y=y \mid Z=z) \\
&\qquad + \beta\sum_{y} y\sum_{x} P(X=x,Y=y \mid Z=z) \\
&= \alpha\sum_{x} xP(X=x \mid Z=z) \\
&\qquad + \beta\sum_{y} yP(Y=y \mid Z=z) \\
&= \alpha E(X \mid Z=z)+\beta E(Y \mid Z=z).
\end{aligned}$$

Therefore:

$$E(\alpha X+\beta Y \mid Z)=\alpha E(X \mid Z)+\beta E(Y \mid Z).$$

■

■ EXAMPLE 6.13

Consider the $n+m$ Bernoulli trials, each trial with success probability p. Calculate the expected number of success in the first n attempts.
Solution: Let Y :=“total number of successes” and, for each $i=1,\cdots,n$, let:

$$X_i := \begin{cases} 1 & \text{if} \quad \text{success is obtained in the } i\text{th attempt} \\ 0 & \qquad \text{otherwise .} \end{cases}$$

It is clear that:

$$X := \sum_{i=1}^{n} X_i = \text{“number of successes in the first } n \text{ attempts”}.$$

As

$$E(X)=E(E(X \mid Y))$$

and

$$\begin{aligned} E(X \mid Y = k) &= E\left(\sum_{i=1}^{n} X_i \mid Y = k\right) \\ &= \sum_{i=1}^{n} E(X_i \mid Y = k) \\ &= \sum_{i=1}^{n} P(X_i = 1 \mid Y = k) \\ &= \frac{kn}{n+m}, \end{aligned}$$

then:

$$E(X) = E\left(\frac{nY}{n+m}\right) = \frac{n}{n+m}E(Y) = \frac{n}{n+m}p(n+m) = np. \qquad \blacktriangle$$

EXAMPLE 6.14

The number of customers entering a supermarket in a given hour is a random variable with mean 100 and standard deviation 20. Each customer, independently of the others, spends a random amount of money with mean \$100 and standard deviation \$50. Find the mean and standard deviation of the amount of money spent during the hour.

Solution: Let N be the number of customers entering the supermarket. Let X_i be the amount spent by the ith customer. Then the total amount of money spent is $Z = \sum_{i=1}^{N} X_i$. The mean is:

$$\begin{aligned} E\left(\sum_{i=1}^{N} X_i\right) &= E\left(E\left(\sum_{i=1}^{N} X_i | N = n\right)\right) \\ &= E(100N) = 10,000. \end{aligned}$$

Using

$$Var(X) = E(Var(X|Y)) + Var(E(X|Y))$$

we get:

$$\begin{aligned} Var\left(\sum_{i=1}^{N} X_i\right) &= E\left(Var\left(\sum_{i=1}^{N} X_i | N = n\right)\right) + Var\left(E\left(\sum_{i=1}^{N} X_i | N = n\right)\right) \\ &= E(50N) + Var(100N) \\ &= 5000 + 10,400 \\ &= 15,400. \end{aligned}$$

Hence the standard deviation is 124.0967. $\blacktriangle$

■ **EXAMPLE 6.15**

A hen lays N eggs, where N has a Poisson distribution with mean λ. The weight of the nth egg is W_n, where $W_1, W_2, \cdots$ are independent and identically distributed random variables with common probability generating function G. Prove that the probability generating function of the total weight $W = \sum_{i=1}^{N} W_i$ is $\exp(-\lambda(1 - G(s)))$.

Solution: The pgf of the total weight is:

$$\begin{aligned}
E(s^W) &= E\left(s^{\sum_{i=1}^{N} W_i} \right) \\
&= E\left(E\left(s^{\sum_{i=1}^{n} W_i} \mid N = n \right) \right) \\
&= E\left[\Pi_{i=1}^{n} E\left(s^{W_i} \mid N = n \right) \right] \\
&= E\left[\Pi_{i=1}^{n} \sum_{w_i} s^{w_i} f_{W_i}(w_i) \right] \\
&= E(G^n(s)) = \sum_{n=0}^{\infty} G^n(s) f_N(n) \\
&= \sum_{n=0}^{\infty} e^{-\lambda} \frac{(\lambda G(s))^n}{n!} \\
&= \exp(-\lambda(1 - G(s))) \ . \quad \blacktriangle
\end{aligned}$$

6.2 CONDITIONAL EXPECTATION GIVEN A σ-ALGEBRA

In this section the concept of conditional expected value of a random variable with respect to a σ-algebra will be worked which generalizes the concept of conditional expected value developed in the previous section.

Definition 6.9 (Conditional Expectation of X Given B) *Let X be a real random variable defined over $(\Omega, \Im, P)$ and let $B \in \Im$ with $P(B) > 0$. The conditional expected value of X given B is defined as*

$$E(X \mid B) := \frac{E(X\mathcal{X}_B)}{P(B)}$$

if the expected value of $Y := X\mathcal{X}_B$ exists.

■ EXAMPLE 6.16

A fair die is thrown twice consecutively. Let X be a random variable that denotes the sum of the results obtained and B be the event that indicates that the first throw is 5. Calculate $E(X \mid B)$.

Solution: The sample space of the experiment is given by:

$$\Omega = \{(a,b) : a, b \in \{1, \cdots, 6\}\}.$$

It is clear that:

$$(X\mathcal{X}_B)(a,b) = \begin{cases} 5+b & \text{if } a = 5,\ b \in \{1, \cdots, 6\} \\ 0 & \text{other cases .} \end{cases}$$

Then

$$\begin{aligned} E(X\mathcal{X}_B) &= \sum_{b=1}^{6} (5+b)\, P(X\mathcal{X}_B = 5+b) \\ &= \frac{1}{36}\sum_{b=1}^{6} (5+b) \\ &= \frac{51}{36} \end{aligned}$$

and

$$P(B) = \frac{1}{6}\,.$$

Therefore:

$$E(X \mid B) = \frac{\frac{51}{36}}{\frac{1}{6}} = \frac{51}{6} = 8.5. \qquad \blacktriangle$$

■ EXAMPLE 6.17

Let X be a random variable with exponential distribution with parameter λ. Calculate $E(X \mid \{X \geq t\})$.

Solution: Given that $X \stackrel{d}{=} Exp(\lambda)$ we have that the density function is given by:

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x) & \text{if } x \geq 0 \\ 0 & \text{other cases .} \end{cases}$$

Therefore:

$$\begin{aligned} P(X \geq t) &= 1 - P(X < t) \\ &= \begin{cases} 1 - \int_0^t \lambda \exp(-\lambda x)dx & \text{if } t \geq 0 \\ 1 & \text{if } t < 0 \end{cases} \\ &= \begin{cases} \exp(-\lambda t) & \text{if } t \geq 0 \\ 1 & \text{if } t < 0\,. \end{cases} \end{aligned}$$

On the other hand,

$$E(X\mathcal{X}_{\{X\geq t\}}) = \begin{cases} \int_t^{\infty} x\lambda \exp(-\lambda x)dx & \text{if} \quad t \geq 0 \\ \int_0^{\infty} x\lambda \exp(-\lambda x)dx & \text{if} \quad t < 0 \end{cases}$$

$$= \begin{cases} t\exp(-\lambda t) + \frac{1}{\lambda}\exp(-\lambda t) & \text{if} \quad t \geq 0 \\ \frac{1}{\lambda} & \text{if} \quad t < 0 \end{cases}$$

and we obtain:

$$E(X \mid \{X \geq t\}) = \begin{cases} t + \frac{1}{\lambda} & \text{if} \quad t \geq 0 \\ \frac{1}{\lambda} & \text{if} \quad t < 0\,. \end{cases} \quad \blacktriangle$$

■ EXAMPLE 6.18

Let X, Y and Z be random variables with joint distribution given by:

$\mathbf{x} := (x, y, z)$	$(0,0,0)$	$(0,0,1)$	$(0,1,1)$	$(1,1,1)$	$(1,1,0)$	$(1,0,1)$
$P(\mathbf{X} = \mathbf{x})$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{2}{9}$

Calculate $E(X \mid Y = 0, Z = 1)$.
Solution:

$$\begin{aligned} E(X \mid Y = 0, Z = 1) &= \sum_x xP(X = x \mid Y = 0, Z = 1) \\ &= P(X = 1 \mid Y = 0, Z = 1) \\ &= \frac{P(X = 1, Y = 0, Z = 1)}{P(Y = 0, Z = 1)} \\ &= \frac{\frac{2}{9}}{\sum_x P(X = x, Y = 0, Z = 1)} \\ &= \frac{1}{2}\,. \quad \blacktriangle \end{aligned}$$

Definition 6.10 (Conditional Expectation of X Given $\mathcal{G}$) *Let X be a real random variable defined over $(\Omega, \Im, P)$ for which $E(X)$ exists. Let $\mathcal{G}$ be a sub-σ-algebra of $\Im$. The conditional expected value of X given $\mathcal{G}$, denoted by $E(X \mid \mathcal{G})$, is a random variable $\mathcal{G}$-measurable so that:*

$$E\left([X - E(X \mid \mathcal{G})]\,\mathcal{X}_G\right) = 0 \text{ for all } G \in \mathcal{G}. \tag{6.1}$$

■ EXAMPLE 6.19

Let $\Omega = \{a, b, c\}$, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{3}$ for all $\omega \in \Omega$. Suppose that X is a real random variable given by

$$X(\omega) = \begin{cases} 0 & \text{if} \quad \omega = a, b \\ 2 & \text{if} \quad \omega = c \end{cases}$$

and let $\mathcal{G} := \{\emptyset, \{a\}, \{b, c\}, \Omega\}$.

We have Y given by

$$Y(\omega) = \begin{cases} 0 & \text{if} \quad \omega = a \\ 1 & \text{if} \quad \omega = b, c \end{cases}$$

which is equal to $E(X \mid \mathcal{G})$. Indeed:

1. Y is $\mathcal{G}$-measurable, due to the fact that:

$$Y^{-1}((-\infty, x]) = \begin{cases} \emptyset & \text{if} \quad x < 0 \\ \{a\} & \text{if} \quad 0 \leq x < 1 \\ \Omega & \text{if} \quad x \geq 1 \,. \end{cases}$$

2. Y satisfies condition (6.1) because:

$$(X - Y)(\omega) = \begin{cases} 0 & \text{if} \quad \omega = a \\ -1 & \text{if} \quad \omega = b \\ 1 & \text{if} \quad \omega = c \,. \end{cases}$$

Therefore,

$$\begin{aligned} (X - Y)\mathcal{X}_{\{a\}} &= 0 = (X - Y)\mathcal{X}_{\emptyset} \\ (X - Y)\mathcal{X}_{\{b,c\}} &= (X - Y) = (X - Y)\mathcal{X}_{\Omega} \end{aligned}$$

and we obtain:

$$\begin{aligned} E\left[(X - Y)\mathcal{X}_{\{a\}}\right] &= 0 = E\left[(X - Y)\mathcal{X}_{\emptyset}\right] \\ E\left[(X - Y)\mathcal{X}_{\{b,c\}}\right] &= E\left[(X - Y)\mathcal{X}_{\Omega}\right] \\ &= E\left[(X - Y)\right] \\ &= (-1)P(b) + 1P(c) = 0 \,. \end{aligned}$$ ▲

■ **EXAMPLE 6.20**

Let $\Omega = \{a, b, c\}$, $\Im = \wp(\Omega)$ and $P(\omega) = \frac{1}{3}$ for all $\omega \in \Omega$. Suppose that X is a real random variable given by

$$X(\omega) = \begin{cases} 0 & \text{if} \quad \omega = a \\ -1 & \text{if} \quad \omega = b \\ 1 & \text{if} \quad \omega = c \end{cases}$$

and let $\mathcal{G} := \{\emptyset, \Omega]$.

It is easy to verify that $Z := 0$ is $\mathcal{G}$-measurable and that it satisfies condition (6.1). Therefore, $Z = E(X \mid \mathcal{G})$. ▲

Definition 6.11 *We define:*

$$L_1 \quad := \quad \{X : X \textit{ is a real random variable defined over } (\Omega, \Im, P) \\ \textit{and with } E(|X|) < \infty\}.$$

In continuation we present some important properties of conditional expectation with respect to a σ-algebra:

Theorem 6.3 *Let $\mathcal{G}$ be a sub-σ-algebra of $\Im$. We have:*

1. *If $X, Y \in L_1$ and $\alpha, \beta \in \mathbb{R}$, then $E(\alpha X + \beta Y) = \alpha E(X) + \beta E(Y)$.*
2. *If X is $\mathcal{G}$-measurable and in L_1, then $E(X \mid \mathcal{G}) = X$. In particular, $E(c \mid \mathcal{G}) = c$ for all c real constant.*
3. *$E(X \mid \{\emptyset, \Omega\}) = E(X)$.*
4. *If $X \geq 0$ and $X \in L_1$, then $E(X \mid \mathcal{G}) \geq 0$.*
5. *If $X, Y \in L_1$ and $X \leq Y$, then $E(X \mid \mathcal{G}) \leq E(Y \mid \mathcal{G})$.*
6. *If $\mathcal{G}_1 \subseteq \mathcal{G}_2 \subseteq \Im$, then for all $X \in L_1$ we have that:*
$$E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) = E(X \mid \mathcal{G}_1) = E(E(X \mid \mathcal{G}_1) \mid \mathcal{G}_2).$$
7. *If $X \in L_1$, then $|E(X \mid \mathcal{G})| \leq E(|X| \mid \mathcal{G})$.*

Proof:

1. Since $Z = E(X \mid \mathcal{G})$ and $W = E(Y \mid \mathcal{G})$ are $\mathcal{G}$-measurable, then $\alpha Z + \beta W$ are also $\mathcal{G}$-measurable. From the definition of the conditional expectation, we have that for all $A \in \mathcal{G}$:
$$\begin{aligned} E((\alpha Z + \beta W)\mathcal{X}_A) &= E(\alpha Z \mathcal{X}_A + \beta W \mathcal{X}_A) \\ &= \alpha E(Z\mathcal{X}_A) + \beta E(W\mathcal{X}_A) \\ &= \alpha E(X\mathcal{X}_A) + \beta E(Y\mathcal{X}_A) \\ &= E((\alpha X + \beta Y)\mathcal{X}_A). \end{aligned}$$

2. By the hypothesis X is $\mathcal{G}$-measurable, and from the definition of conditional expectation, if $Z := E(X \mid \mathcal{G})$, then for all $A \in \mathcal{G}$:

$$E(Z\mathcal{X}_A) = E(X\mathcal{X}_A).$$

3. It is clear that $Z = E(X)$ is measurable with respect to $\Im_0 = \{\emptyset, \Omega\}$. On the other hand, if $A \in \{\emptyset, \Omega\}$ we have that $E(Z\mathcal{X}_A) = E(X\mathcal{X}_A)$.

4. Let $Z = E(X \mid \mathcal{G})$. By the definition, we have that, for all $A \in \mathcal{G}$,

$$E(Z\mathcal{X}_A) = E(X\mathcal{X}_A) \geq 0$$

since $X\mathcal{X}_A \geq 0$ because of $Z \geq 0$.

5. This result follows from the linearity of expectation and the previous result. It is clear that $E(X \mid \mathcal{G}_1)$ is $\mathcal{G}_1$-measurable. Let $Z = E(X \mid \mathcal{G}_2)$ and $W = E(Z \mid \mathcal{G}_1)$. If $A \in \mathcal{G}_1$, then:

$$E(Z\mathcal{X}_A) = E(W\mathcal{X}_A).$$

Since $\mathcal{G}_1 \subseteq \mathcal{G}_2$, then $A \in \mathcal{G}_2$, and it follows that:

$$E(Z\mathcal{X}_A) = E(X\mathcal{X}_A).$$

Therefore, for all $A \in \mathcal{G}_1$:

$$E(X\mathcal{X}_A) = E(W\mathcal{X}_A) .$$

That is,

$$W = E(X \mid \mathcal{G}_1)$$

and we get:

$$E(X \mid \mathcal{G}_1) = E(E(X \mid \mathcal{G}_2) \mid \mathcal{G}_1) .$$

Similarly, if $Y = E(X \mid \mathcal{G}_1)$ and $R = E(Y \mid \mathcal{G}_2)$, then for all $A \in \mathcal{G}_1$ it is true that:

$$E(X\mathcal{X}_A) = E(Y\mathcal{X}_A).$$

Since $A \in \mathcal{G}_2$, we have:

$$E(R\mathcal{X}_A) = E(Y\mathcal{X}_A).$$

Because of this, for all $A \in \mathcal{G}_1$ we have:

$$E(R\mathcal{X}_A) = E(X\mathcal{X}_A).$$

Thus:

$$E(X \mid \mathcal{G}_1) = E(E(X \mid \mathcal{G}_1) \mid \mathcal{G}_2).$$

In particular, if $\mathcal{G}_1 = \{\emptyset, \Omega\}$, then $E(E(X \mid \mathcal{G}_2)) = E(X)$.

6. Let X^+ and X^- be the positive and negative parts of X, respectively. That is:

$$X^+ := \max(X, 0) \text{ and } X^- := \max(-X, 0).$$

Because

$$|X| = X^+ + X^- \quad \text{and} \quad X = X^+ - X^-$$

we have:

$$\begin{aligned}
|E(X \mid \mathcal{G})| &= |E(X^+ - X^- \mid \mathcal{G})| \\
&= |E(X^+ \mid \mathcal{G}) - E(X^- \mid \mathcal{G})| \\
&\leq E(X^+ \mid \mathcal{G}) + E(X^- \mid \mathcal{G}) \\
&= E(X^+ + X^- \mid \mathcal{G}) \\
&= E(|X| \mid \mathcal{G}).
\end{aligned}$$

■

Finally we have the following property whose proof is beyond the scope of this text. Interested readers may refer to Jacod and Protter (2004).

Theorem 6.4 *Let $(\Omega, \Im, P)$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\Im$. If $X \in L_1$ and $(X_n)_{n\geq 1}$ is an increasing sequence of nonnegative real random variables defined over Ω that converges to X a.s., that is,*

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ with probability } 1,$$

then $(E(X_n \mid \mathcal{G}))_{n\geq 1}$ is an increasing sequence of random variables that converges to $E(X \mid \mathcal{G})$.

Theorem 6.5 *Let $(\Omega, \Im, P)$ be a probability space and let $\mathcal{G}$ be a sub-σ-algebra of $\Im$. If $(X_n)_{n\geq 1}$ is a sequence of real random variables in L_1 that converge in probability to 1 and if $|X_n| \leq Z$ for all n, where Z is a random variable in L_1, then:*

$$E\left(\lim_{n\to\infty} X_n \mid \mathcal{G}\right) = \lim_{n\to\infty} E(X_n \mid \mathcal{G}) \text{ with probability } 1.$$

Notation 6.1 *Let X, $Y_1, \cdots, Y_n$ be the real random variables. The expectation $E(X \mid \sigma(Y_1, \cdots, Y_n))$, where $\sigma(Y_1, \cdots, Y_n)$ is the smallest σ-algebra with respect to random variables $Y_1, \cdots, Y_n$, is usually denoted by $E(X \mid Y_1, \cdots, Y_n)$.*

Note 6.5 *Conditional expectation is a very useful application in Bayesian theory of statistics. A classic problem in this theory is obtained when observing data $\mathbf{X} := (X_1, \cdots, X_n)$ whose distribution is determined from the conditional distribution of $\mathbf{X}$ given $\Theta = \theta$, where Θ is considered as a random variable with a specific priori distribution. Using as a base the value of the data $\mathbf{X}$,* *the interesting problem is to estimate the unknown value of θ. An estimator of θ can be any function $d(\mathbf{X})$ of the data. In Bayesian theory we look for choosing $d(\mathbf{X})$ in such a way that the conditional expected value of the square of the distance between the estimator and the parameter is minimized. In other words we look for minimizing $E([\theta - d(\mathbf{X})]^2 \mid \mathbf{X})$.*

Conditioning on $\mathbf{X}$ leaves us with a constant $d(\mathbf{X})$. Along with this and the fact that for any random variable W we have that $E\left[(W-c)^2\right]$ is minimized when $c = E(W)$, we conclude that the estimator minimizing $E([\theta - d(\mathbf{X})]^2 \mid \mathbf{X})$ is given by $d(\mathbf{X}) = E(\theta \mid \mathbf{X})$. This estimator is called the Bayes estimator.

■ EXAMPLE 6.21

The height reached by the son of an individual with a height of x cm is a random variable with normal distribution with mean $x+3$ and variance 2. Which is the best prediction of the height that is expected for the son of the individual with height 170 cm?

Solution: let X be the random variable that denotes the height of the father and let θ be the random variable that denotes the height of the son. According to the information provided, $\theta \stackrel{d}{=} \mathcal{N}(X+3, 2)$. Due to the previous observation, it is known that the best possible predictor of the son's height is $d(X) = E(\theta \mid X)$. Therefore, if $X = 170$, then $\theta \stackrel{d}{=} \mathcal{N}(173, 2)$ and:

$$E(\theta \mid X = 170) = 173 \text{ cm} . \quad \blacktriangle$$

EXERCISES

6.1 Consider a sequence of Bernoulli trials. If the probability of success is a random variable with uniform distribution in the interval $(0, 1)$, what is the probability that n trials are needed?

6.2 Let X be a random variable with uniform distribution over the interval $(0, 1)$ and let Y be a random variable with uniform distribution over $(0, X)$. Determine:

a) The joint probability density function of X and Y.

b) The marginal density function of Y.

6.3 Let Y be a random variable with Poisson distribution with parameter λ. Suppose that Z is a random variable defined by

$$Z := \sum_{i=1}^{Y} X_i$$

where the random variables $X_1, X_2, \cdots$ are mutually independent and independent of Y. Further, suppose that the random variables $X_1, X_2, \cdots$ are identically distributed with Bernoulli distribution with parameter $p \in (0,1)$. Find $E(Z)$ and $Var(Z)$.

6.4 Let X and Y be random variables uniformly distributed over the triangular region limited by $x = 2$, $y = 0$ and $2y = x$, that is, the joint density function of the random variables X and Y is given by:

$$f(x,y) = \begin{cases} \frac{1}{\text{area of the triangle}} & \text{if } (x,y) \text{ is in the triangle} \\ 0 & \text{other cases}. \end{cases}$$

Calculate:

a) $P\left(Y \leq \frac{X}{3}\right)$.

b) $P(Y \geq 0.5)$.

c) $P(X \leq 1.5 \mid Y = 0.5)$.

6.5 The joint density function of X and Y is given by:

$$f(x,y) = \begin{cases} xe^{-x(y+1)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{other cases.} \end{cases}$$

Find the conditional density function of X given that $Y = y$ and the conditional density function for Y given that $X = x$.

6.6 Let $\mathbf{X} = (X, Y)$ be a random vector with density function given by:

$$f(x,y) = \frac{1}{2\pi \times 0.6} \exp\left[-\frac{1}{2 \times 0.36}\left((x-3)^2 - 1.6(x-3)(y-3) + (y-3)^2\right)\right].$$

Calculate:

a) The marginal density functions of X and Y.

b) The conditional density function $f_{Y|X}(y \mid X = 2)$.

c) The value of c so that $P(Y > c \mid X = 2) = 0.05$.

6.7 Suppose that X and Y are discrete random variables with joint probability distribution given by:

$X \backslash Y$	1	2	3	4
1	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
2	0	$\frac{2}{16}$	$\frac{1}{16}$	$\frac{1}{16}$
3	0	0	$\frac{3}{16}$	$\frac{1}{16}$
4	0	0	0	$\frac{4}{16}$

a) Calculate distributions of X and Y.

b) Determine $E(X \mid Y = 1)$ and $E(Y \mid X = 1)$.

6.8 Suppose that X and Y are discrete random variables with joint probability distribution given by:

$X\backslash Y$	1	2	3
1	$\frac{1}{12}$	$\frac{2}{12}$	0
2	0	$\frac{4}{12}$	$\frac{1}{12}$
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{2}{12}$

Verify that $E(E(X \mid Y)) = E(X)$ and $E(E(Y \mid X)) = E(Y)$.

6.9 A fair die is tossed twice consecutively. Let X be a random variable that denotes the number of even numbers obtained and Y the random variable that denotes the number of results obtained that are less than 4. Find $E(XE(Y \mid X))$.

6.10 If $E[Y/X] = 1$, show that:

$$Var[XY] \geq Var[X].$$

6.11 A box contains 8 red balls and 5 black ones. Two consecutive extractions are done without replacement. In the first extraction 2 balls are taken out while in the second extraction 3 balls are taken out. Let X be the random variable that denotes the number of red balls taken out in the first extraction and Y the random variable that denotes the number of red balls taken out in the second extraction. Find $E(Y \mid X = 1)$.

6.12 A player extracts 2 balls, one after the other one, from a box that contains 5 red balls and 4 black ones. For each red ball extracted the player wins two monetary units, and for each black ball extracted the player loses one monetary unit. Let X be a variable that denotes the player's fortune and Y be a random variable that takes the value 1 if the first ball extracted is red and the value 0 if the first ball extracted is black.

a) Calculate $E(X \mid Y)$.

b) Use part (a) to find $E(X)$.

6.13 Assume that taxis are waiting in a queue for passengers to come. Passengers for these taxis arrive independently with interarrival times that are exponentially distributed with mean 1 minute. A taxi departs as soon as two passengers have been collected or 3 minutes have expired since the first passenger has got in the taxi. Suppose you get in the taxi as the first passenger. What is your average waiting time for the departure?

6.14 Suppose you are in Ooty, India, as a tourist and lost at a point with five roads. Out of them, two roads bring you back to the same point after 1 hour of walk. The other two roads bring you back to the same point after 3 hours of travel. The last road leads to the center of the city after 2 hours of walk. Assume that there are no road sign. Assume that you choose a road equally likely at all times independent of earlier choices. What is the mean time until you arrive at the city?

6.15 Suppose that X is a discrete random variable with probability mass function given by $p_X(x) = \frac{x}{3}$, $x = 1, 2$ and Y is a random variable such that:

$$p_{Y|X}(y \mid x) = \binom{x}{y} \left(\frac{1}{2}\right)^x \quad \text{for } y = 0, \cdots, x \text{ and } x = 1, 2 .$$

Find:

a) The joint distribution of X and Y.

b) $E(X \mid Y)$.

6.16 If X has a Bernoulli distribution with parameter p and $E(Y \mid X = 0) = 1$ and $E(Y \mid X = 1) = 2$, what is $E(Y)$?

6.17 Suppose that the joint probability density function of the random variables X and Y is given by:

$$f(x, y) = \begin{cases} e^{-y} & \text{if } x > 0, \ y > x \\ 0 & \text{other cases} . \end{cases}$$

a) Calculate $P(X > 2 \mid Y < 4)$.

b) Calculate $E(X \mid Y = y)$.

c) Calculate $E(Y \mid X = x)$.

d) Verify that $E(X) = E(E(X \mid Y))$ and $E(Y) = E(E(Y \mid X))$.

6.18 Let X and Y be independent random variables. Prove that:

$$E(Y \mid X = x) = E(Y) \text{ for all } x.$$

6.19 Prove that if $E(Y \mid X = x) = E(Y)$ for all x, then X and Y are noncorrelated. Give a counterexample that shows the reciprocal is not true.

Suggestion: You can use the fact that $E(XY) = E(XE(Y \mid X))$.

6.20 Let X and Y be random variables with joint probability density function given by:

$$f(x, y) = \begin{cases} \frac{3}{8}\left(x + y^2\right) & \text{if } 0 < x < 2 , \ 0 < y < 1 \\ 0 & \text{other cases} . \end{cases}$$

Calculate $E(X \mid Y = \frac{1}{4})$.

6.21 The conditional variance of Y given $X = x$ is defined by:

$$Var\left(Y \mid X = x\right) := E\left(Y^2 \mid X = x\right) - \left(E\left(Y \mid X = x\right)\right)^2 .$$

Prove that:

$$Var(Y) = E\left(Var\left(Y \mid X\right)\right) + Var\left(E\left(Y \mid X\right)\right) .$$

6.22 Let X and Y be random variables with joint distribution given by:

$X \backslash Y$	0	1	2
0	$\frac{1}{8}$	$\frac{1}{8}$	$\frac{1}{8}$
1	0	$\frac{1}{8}$	$\frac{1}{8}$
2	0	0	$\frac{3}{8}$

Find $Var\left(Y \mid X\right)$.

6.23 Let X and Y be random variables with joint probability density function given by:

$$f\left(x, y\right) = \begin{cases} \frac{1}{8}\left(x^2 - y^2\right)\exp\left(-x\right) & \text{if } x > 0\text{ , } |y| < x \\ 0 & \text{other cases .} \end{cases}$$

Calculate $E(X \mid Y = 1)$.

6.24 Let (X, Y) be two-dimensional random variables with joint *pdf* given by:

$$f(x, y) = \begin{cases} e^{-y} & \text{if } 0 < x < y < \infty \\ 0 & \text{otherwise .} \end{cases}$$

a) Find the conditional distribution of Y given $X = x$.

b) Find the regression of Y on X.

c) Show that variance of Y for given $X = x$ does not involve x.

6.25 Suppose that the joint probability density function of the random variables X and Y is given by:

$$f\left(x, y\right) = \begin{cases} \frac{6}{7}\left(x^2 + \frac{xy}{2}\right) & \text{if } 0 < x < 1,\ 0 < y < 2 \\ 0 & \text{other cases .} \end{cases}$$

Calculate $E(X \mid Y = y)$.

6.26 Let (X, Y) be a random vector with uniform distribution in a triangle limited by $x \geq 0,\ y \geq 0$ and $x + y \leq 2$. Calculate $E\left(Y \mid X = x\right)$.

6.27 Let X and Y be random variables with joint probability density function given by:

$$f(x,y) = \begin{cases} 8xy & \text{if } 0 < y \leq x < 1 \\ 0 & \text{other cases .} \end{cases}$$

Calculate:

a) $E(X \mid Y = y)$.

b) $E(X^2 \mid Y = y)$.

c) $Var(X \mid Y = y)$.

6.28 Two fair dice are tossed simultaneously. Let X be the random variable that denotes the sum of the results obtained and B the event defined by $B :=$"the sum of the results obtained is divisible by 3". Calculate $E(X \mid B)$.

6.29 Let X and Y be i.i.d. random variables each with uniform distribution over the interval $(0, 2)$. Calculate:

a) $P(X \geq 1 \mid (X + Y) \leq 3)$.

b) $E(X \mid (X + Y) \leq 3)$.

6.30 Let X and Y be random variables with joint density function given by:

$$f(x,y) = \begin{cases} \exp(-x-y) & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{other cases.} \end{cases}$$

Calculate $E(X + Y \mid X < Y)$.

6.31 A fair die is thrown in a successive way. Let X and Y be random variables that denote, respectively, the number of throws required to obtain 2 and 4. Calculate:

a) $E(X)$.

b) $E(X \mid Y = 1)$.

c) $E(X \mid Y = 5)$.

6.32 A box contains 6 red balls and 5 white ones. Two samples are extracted in a consecutive way without replacement of sizes 3 and 5. Let X be the number of white balls in the first sample and Y the number of white balls in the second sample. Calculate $E(X \mid Y = k)$ for $k = 1, 2, 3, 4, 5$.

6.33 Let X be a random variable whose expected value exists. Prove that:

$$E(X) = E(X \mid X < y) P(X < y) + E(X \mid X \geq y) P(X \geq y) \ .$$

6.34 The conditional covariance of X and Y given Z is defined by:

$$Cov\,(X,Y \mid Z) := E\left[(X - E\,(X \mid Z))\,(Y - E\,(Y \mid Z)) \mid Z\right] .$$

a) Prove that:

$$Cov\,(X,Y \mid Z) = E\,(XY \mid Z) - E\,(X \mid Z)\,E\,(Y \mid Z) .$$

b) Verify that:

$$Cov\,(X,Y) = E\left[Cov\,(X,Y \mid Z)\right] + Cov\,(E\left[X \mid Z\right], E\left[Y \mid Z\right]) .$$

6.35 Let X_1 and X_2 be a two i.i.d. random variables each $\mathcal{N}(0,1)$ distributed.

a) Are $X_1 + X_2$ and $X_1 - X_2$ independent random variables? Justify your answers.

b) Obtain $E[X_1^2 + X_2^2 \mid X_1 + X_2 = t]$.

6.36 Let (X, Y) be two-dimensional random variable with joint *pdf*

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{2} y e^{-xy} & \text{if } 0 < x < \infty, 0 < y < 2 \\ 0 & \text{otherwise .} \end{cases}$$

a) Compute $E[(2X + 1) \mid Y = y]$.

b) Find the standard deviation of $[X \mid Y = y]$.

6.37 For $n \geq 1$, let $Y_1^{(n)}$, $Y_2^{(n)}, \ldots$ be i.i.d. random variables with values in $\mathbb{N}_0$, $n \in \mathbb{N}$. Suppose that $m := E(Y_1^{(n)}) < \infty$ and $0 < \sigma^2 := Var(Y_1^{(n)}) < \infty$. Let $Z_0 := 1$ and:

$$Z_{n+1} := \begin{cases} Y_1^{(n)} + Y_2^{(n)} + \cdots + Y_k^{(n)} & \text{if } \; Z_n = k, k > 0 \\ 0 & \text{if } \; Z_n = 0. \end{cases}$$

a) Calculate $E\,(Z_{n+1} \mid Z_n)$ and $E(Z_n)$.

b) Let $f\,(s) := E\left(s^{Z_1}\right) = \sum\limits_k P\,(Z_1 = k)\,s^k$ with $|s| \leq 1$, the probability generating function of Z_1 . Calculate $f_n(s) := E(s^{Z_n})$ in terms of f.

c) Find $Var(Z_n)$.

CHAPTER 7

MULTIVARIATE NORMAL DISTRIBUTIONS

The multivariate normal distributions is one of the most important multidimensional distributions and is essential to multivariate statistics. The multivariate normal distribution is an extension of the univariate normal distribution and shares many of its features. This distribution can be completely described by its means, variances and covariances given in this chapter. The brief introduction to this distribution given will be necessary for students who wish to take the next course in multivariate statistics but can be skipped otherwise.

7.1 MULTIVARIATE NORMAL DISTRIBUTION

Definition 7.1 (Multivariate Normal Distribution) *An n-dimensional random vector $\mathbf{X} = (X_1, \cdots, X_n)$ is said to have a multivariate normal distribution if any linear combination $\sum_{j=1}^{n} \alpha_j X_j$ has a univariate normal distribution (possibly degenerated, as happens, for example, when $\alpha_j = 0$ for all j).*

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

■ **EXAMPLE 7.1**

Suppose that $X_1, \cdots, X_n$ are n independent random variables such that $X_j \stackrel{d}{=} \mathcal{N}(\mu_j, \sigma_j^2)$ for $j = 1, \cdots, n$. Then, if $Y = \sum_{j=1}^{n} \alpha_j X_j$, we have

$$\begin{aligned}
\varphi_Y(t) &= E\left(e^{itY}\right) \\
&= E\left(e^{it[\alpha_1 X_1 + \cdots + \alpha_n X_n]}\right) \\
&= \prod_{j=1}^{n} \Phi_{X_j}(\alpha_j t) \\
&= \prod_{j=1}^{n} \exp\left[i\mu_j\, \alpha_j\, t - \frac{\alpha_j^2\, t^2 \sigma_j^2}{2}\right] \\
&= \exp\left(\sum_{j=1}^{n} \left[i\mu_j\, \alpha_j\, t - \frac{\alpha_j^2\, t^2 \sigma_j^2}{2}\right]\right) \\
&= \exp\left(i\mu t - \frac{1}{2}\sigma^2\, t^2\right)
\end{aligned}$$

where

$$\mu := \sum_{j=1}^{n} \mu_j\, \alpha_j \quad \text{and} \quad \sigma^2 := \sum_{j=1}^{n} \alpha_j^2\, \sigma_j^2 \, .$$

In other words, $Y \stackrel{d}{=} \mathcal{N}(\mu, \sigma^2)$.

Therefore, the vector $\mathbf{X} := (X_1, \cdots, X_n)$ has a multivariate normal distribution. ▲

Note 7.1 *In this chapter, the vector in* $\mathbb{R}^n$ *is represented by a row vector.*

Theorem 7.1 *Let* $\mathbf{X} := (X_1, \cdots, X_n)$ *be a random vector.* $\mathbf{X}$ *has multivariate normal distribution if and only if its characteristic function has the form*

$$\varphi_{\mathbf{X}}(\mathbf{t}) = \exp\left[i\,\langle \mathbf{t}, \mu\rangle - \frac{1}{2}\,\langle \mathbf{t}, \mathbf{t}\Sigma\rangle\right]$$

where $\mu \in \mathbb{R}^n$, Σ *is a positive semidefinite symmetric square matrix and* $\langle\cdot,\cdot\rangle$ *represents the usual inner product of* $\mathbb{R}^n$.

Proof:

$\Longrightarrow$) Let $Y := \sum_{j=1}^{n} \alpha_j X_j$. Clearly

$$
\begin{aligned}
E(Y) &= \sum_{j=1}^{n} \alpha_j \mu_j \\
&= \langle \alpha, \mu \rangle , \quad \text{where } \mu := E(\mathbf{X}) \text{ and } \alpha := (\alpha_1, \cdots, \alpha_n),
\end{aligned}
$$

and

$$
\begin{aligned}
Var(Y) &= Var\left(\sum_{j=1}^{n} \alpha_j X_j\right) \\
&= \sum_{j=1}^{n} Var(\alpha_j X_j) + 2\sum\sum_{j<i} Cov(\alpha_j X_j, \alpha_i X_i) \\
&= \sum_{j=1}^{n} \alpha_j^2 Var(X_j) + 2\sum\sum_{j<i} \alpha_j \alpha_i Cov(X_j, X_i) \\
&= \langle \alpha, \alpha\Sigma \rangle ,
\end{aligned}
$$

where $\alpha := (\alpha_1, \cdots, \alpha_n)$ and Σ is the variance-covariance matrix of $\mathbf{X}$.
Therefore, the characteristic function of Y equals:

$$
\varphi_Y(t) = \exp\left[it \langle \alpha, \mu \rangle - \frac{t^2}{2} \langle \alpha, \alpha\Sigma \rangle\right] .
$$

Then:

$$
\begin{aligned}
\varphi_{\mathbf{X}}(\alpha) &= E\left[\exp\left(i\mathbf{X}\alpha^T\right)\right] \\
&= E\left[\exp\left(i\sum_{j=1}^{n} \alpha_j X_j\right)\right] \\
&= \varphi_Y(1) \\
&= \exp\left[i \langle \alpha, \mu \rangle - \frac{1}{2} \langle \alpha, \alpha\Sigma \rangle\right].
\end{aligned}
$$

$\Longleftarrow$) Let $Y := \sum_{j=1}^{n} \alpha_j X_j$. The characteristic function of Y is then given by

$$\begin{aligned}\varphi_Y(t) &= E\left[\exp\left(itY\right)\right] \\ &= E\left[\exp\left(it\sum_{j=1}^{n}\alpha_j X_j\right)\right] \\ &= E\left[\exp\left(i\mathbf{X}\beta^T\right)\right]\end{aligned}$$

where $\beta := t\alpha$. That is:

$$\begin{aligned}\varphi_Y(t) = \varphi_{\mathbf{X}}(\beta) &= \varphi_{\mathbf{X}}(t\alpha) \\ &= \exp\left[i\left\langle t\alpha, \mu\right\rangle - \frac{1}{2}\left\langle t\alpha, t\alpha\,\Sigma\right\rangle\right] \\ &= \exp\left[it\left\langle \alpha, \mu\right\rangle - \frac{1}{2}t^2\left\langle \alpha, \alpha\,\Sigma\right\rangle\right] .\end{aligned}$$

Then Y has a univariate normal distribution with parameters $\langle\alpha, \mu\rangle$ and $\langle\alpha, \alpha\Sigma\rangle$. Therefore, $\mathbf{X}$ is multivariate normal. ∎

Note 7.2 *It can be easily verified that the vector μ and the matrix Σ from the previous theorem correspond to the expected value and the variance-covariance matrix of $\mathbf{X}$, respectively.*

Notation 7.1 *If $\mathbf{X}$ has a multivariate normal distribution with mean vector μ and variance-covariance matrix Σ, then we write $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$.*

Our next theorem states that any multivariate normal distribution can be obtained by applying a linear transformation to a random vector whose components are independent random variables having all univariate normal distributions. In order to prove this result, the following lemma is required:

Lemma 7.1 *Let $\mathbf{X} := (X_1, \cdots, X_n)$ be a random vector such that $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$. The components X_j, $j = 1, \cdots, n$, are independent if and only if the matrix Σ is diagonal.*

Proof:

$\Longrightarrow$) See the result given in (5.17).

$\Longleftarrow$) Suppose that the matrix Σ is diagonal. Since $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$, then:

$$\begin{aligned}
\varphi_{\mathbf{X}}(\mathbf{t}) &= \exp\left[i\langle \mathbf{t}, \mu\rangle - \frac{1}{2}\langle \mathbf{t}, \mathbf{t}\Sigma\rangle\right] \\
&= \exp\left[i\sum_{j=1}^{n}\mu_j t_j - \frac{1}{2}\sum_{j=1}^{n}\sigma_j t_j^2\right] \\
&= \exp\left[\sum_{j=1}^{n}\left(i\mu_j t_j - \frac{1}{2}\sigma_j t_j^2\right)\right] \\
&= \prod_{j=1}^{n}\exp\left[\left(i\mu_j t_j - \frac{1}{2}\sigma_j t_j^2\right)\right] \\
&= \prod_{j=1}^{n}\varphi_{X_j}(t_j) .
\end{aligned}$$

Therefore, the random variables X_j for $j = 1, \cdots, n$, are independent. ∎

Theorem 7.2 *Let $\mathbf{X} := (X_1, \cdots, X_n)$ be a random vector such that $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$. Then there exist an orthogonal matrix A and independent random variables $Y_1, \cdots, Y_n$ such that either $Y_j = 0$ or $Y_j \stackrel{d}{=} \mathcal{N}(0, \lambda_j)$ for $j = 1, \cdots, n$ so that $\mathbf{X} = \mu + YA$.*

Proof: Since Σ is a positive semidefinite symmetric matrix, there exist a diagonal matrix Λ whose entries are all nonnegative and an orthogonal matrix A such that:

$$\Sigma = A\Lambda A^T.$$

Let $\mathbf{Y} := (\mathbf{X} - \mu)A^T$. Since $\mathbf{X}$ is multivariate normal, so is $\mathbf{Y}$. Additionally, Λ is the variance-covariance matrix of $\mathbf{Y}$. Since this matrix is diagonal, it follows from the previous lemma that the components of $\mathbf{Y}$ are independent. Finally we have:

$$\mathbf{X} = \mu + \mathbf{YA} .$$

∎

Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ is an n-dimensional random vector and that the random variables $X_1, \cdots, X_n$ are independent and identically distributed having a standard normal distribution. The joint probability density function of $X_1, \cdots, X_n$ is given by:

$$\begin{aligned}
f(x_1, \cdots, x_n) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\
&= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_1^2\right) \cdots \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x_n^2\right) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\sum_{j=1}^{n} x_j^2\right) \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}\mathbf{x}^T\right) , \quad \text{where } \mathbf{x} :=(x_1, \cdots, x_n) .
\end{aligned}$$

In addition, it is clear that the vector $\mathbf{X}$ has a multivariate normal distribution. The natural question that arises is: If $\mathbf{X}$ is a random vector with multivariate normal distribution, under what conditions can the existence of a density function for the vector $\mathbf{X}$ be guaranteed? The answer is given in the following theorem:

Theorem 7.3 *Let* $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$. *If* Σ *is a positive definite matrix, then* $\mathbf{X}$ *has a density function given by:*

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}} (\det \Sigma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mu)\Sigma^{-1}(\mathbf{x} - \mu)^T\right] .$$

Proof: Since Σ is a positive definite matrix, all its eigenvalues are positive. Moreover, there exists an orthogonal matrix U such that

$$U\Sigma U^T = \Lambda$$

where $\Lambda = diag(\lambda_i)$ and $\lambda_1, \cdots, \lambda_n$ are the eigenvalues of Σ. In other words, Λ is the diagonal matrix whose entries on the diagonal are precisely the eigenvalues of Σ.

Let $A := U\, diag(\sqrt{\lambda_i})U^T$. Clearly $A^T A = \Sigma$ and A is also a positive definite matrix. Let $\mathbf{h} : \mathbb{R}^{1\times n} \longrightarrow \mathbb{R}^{1\times n}$ be defined by $\mathbf{h}(\mathbf{x}) = \mathbf{x}A + \mu$. The inverse function of $\mathbf{h}$ would then be given by $\mathbf{h}^{-1}(\mathbf{x}) = (\mathbf{x} - \mu)A^{-1}$. The transformation theorem implies that the density function of $\mathbf{X} := \mathbf{Y}A + \mu$, where $\mathbf{Y} = (Y_1, \cdots, Y_n)$, is an n-dimensional random vector such that the random variables $Y_1, \cdots, Y_n$ are independent and identically distributed with a standard normal distribution and is given by:

$$
\begin{aligned}
f_{\mathbf{X}}(\mathbf{x}) &= f_{\mathbf{Y}}\left(h^{-1}(\mathbf{x})\right)\left|\det\frac{\partial h^{-1}}{\partial \mathbf{x}}\right| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\left[h^{-1}(\mathbf{x})\right]\left[h^{-1}(\mathbf{x})\right]^T\right)\left|\det\frac{\partial h^{-1}}{\partial \mathbf{x}}\right| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\left[(\mathbf{x}-\mu)A^{-1}\right]\left[(\mathbf{x}-\mu)A^{-1}\right]^T\right)\left|\det A^{-1}\right| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\left[(\mathbf{x}-\mu)\left(A^TA\right)^{-1}(\mathbf{x}-\mu)^T\right]\right)\left|\det A^{-1}\right| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}\exp\left(-\frac{1}{2}\left[(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T\right]\right)\left|\det A^{-1}\right| \\
&= \frac{1}{(2\pi)^{\frac{n}{2}}}(\det\Sigma)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T\right].
\end{aligned}
$$

■

Note 7.3 (Bivariate Normal Distribution) *As a particular case of the theorem above, suppose that*

$$\mathbf{X} = (X_1, X_2) \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$$

where

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}$$

$$\mu = (\mu_1, \mu_2)\ , \quad \mu_1 := EX_1\ , \quad \mu_2 := EX_2$$

and

$$
\begin{aligned}
\sigma_1^2 &:= Var(X_1)\ , \quad \sigma_2^2 := Var(X_2) \\
\sigma_{12} &= Cov(X_1, X_2) = \sigma_{21} = \rho\sigma_1\sigma_2
\end{aligned}
$$

with ρ representing the correlation coefficient. Therefore:

$$f(\mathbf{x}) = \frac{1}{2\pi}(\det\Sigma)^{-\frac{1}{2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T\right].$$

Since

$$\det\Sigma = \sigma_1^2\sigma_2^2 - \sigma_{12}^2 = \sigma_1^2\sigma_2^2\left(1-\rho^2\right)$$

and

$$\Sigma^{-1} = \frac{1}{\sigma_1^2\sigma_2^2\left(1-\rho^2\right)}\begin{pmatrix} \sigma_2^2 & -\sigma_{12} \\ -\sigma_{21} & \sigma_1^2 \end{pmatrix}$$

we obtain:

$$f(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left[-\tfrac{1}{2(1-\rho^2)}\left\{\left(\tfrac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\tfrac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right) + \left(\tfrac{x_2-\mu_2}{\sigma_2}\right)^2\right\}\right] .$$

We also have that:

$$\begin{aligned} f_{X_1}(x_1) &= \int_{-\infty}^{\infty} f(x_1,x_2)\,dx_2 \\ &= \frac{1}{\sqrt{2\pi}\sigma_1}\exp\left[-\frac{1}{2}\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2\right] \\ f_{X_2}(x_2) &= \int_{-\infty}^{\infty} f(x_1,x_2)\,dx_1 \\ &= \frac{1}{\sqrt{2\pi}\sigma_2}\exp\left[-\frac{1}{2}\left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right] . \end{aligned}$$

In other words, the marginal distributions of $\mathbf{X} = (X_1, X_2)$ *are univariate normal.*

In general, we have:

Theorem 7.4 *All the marginal distributions of* $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$ *are multivariate normal.*

Proof: Suppose that $\mathbf{X} = (X_1, \cdots, X_n)$ and let

$$\widetilde{\mathbf{X}} := (X_{k_1}, \cdots, X_{k_l})$$

where $\{k_1, \cdots, k_l\}$ is a subset of $\{1, \cdots, n\}$. The characteristic function of $\widetilde{\mathbf{X}}$ is given by:

$$\begin{aligned} \varphi_{\widetilde{\mathbf{X}}}(t_{k_1}, \cdots, t_{k_l}) &= E\left(i\sum_{r=1}^{l} t_{k_r} X_{t_{k_r}}\right) \\ &= \varphi_{\mathbf{X}}(t_1, \cdots, t_n) , \quad \text{where } t_j = 0 \text{ if } j \notin \{k_1, \cdots, k_l\} . \end{aligned}$$

Therefore $\widetilde{\mathbf{X}}$ has a multivariate normal distribution. ∎

7.2 DISTRIBUTION OF QUADRATIC FORMS OF MULTIVARIATE NORMAL VECTORS

Let X_i, $i = 1, 2, \cdots, n$ be independent normal random variables with $\mathcal{N}(\mu_i, \sigma_i^2)$, $i = 1, 2, \cdots, n$. It is known that:

$$Y := \sum_{i=1}^{n} \frac{(X_i - \mu_i)^2}{\sigma_i^2} \stackrel{d}{=} \chi^2_{(n)}.$$

Suppose now an n-dimensional random vector $X = (X_1, X_2, \cdots, X_n)$ having multivariate normal distribution with mean vector μ and variance and covariance matrix Σ. Suppose that Σ is a positive definite matrix. From Theorem 7.3, it is known that X has the density function given by

$$f_X(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{n}{2}}}(det\Sigma)^{-\frac{1}{2}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T\right]$$

with $\mathbf{x} \in \mathbb{R}^n$. Now we are interested in finding the distribution of $W = (\mathbf{X}-\mu)\Sigma^{-1}(\mathbf{X}-\mu)^T$. In order to do so, we need to find the moment generating function of W. We have:

$$\begin{aligned}
m_W(t) &= E(e^{Wt}) \\
&= \int e^{t(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T} f_X(\mathbf{x})d\mathbf{x} \\
&= \int e^{t(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T} \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\Sigma}} \exp\left[-\frac{1}{2}(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T\right] d\mathbf{x} \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\Sigma}} \\
&\quad \exp\left[t(x-\mu)\Sigma^{-1}(x-\mu)^T - \frac{1}{2}(x-\mu)\Sigma^{-1}(x-\mu)^T\right] dx_1, dx_2, \cdots, dx_n \\
&= \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\Sigma}} \\
&\quad \exp\left[\frac{-(x-\mu)\Sigma^{-1}(x-\mu)^T(1-2t)}{2}\right] dx_1, dx_2, \cdots, dx_n \ .
\end{aligned}$$

This last integral exists for all values of $t < \frac{1}{2}$.

Now the matrix $(1-2t)\Sigma^{-1}$, $t < \frac{1}{2}$, is positive definite given that Σ is also a positive definite matrix.

On the other hand,

$$det\left[(1-2t)\Sigma^{-1}\right] = (1-2t)^n det(\Sigma^{-1})$$

and consequently the function

$$\frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{\frac{det\Sigma}{(1-2t)^n}}} \exp\left\{\frac{-(\mathbf{x}-\mu)\Sigma^{-1}(\mathbf{x}-\mu)^T(1-2t)}{2}\right\}$$

is the density function of the multivariate normal random variable. When multiplying and dividing the denominator by $(1-2t)^{n/2}$ in the expression given for $m(t)$ we obtain

$$\begin{aligned} m(t) &= (1-2t)^{-\frac{n}{2}} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\frac{det\Sigma}{(1-2t)^n}}} \\ &\quad \exp\left\{ \frac{-(x-\mu)\Sigma^{-1}(x-\mu)^T(1-2t)}{2} \right\} dx_1, dx_2, \cdots, dx_n. \\ &= (1-2t)^{-\frac{n}{2}} \\ &= \frac{1}{(1-2t)^{\frac{n}{2}}}, \quad t < \frac{1}{2}, \end{aligned}$$

which corresponds to the mgf of a random variables with $\mathcal{X}^2_{(n)}$ distribution. We also have that $W \stackrel{d}{=} \mathcal{X}^2_{(n)}$.

Suppose that $X_1, X_2, \cdots, X_n$ are independent with normal distribution $\mathcal{N}(0, \sigma^2)$. Let $X = (X_1, X_2, \cdots, X_n)$ and suppose that A is a real symmetric matrix of order n. We want to find the distribution of XAX^T. In order to find this distribution, the mgf of the variable $\frac{XAX^T}{\sigma^2}$ must be considered. It is clear that:

$$\begin{aligned} m(t) &= E\left(e^{t\frac{XAX^T}{\sigma^2}} \right) \\ &= \int e^{t\frac{\mathbf{x}A\mathbf{x}^T}{\sigma^2}} \frac{1}{(2\pi)^{\frac{n}{2}}\sqrt{det\Sigma}} \cdot e^{-\frac{\mathbf{x}\Sigma^{-1}\mathbf{x}^T}{2}} \, d\mathbf{x} \, . \end{aligned}$$

Given that the random variables $X_1, X_2, \cdots X_n$ are independent with normal distribution $\mathcal{N}(0, \sigma^2)$, we have in this case that

$$\Sigma = \begin{pmatrix} \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \sigma^2 \end{pmatrix}, \quad det\Sigma = \sigma^{2n} \quad \text{and} \quad \Sigma^{-1} = \frac{1}{\sigma^2} I$$

where I_n is the identity matrix of order n.

Therefore:

$$\begin{aligned} m(t) &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{t\mathbf{x}A\mathbf{x}^T}{\sigma^2} - \frac{\mathbf{x}\cdot\mathbf{x}^T}{2\sigma^2} \right] dx_1, dx_2, \cdots, dx_n \\ &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \frac{1}{(\sigma\sqrt{2\pi})^n} \exp\left[\frac{-\mathbf{x}(I-2tA)\mathbf{x}^T}{2\sigma^2} \right] dx_1, dx_2, \cdots, dx_n \, . \end{aligned}$$

Given that $I - 2tA$ is a positive definite matrix and if $|t|$ is sufficiently small, let's say $|t| < h$, we have that the function

$$\frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{det((I-2tA)^{-1}\sigma^2)}} \exp\left[-\frac{\mathbf{x}(I-2tA)\mathbf{x}^T}{2\sigma^2} \right]$$

is the density function of the multivariate normal distribution. Thus:

$$m(t) = [det(I - 2tA)]^{-\frac{1}{2}} \quad \text{with} \quad |t| < h \; .$$

Suppose now that $\lambda_1, \lambda_2, \cdots, \lambda_n$ are eigenvalues of A and let L be an orthogonal matrix of order n such that $L^T AL = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$. Then

$$L^T(I - 2tA)L = \begin{pmatrix} 1 - 2t\lambda_1 & & 0 \\ & \ddots & \\ 0 & & 1 - 2t\lambda_n \end{pmatrix}$$

and therefore:

$$det\left(L^T(I - 2tA)L\right) = \prod_{j=1}^{n}(1 - 2t\lambda_j) \; .$$

Given that

$$det\left(L^T(I - 2tA)L\right) = det(I - 2tA)$$

and because L is an orthogonal matrix, we have

$$det\,(1 - 2tA)) = \prod_{j=1}^{n}(1 - 2t\lambda_j)$$

from which we obtain that:

$$m(t) = \left[\prod_{j=1}^{n}(1 - 2t\lambda_j)\right]^{-\frac{1}{2}} \; , \quad |t| < h \; .$$

Suppose that r is the rank of matrix A with $0 < r \leq n$. Then we have that exactly r of the numbers $\lambda_1, \lambda_2, \cdots, \lambda_n$, let's say $\lambda_1, \cdots, \lambda_r$, are different from zero and the remaining $n - r$ of them are zero. Therefore:

$$m(t) = [(1 - 2t\lambda_1) \cdots (1 - 2t\lambda_r)]^{-\frac{1}{2}} \quad \text{with} \quad |t| < h \; .$$

Under which conditions does the previous mgf correspond to the mgf of a random variable with chi-squared distribution of k degrees of freedom? If this is to be so, then we must have:

$$m(t) = [(1 - 2t\lambda_1) \cdots (1 - 2t\lambda_r)]^{-\frac{1}{2}} = (1 - 2t)^{-\frac{k}{2}} \quad \text{with} \quad |t| < h \; .$$

This implies $(1 - 2t\lambda_1) \cdots (1 - 2t\lambda_r) = (1 - 2t)^k$ and in consequence $k = r$ and $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$. That is, matrix A has r eigenvalues equal to 1 and the other $n - r$ equal to zero, and the rank of the matrix A is r. This implies that matrix A must be *idempotent*,that is, $A^2 = A$.

Conversely, if matrix A has rank r and is *idempotent*, then A has r eigenvalues equal to 1 and $n-r$ real eigenvalues equal to zero and in consequence the mgf of $\frac{XAX^T}{\sigma^2}$ is given by:

$$m(t) = (1-2t)^{-\frac{r}{2}} \quad \text{if } t < \frac{1}{2} \, .$$

In summary:

Theorem 7.5 *Let $X_1, X_2, \cdots, X_n$ be i.i.d. random variables with $\mathcal{N}(0, \sigma^2)$. Let $X = (X_1, X_2, \cdots, X_n)$ and A be a symmetric matrix of order n with rank r. Suppose that $Y := XAX^T$. Then:*

$$\frac{Y}{\sigma^2} \stackrel{d}{=} \mathcal{X}^2_{(r)} \quad \textit{iff} \quad A^2 = A \, .$$

■ EXAMPLE 7.2

Let $Y = X_1X_2 - X_3X_4$ where X_1, X_2, X_3, X_4 are i.i.d. random variables with $\mathcal{N}(0, \sigma^2)$. Is the distribution of the random variable $\frac{Y}{\sigma^2}$ a chi-square distribution? Explain.

Solution: It is clear that $Y = XAX^T$ with:

$$A = \begin{pmatrix} 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} \\ 0 & 0 & -\frac{1}{2} & 0 \end{pmatrix}$$

The random variable $\frac{Y}{\sigma^2}$ does not have $\mathcal{X}^2$ distribution because $A^2 \neq A$.
▲

Suppose that $X_1, X_2, \cdots, X_n$ are i.i.d. random variables with $\mathcal{N}(0, \sigma^2)$. Let A and B be two symmetric matrices of order n and consider the quadratic forms XAX^T and XBX^T. Under what conditions are these quadratic forms independent? To answer this question we must consider the joint mgf of $\frac{XAX^T}{\sigma^2}$ and $\frac{XBX^T}{\sigma^2}$. We have then:

$$\begin{aligned} m(t_1, t_2) &= E\left(e^{\frac{t_1 XAX^T}{\sigma^2} + \frac{t_2 XBX^T}{\sigma^2}}\right) \\ &= \int e^{\frac{t_1 \mathbf{x}A\mathbf{x}^T}{\sigma^2} + \frac{t_2 \mathbf{x}B\mathbf{x}^T}{\sigma^2}} f_X(\mathbf{x}) d\mathbf{x} \\ &= \int e^{\frac{t_1 XAX^T}{\sigma^2} + \frac{t_2 XBX^T}{\sigma^2}} \frac{1}{(2\pi)^{\frac{n}{2}} (det\Sigma)^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{x}\Sigma^{-1}\mathbf{x}^T}{2}\right) d\mathbf{x} \, . \end{aligned}$$

In this case, we have that $det\Sigma = \sigma^{2n}$ and $\Sigma^{-1} = \frac{1}{\sigma^2}I$. So that:

$$\begin{aligned} m(t_1,t_2) &= \int e^{\frac{t_1 \mathbf{x}A\mathbf{x}^T + t_2 \mathbf{x}B\mathbf{x}^T}{\sigma^2}} \frac{1}{\left(\sqrt{2\pi}\sigma\right)^n} \exp\left(-\frac{\mathbf{x}\cdot\mathbf{x}^T}{2\sigma^2}\right) dx \\ &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\left(\frac{t_1 \mathbf{x}A\mathbf{x}^T}{\sigma^2} + \frac{t_2 \mathbf{x}B\mathbf{x}^T}{\sigma^2} - \frac{\mathbf{x}\cdot\mathbf{x}^T}{2\sigma^2}\right)} dx_1, dx_2, \cdots, dx_n \\ &= \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{\left(-\frac{\mathbf{x}(I-2t_1A-2t_2B)\mathbf{x}^T}{2\sigma^2}\right)} dx_1, dx_2, \cdots, dx_n \; . \end{aligned}$$

The matrix $I - 2t_1A - 2t_2B$ is a positive definite matrix if $|t_1|$ and $|t_2|$ are sufficiently small, for example, $|t_1| < h_1$ and $|t_2| < h_2$ with $h_1, h_2 > 0$. Hence:

$$m(t_1,t_2) = [det(I - 2t_1A - 2t_2B)]^{-\frac{1}{2}} \; .$$

If XAX^T and XBX^T are stochastically independent, then $AB = 0$. Indeed, if XAX^T and XBX^T are independent, then $m(t_1,t_2) = m(t_1,0)\cdot m(0,t_2)$ for all t_1, t_2 with $|t_1| < h_1$ and $|t_2| < h_2$. That is,

$$det(I - 2t_1A - 2t_2B) = det(I - 2t_1A)\cdot det(I - 2t_2B)$$

where t_1, t_2 satisfy $|t_1| < h_1$ and $|t_2| < h_2$.

Let $r = rank(A)$ and suppose that $\lambda_1, \lambda_2, \ldots, \lambda_r$ are r eigenvalues of A different than zero. Then there exists an orthogonal matrix L such that:

$$L^TAL = \begin{pmatrix} C_{11} & 0 \\ \vdots & \\ 0 & 0 \end{pmatrix} = C \text{ where } C_{11} = \begin{pmatrix} \lambda_1 & 0 \\ \vdots & \\ 0 & \lambda_r \end{pmatrix} .$$

Suppose that $L^TBL = \begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} = D$. Then, the equation

$$det(I - 2t_1A - 2t_2B) = det(I - 2t_1A)\cdot det(I - 2t_2B)$$

may be rewritten as

$$\begin{aligned} det(L^T)det(I - 2t_1A - 2t_2B)det(L) &= det(L^T)det(I - 2t_1A)(detL)\cdot \\ &\quad\; det(L^T)det(I - 2t_2B)detL \end{aligned}$$

or equivalently:

$$det\left(L^T(I - 2t_1A - 2t_2B)L\right) = det(L^T(I - 2t_1A)L)det(L^T(I - 2t_2B)L) \; .$$

That is:

$$det(I - 2t_1C - 2t_2D) = det(I - 2t_1C)det(I - 2t_2D) \; .$$

Given that the coefficient of $(-2t_1)^r$ on the right side of the previous equation is $\lambda_1\lambda_2\cdots\lambda_r det(I-2t_2D)$ and the coefficient of $(-2t_1)^r$ on the left side of the equation is

$$\lambda_1\lambda_2\cdots\lambda_r det(I_{n-r}-2t_2D_{22})$$

where I_{n-1} is the $(n-r)$-order identity matrix, then, for all t_2 with $|t_2|<h_2$, $det(I-2t_2D)=det(I_{n-r}-2t_2D_{22})$ must be satisfied and consequently the nonzero eigenvalues of the matrices D and D_{22} are equal.

On the other hand, if $A=(a_{ij})_{n\times n}$ is a symmetric matrix, then $\sum\limits_j\sum\limits_i a_{ij}^2$ is equal to the sum of the squares of the eigenvalues of A. Indeed, let L be such that $L^TAL=diag(\lambda_1,\lambda_2,\cdots,\lambda_n)$. Then:

$$tr\left(L^TAL)(L^TAL)\right)=tr(L^TA^2L)=tr(A^2)=\sum_j\sum_i a_{ij}^2 .$$

Therefore, the sum of the squares of the elements of matrix D is equal to the sum of the squares of the elements of matrix D_{22}. Thus:

$$D=\begin{pmatrix}0 & 0\\ 0 & D_{22}\end{pmatrix}.$$

Now $0=CD=L^TAL\cdot L^TBL=L^TABL$ and in consequence $AB=0$.

Suppose now that $AB=0$. Let us verify that $\frac{XAX^T}{\sigma^2}$ and $\frac{XBX^T}{\sigma^2}$ are stochastically independent. We have that:

$$\begin{aligned}(I-2t_1A)(I-2t_2B) &= I-2t_2B-2t_1A+4t_1t_2AB\\ &= I-2t_2B-2t_1A .\end{aligned}$$

That is:

$$det(I-2t_2B-2t_1A)=det(I-2t_1A)det(I-2t_2B) .$$

Therefore:

$$m(t_1,t_2)=m(t_1,0)\cdot m(0,t_2) .$$

In summary, we have the following result:

Theorem 7.6 *Let $X_1,X_2,\cdots,X_n$ - i.i.d. random variables with $\mathcal{N}(0,\sigma^2)$. Let A and B be symmetric matrices and $X=(X_1,X_2,\cdots,X_n)$. Then, the quadratic forms XAX^T and XBX^T are independent if and only if $AB=0$.*

EXERCISES

7.1 Let $\mathbf{X}=(X,Y)$ be a random vector having a bivariate normal distribution with parameters $\mu_X=2$, $\mu_Y=3.1$, $\sigma_X=0.001$, $\sigma_Y=0.02$ and $\rho=0$. Find:

$$P(1.5<X<2.3,\ 7.3<Y<7.8) .$$

7.2 Suppose that X_1 and X_2 are independent $\mathcal{N}(0,1)$ random variables. Let $Y_1 = X_1 + 3X_2 - 2$ and $Y_2 = X_1 - 2X_2 + 1$. Determine the distribution of $\mathbf{Y} = (Y_1, Y_2)$.

7.3 Let $\mathbf{X} = (X_1, X_2)$ be a multivariate normal with $\mu = (5, 10)$ and $\mathbf{\Sigma} = \begin{pmatrix} 1 & \alpha \\ \alpha & 4 \end{pmatrix}$. If $Y_1 = 2X_1 + 2X_2 + 1$ and $Y_2 = 3X_1 - 2X_2 - 2$ are independent, determine the value of α.

7.4 Let $\mathbf{X} = (X_1, \cdots, X_n)$ be an n-dimensional random vector such that $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$, where Σ is a nonsingular matrix. Prove that

$$\mathbf{Y} := (\mathbf{X} - \mu)\, W^{-1}$$

is a random vector with a $\mathcal{N}(\mathbf{0}, I)$ distribution, where I is the identity matrix of order n and W is a matrix satisfying $W^2 = \Sigma$. In this case, we say that the vector $\mathbf{Y}$ has a standard multivariate normal distribution.

7.5 Let $\mathbf{X} = (X, Y)$ be a random vector with bivariate normal distribution. Prove that the conditional distribution of Y, given that $X = x$, is normal with parameters μ given by

$$\mu = E(Y) + \rho(X,Y) \frac{\sqrt{VarX}}{\sqrt{VarY}} (X - EX)$$

and σ^2 given by

$$\sigma^2 = (VarX)(1 - \rho(X,Y)) \ .$$

7.6 Let $\mathbf{X} = (X_1, X_2)$ be multivariate normal with $\mu = (1, -1)$ and $\mathbf{\Sigma} = \begin{pmatrix} 2 & 2 \\ 2 & 3 \end{pmatrix}$. Let $Y_1 = X_1 - X_2 - 2$ and $Y_2 = X_1 + X_2$.

a) Find the distribution of $Y = (Y_1, Y_2)$.

b) Find the density function $f_Y(y_1, y_2)$.

7.7 Suppose that X is multivariate normal $\mathcal{N}(\mu, \Sigma)$ where $\mu = 1$ and:

$$\Sigma = \begin{pmatrix} 3 & 1 \\ 1 & 2 \end{pmatrix} \ .$$

Find the conditional distribution of $X_1 + X_2$ given $X_1 - X_2 = 0$.

7.8 Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector with normal multivariate distribution of parameters $\mu = \mathbf{0}$ and Σ given by:

$$\Sigma = \begin{pmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_3 \\ \rho_2 & \rho_3 & 1 \end{pmatrix} \ .$$

Find $P(X_1 > 0, X_2 > 0, X_3 > 0)$.

7.9 Let $\mathbf{X} = (X_1, X_2, X_3)$ be a random vector with normal multivariate distribution of parameters $\mu = \mathbf{0}$ and Σ given by:

$$\Sigma = \begin{pmatrix} 5 & 2 & -2 \\ 2 & 6 & 3 \\ -2 & 3 & 8 \end{pmatrix}.$$

Find the density function $f(x_1, x_2, x_3)$ of $\mathbf{X}$.

7.10 The random vector $\mathbf{X}$ has three-dimensional normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 3 & 1 \\ -1 & 1 & 5 \end{pmatrix}.$$

Find the distribution of X_2 given that $X_1 - X_3 = 1$ and $X_2 + X_3 = 0$.

7.11 The random vector $\mathbf{X}$ has three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 4 & 0 \\ -1 & 0 & 7 \end{pmatrix}.$$

Find the distribution of X_3 given that $X_1 = 1$.

7.12 The random vector $\mathbf{X}$ has three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & -1 \\ 1 & 3 & 0 \\ -1 & 0 & 5 \end{pmatrix}.$$

Find the distribution of X_2 given that $X_1 + X_3 = 1$.

7.13 Let $\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \boldsymbol{\Sigma})$, where:

$$\mu = \begin{pmatrix} 2 & 0 & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix}.$$

Determine the conditional distribution of $X_1 - X_3$ given that $X_2 = -1$.

7.14 The random vector $\mathbf{X}$ has three-dimensional normal distribution with mean vector μ and covariance matrix $\boldsymbol{\Sigma}$ given by:

$$\mu = \begin{pmatrix} 1 & 0 & -2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \begin{pmatrix} 3 & -2 & 1 \\ -2 & 4 & -1 \\ 1 & -1 & 2 \end{pmatrix}.$$

Find the conditional distribution of X_1 given that $X_1 = -X_2$.

7.15 The random vector $\mathbf{X}$ has three-dimensional normal distribution with expectation $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$ given by:

$$\boldsymbol{\Sigma} = \begin{pmatrix} 2 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 2 \end{pmatrix} .$$

Find the distribution of X_2 given that $X_1 = X_2 = X_3$.

CHAPTER 8

LIMIT THEOREMS

The ability to draw conclusions about a population from a given sample and determine how reliable those conclusions are plays a crucial role in statistics. On that account it is essential to study the asymptotic behavior of sequences of random variables. This chapter covers some of the most important results within the limit theorems theory, namely, the weak law of large numbers, the strong law of large numbers, and the central limit theorem, the last one being called so as a way to assert its key role among all the limit theorems in probability theory (see Hernandez and Hernandez, 2003).

8.1 THE WEAK LAW OF LARGE NUMBERS

When the distribution of a random variable X is known, it is usually possible to find its expected value and variance. However, the knowledge of these two quantities do not allow us to find probabilities such as $P(|X - c| > \epsilon)$ for $\epsilon > 0$. In this regard, the Russian mathematician Chebyschev proved an inequality, appropriately known as *Chebyschev's inequality* (compare with exercise 2.48 from Chapter 2), which offers a bound for such probabilities.

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

Even though in practice it is seldom used, its theoretical importance is unquestionable, as we will see later on.

Chebyschev's inequality is a particular case of Markov's inequality (compare with exercise 2.47 from Chapter 2), which we present next.

Lemma 8.1 (Markov's Inequality) *If X is a nonnegative random variable whose expected value exists, then, for all $a > 0$, we have:*

$$P(X \geq a) \leq \frac{E(X)}{a}\,. \tag{8.1}$$

Proof: Consider the random variable I defined by:

$$I := \begin{cases} 1 & \text{if } X \geq a \\ 0 & \text{otherwise} \end{cases}.$$

Since $X \geq 0$, then $I \leq \frac{X}{a}$, and taking expectations on both sides of this inequality we arrive at the desired expression. ∎

■ EXAMPLE 8.1

From past experience, a teacher knows that the score obtained by a student in the final exam of his subject is a random variable with mean 75. Find an upper bound to the probability that the student gets the score greater than or equal to 85.

Solution: Let X be the random variable defined as follows:

$$X := \text{"Score obtained by the student in the final exam"}.$$

Since X is nonnegative, Markov's inequality implies that:

$$P(X \geq 85) \leq 0.88235\,. \quad \blacktriangle$$

■ EXAMPLE 8.2

Suppose that X is a random variable having a binomial distribution with parameters 5 and $\frac{1}{3}$. Use Markov's inequality to find an upper bound to $P(X \geq 2)$. Compute $P(X \geq 2)$ exactly and compare the two results.

Solution: We know that $E(X) = \frac{5}{3}$, and therefore, from Markov's inequality:

$$P(X \geq 2) \leq \frac{5}{6}\,.$$

On the other hand:

$$\begin{aligned}P(X \geq 2) &= 1 - P(X = 0) - P(X = 1)\\ &= 1 - \binom{5}{0}\left(\frac{1}{3}\right)^0\left(\frac{2}{3}\right)^5 - \binom{5}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^4\\ &= 0.53909\,.\end{aligned}$$

These results show that Markov's inequality can sometimes give a rather rough estimate of the desired probability. ▲

In many cases there is no specific information about the distribution of the random variable X and it is in those cases where Chebyschev's inequality can offer valuable information about the behavior of the random variable.

Theorem 8.1 (Chebyschev's Inequality) *Let X be a random variable such that $Var(X) < \infty$. Then, for any $\epsilon > 0$, we have:*

$$P(|X - E(X)| \geq \epsilon) \leq \frac{1}{\epsilon^2} Var(X)\,.$$

Proof: Let $Y := |X - E(X)|^2$ and $a = \epsilon^2$. Markov's inequality yields

$$P(|X - E(X)| \geq \epsilon) \leq \frac{E\left(|X - E(X)|^2\right)}{\epsilon^2} = \frac{1}{\epsilon^2} Var(X)\ , \qquad (8.2)$$

which completes the proof. ■

Clearly, (8.2) is equivalent to:

$$P(|X - E(X)| < \epsilon) \geq 1 - \frac{1}{\epsilon^2} Var(X)\,.$$

Likewise, taking $\epsilon := \sigma k$, with $k > 0$ and $\sigma := \sqrt{Var(X)}$, we obtain:

$$P(|X - E(X)| \geq \sigma k) \leq \frac{1}{k^2}\,.$$

If in (8.2) $E(X)$ is replaced by a real number C, we obtain:

$$P(|X - C| \geq \epsilon) \leq \frac{E\left(|X - C|^2\right)}{\epsilon^2}\,.$$

The above expression is also called "Chebyschev's inequality" (see Meyer, 1970).

Note 8.1 *In (8.1) and (8.2) we can replace $P(X \geq a)$ and $P(|X - E(X)| \geq \epsilon)$ with $P(X > a)$ and $P(|X - E(X)| > \epsilon)$, respectively, and the inequalities will still hold.*

EXAMPLE 8.3

Is there any random variable X for which

$$P\big(\mu_X - 2\sigma_X \leq X \leq \mu_X + 2\sigma_X\big) = 0.6 \tag{8.3}$$

where μ_X and σ_X are the expected value and standard deviation of X, respectively?
Solution: It follows from Chebyschev's inequality that:

$$P\left(\mu_X - 2\sigma_X \leq X \leq \mu_X + 2\sigma_X\right) = P\left(|X - \mu_X| \leq 2\sigma_X\right) \geq \frac{3}{4}\,.$$

Hence, there cannot be a random variable X satisfying (8.3). ▲

EXAMPLE 8.4

Show that if $Var\,(X) = 0$, then $P\,(X = E(X)) = 1$.
Solution: Chebyschev's inequality implies that for any $n \geq 1$:

$$P\left(|X - E(X)| > \frac{1}{n}\right) = 0\,.$$

Taking the limit when $n \to \infty$, we get:

$$\begin{aligned} 0 &= \lim_{n\to\infty} P\left(|X - E(X)| > \frac{1}{n}\right) = P\left(\lim_{n\to\infty}\left\{|X - E(X)| > \frac{1}{n}\right\}\right) \\ &= P\left(X \neq E(X)\right)\,. \end{aligned}$$

In other words, $P\,(X = E(X)) = 1$. ▲

The weak law of large numbers can be obtained as an application of Chebyschev's inequality. This is one of the most important results in probability theory and was initially demonstrated by Jacob Bernoulli for a particular case. The weak law of large numbers states that the expected value $E(X)$ of a random variable X can be considered an "idealization", for "large enough" n, of the arithmetic mean $\overline{X} := \frac{X_1+\cdots+X_n}{n}$, where $X_1, \cdots, X_n$ are independent and identically distributed random variables with the same distribution of X.

In order to state this law, we need the following concept:

Definition 8.1 (Identically Distributed Random Variables) $X_1, X_2, \cdots$ *is said to be a sequence of independent and identically distributed random variables if:*

1. *For any $n \in \mathbb{Z}^+$ we have that $X_1, \cdots, X_n$ are independent random variables.*

2. *For any $i, j \in \mathbb{Z}^+$, X_i and X_j have the same distribution.*

Note 8.2 *The previous definition can be generalized to random vectors as follows: It is said that $\mathbf{X}_1, \mathbf{X}_2, \cdots$ is a sequence of independent and identically distributed random vectors if:*

1. *For any $n \in \mathbb{Z}^+$ we have that $\mathbf{X}_1, \cdots, \mathbf{X}_n$ are independent random vectors.*

2. *For any $i, j \in \mathbb{Z}^+$, $\mathbf{X}_i$ and $\mathbf{X}_j$ have the same distribution.*

Theorem 8.2 (The Weak Law of Large Numbers (WLLN)) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with mean μ and finite variance σ^2. Then, for any $\epsilon > 0$, we have that*

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2},$$

from which we obtain that for any $\epsilon > 0$:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0 .$$

Proof: Let $\overline{X_n} := \frac{X_1 + \cdots + X_n}{n}$ be the arithmetic mean of the first n random variables. Clearly $E\left(\overline{X_n}\right) = \mu$ and $Var\left(\overline{X_n}\right) = \frac{\sigma^2}{n}$. Chebyschev's inequality implies that for any $\epsilon > 0$ we must have:

$$P\left(\left|\overline{X_n} - E(\overline{X_n})\right| \geq \epsilon\right) \leq \frac{Var\left(\overline{X_n}\right)}{\epsilon^2} ,$$

which is exactly what we wanted to prove. ∎

Note 8.3 *It is possible to prove the weak law of large numbers without the finite variance hypothesis. That result, due to the Russian mathematician Alexander Khinchin (1894–1959), states the following [see Hernandez (2003) for a proof]:*

If $X_1, X_2, \cdots$ is a sequence of independent and identically distributed random variables with mean μ, then, for any $\epsilon > 0$, we have that:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| > \epsilon\right) = 0 .$$

As a special case of the weak law of large numbers we obtain the following result:

Corollary 8.1 (Bernoulli's Law of Large Numbers) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables having a Bernoulli distribution with parameter p. Then, for any $\epsilon > 0$, we have*

$$P\left(\left|\frac{K_n}{n} - p\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2} \tag{8.4}$$

where $K_n := X_1 + \cdots + X_n$.

Proof: The weak law of large numbers implies that:

$$P\left(\left|\frac{K_n}{n} - p\right| \geq \epsilon\right) \leq \frac{p(1-p)}{n\epsilon^2}.$$

Since $p \in (0,1)$, then $p(1-p) \leq \frac{1}{4}$ and consequently (8.4) holds. ∎

Note 8.4 *Bernoulli's law states that when a random experiment with only two possible results, success or failure, is carried out for a number of times large enough, then, for any* $\epsilon > 0$, *the set of results for which the proportion of successes is at a distance greater than* ϵ *of the probability of success* p *tends to zero. Observe that Bernoulli's law does not assert that the relative frequency of successes converges to the success probability as the number of repetitions increases. Even when the latter is true, it cannot be directly inferred from Bernoulli's law.*

Note 8.5 *Bernoulli's law was proved by means of the weak law of large numbers, which in turn rests upon Chebyschev's inequality. However, the original proof given by Bernoulli states that, for arbitrary* $\epsilon > 0$ *and* $0 < \delta < 1$, *we have*

$$P\left(\left|\frac{K_n}{n} - p\right| < \epsilon\right) > 1 - \delta$$

where $n \geq \left[\frac{(1+\epsilon)}{\epsilon^2}\right] \ln\left(\frac{1}{\delta}\right) + \frac{1}{\epsilon}$. *For example, if* $\epsilon = 0.03$ *and* $\delta = 0.00002$, *then* $n \geq 12416$ *implies:*

$$P\left(\left|\frac{K_n}{n} - p\right| < 0.03\right) > 0.99998 .$$

That is, if the experiment is repeated at least 12,416 *times, we can be at least* 99.998% *sure that the success ratio will be less than* 3% *away from the success probability* p.

According to Hernandez (2003), Italian mathematician Francesco Paolo Cantelli (1875–1966) proved an even stronger result, namely, that if $n \geq \left(\frac{2}{\epsilon^2}\right) \ln\left(\frac{4}{\delta\epsilon^2}\right) + 2 := N$, *then*

$$P\left(\bigcap_{n=N}^{\infty}\left(\left|\frac{K_n}{n} - p\right| < \epsilon\right)\right) > 1 - \delta$$

and thus, for $\epsilon = 0.03$ *and* $\delta = 0.00002$, $n \geq 42711$ *implies that:*

$$P\left(\bigcap_{n=N}^{\infty}\left(\left|\frac{K_n}{n} - p\right| < 0.03\right)\right) > 0.99998 .$$

8.2 CONVERGENCE OF SEQUENCES OF RANDOM VARIABLES

Let $X, X_1, X_2, \cdots$ be real-valued random variables all defined over the same probability space. In this section three important and common modes of convergence of the sequence $(X_n)_n$ to X will be defined. It is important, however, to clarify that there exist other modes of convergence different from those studied here (in this regard, see, Bauer, 1991).

Definition 8.2 (Convergence in Probability) *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. It is said that $(X_n)_n$ converges in probability to X and write*

$$X_n \xrightarrow[n\to\infty]{P} X$$

if, for any $\epsilon > 0$, the following condition is met:

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0 \ .$$

Note 8.6 *Making use of the last definition, we can express the weak law of large numbers as follows: If $X_1, X_2, \cdots$ is a sequence of independent and identically distributed random variables with mean μ and finite variance σ^2, then*

$$\overline{X_n} \xrightarrow[n\to\infty]{P} \mu \ ,$$

where $\overline{X_n}$ represents the arithmetic mean of the first n random variables.

■ EXAMPLE 8.5

Let $(X_n)_{n\in\mathbb{Z}^+}$ be a sequence of random variables such that $P(X_n = 0) = 1 - \frac{1}{n}$ and $P(X_n = n) = \frac{1}{n}$, for $n = 1, 2, \cdots$. Let $\epsilon > 0$. Then:

$$P(|X_n| > \epsilon) = \begin{cases} \frac{1}{n} & \text{if } \epsilon < n \\ 0 & \text{if } \epsilon \geq n \end{cases} .$$

Therefore:

$$\lim_{n\to\infty} P(|X_n| > \epsilon) = 0 \ .$$

Thus:

$$X_n \xrightarrow[n\to\infty]{P} 0 \ . \quad \blacktriangle$$

■ **EXAMPLE 8.6**

A fair dice is thrown once. For each $n = 1, 2, \cdots$, we define the random variable X_n by

$$X_n := \begin{cases} 1 & \text{if the result obtained is tail} \\ 0 & \text{if the result obtained is head} \end{cases}$$

and let X be the random variable given by:

$$X := \begin{cases} 1 & \text{if the result obtained is head} \\ 0 & \text{if the result obtained is tail .} \end{cases}$$

Clearly, for all $n = 1, 2, \cdots$, we have:

$$|X_n - X| = 1 \ .$$

Hence

$$P\left(|X_n - X| > \frac{1}{2}\right) = 1$$

and therefore:

$$X_n \underset{n\to\infty}{\overset{P}{\not\longrightarrow}} X \ . \qquad \blacktriangle$$

A useful result to establish the convergence of a sequence of random variables is the following theorem, whose proof is omitted. We encourage the interested reader to see Jacod and Protter (2004) for a proof.

Theorem 8.3 *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Then $X_n \xrightarrow[n\to\infty]{P} X$ if and only if the following expression holds:*

$$\lim_{n\to\infty} E\left(\frac{|X_n - X|}{1 + |X_n - X|}\right) = 0 \ .$$

Definition 8.3 (Convergence in L^r) *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Let r be a positive integer such that $E(|X|^r) < \infty$ and $E(|X_n|^r) < \infty$ $\forall n$. Then $X_n \xrightarrow[n\to\infty]{L^r} X$ if and only if the following expression holds:*

$$\lim_{n\to\infty} E(|X_n - X|^r) = 0 \ .$$

■ EXAMPLE 8.7

Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with mean μ and finite variance σ^2. Let

$$\overline{X} := \frac{X_1 + \cdots + X_n}{n} = \frac{S_n}{n}$$

where $S_n = X_1 + \cdots + X_n$. Now:

$$\begin{aligned} E\left(\left|\frac{S_n}{n} - \mu\right|^2\right) &= E\left(\frac{(S_n - n\mu)^2}{n^2}\right) \\ &= \frac{1}{n^2} Var(S_n) = \frac{1}{n^2} n\sigma^2 \xrightarrow[n\to\infty]{} 0 \,. \end{aligned}$$

Hence:

$$\overline{X} \xrightarrow[n\to\infty]{L^2} \mu. \qquad \blacktriangle$$

■ EXAMPLE 8.8

Consider Example 8.5. Since $P(X_n = 0) = 1 - \frac{1}{n}$ and $P(X_n = n) = \frac{1}{n}$, for $n = 1, 2, \cdots$, we get $E(X_n) = 1$. Hence, $X_n \overset{L^1}{\underset{n\to\infty}{\not\longrightarrow}} 0$ whereas $X_n \xrightarrow[n\to\infty]{P} 0$. ▲

Theorem 8.4 *Let $r > 0$, $X, X_1, X_2, \ldots$ be a sequence of random variables such that $X_n \xrightarrow[n\to\infty]{L^r} X$. This implies $X_n \xrightarrow[n\to\infty]{P} X$.*

Proof: Chebyschev's inequality implies that for any $\epsilon > 0$ we must have:

$$P(|X_n - X| \geq \epsilon) \leq \frac{E(|X_n - X|^r)}{\epsilon^r} \xrightarrow[n\to\infty]{} 0 \,.$$

Hence, the proof. ■

■ EXAMPLE 8.9

Let $(\Omega, \Im, P) = ([0,1], \mathcal{B}([0,1]), \lambda([0,1]))$ where $\mathcal{B}([0,1])$ is the Borel σ-algebra on $[0,1]$ and $\lambda([0,1])$ is the Lebesgue measure on $[0,1]$. Let

$$X_n(w) = \begin{cases} 2^n & \text{if } 0 < w < 1/n \\ 0 & \text{otherwise}\,. \end{cases}$$

Then for all $\epsilon > 0$:

$$P(|X_n| \geq \epsilon) = P((0, 1/n)) = \frac{1}{n} \xrightarrow[n\to\infty]{} 0 .$$

That is, $X_n \xrightarrow[n\to\infty]{P} 0$. But:

$$E(|X_n|^r) = 2^{nr}\frac{1}{n} \xrightarrow[n\to\infty]{} \infty . \qquad \blacktriangle$$

Another mode of convergence that will be covered in this section is almost sure convergence, a concept defined below.

Definition 8.4 (Almost Sure Convergence) *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. We say that $(X_n)_n$ converges almost surely (or with probability 1) to X and write*

$$X_n \xrightarrow[n\to\infty]{a.s.} X$$

if the following condition is met:

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1 .$$

In other words, if $A := \{\omega \in \Omega : X_n(\omega) \longrightarrow X(\omega)\}$, then:

$$X_n \xrightarrow[n\to\infty]{a.s.} X \quad \textit{if and only if} \quad P(A) = 1 .$$

■ EXAMPLE 8.10

Let $(\Omega, \Im, P)$ be an arbitrary probability space and let $(X_n)_n$ be the sequence random variables defined over $(\Omega, \Im, P)$ as follows:

$$\begin{array}{rccc} X_n : & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & \frac{1}{n} + 1 \end{array}$$

Clearly, for all $\omega \in \Omega$, we have:

$$\lim_{n\to\infty} X_n(\omega) = 1.$$

Therefore,

$$P(\{\omega \in \Omega : X_n(\omega) \longrightarrow 1\}) = P(\Omega) = 1 ,$$

which implies:

$$X_n \xrightarrow[n\to\infty]{a.s.} 1 . \qquad \blacktriangle$$

8.3 THE STRONG LAW OF LARGE NUMBERS

The following result, known as *the strong law of large numbers*, states that the average of a sequence of independent and identically distributed random variables converges, with probability 1, to the mean of the distribution.

The proof of this law requires the following lemma [see Jacod and Protter (2004) for a demonstration]:

Lemma 8.2 *Let $X_1, X_2, \cdots$ be a sequence of random variables. We have that:*

1. *If all the random variables X_i are positive, then*

$$E\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} E\left(X_n\right) \tag{8.5}$$

where the expressions in (8.5) are either both finite or both infinite.

2. *If $\sum_{n=1}^{\infty} E\left(|X_n|\right) < \infty$, then $\sum_{n=1}^{\infty} X_n$ converges almost surely and (8.5) holds.*

Theorem 8.5 (The Strong Law of Large Numbers (SLLN)) *Let $X_1, X_2, \cdots$ be a sequence of independent and identically distributed random variables with finite mean μ and finite variance σ^2. Then:*

$$\overline{X_n} = \frac{X_1 + X_2 + \cdots + X_n}{n} \xrightarrow[n\to\infty]{a.s.} \mu .$$

Proof: (Jacod and Protter, 2004) Without loss of generality, we can assume that $\mu = 0$.

Since $E\left(\overline{X_n}\right) = 0$,

$$E\left(\left[\overline{X_n}\right]^2\right) = \frac{1}{n^2} \sum_{1 \le j,k \le n} E\left(X_j X_k\right) \tag{8.6}$$

and $E\left(X_j X_k\right) = 0$ for $j \neq k$, then (9.7) equals:

$$\begin{aligned} E\left(\left[\overline{X_n}\right]^2\right) &= \frac{1}{n^2} \sum_{j=1}^{n} E(X_j^2) \\ &= \frac{\sigma^2}{n} . \end{aligned}$$

Therefore:

$$\lim_{n\to\infty} E\left(\left[\overline{X_n}\right]^2\right) = 0 .$$

Moreover, $E\left(\left[\,\overline{X_n}\,\right]^2\right) = \frac{\sigma^2}{n}$ implies that:

$$\sum_{n=1}^{\infty} E\left(\left[\,\overline{X_{n^2}}\,\right]^2\right) = \sum_{n=1}^{\infty} \frac{\sigma^2}{n^2} < \infty\,.$$

From Lemma 8.2, we have

$$\sum_{n=1}^{\infty} \left[\,\overline{X_{n^2}}\,\right]^2 < \infty \quad \text{with probability } 1$$

and consequently $\lim_{n\to\infty} \overline{X_{n^2}} = 0$ with probability 1.

Let $n \in \mathbb{N}$ and k_n be an integer such that:

$$[k_n]^2 \le n < [k_n + 1]^2\ .$$

Then

$$\overline{X_n} - \frac{[k_n]^2}{n}\overline{X_{k_n^2}} = \frac{1}{n} \sum_{j=k_n^2+1}^{n} X_j\ ,$$

so that:

$$\begin{aligned}
E\left(\left[\overline{X_n} - \frac{[k_n]^2}{n}\overline{X_{k_n^2}}\right]^2\right) &= \frac{n - k_n^2}{n^2}\sigma^2 \\
&\le \frac{2k_n + 1}{n^2}\sigma^2 \\
&\le \frac{2\sqrt{n} + 1}{n^2}\sigma^2 \\
&\le \frac{3\sigma^2}{\sqrt[2]{n^3}}\ .
\end{aligned}$$

Accordingly,

$$\sum_{n=1}^{\infty} E\left(\left[\overline{X_n} - \frac{[k_n]^2}{n}\overline{X_{k_n^2}}\right]^2\right) \le \sum_{n=1}^{\infty} \frac{3\sigma^2}{\sqrt[2]{n^3}} < \infty$$

and another application of Lemma 8.2 yields

$$\sum_{n=1}^{\infty} \left[\overline{X_n} - \frac{[k_n]^2}{n}\overline{X_{k_n^2}}\right]^2 < \infty \quad \text{with probability } 1$$

from which:

$$\lim_{n\to\infty} \left[\overline{X_n} - \frac{[k_n]^2}{n}\overline{X_{k_n^2}}\right] = 0 \quad \text{with probability } 1\ .$$

Since $\lim_{n\to\infty} \overline{X_{k_n^2}} = 0$ and $\lim_{n\to\infty} \frac{[k_n]^2}{n} = 1$, then, with probability 1, we conclude that $\lim_{n\to\infty} \overline{X_n} = 0$, which completes the proof for the case $\mu = 0$. If $\mu \neq 0$, it suffices to consider the random variables defined, for $i = 1, 2, \cdots$, by

$$Z_i := X_i - \mu$$

and apply the result to the sequence thus obtained. ■

The last mode of convergence presented in this text, known as *convergence in distribution*, is most widely used in applications.

Definition 8.5 (Convergence in Distribution) *Let $X, X_1, X_2, \cdots$ be real-valued random variables with distribution functions $F, F_1, F_2, \cdots$, respectively. It is said that $(X_n)_n$ converges in distribution to X, written as*

$$X_n \xrightarrow[n\to\infty]{d} X \ ,$$

if the following condition holds:

$$\lim_{n\to\infty} F_n(x) = F(x) \qquad \textit{for any point } x \textit{ where } F \textit{ is continuous}$$

or, equivalently,

$$X_n \xrightarrow[n\to\infty]{d} X \quad \textit{if and only if} \quad \lim_{n\to\infty} P(X_n \le x) = P(X \le x)$$

for any point x where F is continuous.

Note that convergence in distribution, just like the other modes of convergence previously discussed, reduces itself to the convergence of a sequence of real numbers and not to the convergence of a sequence of events.

■ EXAMPLE 8.11

Let $(\Omega, \Im, P)$ be a probability space, $(X_n)_n$ be the sequence of random variables over $(\Omega, \Im, P)$ defined as

$$\begin{array}{rccc} X_n : & \Omega & \longrightarrow & \mathbb{R} \\ & \omega & \longmapsto & \frac{1}{n} \end{array}$$

and $X = 0$. Assume:

$$\begin{aligned} F_n(x) &= \begin{cases} 1 & \text{if } x \ge \frac{1}{n} \\ 0 & \text{otherwise} \end{cases} \\ F(x) &= \begin{cases} 1 & \text{if } x \ge 0 \\ 0 & \text{otherwise} . \end{cases} \end{aligned}$$

We have for each $x \neq 0$:

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

In other words:

$$X_n \xrightarrow[n\to\infty]{d} 0. \quad \blacktriangle$$

Note 8.7 *If $X, X_1, X_2, \cdots$ are random variables with values in $\mathbb{N}$, then:*

$$X_n \xrightarrow[n\to\infty]{d} X \text{ if and only if } \lim_{n\to\infty} P(X_n = k) = P(X = k) \text{ for all } k \in \mathbb{N}.$$

The following theorem relates convergence in distribution with convergence of the characteristic functions. Its proof is beyond the scope of this text and the reader may refer to Hernandez (2003) or Rao (1973).

Theorem 8.6 (Lévy and Cramer Continuity Theorem) *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$ and having the characteristic functions $\phi, \phi_1, \phi_2, \cdots$ respectively. Then:*

$$X_n \xrightarrow[n\to\infty]{d} X \text{ if and only if } \lim_{n\to\infty} \phi_n(t) = \phi(t) \text{ for all } t \in \mathbb{R}.$$

We proceed now to establish the major links between the convergence modes defined so far, namely, we will see that:

$$X_n \xrightarrow[n\to\infty]{a.s.} X \implies X_n \xrightarrow[n\to\infty]{P} X \implies X_n \xrightarrow[n\to\infty]{d} X.$$

Theorem 8.7 *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Then:*

$$X_n \xrightarrow[n\to\infty]{a.s.} X \implies X_n \xrightarrow[n\to\infty]{P} X.$$

Proof: Let $\epsilon > 0$. The sets

$$A_k := \{\omega \in \Omega : |X_n(\omega) - X(\omega)| < \epsilon \text{ for all } n \geq k\}$$

form an increasing sequence of events whose union

$$A_\infty := \bigcup_{k=1}^{\infty} A_k$$

contains the set

$$A := \left\{\omega \in \Omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\}.$$

By hypothesis $P(A) = 1$, from which we infer that $P(A_\infty) = 1$. On the other hand, it follows from the continuity of P that:

$$\lim_{k\to\infty} P(A_k) = P(A_\infty) = 1 \ .$$

Since

$$0 \le P(|X_k - X| \ge \epsilon) \le P(A_k^c)$$

and

$$\lim_{k\to\infty} P(A_k^c) = 0$$

we conclude that:

$$X_n \xrightarrow[n\to\infty]{P} X \ .$$

■

Theorem 8.8 *Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$ with distribution functions $F, F_1, F_2, \cdots$ respectively. Then:*

$$X_n \xrightarrow[n\to\infty]{P} X \implies X_n \xrightarrow[n\to\infty]{d} X \ .$$

Proof: Let x be a point of continuity of F. We need to prove that $\lim_{n\to\infty} F_n(x) = F(x)$. Since F is continuous at x, then, for any $\epsilon > 0$, there exists a $\delta > 0$ such that:

$$F(x+\delta) - F(x) < \frac{\epsilon}{2}$$

$$F(x) - F(x-\delta) < \frac{\epsilon}{2} \ .$$

On the other hand, since $X_n \xrightarrow[n\to\infty]{P} X$, then there exists $n(\epsilon) \in \mathbb{N}$ such that, for $n \ge n(\epsilon)$, the following condition is satisfied:

$$P(|X_n - X| > \delta) < \frac{\epsilon}{2} \ .$$

Furthermore, we have that:

$$\begin{aligned}(X_n \le x) &= [(X_n \le x) \cap (|X_n - X| \le \delta)] \cup [(X_n \le x) \cap (|X_n - X| > \delta)] \\ &\subseteq (X \le x + \delta) \cup (|X_n - X| > \delta) \ .\end{aligned}$$

Analogously:

$$(X \le x - \delta) \subseteq (X_n \le x) \cup (|X_n - X| > \delta) \ .$$

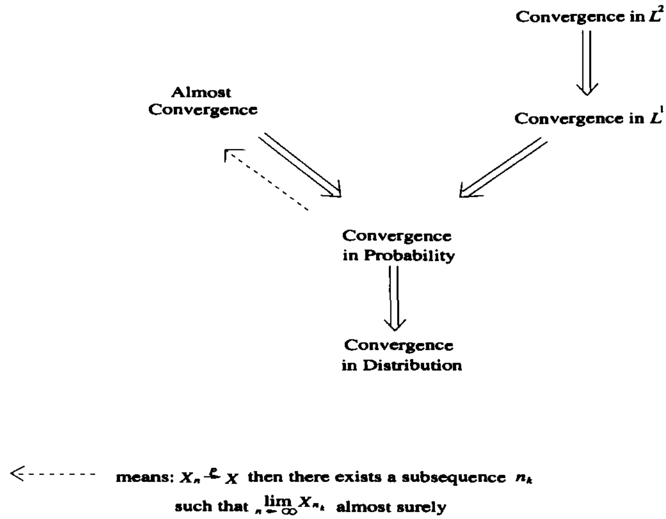


Figure 8.1 Relationship between the different modes of convergence

Thus:

$$\begin{aligned}
F(x) - \epsilon &< F(x-\delta) - \frac{\epsilon}{2} \\
&\le P(X_n \le x) + P(|X_n - X| > \delta) - \frac{\epsilon}{2} \\
&\le P(X_n \le x) \\
&\le F(x+\delta) + \frac{\epsilon}{2} \\
&< F(x) + \epsilon \; .
\end{aligned}$$

That is, for all $n \ge n(\epsilon)$, we have that

$$|F_n(x) - F(x)| < \epsilon$$

and consequently:

$$\lim_{n\to\infty} F_n(x) = F(x) \; .$$

■

Note 8.8 *The converse of the above theorem is not, in general, true. Consider, for example, the random variables given in Example 8.6. In that case, we have* $X_n \xrightarrow[n\to\infty]{d} X$, *but* $X_n \not\xrightarrow[n\to\infty]{P} X$.

Figure 8.1 summarizes the relationship between the different modes of convergence. In this figure, the dotted arrow means that the convergence in probability does not imply the convergence in almost surely. However, we can *prove that there is a subsequence of the original sequence that converges almost surely.*

All the definitions given so far can be generalized to random vectors in the following way:

Definition 8.6 (Convergence of Random Vectors) *Let* $\mathbf{X}, \mathbf{X}_1, \mathbf{X}_2, \cdots$ *be* k*-dimensional random vectors,* $k \in \mathbb{N}$*, having the distribution functions* $F_{\mathbf{X}}, F_{\mathbf{X}_1}, F_{\mathbf{X}_2}, \cdots$*, respectively. Then, it is said that:*

1. $(\mathbf{X}_n)_n$ *converges in probability to* $\mathbf{X}$*, written as* $\mathbf{X}_n \xrightarrow[n\to\infty]{P} \mathbf{X}$*, if the ith component of* $\mathbf{X}_n$ *converges in probability to the ith component of* $\mathbf{X}$ *for each* $i = 1, 2, \cdots, k$*.*

2. $(\mathbf{X}_n)_n$ *converges almost surely (or with probability* 1*) to* $\mathbf{X}$*, written as* $\mathbf{X}_n \xrightarrow[n\to\infty]{a.s.} \mathbf{X}$*, if the ith component of* $\mathbf{X}_n$ *converges almost surely to the ith component of* $\mathbf{X}$ *for each* $i = 1, 2, \cdots, k$*.*

3. $(\mathbf{X}_n)_n$ *converges in distribution to* $\mathbf{X}$*, written as* $\mathbf{X}_n \xrightarrow[n\to\infty]{d} \mathbf{X}$*, if* $\lim_{n\to\infty} F_{\mathbf{X}_n}(\mathbf{x}) = F_{\mathbf{X}}(\mathbf{x})$ *for all points* $\mathbf{x} = (x_1, \cdots, x_k) \in \mathbb{R}^k$ *of continuity of* $F_{\mathbf{X}}(\cdot)$*.*

8.4 CENTRAL LIMIT THEOREM

This section is devoted to one of the most important results in probability theory, the central limit theorem, whose earliest version was first proved in 1733 by the French mathematician Abraham DeMoivre. In 1812 Pierre Simon Laplace, also a French mathematician, proved a more general version of the theorem. The version known nowadays was proved in 1901 by the Russian mathematician Liapounoff.

The central limit theorem states that the sum of independent and identically distributed random variables has, approximately, a normal distribution whenever the number of random variables is large enough and their variance is finite and different from zero.

Theorem 8.9 (Univariate Central Limit Theorem (CLT)) *Let* $X_1, X_2, \cdots$ *be a sequence of independent and identically distributed random variables with mean* μ *and finite variance* σ^2*. Let* $S_n := \sum_{j=1}^{n} X_j$ *and* $Y_n := \frac{S_n - n\mu}{\sigma\sqrt{n}}$*. Then, the sequence of random variables* $Y_1, Y_2, \cdots$ *converges in distribution to a random variable* Y *having a standard normal distribution. In other words:*

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x) \text{ for all } x \in \mathbb{R} .$$

Proof: Without loss of generality, it will be assumed that $\mu = 0$. Let ϕ be the characteristic function of the random variables $X_1, X_2, \cdots$. Since the random variables are independent and identically distributed, we have that:

$$
\begin{aligned}
\phi_{Y_n}(t) &= E\left(\exp\left[itY_n\right]\right) \\
&= E\left(\exp\left[it\frac{X_1 + \cdots + X_n}{\sigma\sqrt{n}}\right]\right) \\
&= E\left(\prod_{j=1}^{n} \exp\left[it\frac{X_j}{\sigma\sqrt{n}}\right]\right) \\
&= \prod_{j=1}^{n} E\left(\exp\left[it\frac{X_j}{\sigma\sqrt{n}}\right]\right) \\
&= \prod_{j=1}^{n} \phi\left(\frac{t}{\sigma\sqrt{n}}\right) \\
&= \left[\phi\left(\frac{t}{\sigma\sqrt{n}}\right)\right]^n .
\end{aligned}
$$

Expansion of ϕ in a Taylor series about zero yields:

$$
\begin{aligned}
\phi(t) &= E(e^{itX}) \\
&= E\left(1 + itX - \frac{t^2X^2}{2} + \frac{it^3X^3}{6} - \cdots\right) \\
&= 1 + 0 - \frac{\sigma^2 t^2}{2} + t^2 o(t) .
\end{aligned}
$$

Thus:

$$
\begin{aligned}
\phi_{Y_n}(t) &= \left[1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 o(t)\right]^n \\
&= \exp\left[n \ln\left(1 - \frac{\sigma^2}{2}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \left(\frac{t}{\sigma\sqrt{n}}\right)^2 o(t)\right)\right] .
\end{aligned}
$$

Taking the limit when $n \to \infty$ we get:

$$
\lim_{n\to\infty} \phi_{Y_n}(t) = \exp\left[-\frac{t^2}{2}\right] .
$$

The Lévy-Cramer continuity theorem implies that

$$
Y_n \xrightarrow[n\to\infty]{d} Y ,
$$

where Y is a random variable having a standard normal distribution. ∎

Note 8.9 *The de Moivre-Laplace theorem (4.6) is a particular case of the central limit theorem. Indeed, if* $X_1, X_2, \cdots$ *are independent and identically distributed random variables having a Bernoulli distribution of parameter* p*, then the conditions given in (8.9) are satisfied.*

Note 8.10 *Let* $X_1, X_2, \cdots$ *be a sequence of independent and identically distributed random variables with mean* μ *and finite positive variance* σ^2*. The central limit theorem asserts that, for large enough* n*, the random variable* $S_n := \sum_{j=1}^{n} X_j$ *has, approximately, a normal distribution with parameters* $n\mu$ *and* $n\sigma^2$.

Note 8.11 *The central limit theorem can be applied to most of the classic distributions, namely: binomial, Poisson, negative binomial, gamma, Weibull, etc., since all of them satisfy the hypotheses required by the theorem. However, it cannot be applied to Cauchy distribution since it does not meet the requirements of the central limit theorem.*

Note 8.12 *There are many generalizations of the central limit theorem, among which the Lindenberg-Feller theorem stands out in that it allows for the sequence* $X_1, X_2, \cdots$ *of independent random variables with different means* $E(X_i) = \mu_i$ *and different variances* $Var(X_i) = \sigma_i^2$.

Note 8.13 (Multivariate Central Limit Theorem) *The central limit theorem can be generalized to sequences of random variables as follows:*

Let $\mathbf{X}_1, \mathbf{X}_2, \cdots$ *be a sequence of* k*-dimensional independent and identically distributed random vectors, with* $k \in \mathbb{N}$*, having mean vector* μ *and variance-covariance matrix* Σ*, where* Σ *is positive definite. Let*

$$\overline{\mathbf{X}_n} := \frac{\mathbf{X}_1 + \mathbf{X}_2 + \cdots + \mathbf{X}_n}{n}$$

be the vector of arithmetic means. Then

$$\sqrt{n}\left(\overline{\mathbf{X}_n} - \mu\right) \xrightarrow[n \longrightarrow \infty]{d} \mathbf{X} \ ,$$

where $\mathbf{X}$ *is a* k*-dimensional random vector having a multivariate normal distribution with mean vector* $\mathbf{0}$ *and variance-covariance matrix* Σ*. (See Hernandez, 2003).*

Some applications of the central limit theorem are given in the examples below.

■ EXAMPLE 8.12

A fair dice is tossed 1000 times. Find the probability that the number 4 appears at least 150 times.

Solution: Let $X :=$ "Number of times that the number 4 is obtained". We know that $X \stackrel{d}{=} \mathcal{B}\left(1000, \frac{1}{6}\right)$. Applying the de Moivre-Laplace theorem, we can approximate this distribution with a normal distribution of mean $\frac{1000}{6}$ and variance $\frac{5000}{36}$. Therefore:

$$\begin{aligned}
P(X \geq 150) &= P\left(\frac{X - \frac{500}{3}}{\frac{25\sqrt{2}}{3}} \geq \frac{150 - \frac{500}{3}}{\frac{25\sqrt{2}}{3}}\right) \\
&\approx 1 - \Phi(-1.4142) \\
&\approx 1 - 0.07865 \\
&\approx 0.9213 . \quad \blacktriangle
\end{aligned}$$

■ EXAMPLE 8.13

Suppose that for any student the time required for a teacher to grade the final exams of his probability course is a random variable with mean 1 hour and standard deviation 0.4 hours. If there are 100 students in the course, what is the probability that he needs more than 110 hours to grade the exam?

Solution: Let X_i be the random variable representing the time required by the teacher to grade the ith student of his course. We need to calculate the probability that $T := \sum_{i=1}^{100} X_i$ greater than 110. By the central limit theorem, we have:

$$\begin{aligned}
P(T > 110) &= P\left(\frac{T - 100}{\sqrt{100 \times 0.4}} > \frac{110 - 100}{\sqrt{100 \times 0.4}}\right) \\
&\approx 1 - \Phi(1.5811) \\
&\approx 1 - 0.9429 \\
&\approx 0.0571 .
\end{aligned}$$

■ EXAMPLE 8.14

(Wackerly et al., 2008) Many raw materials, like iron ore, coal and unrefined sugar, are sampled to determine their quality by taking small periodic samples of the material as it moves over a conveyor belt. Later on, the small samples are gathered and mixed into a compound sample which is then submitted for analysis. Let Y_i be the volume of the ith small sample in a particular lot, and assume that $Y_1, Y_2, \cdots, Y_n$ is a random sample where each Y_i has mean μ (in cubic inches) and variance

σ^2. The average volume of the samples can be adjusted by changing the settings of the sampling equipment. Suppose that the variance σ^2 of the volume of the samples is approximately 4. The analysis of the compound sample requires that, with a 0.95 probability, $n = 50$ small samples give rise to a compound sample of 200 cubic inches. Determine how the average sample volume μ must be set in order to meet this last requirement.

Solution: The random variables $Y_1, Y_2, \cdots, Y_{50}$ are independent and identically distributed having mean μ and variance $\sigma^2 = 4$. By the central limit theorem, $\sum_{i=1}^{50} Y_i$ has, approximately, a normal distribution with mean 50μ and variance 200. We wish to find μ such that $P\left(\sum_{i=1}^{50} Y_i > 200\right) = 0.95$. Accordingly:

$$\begin{aligned} 0.95 &= P\left(\sum_{i=1}^{50} Y_i > 200\right) \\ &= P\left(\frac{\sum_{i=1}^{50} Y_i - 50\mu}{10\sqrt{2}} > \frac{200 - 50\mu}{10\sqrt{2}}\right) \\ &\approx 1 - \Phi\left(\frac{200 - 50\mu}{10\sqrt{2}}\right). \end{aligned}$$

That is,

$$\Phi\left(\frac{200 - 50\mu}{10\sqrt{2}}\right) \approx 0.05 ,$$

which in turn implies

$$\frac{200 - 50\mu}{10\sqrt{2}} \approx -1.64 ,$$

from which we obtain:

$$\mu \approx 4.46 . \quad \blacktriangle$$

EXERCISES

8.1 Suppose that 50 electronic components say $D_1, D_2, \cdots, D_{50}$, are used in the following manner. As soon as D_1 fails, D_2 becomes operative. When D_2 fails, D_3 becomes operative, etc. Assume that the time to failure of D_i is an exponentially distributed random variable with parameter 0.05 per hour. Let T be the total time of operation of the 50 devices. Compute the approximate probability that T exceeds 500 hours using the central limit theorem?

8.2 Let $X_1, X_2, \cdots, X_n$ be n independent Poisson distributed random variables with means $1, 2, \cdots, n$, respectively. Find an x in terms of t such that

$$P\left(\frac{S_n - \frac{n^2}{2}}{n} \leq t\right) \approx \Phi(x) \quad \text{for sufficiently large} \quad n$$

where Φ is the *cdf* of $\mathcal{N}(0,1)$.

8.3 Suppose that $X_i, i = 1, 2, \cdots, 450$, are independent random variables each having a distribution $\mathcal{N}(0,1)$. Evaluate $P(X_1^2 + X_2^2 + \cdots + X_{450}^2 > 495)$ approximately.

8.4 Let Y be a gamma distributed random variable with *pdf*

$$f(y) = \begin{cases} \frac{1}{\Gamma(p)} e^{-y} y^{p-1} & \text{if } y > 0 \\ 0 & \text{otherwise}. \end{cases}$$

Find the limiting distribution of $\frac{Y - E(Y)}{\sqrt{Var(Y)}}$ as $p \to \infty$.

8.5 Let $X \sim \mathcal{B}(n,p)$. Use the CLT to find n such that:

$$P[X > n/2] \geq 1 - \alpha.$$

Calculate the value of n when $\alpha = 0.90$ and $p = 0.45$.

8.6 A fair coin is flipped 100 consecutive times. Let X be the number of heads obtained. Use Chebyschev's inequality to find a lower bound for the probability that $\frac{X}{100}$ differs from $\frac{1}{2}$ by less than 0.1.

8.7 Let $X_1, X_2, \cdots, X_{50}$ be independent and identically distributed random variables having a Poisson distribution with mean 2.5. Compute:

$$P\left(\sum_{i=1}^{50} X_i > 25\right).$$

8.8 **(Chernoff Bounds)** Suppose that the moment generating function $m_X(t)$ of a random variable X exists. Use Markov's inequality to show that for any $a \in \mathbb{R}$ the following holds:

a) $P(X \geq a) \leq \exp(-ta)\, m_X(t)$ for all $t > 0$.

b) $P(X \leq a) \leq \exp(-ta)\, m_X(t)$ for all $t < 0$.

8.9 **(Jensen Inequality)** Let f be a twice-differentiable convex real-valued function, that is, $f''(x) \geq 0$ for all x. Prove that if X is a real-valued random variable, then

$$E(f(X)) \geq f(E(X))$$

subject to the existence of the expected values.
Hint: Consider the Taylor polynomial of f about $\mu = E(X)$.

8.10 If X is a nonnegative random variable with mean 2, what can be said about $E(X^3)$ and $E(\ln X)$.

8.11 Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Show the following properties about convergence in probability:

a) $X_n \xrightarrow[n\to\infty]{P} X$ if and only if $(X_n - X) \xrightarrow[n\to\infty]{P} 0$.

b) If $X_n \xrightarrow[n\to\infty]{P} X$ and $X_n \xrightarrow[n\to\infty]{P} Y$, then $P(X = Y) = 1$.

c) If $X_n \xrightarrow[n\to\infty]{P} X$, then $(X_n - X_m) \xrightarrow[n,m\to\infty]{P} 0$.

d) If $X_n \xrightarrow[n\to\infty]{P} X$ and $Y_n \xrightarrow[n\to\infty]{P} Y$, then $(X_n + Y_n) \xrightarrow[n\to\infty]{P} (X + Y)$.

e) If $X_n \xrightarrow[n\to\infty]{P} X$ and k is a real constant, then $kX_n \xrightarrow[n\to\infty]{P} kX$.

f) If $X_n \xrightarrow[n\to\infty]{P} X$ and $Y_n \xrightarrow[n\to\infty]{P} Y$, then $X_nY_n \xrightarrow[n\to\infty]{P} XY$.

8.12 Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Prove that:

a) If $X_n \xrightarrow[n\to\infty]{a.s.} X$ and $Y_n \xrightarrow[n\to\infty]{a.s.} Y$, then $(X_n + Y_n) \xrightarrow[n\to\infty]{a.s.} (X + Y)$.

b) If $X_n \xrightarrow[n\to\infty]{a.s.} X$ and $Y_n \xrightarrow[n\to\infty]{a.s.} Y$, then $X_nY_n \xrightarrow[n\to\infty]{a.s.} XY$.

8.13 Let f be a continuous real-valued function and let $X, X_1, X_2, \cdots$ be real-valued random variables over the same probability space $(\Omega, \Im, P)$. Show that if $X_n \xrightarrow[n\to\infty]{a.s.} X$, then $f(X_n) \xrightarrow[n\to\infty]{a.s.} f(X)$.
Note: *The result still holds if almost sure convergence is replaced by convergence in probability.*

8.14 Let $X, X_1, X_2, \cdots$ be real-valued random variables defined over the same probability space $(\Omega, \Im, P)$. Show that if $X_n \xrightarrow[n\to\infty]{d} k$, where k is a real constant, then $X_n \xrightarrow[n\to\infty]{P} k$.

8.15 Suppose that a Bernoulli sequence has length equal to 100 with a success probability of $p = 0.7$. Let X :="Number of successes obtained". Compute $P(X \in [65, 80])$.

8.16 A certain medicine has been shown to produce allergic reactions in 1% of the population. If this medicine is dispensed to 500 people, what is the probability that at most 10 of them show any allergy symptoms?

8.17 A fair dice is rolled as many times as needed until the sum of all the results obtained is greater than 200. What is the probability that at least 50 tosses are needed?

8.18 A Colombian coffee exporter reports that the amount of impurities in 1 pound of coffee is a random variable with mean 3.4 mg and standard deviation 4 mg. In a sample of 100 pounds of coffee from this exporter, what is the probability that the sample mean is greater than 4.5 mg?

8.19 (Ross, 1998) Let X be a random variable having a gamma distribution with parameters n and 1. How large must n be in order to guarantee that:

$$P\left(\left|\frac{X}{n} - 1\right| > 0.01\right) < 0.01 \ ?$$

8.20 How many tosses of a fair coin are needed so that the probability that the average number of heads obtained differs at most 0.01 from 0.5 is at least 0.90?

8.21 Let X be a nonnegative random variable. Prove:

$$E(X) \leq \left[E\left(X^2\right)\right]^{\frac{1}{2}} \leq \left[E\left(X^3\right)\right]^{\frac{1}{3}} \leq \cdots .$$

8.22 Let $(X_n)_{n\geq 1}$ be a sequence of random variables, and let c be a real constant verifying, for any $\epsilon > 0$, the following condition:

$$\lim_{n\to\infty} P\left(|X_n - c| > \epsilon\right) = 0 .$$

Prove that for any bounded continuous function g we have that:

$$\lim_{n\to\infty} E\left(g\left(X_n\right)\right) = g\left(c\right) .$$

8.23 Examine the nature of convergence of $\{X_n\}$ defined below for the different values of k for $n = 1, 2, \cdots$:

$$\begin{aligned} P(X_n = n^k) &= \frac{1}{n} \\ P(X_n = 0) &= 1 - \frac{2}{n} \\ P(X_n = -n^k) &= \frac{1}{n}. \end{aligned}$$

8.24 Show that the convergence in the rth mean does not imply almost sure convergence for the sequence $\{X_n\}$ defined below:

$$P(X_n = 0) = 1 - \frac{1}{n}$$
$$P(X_n = -n^{\frac{1}{2r}}) = \frac{1}{n}.$$

8.25 Use CLT to show that $\lim_{n\to\infty} e^{-n} \sum_{i=0}^{n} \frac{n^i}{i!} \simeq 0.5$.

8.26 Let $(X_n)_n$ be a sequence of random variables such that $X_n \stackrel{d}{=} Exp(n)$. Show that $X_n \xrightarrow{P} 0$.

8.27 For each $n \geq 1$, let X_n be an uniformly distributed random variable over set $\left\{0, \frac{1}{n}, \frac{2}{n}, \cdots, \frac{n-1}{n}, 1\right\}$, that is:

$$P\left(X_n = \frac{k}{n}\right) = \frac{1}{n+1}, \quad k = 0, 1, \cdots, n\ .$$

Let U be a random variable with uniform distribution in the interval $[0,1]$. Show that $X_n \xrightarrow{d} U$.

8.28 Let $(X_n)_n$ be a sequence of i.i.d. random variables with $E(X_1) = Var(X_1) = \lambda \in (0,\infty)$ and $P(X_1 > 0) = 1$. Show that, for $n \to \infty$,

$$\sqrt{n}.\frac{Y_n - \lambda}{\sqrt{\lambda}} \xrightarrow{d} \mathcal{N}(0,1)$$

where:

$$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k\ .$$

8.29 Let $(X_n)_n$ be a sequence of random variables such that $X_n \stackrel{d}{=} Exp(n)$. Show that $X_n \xrightarrow{P} 0$.

8.30 Let $(\Omega, \Im, P) = ([0,1], \mathcal{B}(\mathbb{R}) \cap [0,1], \mathcal{U}([0,1]))$. Prove the following statements:

a) If $(X_n)_n$ is a sequence of random variables with $X_n \stackrel{d}{=} \mathcal{U}\left(\left[\frac{1}{2} - \frac{1}{n}, \frac{1}{2} + \frac{1}{n}\right]\right)$, then $X_n \xrightarrow{d} X$ with $X = \frac{1}{2}$.

b) The sequence $(X_n)_n$ with $X_n = \mathcal{X}_{[0,\frac{1}{2}+\frac{1}{n}]}$ converges in distribution but does not converge in probability to $X = \mathcal{X}_{[\frac{1}{2},1]}$.

8.31 Let $(X_n)_n$ be a sequence of i.i.d. random variables with $P(X_n = 1) = P(X_n = -1) = \frac{1}{2}$. Show that

$$\frac{1}{n}\sum_{j=1}^{n} X_j$$

converges in probability to 0 .

CHAPTER 9

INTRODUCTION TO STOCHASTIC PROCESSES

In the last eight chapters, we have studied probability theory, which is the mathematical study of random phenomena. A random phenomenon occurs through a stochastic process. In this chapter we introduce stochastic processes that can be defined as a collection of random variables indexed by some parameter. The parameters could be time (or length, weight, size, etc.). The word "stochastic[1]" means *random* or *chance*. The theory of stochastic processes turns out to be a useful tool in solving problems belonging to diverse disciplines such as engineering, genetics, statistics, economics, finance, etc. This chapter discusses Markov chains, Poisson processes, and renewal processes. In the next chapter, we introduce some important stochastic processes for financial mathematics, Wiener processes, martingales, and stochastic integrals.

[1]The word stochastic comes from Greek *stokhastikos*.

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

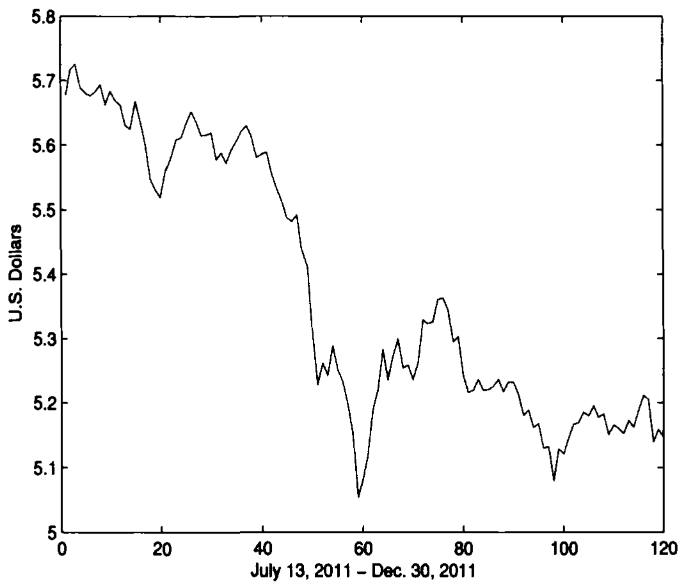


Figure 9.1 Sample path for the U.S. dollar value of 10,000 Colombian pesos

9.1 DEFINITIONS AND PROPERTIES

Definition 9.1 (Stochastic Process) *A real stochastic process is a collection of random variables* $\{X_t;\ t \in T\}$ *defined on a common probability space* $(\Omega, \Im, P)$ *with values in* $\mathbb{R}$*.* T *is called the index set of the process or parametric space, which is usually a subset of* $\mathbb{R}$*. The set of values that the random variable* X_t *can take is called the state space of the process and is denoted by* S*.*

The mapping defined for each fixed $\omega \in \Omega$,

$$\begin{aligned} X(\omega) \ : \ & T \to S \\ & t \mapsto X_t(\omega) \end{aligned}$$

is called a *sample path* of the process over time or a *realization* of the stochastic process.

■ EXAMPLE 9.1

The U.S. dollar value of 10,000 Colombian pesos at the end of each day is a stochastic process $\{X_t; t \in T\}$. The sample path of $\{X_t; t \in T\}$

between July 13, 2011 and December 30, 2011, is shown in Figure 9.1.
▲

Stochastic processes are classified into four types, depending upon the nature of the state space and parameter space, as follows:

■ **EXAMPLE 9.2**

1. Discrete-state, discrete-time stochastic process

   (a) The number of individuals in a population at the end of year t can be modeled as a stochastic process $\{X_t; t \in T\}$, where the index set $T = \{0, 1, 2, \cdots\}$ and the state space $S = \{0, 1, 2, \cdots\}$.

   (b) A motor insurance company reviews the status of its customers yearly. Three levels of discounts are possible $(0, 10\%, 25\%)$ depending on the accident record of the driver. Let X_t be the percentage of discount at the end of year t. Then the stochastic process $\{X_t; t \in T\}$ has $T = \{0, 1, 2, \cdots\}$ and $S = \{0, 10, 25\}$.

2. Discrete-state, continuous-time stochastic process

   (a) The number of incoming calls X_t in an interval $[0, t]$. Then the stochastic processes $\{X_t; t \in T\}$ has $T = \{t : 0 \leq t < \infty\}$ and $S = \{0, 1, \cdots\}$.

   (b) The number of cars X_t parked at a commercial center in the time interval $[0, t]$. Then the stochastic processes $\{X_t; t \in T\}$ has $T = \{t : 0 \leq t < \infty\}$ and $S = \{0, 1, \cdots\}$.

3. Continuous-state, discrete-time stochastic process

   (a) The share price for an asset at the close of trading on day t with $T = \{0, 1, 2, \cdots\}$ and $S = \{x : 0 \leq x < \infty\}$. Then it is a discrete-time stochastic process with the continuous-state space.

4. Continuous-state, continuous-time stochastic process

   (a) The value of the Dow-Jones index at time t where $T = \{t : 0 \leq t < \infty\}$ and $S = \{x : 0 \leq x < \infty\}$. Then it is a continuous-time stochastic process with the continuous-state space. ▲

Definition 9.2 (Finite-Dimensional Distributions of the Process) *Let $\{X_t; t \in T\}$ be a stochastic process and $\{t_1, t_2, \cdots, t_n\} \subset T$ where $t_1 < t_2 < \cdots < t_n$. The finite dimensional distribution of the process is defined by:*

$$F_{t_1 \cdots t_n}(x_1, \cdots, x_n) := P(X_{t_1} \leq x_1, \cdots, X_{t_n} \leq x_n),\ x_i \in \mathbb{R},\ i = 1, \cdots, n\ .$$

The family of all finite-dimensional distributions determines many important properties of the stochastic process. Under some conditions it is possible to show that a stochastic process is uniquely determined by its finite-dimensional distributions (advanced reader may refer to Breiman, 1992).

We now discuss a few important characteristics and classes of stochastic processes.

Definition 9.3 (Independent Increments) *If, for all $t_0, t_1, t_2, \cdots, t_n$ such that $t_0 < t_1 < t_2 < \cdots < t_n$, the random variables $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \cdots, X_{t_n} - X_{t_{n-1}}$ are independent (or equivalently $X_{t+\tau} - X_\tau$ is independent of X_s for $s < \tau$), then the process $\{X_t; t \in T\}$ is said to be a process with independent increments.*

Definition 9.4 (Stationary Increments) *A stochastic process $\{X_t; t \in T\}$ is said to have stationary increments if $X_{t_2+\tau} - X_{t_1+\tau}$ has the same distribution as $X_{t_2} - X_{t_1}$ for all choices of t_1, t_2 and $\tau > 0$.*

Definition 9.5 (Stationary Process) *If for arbitrary $t_1, t_2, \cdots, t_n$, such that $t_1 < t_2 < \cdots < t_n$, the joint distributions of the vector random variables $(X_{t_1}, X_{t_2}, \cdots, X_{t_n})$ and $(X_{t_1+h}, X_{t_2+h}, \cdots, X_{t_n+h})$ are the same for all $h > 0$, then the stochastic process $\{X_t; t \in T\}$ is said to be a stationary stochastic process of order n (or simply a stationary process). The stochastic process $\{X_t; t \in T\}$ is said to be a strong stationary stochastic process or strictly stationary process if the above property is satisfied for all n.*

■ EXAMPLE 9.3

Suppose that $\{X_n; n \geq 1\}$ is a sequence of independent and identically distributed random variables. We define the sequence $\{Y_n; n \geq 1\}$ as

$$Y_n = X_n + aX_{n-1}$$

where a is a real constant. Then it is easily seen that $\{Y_n; n \geq 1\}$ is strictly stationary. ▲

Definition 9.6 (Second-Order Process) *A stochastic process $\{X_t; t \in T\}$ is called a second-order process if $E\left((X_t)^2\right) < \infty$ for all $t \in T$.*

■ EXAMPLE 9.4

Let Z_1 and Z_2 be independent normally distributed random variables, each having mean 0 and variance σ^2. Let $\lambda \in \mathbb{R}$ and:

$$X_t = Z_1 \cos(\lambda t) + Z_2 \sin(\lambda t), \quad t \in \mathbb{R}.$$

$\{X_t; t \in T\}$ is a second-order stationary process. ▲

Definition 9.7 (Covariance Stationary Process) *A second-order stochastic process $\{X_t; t \in T\}$ is called covariance stationary or weakly stationary if its mean function $m(t) = E[X_t]$ is independent of t and its covariance function $Cov(X_s, X_t)$ depends only on the difference $\mid t-s \mid$ for all $s, t \in T$. That is:*

$$Cov(X_s, X_t) = f(\mid t-s \mid) \ .$$

■ EXAMPLE 9.5

Let $\{X_n; n \geq 1\}$ be uncorrelated random variables with mean 0 and variance 1. Then $Cov(X_m, X_n) = E(X_m X_n)$ equals 0 if $m \neq n$ and 1 if $m = n$. Then this shows that $\{X_n; n \geq 1\}$ is a covariance stationary process. ▲

Definition 9.8 (Evolutionary Process) *A stochastic process which is not stationary (in any sense) is said to be an evolutionary stochastic process.*

■ EXAMPLE 9.6

Consider the process $\{X_t; t \geq 0\}$, where $X_t = A_1 + A_2 t$, A_1 and A_2 are independent random variables with $E(A_i) = \mu_i$, $Var(A_i) = \sigma_i^2$ for $i = 1, 2$. It easy to see that:

$$\begin{aligned} E(X_t) &= \mu_1 + \mu_2 t \\ Var(X_t) &= \sigma_1^2 + \sigma_2^2 t^2 \\ Cov(X_s, X_t) &= \sigma_1^2 + st\sigma_2^2 \ . \end{aligned}$$

These are functions of t and the process is evolutionary. ▲

The class of processes defined below, known as Markov processes is a very important process both in applications as well as in the development of the theory of stochastic processes.

Definition 9.9 (Markov Process) *Let $\{X_t \ t \geq 0\}$ be a stochastic process defined over a probability space $(\Omega, \Im, P)$ and with state space $(\mathbb{R}, \mathcal{B})$. We say that $\{X_t \ t \geq 0\}$ is a Markov process if for any $0 \leq t_1 < t_2 < \cdots < t_n$ and for any $B \in \mathcal{B}$:*

$$P\left(X_{t_n} \in B \mid X_{t_1}, \cdots, X_{t_{n-1}}\right) = P\left(X_{t_n} \in B \mid X_{t_{n-1}}\right) \ .$$

Note 9.1

1. *Roughly a Markov process is a process such that, given the value X_s, the distribution of X_t for $t > s$ does not depend on the values of X_u, $u < s$.*

2. *Any stochastic process which has independent increments is a Markov process.*

A discrete-state Markov process is known as a Markov chain. Based on the values of the index set T, the Markov chain is classified as a discrete-time Markov chain (DTMC) or a continuous-time Markov chain (CTMC).

In the next section we will work with discrete-time parameter Markov processes and with discrete-state space, that is, we will continue our study of the so-called Markov chains with discrete-time parameter.

9.2 DISCRETE-TIME MARKOV CHAIN

This section intends to present the basic concepts and theorems related to the theory of discrete-time Markov chains with examples. This section was inspired by the class notes of the courses Stochastik I und II given by Professor H.J. Schuh at the Johannes Gutenberg Universität (Mainz, Germany).

Definition 9.10 (Discrete-Time Markov Chain) *A sequence of random variables $(X_n)_{n\in\mathbb{N}}$ with discrete-state space is called a discrete-time Markov chain if it satisfies the conditions*

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j \mid X_n = i) \quad (9.1)$$

for all $n \in \mathbb{N}$ and for all $i_0, i_1, \cdots, i_{n-1}, i, j \in S$ with:

$$P(X_0 = i_0, \cdots, X_n = i_n) > 0.$$

In other words, the condition (9.1) *implies the following: if we know the present state "$X_n = i$", the knowledge of past history "$X_{n-1}, X_{n-2}, \cdots, X_0$" has no influence on the probabilistic structure of the future state X_{n+1}.*

■ EXAMPLE 9.7

Suppose that a coin is tossed repeatedly and let:

$$X_n := \text{"number of heads obtained in the first } n \text{ tosses"}.$$

It is clear that the number of heads obtained in the first $n+1$ tosses only depends on the knowledge of the number of heads obtained in the first n tosses and therefore:

$$P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1) = P(X_{n+1} = j \mid X_n = i)$$

for all $n \in \mathbb{N}$ and for all $i, j, i_{n-1}, \cdots, i_1 \in \mathbb{N}$. ▲

■ **EXAMPLE 9.8**

Three boys A, B and C throw a ball to each other. A always throws the ball to B and B always throws the ball to C where C is equally likely to throw the ball to A or B. Let $X_n :=$ "the boy who has the ball in the nth throw". The state space of the process is $S = \{A, B, C\}$ and it is clear that $\{X_n; n \geq 0\}$ is a Markov chain, since the person throwing the ball is not influenced by those who previously had the ball. ▲

■ **EXAMPLE 9.9**

Consider a school which consists of 400 students with 250 of them being boys and 150 girls. Suppose that the students are chosen randomly one followed by the other for a medical examination. Let X_n be the sex of the nth student chosen. It is easy to see that $\{X_n;\ n = 1, 2, \cdots, 400\}$ is not a discrete-time Markov chain. ▲

■ **EXAMPLE 9.10**

Let $Y_0, Y_1, \cdots, Y_n$ be nonnegative, independent and identically distributed random variables. The sequence $\{X_n; n \geq 0\}$ with

$$\begin{aligned} X_0 &:= Y_0 \\ X_n &:= X_0 + Y_1 + \cdots + Y_n \text{ for } n \geq 1 \end{aligned}$$

is a Markov chain because:

$$\begin{aligned}
&P(X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0) \\
&= \frac{P(X_{n+1} = j, X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0)}{P(X_n = i, X_{n-1} = i_{n-1}, \cdots, X_1 = i_1, X_0 = i_0)} \\
&= \frac{P(X_0 = i_0, Y_1 = i_1 - i_0, \cdots, Y_n = i - i_{n-1}, Y_{n+1} = j - i)}{P(X_0 = i_0, Y_1 = i_1 - i_0, \cdots, Y_n = i - i_{n-1})} \\
&= \frac{P(X_0 = i_0)P(Y_1 = i_1 - i_0) \cdots P(Y_n = i - i_{n-1})P(Y_{n+1} = j - i)}{P(X_0 = i_0)P(Y_1 = i_1 - i_0) \cdots P(Y_n = i - i_{n-1})} \\
&= P(Y_{n+1} = j - i) = P(X_{n+1} = j \mid X_n = i). \quad ▲
\end{aligned}$$

■ **EXAMPLE 9.11 A Simple Queueing Model**

Let $0, 1, 2, \cdots$ be the times at which an elevator starts. It is assumed that the elevator can transport only one person at a time. Between the times n and $n+1$, Y_n people who want to get into the elevator arrive. Also assume that the random variables Y_n, $n = 0, 1, 2, \cdots$, are independent. The queue length X_n immediately before the start of the elevator at time n is equal to:

$$X_n = \max(0, X_{n-1} - 1) + Y_{n-1}, \; n \geq 1.$$

Suppose that $X_0 = 0$. Since X_i with $i \leq n$ can be expressed in terms of $Y_1, Y_2, \cdots, Y_{n-1}$, we have that Y_n is independent of $(X_0, X_1, \cdots, X_n)$ as well. Thus:

1. If $i_n \geq 1$, then:

$$\begin{aligned}
&P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) \\
&\quad = \frac{P(X_{n+1} = i_{n+1}, X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0)}{P(X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0)} \\
&\quad = \frac{P(Y_n = i_{n+1} - i_n + 1) P(X_n = i_n, \cdots, X_0 = i_0)}{P(X_n = i_n, \cdots, X_0 = i_0)} \\
&\quad = P(Y_n = i_{n+1} - i_n + 1) \\
&\quad = P(X_{n+1} = i_{n+1} \mid X_n = i_n) .
\end{aligned}$$

The second equality follows from the fact that $i_n \geq 1$, then $i_{n+1} = X_{n+1} = X_n - 1 + Y_n$, which implies that $Y_n = i_{n+1} - X_n + 1$.

2. If $i_n = 0$, then $X_{n+1} = Y_n$, and in this case:

$$\begin{aligned}
&P(X_{n+1} = i_{n+1} \mid X_n = i_n, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) \\
&\quad = \frac{P(Y_n = i_{n+1}, X_n = i_n, \cdots, X_0 = i_0)}{P(X_n = i_n, \cdots, X_0 = i_0)} \\
&\quad = P(X_{n+1} = i_{n+1} \mid X_n = 0) . \quad \blacktriangle
\end{aligned}$$

Lemma 9.1 *If $\{X_n; n \geq 0\}$ is a Markov chain, then for all n and for all $i_0, i_1, \cdots, i_n \in S$ we have:*

$$\begin{aligned}
P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) &= P(X_0 = i_0) P(X_1 = i_1 \mid X_0 = i_0) \\
&\quad \cdots P(X_n = i_n \mid X_{n-1} = i_{n-1}) .
\end{aligned}$$

Proof:

$$\begin{aligned}
&P(X_0 = i_0) P(X_1 = i_1 \mid X_0 = i_0) \cdots P(X_n = i_n \mid X_{n-1} = i_{n-1}) \\
&= P(X_0 = i_0) \frac{P(X_1 = i_1, X_0 = i_0)}{P(X_0 = i_0)} \cdots \frac{P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n)}{P(X_0 = i_0, X_1 = i_1, \cdots, X_{n-1} = i_{n-1})} \\
&= P(X_0 = i_0, X_1 = i_1, \cdots, X_n = i_n) .
\end{aligned}$$

■

The previous lemma states that to know the joint distribution of $X_0, X_1, \cdots, X_n$, it is enough to know $P(X_0 = i_0)$, $P(X_1 = i_1 \mid X_0 = i_0)$, etc. Moreover, for any finite set $\{j_1, j_2, \cdots, j_l\}$ of subindices, any probability that involves the variables $X_{j1}, X_{j2}, \cdots, X_{jl}$ with $j_1 < j_2 < \cdots < j_l$, $l = 1, 2, \cdots, n$, can be obtained from:

$$P(X_0 = i_0, \cdots, X_n = i_n).$$

Then it follows that one can know the joint distribution of the random variables $X_{j1}, X_{j2}, \cdots, X_{jl}$ from the knowledge of the values of $P(X_0 = i_0)$, $P(X_1 = i_1 \mid X_0 = i_0)$, etc. These probabilities are so important that they have a special name.

Definition 9.11 (Transition Probability) *Let $\{X_n; n \geq 0\}$ be a Markov chain. The probabilities*

$$p_{ij} := P(X_{n+1} = j \mid X_n = i)$$

(if defined) are called transition probabilities.

Definition 9.12 *A Markov chain $\{X_n; n \geq 0\}$ is called homogeneous or a Markov chain with stationary probabilities if the transition probabilities do not depend on n.*

Note 9.2 *In this book we only consider homogeneous Markov chains unless otherwise specially mentioned.*

Definition 9.13 *The probability distribution $\pi := (\pi_i)_{i \in S}$ with*

$$\pi_i := P(X_0 = i)$$

is called the initial distribution.

Because of the dual subscripts it is convenient to arrange the transition probabilities in a matrix form.

Definition 9.14 (Transition Probability Matrix) *The matrix*

$$P = (p_{ij}) = \begin{pmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{12} & \cdots \\ p_{20} & p_{21} & p_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is called the transition probability matrix or stochastic matrix. Note that:

$$\begin{aligned} p_{ij} &\geq 0 \text{ for all } i, j \in S \\ \sum_j p_{ij} &= 1 \text{ for all } i \in S\ . \end{aligned}$$

■ **EXAMPLE 9.12**

From Example 9.11 we have that:

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) = P(Y_n = j - i + 1) \text{ for } i > 0$$
$$p_{0j} = P(Y_n = j) \text{ for } i = 0 .$$

If we define $p_j := P(Y_n = j)$, then the transition probability matrix P of the Markov chain is given by:

$$P = \begin{pmatrix} p_0 & p_1 & p_2 & p_3 & \cdots \\ p_0 & p_1 & p_2 & p_3 & \cdots \\ 0 & p_0 & p_1 & p_2 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \end{pmatrix} . \qquad \blacktriangle$$

■ **EXAMPLE 9.13**

On any given day Gary is cheerful (C), normal (N) or depressed (D). If he is cheerful today, then he will be C, N or D tomorrow with probabilities 0.5, 0.4, 0.1, respectively. If he is feeling so-so today, then he will be C, N or D tomorrow with probabilities 0.3, 0.4, 0.3. If he is glum today, then he will be C, N, or D tomorrow with probabilities 0.2, 0.3, 0.5. Let X_n denote Gary's mood on the nth day. Then $\{X_n;\ n \geq 0\}$ is a three-state discrete-time Markov chain (state $0 = C$, state $1 = N$, state $2 = D$) with transition probability matrix

$$P = \begin{pmatrix} 0.5 & 0.4 & 0.1 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.3 & 0.5 \end{pmatrix} . \qquad \blacktriangle$$

■ **EXAMPLE 9.14 Simple Random Walk Model**

A discrete-time Markov chain whose state space is given by the integers $i = 0, \pm1, \pm2, \cdots$ is said to be a random walk if for some number $0 < p < 1$:

$$p_{i,i+1} = p = 1 - p_{i,i-1}, \quad i = 0, \pm1, \pm2, \cdots .$$

The preceding DTMC is called a simple random walk for we may think of it as being a model for an individual walking on a straight line who at each point of time either takes one step to the right with probability p or one step to the left with probability $1 - p$. The one-step transition probability matrix is given by:

$$P = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \cdots & 0 & 1-p & 0 & p & 0 & \cdots \\ \cdots & \cdots & 0 & 1-p & 0 & p & \cdots \\ \cdots & \cdots & \ddots & \ddots & \ddots & \ddots & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} . \quad \blacktriangle$$

■ EXAMPLE 9.15 Gambler's Ruin

Suppose that we have two players A and B and that player A who started a game has a capital of $x \in \mathbb{N}$ dollars and player B has a capital of $y \in \mathbb{N}$ dollars. Let $a := x + y$. In each round of the game, either A wins one dollar from B with probability p or B wins one dollar from A with probability q, with $p + q = 1$. The game goes on until one of the players loses all of his capital, that is, until $X_n = 0$ or $X_n = a$, since $X_n =$ "The capital of the player A in the nth round". In this case we have $T = \mathbb{N}$ and $S = \{0, 1, \cdots, a\}$. It is easy to verify that $\{X_n; n \geq 0\}$ is a Markov chain. Next, we will see its initial distribution and transition matrix. We have that $P(X_0 = x) = 1$, and hence

$$\pi = (0, \cdots, 0, 1, 0, \cdots, 0) =: \epsilon_x$$

where the 1 appears in the xth component and the matrix P is equal to:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ q & 0 & p & 0 & \dots & 0 \\ 0 & q & 0 & p & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & & \\ 0 & 0 & 0 & 0 & \dots & 1 \end{pmatrix} . \quad \blacktriangle$$

■ EXAMPLE 9.16

Let $\{X_n;\ n \geq 0\}$ be a Markov chain with states $0, 1, 2$ and with transition probability matrix

$$P = \begin{pmatrix} 0.75 & 0.25 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0 & 0.75 & 0.25 \end{pmatrix} .$$

The initial distribution is $P(X_0 = i) = \frac{1}{3}$, $i = 0, 1, 2$. Then:

$$\begin{aligned} P(X_1 = 1 \mid X_0 = 2) &= 0.75 \\ P(X_2 = 2 \mid X_1 = 1) &= 0.25 \\ P(X_2 = 2, X_1 = 1 \mid X_0 = 2) &= P(X_2 = 2 \mid X_1 = 1) P(X_1 = 1 \mid X_0 = 2) \\ &= \frac{3}{16}. \quad \blacktriangle \end{aligned}$$

We now explain how Markov chains can be represented as graphs. In a Markov chain, a set of states can be represented by a "network" in which states are vertices and one-step transitions between states are represented by directed arcs. Each of the transitions corresponds to a probability. This graphical representation for the Markov chain is called a *state transition diagram*. If for some transition the probability of occurrence is zero, then it indicates that the transition is not possible and the corresponding arc is not drawn.

For example, if $\{X_n; n \geq 1\}$ is a Markov chain with $S = \{0, 1, 2, 3\}$ and transition probability matrix

$$P = \begin{pmatrix} \frac{1}{5} & \frac{2}{5} & 0 & \frac{2}{5} \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 \\ \frac{5}{6} & 0 & 0 & \frac{1}{6} \\ \frac{2}{13} & \frac{3}{13} & \frac{5}{13} & \frac{3}{13} \end{pmatrix},$$

then the state transition diagram is shown in Figure 9.2. In the study of Markov chains the following equations, called Chapman-Kolmogorov equations are very important.

Theorem 9.1 (Chapman-Kolmogorov Equations) *If the sequence of random variables $\{X_n; n \geq 0\}$ is a Markov chain and if $k < m < n$, then we have for all h, $j \in S$:*

$$P(X_n = j \mid X_k = h) = \sum_{i \in S} P(X_n = j \mid X_m = i) P(X_m = i \mid X_k = h).$$

Proof:

$$\begin{aligned} P(X_k = h, X_n = j) &= \sum_{i \in S} P(X_k = h, X_m = i, X_n = j) \\ &= \sum_{i \in S} P(X_k = h, X_m = i) P(X_n = j \mid X_k = h, X_m = i) \\ &= \sum_{i \in S} P(X_m = i, X_k = h) P(X_n = j \mid X_m = i) \\ &= \sum_{i \in S} P(X_m = i \mid X_k = h) P(X_k = h) P(X_n = j \mid X_m = i) \ . \end{aligned}$$

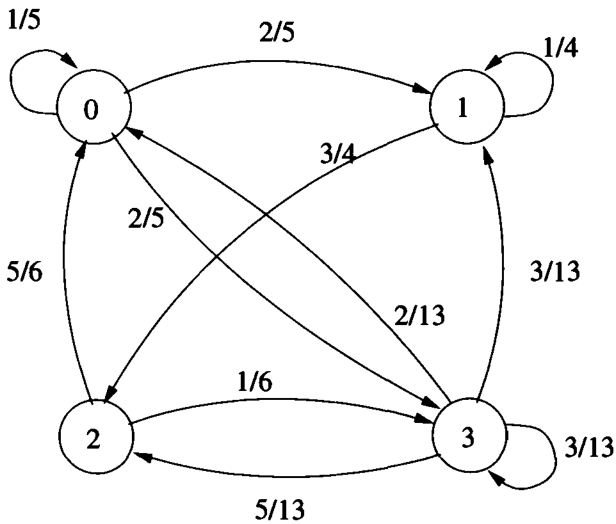


Figure 9.2 State transition diagram

That is:

$$\frac{P(X_k = h, X_n = j)}{P(X_k = h)} = \sum_{i \in S} P(X_m = i \mid X_k = h) P(X_n = j \mid X_m = i) .$$

■

To give an interpretation to the Chapman-Kolmogorov equations, we introduce the following concept:

Definition 9.15 *The probability*

$$p_{ij}^{(m)} = P(X_{n+m} = j \mid X_n = i), \ m \in \mathbb{N}, \ i, j \in S,$$

is the m-step transition probability from i to j:

$$p_{ij}^{(0)} := P(X_n = j \mid X_n = i) = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} =: \delta_{ij}$$

where δ_{ij} is Kronecker's delta. The matrix $P^{(m)} := \left(p_{ij}^{(m)}\right)_{ij \in S}$ is the m-step transition matrix.

■ EXAMPLE 9.17

Consider the simple random walk of Example 9.14. The state transition diagram is shown in Figure 9.3. Assume that the process starts at the

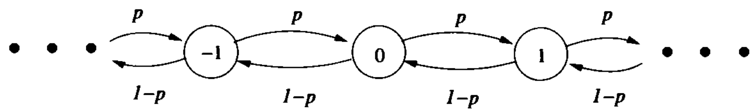


Figure 9.3 State transition diagram

origin. We have:

$$p_{ij}^{(n)} = \begin{cases} \binom{n}{\frac{n+j-i}{2}} p^{\frac{n+j-i}{2}} q^{\frac{n-j+i}{2}} & \text{if} \quad n+j-i \text{ is even} \\ 0 & \text{otherwise.} \end{cases}$$

This can be seen by noting that there will be $\frac{n+j-i}{2}$ positive steps and $\frac{n-j+i}{2}$ negative steps in order to go from state i to j in n steps when $n+j-i$ is even. ▲

Corollary 9.1 *The m-step transition probability matrix is the mth power of the transition matrix P.*

Corollary 9.2 *Let $\{X_n; n \geq 0\}$ be a Markov chain with transition probability matrix P and initial distribution π. Then for each $n \geq 1$ and for each $k \in S$ we have:*

$$P(X_n = k) = \sum_{j \in S} p_{jk}^{(n)} \pi_j \ .$$

Proof:

$$\begin{aligned} P(X_n = k) &= \sum_{j \in S} P(X_n = k, X_0 = j) \\ &= \sum_{j \in S} P(X_n = k \mid X_0 = j) P(X_0 = j) \\ &= \sum_{j \in S} p_{jk}^{(n)} \pi_j \ . \end{aligned}$$

■

Note 9.3 *The Chapman-Kolmogorov equations provide a procedure to compute the n-step transition probabilities. Indeed it follows that:*

$$p_{ij}^{(n+m)} = \sum_{k=0}^{\infty} p_{ik}^{(n)} p_{kj}^{(m)} \quad \textit{for all } n, m \geq 0, \ \textit{all } i, j \in S \ .$$

In matrix form, we can write these as

$$P^{(n+m)} = P^{(n)} \cdot P^{(m)}$$

where $\cdot$ denotes matrix multiplication. Here, $P^{(n)}$ is the matrix consisting of n-step transition probabilities.

Note that beginning with

$$P^{(2)} = P^{(1+1)} = P \cdot P = P^2$$

and continuing by induction, we can show that the n-step transition matrix can be obtained by multiplying matrix P by itself n times, that is:

$$P^{(n)} = P^{(n-1+1)} = P^{(n-1)} \cdot P = P^n \ .$$

■ **EXAMPLE 9.18**

Let $\{X_n; n \geq 1\}$ be a Markov chain with state space $S = \{0, 1\}$, initial distribution $\pi = \left(\frac{1}{2}, \frac{1}{2}\right)$ and transition matrix

$$P = \begin{pmatrix} 0.1 & 0.9 \\ 0.3 & 0.7 \end{pmatrix} \ .$$

Then

$$P(X_3 = 0) = \sum_{j \in S} p_{j0}^{(2)} \pi_j^1$$

with

$$P \times P = \begin{pmatrix} 0.28 & 0.72 \\ 0.24 & 0.76 \end{pmatrix}$$

so that:

$$P(X_3 = 0) = 0.26 \ . \qquad \blacktriangle$$

9.2.1 Classification of States

One of the fundamental problems in the study of Markov chains is the analysis of its asymptotic behavior. As we will see later, it depends on whether the chain returns to its starting point with probability 1 or not. For this analysis we need to classify the states, which is the objective of this section.

Definition 9.16 (Accessibility) *State j is said to be accessible from state i in $n \geq 0$ steps if $p_{ij}^{(n)} > 0$. This is written as $i \to j[n]$. We say that state j is accessible from state i if there exists $n \geq 0$ such that $p_{ij}^{(n)} > 0$. In this case we write $i \to j$.*

Lemma 9.2 *The relation "$\to$" is transitive.*

Proof: Suppose that $i \to j$ and $j \to k$. Then there exist nonnegative integers r and l such that:

$$p_{ij}^{(r)} > 0 \quad \text{and} \quad p_{jk}^{(l)} > 0.$$

From the Chapman-Kolmogorov equation, we have:

$$p_{ik}^{(r+l)} \geq p_{ij}^{(r)} p_{jk}^{(l)} > 0.$$

Therefore, $i \to k$. ∎

Definition 9.17 *States i and j communicate if $i \to j$ and $j \to i$. This is written as $i \leftrightarrow j$.*

It is easy to verify that "$i \leftrightarrow j$" is an equivalence relation over S and thus the equivalence classes

$$C(i) := \{j \in S : i \leftrightarrow j\},\ i \in S\ ,$$

form a partition of S.

Definition 9.18 *A Markov chain is said to be irreducible if the state space consists of only one class, that is, all states communicate with each other.*

A Markov chain for which there is only one communication class is called an irreducible Markov chain: Note that, in this case, all states communicate.

■ EXAMPLE 9.19

Let $\{X_n; n \geq 0\}$ be a Markov chain with state space $S = \{1, 2, 3\}$, initial distribution $\pi = (1, 0, 0)$ and transition matrix

$$P = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} .$$

It is to be observed that $C(1) = C(2) = C(3) = \{1, 2, 3\}$, that is, $(X_n)_{n \geq 0}$ is an irreducible Markov chain. ▲

Definition 9.19 (Absorbing State) *A state i is said to be an absorbing state if $p_{ii} = 1$ or equivalently $p_{ij} = 0$ for all $j \neq i$.*

■ EXAMPLE 9.20

Consider a Markov chain consisting of the four states 0, 1, 2, 3 and having transition probability matrix

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 1 \end{pmatrix} .$$

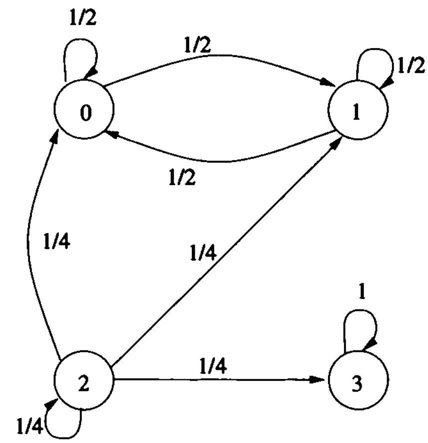


Figure 9.4 State transition diagram for Example 9.20

The state transition diagram for this Markov chain is shown in Figure 9.4. The classes of this Markov chain are $\{0, 1\}$ and $\{3\}$. Note that while state 0 (or 1) is accessible from state 2, the reverse is not true. Since state 3 is an absorbing state, that is, $p_{33} = 1$, no other state is accessible from it. ▲

Definition 9.20 *Let $i \in S$ fixed. The period of i is defined as follows:*

$$\lambda(i) := GCD\left\{n \geq 1 : p_{ii}^{(n)} > 0\right\}$$

where GCD stands for greatest common divisor. If $p_{ii}^{(n)} = 0$ for all $n \geq 1$, then we define $\lambda(i) := 0$.

Definition 9.21 (Aperiodic) *State i is called aperiodic when $\lambda(i) = 1$.*

Definition 9.22 *A Markov chain with all states aperiodic is called an aperiodic Markov chain.*

■ EXAMPLE 9.21

In the example of the gambler's ruin, we have that:

$$\begin{aligned}\lambda(0) &= GCD\left\{n \geq 1 : p_{00}^{(n)} > 0\right\} \\ &= 1 = GCD\left\{n \geq 1 : p_{aa}^{(n)} > 0\right\} = \lambda(a) \\ \lambda(1) &= GCD\left\{n \geq 1 : p_{11}^{(n)} > 0\right\} = 2\ .\end{aligned}$$

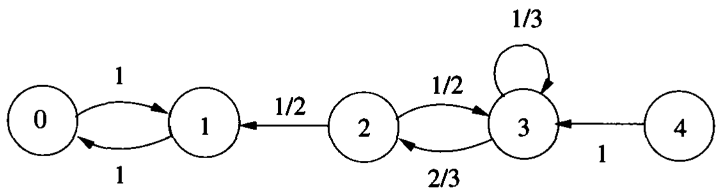


Figure 9.5 State transition diagram for Example 9.22

The following theorem states that if two states are communicating with each other, then they have the same period. ▲

Theorem 9.2 *If* $i \leftrightarrow j$*, then* $\lambda(i) = \lambda(j)$.

Proof: Suppose that $j \to j[n]$ and let $k, m \in \mathbb{N}$ such that $i \to j[k]$ and $j \to i[m]$. Then $i \to i[m+k]$ and $i \to i[m+n+k]$. Therefore, $\lambda(i)$ divides both $m+k$ and $m+n+k$ and from this it follows that $\lambda(i)$ divides n. That is, $\lambda(i)$ is a common divisor for all n such that $j \to j[n]$ and this implies that $\lambda(i) \leq \lambda(j)$. Similarly it can be shown that $\lambda(j) \leq \lambda(i)$. ∎

■ EXAMPLE 9.22

Let $\{X_n; n \geq 1\}$ be a Markov chain with state space $S = \{0, 1, 2, 3, 4\}$ and transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The graphical representation of the chain is shown in Figure 9.5. Then:

$$\begin{aligned} C(0) &= C(1) = \{0, 1\} \\ C(2) &= C(3) = \{2, 3\} \\ C(4) &= \{4\}\ . \end{aligned}$$

It is clear that:

$$\begin{aligned} \lambda(0) &= \lambda(1) = 2 \\ \lambda(2) &= \lambda(3) = 1 \\ \lambda(4) &= 0\ . \end{aligned}$$ ▲

Suppose that the Markov chain $\{X_n; n \geq 0\}$ starts with the state j. Let τ_k be the first passage time to state k:

$$\tau_k = \min\{n \geq 1 : X_n = k\} .$$

If $\{n \geq 1 : X_n = k\} = \emptyset$, we define $\tau_k = \infty$. For $n \geq 1$, define:

$$\begin{aligned} f_{jk}^{(n)} &= P(\tau_k = n \mid X_0 = j) \\ &= P\{X_n = k, X_\nu \neq k, \nu = 1, \cdots, n-1 \mid X_0 = j\} . \end{aligned}$$

Let $p_{jk}^{(n)}$ be the probability that the chain reaches the state k (not necessarily for the first time) after n transitions. A relation between $f_{jk}^{(n)}$ and $p_{jk}^{(n)}$ is as follows.

Theorem 9.3 (First Entrance Theorem)

$$p_{jk}^{(n)} = \sum_{r=0}^{n} f_{jk}^{(r)}\, p_{kk}^{(n-r)} \tag{9.2}$$

where $f_{jk}^{(0)} = 0, \quad p_{kk}^{(0)} = 1, \quad f_{jk}^{(1)} = p_{jk}$.

If $j = k$ we refer to $f_{jj}^{(n)}$ as the probability that the first return to state j occurs at time n. By definition, we have $f_{jj}^{(0)} = 0$.

For fixed states j and k, we have

$$f_{jk} = \sum_{n=1}^{\infty} f_{jk}^{(n)}, \tag{9.3}$$

which is the probability that starting with state j the system ever reaches state k. If $j = k$, we let $f_{jj} = \sum_{n=1}^{\infty} f_{jj}^{(n)}$ to denote the probability of ultimately returning to state j.

Definition 9.23 (Recurrent State) *A state j is said to be persistent or recurrent if $f_{jj} = 1$ (i.e., return to state j is certain).*

Definition 9.24 (Transient State) *A state j is said to be transient if $f_{jj} < 1$ (i.e., return to state j is uncertain).*

■ EXAMPLE 9.23

Consider the Markov chain with state space $S = \{0, 1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{pmatrix} .$$

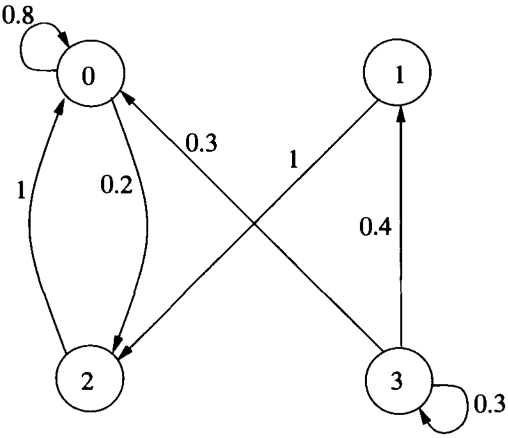


Figure 9.6 State transition diagram for Example 9.23

The state transition diagram for this Markov chain is shown in Figure 9.6.

In this case:

$$\begin{aligned}
f_{00} &= f_{00}^{(1)} + f_{00}^{(2)} + f_{00}^{(3)} + \cdots = 0.8 + (0.2)(1.0) + 0 = 1 \\
f_{11} &= f_{11}^{(1)} + f_{11}^{(2)} + f_{11}^{(3)} + \cdots = 0 \\
f_{22} &= f_{22}^{(1)} + f_{22}^{(2)} + f_{22}^{(3)} + \cdots \\
&= 0 + (1.0)(0.2) + (1.0)(0.8)(0.2) + (1.0)(0.8)^2(0.2) + \cdots = 1 \\
f_{33} &= f_{33}^{(1)} + f_{33}^{(2)} + f_{33}^{(3)} + \cdots = 0.3 \,.
\end{aligned}$$

We have $C(0) = \{0, 2\}$, $C(1) = \{1\}$, $C(3) = \{3\}$. The states 0 and 2 are recurrent and states 1 and 3 are transient. ▲

■ EXAMPLE 9.24

Consider the Markov chain with state space $S = \{1, 2, 3, 4, 5\}$ and transition matrix

$$P = \begin{pmatrix} 0.5 & 0 & 0 & 0.5 & 0 \\ 0 & 0.6 & 0 & 0 & 0.4 \\ 0.3 & 0 & 0.7 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

In this case, using Definition 9.17, $C(1) = \{1, 3, 4\}$, $C(2) = \{2, 5\}$. All the states are recurrent. ▲

Note 9.4 *The transition matrix of a finite Markov chain can always be written in the form*

$$P = \begin{pmatrix} R & 0 \\ A & B \end{pmatrix} \tag{9.4}$$

where R corresponds to the submatrix that gives the probability of transition between recurrent states, A is the submatrix whose components are the probabilities of transition from transient states to recurrent states, 0 is the zero matrix and B is the submatrix of transition probabilities between transient states. If among the recurrent states there is an absorbing state, it is placed first in this reordering of the states. This representation of the transition matrix is known as the canonical form of the transition matrix.

■ **EXAMPLE 9.25**

Consider the Markov chain with state space $S = \{0, 1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0.8 & 0 & 0.2 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0.3 & 0.4 & 0 & 0.3 \end{pmatrix}.$$

We have the canonical form of the transition matrix:

$$P = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.3 & 0 & 0.4 & 0.3 \end{pmatrix}. \qquad \blacktriangle$$

Note 9.5 *In the study of finite Markov chains, we are interested in answering questions such as:*

1. *What is the probability that starting from a transient state i, the chain reaches the recurrent state j at least once?*

2. *Given that the chain is in a transient state, what is the mean number of visits to another transient state j before reaching a recurrent state?*

To answer these questions we introduce some results (without proof). The interested reader may refer to Bhat and Miller (2002).

Definition 9.25 (Fundamental Matrix) *The matrix $M = (I - B)^{-1}$, as in (9.4), is called the fundamental matrix.*

Theorem 9.4 *For a finite Markov chain with transition matrix P partitioned as* (9.4), *we have that $(I-B)^{-1}$ exists and is equal to:*

$$(I-B)^{-1} = I + B + B^2 + \cdots = \sum_{r=0}^{\infty} B^r \, .$$

Proof: we have

$$(I-B)(I+B+\cdots+B^{n-1}) = I - B^n \tag{9.5}$$

and since $B^n \underset{n\to\infty}{\longrightarrow} 0$, then:

$$\det(I-B^n) \underset{n\to\infty}{\longrightarrow} 1 \, .$$

Thus for sufficient large n, we have that $\det(I-B^n) \neq 0$ and consequently

$$\det\left((I-B)(I+B+\cdots+B^{n-1})\right) \neq 0$$

so that $\det(I-B) \neq 0$, which implies that $(I-B)^{-1}$ exists.
Multiplying (9.5) by $(I-B)^{-1}$ we obtain

$$I + B + \cdots + B^{n-1} = (I-B)^{-1}(I-B^n)$$

and thus, letting n tend to ∞, we get:

$$\sum_{r=0}^{\infty} B^r = (I-B)^{-1} \, .$$

■

Definition 9.26 *Let i, j be transient states. Suppose that the chain starts from the state i and let N_{ij} be the random variable defined as: N_{ij} = "The number of times the chain visits state j before reaching possibly a recurrent state". We define:*

$$\mu_{ij} = E(N_{ij}) \, .$$

Theorem 9.5 *Let $i, j \in T$ where T denotes the set of transient states. Then the fundamental matrix M is given by:*

$$M = (\mu_{ij})_{i,j\in T} \, .$$

■ EXAMPLE 9.26

Let us consider a Markov chain with states space $S = \{0,1,2,3\}$ and probability transition matrix

$$P = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} .$$

For a Markov chain, starting from state 0, determine the expected number of visits that the chain makes to state 0 before reaching a recurrent class.

Solution: It can be easily seen that the chain classes are given by $C(2) = \{2,3\}$ and $C(0) = \{0,1\}$, the transient states are 0 and 1 and the recurrent states are 2 and 3. Recognizing matrix P (with the order of states 2 3 0 1) we have:

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix} .$$

That is,

$$R = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} , \quad A = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} , \quad B = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix}$$

and:

$$\begin{aligned} M = (I-B)^{-1} &= \left(\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & 0 \end{pmatrix} \right)^{-1} \\ &= \begin{pmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{pmatrix}^{-1} \\ &= \begin{pmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{4}{3} \end{pmatrix} . \end{aligned}$$

Hence, $\mu_{00} = \frac{4}{3}$. ▲

Definition 9.27 *g_{ij} is the probability that starting from the transient state i the chain reaches the recurrent state j at least once. We define:*

$$G = (g_{ij}) .$$

Theorem 9.6 *Let M and A be the matrices defined earlier. Then the matrix G defined above satisfies the relation*

$$G = MA .$$

■ EXAMPLE 9.27

Consider the education process of a student after schooling. The student begins with a bachelor's program and after its completion, proceeds to

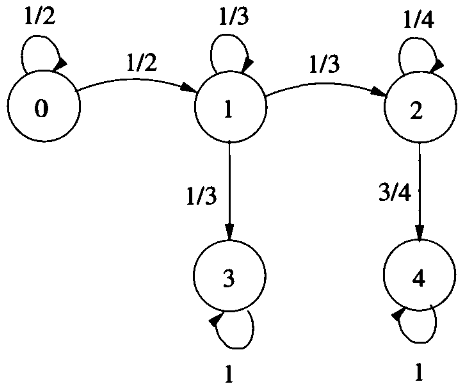


Figure 9.7 State transition diagram for Example 9.27

a master's program. On completion of the bachelor's program, the student can join job A whereas on completion of the master's program, the student joins job B. The education of the student is modeled as a stochastic process $\{X_t; t \geq 0\}$ with state space $S = \{0, 1, 2, 3, 4\}$ where states 0, 1 and 2 represent the stages of educational qualification achieved in chronological order and states 3 and 4 represent job A and job B, respectively. The corresponding state transition diagram is shown in Figure 9.7.

(i) Obtain the expected number of visits before being absorbed in job A or job B.

(ii) Find the probabilities of absorption in job A and job B.

Solution: The education of a student is modeled as a stochastic process $\{X_n; n \geq 0\}$ with state space $S = \{0, 1, 2, 3, 4\}$ where states 0, 1 and 2 represent the stages of educational qualification achieved in chronological order and states 3 and 4 represent job A and job B, respectively. The state transition diagram is shown in Figure 9.7. The states 0,1,2 are transient states and states 3 and 4 are absorbing states.

The probability transition matrix for this model is given by

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & \frac{3}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

where the submatrix I gives the probability of transition between recurrent states 3 and 4, the submatrix $A = \begin{pmatrix} 0 & 0 \\ \frac{1}{3} & 0 \\ 0 & \frac{3}{4} \end{pmatrix}$ gives the probabilities of transition from transient states 0, 1 and 2 to recurrent states 3 and 4 and the submatrix $B = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & \frac{1}{4} \end{pmatrix}$ of transition probabilities between transient states 0, 1 and 2.

(i) From Theorem 9.4, since the probability of transition from i to j in exactly k steps is the (i, j)th component of B^k, it is easy to see that:

$$M = \begin{pmatrix} 2 & \frac{3}{2} & \frac{2}{3} \\ 0 & \frac{3}{2} & \frac{2}{3} \\ 0 & 0 & \frac{4}{3} \end{pmatrix} .$$

(ii) From Theorem 9.6, we get the probability of absorption:

$$G = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{pmatrix} . \qquad \blacktriangle$$

Definition 9.28 *For $i, j \in S$ and $0 \le s \le 1$ we define:*

$$F_{ij}(s) := \sum_{n=0}^{\infty} f_{ij}^{(n)} s^n$$
$$p_{ij}(s) := \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n .$$

Note that $f_{ij} = F_{ij}(1)$.

Definition 9.29

$$p_{ij}^* := \sum_{n=0}^{\infty} p_{ij}^{(n)} = p_{ij}(1),$$

where $i, j \in S$.

The following note provides an interpretation of the p_{ij}^*.

Note 9.6

$$\begin{aligned}
p^*_{ij} &= \sum_{n=0}^{\infty} P\left(X_n = j \mid X_0 = i\right) \\
&= \sum_{n=0}^{\infty} E\left(\mathcal{X}_{\{X_n=j\}} \mid X_0 = i\right) \\
&= E\left(\sum_{n=0}^{\infty} \mathcal{X}_{\{X_n=j\}} \mid X_0 = i\right) \\
&= \text{“\textit{expected number of visits that the chain makes to the state } } j \\
&\quad \text{\textit{starting at} } i\text{”} .
\end{aligned}$$

The theorem presented below, along with its corollary, provides a criterion for determining whether or not a state is recurrent.

Theorem 9.7

$$p_{ij}(s) - \delta_{ij} = F_{ij}(s) p_{jj}(s), \ \ i, j \in S.$$

In particular:

$$p_{ii}(s) = \frac{1}{1 - F_{ii}(s)}, \ \ i \in S.$$

Proof:

$$\begin{aligned}
p^{(n)}_{ij} &= \sum_{k=1}^{n} P(X_n = j, X_k = j, X_l \neq j \text{ for all } 0 \leq l \leq k-1 \mid X_0 = i) \\
&= \sum_{k=1}^{n} P(X_n = j \mid X_k = j) P(X_k = j, X_l \neq j \text{ for all } 0 \leq l \leq k-1 \mid X_0 = i) \\
&= \sum_{k=1}^{n} p^{(n-k)}_{jj} f^{(k)}_{ij} \\
&= \sum_{k=0}^{n} p^{(n-k)}_{jj} f^{(k)}_{ij}, \ \ n \geq 1.
\end{aligned}$$

Then:

$$
\begin{aligned}
p_{ij}(s) - \delta_{ij} &= \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n - \delta_{ij} \\
&= \sum_{n=1}^{\infty} p_{ij}^{(n)} s^n \\
&= \sum_{n=1}^{\infty} \left[\sum_{k=0}^{n} p_{jj}^{(n-k)} f_{ij}^{(k)} \right] s^n \\
&= \sum_{k=0}^{\infty} f_{ij}^{(k)} \sum_{n=k}^{\infty} p_{jj}^{(n-k)} s^n \\
&= \left(\sum_{k=0}^{\infty} f_{ij}^{(k)} s^k \right) \left(\sum_{n=0}^{\infty} p_{jj}^{(n)} s^n \right) \\
&= F_{ij}(s) p_{jj}(s).
\end{aligned}
$$

■

Corollary 9.3 *Let $i \in S$. Then state i is recurrent if and only if $p_{ii}^* = \infty$ or equivalently: i is transient if and only if $p_{ii}^* < \infty$.*

Proof:

$$
\begin{aligned}
p_{ii}^* &= \sum_{n=0}^{\infty} p_{ii}^{(n)} = p_{ii}(1) \\
&= \frac{1}{1 - F_{ii}(1)} = \frac{1}{1 - f_{ii}} = \begin{cases} \infty & \text{if } i \text{ is recurrent} \\ < \infty & \text{if } i \text{ is transient}. \end{cases}
\end{aligned}
$$

■

Definition 9.30 *A property of states is called a solidarity or class property, if whenever the state i has property $i \leftrightarrow j$, then the state j also has the property.*

The following theorem proves that the recurrence, transience and period of a state are class properties.

Theorem 9.8 *Suppose that $i \leftrightarrow j$. Then i is recurrent if and only if j is recurrent.*

Proof: Since $i \leftrightarrow j$, there exist $m, n \geq 0$ such that:

$$
p_{ij}^{(n)} > 0 \quad \text{and} \quad p_{ji}^{(m)} > 0 .
$$

Assuming that i is recurrent, then

$$
p_{jj}^* = \sum_{k=0}^{\infty} p_{jj}^{(k)} \geq \sum_{k=0}^{\infty} p_{jj}^{(n+m+k)} \geq \sum_{k=0}^{\infty} p_{ji}^{(m)} p_{ii}^{(k)} p_{ij}^{(n)},
$$

that is:

$$p^*_{jj} \geq p^{(m)}_{ji} p^{(n)}_{ij} p^*_{ii} \ .$$

Therefore, if $p^*_{ii} = \infty$, then $p^*_{jj} = \infty$. ■

■ EXAMPLE 9.28

Let $\{X_n; n \geq 0\}$ be a random walk in $\mathbb{Z}$ defined as

$$X_0 := 0 \quad \text{and} \quad X_n := \sum_{j=1}^{n} Y_j,$$

where $Y_1, Y_2, \cdots$ are independent and identically distributed random variables with:

$$P(Y_1 = 1) = p \quad \text{and} \quad P(Y_1 = -1) = 1 - p =: q \quad \text{with} \quad 0 < p < 1 \ .$$

It is clear that $(X_n)_{n \geq 0}$ is a Markov chain with state space $S = \mathbb{Z}$. For all $i, j \in S$, we have that $i \leftrightarrow j$. Therefore, $\{X_n; n \geq 0\}$ is an irreducible Markov chain. Now

$$p^{(2n)}_{00} = \binom{2n}{n} p^n q^n \text{ and } p^{(2n+1)}_{00} = 0 \text{ for all } n \geq 1 \ .$$

Stirling's formula (see Feller, 1968 Vol. I, page 52) shows that

$$n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n ,$$

where $a_n \sim b_n$ implies that $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$. Using Stirling's formula in the above expression for $p^{(2n)}_{00}$, we obtain:

$$p^{(2n)}_{00} \sim \frac{1}{\sqrt{\pi n}} (4pq)^n \ .$$

If we have $p \neq q$, then $4pq < 1$, and therefore $p^*_{00} < \infty$, that is, if $p \neq q$, the chain is transient. However, if $p = q = \frac{1}{2}$, then $p^*_{00} = \infty$, that is, if $p = q = \frac{1}{2}$, the chain is recurrent. ▲

The following theorem proves that if the state j is transient, then the expected number of visits by the chain to the state j is finite, regardless of the state from which the chain has started.

Theorem 9.9 *If j is transient, then $p^*_{ij} < \infty$ for $i \in S$.*

Proof: We know that

$$p_{ij}(s) - \delta_{ij} = F_{ij}(s) p_{jj}(s)$$

for all $0 \le s \le 1$. Then

$$p^*_{ij} - \delta_{ij} = f_{ij} p^*_{jj} < \infty$$

since j is transient. Therefore, $p^*_{ij} < \infty$. ∎

From the above theorem, it follows that, if j is transient, then:

$$\lim_{n \to \infty} p^{(n)}_{ij} = 0 .$$

The following question arises naturally: What happens to $\lim_{n\to\infty} p^{(n)}_{ij}$ when state j is recurrent? To answer this question, we require the following concepts

Definition 9.31 *For $i \in S$, define:*

$$\mu_i := E(\tau_i) = \begin{cases} \sum_{n=1}^{\infty} n f^{(n)}_{ii} & \text{if } i \text{ is recurrent} \\ \infty & \text{if } i \text{ is transient} . \end{cases}$$

That is, μ_i represents the expected return time of the chain to state i given that the chain started from state i.

Definition 9.32 *The recurrent state i is said to be positive recurrent if $\mu_i < \infty$ and is said to be null recurrent if $\mu_i = \infty$. It is clear from the definition of $F_{ii}(s)$ that:*

$$\mu_i = \left. \frac{dF_{ii}(s)}{ds} \right|_{s=1} = F'_{ii}(1) .$$

■ EXAMPLE 9.29

Let us consider Example 9.28, with $p = q = \frac{1}{2}$. In this case, it is known that:

$$p^{(2n)}_{00} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} = (-1)^n \binom{-\frac{1}{2}}{n} 2^{2n} \left(\frac{1}{2}\right)^{2n} .$$

Therefore:

$$p_{00}(s) = \sum_{n=0}^{\infty} \binom{-\frac{1}{2}}{n} (-1)^n s^{2n} = (1 - s^2)^{-\frac{1}{2}} .$$

Thus,

$$F_{00}(s) = 1 - \sqrt{1 - s^2}$$

and it follows that 0 is a null recurrent state. ▲

Theorem 9.10 *Let $i \in S$ be a recurrent state with $\lambda(i) = 1$. Then:*

$$\lim_{n \to \infty} p^{(n)}_{ii} = \frac{1}{\mu_i} .$$

Therefore, if i is null recurrent, then $p_{ii}^{(n)} \to 0$ as $n \to \infty$.

Proof: For the proof, the reader is refer to Resnick (1994). ∎

It is to be observed that if $i \in S$ is a null recurrent state with $\lambda(i) = 1$, then:

$$\lim_{n\to\infty} p_{ii}^{(n)} = 0 \ .$$

Theorem 9.11 *If $j \to i$ and j is recurrent, then it satisfies:*

1. $f_{ij} = 1$.
2. $p_{ij}^* = \infty$.

Proof:

1. Suppose that $p_{ji}^{(r)} > 0$, with r being the least positive number with this property. We have:

   $$f_{ij} = P(X_n = j \text{ fo'r some } n \mid X_0 = i).$$

   Therefore:

   $$0 \leq p_{ji}^{(r)}(1 - f_{ij}) \leq P(X_n \neq j \text{ for all } n \mid X_0 = j) = 0.$$

   Which implies:

   $$p_{ji}^{(r)}(1 - f_{ij}) = 0.$$

   Since $p_{ji}^{(r)} > 0$, it follows that $f_{ij} = 1$.

2. $p_{ij}^* = f_{ij}p_{jj}^* = \infty$; then $p_{ij}^* = \infty$.

∎

Theorem 9.12 *Let $\{X_n; n \geq 0\}$ be an irreducible Markov chain whose states are recurrent and aperiodic. Then*

$$\lim_{n\to\infty} p_{ij}^{(n)} = \frac{1}{\mu_j}$$

independent of the initial state i.

Proof: For the proof, the reader is refer to Resnick (1994). ∎

9.2.2 Measure of Stationary Probabilities

To complete the count of significant results in the context of the theory of Markov chains, we present the concept of invariant probability measure on the state space S and give a necessary and sufficient condition for the existence of such a measure.

Definition 9.33 *Let $\{X_n; n \geq 0\}$ be a Markov chain with state space S. A probability measure $\pi = (\pi_h)_{h \in S}$ over S is called invariant or stationary if*

$$\pi_j = \sum_{i \in S} \pi_i p_{ij} \tag{9.6}$$

for all $j \in S$. In other words, π is invariant if $\pi = \pi P$.

Theorem 9.13 *Let $\{X_n; n \geq 0\}$ be an irreducible, aperiodic Markov chain with state space S. Then there exists an invariant probability measure over S if and only if the chain is positive recurrent. The determined probability measure is unique and satisfies the condition* (9.6).

Without loss of generality we assume that $S = \mathbb{N}$.

Proof: Let us assume that there exists an invariant probability measure π over S that satisfies (9.6) since $\pi = \pi P$. Then $\pi = \pi P^n$. That is:

$$\pi_j = \sum_i \pi_i p_{ij}^{(n)} \ .$$

If the states of a Markov chain are transient or null recurrent, then

$$\lim_{n \to \infty} p_{ij}^{(n)} = 0$$

and therefore $\pi_j = 0$ for all j, which is absurd, since $\sum_{j=1}^{\infty} \pi_j = 1$.
Suppose that the chain is positive recurrent and let:

$$\pi_j := \frac{1}{\mu_j} \ j \in S.$$

1. Since the matrix P^n is stochastic:

   $$1 = \sum_{j=0}^{\infty} p_{ij}^{(n)} \geq \sum_{j=0}^{m} p_{ij}^{(n)} \underset{n \to \infty}{\longrightarrow} \sum_{j=0}^{m} \pi_j \text{ for all } m.$$

   Therefore, $\sum_{j=0}^{\infty} \pi_j \leq 1$.

   On the other hand, using the Chapman-Kolmogorov equation:

   $$p_{ij}^{(n+1)} \geq \sum_{k=0}^{m} p_{ik}^{n} p_{kj}.$$

   Then $\pi_j \geq \sum_{k=0}^{m} \pi_k p_{kj}$ for all m. That is:

   $$\sum_{k=0}^{\infty} \pi_k p_{kj} \leq \pi_j \text{ for all } j \in S.$$

Suppose that for some j we have $\sum_{k=0}^{\infty} \pi_k p_{kj} < \pi_j$. Then

$$\sum_{j=0}^{\infty} \pi_j > \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \pi_k p_{kj} = \sum_{k=0}^{\infty} \pi_k \sum_{j=0}^{\infty} p_{kj} = \sum_{k=0}^{\infty} \pi_k,$$

which contradicts the fact that j is recurrent. Therefore, for all j it satisfies:

$$\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}.$$

That is, $\pi P = \pi$. Hence if $\pi P^n = \pi$ and

$$\pi_j = \sum_{k=0}^{\infty} \pi_k p_{kj}^{(n)} \quad \forall \; j \in S, \forall \; n \geq 1 \; ,$$

then for all $\epsilon > 0$ there exists $n_0 \geq 1$ such that:

$$\pi_j = \sum_{k=n_0+1}^{\infty} \pi_k p_{kj}^{(n)} < \epsilon \; .$$

Hence:

$$\pi_j \quad \leq \quad \sum_{k=0}^{n_0} \pi_k p_{kj}^{(n)} + \epsilon \underset{n\to\infty}{\longrightarrow} \sum_{k=0}^{n_0} \pi_k \pi_j + \epsilon \tag{9.7}$$

$$\pi_j \quad \geq \quad \sum_{k=0}^{n_0} \pi_k p_{kj}^{(n)} \underset{n\to\infty}{\longrightarrow} \pi_j \sum_{k=0}^{n_0} \pi_k \; . \tag{9.8}$$

Since $\epsilon > 0$ is arbitrary, (9.7) and (9.8) imply that

$$\pi_j = \pi_j \sum_{k=0}^{\infty} \pi_k,$$

that is, $\sum_{k=0}^{\infty} \pi_k = 1$.

2. If it is assumed that there exists another stationary distribution $(r_k)_{k \in S}$, then it is obtained similarly that:

$$r_j = \sum_{k=0}^{\infty} r_k p_{kj}^{n}.$$

Now taking the limit as $n \to \infty$, we obtain:

$$r_j = \sum_{k=0}^{\infty} r_k \pi_j = \pi_j \sum_{k=0}^{\infty} r_k = \pi_j \quad \forall j \in S.$$

■

■ **EXAMPLE 9.30**

Let $\{X_n; n \geq 0\}$ be a Markov chain with state space $S = \{1, 2, 3\}$ and transition matrix

$$P = \begin{pmatrix} 0 & \frac{3}{4} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix}.$$

$\{X_n; n \geq 0\}$ is an irreducible, aperiodic Markov chain, and also S is finite so that the chain is positive recurrent. Thus, there exists a stationary probability measure $\pi = (\pi_j)_{j \in S}$ over S. Find these stationary probabilities π_j.

Solution: Since $\pi P = \pi$, we obtain from the system:

$$\begin{cases} \pi_1 + \pi_2 + \pi_3 = 1 \\ \pi_1 = \frac{1}{2}\pi_2 + \pi_3 \\ \pi_2 = \frac{3}{4}\pi_1 \\ \pi_3 = \frac{1}{4}\pi_1 + \frac{1}{2}\pi_2. \end{cases}$$

Solving the system, we obtain:

$$\pi_1 = \frac{8}{19}, \; \pi_2 = \frac{6}{19}, \; \pi_3 = \frac{5}{19}.$$

It is known that

$$\lim_{n \to \infty} p_{ij}^{(n)} = \pi_j$$

for $j = 1, 2, 3$ independent of i; this implies, in particular, that the probability that, for some n sufficiently large, the chain is in state 1 given that it started from state i is equal to $\frac{8}{19}$ independent of the initial state i. ▲

9.3 CONTINUOUS-TIME MARKOV CHAINS

In the previous section it was assumed that the time parameter t was discrete. This assumption may be appropriate in some cases, but in situations such as queueing models, the time parameter should be considered as continuous because the process evolves continuously over time. In probability theory, a continuous-time Markov process is a stochastic process $\{X_t; t \geq 0\}$ that satisfies the Markov property. It is the continuous-time adaptation of a Markov chain and hence it is called a continuous-time Markov chain (CTMC). In this section we are going to study the definition and basic properties of the CTMC with some examples.

Definition 9.34 *Let $X = \{X_t; t \geq 0\}$ be a stochastic process with countable state space S. We say that the process is a continuous-time Markov chain if:*

$$P\left(X_{t_n} = j \mid X_{t_1} = i_1, \cdots, X_{t_{n-1}} = i_{n-1}\right) = P\left(X_{t_n} = j \mid X_{t_{n-1}} = i_{n-1}\right)$$

for all $j, i, \cdots, i_{n-1} \in S$ *and for all* $0 \leq t_1 < t_2 < \cdots < t_n$.

For Markov chains with discrete-time parameter we saw that the n-step transition matrix can be expressed in terms of the transition matrix raised to the power of n. In the continuous-time case there is no exact analog of the transition matrix P since there is no implicit unit of time. We will see in this section that there exists a matrix Q called the infinitesimal generator of the Markov chain which plays the role of P.

Definition 9.35 *We say that the continuous-time Markov chain is homogeneous if and only if the probabilities* $P(X_{t+s} = j \mid X_s = i)$ *is independent of* s *for all* t.

Definition 9.36 *The probability*

$$p_{ij}(t) = P(X_{t+s} = j \mid X_s = i)$$

where $s, t \geq 0$ *is called the transition probability for the continuous-time Markov chain.*

Let $p_{ij}(t)$ be the probability of transition from state i to state j in an interval of length t. We denote:

$$P(t) = (p_{ij}(t)) \qquad \text{for all } i, j \in S \ .$$

We say that $P(t)$ is a transition probability matrix. It is easy to verify that it satisfies the following conditions:

1. $p_{ij}(0) = \delta_{ij}$.
2. $\lim_{t \to 0^+} p_{ij}(t) = \delta_{ij}$.
3. For any $t \geq 0$, $i, j \in S$, $0 \leq p_{ij}(t) \leq 1$ and $\sum_{k \in S} p_{ik}(t) = 1$.
4. For all $i, j \in S$, for any $s, t \geq 0$:

$$p_{ij}(t+s) = \sum_{k \in S} p_{ik}(t) . p_{kj}(s) \ . \tag{9.9}$$

The above equation is called a Chapman-Kolmogorov equation for continuous-time Markov chains.

Note 9.7

1. *From part 2 of the above observation we get*

$$\lim_{t \to 0^+} P(t) = I$$

where I *is the identity matrix.*

2. *From part 4 of the above observation we get*

$$P(s+t) = P(s).P(t), \tag{9.10}$$

that is, the family of transition matrices forms a semigroup.

Note 9.8 *The following properties of transition probabilities are extremely important for applications of continuous-time Markov chains. They are outlined here without proof and the reader may refer to Resnick (1994).*

1. $p_{ij}(t)$ *is uniformly continuous on* $[0, \infty)$.

2. *For each* $i \in S$ *we have that*

$$\lim_{t \to 0^+} \frac{1 - p_{ii}(t)}{t} = q_i$$

exists (but may be equal to $+\infty$*).*

3. *For all* $i, j \in S$ *with* $i \neq j$*, we have that the following limit exists:*

$$\lim_{t \to 0^+} \frac{p_{ij}(t)}{t} = q_{ij} < \infty \ .$$

Definition 9.37 *The matrix*

$$Q = \begin{bmatrix} -q_0 & q_{01} & q_{02} & \cdots \\ q_{10} & -q_1 & q_{12} & \cdots \\ q_{20} & q_{21} & -q_2 & \cdots \\ \vdots & \vdots & \vdots & \end{bmatrix}$$

is called the infinitesimal generator of the Markov chain $\{X_t; t \geq 0\}$.

Since $P(0) = I$, we conclude that:

$$Q = P'(0).$$

Note 9.9 *Suppose that* S *is finite or countable. The matrix* $Q = (q_{ij})_{i,j \in S}$ *satisfies the following properties:*

1. $q_{ii} \leq 0$ *for all* i.

2. $q_{ij} \geq 0$ *for all* $i \neq j$.

3. $\sum_j q_{ij} = 0$ *for all* i.

The infinitesimal generator Q of the Markov chain $\{X_t; t \geq 0\}$ plays an essential role in the theory of continuous-time Markov chains as will be shown in what follows.

Definition 9.38

1. *A state $i \in S$ is called an absorbing state if $q_i = 0$.*
2. *If $q_i < \infty$ and $q_i = \sum_{j \neq i} q_{ij}$, then the state i is called stable or regular.*
3. *A state $i \in S$ is called an instantaneous state if $q_i = \infty$.*

Theorem 9.14 *Suppose that $q_i < \infty$ for each $i \in S$. Then the transition probabilities $p_{ij}(t)$ are differentiable for all $t \geq 0$ and $i, j \in S$ and satisfy the following equations:*

1. *(Kolmogorov forward equation)*

$$p'_{ij}(t) = -q_j p_{ij}(t) + \sum_{k \neq j} q_{kj} p_{ik}(t) \ .$$

2. *(Kolmogorov backward equation)*

$$p'_{ij}(t) = -q_i p_{ij}(t) + \sum_{k \neq i} q_{ik} p_{kj}(t)$$

or equivalently:

$$\begin{aligned} P'(t) &= P(t)Q \\ P'(t) &= QP(t) \ . \end{aligned}$$

The initial condition for both the equations is $P(0) = I$.

Proof: *For $h > 0$ and $t \geq 0$ it is satisfied that:*

$$\begin{aligned} \frac{p_{ij}(h+t) - p_{ij}(t)}{h} &= \frac{1}{h}\left(\sum_{k \in S} p_{ik}(h) p_{kj}(t) - p_{ij}(t)\right) \\ &= \frac{1}{h}\left(p_{ii}(h) p_{ij}(t) - p_{ij}(t)\right) + \frac{1}{h}\left(\sum_{\substack{k \in S \\ k \neq i}} p_{ik}(h) p_{kj}(t)\right) \ . \end{aligned}$$

If $h \downarrow 0$, then we have

$$\begin{aligned} \lim_{h \downarrow 0} \frac{p_{ij}(h+t) - p_{ij}(t)}{h} &= -q_i p_{ij}(t) + \sum_{\substack{k \in S \\ k \neq i}} q_{ik} p_{kj}(t) \\ &= \sum_{k \in S} q_{ik} p_{kj}(t) \end{aligned}$$

where $q_{ii} = -q_i$. ■

Formally the solution of the above equations can be cast in the form

$$P(t) = e^{Qt} = I + \sum_{n=1}^{\infty} \frac{t^n Q^n}{n!} .$$

If Q is a finite-dimensional matrix, the above series is convergent and has a unique solution for the system of equations. If Q is infinite dimensional we cannot say anything. Suppose that Q is a finite-dimensional matrix and diagonalizable and let $\beta_0, \beta_1, \cdots, \beta_n$ be the distinct eigenvalues of the matrix Q. Then there exists a matrix A such that

$$Q = A. \begin{pmatrix} \beta_0 & & 0 \\ & \ddots & \\ 0 & & \beta_n \end{pmatrix} .A^{-1}$$

and in this case:

$$P(t) = A \begin{pmatrix} e^{\beta_0 t} & & 0 \\ & \ddots & \\ 0 & & e^{\beta_n t} \end{pmatrix} .A^{-1} .$$

Note 9.10 *For a given matrix Q we can define a stochastic matrix P as follows:*

1. *If $q_i \neq 0$,*

$$p_{ij} := \begin{cases} \frac{q_{ij}}{q_i}, & i \neq j \\ 0, & i = j . \end{cases} \tag{9.11}$$

2. *If $q_i = 0$, then $P_{ij} := \delta_{ij}$.*

■ EXAMPLE 9.31

Let:

$$Q = \begin{pmatrix} -5 & 3 & 2 \\ 1 & -2 & 1 \\ 4 & 0 & -4 \end{pmatrix}.$$

Then:

$$P = \begin{pmatrix} 0 & \frac{3}{5} & \frac{2}{5} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 1 & 0 & 0 \end{pmatrix} .$$

Since

$$q_1 = \sum_{j \neq 1} q_{1j} = 5$$

we have, for example:

$$p_{12} = \frac{q_{12}}{q_1} = \frac{3}{5}. \qquad ▲$$

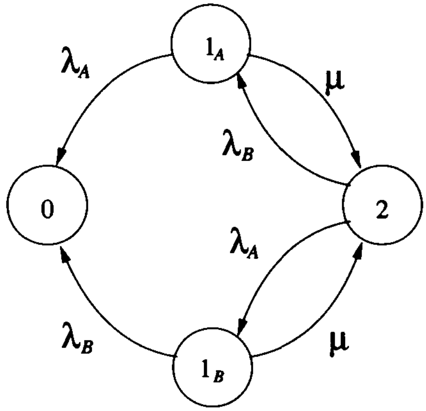


Figure 9.8 State transition diagram for Example 9.32

■ EXAMPLE 9.32

Consider a two-unit system. Unit A has a failure rate λ_A and unit B has failure rate λ_B. There is one repairman and the repair rate of each of the units is μ. When both the machines fail, the system comes to a stop. In this case, $\{X_t; t \geq 0\}$ is a continuous-time Markov chain with state space $S = \{0, 1_A, 1_B, 2\}$ where 0 denotes both the units have failed, 1_A denotes unit A is working and unit B has failed, 1_B denotes unit A has failed and unit B is working and 2 denotes both the units are working. The corresponding infinitesimal generator matrix is given by:

$$Q = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \lambda_A & -(\lambda_A + \mu) & 0 & \mu \\ \lambda_B & 0 & -(\lambda_B + \mu) & \mu \\ 0 & \lambda_B & \lambda_A & -(\lambda_A + \lambda_B) \end{pmatrix}.$$

The state 0 is an absorbing state. The state transition diagram for this Markov chain is shown in Figure 9.8.

Then the transition probability matrix P is:

$$P = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{\lambda_A}{\lambda_A+\mu} & 0 & 0 & \frac{\mu}{\lambda_A+\mu} \\ \frac{\lambda_B}{\lambda_B+\mu} & 0 & 0 & \frac{\mu}{\lambda_B+\mu} \\ 0 & \frac{\lambda_B}{\lambda_A+\lambda_B} & \frac{\lambda_A}{\lambda_A+\lambda_B} & 0 \end{pmatrix}. \qquad \blacktriangle$$

Definition 9.39 *Let $\{X_t; t \geq 0\}$ be a continuous-time Markov chain with transition probability matrix $(P(t))_{t\geq 0}$. A measure μ defined over the state space S is called an invariant measure for $\{X_t; t \geq 0\}$ if and only if, for all*

$t \geq 0$, μ satisfies

$$\mu = \mu P(t)$$

that is, for each $j \in S$, μ satisfies:

$$\mu(j) = \sum_{i \in S} \mu(i) p_{ij}(t).$$

If $\sum_{j \in S} \mu(j) = 1$, then μ is called a stationary distribution.

■ EXAMPLE 9.33

Let $\{X_t; t \geq 0\}$ be a continuous-time Markov chain with state space $S = \{0, 1\}$ and transition matrix given by:

$$P(t) = \begin{pmatrix} \frac{2}{3} + \frac{1}{3}e^{-3t} & \frac{1}{3} - \frac{1}{3}e^{-3t} \\ \frac{2}{3} - \frac{2}{3}e^{-3t} & \frac{1}{3} + \frac{2}{3}e^{-3t} \end{pmatrix}.$$

It is easy to verify that $\mu = \left(\frac{2}{3}, \frac{1}{3}\right)$ is a stationary distribution for $\{X_t; t \geq 0\}$. ▲

Definition 9.40 *Let $\{X_t; t \geq 0\}$ be a continuous-time Markov chain with infinitesimal generator Q and initial probability distribution λ on S.*
The discrete-time Markov chain with initial probability distribution λ and transition probability matrix P [given by (9.11)] is called the embedded Markov chain.

Now we will make use of the *embedded* Markov chain to give conditions that will guarantee the existence and uniqueness of a stationary distribution.

Theorem 9.15 *Let $\{X_t; t \geq 0\}$ be a continuous-time Markov chain with infinitesimal matrix Q and let $\{Y_n; n \in \mathbb{N}\}$ be the corresponding embedded Markov chain. If $\{Y_n; n \in \mathbb{N}\}$ is irreducible and positive recurrent with $\widetilde{\lambda} := inf\{\lambda_i : i \in S\} > 0$, then there exists a unique stationary distribution for $\{X_t; t \geq 0\}$.*

Proof: From the theory developed for discrete-time Markov chains, we have that the transition matrix P_Y of the embedded Markov chain $\{Y_n; n \in \mathbb{N}\}$ has a unique stationary distribution v with $vP_Y = v$. The infinitesimal generator Q satisfies:

$$Q = \Lambda(P_Y - I_S).$$

Since I_s is the identity matrix of order $|S|$ and

$$\Lambda = \text{diag}(\lambda_i : i \in S)$$

then $\mu := v\Lambda^{-1}$ is a stationary measure for the process $\{X_t; t \geq 0\}$. Because $\widetilde{\lambda} > 0$, by hypothesis, μ is finite. Let $\pi = (\pi_j : j \in S)$ be defined by

$$\pi_j := \frac{\mu_j}{\sum\limits_{i \in S} \mu_i}$$

or equivalently:

$$\pi_j := \frac{v_j/\lambda_j}{\sum\limits_{i \in S} v_i/\lambda_i} .$$

Thus we obtain a stationary distribution for $\{X_t; t \geq 0\}$ which is unique since v is unique. In addition, from the relation between $\{Y_n; n \in \mathbb{N}\}$ and $\{X_t; t \geq 0\}$ and from the way the invariant measure $\{Y_n; n \in \mathbb{N}\}$ was constructed, we conclude that:

$$\lim_{t \to \infty} P(X_t = j) = \pi_j, \quad j \in S.$$

■

■ **EXAMPLE 9.34 Birth-and-Death Processes**

A *birth-and-death process (BDP)* is a continuous-time Markov chain $\{X_t; t \geq 0\}$ with state space $S = \mathbb{N}$ such that the elements $q_{i,i-1}, q_{ii}$ and $q_{i,i+1}$ of the intensity matrix Q are the only ones that can be different from zero. Let

$$\lambda_i := q_{i,i+1} \quad \text{and} \quad \mu_i := q_{i,i-1}$$

be the birth and death rates, respectively, as they are known. The matrix Q is given by:

$$Q = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \cdot & \cdot & \end{pmatrix} .$$

It is clear that $\lambda_i h + o(h)$ represents the probability of a birth in the interval of infinitesimal length $(t, t+h)$ given that $X_t = i$. Similarly $\mu_i h + o(h)$ represents the probability of a death in the interval of infinitesimal length $(t, t+h)$ given that $X_t = i$. From the Kolmogorov backward equations, we obtain:

$$\begin{aligned} p'_{0j}(t) &= -\lambda_0 p_{0j}(t) + \lambda_0 p_{1j}(t), \quad j = 0, 1, 2, \cdots \\ p'_{ij}(t) &= -(\lambda_i + \mu_i) p_{ij}(t) + \lambda_i p_{i+1,j}(t) + \mu_i p_{i-1,j}(t) \quad \text{for} \quad i \geq 1. \end{aligned}$$

Similarly, for the forward Kolmogorov equations, we obtain:

$$\begin{aligned} p'_{i0}(t) &= -\lambda_0 p_{i0}(t) + \mu_1 p_{i1}(t), \quad i \in S \\ p'_{ij}(t) &= -(\lambda_j + \mu_j)p_{ij}(t) + \lambda_{j-1}p_{i,j-1}(t) + \mu_{j+1}p_{i,j+1}(t) \text{ for } j \geq 1 \ . \end{aligned}$$

These equations can be solved explicitly for some special cases, as we will show later.

Next we will suppose that the state space S is finite and that $\lambda_i > 0, \mu_i > 0$ for $i \in S$. The embedded Markov chain is irreducible and positive recurrent. Hence there exists a stationary distribution for $\{X_t; t \geq 0\}$, say $\pi = (\pi_0, \pi_1, \cdots, \pi_m)$. π is the solution of the system $\pi Q = 0$, which is given by:

$$\pi_i = \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} \pi_0, \quad i = 1, 2, \cdots, m \ .$$

Also $\sum_{i=0}^{m} \pi_i = 1$. Then:

$$\pi_0 = \frac{1}{1 + \sum_{i=1}^{m} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}} \ . \qquad \blacktriangle$$

■ **EXAMPLE 9.35 Linear Pure Death Process**

In the previous example of a birth-and-death process, suppose that $\lambda_0 = \lambda_1 = \cdots = 0$ and that $\mu_i = i\mu$. Let us assume that the initial size of the population is $N > 0$. Then the backward and forward Kolmogorov equations are respectively

$$\begin{aligned} p'_{ij}(t) &= i\mu\left(p_{i-1,j}(t) - p_{ij}(t)\right), \quad j < i \leq N \\ p'_{ii}(t) &= -i\mu\, p_{ii}(t) \qquad i \leq N \end{aligned}$$

and

$$\begin{aligned} p'_{ij}(t) &= j\mu\left(p_{i-1,j}(t) - p_{ij}(t)\right), \quad j < i \leq N \\ p'_{ii}(t) &= -i\mu\, p_{ii}(t) \qquad i \leq N. \end{aligned}$$

We obtain that

$$p_{ij}(t) = \binom{i}{j} \left(e^{-\mu t}\right)^j \left(1 - e^{-\mu t}\right)^{i-j}, \quad 0 \leq j \leq i \ ,$$

and therefore:

$$P(X_t = k) = p_{Nk}(t) = \binom{N}{k} \left(e^{-\mu t}\right)^k \left(1 - e^{-\mu t}\right)^{N-k}, \quad k = 0, 1, \cdots, N \ .$$

In other words, X_t has a binomial distribution with parameters N and $e^{-\mu t}$. $\blacktriangle$

■ EXAMPLE 9.36 Poisson Process

Suppose that in the birth-and-death process we assume that $\mu_1 = \mu_2 = \cdots = 0$ and $\lambda_i = \lambda$ for all $i \in S$. Further we assume that $X_0 = 0$ and $p_i(t) = P(X_t = i \mid X_0 = 0)$. We obtain both Kolmogorov equations as follows:

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) \\ p_j'(t) &= -\lambda p_j(t) + \lambda p_{j-1}(t) \ . \end{aligned}$$

By solving the above systems of equations, we obtain:

$$P(X_t = n) = e^{-\lambda t}\frac{(\lambda t)^n}{n!}, \quad n = 0, 1, 2, \cdots \ .$$

The random variable X_t has a Poisson distribution with parameter λt. The process $\{X_t; t \geq 0\}$ is called a *Poisson process* with parameter λ. ▲

■ EXAMPLE 9.37

Assume that individuals remain healthy for an exponential time with mean $\frac{1}{\lambda}$ before becoming sick. Assume also that it takes an exponential time to recover from sick to healthy again with mean sick time of $\frac{1}{\mu}$. If the individual starts healthy at time 0, then we are interested in the probabilities of being sick and healthy in future times. Let state 0 denote the healthy state and state 1 denote the sick state.

We have a birth-and-death process with

$$\lambda_0 = \lambda, \text{ and } \mu_1 = \mu$$

and all other λ_i, μ_i are zero.

From the Kolmogorov backward equations, we have:

$$\begin{aligned} p_{00}'(t) &= \lambda p_{10}(t) - \lambda p_{00}(t) \\ p_{10}'(t) &= \mu p_{00}(t) - \mu p_{10}(t) \\ p_{01}'(t) &= \lambda p_{11}(t) - \lambda p_{01}(t) \\ p_{11}'(t) &= \mu p_{01}(t) - \mu p_{11}(t) \ . \end{aligned}$$

It can be shown that:

$$\begin{aligned}
p_{00}(t) &= \frac{\mu}{\mu+\lambda} + \frac{\lambda}{\mu+\lambda} e^{-(\mu+\lambda)t} \\
p_{10}(t) &= \frac{\mu}{\mu+\lambda} \left[1 - e^{-(\mu+\lambda)t}\right] \\
p_{01}(t) &= \frac{\lambda}{\mu+\lambda} \left[1 - e^{-(\mu+\lambda)t}\right] \\
p_{11}(t) &= \frac{\lambda}{\mu+\lambda} + \frac{\mu}{\mu+\lambda} e^{-(\mu+\lambda)t} .
\end{aligned}$$

The stationary probability for the above equation is

$$\pi_0 = \lim_{t\to\infty} p_{00}(t) = \frac{\mu}{\mu+\lambda}$$

and

$$\pi_1 = \lim_{t\to\infty} p_{11}(t) = \frac{\lambda}{\mu+\lambda} . \qquad \blacktriangle$$

9.4 POISSON PROCESS

In Example 9.36 of the previous section, we defined the Poisson process. This process, named after the French mathematician Siméon Denis Poisson (1781–1840), is one of the most widely used mathematical models. This process was also used by the Swedish mathematician Filip Lundberg (1876–1965) in 1903 in his doctoral thesis "Approximerad framställning af sannolikhetsfunktionen/ Återförsäkring af kollektivrisker" (Approximations of the probability function/ Reinsure of collective risks) to determine the ruin probability of an insurance company. Later the Swedish mathematician, actuary, and statistician Harald Cramér (1893–1985) and his pupils developed the ideas of Lundberg and constructed the so-called ruin process or Cramer-Lundberg model, which allows us to describe, at each instant, the reserve of an insurance company. Poisson processes have applications not only in risk theory but also in many other areas, for example, reliability theory and queueing theory.

Note 9.11 *Let $\{N_t; t \geq 0\}$ be a Poisson process with parameter $\lambda > 0$. Then it satisfies the following conditions:*

1. $N_0 = 0$.
2. *It has independent and stationary increments.*
3. *It has unit jumps, i.e.,*

$$\begin{aligned}
P(N_h = 1) &= \lambda h + o(h) \\
P(N_h \geq 2) &= o(h)
\end{aligned}$$

where $N_h := N_{t+h} - N_t$.

The interested reader may refer to Hoel et al. (1972).

■ EXAMPLE 9.38

The examples for the Poisson processes are as follows:

1. The number of particles emitted by a certain radioactive material undergoing radioactive decay during a certain period.
2. The number of telephone calls originated in a given locality during a certain period.
3. The occurrence of accidents at a given road over a certain period.
4. The breakdowns of a machine over a certain period of time. ▲

■ EXAMPLE 9.39

Suppose that accidents in Delhi roads involving Blueline buses obey a Poisson process with 9 accidents per month of 30 days. In a randomly chosen month of 30 days:

1. What is the probability that there are exactly 4 accidents in the first 15 days?
2. Given that exactly 4 accidents occurred in the first 15 days, what is the probability that all 4 occurred in the last 7 days out of these 15 days?

Solution: Let $N_t :=$ "number of accidents in the time interval $(0, t]$" where the time t is measured in days.

Given that $N_t \stackrel{d}{=} \mathcal{P}(\frac{9}{30}t)$, the probability of exactly 4 accidents in the first 15 days is:

$$P(N_{15} = 4) = e^{-4.5}\frac{(4.5)^4}{4!} = 0.1898.$$

Now:

$P(4 \text{ accidents occurred in last 7 days} \mid 4 \text{ accidents occurred in 15 days})$

$$= \frac{P(4 \text{ accidents in 15 days where all 4 are in the last 7 days})}{P(4 \text{ accidents in 15 days})}$$

$$= \frac{P(\text{no accident in the first 8 days and 4 accidents in the next 7 days})}{P(4 \text{ accidents in 15 days})}.$$

Assume:

$$\begin{aligned} P(\text{no accident in 8 days}) &= P(N(8) = 0) &&= e^{-\frac{9}{30}8} = 0.0.0907 \\ P(4 \text{ accidents in 7 days}) &= P(N(7) = 4) &&= \frac{e^{-\frac{9}{30}7}(\frac{9}{30}7)^4}{4!} = 0.0992\ , \end{aligned}$$

the required probability is:

$$\frac{0.3247 \times 0.0315}{0.1898} = 0.0474 . \quad \blacktriangle$$

■ EXAMPLE 9.40

Suppose that incoming calls in a call center arrive according to a Poisson process with intensity of 30 calls per hour. What is the probability that no call is received in a 3-minute period? What is the probability that more than 5 calls are received in a 5-minute interval?

Solution: Let N_t := "number of calls received in the time interval $(0, t]$" where the time t is measured in minutes.

From the data given above, it is known that $N_t \stackrel{d}{=} \mathcal{P}(\frac{1}{2}t)$. Therefore:

$$\begin{aligned} P(N_3 = 0) &= e^{-0.5\times 3} = 0.22313 \\ P(N_5 \geq 6) &= \sum_{k=6}^{\infty} e^{-2.5}\frac{(2.5)^k}{k!} = 0.042. \quad \blacktriangle \end{aligned}$$

Definition 9.41 *Let $\{N_t; t \geq 0\}$ be a Poisson process with parameter $\lambda > 0$. If T_n is the time between the $(n-1)$th and nth event, then $\{T_n;\ n = 1, 2, \cdots\}$ are the interarrival times or holding times of N_t, and $S_n = \sum_{i=1}^{n} T_i$, for $n \geq 1$ is the arrival time of the nth event or the waiting time to the nth event.*

Theorem 9.16 *The T_i's have an exponential distribution with expected value $\frac{1}{\lambda}$.*

Proof: Let T_1 is the time of the first event. Then:

$$P(T_1 > t) = P(N_t = 0) = e^{-\lambda t} .$$

Thus T_1 has an exponential distribution with expected value $\frac{1}{\lambda}$. Now:

$$\begin{aligned} P(T_2 > t | T_1 = s) &= P(0 \text{ events in } (s, s+t] | T_1 = s) \\ &= P(0 \text{ events in } (s, s+t]) \quad \text{(independent increments)} \\ &= P(0 \text{ events in } (0, t]) \quad \text{(stationary increments)} \\ &= e^{-\lambda t} . \end{aligned}$$

Thus T_2 also has an exponential distribution with expected value $\frac{1}{\lambda}$. Note also that T_1 and T_2 are independent and in general we have that the interarrival times T_n, $n = 1, 2, \cdots$, are independent and identically distributed random variables each with expected value $\frac{1}{\lambda}$. ■

Note 9.12 *From the above theorem we have that S_n, the arrival time of the nth event, has a gamma(n, λ) distribution. Therefore the probability distribution function of S_n is given by:*

$$f_{S_n}(t) = \lambda e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!}, \quad t \geq 0 .$$

■ EXAMPLE 9.41

Suppose that people immigrate into a territory at a Poisson rate $\lambda = 1$ per day.

1. What is the expected time until the tenth immigrant arrives?
2. What is the probability that the elapsed time between the tenth and the eleventh arrival exceeds two days?

Solution:
$E(S_{10}) = \frac{10}{\lambda} = 10$ days.
$P(T_{11} > 2) = e^{-2\lambda} = e^{-2} \approx 0.1353$. ▲

Suppose that exactly one event of a Poisson process occurs during the interval $(0, t]$. Then the conditional distribution of T_1 given that $N_t = 1$ is uniformly distributed over the interval $(0, t]$ is:

$$\begin{aligned} P(T_1 \leq s \mid N_t = 1) &= \frac{P(T_1 \leq s, N_t = 1)}{P(N_t = 1)} \\ &= \frac{P(N_s = 1, N_{t-s} = 0)}{P(N_t = 1)} \\ &= \frac{P(N_s = 1)P(N_{t-s} = 0)}{P(N_t = 1)} \\ &= \frac{\lambda s e^{-\lambda s} . e^{-\lambda(t-s)}}{\lambda t e^{-\lambda t}} \\ &= \frac{s}{t} . \end{aligned}$$

Generally:

Theorem 9.17 *Let N_t be a Poisson process with parameter λ. Then the joint conditional density of $T_1, T_2, \cdots, T_n$ given $N_t = n$ is*

$$f_{T_1, T_2, \cdots, T_n | N_t = n}(t_1, t_2, \cdots, t_n) = \begin{cases} \frac{n!}{t^n} & \textit{if } 0 < t_1 < t_2 < \cdots < t_n < t \\ 0 & \textit{other cases} . \end{cases}$$

Proof: Left as an exercise. ■

An application of the previous theorem is as follows: Consider a Poisson process $\{N_t; t \geq 0\}$ with intensity $\lambda > 0$. Suppose that if an event occurs, it is classified as a type-I event with probability $p(s)$, where s is the time at which the event occurs, otherwise it is a type-II event with probability $1-p(s)$. We now prove the following theorem.

Theorem 9.18 *(Ross, 1996) If $N_t^{(i)}$ represents the number of type-i events that occur in the interval $(0,t]$ with $i=1,2$, then $N_t^{(1)}$ and $N_t^{(2)}$ are independent Poisson random variables having parameters λp and $\lambda(1-p)$, respectively, where:*

$$p = \frac{1}{t}\int_0^t p(s)ds \ .$$

Proof: In the interval $(0,t]$, let there be n type-I events and m type-II events, so that we have a total of $n+m$ events in the interval $(0,t]$:

$$\begin{aligned}
P(N_t^{(1)} = n, N_t^{(2)} = m) &= \sum_k P(N_t^{(1)} = n, N_t^{(2)} = m \mid N_t = k)P(N_t = k) \\
&= P(N_t^{(1)} = n, N_t^{(2)} = m \mid N_t = n+m)P(N_t = n+m) \\
&= P(N_t^{(1)} = n, N_t^{(2)} = m \mid N_t = n+m)e^{-\lambda t}\frac{(\lambda t)^{n+m}}{(n+m)!} .
\end{aligned}$$

We consider an arbitrary event that occurred in the interval $(0,t]$. If it has occurred at time s, the probability that it would be a type-I event is $p(s)$. In other words, we know that the time at which this event occurs is uniformly distributed in the interval $(0,t]$. Therefore, the probability that it will be a type-I event is

$$\begin{aligned}
p &:= P(\text{“type-1”}) \\
&= E(P(\text{“type-1”} \mid T = s) \\
&= \int_0^t p(s)\frac{1}{t}ds
\end{aligned}$$

independently of other events. Hence $P(N_t^{(1)} = n, N_t^{(2)} = m \mid N_t = n+m)$ is the probability of n successes and m failures in a Bernoulli experiment of $n+m$ trails with probability of success p in each trial. We get

$$\begin{aligned}
P(N_t^{(1)} = n, N_t^{(2)} = m) &= \binom{n+m}{n} p^n q^m e^{-\lambda t}\frac{(\lambda t)^{n+m}}{(n+m)!} \\
&= e^{-\lambda p t}\frac{(\lambda p t)^n}{n!} e^{-\lambda(1-p)t}\frac{(\lambda(1-p)t)^m}{m!} .
\end{aligned}$$

■

■ EXAMPLE 9.42

If immigrants arrive to area A at a Poisson rate of ten per week and each immigrant is of English descent with probability $\frac{1}{12}$, then what is the probability that no person of English descent will emigrate to area A during the month of February?

Solution: By the previous proposition it follows that the number of Englishmen emigrating to area A during the month of February is Poisson distributed with mean $4 \cdot 10 \cdot \frac{1}{12} = \frac{10}{3}$. Hence the desired probability is $e^{\frac{-10}{3}}$. ▲

In the following algorithm, we simulate the sample path for the Poisson process using interarrival times which are i.i.d. exponentially distributed random variables.

Algorithm 9.1

Input: λ, T, where T is the maximum time unit.

Output: $PP(k)$ for $k = 0(1)T$.

Initialization:

$PP(0) := 0.$

Iteration:

For $k = 0(1)T - 1$ do:
$U(k+1) = rand(0,1)$
$PP(k+1) = PP(k) - \frac{1}{\lambda} \times \log(1 - U(k+1))$.

where $rand(0,1)$ is the uniform random number generated in the interval $(0,1)$. Using the previous algorithm we obtain the sample path of Poisson process as shown in Figure 9.9 for $\lambda = 2.5$ and $T = 10$.

In the Poisson process, we assume that the intensity λ is a constant. If we assume a time-dependent intensity, that is, $\lambda = \lambda(t)$, we get a nonhomogeneous Poisson process. More precisely, we have the following definition:

Definition 9.42 (Nonhomogeneous Poisson Process) *A Markov process $\{N_t; t \geq 0\}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t)$, $t \geq 0$, if:*

1. $N_0 = 0$.

2. *$\{N_t; t \geq 0\}$ has independent increments.*

3. *For $0 \leq s < t$, the random variable $N_t - N_s$ has a Poisson distribution with parameter $\int_s^t \lambda(u)du$. That is,*

$$P(N_t - N_s = k) = \frac{\left(\int_s^t \lambda(u)du\right)^k}{k!} e^{-\int_s^t \lambda(u)du}$$

for $k = 0, 1, 2, \cdots$.

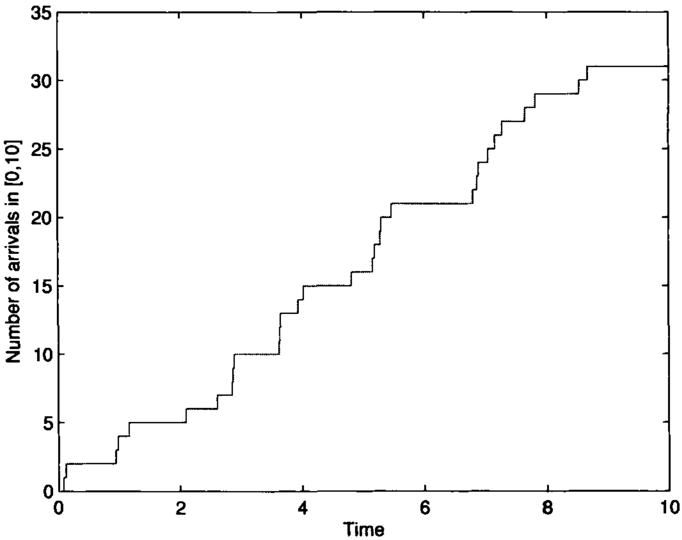


Figure 9.9 Sample path of Poisson process

Definition 9.43 *The function*

$$m(t) = \int_0^t \lambda(u)du$$

is called the mean value function

Note 9.13 *If $\{N_t; t \geq 0\}$ is a nonhomogeneous Poisson process with mean value function $m(t)$, then $\{N_{m^{-1}(t)}; t \geq 0\}$ is a homogeneous Poisson process with intensity $\lambda = 1$. This follows since N_t is a Poisson random variable with mean $m(t)$, and if we let $X_t = N_{m^{-1}(t)}$, then X_t is Poisson with mean $m(m^{-1}(t)) = t$.*

■ EXAMPLE 9.43

(Ross, 2007) John owns a cafeteria which is open from 8 AM. From 8 AM to 11 AM, the arrival rate of customers grows linearly from 5 customers per hour at 8 AM to 20 customers per hour at 11 AM. From 11 AM to 1 PM the arrival rate of customers is 20 customers per hour. From 1 PM to 5 PM the arrival rate of customers decreases linearly until it reaches a value of 12 customers per hour. If we assume that customers arrive at the cafeteria in nonoverlapping time intervals and are independent of each other, then:

1. What is the probability that no customer arrives at the cafeteria between 8:30 AM and 9:30 AM?

2. What is the expected number of customers in the same period of time?

Solution: Let N_t ="Number of customers arriving to the cafeteria in the time interval $(0, t]$".

An adequate model for this situation is the nonhomogeneous Poisson process with intensity function $\lambda(t)$ given by

$$\lambda(t) = \begin{cases} 5+5t & \text{if } \ 0 \le t \le 3 \\ 20 & \text{if } \ 3 \le t \le 5 \\ 20-2(t-5) & \text{if } \ 5 \le t \le 9 \end{cases}$$

and $\lambda(t) = \lambda(9-t)$ for $t > 9$. Since $\left(N_{\frac{3}{2}} - N_{\frac{1}{2}}\right) \stackrel{d}{=} P\left(m\left(\frac{3}{2}\right) - m\left(\frac{1}{2}\right)\right)$ where $m(t) = \int_0^t \lambda(s)\, ds$, then:

$$P\left(N_{\frac{3}{2}} - N_{\frac{1}{2}} = 0\right) = e^{-10}.$$

The expected number of clients in this period of time is 10. ▲

Definition 9.44 (Compound Poisson Process) *A stochastic process* $\{X_t; t \ge 0\}$ *is a compound Poisson process if*

$$X_t = \sum_{i=1}^{N_t} Y_i, \quad t \ge 0\ , \tag{9.12}$$

where $\{N_t; t \ge 0\}$ *is a Poisson process and* $\{Y_i;\ \ i = 1, 2, \cdots\}$ *are independent and identically distributed random variables.*

Note 9.14

1. If $Y_t = 1$, *then* $X_t = N_t$ *is a Poisson process.*

2.

$$\begin{aligned} E(X_t) &= E\left(E(X_t|N_t)\right) \\ &= E\left(E\left(\sum_{i=1}^{N_t} Y_i|N_t\right)\right) \\ &= E\left(N_t E(Y_i)\right) \\ &= \lambda t E(Y_i)\ . \end{aligned}$$

3.

$$\begin{aligned} Var(X_t) &= E\left(Var(X_t|N_t)\right) + Var\left(E(X_t|N_t)\right) \\ &= E\left(N_t Var(Y_i)\right) + Var\left(N_t E(Y_i)\right) \\ &= \lambda t Var(Y_i) + E(Y_i)^2 \lambda t \\ &= \lambda t \left(Var(Y_i) + E(Y_i)^2\right) \\ &= \lambda t \left(E(Y_i^2)\right) \ . \end{aligned}$$

■ **EXAMPLE 9.44**

In life insurance, total claims are often modeled using a compound Poisson distribution. Claim numbers are usually assumed to occur according to a Poisson process and claim amounts have an appropriate density such as a log-normal or gamma. ▲

■ **EXAMPLE 9.45**

Suppose that buses arrive at a sporting event in accordance with a Poisson process, and suppose that the numbers of customers in each bus are assumed to be independent and identically distributed. Then $\{X_t; t \geq 0\}$ is a compound Poisson process where X_t denotes the number of customers who have arrived by t. In equation (9.12), Y_i represents the number of customers in the ith bus. ▲

■ **EXAMPLE 9.46**

Suppose customers leave a supermarket in accordance with a Poisson process. If Y_i, the amount spent by the ith customer, $i = 1, 2, \cdots$, are independent and identically distributed, then $\{X_t : t \geq 0\}$ is a compound Poisson process where X_t denotes the total amount of money spent by time t. ▲

9.5 RENEWAL PROCESSES

The stochastic process $\{N_t; t \geq 0\}$ is called a counting process if N_t represents the number of events that have occurred up to time t.

In the previous section we dealt with the Poisson process, which is a counting process for which the periods of time between occurrences are i.i.d. random variables with exponential distribution. A possible generalization is to consider a counting process for which the periods of time between occurrences are i.i.d. random variables with arbitrary distribution. Such counting processes are known under the name *renewal processes.*

We have the following formal definition:

Definition 9.45 *Let $\{T_n; n \geq 1\}$ be an i.i.d sequence of nonnegative random variables with common distribution F, where $F(0) = P(T_n = 0) < 1$.*
The process $\{S_n; n \geq 1\}$ given by

$$\begin{aligned} S_0 &: = 0 \\ S_n &: = T_1 + T_2 + \cdots + T_n \end{aligned}$$

is called a renewal process with duration or length-of-life distribution F.

If we interpret T_n as the period of time between the $(n-1)$th and the nth occurrence of a certain event, then the random variable S_n represents the nth holding time. Therefore, if we define the process $\{N_t : t \geq 0\}$ as

$$N_t := \sup\{n : S_n \leq t\}$$

then it is clear that N_t represents the number of events that have occurred (or renewals) up to time t.

From now on we will call, indistinctly, the processes $\{S_n; n \geq 1\}$ and $\{N_t; t \geq 0\}$ as renewal processes.

■ EXAMPLE 9.47

The well-known example of renewal processes is the repeated replacement of light bulbs. It is supposed that as soon as a light bulb burns out, it is instantaneously replaced by a new one. We assume that successively replaced bulbs are random variables having the same distribution function F. Let $S_n = T_1 + \cdots + T_n$, where T_i's are the random life of the bulb with distribution function F. We have a renewal process $\{N_t; t \geq 0\}$ where N_t represents the number of bulbs replaced up to time t. ▲

In the next theorem, we give a relation between the distributions of N_t and S_n.

Theorem 9.19 *For $k \in \mathbb{N}$ and $t \geq 0$ we have that:*

$$\{N_t \geq k\} \text{ if and only if } \{S_k \leq t\} .$$

Proof: It is clear that $N_t \geq k$ if and only if in the time interval $[0, t]$ at least k renewals have occurred, which in turn occurs if and only if the kth renewal has occurred on or before t, that is, if $(S_k \leq t)$. ■

From the previous result, it follows for $t \geq 0$ and $k = 1, 2, \cdots$ that

$$\begin{aligned} P(N_t = k) &= P(N_t \geq k) - P(N_t \geq k+1) \\ &= P(S_k \leq t) - P(S_{k+1} \leq t) \\ &= F_k(t) - F_{k+1}(t) \end{aligned}$$

where F_k denotes the kth convolution of F with itself.

Similarly, the expected number of renewals up to time t follows from the next expression:

$$\begin{aligned} m(t) &:= E(N_t) = \sum_{k=1}^{\infty} kP(N_t = k) \\ &= \sum_{k=1}^{\infty} P(N_t \geq k) = \sum_{k=1}^{\infty} P(S_k \leq t) = \sum_{k=1}^{\infty} F_k(t) . \end{aligned}$$

The function $m(t)$ is called the *mean value function or renewal function.*

■ EXAMPLE 9.48

Consider $\{T_n; n = 1, 2, \cdots\}$ with $P(T_n = i) = p(1-p)^{i-1}$. Here, T_n denotes the number of trials until the first success. Then S_n is the number of trials until the nth success:

$$P(S_n = k) = \binom{k-1}{n-1} p^n (1-p)^{k-n}, \quad k \geq n .$$

The renewal process $\{N_t; t \geq 0\}$, expressing the number of successes, is given by:

$$P(N_t = n) = \sum_{k=n}^{[t]} P(S_n = k) - \sum_{k=n+1}^{[t]} P(S_{n+1} = k) .$$

After simplification, we have

$$P(N_t = n) = \binom{[t]}{n-1} p^n (1-p)^{[t]-n}$$

where $[t]$ denotes the greatest integer function of t. ▲

■ EXAMPLE 9.49

Consider a renewal process $\{N_t; t \geq 0\}$ with interarrival time distribution F. Suppose that the interarrival time has a uniform distribution between 0 and 1. Then the n-fold convolution of F is given by

$$\begin{aligned} F_n(t) &= P(S_n \leq t) \\ &= \int_0^t P(t_n \leq t - x) \frac{x^{n-2}}{(n-2)!} dx \\ &= \frac{t^n}{n!}, \quad 0 \leq t \leq 1 . \end{aligned}$$

The renewal function for $0 \leq t \leq 1$ is given by:

$$m(t) = \sum_{k=1}^{\infty} \frac{t^n}{n!} = e^t - 1 \ . \qquad \blacktriangle$$

It can be shown that the renewal function $m(t)$, $0 \leq t < \infty$, uniquely determine the interarrival time distribution F.

In order to describe more accurately a renewal process we will introduce the following concepts:

Definition 9.46 *For $t > 0$ define:*

1. *The age:*

$$\delta_t := t - S_{N_t}.$$

2. *Residual lifetime or overshoot:*

$$\gamma_t := S_{N_t+1} - t.$$

3. *Total lifetime:*

$$\beta_t := \delta_t + \gamma_t.$$

For instance, consider the case of an individual who arrives, at time t, at a train station. Trains arrive at the station according to a renewal process $\{N_t; t \geq 0\}$. In this case γ_t represents the waiting time of the individual at the station until the arrival of the train, δ_t the amount of time that has elapsed since the arrival of the last train and the arrival of the individual at the station and β_t the total time elapsed between the train that the individual could not take and the arrival of the train that he is expected to take.

Definition 9.47 (Renewal Equation) *Suppose $\{N_t; t \geq 0\}$ is a renewal process and T_1 denotes the first renewal time. If T_1 has density function f, then:*

$$\begin{aligned} m(t) &= E(N_t) \\ &= E(E(N_t \mid T_1)) \\ &= \int_0^t [1 + E(N_{t-y})] f(y)\, dy \ . \end{aligned}$$

That is,

$$m(t) = F(t) + \int_0^t m(t-y) f(y)\, dy$$

where F denotes the distribution function of T_1. The above equation is known as the renewal equation. In the discrete case, the renewal equation is:

$$m(t) = F(t) + \sum_{k=0}^{[t]} m(t-k) P(T_1 = k) \ .$$

■ **EXAMPLE 9.50**

Consider the residual lifetime γ_t at time t. Let $g(t) = E(\gamma(t))$. Using the renewal equation, we can find $g(t)$. Conditioning on T_1 gives:

$$g(t) = \int_0^\infty E\left(\gamma(t) \mid T_1 = x\right) dF(x) \ .$$

Now,

$$E\left(\gamma(t) \mid T_1 = x > t\right) = x - t$$

while for $T_1 = x < t$:

$$E\left(\gamma(t) \mid T_1 = x < t\right) = g(t - x) \ .$$

Hence,

$$g(t) = \int_t^\infty (x - t) dF(x) + \int_0^t g(t - x) dF(x)$$

is the renewal equation for $g(t)$. ▲

■ **EXAMPLE 9.51**

Consider a maintained system in which repair restores the system to as good as new condition and repair times are negligible in comparison to operating times. Let $T_i, i = 1, 2, \cdots$, represent the successive lifetimes of a system that, upon failure, is replaced by a new one or is overhauled to as-new condition. Let $N(t)$ be the number of failures that have occurred up to time t, i.e., $N_t = \max\{n : T_1 + T_2 + \cdots + T_n \leq t\}$.

The renewal function $m(t) = E(N_t)$ is the expected number of failures that have occurred by time t. It satisfies the equation

$$m(t) = F(t) + \int_0^t m(t - u) dF(u).$$

Suppose $T_i, i = 1, 2, \cdots$, follows exponential distribution with parameter λ so that $F(t) = 1 - e^{-\lambda t}, t \geq 0$. Then the renewal process $\{N_t; t \geq 0\}$ is a Poisson process with

$$P(N_t = k) = \frac{(\lambda t)^k}{k!} e^{-\lambda t}, \quad k = 0, 1, \cdots$$

and $m(t) = \lambda t$. ▲

Next we will discuss the asymptotic behavior of a renewal process. In the first place, we will see that, with probability 1, the total number of renewals, as the time tends to infinity, is infinity. As a matter of fact:

Theorem 9.20 *For each renewal process* $\{N_t : t \geq 0\}$ *we have that:*

$$P\left(\lim_{t\to\infty} N_t = \infty\right) = 1 \ .$$

Proof: *For all* $n \geq 1$, *we have*

$$\begin{aligned} 1 &= \lim_{t\to\infty} P(S_n \leq t) \\ &= \lim_{t\to\infty} P(N_t \geq n) \\ &= P\left(\lim_{t\to\infty} N_t \geq n\right) \end{aligned}$$

so that:

$$P\left(\lim_{t\to\infty} N_t = \infty\right) = 1 \ .$$

■

Theorem 9.21 *With probability* 1, *we have*

$$\frac{N_t}{t} \longrightarrow \frac{1}{\mu} \quad \textit{when } t \longrightarrow \infty$$

where $\mu := E(T_1)$.

Proof: Since for any $t > 0$

$$N_t = n \iff S_n \leq t < S_{n+1}$$

then:

$$\frac{S_{N_t}}{N_t} \leq \frac{t}{N_t} < \frac{S_{N_t+1}}{N_t+1}\frac{N_t+1}{N_t} \ .$$

By using the strong law of large numbers, we have that:

$$\lim_{t\to\infty} \frac{S_{N_t}}{N_t} = \mu \ .$$

Then:

$$\lim_{t\to\infty} \frac{t}{N_t} = \mu \ .$$

■

Our next goal is to prove that the expected number of renewals per unit time tends to μ when $t \to \infty$. We could think that this result is a consequence of the previous theorem. Unfortunately that is not true, since if $Y, Y_1, Y_2, \cdots$ are random variables such that $Y_n \to Y$ with probability 1, we do not necessarily have that $E(Y_n) \to E(Y)$.

Theorem 9.22 (The Elementary Renewal Theorem) *Let* $\{N_t; t \geq 0\}$ *be a renewal process. Then:*

$$\frac{m(t)}{t} = \frac{E(N_t)}{t} \longrightarrow \frac{1}{\mu} \quad \textit{when } t \longrightarrow \infty \ .$$

Proof: We have that:

$$\begin{aligned} E\left(S_{N_t+1}\right) &= E\left(N_t+1\right)E\left(T_1\right) \\ &= \left(m\left(t\right)+1\right)\mu \ . \end{aligned}$$

It follows that:

$$\frac{m\left(t\right)}{t} = \frac{1}{\mu t}E\left(S_{N_t+1}-t\right)+\frac{1}{\mu}-\frac{1}{t} \ .$$

Because $S_{N_t+1} > t$, we have $\left(m\left(t\right)+1\right)\mu > E\left(t\right) = t$, and we obtain that:

$$\liminf_{t\to\infty}\frac{m\left(t\right)}{t} \geq \frac{1}{\mu} \ .$$

Since

$$S_{N_t+1}-t \leq S_{N_t+1}-S_{N_t} = T_{N_t+1} \ ,$$

it follows that

$$\frac{m\left(t\right)}{t} \leq \frac{1}{\mu}+\frac{1}{\mu t}E\left(T_{N_t+1}\right) \ .$$

If the random variables $T_1, T_2, \cdots$ are uniformly bounded, then there exists a number $M > 0$ such that

$$P\left(T_n \leq M\right) = 1 \ .$$

Thus $E\left(T_{N_t+1}\right) \leq M$ and we obtain

$$\limsup_{t\to\infty}\frac{m\left(t\right)}{t} \leq \frac{1}{\mu}$$

completing the proof in this case.

If the variables are not bounded, we can consider a fixed number K and construct the *truncated* process $\left(\overline{T_n}\right)_{n\geq 1}$ given by

$$\overline{T_n} := \begin{cases} T_n & \text{if} \quad T_n \leq K \\ K & \text{if} \quad T_n > K \end{cases}$$

for $n = 1, 2, \cdots$. It is clear that $\overline{T_n} \nearrow T_n$ as $K \to \infty$. For the renewal process $\{\overline{N_t}; t \geq 0\}$ is determined by the process $\left(\overline{T_n}\right)_{n\geq 1}$ and consequently we have:

$$\limsup_{t\to\infty}\frac{\overline{m}\left(t\right)}{t} \leq \frac{1}{E\left(\overline{T_1}\right)} \ .$$

Since

$$m\left(t\right) \leq \overline{m}\left(t\right)$$

it follows that:

$$\limsup_{t\to\infty}\frac{m\left(t\right)}{t} \leq \frac{1}{E\left(\overline{T_1}\right)} \ .$$

Allowing $K \to \infty$ and using the monotone convergence theorem, it follows that:

$$E\left(\overline{T_n}\right) \nearrow \mu .$$

■

Next, we will show that N_t has, asymptotically, a Gaussian distribution.

Theorem 9.23 *Suppose $\mu = E(T_1)$ and $\sigma^2 = Var(T_1)$ are finite. Then:*

$$\lim_{t\to\infty} P\left(\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} \le y\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left(-\frac{x^2}{2}\right) dx .$$

Proof: Fix x and let $n \to \infty$ and $t \to \infty$ so that:

$$-x = \frac{t - n\mu}{\sigma\sqrt{n}} .$$

Observing that $T_1, T_2, \cdots$ are i.i.d. random variables, the central limit theorem implies that:

$$\begin{aligned}
\lim_{n\to\infty} P(S_n > t) &= \lim_{n\to\infty} P\left(S_n > n\mu - x\sigma\sqrt{n}\right) \\
&= \lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} > -x\right) \\
&= \frac{1}{\sqrt{2\pi}}\int_{-x}^{\infty} \exp\left(-\frac{u^2}{2}\right) du \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \exp\left(-\frac{u^2}{2}\right) du .
\end{aligned}$$

Since

$$\sqrt{n} = \frac{x\sigma + \sqrt{(x\sigma)^2 + 4t\mu}}{2\mu}$$

we have

$$\frac{n - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} = x\sqrt{\frac{n\mu}{t}} = x\left(1 + \frac{x\sigma\sqrt{n}}{t}\right)^{\frac{1}{2}} \longrightarrow x$$

when $t \to \infty$. Therefore

$$P(N_t < n) = P\left(\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} < \frac{n - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}}\right) \longrightarrow P\left(\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} < x\right)$$

and it follows that:

$$\lim_{t\to\infty} P\left(\frac{N_t - \frac{t}{\mu}}{\sigma\sqrt{\frac{t}{\mu^3}}} \le y\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left(-\frac{x^2}{2}\right) dx .$$

■

Definition 9.48 *A nonnegative random variable X is said to be a lattice if there exists $d \geq 0$ such that:*

$$\sum_{n=0}^{\infty} P(X = nd) = 1 .$$

That is, X is a lattice if it only takes on integral multiples of some nonnegative number d. The largest d having this property is called the period of X. If X is a lattice and F is the distribution function of X, then we say that F is a lattice.

■ **EXAMPLE 9.52**

(*a*) If $P(X = 2) = P(X = 4) = \frac{1}{2}$, then X is a lattice with period 2.

(*b*) If $P(X = 4\pi) = P(X = 6\pi) = \frac{1}{2}$, then X is a lattice with period 2π.

(*c*) If $P\left(X = \sqrt{2}\right) = P\left(X = \sqrt{5}\right) = \frac{1}{2}$, then X is not a lattice.

(*d*) If $X \stackrel{d}{=} \exp(\lambda)$, then X is not a lattice. ▲

We will now state a result known as Blackwell's theorem without proof. If F is not a lattice, then the expected number of renewals in an interval of length a from the origin is approximately $\frac{a}{\mu}$. In the case of F being a lattice with period d, the Blackwell theorem asserts that the expected number of renewals up to nd tends to $\frac{d}{\mu}$ as $n \to \infty$.

Theorem 9.24 *Let $\{N_t; t \geq 0\}$ be a renewal process having renewal function $m(t)$.*

(*a*) *If F is not a lattice, then*

$$\lim_{t \to \infty} (m(t+a) - m(t)) = \frac{a}{\mu}$$

for $a > 0$ fixed.

(*b*) *If F is a lattice with period d, then*

$$\lim_{n \to \infty} E(R_n) = \frac{d}{\mu}$$

where R_n is a random variable denoting the number of renewals up to nd.

The Blackwell theorem is equivalent to the following theorem, known as the key renewal theorem (Smith, 1953), and will be stated without proof.

Theorem 9.25 *Let F be the distribution function of a positive random variable with mean μ. Suppose that h is a function defined on $[0,\infty)$ such that:*

(1) $h(t) \geq 0$ for all $t \geq 0$.

(2) $h(t)$ is nonincreasing.

(3) $\int_0^\infty h(t)\,dt < \infty$

and let Z be the solution of the renewal equation

$$Z(t) = h(t) + \int_0^t Z(t-x)\,dF(x) \ .$$

Then:

(a) If F is not a lattice:

$$\lim_{t\to\infty} Z(t) = \begin{cases} \dfrac{1}{\mu}\displaystyle\int_0^\infty h(x)\,dx & \text{if} \quad \mu < \infty \\ 0 & \text{if} \quad \mu = \infty \ . \end{cases}$$

(b) If F is a lattice with period d, then for $0 \leq c < d$ we have that:

$$\lim_{n\to\infty} Z(c+nd) = \begin{cases} \dfrac{1}{\mu}\displaystyle\sum_{n=0}^{\infty} h(c+nd) & \text{if} \quad \mu < \infty \\ 0 & \text{if} \quad \mu = \infty \ . \end{cases}$$

Next we are going to use the key renewal theorem to find the distributions of the random variables $\delta_t := t - S_{N_t}$ (age) and β_t (total lifetime).

Theorem 9.26 *For $t \in (x,\infty)$ we have the function $z(t) = P(\delta_t > x)$ satisfying the equation*

$$z(t) = 1 - F(t+x) + \int_0^t z(t-u)\,dF(u)$$

so that, if $\mu < \infty$, then:

$$\lim_{t\to\infty} z(t) = \frac{1}{\mu}\int_x^\infty (1 - F(y))\,dy \qquad \text{for} \ \ x > 0 \ .$$

Proof: Conditioning on the value of T_1, we have for fixed $x > 0$:

$$P(\delta_t > x) = E\left(E\left(\mathcal{X}_{\{\delta_t > x\}} \mid T_1\right)\right) \ .$$

Since

$$\begin{aligned} E\left(\mathcal{X}_{\{\delta_t>x\}} \mid T_1=u\right) &= P\left(\delta_t>x \mid T_1=u\right) \\ &= \begin{cases} 1 & \text{if } u>t+x \\ 0 & \text{if } t<u\leq t+x \\ P\left(\delta_{t-u}>x\right) & \text{if } 0<u\leq t \end{cases} \end{aligned}$$

we write:

$$\begin{aligned} P\left(\delta_t>x\right) &= \int_0^\infty P\left(\delta_t>x \mid T_1=u\right) dF(u) \\ &= 1-F(t+x)+\int_0^t P\left(\delta_{t-u}>x\right) dF(u) \ . \end{aligned}$$

Thus

$$z(t)=1-F(t+x)+\int_0^t z(t-u)\,dF(u)$$

so that the function $h(t) := 1-F(t+x)$ satisfies the key renewal theorem conditions, and as a consequence:

$$\lim_{t\to\infty} z(t)=\frac{1}{\mu}\int_0^\infty (1-F(u+x))\,du=\frac{1}{\mu}\int_x^\infty (1-F(y))\,dy \ .$$

■

Theorem 9.27 *The function $g(t)=P(\beta_t>x)$ satisfies the renewal equation*

$$g(t)=1-F(t\vee x)+\int_0^t g(t-u)\,dF(u)$$

where $t\vee x := \max(t,x)$. *Consequently,*

$$\lim_{t\to\infty} g(t)=\frac{1}{\mu}\int_x^\infty y\,dF(y)$$

if $\mu<\infty$.

Proof: Since

$$g(t)=E\left(E\left(\mathcal{X}_{\{\beta_t>x\}} \mid T_1\right)\right)$$

and

$$E\left(\mathcal{X}_{\{\beta_t>x\}} \mid T_1=u\right)=\begin{cases} 1 & \text{if } u>x\vee t \\ g(t-u) & \text{if } u\leq t \\ 0 & \text{otherwise} \end{cases}$$

we have that:

$$g(t)=1-F(t\vee x)+\int_0^t g(t-u)\,dF(u) \ .$$

Applying the key renewal theorem with $h(t) = 1 - F(t \vee x)$ we obtain:

$$\lim_{t \to \infty} g(t) = \frac{1}{\mu} \int_0^\infty [1 - F(\tau \vee x)] \, d\tau = \frac{1}{\mu} \int_x^\infty y dF(y) \; .$$

■

9.6 SEMI-MARKOV PROCESS

In this section, we give a brief account of the semi-Markov process (SMP) and the Markov regenerative process (MRGP). Knowledge of these concepts are required for building queueing models, introduced in the next chapter.

Definition 9.49 (Semi-Markov Process) *Let $\{Y_n; n \in \mathbb{N}\}$ be a stochastic process with state space $S = \mathbb{N}$ and $\{V_n; n \in \mathbb{N}\}$ a sequence of nonnegative random variables. Let*

$$\begin{aligned} U_0 &:= 0 \\ U_1 &:= V_1 \\ &\vdots \\ U_n &:= \sum_{i=1}^n V_i \quad for \;\; n \geq 2 \end{aligned}$$

and

$$U(t) := max\{n \geq 1 : U_n \leq t\}, \quad t \geq 0 \; .$$

The continuous-time stochastic process $\{X_t; t \geq 0\}$ defined by

$$X_t := Y_{U(t)}, \quad t \geq 0 \; ,$$

is called a semi-Markov process if the following properties hold:

(a) For all $n \geq 0$:

$$\begin{aligned} &P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i_n, \cdots, Y_0 = i_0, V_n \leq t_n, \cdots, V_1 \leq t_1) \\ &\quad = P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i_n) \quad (Markov\ property) \; . \end{aligned}$$

(b)

$$P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i) = P(Y_1 = j, V_1 \leq t \mid Y_0 = i) \quad (time\ homogeneity) \; .$$

The semi-Markov process $\{X_t; t \geq 0\}$ is also known as a Markov renewal process.

Note 9.15 *$\{Y_n, n \in \mathbb{N}\}$ is a homogeneous Markov chain with transition probabilities given by:*

$$P_{ij} = \lim_{t \to \infty} P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i).$$

$\{Y_n, n \in \mathbb{N}\}$ is known as the embedded Markov chain of the semi-Markov process $\{X_t; t \geq 0\}$.

Definition 9.50 *$M(t) = (m_{ij}(t))_{i,j \in S}$ with*

$$m_{ij}(t) := P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i)$$

is called the semi-Markov kernel.

Definition 9.51 *Let $\{X_t; t \geq 0\}$ be a semi-Markov process. Then:*

1. *The sojourn time distribution for the state i defined as:*

$$H_i(t) := P(V_{n+1} \leq t \mid Y_n = i) = \sum_{j \in S} m_{ij}(t) \ .$$

2. *The mean sojourn time in state i is defined as:*

$$\mu_i := \begin{cases} \int_0^\infty x h_i(x) dx & \text{if the sojourn time distribution is continuous} \\ \sum_j x_j P(V_{n+1} = x_j \mid Y_n = i) & \text{if the sojourn time distribution is discrete.} \end{cases}$$

3. $$F_{ij}(t) = P(V_{n+1} \leq t \mid Y_n = i, Y_{n+1} = j) \ .$$

Note 9.16 *For $i, j \in S$, we have:*

$$F_{ij}(t) = \frac{m_{ij}(t)}{P_{ij}} \ .$$

Note 9.17 *A continuous-time Markov chain is a semi-Markov process with:*

$$F_{ij}(t) = 1 - e^{-\lambda_i t}, \quad t \geq 0 \ .$$

Therefore:

$$\begin{aligned} H_i(t) &= \sum_{j \in S} m_{ij}(t) \\ &= \sum_{j \in S} P_{ij} \left(1 - e^{-\lambda_i t}\right) \\ &= 1 - e^{-\lambda_i t} \\ \mu_i &= \int_0^\infty x_i h_i(t) dt = \frac{1}{\lambda_i} \ . \end{aligned}$$

■ EXAMPLE 9.53

Consider a stochastic process $\{X_t; t \geq 0\}$ with state space $S = \{1, 2, 3, 4\}$. Assume that the time spent in states 1 and 2 follow exponential distributions with parameter $\lambda_1 = 2$ and $\lambda_2 = 3$, respectively. Further, assume that the time spent in states 3 and 4 follow general distributions with distribution function $H_3(t)$ and $H_4(t)$ (sojourn time distribution) respectively and are given by:

$$H_3(t) = \begin{cases} 0 & \text{if } \ t < 1 \\ t-1 & \text{if } \ 1 \leq t < 2 \\ 1 & \text{if } \ 2 \leq t < \infty; \end{cases} \qquad H_4(t) = \begin{cases} 0 & \text{if } \ t < 2 \\ t-2 & \text{if } \ 2 \leq t < 3 \\ 1 & \text{if } \ 3 \leq t < \infty\,. \end{cases}$$

From Definition 9.49, $\{X_{t},; t \geq 0\}$ is a semi-Markov process with state space S. ▲

The following theorem describes the limiting behaviour of SMPs (see Cinlar, 1975).

Theorem 9.28 *Let $\{X_t; t \geq 0\}$ be a SMP with $\{Y_n; n \in \mathbb{N}\}$ its embedded Markov chain. Suppose that $\{Y_n; n \in \mathbb{N}\}$ is irreducible, aperiodic and positive recurrent. Then*

$$\lim_{t \to \infty} P(X_t = j) = \frac{\pi_j \mu_j}{\sum_{k \in S} \pi_k \mu_k} \tag{9.13}$$

where μ_j is the mean sojourn time in state j and $(\pi_j)_{j \in S}$ is the stationary distribution of $\{Y_n; n \in \mathbb{N}\}$.

■ EXAMPLE 9.54

Consider Example 9.53. Find the steady-state distribution of $\{X_t; t \geq 0\}$.

Solution: $\{X_t; t \geq 0\}$ is a SMP with state space $S = \{1, 2, 3, 4\}$. $\{Y_n; n \in \mathbb{N}\}$ is a homogeneous Markov chain with transition probability matrix

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}.$$

Since $\{Y_n; n \in \mathbb{N}\}$ is irreducible, aperiodic and positive recurrent, the stationary distribution $\pi = (\pi_i)_{i \in S}$ exists and is obtained by solving

$$\pi = \pi P \text{ and } \sum_{i \in S} \pi_i = 1$$

to get

$$\pi_1 = \pi_2 = \pi_3 = \pi_4 = \frac{1}{4}.$$

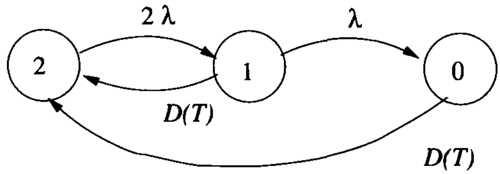


Figure 9.10 State transition diagram for Example 9.55

The mean sojourn times for states $i = 1, 2, 3, 4$ are given by:

$$\mu_1 = \frac{1}{2}; \quad \mu_2 = \frac{1}{3}; \quad \mu_3 = \frac{3}{2}; \quad \mu_4 = \frac{5}{2}.$$

Using equation (9.13), the steady-state probabilities for the SMP are given by:

$$\begin{aligned}
\lim_{t\to\infty} P(X_t = 1) &= \frac{3}{29} \\
\lim_{t\to\infty} P(X_t = 2) &= \frac{2}{29} \\
\lim_{t\to\infty} P(X_t = 3) &= \frac{9}{29} \\
\lim_{t\to\infty} P(X_t = 4) &= \frac{15}{29} . \quad \blacktriangle
\end{aligned}$$

■ EXAMPLE 9.55

Consider a system with two components. Each component can fail with failure time exponentially distributed with parameter λ. The system has one repair unit and the repair time is constant with time T. Note that when both the components fail the system will be completely down. Assume that both the components can be repaired simultaneously and the repair time for both the components is a constant T. Let X_t denote the number of components working at time t. Figure 9.10 shows the state transition diagram for this model.

From Definition 9.49, we conclude that $\{X_t; t \geq 0\}$ is a semi-Markov process. Now:

$$\begin{aligned}
K_{i,j}(t) &= P(X_1 = j, V_1 < t \mid X_0 = i) \\
&= \begin{bmatrix} 0 & 1 - e^{-2\lambda t} & 0 \\ \frac{1}{T}\int_0^T e^{-\lambda t} dt & 0 & \frac{1}{T}(1 - e^{-\lambda t}) \\ 1 & 0 & 0 \end{bmatrix} .
\end{aligned}$$

By using the steps in Theorem 9.28, the steady state probabilities can be obtained. ▲

Definition 9.52 (Markov Regenerative Process) *Let $\{Y_n; n \in \mathbb{N}\}$ be a stochastic process with state space $S = \mathbb{N}$ and $\{V_n; n \in \mathbb{N}\}$ a sequence of nonnegative random variables. Let*

$$\begin{aligned} U_0 &:= 0 \\ U_1 &:= V_1 \\ &\vdots \\ U_n &:= \sum_{i=1}^{n} V_i \quad \textit{for} \;\; n \geq 2 \end{aligned}$$

and

$$U(t) := max\{n \geq 1 : U_n \leq t\}, \quad t \geq 0 \ ,$$

The process $\{X_t; t \geq 0\}$ given by

$$X_t := Y_{U(t)}, \quad t \geq 0 \ ,$$

is called a Markov regenerative process if:

(a) There exists $S' \subset S$ such that $\{Y_n; n \in \mathbb{N}\}$ is a homogeneous Markov chain with state space S'.

(b) For all $n \geq 0$:

$$\begin{aligned} P(Y_{n+1} = j, \quad & V_{n+1} \leq t \mid Y_n = i, \cdots, Y_0 = i_0, V_n \leq t_n, \cdots, V_1 \leq t_1) \\ &= P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i) \ \ (\textit{Markov property}) \\ &= P(Y_1 = j, V_1 \leq t \mid Y_0 = i) \ \ (\textit{time homogeneity}) \ . \end{aligned}$$

(c) $P(X_{V_n + t} = j \mid X_u, 0 \leq u \leq V_n, Y_n = i) = P(X_t = j \mid Y_0 = i)$.

Note 9.18 *$\{Y_n; n \in \mathbb{N}\}$ is a homogeneous Markov chain with state space S'. Its transition probabilities are given by:*

$$P_{ij} = \lim_{t \to \infty} P(Y_{n+1} = j, V_{n+1} \leq t \mid Y_n = i), \quad i, j \in S' \ .$$

Definition 9.53 $K(t) = (k_{ij}(t))_{i,j \in S'}$ *with*

$$k_{ij}(t) := P(Y_1 = j, V_1 \leq t \mid Y_0 = i)$$

is called the global kernel of MRGP.

Definition 9.54 $E(t) = (e_{ij}(t))_{i \in S', j \in S}$ *with*

$$e_{ij}(t) := P(Y_1 = j, V_1 > t \mid Y_0 = i)$$

is called the local kernel of MRGP. The matrix $E(t)$ describes the behavior of the MRGP between two regeneration epochs of the embedded Markov chain $\{Y_n, n \in \mathbb{N}\}$.

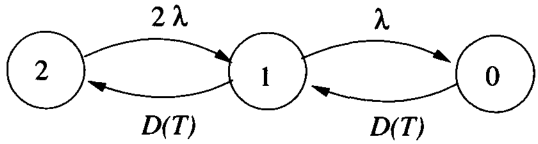


Figure 9.11 State transition diagram for Example 9.56

Definition 9.55 *Let $\{X_t; t \geq 0\}$ be a Markov regenerative process. Then:*

1. *The sojourn time distribution for state i is defined as:*

$$H_i(t) := P(V_{n+1} \leq t \mid Y_n = i) = \sum_{j \in S'} k_{ij}(t), \quad i \in S' .$$

2. *The mean sojourn time in state $j \in S$ between two successive regeneration epochs given that it started in state $i \in S'$ after the last regeneration is defined as:*

$$\begin{aligned} \alpha_{i,j} &:= E(\textit{time in state } j \textit{ during } (0, V_1) \mid Y_0 = i) \\ &= \int_0^{\infty} e_{i,j}(t)dt, i \in S', j \in S . \end{aligned} \tag{9.14}$$

3. *The mean sojourn time in state $j \in S$ between two successive regeneration epochs is defined as:*

$$\mu_j = \sum_{i \in S'} \alpha_{i,j}, \quad j \in S. \tag{9.15}$$

The following theorem describes the limiting behavior of MRGPs (see Cinlar, 1975).

Theorem 9.29 *Let $\{X_t; t \geq 0\}$ be a MRGP with state space S and $\{Y_n; n \in \mathbb{N}\}$ its embedded Markov chain with state space $S' \subset S$. Suppose that $\{Y_n; n \in \mathbb{N}\}$ is irreducible, aperiodic and positive recurrent. Then*

$$\lim_{t \to \infty} P(X_t = j) = \frac{\sum_{k \in S'} \pi_k \alpha_{k,j}}{\sum_{k \in S'} \pi_k \mu_k}, \; j \in S , \tag{9.16}$$

where $(\pi_k)_{k \in S'}$ is the stationary distribution of $\{Y_n; n \in \mathbb{N}\}$, μ_k is defined in equation (9.15) and $\alpha_{k,j}$ is defined in equation (9.14).

Markov regenerative processes have wide applications in queueing models which are presented in the next chapter.

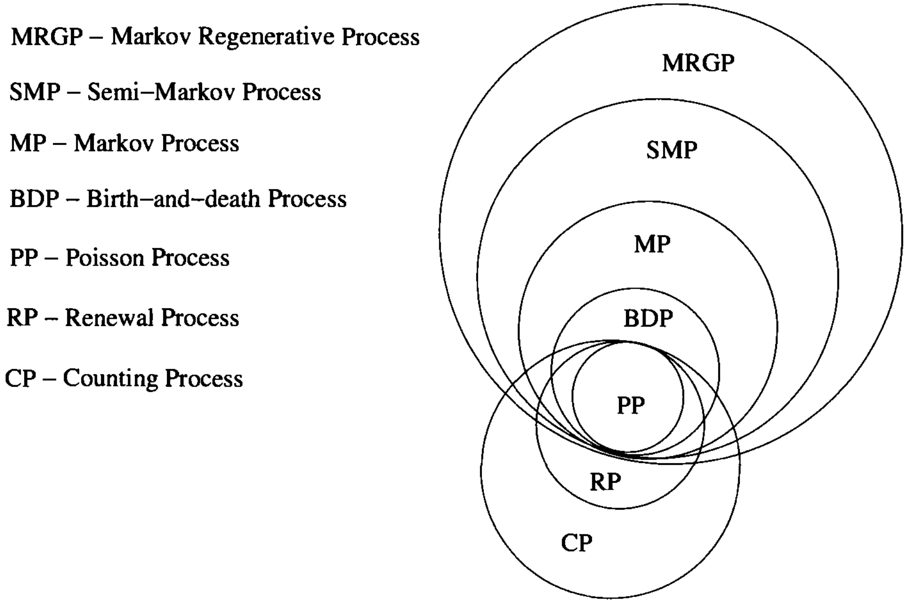


Figure 9.12 Relationship between various stochastic processes

■ **EXAMPLE 9.56**

Consider Example 9.55. Assume that when both the components fail, they cannot be repaired simultaneously. Assume that repair time is constant with time T. Let X_t denote the number of components working at time t. Figure 9.11 shows the state transition diagram for this model.

Based on the transition of the system moving from one state to other states, we conclude that $\{X_t; t \geq 0\}$ is a regenerative process. By using the steps in Theorem 9.29, the steady state probabilities can be obtained. ▲

To conclude this chapter, we present Figure 9.12 to indicate the relationship of inclusion and exclusion among various important stochastic processes. For more details, readers may refer to Grimmett and Stirzaker (2001), Karlin and Taylor (1975), Medhi (1994), Lawler (2006), Resnick (1994), Ross (1996), and Ross (2007).

EXERCISES

9.1 Let $X_t = A\cos(\eta t + \phi)$ be a cosine process, where A, η are real-valued random variables and ϕ is uniformly distributed over $(0, 2\pi)$ and is independent of A and η.

a) Compute the covariance function of the process $\{X_t; t \in \mathbb{R}\}$.

b) Is $\{X_t; t \in \mathbb{R}\}$ covariance stationary.

9.2 Consider the process $X_t = A\cos(wt) + B\sin(wt)$ where A and B are uncorrelated random variables with mean 0 and variance 1 and w is a positive constant. Is $\{X_t; t \geq 0\}$ covariance/wide-sense stationary?

9.3 Show that an i.i.d. sequence of continuous random variables with common density function f is strictly stationary.

9.4 Show that a stochastic process with independent increments has the Markov property.

9.5 Suppose that a taxi continuously moves between the airport and two hotels according to a Markov chain with a transition probability matrix given by:

$$P = \begin{array}{c} \\ Airport \\ HotelA \\ HotelB \end{array} \begin{array}{c} \begin{array}{ccc} Airport & HotelA & HotelB \end{array} \\ \begin{pmatrix} 0 & 0.6 & 0.4 \\ 0.8 & 0 & 0.2 \\ 0.9 & 0.1 & 0 \end{pmatrix} \end{array} .$$

a) Argue that this Markov chain has a stationary distribution.

b) In the long run, what proportion of time will the taxi be in each of the three locations?

c) Suppose that you are at the airport and just about to get into this taxi when a very rude person barges ahead of you and steals the cab. If each transition that the taxi makes takes 25 minutes, how long do you expect to have to wait at the airport until the taxi returns?

9.6 One way of spreading information on a network uses a rumor-spreading paradigm. Suppose that there are 5 hosts currently on the network. Initially, one host begins with a message. In every round, each host that has the message contacts another host chosen independently and uniformly at random from the other 4 hosts and sends the message to that host. The process stops when all hosts have the message.

a) Model this process as a discrete-time Markov chain.

b) Find the transition probability matrix for the chain.

c) Classify the states of the chain as transient, recurrent or null recurrent.

9.7 Consider a Markov chain with state space $\{0, 1, 2, 3, 4\}$ and transition matrix

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} .$$

a) Classify the states of the chain.

b) Determine the stationary distribution for states 1 and 2.

c) For the transient states, calculate μ_{ij}, the expected number of visits to transient state j, given that the process started in transient state i.

d) Find $P(X_5 = 2 \mid X_3 = 1)$.

9.8 The transition probability matrix of a discrete-time Markov chain $\{X_n; n = 1, 2, \cdots\}$ having three states 1, 2 and 3 is

$$P = \begin{pmatrix} 0.3 & 0.4 & 0.3 \\ 0.6 & 0.2 & 0.2 \\ 0.5 & 0.4 & 0.1 \end{pmatrix}$$

and the initial distribution is

$$\pi = (0.7, 0.2, 0.1) .$$

Find:

a) $P(X_2 = 3)$.

b) $P(X_3 = 2, X_2 = 3, X_1 = 3, X_0 = 2)$.

9.9 Consider a DTMC model which arises in an insurance problem. To compute insurance or pension premiums for professional diseases such as silicosis, we need to compute the average degree of disability at preassigned time periods. Suppose that we retain m degrees of disability $S_1, S_2, \cdots, S_m$. Assume that an insurance policy holder can go from degree S_i to degree S_j with a probability p_{ij}. This strong assumption leads to the construction of the DTMC model in which $P = [p_{ij}]$ is the one-step transition probability matrix related to the degree of disability. Using real observations recorded in India, we considered the following transition matrix P:

$$P = \begin{pmatrix} 0.90 & 0.10 & 0 & 0 & 0 \\ 0 & 0.95 & 0.05 & 0 & 0 \\ 0 & 0 & 0.90 & 0.05 & 0.05 \\ 0 & 0 & 0 & 0.90 & 0.10 \\ 0 & 0 & 0.05 & 0.05 & 0.90 \end{pmatrix} .$$

a) Classify the states of the chain as transient, positive recurrent or null recurrent along each period.

b) Find the limiting distribution for the degree of disability.

9.10 Let $p_{00} = 1$ and, for $j > 0$, $p_{jj} = p$, $p_{j,j-1} = q$ where $p + q = 1$, define the transition probability matrix of DTMC.

a) Find $f_{j0}^{(n)}$, the probability that absorption takes place exactly at the nth step given that the initial state is j.

b) Find the expectation of this distribution.

9.11 Assume a DTMC $\{X_n; n = 1, 2, \cdots\}$ with state space $E = \{0, 1, 2\}$ and transition probability matrix P given by:

$$P = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}.$$

a) Classify the states of the Markov chain.

b) Find the distribution of X_n given that the initial probability vector is $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$.

c) Examine whether there exists a limiting distribution for this Markov chain.

9.12 Consider a time homogeneous discrete-time Markov chain with transition probability matrix P and states $\{0, 1, 2, 3, 4\}$ where:

$$P = \begin{pmatrix} 0.5 & 0 & 0.5 & 0 & 0 \\ 0.25 & 0.5 & 0.25 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0.5 & 0.5 \end{pmatrix}.$$

a) Classify the states of the Markov chain as positive recurrent, null recurrent or transient.

b) Discuss the behavior of $p_{ij}^{(n)}$ as $n \to \infty$ for all $i, j = 0, 1, \cdots, 4$.

9.13 Consider a DTMC with states $\{0, 1, 2, 3, 4\}$. Suppose $p_{0,4} = 1$ and suppose that when the chain is in state i, $i > 0$, the next state is equally likely to be any of the states $0, 1, \cdots, i - 1$.

a) Discuss the nature of the states of this Markov chain.

b) Discuss whether there exists a limiting distribution and find it if it exists.

9.14 A factory has two machines and one repair crew. Assume that probability of any one machine breaking down on a given day is α. Assume that if the repair crew is working on a machine, the probability that they will complete the repairs in one more day is β. For simplicity, ignore the probability of a repair completion or a breakdown taking place except at the end of a day. Let X_n be the number of machines in operation at the end of the nth day. Assume that the behavior of X_n can be modeled as a DTMC.

a) Find the transition probability matrix for the chain.

b) If the system starts out with both machines operating, what is the probability that both will be in operation two days later?

9.15 Consider a gambler who at each play of the game has probability p of winning one unit and probability $q = 1 - p$ of losing one unit. Assume that successive plays of the game are independent. Suppose the gambler's fortune is presently i, and suppose that we know that the gambler's fortune will eventually reach N (before it goes to 0). Given this information, show that the probability he wins the next game is:

$$\begin{cases} \frac{p[1-(q/p)^{i+1}]}{1-(q/p)^i} & \text{if } p \neq \frac{1}{2} \\ \frac{i+1}{2i} & \text{if } p = \frac{1}{2} \, . \end{cases}$$

9.16 Consider a DTMC on the nonnegative integers such that, starting from i, the chain goes to state $i + 1$ with probability $p, 0 < p < 1$, and goes to state 0 with probability $1 - p$.

a) Show that this DTMC is irreducible and recurrent.

b) Show that this DTMC has a unique steady state distribution π and find π.

9.17 Suppose a virus can exist in 3 different strains and in each generation it either stays the same or, with probability p, mutates to another strain, which is chosen at random. What is the probability that the strain in the nth generation is the same as that in the 0th?

9.18 There are two drive-through service windows at a very busy restaurant, in series, served by a single line. When there is a backlog of waiting cars, two cars begin service simultaneously. Out of the two cars being served, the front customer can leave if he finishes service before the rear customer, but the rear customer has to wait until the first customer finishes service. Consequently, each service window will sometimes be idle if their customer completes service before the customer at the other window. Assume there is an infinite backlog of waiting cars and that service requirements of the cars (measured in seconds) are exponential random variables with a mean of 120 seconds. Draw the state transition diagram that describes whether each service window is busy. What is the stationary probability that both service windows are busy?

9.19 Consider a taxi station where taxis and customers arrive independently in accordance with Poisson processes with respective rates of one and two per minute. A taxi will wait no matter how many other taxis are in the system. However, an arriving customer that does not find a taxi waiting leaves. Note that a taxi can accommodate only one customer. Let $\{X_t; t \geq 0\}$ denote the number of taxis waiting for customers at time t.

a) Draw the state transition diagram for this process.

b) Find the average time a taxi waits in the system.

c) Find the proportion of arriving customers that get taxis.

9.20 A particle starting at position A *makes a random to-and-from motion in a circle over* A, B, C, D, A *such that if it reaches a position* $I(= A, B, C$ *or* $D)$ *at any step, in the next step, either it stays in the same position with probability* $\frac{1}{2}$ *or else jumps to the right (clockwise) or left (anticlockwise) with probability* $\frac{1}{3}$ *or* $\frac{1}{6}$ *respectively.*

a) Draw the state transition diagram for the process.

b) Determine the probability that the process will never return to state A *having started at* A *initially. Give justification.*

c) Find the limiting distribution for the process.

9.21 *Customers arrive at a bank at a Poisson rate* λ. *Suppose two customers arrive in the first hour. What's the probability that:*

a) Both arrived in the first 20 minutes?

b) At least one arrived in the first 20 minutes?

9.22 *An insurance company pays out claims on its life insurance policies in accordance with a Poisson process having rate* $\lambda = 5$ *per week. If the amount of money paid on each policy is exponentially distributed with mean $2000, what is the mean and variance of the amount paid by the insurance company in a four-week span?*

9.23 *Show that if* $N_1(t)$ *and* $N_2(t)$ *are independent Poisson processes with rate* λ_1 *and* λ_2, *respectively, then* $N_t = N_1(t) + N_2(t)$ *is a Poisson process with rate* $\lambda_1 + \lambda_2$.

9.24 *Let $\{N_t; t \geq 0\}$ be a Poisson process with intensity parameter λ. Suppose each arrival is "registered" or recorded with probability p, independent of other arrivals. Let M_t be the counting process of registered or recorded events. Show that M_t is a Poisson process with parameter λp.*

9.25 *Consider the IIT Delhi Open House program. Assume that students from various schools arrive at the reception at the instants of a Poisson process with rate 2 per minute. At the reception main door, two program representatives separately explain the program to students entering the hall. Each explanation takes a time (in minutes) which is assumed to be exponentially distributed with parameter 1 and is independent of other explanations. After the explanation, the students enter the hall. If both representatives are busy the student goes directly into the hall. Let X_t be the number of busy representatives at time t. Without loss of generality, assume that the system is modeled as a birth-and-death process.*

a) *Write the generator matrix Q.*

b) *Write the forward Kolmogorov equations for the birth-and-death process $\{X_t; t \geq 0\}$.*

c) *Derive the stationary distribution of the process.*

9.26 *Consider the New Delhi International Airport. Suppose that it has three runways. Airplanes have been found to arrive at the rate of 20 per hour. It is estimated that each landing takes 3 minutes. Assume a Poisson process for arrivals and an exponential distribution for landing times. Without loss of generality, assume that the system is modeled as a birth-and-death process.*

a) *What is the steady state probability that there is no waiting time to land?*

b) *What is the expected number of airplanes waiting to land?*

c) *Find the expected waiting time to land?*

9.27 *A cable car starts off with n riders. The times between successive stops of the car are independent exponential random variables, each with rate λ. At each stop, one rider gets off. This takes no time and no additional riders get on. Let X_t denote the number of riders present in the car at time t.*

a) *Write down the Kolmogorov forward equations for the process $\{X_t; t \geq 0\}$.*

b) *Find the mean and variance of the number of riders present in the car at any time t.*

9.28 *Let $\{X_t; t \geq 0\}$ be a pure birth process with $\lambda_n = n\lambda,\ n = 1, 2, \cdots,$ $\lambda_0 = \lambda; \mu_n = 0,\ n = 1, 2, \cdots.$*

a) Find the conditional probability that $X_t = n$ *given that* $X_0 = i$ $(1 \le i \le n)$.

b) Find the mean of this conditional distribution.

9.29 *Suppose that an office receives two different types of inquiry: persons who walk in off the street and persons who call by telephone. Suppose the two types of arrival are described by independent Poisson processes with rate* α_1 *and* α_2 *for walk-in and the callers, respectively. What is the distribution of the number of telephone calls received before the first walk-in customer.*

9.30 *An insurance company wishes to test the assumption that claims of a particular type arrive according to a Poisson process model. The times of arrival of the next 20 incoming claims of this type are to be recorded, giving a sequence* $T_1, \cdots, T_{20}$.

a) Give reasons why tests for the goodness of fit should be based on the interarrival times $X_i = T_i - T_{i-1}$ *rather than on the arrival times* T_i.

b) Write down the distribution of the interarrival times if the Poisson process model is correct and state one statistical test which could be applied to determine whether this distribution is realized in practice.

c) State the relationship between successive values of the interarrival times if the Poisson process model is correct and state one method which could be applied to determine whether this relationship holds in practice.

9.31 *Let us consider that the buses arrive at a particular stop according to a Poisson process with rate 10 per hour. A student starts waiting at the stop at 1:00.*

a) What is the probability that no buses arrive in the next half-hour?

b) How many buses are expected to arrive in 2 hours?

c) What is the variance of the number of buses that arrive over a 5-hour period?

d) What is the probability that no buses arrive in the next half-hour given that one hasn't arrived for over 2 hours?

e) If you wait until 2 o'clock and no buses have arrived, what is the probability that a student still has to wait at least a further half-hour before one comes?

9.32 *According to an interest rate model which operates in continuous time, the interest rate* r_t *may change only by upward jumps of fixed size* j_u *or by downward jumps of fixed size* j_d *(where* $j_d < 0$*), occurring independently according to Poisson processes* $N_u(t)$ *(with rate* λ_u*) and* $N_d(t)$ *(with rate* λ_d*).*

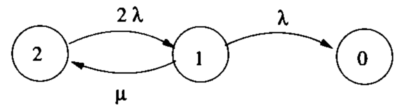


Figure 9.13 State transition diagram for Exercise 9.33

Let T_u denote the time of the first up jump in the interest rate, T_d the time of the first down jump, $T = \min(T_u, T_d)$ the time of the first jump. Further, let I be defined as an indicator taking the value 1 if the first jump is an up jump or 0 otherwise.

a) Determine expressions for the probabilities $P\{T_u > t\}, P\{T_d > t\}$ and $P\{T > t\}$.

b) Determine the distribution of I.

c) Show, by evaluating $P\{T > t \text{ and } I = 1\}$, that I and T are independent random variables.

d) Calculate the expectation and variance of the interest rate at time t given the current rate r_0. Hint: $r_t = r_0 + j_u N_u(t) + j_d N_d(t)$

e) Show that $\{r_t; t \geq 0\}$ is a process with stationary and independent increments.

9.33 *Consider Example 9.32. Assume that both units have a failure rate λ. The state transition diagram is shown in Figure 9.13.*

a) Write the forward Kolmogorov equations for this system.

b) Determine the reliability of the system.

9.34 *Consider an n-unit system in which 1 unit is active and the other $n-1$ units are inactive. There is one repairman and the repair rate of each unit is μ. The failure and repair rates are λ and μ, respectively. This system can be modeled as a birth-and-death process $\{X_t; t \geq 0\}$. The state of the system is the number of working components. Determine the steady state probabilities of the states of the underlying system.*

9.35 *Consider a service station with two identical computers and two technicians. Assume that when both computers are in good condition, most of the work load is on one computer, exposed to a failure rate $\lambda = 1$, while the other computer's failure rate is $\lambda = 0.5$. Further assume that, if one of the computer fails, the other one takes the full load, thus exposed to a failure rate $\lambda = 2$. Among the technicians, one is with repair rate $\mu = 2$ while the second is with repair rate $\mu = 1$. If both work simultaneously on the same computer,* *the total repair rate is $\mu = 2.5$. Note that, at any given moment, they work so that their repair rate is maximized.*

a) Determine the infinitesimal generator matrix Q.

b) Draw the state transition diagram of the system.

c) Determine the steady state probabilities and the system availability.

9.36 *Consider a service station with two computers and a technician. Suppose that computer A fails on the average once per hour and computer B twice per hour. The technician can repair a computer in 3 minutes on average.*

a) Determine the infinitesimal generator matrix Q.

b) Draw the state transition diagram of the system.

c) Determine the steady state probabilities and the system availability.

9.37 *The occurrences of successive failures can be described by a Poisson process with mean time between arrivals of 50 minutes.*

a) What is the probability that exactly 3 failures occur in the first 1000 minutes?

b) Find the probability of no failure in the first 500 minutes.

c) What is the probability that the first failure occurs after 400 minutes?

d) Compute the probability that the 4th failure occurs after 2000 minutes?

9.38 *Consider an n-unit parallel redundant system. The system is operates when at least one of the n units is operating. Then the system fails when all units are down simultaneously. It will begin to operate again immediately by replacing all failed units with new ones. Assume that each unit operates independently. Also assume that each unit has an identical failure distribution which follows exponential distribution with parameter λ. Obtain the renewal function.*

9.39 *Consider Example 9.34. Suppose that Markov chain $\{X_t; t \geq 0\}$ with state space S is irreducible and positive recurrent. Determine the stationary distribution.*

CHAPTER 10

INTRODUCTION TO QUEUEING MODELS

10.1 INTRODUCTION

The queueing theory is considered to be a branch of applied probability theory and is often used to describe the more specialized mathematical models for waiting lines or queues. The concept of queueing theory has been developed largely in the context of telephone traffic engineering originated by A. K. Erlang in 1909. Queueing models find applications in a wide variety of situations that may be encountered in health care, engineering, and operations research (Gross and Harris, 1998). In this chapter, the reader is introduced to the fundamental concepts of queueing theory and some of the basic queueing models which are useful in day-to-day real life. Important performance measures such as queue length, waiting time and loss probability are studied for some queueing models.

Queueing systems are comprised of customer(s) waiting for service and server(s) who serve the customer. They are frequently observed in some areas of day-to-day life, for example:

1. People waiting at the check-in counter of an airport

Introduction to Probability and Stochastic Processes with Applications, First Edition.
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

2. Aeroplanes arriving in an airport for landing
3. Online train ticket reservation system
4. People waiting to be served at a buffet
5. Customers waiting at a barber shop for a hair cut
6. Sequence of emails awaiting processing in a mail server ▲

Certain factors which affect the performance of queueing systems are as follows:

1. Arrival pattern
2. Service pattern
3. Number of servers
4. Maximum system capacity
5. Population size
6. Queue discipline

To incorporate these features, David G. Kendall introduced a queueing notation $A/B/C/X/Y/Z$ in 1953 where:

A	is the interarrival time distribution
B	is the service time distribution
C	is the number of servers
X	is the system capacity
Y	is the population size
Z	is the queue discipline

The symbols traditionally used for some common probability distributions for A and B are given as:

D	Deterministic (constant)
E_k	Erlang-k
G	General service time
GI	General independent interarrival time
H_k	k-Stage hyperexponential
M	Exponential (Markovian)
PH	Phase type

In this chapter, an infinite customer population and service in the order of arrival [first-in first-out (FIFO)] are default assumptions. Hence, for example $M/M/1/\infty$, where M stands for Markovian or memoryless. The first M denotes arrivals following a Poisson process, the second M denotes service

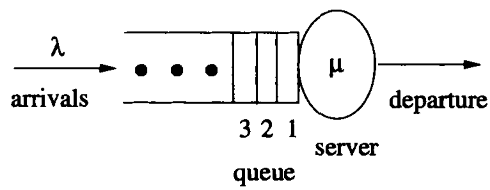


Figure 10.1 $M/M/1/\infty$ queueing system

time following exponential distribution, 1 refers to a single server and ∞ refers to infinite system capacity. Also there is an additional default assumption: interarrival and service times are independent. In a $GI/D/c/N$ queueing model, GI denotes a general independent interarrival time distribution, D denotes a deterministic (constant) service time distribution, c denotes the number of servers and N refers to the number of waiting spaces including one in service.

The queueing systems in real life are quite complex as compared to the existing basic queueing models. As the real-life system complexity increases, it becomes more difficult to analyze the corresponding queueing system. A model should be made as simple as possible but at the same time it should be close to reality. In this chapter, we describe simple queueing models and analyze them by studying their desired characteristics. The real-world phenomenon can be popularly mapped to one of these queueing models.

10.2 MARKOVIAN SINGLE-SERVER MODELS

10.2.1 $M/M/1/\infty$ Queueing System

Definition 10.1 ($M/M/1/\infty$ Queueing System) *The $M/M/1/\infty$ or simply $M/M/1$ queueing system describes a queueing system with both interarrival and service times following exponential distribution with parameters λ and μ, respectively, one server, unlimited queue size with FIFO queueing discipline and unlimited customer population. It is shown in Figure 10.1 with arrival rate λ and service rate μ.*

Theorem 10.1 *Let X_t be a random variable denoting the number of customers in the $M/M/1$ queueing system at any time t. Define:*

$$P_n(t) = P(X_t = n), \ n = 0, 1, 2, \cdots , \quad t \geq 0 \, .$$

Let $\lambda/\mu = \rho$. When $\rho < 1$, the steady state probabilities are given by:

$$P_n = \lim_{t \to \infty} P(X_t = n) = (1 - \rho)\rho^n, \ n = 0, 1, 2, \cdots \, .$$

Proof: The state transition diagram for the $M/M/1$ queueing system is shown in Figure 10.2. The stochastic process $\{X_t; t \geq 0\}$ in the $M/M/1$

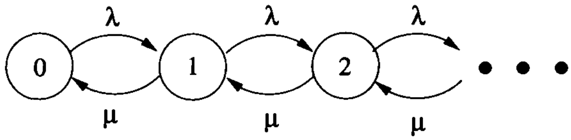


Figure 10.2 State transition diagram for $M/M/1/\infty$ queueing system

queueing system can be modeled as a homogeneous continuous-time Markov chain (CTMC) (refer to Section 9.3). In particular, this system can be modeled by a birth-and-death process (BDP) with birth rates $\lambda_n = \lambda, n = 0, 1, \cdots$, and death rates $\mu_n = \mu, n = 1, 2, \cdots$. When $\rho < 1$, the underlying CTMC is irreducible and positive recurrent and has stationary distribution. Note that the stationary distribution holds good only for the system in equilibrium, which is attained asymptotically. Moreover, when the process attains equilibrium, its behavior becomes independent of time and its initial state. Hence, the steady state balance equations can be obtained as discussed in Section 9.4 and is given by:

$$\begin{aligned} 0 &= -\lambda P_0 + \mu P_1 \\ 0 &= \lambda P_{n-1} - (\lambda + \mu) P_n + \mu P_{n+1}, \quad n = 1, 2, \cdots . \end{aligned}$$

Solving the above equations, we get

$$\begin{aligned} P_1 &= \frac{\lambda}{\mu} P_0 \\ P_{n+1} &= \frac{\lambda}{\mu} P_n, \; n = 1, 2, \cdots \end{aligned}$$

so that:

$$P_{n+1} = \left(\frac{\lambda}{\mu}\right)^{n+1} P_0, \; n = 1, 2, \cdots .$$

The value of P_0 can be computed by using the fact that the sum of all the probabilities must be equal to 1, i.e., $\sum_{n=0}^{\infty} P_n = 1$. Hence, when $\rho < 1$:

$$P_0 = \frac{1}{1 + \frac{\lambda}{\mu} + (\frac{\lambda}{\mu})^2 + \cdots} = 1 - \rho .$$

The steady state probabilities are given by:

$$P_n = (1 - \rho)\rho^n, \; n = 0, 1, 2, \cdots . \tag{10.1}$$

■

Note 10.1 *Equation (10.1) represents the probability mass function of a discrete random variable denoting the number of customers in the system in the* *long run. Clearly this distribution follows a geometric distribution with parameter $1-\rho$. Moreover, as $t \to \infty$, the system state depends neither on t nor on the initial state. The sequence $\{P_n,\ n = 0, 1, \cdots\}$ is called the steady state or stationary distribution. Here, $\rho < 1$ is a necessary and sufficient condition for the steady state solution to exist; otherwise the size of the queue will increase without limit as time advances and therefore steady state solution will not exist.*

Note 10.2 *In case the arrivals follow a Poisson process, a customer arriving to the queue in the steady state sees exactly the same statistics of the number of customers in the system as for "real random times". In a more condensed form, this is expressed as Poisson Arrivals See Time Averages, abbreviated PASTA (Wolff, (1982)). Since $M/M/1$ queueing system satisfies the PASTA property, an arriving customer finds n customers in the system with probability P_n.*

Theorem 10.2 *Let L_s be the average number of customers in the $M/M/1$ queueing system and L_q be the average number of customers in the queue. Then:*

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1-\rho}; \quad L_q = \frac{\rho^2}{1-\rho} .$$

Proof: Using (10.1), the mean number of customers can be found. Note that the number of customers in the system is the sum of the number of customers in the queue and the number of customers in service. Hence:

$$L_s = \sum_{n=0}^{\infty} nP_n = \sum_{n=1}^{\infty} n(1-\rho)\rho^n . \tag{10.2}$$

After simple calculations, we get:

$$L_s = \frac{\lambda}{\mu - \lambda} = \frac{\rho}{1-\rho}.$$

As expected, these equations show that with increasing load, i.e., as $\rho \to 1$, the mean number of customers in the system grows and the probability of an idle system decreases. Similarly, the average number of customers in the queue can be computed as:

$$\begin{aligned} L_q &= \sum_{n=1}^{\infty}(n-1)P_n \\ &= L_s - (1 - P_0) \\ &= \frac{\rho^2}{1-\rho} . \end{aligned}$$

∎

Note 10.3 *The probability that the server is busy is another performance measure of the queueing system. The probability that the server is busy when the system is in equilibrium is known as the utilization factor* ρ. *Clearly, this utilization factor for the* $M/M/1$ *queue is equal to* $1 - P_0 = \rho$, *which is also the traffic intensity.*

Note 10.4 *The number of customers in the system is of importance from the management's perspective and interest. Besides, the average queue size and average system size are also important parameters that represent the quality of service. Two more measures important from the customer's point of view are the average time spent in the system* (T_s) *and the average time spent in the queue* (T_q). *One of the most significant contributions in queueing theory is Little's formula, which gives the relation between the average number of customers in the system* (L_s) *and the average time spent in the system* (T_s) *and also between the average number of customers in the queue* (L_q) *and* T_q:

$$L_s = \lambda T_s; \qquad L_q = \lambda T_q \ . \tag{10.3}$$

It is justified that, using the average time spent in the system, T_s, *the average number of the customers during this time is* λT_s, *where* λ *is the average number of arrivals per unit time. It is very important to note that no assumption is made on the interarrival distribution, the service time distribution and the queue discipline. The first formal proof for Little's result appeared in Little (1961). Usually the average waiting time of a customer is difficult to calculate directly. Relation (10.3) comes to our rescue, because the evaluation of mean number of customers in the system is relatively easy. As an example, it is derived for the* $M/M/1$ *queueing system from (10.2):*

$$T_s = \frac{\rho}{\lambda(1-\rho)} = \frac{1}{\mu - \lambda} \ . \tag{10.4}$$

It can be deduced that:

$$T_s = T_q + \frac{1}{\mu} \ .$$

Note 10.5 *The variance of the four measures are given by (left as an exercise for the reader):*

$$\begin{aligned}
Var(\textit{Number of customers in the system}) &= \frac{\rho}{(1-\rho)^2} \\
Var(\textit{Number of customers in the queue}) &= \frac{\rho(1+\rho-\rho^2)}{(1-\rho)^2} \\
Var(\textit{Waiting time of customers in the system}) &= \frac{1}{(1-\rho)^2\mu^2} \\
Var(\textit{Waiting time of customers in the queue}) &= \frac{\rho(2-\rho)}{(1-\rho)^2\mu^2} \ .
\end{aligned}$$

Theorem 10.3 *Let W be the random variable which denotes the waiting time of a customer in the $M/M/1$ queueing system. Then, the distribution of waiting time is:*

$$P(W \le t) = \begin{cases} 0 & \text{if } t < 0 \\ 1-\rho & \text{if } t = 0 \\ 1-\rho e^{-(\mu-\lambda)t} & \text{if } t > 0 \,. \end{cases} \tag{10.5}$$

Proof: An arriving customer has to wait only when there is one or more customer in the system. Hence the waiting time of the customer will be 0 if the system is empty. Thus, the probability of the event $W = 0$ is equivalent to the probability of the system being empty, i.e.:

$$P(W = 0) = P(\text{system is empty}) = 1 - \rho \,.$$

When the system is not empty, i.e., there is one or more customers in the system, say n, then the arriving customer has to wait until all the n customers get serviced. Note that as the service time follows exponential distribution with parameter μ, which has the memory less property, the residual (remaining) service time of the customer in service also follows exponential distribution with parameter μ. Hence, the waiting time of a customer who arrives when n customers are already in the system follows a gamma distribution with parameters n and μ. Thus:

$$P(0 < W \le t) = \sum_{n=1}^{\infty} P(0 < W \le t \mid N = n) P(N = n) \,. \tag{10.6}$$

We know that:

$$P(N = n) = (1-\rho)\rho^n \tag{10.7}$$

$$P(0 < W \le t \mid N = n) = \int_0^t \frac{\mu^n x^{n-1} e^{-\mu x}}{(n-1)!} dx, \quad t > 0. \tag{10.8}$$

Substituting (10.7) and (10.8) in (10.6), we get:

$$\begin{aligned} P(0 < W \le t) &= \sum_{n=1}^{\infty} \int_0^t \frac{\mu^n x^{n-1} e^{-\mu x}}{(n-1)!} \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n dx \\ &= \rho\left(1 - e^{-(\mu-\lambda)t}\right) \,. \end{aligned}$$

Hence:

$$P(W \le t) = P(W = 0) + P(0 < W \le t) = 1 - \rho e^{-(\mu-\lambda)t} \,.$$

■

Similarly, we obtain the distribution of the total time spent by a customer in the system.

Theorem 10.4 *Let T denote the random variable for the total time spent by a customer in the $M/M/1$ queueing system. Then the distribution of total time spent is given by:*

$$P(0 < T \leq t) = 1 - e^{-\mu(1-\rho)t} .$$

Proof: As we evaluate the waiting time distribution, we have:

$$\begin{aligned} P(0 < T \leq t) &= \sum_{n=0}^{\infty} P(T \leq t \mid N = n) P(N = n) \\ &= \sum_{n=0}^{\infty} \int_0^t \frac{\mu^{n+1} x^n e^{-\mu x}}{n!} \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^n dx \\ &= 1 - e^{-\mu(1-\rho)t} . \end{aligned}$$

■

Note 10.6 *This shows that T has an exponential distribution with parameter $\mu - \lambda$. Hence, the average total time spent by a customer in the system is given by:*

$$T_s = E(T) = \frac{1}{\mu - \lambda} .$$

The above result is the same as that obtained in equation (10.4) using Little's formula. It can also be verified that $T_s = T_q + \frac{1}{\mu}$.

■ EXAMPLE 10.1

Consider a unisex hair salon where customers are served on a first-come, first-served basis. The data show that customers arrive according to a Poisson process with a mean arrival rate of 5 customers per hour. Because of its excellent reputation, customers are always willing to wait. The data further show that the customer processing time is exponentially distributed with an average of 10 minutes per customer. This system can be modeled as a $M/M/1$ queueing system. Then, in the long run:

(a) What is the average number of customers in the shop?

(b) What is the average number of customers waiting for a haircut?

(c) What is the probability that an arrival can walk right in without having to wait at all?

Solution: Let X_t be the random variable denoting the number of customers in the salon at any time t. Then the system can be modeled as a $M/M/1$ queueing system with $\{X_t; t \geq 0\}$ as the underlying stochastic process. Here $\lambda = 5$ per hour and $\mu = 6$ per hour. Hence, $\rho = \frac{5}{6}$.

(a) The average number of customers in the shop:

$$L_s = \frac{\rho}{1-\rho} = 5 \ .$$

(b) The average number of customers waiting for a haircut:

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{25}{6} \ .$$

(c) The probability that an arrival can walk right in without having to wait at all:

$$P_0 = 1 - \rho = 0.1667 \ . \qquad \blacktriangle$$

■ EXAMPLE 10.2

Consider the New Delhi International Airport. Assume that it has one runway which is used for arrivals only. Airplanes have been found to arrive at a rate of 10 per hour. The time (in minutes) taken for an airplane to land is assumed to follow exponential distribution with mean 3 minutes. Assume that arrivals follow a Poisson process. Without loss of generality, assume that the system is modeled as a $M/M/1$ queueing system.

(a) What is the steady state probability that there is no waiting time to land?

(b) What is the expected number of airplanes waiting to land?

(c) Find the expected waiting time to land.

Solution: We model the given problem as a $M/M/1$ queueing system where each state represents the number of airplanes landing. The arrival rate is 10 per hour and the service rate is 20 per hour.

(a) The required probability that an airplane does not have to wait to land is:

$$P(\text{runway is available}) = P_0 = 1 - \rho = 0.5 \ .$$

(b) From equation (10.3), the expected number of airplanes waiting to land is:

$$L_q = 0.5 \ .$$

(c) Using Little's formula, the expected waiting time to land is:

$$T_q = 0.05 \text{ hours} \ . \qquad \blacktriangle$$

EXAMPLE 10.3

Consider the online ticket reservation system of Indian Railways. Assume that customers arrive according to a Poisson process at an average rate of 100 per hour. Also assume that the time taken for each reservation by a computer server follows an exponential distribution. Find out at what average rate the computer server should issue an e-ticket in order to ensure that a customer will not wait more than 45 seconds with a probability of 0.95.

Solution: $P(W \leq 45) = 0.95$. Using equation (10.5), $1 - \rho e^{-45(\mu-\lambda)} = 0.95$ where $\lambda = \frac{100}{60 \times 60} = \frac{1}{36}$ per second. From the stability condition, $\mu > \frac{1}{36} = 0.02778$. Simple calculations yield:

$$\begin{aligned} \frac{\lambda}{\mu} e^{-45(\mu-\lambda)} &= 0.05 \\ 45\mu + \log \mu &= 0.9947 \,. \end{aligned}$$

Solving, we get:

$$\mu = 0.0508 \,.$$

Hence, the server issues the e-ticket at an average 19.685 seconds in order to ensure that a customer does not wait more than 45 seconds.

▲

EXAMPLE 10.4

In a mobile handset manufacturing factory, components arrive according to a Poisson process with rate λ. Assume that the testing time of the components is exponential with mean $1/\mu$ and there is one testing machine for the testing purpose. During the testing period, with probability p the product is considered to be faulty and sent back to the production unit queue.

(a) Determine the number of handsets in the queue of the production unit on average in the steady state.

(b) Find the average response time of a production unit in the steady state.

Solution: This system can be modeled as a $M/M/1$ queueing model with retrials. The state transition diagram for the underlying queueing model is shown in Figure 10.3.

The balance equations for this queueing model are:

$$\begin{aligned} \lambda P_0 &= (1-p)\mu P_1 \\ (\lambda + (1-p)\mu)P_k &= \lambda P_{k-1} + (1-p)\mu P_{k+1}, \; k = 1, 2, \cdots \,. \end{aligned}$$

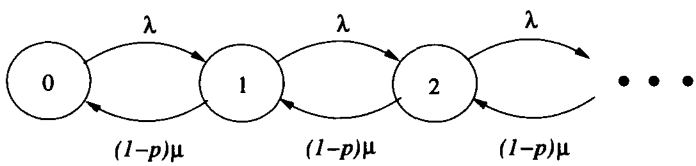


Figure 10.3 State transition diagram

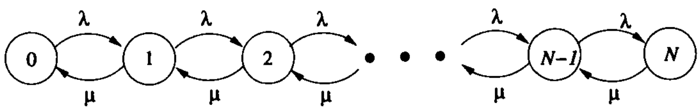


Figure 10.4 $M/M/1/N$ queueing system

(a) The average number of customers in the queue is given by:

$$L_q = \frac{\rho^2}{1-\rho}$$

where $\rho = \frac{\lambda}{(1-p)\mu}$.

(b) The average response time is given by:

$$T_s = \frac{1}{(1-p)\mu - \lambda} \; . \qquad \blacktriangle$$

10.2.2 $M/M/1/N$ Queueing System

We consider the single-server Markovian queueing system in which only a finite number of customers are allowed.

Definition 10.2 ($M/M/1/N$ Queueing System) *This system is a type of $M/M/1/\infty$ queue with at most N customers allowed in the system. When there are N customers in the system, i.e., one customer is under service and the remaining $N-1$ customers are waiting in the queue, new arriving customers are blocked. The state transition diagram for this system is shown in Figure 10.4.*

Theorem 10.5 *The state probabilities in equilibrium for the $M/M/1/N$ queueing system are given by:*

$$P_n = \begin{cases} \dfrac{(1-\rho)\rho^n}{1-\rho^{N+1}} & \text{if } \rho \neq 1 \\[2ex] \dfrac{1}{N+1} & \text{if } \rho = 1, \end{cases} \qquad n = 0, 1, \cdots, N \; . \tag{10.9}$$

Proof: Let X_t be a random variable denoting the number of customers in the system at any time t. For the $M/M/1/N$ queueing system, $\{X_t; t \geq 0\}$ is a BDP with birth rates $\lambda_n = \lambda$, $n = 0, 1, \cdots, N-1$, and death rates $\mu_n = \mu$, $n = 1, 2, \cdots, N$. Since the BDP $\{X_t; t \geq 0\}$ with finite state space $S = \{0, 1, \cdots, N\}$ is irreducible and positive recurrent, the equilibrium probabilities, denoted by $(P_i)_{i \in S}$, exist (refer to Example 9.34). Hence $(P_i)_{i \in S}$ are given by:

$$P_i = \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i} P_0, \quad i = 1, 2, \cdots, N,$$

with:

$$P_0 = \frac{1}{1 + \sum_{i=1}^{N} \frac{\lambda_0 \lambda_1 \cdots \lambda_{i-1}}{\mu_1 \mu_2 \cdots \mu_i}} .$$

Substituting $\lambda_n = \lambda$, $n = 0, 1, \cdots, N-1$, and $\mu_n = \mu$, $n = 1, 2, \cdots, N$, in the above equations, we obtain the state probabilities in equilibrium. ■

Note 10.7 *The state probability P_N is called the blocking probability since the customers are blocked when the system is in state N. In this finite-state queueing model, the effective arrival rate is given as $\lambda_{eff} = \lambda(1 - P_N)$.*

Note 10.8 *Using the state probabilities, one can compute the average number of customers in the system, which is given by:*

$$L_s = \sum_{n=0}^{N} n P_n = \begin{cases} \dfrac{\rho}{1-\rho} - \dfrac{(N+1)\rho^{N+1}}{1-\rho^{N+1}} & \text{if } \rho \neq 1 \\[2ex] \dfrac{N}{2} & \text{if } \rho = 1 . \end{cases}$$

Using Little's formula, the average time spent by a customer in the system can be calculated as

$$T_s = \frac{L_s}{\lambda_{eff}}$$

where λ_{eff} is the effective arrival rate to the system.

The waiting time distribution for this $M/M/1/N$ queueing system is left as an exercise for the reader.

■ EXAMPLE 10.5

Consider a message switching center in Vodafone. Assume that traffic arrives as a Poisson process with average rate of 180 messages per minute. The line has a transmission rate of 600 characters per second. Suppose that the message length discipline follows an exponential distribution with an average length of 176 characters. The arriving messages are buffered for transmission where buffer capacity is, say, N. What is the minimum N to guarantee that $P_N < 0.005$.

Solution: This system can be modeled as a $M/M/1/N$ queueing system, where each state represents the number of messages in the switching center. In this example, $\lambda = 3 \times 176 = 528$ characters per second and $\mu = 600$ characters per second. Hence, $\rho = \frac{\lambda}{\mu} = 0.88$. From equation (10.9), the probability that all the buffers are filled is:

$$P_N = \frac{(1-\rho)\rho^N}{1-\rho^{N+1}} \ .$$

The above equation can be solved for $P_N = 0.005$ with $\rho = 0.88$ to obtain $N = 26$. ▲

EXAMPLE 10.6

Consider a $M/M/1/2$ queueing system. Let $P_n(t)$ be the probability that there are n customers in the system at time t given that there was no customers at time 0. Find the time-dependent probabilities $P_n(t)$ for $n = 0, 1, 2$.

Solution: The Kolmogorov forward equations for the $M/M/1/2$ queueing system are:

$$\begin{aligned} p_0'(t) &= -\lambda p_0(t) + \mu p_1(t) \\ p_1'(t) &= \lambda p_0(t) - (\lambda+\mu)p_1(t) + \mu p_2(t) \\ p_2'(t) &= \lambda p_1(t) - \mu p_2(t) \ . \end{aligned}$$

Taking the Laplace transform on both sides with the initial condition that the system begins in state 0, we obtain:

$$\begin{aligned} sp_0^*(s) - 1 &= -\lambda p_0^*(s) + \mu p_1^*(s) \\ p_1^*(s) &= \lambda p_0^*(s) - (\lambda+\mu)p_1^*(s) + \mu p_2^*(s) \\ p_2^*(s) &= \lambda p_1^*(s) - \mu p_2^*(s) \ . \end{aligned}$$

In matrix form

$$A(s)\begin{bmatrix} p_0^*(s) \\ p_1^*(s) \\ p_2^*(s) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

where:

$$A(s) = \begin{bmatrix} s+\lambda & -\mu & 0 \\ -\lambda & s+\lambda+\mu & -\mu \\ 0 & -\lambda & s+\mu \end{bmatrix}$$

For $\lambda \neq \mu$, the roots of $det(A(s)) = 0$ are:

$$s_0 = 0; \quad s_1 = \lambda + \mu - \sqrt{\lambda\mu}; \quad s_2 = \lambda + \mu + \sqrt{\lambda\mu} \ .$$

Solving the above equations for $P_n^*(s), n = 0, 1, 2$, by using matrix inversion, we get:

$$\begin{aligned}
p_0^*(s) &= \frac{\lambda^2}{s(s^2+2s\mu+2\lambda s+\mu^2+\lambda\mu+\lambda^2)} \\
p_1^*(s) &= \frac{\lambda(s+\mu)}{s(s^2+2*s\mu+2\lambda s+\mu^2+\lambda\mu+\lambda^2)} \\
p_2^*(s) &= \frac{s^2+2s\mu+s\lambda+\mu^2}{s(s^2+2s\mu+2\lambda s+\mu^2+\lambda\mu+\lambda^2)} .
\end{aligned}$$

Applying partial fraction techniques and then taking the inverse Laplace transform, for $\lambda \neq \mu$, we get:

$$\begin{aligned}
p_0(t) &= \tfrac{\lambda^2}{\lambda^2+\lambda\mu+\mu^2} + \frac{\lambda^2}{2\lambda\mu - 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu-\sqrt{\lambda\mu})t} \\
&\quad + \frac{\lambda^2}{2\lambda\mu + 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu+\sqrt{\lambda\mu})t} \\
p_1(t) &= \tfrac{\lambda\mu}{\lambda^2+\lambda\mu+\mu^2} + \frac{\lambda(-\lambda+\sqrt{\lambda\mu})}{2\lambda\mu - 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu-\sqrt{\lambda\mu})t} \\
&\quad + \frac{\lambda(-\lambda-\sqrt{\lambda\mu})}{2\lambda\mu - 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu+\sqrt{\lambda\mu})t} \\
p_2(t) &= \tfrac{\mu^2}{\lambda^2+\lambda\mu+\mu^2} - \frac{\lambda\sqrt{\lambda\mu}}{2\lambda\mu - 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu-\sqrt{\lambda\mu})t} \\
&\quad + \frac{\lambda\sqrt{\lambda\mu}}{2\lambda\mu - 2(\lambda+\mu)\sqrt{\lambda\mu}} e^{-(\lambda+\mu+\sqrt{\lambda\mu})t} . \qquad \blacktriangle
\end{aligned}$$

■ EXAMPLE 10.7

A person repairing motor cars finds that the time spent on repairing a motor car follows an exponential distribution with mean 40 minutes. The shop can accommodate a maximum of 7 motor cars including the one under repair. The motor cars are repaired in the order in which they arrive, and the arrivals can be approximated by a Poisson process with an average rate of 15 per 10-hour day. What is the probability of the repair person being idle in the long run? Also, find the proportion of time the shop is not full in the long run.

Solution: This repair system can be modeled as a $M/M/1/N$ queueing model where each state of the system represents the number of motor cars being repaired. Here $N = 7$, $\lambda = \frac{1}{40}$ and $\mu = \frac{1}{40}$. Hence, from equation (10.9), we get:

$$P_0 = \frac{1}{N+1} = \frac{1}{8}.$$

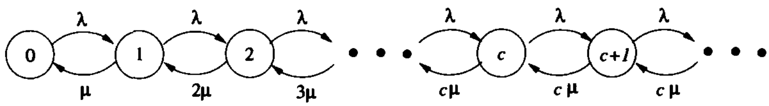


Figure 10.5 $M/M/c/\infty$ queueing system

Further, the expected idle time of the repair person is P_0. The proportion of time that the shop is not full is:

$$1 - P_7 = 1 - \frac{1}{N+1} = 1 - \frac{1}{8} = \frac{7}{8}\,.$$

Thus, the shop is not full 87.5% of the time in the long run. ▲

10.3 MARKOVIAN MULTISERVER MODELS

10.3.1 $M/M/c/\infty$ Queueing System

We consider a multiserver queueing system $M/M/c/\infty$. The state transition diagram for this system is shown in Figure 10.5. The corresponding steady state equations are given by:

$$\begin{aligned}
0 &= \lambda P_0 - \mu P_1 \\
0 &= \lambda P_{n-1} - (\lambda + n\mu)P_n + (n+1)\mu P_{n+1}, n = 1, 2, \cdots, c-1 \\
0 &= \lambda P_{n-1} - (\lambda + c\mu)P_n + c\mu P_{n+1}, n = c, c+1, \cdots\,.
\end{aligned}$$

The steady state probabilities will exist for $\rho < 1$, where $\rho = \frac{\lambda}{c\mu}$. Solving the above equations, we get the steady state probabilities

$$P_n = \begin{cases} \frac{1}{n!}(\frac{\lambda}{\mu})^n P_0 & 1 \leq n \leq c \\ \frac{1}{c^{n-c}c!}(\frac{\lambda}{\mu})^n P_0 & n \geq c\,. \end{cases} \tag{10.10}$$

Using the fact that $\sum_{n=0}^{\infty} P_n = 1$, we get:

$$P_0 = \left[\sum_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \sum_{n=c}^{\infty} \frac{1}{c^{n-c}c!}\left(\frac{\lambda}{\mu}\right)^n\right]^{-1}\,. \tag{10.11}$$

Note 10.9 *The probability that an arriving customer has to wait is given by*

$$\sum_{n=c}^{\infty} P_n = \sum_{n=c}^{\infty} \frac{1}{c^{n-c}c!}\left(\frac{\lambda}{\mu}\right)^n P_0$$

where P_0 is given by (10.11). This is known as the Erlang-C formula.

Note 10.10 *The average number of customers in the queue is given by:*

$$L_q = \sum_{n=c+1}^{\infty} (n-c)P_n = \frac{\rho}{c!(1-\rho)^2}\left(\frac{\lambda}{\mu}\right)^c P_0 \ . \qquad (10.12)$$

Using Little's formula we evaluate T_q, T_s and L_s:

$$T_q = \frac{L_q}{\lambda} = \left(\frac{(\lambda/\mu)^c}{c!(c\mu)(1-\rho)^2}\right) P_0 \qquad (10.13)$$

$$T_s = T_q + \frac{1}{\mu} = \left(\frac{(\lambda/\mu)^c}{c!(c\mu)(1-\rho)^2}\right) P_0 + \frac{1}{\mu} \qquad (10.14)$$

$$L_s = L_q + \frac{\lambda}{\mu} = \frac{\rho}{c!(1-\rho)^2}\left(\frac{\lambda}{\mu}\right)^c P_0 + \frac{\lambda}{\mu} \ . \qquad (10.15)$$

■ EXAMPLE 10.8

Consider Example 10.2. Further assume that it has three runways. In this scenario, assume that the arrival rate is 20 per hour and the service rate is 20 per hour. Without loss of generality, assume that the system is modeled as a $M/M/3$ queueing system.

(a) What is the steady state probability that there is no waiting time to land?

(b) What is the expected number of airplanes waiting to land?

(c) Find the expected waiting time to land.

Solution: We model the problem as a $M/M/3$ queueing system where each state represents the number of airplanes landing. The arrival rate is 20 per hour and the service rate is 20 per hour.

(a) The required probability that an airplane does not have to wait to land is

$$P(\text{at least one runway is available})$$

$$\begin{aligned} &= P(\text{Less than 3 airplanes are landing}) \\ &= P_0 + P_1 + P_2 \end{aligned}$$

where P_i is the steady state probability of the system being in state i. Substituting the values of P_i using equations (10.10) and (10.11), the desired probability is 0.909 .

(b) From equation (10.12), the expected number of airplanes waiting to land is:

$$L_q = \sum_{n=4}^{\infty}(n-3)P_n = 0.04535 \ .$$

(c) From equation (10.13), the expected waiting time to land is:

$$T_q = 0.0022725 \ . \qquad \blacktriangle$$

■ EXAMPLE 10.9

Consider Walmart supermarket with 7 counters. Assume that 9 customers arrive on an average every 5 minutes while each cashier in the counter can serve on an average three customers in 5 minutes. Assume that arrivals follow a Poisson process and service times follow exponential distribution. Find:

(a) Average number of customers in the queue

(b) Average time a customer spends in the system

[c) Average number of customers in the system

(d) Optimal number of counters so that the proportion of time a customer has to wait is at most 10 seconds

Solution: This system is modeled as a $M/M/7/\infty$ queueing system. Here, $\lambda = \frac{9}{5}$ and $\mu = \frac{3}{5}$. The state probabilities are given by:

$$P_n = \begin{cases} \frac{1}{n!}(\frac{\lambda}{\mu})^n P_0 & 1 \leq n \leq 7 \\ \frac{1}{7^{n-7}7!}\left(\frac{\lambda}{\mu}\right)^n P_0 & n \geq 7 \ . \end{cases}$$

Using the fact $\sum_{n=0}^{\infty} P_n = 1$, we get $P_0 = 0.0496$.

(a) Average number of customers in the queue:

$$L_q = \sum_{n=8}^{\infty}(n-7)P_n = 0.02825 \ .$$

(b) Average time a customer spends in the system: $T_s = 0.015694$.

(c) Average number of the customers in the system: $L_s = 3.02825$.

(d) Using the fact that $\frac{\lambda}{c\mu} < 1$, we get $c > 3$. Further, since it is required that $T_q \leq 10$ seconds $= 0.1667$ minutes, by trial and error, we get that, for $c = 4$, $T_q = 0.2121$ and, for $c = 5$, $T_q = 0.0786$. Hence the optimal number of servers that must be installed in order that $T_q < 10$ seconds is $c = 5$. $\blacktriangle$

■ **EXAMPLE 10.10**

Suppose that Air India airlines is planning a new customer service center at London. Assume that each agent can serve a caller with an exponentially distributed service time with an average of 4 minutes. Also assume that calls arrive randomly as a Poisson process at an average rate of 30 calls per hour. Also assume that the system has a large message buffer system to hold calls that arrive when no agents are free. How many agents should be provided so that the average waiting time for those who must wait does not exceed 2 minutes?

Solution: This system can be modeled as a $M/M/c/\infty$ queueing system where $\lambda = \frac{1}{2}$ and $\mu = \frac{1}{4}$. We require that for such a system to be stable, $\frac{\lambda}{c\mu} < 1 \Rightarrow c > 2$. By trial and error, we obtain, for $c = 3$, $T_q = 0.5925$. Hence the number of agents who must be provided so that the average waiting time $T_q < 2$ is $c = 3$. ▲

■ **EXAMPLE 10.11**

A situation in which a customer refuses to enter a queueing system because the queue is too long is said to be balking. On the other hand, a customer who enters the system but leaves the queue after some time without receiving service because of excessive waiting time is said to be reneging. Assume that any customer who leaves the system and then decides to return after some time is a new arrival. Consider the $M/M/c$ queueing system with the following:

1. (Balking) An arriving customer who finds that all the servers are busy may join the system with probability p or balk (does not join) with probability $1 - p$.

2. (Reneging) A customer already in the queue may renege, i.e., leave the system without obtaining service. It is assumed that the time spent by such a customer in the system follows an exponential distribution with parameter β.

Find the steady state probability of the $M/M/c$ system with:

(a) Only balking is possible.

(b) Only reneging is possible.

(c) Both balking and reneging are possible.

Solution:

(a) Consider the $M/M/c$ with balking. Note that the arriving customer will enter the system if any server is not busy. This model can be viewed

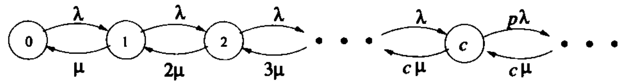


Figure 10.6 $M/M/c$ queueing system with balking

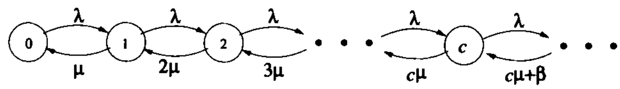


Figure 10.7 $M/M/c$ queueing system with reneging

as a birth-and-death process with the arrival rate for the system being $\lambda_n = \lambda, n = 0, 1, \cdots, c-1, \lambda_n = p\lambda, n = c, c+1, \cdots$. The service rate is $\mu_n = n\mu, n = 1, 2, \cdots, c, \mu_n = c\mu, n = c+1, c+2, \cdots$. The state transition diagram for this model is shown in Figure 10.6.

The steady state probability is given as:

$$P_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 & 1 \le n \le c \\ \frac{1}{c!}\left(\frac{p}{c}\right)^{n-c}\left(\frac{\lambda}{\mu}\right)^n P_0 & n > c\,. \end{cases}$$

(b) Consider the $M/M/c$ system with reneging. In this case, the customer will not leave the system when the customer is under service. Hence, the arrival rate is $\lambda_n = \lambda, n = 0, 1, \cdots$. But the service rate is $\mu_n = n\mu, n = 1, 2, \cdots, c$, $\mu_n = (n-c)\beta + c\mu, n = c+1, c+2, \cdots$. The state transition diagram for this model is shown in Figure 10.7. Accordingly, the steady state probability is given as:

$$P_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 & 1 \le n \le c \\ \\ \frac{\lambda^n}{c!\mu^c c^{n-c}(c\mu+(n-c)\beta)^{n-c}} P_0 & n > c\,. \end{cases}$$

(c) Consider the $M/M/c$ with both balking and reneging. In this case, the arrival rate is $\lambda_n = \lambda, n = 0, 1, \cdots, c-1, \lambda_n = p\lambda, n = c, c+1, \cdots$. The service rate is $\mu_n = n\mu, n = 1, 2, \cdots, c$, $\mu_n = (n-c)\beta + c\mu, n = c+1, c+2, \cdots$. Hence, the steady state probability is given as

$$P_n = \begin{cases} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n P_0 & 1 \le n \le c \\ \\ \frac{p^{n-c}}{c!\mu^c c^{n-c}} \frac{\lambda^n}{(c\mu+(n-c)\beta)^{n-c}} P_0 & n > c \end{cases}$$

where P_0 is obtained from the fact that $\sum_{n=0}^{\infty} P_n = 1$. ▲

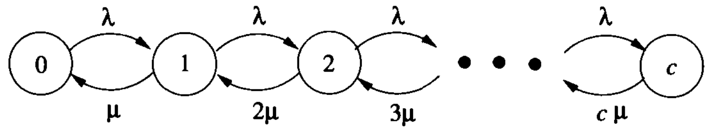


Figure 10.8 $M/M/c/c$ loss system

10.3.2 $M/M/c/c$ Loss System

This loss system is also known as the Erlang loss system. The state transition diagram for this loss system is shown in Figure 10.8. Solving the balance equations for steady state probabilities, we get

$$P_n = \frac{\rho^n/n!}{\sum_{i=0}^{c} \rho^i/i!}, \quad n = 0, 1, \cdots, c$$

where $\rho = \frac{\lambda}{\mu}$.

The above state probabilities indicate that the system state follows a truncated Poisson distribution with parameter ρ. The average number of customers in the system is given by:

$$L_s = \sum_{n=0}^{c} nP_n = \rho(1 - P_c) \ .$$

Definition 10.3 (Erlang-B Formula) *The blocking probability P_c is also called the Erlang-B formula and is given by:*

$$B(c, \rho) = P_c = \frac{\rho^c/c!}{\sum_{i=0}^{c} \rho^i/i!} \ . \tag{10.16}$$

Note 10.11 *It is a fundamental result used for telephone traffic engineering problems and can be used to select the appropriate number of trunks (servers) needed to ensure a small proportion of lost calls (customers). Observe that the steady state distribution depends on the service time distribution only through its mean (since $\rho = \lambda/\mu$). Hence, the Erlang-B formula also holds good for the $M/G/c/c$ loss system where service time follows a general distribution. The important property of $B(c, \rho)$ is illustrated in Figures 10.9 and 10.10. From Figure 10.9, for a fixed ρ, blocking probability, $B(c, \rho)$ monotonically decreases to zero as the number of servers increases. From Figure 10.10, for a fixed number of servers c, the blocking probability $B(c, \rho)$ monotonically increases to unity as ρ increases.*

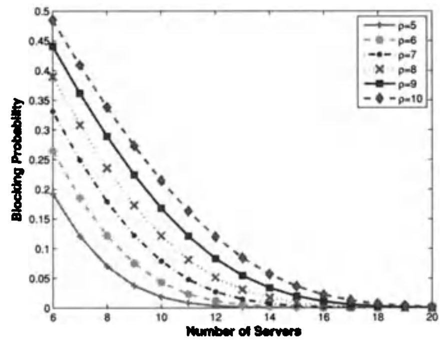


Figure 10.9 Erlang-B formula versus number of servers

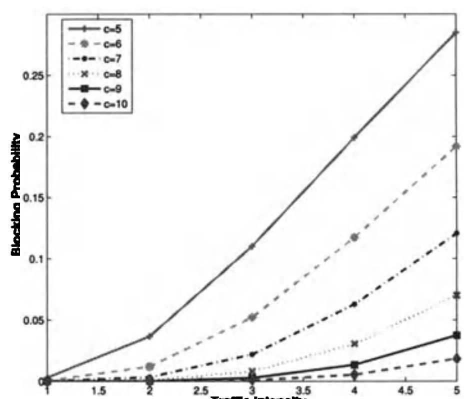


Figure 10.10 Blocking probability versus ρ

■ **EXAMPLE 10.12**

A rural telephone switch has C circuits available to carry C calls. A new call is blocked if all circuits are busy. Suppose calls have duration (in minutes) which has exponential distribution with mean 1 and interarrival time (in minutes) of calls is also exponential with mean 1. Assume that calls arrive independently. Given that $\lambda = 1$ and $\mu = 1$, so that $\rho = \frac{\lambda}{\mu} = 1$. Using Erlang-B formula given in (10.16), the probability that the system is blocked in the steady state is:

$$P_C = \frac{\frac{1}{C!}}{\sum_{i=0}^{C} \frac{1}{i!}} . \qquad \blacktriangle$$

■ **EXAMPLE 10.13**

Suppose that Spencer supermarket in Chennai has decided to construct a parking system in the supermarket. An incoming four wheeler is not allowed inside the parking system when all the lots are in use. On average 100 four wheelers arrive to the system per hour following a Poisson process and the average time that the four wheeler spends in the parking lot is 4 minutes. Enough parking lots are to be provided to ensure that the probability of the system being full does not exceed 0.005. How many lots should be constructed?

Solution: This system can be modeled as a $M/G/c/c$ system with:

$$\rho = \frac{100}{60} \times 4 = 6.67 \text{ erlangs.}$$

Using the Erlang-B formula in equation (10.16), we obtain the blocking probability

$$P_B = \begin{cases} 0.0104 & \text{when } c = 13 \\ 0.005 & \text{when } c = 14 . \end{cases}$$

The smallest c such that $P_B \leq 0.005$ is 14. Hence, we conclude that 14 lots are required. ▲

10.3.3 $M/M/c/K$ Finite-Capacity Queueing System

In this queueing model, the system has a finite capacity of size K and we assume that $c < K$. The departure rates are state dependent and are given by:

$$\mu_n = \begin{cases} n\mu, & n = 1, 2, \cdots, c-1 \\ c\mu, & n \geq c . \end{cases}$$

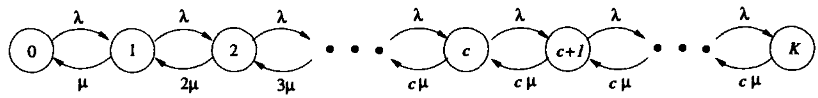


Figure 10.11 $M/M/c/K$ queueing system

The state transition diagram for this system is shown in Figure 10.11. As evaluated before, the steady state probability of the system is given by:

$$P_n = \begin{cases} \frac{1}{n!}(\frac{\lambda}{\mu})^n P_0 & 1 \leq n \leq c \\ \frac{1}{c^{n-c}c!}(\frac{\lambda}{\mu})^n P_0 & c < n \leq K \ . \end{cases}$$

Using the fact $\sum_{n=0}^{K} P_n = 1$, we get

$$P_0 = \begin{cases} \left(\sum_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c \frac{1-\rho^{K-c+1}}{1-\rho}\right)^{-1} & \text{if} \quad \rho \neq 1 \\ \left(\sum_{n=0}^{c-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{c!}\left(\frac{\lambda}{\mu}\right)^c (K-c+1)\right)^{-1} & \text{if} \quad \rho = 1 \end{cases}$$

where $\rho = \frac{\lambda}{c\mu}$.

Further, the average number of customers in the queue is given by:

$$\begin{aligned} L_q &= \sum_{n=c+1}^{K} (n-c)P_n \\ &= \frac{P_0(\lambda/\mu)^c \rho}{c!(1-\rho)^2}\left[1-\rho^{K-c+1} - (1-\rho)(K-c+1)\rho^{K-c}\right] \ . \end{aligned}$$

Using Little's formula, we can evaluate

$$L_s = L_q + \frac{\lambda_{eff}}{\mu}, \ T_s = \frac{L_s}{\lambda_{eff}}, \ T_q = T_s - \frac{1}{\mu}$$

where $\lambda_{eff} = \lambda(1 - P_K)$.

10.3.4 $M/M/\infty$ Queueing System

This is a Markovian queueing model without any queue. There are infinitely many servers such that every incoming customer finds an idle server immediately. The state transition diagram for this loss system is shown in Figure 10.12.

One can easily obtain the stationary distribution P_n as given by:

$$P_n = \frac{\rho^n e^{-\rho}}{n!}, \quad n = 0, 1, \cdots \ .$$

This follows a Poisson distribution with parameter ρ. Hence, the mean and variance of the number of customers in the steady state are the same as ρ.

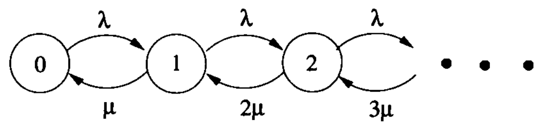


Figure 10.12 $M/M/\infty$ system

Since there is no queueing in the $M/M/\infty$ system, all waiting times are zero and the mean sojourn time in the system equals $1/\mu$. This means that all customers passing through such a system independently spend an exponentially distributed time.

One can observe that the expression of P_0 for the $M/M/\infty$ queue is the limit of the respective expression for the $M/M/c/c$ model as c tends to infinity. Further, the stationary distribution for the $M/M/c/c$ queue converges to the stationary distribution of $M/M/\infty$ for increasing c.

■ **EXAMPLE 10.14**

Consider a buffet served in a large area. The diners serve themselves. It is observed that a diner arrives every 10 seconds following a Poisson process and it takes about $\frac{1}{8}$ of a minute on average for the diner to serve himself. Find the average number of customers serving themselves and the average time spent.

Solution: According to the question, $\lambda = 6$ per minute and $\mu = 8$ per minute. Hence $\rho = \frac{6}{8} = 0.75$. The steady state probability of the system is:

$$p_n = \frac{(0.75)^n e^{-0.75}}{n!}, n = 0, 1, \cdots .$$

Hence the average number of customers serving themselves is:

$$L_s = \rho = 0.75 .$$

The average time spent in the system is:

$$T_s = 0.1250 \text{ minutes} . \quad \blacktriangle$$

10.4 NON-MARKOVIAN MODELS

In the previous sections, various characteristics for Markovian models were studied. In Markovian models, the analysis is conducted using the memoryless property of exponential distribution. In this section, we study non-Markovian models which exhibit a memoryless property only at certain time epochs. These random time epochs at which the system probabilistically restarts itself are known as regeneration time points. At regeneration time points, the memory of the elapsed time is erased. The times between the regeneration time points constitute an embedded renewal sequence. We deal with specific non-Markovian models such as $M/G/1$, $GI/M/1$, $M/G/1/N$ and $GI/M/1/N$, where G represents a general or arbitrary distribution.

10.4.1 $M/G/1$ Queueing System

Definition 10.4 ($M/G/1$ Queueing System) *This system is a type of $M/M/1/\infty$ queue where service times follow a general distribution. Assume that the interarrival time follows an exponential distribution with mean $\frac{1}{\lambda}$ and the distribution of service time is $F(t)$ with pdf $f(t)$ and mean $\frac{1}{\mu}$.*

Theorem 10.6 *Consider X_t as the number of customers in the system at time t in the $M/G/1$ queueing system. Let $\rho = \frac{\lambda}{\mu}$. Prove that, for $\rho < 1$, the generating function $V(z)$ for the steady state probabilities are given by:*

$$V(z) = \frac{(1-\rho)(z-1)f^*(\lambda(1-z))}{z - f^*(\lambda(1-z))} \ .$$

This is known as the Pollaczek-Khinchin (P-K) formula.

Proof: Observe that the state of the system after an arrival of a customer depends not only on the number of the customers X_t at that time but also on the remaining service time of the customer receiving service, if any. Also note that the state of the system after a service completion depends only on the state of the system at that time.

Let X_n be the number of customers in the system at the departure instant of the nth customer. Let t_n denote the departure instant of the nth customer. Suppose $Y_t = X_n, \quad t_n \le t < t_{n+1}$. Then, by Definition 9.49, $\{Y_t; t \ge 0\}$ will be a semi-Markov process having embedded discrete-time Markov chain (DTMC) $\{X_n, n = 0, 1, \cdots\}$.

To obtain steady state probabilities of Y_t, we need to compute the transition probabilities of the embedded DTMC $\{X_n; n = 0, 1, \cdots\}$. Since these points t_n are the regeneration points of the process $\{X_t; t \ge 0\}$, the sequence of points $\{t_n;\ n = 0, 1, \cdots\}$ forms a renewal process.

Let A_n be a random variable denoting the number of customers who arrive during the service time of the nth customer. We have:

$$X_{n+1} = \begin{cases} A_{n+1}, & X_n = 0 \\ X_n - 1 + A_{n+1}, & X_n \ge 1 \ . \end{cases}$$

Since service times of all the customers, denoted by S, have the same distribution, the distribution of A_n is the same for all n. Denote, for all n,

$$\begin{aligned} a_r &= P(A_n = r) \\ &= \int_0^\infty \frac{e^{-\lambda t}(\lambda t)^r}{r!} dF(t), \quad r = 0, 1, \cdots \ . \end{aligned} \tag{10.17}$$

Therefore:

$$\begin{aligned}
P_{ij} &= P(X_{n+1} = j \mid X_n = i) \\
&= \begin{cases} a_j & \text{if} \quad j \geq 0, i = 0 \\ a_{j-i+1} & \text{if} \quad i \geq 1, j \geq i-1 \\ 0 & \text{if} \quad i \geq 1, j < i-1 \, . \end{cases}
\end{aligned}$$

Denoting $P = [P_{ij}]$, we have:

$$P = [P_{ij}] = \begin{bmatrix} a_0 & a_1 & a_2 & \cdot & \cdot & \cdot \\ a_0 & a_1 & a_2 & \cdot & \cdot & \cdot \\ 0 & a_0 & a_1 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix} . \tag{10.18}$$

Note that $\{X_n, n = 0, 1, \cdots\}$ is an irreducible Markov chain. When $\rho = \frac{\lambda}{\mu} < 1$, the chain is positive recurrent. Hence the Markov chain is ergodic. The limiting probabilities

$$v_j = \lim_{n \to \infty} P_{ij}^{(n)}, \quad j = 0, 1, 2, \cdots ,$$

exist and are independent of the initial state i. The probability vector $v = [v_0, v_1, \cdots]$ is given as the unique solution of $v = vP$ and $\sum_j v_j = 1$. The solution can be obtained by using the generating functions of v_j's and a_j's.

Define:

$$A(z) = \sum_{j=0}^{\infty} a_j z^j \text{ and } V(z) = \sum_{j=0}^{\infty} v_j z^j$$

$$\begin{aligned}
A(z) &= \sum_{j=0}^{\infty} a_j z^j \\
&= \sum_{j=0}^{\infty} z^j \left(\int_0^{\infty} \frac{e^{-\lambda t} (\lambda t)^j}{j!} dF(t) \right) \\
&= \int_0^{\infty} e^{-t(\lambda - \lambda z)} f(t) dt \\
&= f^*(\lambda - \lambda z) \, .
\end{aligned}$$

Here, $f^*(s)$ is the Laplace transform of $f(t)$.

The expected number of arrivals is given by:

$$\begin{aligned} A'(1) &= \frac{d}{dz}A(z)\Big|_{z=1} \\ &= -\lambda\frac{d}{ds}\left(\int_0^\infty e^{-st}f(t)dt\right)\Big|_{s=0} \\ &= -\lambda\left(\int_0^\infty -tf(t)dt\right) \\ &= \frac{\lambda}{\mu} = \rho \,. \end{aligned}$$

From $v = vP$, we get:

$$v_j = v_0 a_j + \sum_{i=0}^{j+1} v_i a_{j-i+1}, \quad j = 0, 1, 2, \cdots \,.$$

Multiplying by z_j on both sides and taking the sum, we get:

$$\begin{aligned} V(z) &= v_0 \sum_{j=0}^{\infty} a_j z^j + \sum_{i=1}^{\infty}\left(\sum_{j=i-1}^{\infty} v_i a_{j-i+1} z^j\right) \\ &= v_0 A(z) + \frac{1}{z}[V(z) - v_0]A(z) \\ V(z) &= \frac{v_0 A(z)(z-1)}{z - A(z)} \,. \end{aligned}$$

Using $V(1) = A(1) = 1$,

$$\begin{aligned} V(1) &= \lim_{z\to 1} v_0 \left(\frac{A'(z)(z-1) + A(z)}{1 - A'(z)}\right) \\ 1 &= \frac{v_0 A(1)}{1 - A'(1)} = \frac{v_0}{1 - A'(1)} \end{aligned}$$

provided $A'(1)$ is finite and less than 1. Taking $\rho = A'(1) < 1$ we get $v_0 = 1-\rho$. Hence:

$$V(z) = \frac{(1-\rho)(z-1)f^*(\lambda(1-z))}{z - f^*(\lambda(1-z))} \,.$$

■

Note 10.12 *The average number of customers in the system in the steady state is given by:*

$$L_s = \frac{dV(z)}{dz}\Big|_{z=1} = \rho + \frac{\lambda^2 E(S^2)}{2(1-\rho)} \,.$$

This is known as the P-K mean value formula. Here, $E(S^2)$ is the second-order moment about the origin for the service time. This result holds true for *all scheduling disciplines in which the server is busy if the queue is nonempty. When σ_S^2 is the variance of the service time distribution, we get:*

$$L_s = \rho + \frac{\lambda^2\sigma_S^2 + \rho^2}{2(1-\rho)} \ . \tag{10.19}$$

Note 10.13 *In this derivation, we assume FIFO scheduling to simplify the analysis. However, the above formulas are valid for any scheduling discipline in which the server is busy if the queue is nonempty, no customer departs from the queue before completing service and the order of service is not dependent on the knowledge about service times. Hence, using Little's formula, the average time spent in the system can be calculated.*

Note 10.14 *Using the P-K mean value formula, other measures such as L_q, T_q and T_s can be obtained as follows:*

$$L_q = L_s - \lambda E(S), \quad E(S) = \frac{1}{\mu}$$

$$T_q = \frac{L_q}{\lambda}, \quad T_s = \frac{L_s}{\lambda} \ .$$

The formula states that mean waiting time is given by:

$$T_q = \frac{\rho(1 + \mu^2\sigma_S^2)}{2\mu(1-\rho)} \ .$$

Hence, the average time spent in the system is $T_s = T_q + \frac{1}{\mu}$, i.e.:

$$T_s = \frac{1 + \mu^2\sigma_S^2}{2(\mu - \lambda)} + \frac{1}{\mu} \ .$$

Note 10.15 ***(M/D/1 Queue)*** *When the service time is a constant, i.e., the service time follows a deterministic distribution, $\sigma_S^2 = 0$ and the P-K formula reduces to*

$$L_s = \rho + \frac{\rho^2}{2(1-\rho)}$$

where $\rho = \lambda/\mu$ and $1/\mu$ is the constant service time.

Note 10.16 ***(M/M/1 Queue)*** *When service time is exponentially distributed with mean $\frac{1}{\mu}$ and $\sigma_S^2 = \frac{1}{\mu^2}$:*

$$L_s = \rho + \frac{2\rho^2}{2(1-\rho)} = \frac{\rho}{1-\rho} \ .$$

■ **EXAMPLE 10.15**

Consider Example 10.3. Assume that customers arrive in a Poisson process at an average rate of 100 per hour. Also assume that the time taken (in seconds) for each reservation by a computer server follows a uniform distribution with parameters 20 and 30.

(a) Find the average waiting time of the customers.

(b) Find the probability that the system is empty in the long run.

Solution: The arrival rate is $\lambda = \frac{1}{36}$ and the average service time is $\frac{1}{\mu} = 25$ seconds. Hence $\rho = \frac{25}{36}$.

(a) By using the P-K mean formula and Little's formula, the average waiting time evaluates to 42.044 .

(b) The probability that the system is empty in the long run is:

$$v_0 = 1 - \rho = 0.30556. \qquad \blacktriangle$$

■ **EXAMPLE 10.16**

Consider an automatic transaction machine (ATM) with only one counter. Assume that the customers arrive according to a Poisson process with 5 customers per hour and may wait outside the counter if the ATM is busy. If the service time for the customers follows $Weibull(2, \frac{1}{10})$, determine L_s and T_q.

Solution: This problem is modeled as a $M/G/1$ queueing system with $\lambda = 5$ per hour and $1/\mu = 0.0886$. In this case, service time S follows $Weibull(2, \frac{1}{10})$. Using Theorem 4.5, we have $E(S) = 0.0886$ and $Var(S) = 0.01785$.

Using the P-K mean formula in equation (10.19), we get:

$$L_s = 1.02 \ .$$

Then, using Little's formula, we get:

$$T_q = \frac{L_q}{\lambda} = \frac{L_s - \rho}{\lambda} = 0.1135 \ . \qquad \blacktriangle$$

10.4.2 $GI/M/1$ Queueing System

The model to be studied next is the $GI/M/1$ model in which the arrivals are independent and interarrival times follow a general distribution, but service times follow a Markovian property. Assume that the *cdf* of the interarrival time is $F(t)$ with mean $\frac{1}{\lambda}$. Consider X_t as the number of customers in the system at time t. Note that the state of the system after the service completion of a customer depends not only on the state of the system X_t at that time but also on the remaining arrival time of the next customer, if any. However, the state of the system after the arrival of a customer depends only on the state of the system at that time.

Let X_n be the number of customers in the system at the arrival instant of the nth customer. Suppose t_n is the instant at which the nth customer arrives. Since these points t_n are the regeneration points of the process $\{X_t;\ t \geq 0\}$, the sequence of points $\{t_n;\ n = 0, 1, 2, \cdots\}$ forms a renewal process. Then $\{X_n; n = 0, 1, \cdots\}$ is an embedded Markov chain with state space $S' = \{1, 2, \cdots\}$. By Definition 9.52, $\{X_t; t \geq 0\}$ is a Markov regenerative process having embedded Markov chain $\{X_n; n = 0, 1, \cdots\}$.

Let A_n be the number of customers served during the interarrival time of the $(n+1)$th arrival. We then have:

$$X_{n+1} = X_n + 1 - A_n \quad \text{where} \quad A_n \leq X_n + 1, \quad X_n \geq 0.$$

Since the interarrival times are assumed to be independent and have the same distribution, the distribution of A_n is the same for all n. Denote for all n:

$$\begin{aligned} b_r &= P(A_n = r) \\ &= \int_0^\infty \frac{e^{-\mu t}(\mu t)^r}{r!} dF(t)\ , \quad r = 0, 1, 2, \cdots\ . \end{aligned}$$

Therefore:

$$\begin{aligned} P_{ij} &= P(X_{n+1} = j \mid X_n = i) \\ &= \begin{cases} b_{i+1-j} & i+1 \geq j \geq 1 \\ 1 - \sum_{k=0}^{i} b_k & j = 0, i \geq 0 \\ 0 & i+1 < j \text{ or } i < 0 \text{ or } j < 0\ . \end{cases} \end{aligned}$$

The transition probability matrix is given by:

$$P = [P_{ij}] = \begin{pmatrix} 1 - b_0 & b_0 & 0 & 0 & 0 & . & . & . \\ 1 - \sum_{k=0}^{1} b_k & b_1 & b_0 & 0 & 0 & . & . & . \\ 1 - \sum_{k=0}^{2} b_k & b_2 & b_1 & b_0 & 0 & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \end{pmatrix}.$$

When the embedded DTMC is irreducible and ergodic, the limiting probabilities that an arrival finds n in the system, denoted by π_n, $n = 0, 1, 2, \cdots$ are given as the unique solutions of

$$\pi P = \pi \quad \text{and} \quad \sum_{k=0}^{\infty} \pi_k = 1,$$

which gives:

$$\pi_0 = \sum_{j=0}^{\infty} \pi_j \left(1 - \sum_{k=0}^{j} b_k\right)$$

$$\pi_i = \sum_{k=0}^{\infty} \pi_{i+k-1} b_k \ , \quad i \geq 1 \ .$$

Denoting $z\pi_i = \pi_{i+1}$, we obtain for $i \geq 1$:

$$\pi_{i-1}(z - b_0 - zb_1 - z^2 b_2 - \cdots) = 0 \ .$$

Thus, for a nontrivial solution

$$z - \sum_{n=0}^{\infty} b_n z^n = 0, \ .$$

it becomes

$$G(z) = z$$

where $G(z)$ is the probability generating function of the $\{b_n\}$. It can be shown that $G(z) = F^*(\mu(1-z))$, where $F^*(z)$ is the Laplace-Steiltjes transform of the interarrival time *cdf* $F(t)$. Hence, the above equation may also be written as:

$$F^*(\mu(1-z)) = z \ . \tag{10.20}$$

The condition $\rho = \frac{\lambda}{\mu} < 1$ is a necessary and sufficient condition for the existence of the stationary solution (refer to Gross and Harris, 1998). When $\rho < 1$, the stationary probability of n in the system just prior to an arrival is given by

$$\pi_i = (1 - r_0) r_0^i, \quad i = 0, 1, 2, \ldots,$$

where r_0 is the positive root between 0 and 1 of equation (10.20).

Note 10.17 *It may be noted that this state distribution will be the equilibrium distribution that will be seen if the system is examined just before the instant of arrival of a customer to the system.*

Note 10.18 *Consider an arrival to the $GI/M/1$ queue. Let W be a random variable denoting the waiting time in the queue. It takes the value zero if an arriving customer finds the system to be empty on arrival (with probability π_0). If an arriving customer finds n jobs in the system (including the one currently in service), then the customer must wait for all of them to get served before his own service can begin. Note that the residual service time for the job that is currently in service also follows an exponential distribution. The mean waiting time in the queue, denoted by T_q, is given by:*

$$T_q = \sum_{n=1}^{\infty} \frac{n}{\mu} (1 - r_0) r_0^n = \frac{r_0}{\mu(1 - r_0)} \ .$$

Note that the mean results will be the same regardless of the service discipline being followed other than the FIFO discipline.

The waiting time distribution can be obtained using conditional distributions. Note that a customer who finds n customers in the system (including the one in service) on his/her arrival will encounter a random waiting time that will be equal to the sum of n independent exponentially distributed random variables. From this conditional distribution, the waiting time distribution is given by:

$$P(W \le t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 - r_0 & \text{if } t = 0 \\ 1 - r_0 e^{-\mu(1-r_0)t} & \text{if } t > 0\,. \end{cases} \tag{10.21}$$

■ **EXAMPLE 10.17**

In a mobile handset manufacturing factory, a component arrives for testing every 3 seconds. It is assumed that the time (in seconds) for testing the component is exponentially distributed with parameter 4.

(a) Find the waiting time distribution of a component in the queue.

(b) Find the probability that there are no components available for testing in the long run.

Solution: Given $\lambda = \frac{1}{3}$ and $\mu = 4$.

(a) Using equation (10.21), the probability distribution of the waiting time of a component in the queue is:

$$P(W \le t) = 1 - \frac{11}{36} e^{-44t/12}\,.$$

(b) The long-run probability that the system is empty is $\pi_0 = (1 - b_0)$ where $b_0 = \int_0^3 e^{-4t} dt = 0.25$. ▲

10.4.3 $M/G/1/N$ Queueing System

The non-Markovian queueing model to be discussed next is a finite-capacity non-Markovian queueing model. The first model in this list is the $M/G/1/N$ queueing model. This is similar to $M/G/1$, but now the system capacity is restricted to N.

The analysis of the $M/G/1/N$ queue is similar to that of the $M/G/1/\infty$ queue. But the main results of the $M/G/1/\infty$ queue will not be applicable to the $M/G/1/N$ queue. Let us examine the results for the $M/G/1/N$ queue in detail. Let X_t denote the number of customers in the system at time t. Let T_n, $n = 1, 2, \cdots$, be the random variable denoting the departure time instants of the nth customer.

The P-K formula will no longer be applicable to the $M/G/1/N$ queue as the expected number of arrivals during a service period will depend on the system size. Note that the number of customers in the $M/G/1/N$ system constitutes a Markov regenerative process $\{X_t; t \geq 0\}$ with $S = \{0, 1, \cdots, N\}$ and embedded Markov chain $\{X_{t_n^+}; n = 0, 1, \cdots\}$ where t_n is the nth customer departure instant and $X_{t_n^+} = X_n$ the corresponding number of customers left behind in the system by the departing customer. $\{X_n; n = 0, 1, \cdots\}$ is an embedded Markov chain (EMC) with state space $S' = \{0, 1, 2, \cdots, N-1\}$.

The one-step transition probability matrix is now truncated at $N-1$ since we are observing the state just after a departure. Using P given in (10.18), it is given by

$$P = [P_{ij}] = \begin{pmatrix} a_0 & a_1 & a_2 & . & . & . & 1 - \sum_{n=0}^{N-2} a_n \\ a_0 & a_1 & a_2 & . & . & . & 1 - \sum_{n=0}^{N-2} a_n \\ 0 & a_0 & a_1 & . & . & . & 1 - \sum_{n=0}^{N-3} a_n \\ 0 & 0 & a_0 & . & . & . & 1 - \sum_{n=0}^{N-4} a_n \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ 0 & 0 & 0 & . & . & . & 1 - a_0 \end{pmatrix}$$

where the a's are defined in (10.17). Assume that $\{X_n, n = 1, 2, \cdots\}$ is irreducible, aperiodic and positive recurrent. Then the steady state probability vector $v = (v_k)$ of the embedded Markov chain is given by

$$v = vP, \sum_{k \in S'} v_k = 1,$$

where $P = K(\infty)$ is the transition probability matrix of the EMC. Define the matrix $(E_{i,j}(t))_{i \in S', j \in S}$ as follows:

$$E_{i,j}(t) = P(X_t = j, T_1 > t \mid X_0 = i)$$

$$\alpha_{i,j} = E[\text{time spent in } j \text{ during } (0, T_1) \mid X_0 = i] = \int_0^\infty E_{i,j}(t)dt\ . \quad (10.22)$$

Finally, the steady state probabilities for $M/G/1/N$ are

$$\pi_j = \lim_{t \to \infty} P(X_t = j) = \frac{\sum_{k \in S'} v_k \alpha_{k,j}}{\sum_{k \in S'} v_k \beta_k}$$

where $\beta_k = \sum_{l \in S} \alpha_{k,l}$.

■ **EXAMPLE 10.18**

A machine is used for packing items. The machine can pack at most 2 items at a time. Items arrive as a Poisson process with rate $\lambda = 2$. The time required for packing an item follows a general distribution. Find the steady state probability of the system when the time required for packing follows a *Weibull*(2,1/2).

Solution: Let X_t be the number of items for packing at time t. The system can be modeled as a $M/G/1/2$ queueing system where each state represents a number of items for packing. Using the MRGP theory as defined in Section 9.4, we obtain the steady state probabilities. In this model, service completion time instants are the only regeneration time points. Hence, the time points of entering into states 0 and 1 are the only regeneration time instants. Therefore, in this case $S' = \{0, 1\}$ and $S = \{0, 1, 2\}$. When service time follows a $Weibull(2, \frac{1}{2})$:

$$F(t) = 8te^{-4t^2}, \ t \geq 0 \ .$$

Define the stochastic process $\{Y_t, t \geq 0\}$ during two successive regenerative time instants. Then $\{Y_t; t \geq 0\}$ is a CTMC which is known as the subordinated CTMC of the Markov regenerative process $\{X_t; t \geq 0\}$. Let $p_{ij}(t)$ be the probability that $Y_t = j$ given that $Y_0 = i$. We first obtain the global and local kernel. The global kernel is defined as

$$K(t) = [k_{ij}(t)], i, j \in S',$$

where

$$\begin{aligned} k_{ij}(t) &= P(X_1 = j, \ T_1 \leq t \mid X_0 = i), \ \ i, j \in S' \\ &= \begin{cases} \displaystyle\int_0^t (p_{00}(x) + p_{01}(x))\, dF(x), & i, j = 0 \\ \displaystyle\int_0^t p_{ij+1}(x) dF(x), & i, j \in S' \end{cases} \end{aligned}$$

and the local kernel is given by $E(t) = [e_{ij}(t)]_{i \in S', \ j \in S}$, where:

$$e_{ij}(t) = P(X_1 = j, \ T_1 \geq t \mid X_0 = i) = p_{ij}(t)(1 - F(t)), \ i \in S', \ j \in S \ .$$

Here $F(t)$ is the cumulative distribution function of the general distribution and $p_{ij}(t)$ are the time-dependent probabilities of the subordinated CTMC and are evaluated as:

$$P(t) = \begin{pmatrix} e^{-\lambda t} & \lambda t e^{-\lambda t} & 1 - e^{-\lambda t} - \lambda t e^{-\lambda t} \\ 0 & e^{-\lambda t} & 1 - e^{-\lambda t} \\ 0 & 0 & 1 \end{pmatrix} .$$

With this, the elements of $K(t)$ are evaluated as:

$$\begin{aligned}
k_{00}(t) &= \int_0^t 8x\,(\lambda x+1)\,e^{-\lambda x-4x^2}dx \\
k_{01}(t) &= \int_0^t 8x\left(1-e^{-\lambda x}-\lambda x e^{-\lambda x}\right)e^{-4x^2}dx \\
k_{10}(t) &= \int_0^t 8xe^{-\lambda x-4x^2}dx \\
k_{11}(t) &= \int_0^t 8x(1-e^{-\lambda x})e^{-4x^2}dx\ .
\end{aligned}$$

Substituting $\lambda = 2$, we get:

$$\begin{aligned}
\lim_{t\to\infty} k_{00}(t) &= 0.7728 \\
\lim_{t\to\infty} k_{01}(t) &= 0.2272 \\
\lim_{t\to\infty} k_{10}(t) &= 0.4544 \\
\lim_{t\to\infty} k_{11}(t) &= 0.5456\ .
\end{aligned}$$

Then the steady state probability vector $v = (v_k)$ of the EMC is

$$v = vP,\ \sum_{k\in S'} v_k = 1,$$

where $P = K(\infty)$ is the transition probability matrix of the EMC. Solving, we get $v_0 = v_1 = 1/2$.

Now, the matrix $E = [e_{i,j}(t), i \in S', j \in S]$:

$$\begin{aligned}
e_{00}(t) &= e^{-\lambda t-4t^2} \\
e_{01}(t) &= \lambda te^{-\lambda t-4t^2} \\
e_{02}(t) &= \left(1-e^{-\lambda t}-\lambda te^{-\lambda t}\right)e^{-4t^2} \\
e_{10}(t) &= 0 \\
e_{11}(t) &= e^{-\lambda t-4t^2} \\
e_{12}(t) &= (1-e^{-\lambda t})e^{-4t^2}\ .
\end{aligned}$$

Substituting $\lambda = 2$ and using $\alpha_{i,j} = \int_0^\infty e_{i,j}(t)dt$, we get:

$$\begin{aligned}
\alpha_{00} &= 0.2728 \\
\alpha_{01} &= 0.1136 \\
\alpha_{02} &= 0.0567 \\
\alpha_{10} &= 0 \\
\alpha_{11} &= 0.2728 \\
\alpha_{12} &= 0.1703\ .
\end{aligned}$$

Finally, we compute the steady state probabilities for the $M/G/1/N$ as given below:

$$\pi_j = \frac{\sum_{k\in S'} v_k \alpha_{k,j}}{\sum_{k\in S'} v_k \beta_k}$$

where $\beta_0 = \beta_1 = \sqrt{2\pi}/8$. Hence:

$$\pi_0 = 0.3078, \quad \pi_1 = 0.436, \quad \pi_2 = 0.2562 \ . \quad \blacktriangle$$

10.4.4 $GI/M/1/N$ Queueing System

This queueing model is similar to $GI/M/1$, but now the system capacity is restricted to N. Let X_t denote the number of customers in the system at time t. Let T_n, $n = 1, 2, \cdots$, be the random variable denoting the arrival time instant of the nth customer. Here, the arrival time instants are the only regeneration time epochs. Hence, $\{X_t; t \geq 0\}$ is not a semi-Markov process, but a MRGP with $\{X_n; n = 1, 2, \cdots\}$ is an embedded Markov chain. $S = \{0, 1, 2, \cdots, N\}$ is the set of states at all time instants, and $S' = \{1, 2, \cdots, N\}$ is the set of states only at regeneration time instants. Following the theory of MRGP, we now proceed to determine the global kernel $K(t)$ and local kernel $E(t)$ matrices for the process. The elements of the global and local kernel are defined as:

$$k_{ij}(t) = P(X_1 = j,\ T_1 \leq t \mid X_0 = i), \quad i, j \in S'$$

$$e_{ij}(t) = P(X_1 = j,\ T_1 \geq t \mid X_0 = i), \quad i \in S',\ j \in S\ .$$

The global kernel matrix is

$$K(t) = \begin{pmatrix} k_{11}(t) & k_{12}(t) & 0 & 0 & . & . & 0 \\ k_{21}(t) & k_{22}(t) & k_{23}(t) & 0 & . & . & 0 \\ k_{31}(t) & k_{32}(t) & k_{33}(t) & k_{34}(t) & . & . & 0 \\ k_{41}(t) & k_{42}(t) & k_{43}(t) & k_{44}(t) & . & . & 0 \\ . & . & . & . & . & . & . \\ . & . & . & . & . & . & . \\ k_{N1}(t) & k_{N2}(t) & k_{N3}(t) & . & & . & . & k_{NN}(t) \end{pmatrix}$$

where

$$k_{ij}(t) = \begin{cases} \displaystyle\int_0^t p_{ij-1}(x) dF(x)\ , & i, j \in S' \\ \displaystyle\int_0^t \left(p_{NN-1}(x) + p_{NN}(x)\right) dF(x)\ , & i, j = N \end{cases}$$

and the p_{ij}'s are the transition probabilities of the subordinated CTMC. The state transition diagram for the subordinated CTMC is shown in Figure 10.13. The local kernel matrix is

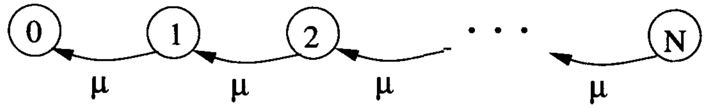


Figure 10.13 Subordinated CTMC

$$E(t) = \begin{pmatrix} e_{10}(t) & e_{11}(t) & 0 & 0 & 0 & . & . & 0 \\ e_{20}(t) & e_{21}(t) & e_{22}(t) & 0 & 0 & . & . & 0 \\ e_{30}(t) & e_{31}(t) & e_{32}(t) & e_{33}(t) & 0 & . & . & 0 \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ e_{N0}(t) & e_{N1}(t) & e_{N2}(t) & e_{N3}(t) & e_{N4}(t) & . & . & e_{NN}(t) \end{pmatrix}$$

where $e_{ij}(t) = p_{ij}(1 - F(t))$.

The limiting behavior is obtained by taking the limit as t approaches infinity. We require two new process characteristics to be defined, i.e., α_{ij}, the mean time spent by the MRGP in state j between two successive regeneration instants, given that it started in state i after the last regeneration,

$$\alpha_{ij} = E[\text{time in } j \text{ during } (0, T_1) \mid X_0 = i] = \int_0^\infty e_{ij}(t)dt,$$

and the steady state probability vector $v = [v_k]$ of the EMC,

$$v = vP \text{ and } \sum_{k=1}^{N} v_k = 1$$

where $P = K(\infty)$ is the transition probability matrix of the embedded Markov chain. Then the steady state probability of the MRGP is given by

$$\pi_{ij} = \frac{\sum_{k=1}^{N} v_k \alpha_{kj}}{\sum_{k=1}^{N} v_k \beta_k}$$

where $\beta = \sum_{l=0}^{N} \alpha_{kl}$.

Now, let us discuss the time-dependent behavior of the $G/M/1/N$ queueing system. Define $V(t) = (V_{ij}(t))$ as the matrix of conditional state probabilities of the MRGP and the following equations for $V(t)$ are matrix equations and

$$V_{ij}(t) = P\{X_t = j \mid X_0 = Y_0 = i\}, \quad i, j \in S,$$

where $\{Y_n,\ n \geq 0\}$ is the embedded Markov chain of the MRGP $\{X_t;\ t \geq 0\}$. Also,

$$\begin{aligned} V(t) &= E(t) + \int_0^t dK(s)V(t-s) \\ &= E(t) + K(t) * V(t) \end{aligned}$$

where $*$ denotes the convolution operator. The solution method for the time-dependent state distribution is outlined below:

1. Calculate $K(t)$ and $E(t)$.

2. Compute $\tilde{K}(s)$ and $\tilde{E}(s)$ where $\tilde{K}(s)$ and $\tilde{E}(s)$ are the Laplace-Steiltjes transform obtained as:

$$\tilde{K}(s) = \int_0^{\infty} e^{-st} dK(t) \text{ and } \tilde{E}(s) = \int_0^{\infty} e^{-st} dE(t) .$$

3. Solve the following linear system for $\tilde{V}(s)$:

$$[I - \tilde{K}(s)]\tilde{V}(s) = \tilde{E}(s)$$

where

$$\tilde{V}(s) = \int_0^{\infty} e^{-st} dV(t) .$$

4. Using inverse Laplace transforms, invert $\tilde{V}(s)$ to obtain $V(t)$.

5. Using $p(t)_{1\times S} = p(0)_{1\times S'}\ V(t)_{S'\times S}$, obtain the time-dependent state probabilities of the MRGP.

■ EXAMPLE 10.19

Consider a system of two communicating satellites. The lifetime of a satellite follows an exponential distribution with parameter λ whereas the time to repair and resend the satellite follows a general distribution with mean $1/\mu$. Our interest is to determine the steady state probabilities when time to repair and resend the satellite follows a *Weibull*(2,1/2) distribution. Next we analyze a special case of the $GI/M/1/N$ queue with $N = 2$. In this queueing model the arrival time instants are the only regeneration time instants. Hence, the time points of entering into the states 1 and 2 are the only regeneration time instants. Therefore, in this case $S' = \{1, 2\}$ and $S = \{0, 1, 2\}$. We solve this MRGP by taking F as $Weibull(2, \frac{1}{2})$. Here interarrival time is $Weibull(2, \frac{1}{2})$. Hence:

$$F(t) = 8te^{-4t^2},\ t \geq 0 .$$

Proceeding as above, we first evaluate the global and local kernel for this MRGP. The global kernel is defined as

$$K(t) = [k_{ij}(t)],\ i,\ j \in S',$$

where

$$k_{ij}(t) = \begin{cases} \displaystyle\int_0^t p_{ij-1}(x)dT(x)\ , & (i,j) = (1,1),(1,2),(2,1) \\ \displaystyle\int_0^t (p_{21}(x) + p_{22})(x)dT(x)\ , & (i,j) = (2,2) \end{cases}$$

and the local kernel is given by

$$E(t) = [e_{ij}(t)]_{\in S',\ j\in S}\ ,$$

where

$$e_{ij}(t) = p_{ij}(t)(1 - F(t))\ .$$

Here $F(t)$ is the cumulative distribution function of the Weibull distribution and $p_{ij}(t)$ are the time-dependent probabilities of the subordinated CTMC and are evaluated as:

$$P(t) = \begin{pmatrix} 1 & 0 & 0 \\ e^{-\mu t} & 1 - e^{-\mu t} & 0 \\ e^{-\mu t} & \mu t e^{-\mu t} & 1 - e^{-\mu t} - \mu t e^{-\mu t} \end{pmatrix}\ .$$

With this, the elements of $K(t)$ are evaluated as:

$$\begin{aligned} k_{11}(t) &= \int_0^t (1 - e^{\mu x})\, 8xe^{4x^2} dx \\ &= e^{-4t^2 - \mu t} - e^{-4t^2} + \mu \int_0^t e^{-4x^2 - \mu x} dx \\ k_{12}(t) &= \int_0^t 8xe^{-\mu x - 4x^2} dx \\ k_{21}(t) &= \int_0^t \left(1 - e^{-\mu x} - \mu x e^{-\mu x}\right) 8xe^{-4x^2} dx \\ &= \left(1 - e^{4t^2}\right) - \int_0^t \left(e^{-\mu x} + \mu x e^{-\mu x}\right) 8xe^{4x^2} dx \\ k_{22}(t) &= \int_0^t 8\mu x^2 e^{-\mu x - 4x^2} dx + \int_0^t 8xe^{-\mu x - 4x^2} dx\ . \end{aligned}$$

Substituting $\mu = 2$, we get:

$$\begin{aligned} \lim_{t\to\infty} k_{11}(t) &= 0.5456 \\ \lim_{t\to\infty} k_{12}(t) &= 0.4544 \\ \lim_{t\to\infty} k_{21}(t) &= 0.2172 \\ \lim_{t\to\infty} k_{22}(t) &= 0.7728\ . \end{aligned}$$

Then the steady state probability vector $v = (v_k)$ of the EMC is

$$v = vP, \sum_{k \in S'} v_k = 1,$$

where $P = K(\infty)$ is the transition probability matrix of the EMC. Solving, we get $v_0 = v_1 = 1/2$. Now, the matrix $E = [e_{i,j}(t), i \in S', j \in S]$:

$$\begin{aligned}
e_{10}(t) &= \left(1 - e^{-\mu t}\right) e^{-4t^2} \\
e_{11}(t) &= e^{-\mu t - 4t^2} \\
e_{12}(t) &= 0 \\
e_{20}(t) &= \left(1 - e^{-\mu t} - \mu t e^{-\mu t}\right) e^{-4t^2} \\
e_{21}(t) &= \mu t e^{-\mu t - 4t^2} \\
e_{22}(t) &= e^{-\mu t - 4t^2} \ .
\end{aligned}$$

Substituting $\mu = 2$ and using $\alpha_{i,j} = \int_0^\infty e_{i,j}(t)dt$, we get:

$$\begin{aligned}
\alpha_{10} &= 0.1703 \\
\alpha_{11} &= 0.2728 \\
\alpha_{12} &= 0 \\
\alpha_{20} &= 0.0567 \\
\alpha_{21} &= 0.1136 \\
\alpha_{22} &= 0.2728 \ .
\end{aligned}$$

The steady state probabilities for the $GI/M/1/2$ are given as

$$\pi_j = \frac{\sum_{k \in S'} v_k \alpha_{k,j}}{\sum_{k \in S'} v_k \beta_k}, \quad j \in S \ ,$$

where $\beta_0 = \beta_1 = \sqrt{2\pi}/8$. Hence:

$$\pi_0 = 0.2561, \quad \pi_1 = 0.436, \quad \pi_2 = 0.3078 \ . \qquad \blacktriangle$$

Queueing modeling is an important tool used to evaluate system performance in communication networks. It has a wide spectrum of applications in science and engineering. In this chapter we have studied some simple queueing models. For advanced queueing models such as non-Markovian, bulk and priority queues and queueing networks, the reader may refer to Gross and Harris (1998), Medhi (2003), Bhat (2008), and Bolch et al. (2006).

EXERCISES

10.1 Find the service rate for a $M/M/1$ queue where customers arrive at a rate of 3 per minute, given that 95% of the time the queue contains less than 10 customers.

10.2 The arrival of a patient at a doctor's clinic follows a Poisson process with rate 8 per hour. The time taken by the doctor to examine a patient is exponential distribution with mean 6 minutes. Given that no patients are returned, find:

a) Probability that the patient has to wait on arrival.

b) Expected total time spent (including the service time) by any visiting patient.

10.3 It is assumed that the duration (in minutes) of a telephone conversation at a telephone booth consisting of only one telephone follows an exponential distribution with parameter $\mu = \frac{1}{4}$. If a person arrives at the telephone booth 3 minutes after a call started, then find his expected waiting time.

10.4 There are two servers at a service center providing service at an exponential rate of two services per hour. If the arrival rate of customers is 3 per hour and the system capacity is at most 3, then:

a) Find the fraction of potential customers that enter the system?

b) If there was only a single server with service rate twice as fast, i.e., μ=4, then what will be the value of part (a)?

10.5 There are N spaces in a parking lot. Traffic arrives in a Poisson process with rate λ, but only as long as empty spaces are available. The occupancy times have an exponential distribution with mean $1/\mu$. If X_t denotes the number of occupied parking spaces at any time t, then:

a) Determine infinitesimal generator matrix Q and the forward Kolmogorov equations for the Markov process $\{X_t; t \geq 0\}$.

b) Determine the limiting probability distribution of the stochastic process $\{X_t; t \geq 0\}$.

10.6 An automobile emission inspection station has three inspection stalls, each with room for only one car. It is assumed that when a stall becomes vacant, the car standing first in the waiting line pulls up to it.Only four cars are allowed to wait at a time (seven in the station). Arrival occurs in a Poisson way with mean of one car per minute during the peak periods. The service time is exponential with mean 6 minutes.

1. Determine the average number of cars in the system during peak periods.

2. Determine the average waiting time (including service).

10.7 Show that the average time spent in a $M/M/1$ system with arrival rate λ and service rate 2μ is lesser than the average time spent in a $M/M/2$ system with arrival rate λ and each having service rate μ which is lesser than the average time spent in two independent $M/M/1$ queues with each having arrival rate $\lambda/2$ and equal service rate μ.

10.8 Suppose that you arrive at an ATM to find seven others, one being served (first-come-first-service basis) and the other six waiting in line. You join the end of the line. Assume that service times are independent and exponentially distributed with rate μ.

a) Model this situation as a birth-and-death process.

b) What is the expected amount of time you will spend in the ATM?

10.9 A telephone switching system consists of c trunks with an infinite caller population. Calls arrive in a Poisson process with rate λ and each call holding time is exponentially distributed with average $1/\mu$. Let $\rho = \lambda/\mu$. It is assumed that an incoming call is lost if all the trunks are busy.

a) Draw the state transition diagram for the system.

b) Derive an expression for π_n, the steady state probability that n trunks are busy.

c) Also find the number of trunks n, such that $\pi_n \leq 0.001$.

10.10 For the $M/M/1/N$ queueing system, show that as limit $N \to \infty$ there are two possibilities: either $\rho < 1$ and P converges to the stationary distribution of the $M/M/1$ queue or $\rho \geq 1$ and P goes to infinity.

10.11 Consider a telephone switching system with N subscribers. Each subscriber can attempt a call from time 0. Assume that there are c $(< N)$ channels in the telephone switching system to handle the calls and assume that each call needs one channel. Assume that the arrival of a call from each customer follows a Poisson process with rate λ and each call duration follows an independent exponential distribution with rate μ.

a) Draw the state transition diagram for this queueing system.

b) Suppose that $c = 10$. Find λ such that no call is waiting.

10.12 For an $M/M/c/\infty$ queue, prove that the distribution of waiting time in the queue is given by

$$P[w_q = 0] = 1 - \frac{\sum_{n=c}^{\infty} P_n}{1 - \rho}$$

and

$$f_{W_q}(t) = c\mu\ P_c\ e^{-(c\mu-\lambda)t},\quad 0 < t < \infty\ .$$

10.13 A toll bridge with c booths at the entrance can be modeled as a c server queue with infinite capacity. Assuming the service times are independent exponential random variables with mean 1 second, sketch the state transition diagram for a continuous-time Markov chain for the system. Find the limiting state probabilities. What is the maximum arrival rate such that the limiting state probabilities exist?

10.14 Assume that taxis are waiting outside a station, in a queue, for passengers to come. Passengers for these taxis arrive according to a Poisson process with an average of 60 passengers per hour. A taxi departs as soon as two passengers have been collected or 3 minutes have expired since the first passenger has got in the taxi. Suppose you get in the taxi as the first passenger. What is your average waiting time for the departure?

10.15 Consider a Markovian queueing model with finite population, say N. In reliability theory, this model is known as the machine repair problem. In this case, customers are treated as N machines which are prone to failure. The lifetime of each machine is assumed to be exponentially distributed with parameter λ. There are c repairmen who repair the broken machines sequentially. The repair times are assumed to be exponential distributions with parameter μ. Obtain the steady state probabilities for this system.

10.16 For an $M/M/1/N$ queueing system, show that the waiting time distribution $W(t)$ is

$$P(W \le t) = 1 - \sum_{n=0}^{N-1} \frac{P_n}{1-P_N} e^{-\mu t} \sum_{k=0}^{n} \frac{(\mu t)^k}{k!}$$

where $P_n, n = 0, 1, \cdots, N-1$, is the steady state probability.

10.17 Find the stationary distribution for a $M/M/c/c/K$ loss system with K population.

10.18 Consider a $M/E_k/1$ queueing system with arrival rate λ and mean service time $\frac{1}{\mu}$. Let $\rho = \frac{\lambda}{\mu}$. Then show that the probability generating function of the distribution of the system state in equilibrium is given by:

$$G(z) = \frac{(1-\rho)(1-z)}{1 - z\left[1 + \frac{\rho(1-z)}{k}\right]^k}\ .$$

CHAPTER 11

STOCHASTIC CALCULUS

Stochastic calculus plays an essential role in modern mathematical finance and risk management. The objective of this chapter is to develop conceptual ideas of stochastic calculus in order to provide a motivational framework. This chapter presents an informal introduction to martingales, Brownian motion, and stochastic calculus. Martingales were first defined by Paul Lévy (1886–1971). The mathematical theory of martingales has been developed by American mathematician Joseph Doob (1910–2004). We begin with the basic notions of martingales and its properties.

11.1 MARTINGALES

The martingale is a strategy in a roulette game in which, if a player loses a round of play, then he doubles his bet in the following games so that if he wins he would recover from his previous losses. Since it is true that a large losing sequence is a rare event, if the player continues to play, it is possible for the player to win, and thus this is apparently a good strategy. However, the player could run out of funds as the game progresses, and therefore the player

Introduction to Probability and Stochastic Processes with Applications, First Edition.
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

cannot recover the losses he has previously accumulated. One must also take into account the fact that casinos impose betting limits.

Formally, suppose that a player starts a game in which he wins or loses with the same probability of $\frac{1}{2}$. The player starts betting a single monetary unit. The strategy is progressive where the player doubles his bet after each loss in order to recoup the loses. A possible outcome for the game would be the following:

Bet	1	2	4	8	16	1	1
Outcome	F	F	F	F	W	W	F
Profit	-1	-3	-7	-15	1	2	1

Here W denotes "Win" and F denotes "Failure". This shows that every time the player wins, he recovers all the previous losses and it is also possible to increase his wealth to one monetary unit. Moreover, if he loses the first n bets and wins the $(n+1)$th, then his wealth after the nth bet is equal to:

$$-1-2-2^2-\cdots-2^{n-1}+2^n=-\sum_{k=0}^{n-1}2^k+2^n=1.$$

This would indicate a win for the player. Nevertheless, as we shall see later, to carry out this betting strategy successfully, the player would need on average infinite wealth and he would have to bet infinitely often (Rincón, 2011).

In probability theory, the notion of a martingale describes a fair game. Suppose that the random variable X_m denotes the wealth of a player in the mth round of the game and the σ-field $\Im_m$ has all the knowledge of the game at the mth round. The expectation of X_n (with $n \geq m$), given the information in $\Im_m$, is equal to the fortune of the player up to time m. Then the game is fair. Using probability terms, we have, with probability 1:

$$E\left(X_n \mid \Im_m\right)=X_m \text{ for all } m \leq n.$$

A stochastic process $\{X_t; t \geq 0\}$ satisfying the above equation is called a *discrete-time martingale*. Formally we have the following definitions:

Definition 11.1 *Let $(\Omega, \Im, P)$ be a probability space. A filtration is a collection of sub-σ-algebras $(\Im_n)_{n\geq 0}$ of $\Im$ such that $\Im_m \subseteq \Im_n$ for all $m \leq n$. We say that the sequence $\{X_n; n \geq 0\}$ is adapted to the filtration $(\Im_n)_{n\geq 0}$ if for each n the random variable X_n is $\Im_n$-measurable, that is, $\{\omega \in \Omega : X_n(\omega) \leq a\} \in \Im_n$ for all $a \in \mathbb{R}$.*

Definition 11.2 *Let $\{X_n; n \geq 0\}$ be a sequence of random variables defined on the probability space $(\Omega, \Im, P)$ and $(\Im_n)_{n\geq 0}$ be a filtration in $\Im$. Suppose that $\{X_n; n \geq 0\}$ is adapted to the filtration $(\Im_n)_{n\geq 0}$ and $E(X_n)$ exists for all n. We say that:*

(a) *$\{X_n; n \geq 0\}$ is a $(\Im_n)_n$-martingale if and only if $E\left(X_n \mid \Im_m\right) = X_m$ a.s. for all $m \leq n$.*

(b) $\{X_n; n \geq 0\}$ *is a* $(\Im_n)_n$*-submartingale if and only if* $E(X_n \mid \Im_m) \geq X_m$ *a.s. for all* $m \leq n$.

(c) $\{X_n; n \geq 0\}$ *is a* $(\Im_n)_n$*-supermartingale if and only if* $E(X_n \mid \Im_m) \leq X_m$ *a.s. for all* $m \leq n$.

Note 11.1 *The sequence* $\{X_n; n \geq 0\}$ *is obviously adapted to the canonical filtration or natural filtration. That is to say that the filtration* $(\Im_n)_{n \geq 0}$ *is given by* $\Im_n = \sigma(X_1, X_2, \cdots, X_n)$*, where* $\sigma(X_1, X_2, \cdots, X_n)$ *is the smallest* σ*-algebra with respect to which the random variables* $X_1, X_2, \ldots, X_n$ *are* $\Im_n$*-measurable. When we speak of martingales, supermartingales and submartingales, with respect to the canonical filtration, we will not explicitly mention it. In other words, if we say: "*$(X_n)_n$ *is a (sub-, super-) martingale" and we do not reference the filtration, it is assumed that the filtration is the canonical filtration.*

Note 11.2 *If* $\{X_n; n \geq 0\}$ *is a* $(\Im_n)_n$*-martingale, it is enough to see that:*

$$E(X_{n+1} \mid \Im_n) = X_n \text{ for all } n \in \mathbb{N}.$$

Note 11.3 *If* $\{X_n; n \geq 0\}$ *is a* $(\Im_n)_n$*-submartingale, then* $\{-X_n; n \geq 0\}$ *is a* $(\Im_n)_n$*-supermartingale. Thus, in general, with very few modifications, every proof made for submartingales is also valid for supermartingales and vice versa.*

■ EXAMPLE 11.1

Let $\{X_n; n \geq 0\}$ be a martingale with respect to $(\Im_n)_{n \geq 0}$ and $(\mathcal{G}_n)_{n \geq 0}$ be a filtration such that $\mathcal{G}_n \subseteq \Im_n$ for all n. If X_n is $\Im_n$-measurable, then $\{X_n; n \geq 0\}$ is a martingale with respect to $(\mathcal{G}_n)_n$. Indeed:

$$\begin{aligned} E(X_{n+1} \mid \mathcal{G}_n) &= E(E(X_{n+1} \mid \Im_n) \mid \mathcal{G}_n) \\ &= E(X_n \mid \mathcal{G}_n) \\ &= X_n. \end{aligned}$$

Therefore, every $(\mathcal{G}_n)_n$-martingale is a martingale with respect to the canonical filtration. ▲

■ EXAMPLE 11.2 Random Walk Martingale

Let $Z_1, Z_2, \cdots$ be a sequence of i.i.d. random variables on a probability space $(\Omega, \Im, P)$ with finite mean $\mu = E(Z_1)$, and let $\Im_n = \sigma(Z_1, \cdots, Z_n)$,

$n \geq 1$. Let $X_n = Z_1 + \cdots + Z_n$, $n \geq 1$. Then, for all $n \geq 1$,

$$\begin{aligned} E\left(X_{n+1} \mid \Im_n\right) &= E\left(X_n + Z_{n+1} \mid \Im_n\right) \\ &= E\left(X_n \mid \Im_n\right) + E\left(Z_{n+1} \mid \Im_n\right) \\ &= X_n + E\left(Z_{n+1}\right) \\ &= X_n + \mu \end{aligned}$$

so that:

$$\begin{aligned} E\left(X_{n+1} \mid \Im_n\right) &= X_n \quad \text{if} \quad \mu = 0 \\ &> X_n \quad \text{if} \quad \mu > 0 \\ &< X_n \quad \text{if} \quad \mu < 0. \end{aligned}$$

Thus, $\{X_n; n \geq 1\}$ is a martingale if $\mu = 0$, a submartingale if $\mu > 0$ and a supermartingale if $\mu < 0$. ▲

■ EXAMPLE 11.3 Second-Moment Martingale

Let $Z_1, Z_2, \cdots$ be a sequence of i.i.d. random variables on a probability space $(\Omega, \Im, P)$ with finite mean $\mu = E(Z_1)$ and variance $\sigma^2 = Var(Z_1)$. Let $\Im_n = \sigma(Z_1, \cdots, Z_n)$, $n \geq 1$. Let $Y_n = \sum_{i=1}^{n} (Z_i - \mu)^2$ and $\widetilde{Y}_n = Y_n - n\sigma^2$. It is easily verified that $\{Y_n; n \geq 1\}$ is a submartingale and $\left\{\widetilde{Y}_n; n \geq 1\right\}$ is a martingale. Assume:

$$\begin{aligned} E\left(\widetilde{Y}_{n+1} \mid \Im_n\right) &= E\left(Y_{n+1} - (n+1)\sigma^2 \mid \Im_n\right) \\ &= E\left(Y_n + (Z_{n+1} - \mu)^2 \mid \Im_n\right) - (n+1)\sigma^2 \\ &= E\left(Y_n \mid \Im_n\right) + E\left((Z_{n+1} - \mu)^2\right) - (n+1)\sigma^2 \\ &= Y_n - n\sigma^2 = \widetilde{Y}_n. \quad ▲ \end{aligned}$$

■ EXAMPLE 11.4

Let $X_1, X_2, \cdots$ be a sequence of independent random variables with $E(X_n) = 1$ for all n. Let $\{Y_n; n \geq 1\}$ be:

$$Y_n := \prod_{i=1}^{n} X_i.$$

If $\Im_n = \sigma(X_1, \cdots, X_n)$, it is clear that:

$$\begin{aligned} E(Y_{n+1} \mid \Im_n) &= E(Y_n X_{n+1} \mid \Im_n) \\ &= Y_n E(X_{n+1} \mid \Im_n) \\ &= Y_n E(X_{n+1}) = Y_n. \end{aligned}$$

That is, $\{Y_n; n \geq 1\}$ is a martingale with respect to $(\Im_n)_n$. ▲

■ EXAMPLE 11.5 Polya Urn Model

Suppose that an urn has one red ball and one black ball. A ball is drawn at random from the urn and is returned along with a ball of the same color. The procedure is repeated many times. Let X_n denote the number of black balls in the urn after n drawings. Then $X_0 = 1$ and $\{X_n; n \geq 0\}$ is a Markov chain with transitions

$$\begin{aligned} P(X_{n+1} = k + 1 \mid X_n = k) &= \frac{k}{n+2} \\ P(X_{n+1} = k \mid X_n = k) &= \frac{n+2-k}{n+2} \end{aligned}$$

and

$$E(X_{n+1} \mid \Im_n) = X_n + \frac{X_n}{n+2}.$$

Let $M_n = \frac{X_n}{n+2}$ be the proportion of black balls after n drawings. Then $\{M_n; n \geq 0\}$ is a martingale, since:

$$\begin{aligned} E(M_{n+1} \mid \Im_n) &= E\left(\frac{X_{n+1}}{n+3} \mid \Im_n\right) \\ &= \frac{1}{n+3}\left(X_n + \frac{X_n}{n+2}\right) \\ &= \frac{X_n}{n+2} = M_n. \end{aligned}$$ ▲

■ EXAMPLE 11.6 Doob's Martingale

Let X be a random variable with $E(|X|) < \infty$, and let $\{\Im_n\}_{n\geq 1}$ be a filtration. Define $X_n = E(X \mid \Im_n)$ for $n \geq 1$. Then $\{X_n, n \geq 0\}$ is a martingale with respect to $\{\Im_n\}_{n\geq 0}$:

$$\begin{aligned} E(|X_n|) &= E(|E(X \mid \Im_n)|) \\ &\leq E(E(|X|)) \\ &= E(|X|) < \infty. \end{aligned}$$

Also,

$$\begin{aligned} E\left(X_{n+1} \mid \Im_n\right) &= E\left(E\left(X \mid \Im_{n+1}\right) \mid \Im_n\right) \\ &= E\left(X \mid \Im_n\right) \\ &= X_n . \quad \blacktriangle \end{aligned}$$

As we know that every martingale is also a submartingale and a supermartingale, the following theorem provides a method for getting a submartingale from a martingale.

Theorem 11.1 *Let $\{M_n; n \geq 0\}$ be a martingale with respect to the filtration $(\Im_n)_{n\geq 0}$. If $\phi(\cdot)$ is a convex function with $E(|\phi(M_n)|) < \infty$ for all n, then $\{\phi(M_n); n \geq 0\}$ is a submartingale.*

Proof: By Jensen's inequality (Jacod and Protter, 2004):

$$\begin{aligned} E\left(\phi\left(M_{n+1}\right) \mid \Im_n\right) &\geq \phi\left(E\left(M_{n+1} \mid \Im_n\right)\right) \\ &\geq \phi\left(M_n\right) . \end{aligned}$$

■

■ EXAMPLE 11.7

Let $\{M_n; n \geq 0\}$ be a nonnegative martingale with respect to the filtration $(\Im_n)_{n\geq 0}$. Then $\{M_n^2; n \geq 0\}$ and $\{-\log M_n; n \geq 0\}$ are submartingales. ▲

■ EXAMPLE 11.8

Let $\{Y_n; n \geq 1\}$ be an arbitrary collection of random variables with $E[|Y_n|] < \infty$ for all $n \geq 1$. Let $\Im_n = \sigma\left(Y_1, \cdots, Y_n\right)$, $n \geq 1$. For $n \geq 1$, define

$$X_n = \sum_{j=1}^{n} \left[Y_j - E\left(Y_j \mid \Im_{j-1}\right)\right] \tag{11.1}$$

where $\Im_0 = \{\emptyset, \Omega\}$. Then, for each $n \geq 1$, X_n is $\Im_n$-measurable with $E[|X_n|] < \infty$. Also, for $n \geq 1$:

$$\begin{aligned} E\left(X_{n+1} \mid \Im_n\right) &= \sum_{j=1}^{n+1} E\left(\left[Y_j - E\left(Y_j \mid \Im_{j-1}\right)\right] \mid \Im_n\right) \\ &= \sum_{j=1}^{n} \left[Y_j - E\left(Y_j \mid \Im_{j-1}\right)\right] \\ &\quad + \left[E\left(Y_{n+1} \mid \Im_n\right) - E\left(Y_{n+1} \mid \Im_n\right)\right] \\ &= X_n . \end{aligned}$$

Hence $\{X_n; n \geq 1\}$ is a martingale. Thus, it is possible to construct a martingale sequence starting from any arbitrary sequence of random variables. ▲

■ EXAMPLE 11.9

Let $\{X_n; n \geq 0\}$ be a martingale with respect to the filtration $(\Im_n)_{n\geq 0}$ and let $\{Y_n; n \geq 0\}$ be defined by:

$$Y_{n+1} := X_{n+1} - X_n, \ n = 0, 1, 2, \cdots .$$

It is clear that:

$$E(Y_{n+1} \mid \Im_n) = 0 \ \forall \ n = 0, 1, 2, \cdots .$$

Suppose that $\{C_n; n \geq 1\}$ is a predictable stochastic process, that is, C_n is a $\Im_{n-1}$-measurable random variable for all n. We define a new process $\{Z_n; n \geq 0\}$ as:

$$\begin{aligned} Z_0 &:= 0 \\ Z_n &:= \sum_{i=1}^{n} C_i Y_i, \ n \geq 1. \end{aligned}$$

The process $\{Z_n; n \geq 0\}$ is a martingale with respect to filtration $\{\Im_n\}_{n\geq 0}$ and is called a *martingale transformation of the process* Y*, denoted by* $Z = C \cdot Y$*. The martingale transforms are the discrete analogues of stochastic integrals. They play an important role in mathematical finance in discrete time(see Section 12.3).* ▲

Note 11.4 *Suppose that* $\{C_n; n \geq 1\}$ *represents the amount of money a player bets at time* n *and* $Y_n := X_n - X_{n-1}$ *is the amount of money he can win or lose in each round of the game. If the bet is a monetary unit and* X_0 *is the initial wealth of the player, then* X_n *is the player's fortune at time* n *and* Z_n *represents the player's fortune by using the game strategy* $\{C_n; n \geq 1\}$*. The previous example shows that if* $\{X_n; n \geq 0\}$ *is a martingale and the game is fair, it will remain so no matter what strategy the player follows.*

■ EXAMPLE 11.10

Let $\xi_1, \xi_2, \cdots$ be i.i.d. random variables and suppose that for a fixed t:

$$m(t) := E\left(e^{t\xi_1}\right) < \infty.$$

The sequence of random variables $\{X_n; n \geq 0\}$ with $X_0 := 1$ and

$$X_n = X_n(t) := \frac{1}{(m(t))^n} \exp\left(t \sum_{j=1}^{n} \xi_j\right), \ n \geq 1,$$

is a martingale. ▲

■ EXAMPLE 11.11

Let $\xi_1, \xi_2, \cdots$ and $X_n(t)$ be as in the example above. We define the random variables $Y_n^{(k)}$ as:

$$Y_n^{(k)} = Y_n^{(k)}(t) := \frac{d^k X_n(t)}{dt^k}\Big|_{t=0}$$

We have that $\left\{Y_n^{(k)} : n \geq 1\right\}$ is a martingale. ▲

Definition 11.3 *A random variable τ with values $\{1, 2, \cdots\} \cup \{\infty\}$ is a stopping time with respect to the filtration $(\Im_n)_{n\geq 1}$ if $\{\tau \leq n\} \in \Im_n$ for each $n \geq 1$.*

Note 11.5 *The condition given in the previous definition is equivalent to $\{\tau = n\} \in \Im_n$ for each $n \geq 1$.*

■ EXAMPLE 11.12 First Arrival Time

Let $X_1, X_2, \cdots$ be a sequence of random variables adapted to the filtration $(\Im_n)_{n\geq 1}$. Suppose that A is a Borel set of $\mathbb{R}$ and consider the random variable defined by

$$\tau := \min\{n \geq 1 : X_n \in A\}$$

with $\min(\emptyset) := \infty$. It is clear that τ is a stopping time since:

$$\{\tau = n\} = \{X_1 \notin A\} \cap \cdots \cap \{X_{n-1} \notin A\} \cap \{X_n \in A\} \in \Im_n.$$

In particular we have that, for the gambler's ruin case, the time τ at which the player reaches the set $A = \{0, a\}$ for the first time is a stopping time. ▲

■ EXAMPLE 11.13 Martingale Strategy

Previously we observed that if a player who follows the martingale strategy loses the first n bets and wins the $(n+1)$th bet, then his wealth X_{n+1} after the $(n+1)$th bet is:

$$-1 - 2 - 2^2 - \cdots - 2^{n-1} + 2^n = -\sum_{k=1}^{n-1} 2^k + 2^n = 1.$$

Suppose that τ is the stopping time at which the player wins for the first time. It is of our interest to know what is, on average, his deficit for that time. That is, we want to determine the value $E(X_{\tau-1})$ from the previous equation. We have:

$$E(X_{\tau-1}) = \sum_{n=0}^{\infty} \left(-1-2-2^2-\cdots-2^{n-1}\right)\frac{1}{2^{n+1}} = -\infty.$$

Therefore, on average, a player must have an infinite capital to fulfill the strategy. ▲

Let $\{X_n; n \geq 1\}$ be a martingale with respect to the filtration $(\Im_n)_{n\geq 1}$. We know that $E(X_n) = E(X_1)$ for any $n \geq 1$. Nevertheless, if τ is a stopping time, it is not necessarily satisfied that $E(X_\tau) = E(X_1)$. Our next objective is to determine the conditions under which $E(X_\tau) = E(X_1)$, where τ is the stopping time.

Definition 11.4 *Let τ be a stopping time with respect to the filtration $(\Im_n)_{n\geq 0}$ and let $\{X_n; n \geq 0\}$ be a martingale with respect to the same filtration. We define the stopped process $\{X_{\tau\wedge n}; n \geq 0\}$ as follows:*

$$X_{\tau\wedge n}(\omega) = X_n(\omega)\mathcal{X}_{\{\tau(\omega)\geq n\}} + X_{\tau(\omega)}\mathcal{X}_{\{\tau(\omega)<n\}}.$$

Theorem 11.2 *If $\{X_n; n \geq 1\}$ is a martingale with respect to $(\Im_n)_{n\geq 0}$ and if τ is a stopping time with respect to $(\Im_n)_{n\geq 0}$, then $\{X_{\tau\wedge n}; n \geq 0\}$ is a martingale.*

Proof: Refer to Jacod and Protter (2004). ■

Theorem 11.3 (Optional Stopping Theorem) *Let $\{X_n; n \geq 0\}$ be a martingale with respect to the filtration $(\Im_n)_{n\geq 1}$ and let τ be a stopping time with respect to $(\Im_n)_{n\geq 1}$. If*

1. *$\tau < \infty$ a.s.,*
2. *$E(X_\tau)$ and $< \infty$*
3. *$\lim_{n\to\infty} E\left(X_n\mathcal{X}_{\{\tau>n\}}\right) = 0$,*

then $E(X_\tau) = E(X_n)$ for all $n \geq 1$.

Proof: Since for any $n \geq 1$ it is satisfied that

$$X_\tau = X_{\tau\wedge n} + (X_\tau - X_n)\,\mathcal{X}_{\{\tau>n\}}$$

and since the process $\{X_n; n \geq 0\}$ and $\{X_{\tau\wedge n}; n \geq 0\}$ are both martingales, we have:

$$\begin{aligned} E(X_\tau) &= E(X_{\tau\wedge n}) + E\left((X_\tau - X_n)\,\mathcal{X}_{\{\tau>n\}}\right) \\ &= E(X_n) + E\left(X_\tau\mathcal{X}_{\{\tau>n\}}\right) - E\left(X_n\mathcal{X}_{\{\tau>n\}}\right). \qquad (11.2) \end{aligned}$$

On the other hand by the hypothesis

$$\lim_{n\to\infty} E\left(X_n \mathcal{X}_{\{\tau>n\}}\right) = 0$$

and

$$E(X_\tau) = E\left(\sum_{j=1}^{\infty} X_j \mathcal{X}_{\{\tau=j\}}\right) = \sum_{j=1}^{\infty} E\left(X_j \mathcal{X}_{\{\tau=j\}}\right) < \infty,$$

it follows that the tail of the series, which is $E\left(X_\tau \mathcal{X}_{\{\tau>n\}}\right)$, tends to zero as n tends to ∞. Therefore, taking the limit as $n \to \infty$ in (11.2), we obtain:

$$E(X_\tau) = E(X_n) \text{ for all } n \geq 1.$$

■

Note 11.6 *Suppose that $\{X_n; n \geq 0\}$ is a symmetric random walk in $\mathbb{Z}$ with $X_0 := 0$ and that N is a fixed positive integer and let τ be the stopping time defined by:*

$$\tau := \min\{n \geq 1 : |X_n| = N\}.$$

It is easy to verify that the process $\{X_n; n \geq 0\}$ and the process $\{X_n^2 - n; n \geq 0\}$ are martingales. Moreover, it is possible to show that the stopping theorem hypotheses are satisfied. Consequently, we get

$$E\left(X_\tau^2 - \tau\right) = E\left(X_1^2 - 1\right) = 0$$

from which we have:

$$E(\tau) = E\left(X_\tau^2\right) = E\left(|X_\tau|^2\right) = E\left(N^2\right) = N^2.$$

That is, the random walk needs on average N^2 steps to reach the level N.

The following results on convergence of martingales, which we state without proof, provide many applications in stochastic calculus and mathematical finance.

Theorem 11.4 *Let $\{X_n; n \geq 0\}$ be a submartingale with respect to $(\Im_n)_{n\geq 0}$ such that $\sup_n E(|X_n|) < \infty$. Then there exists a random variable X having $E(|X|) < \infty$ such that:*

$$\lim_{n\to\infty} X_n = X \quad a.s.$$

Note 11.7 *There is a similar result for supermartingales because if $\{X_n; n \geq 0\}$ is a supermartingale with respect to $(\Im_n)_{n\geq 0}$, then $\{-X_n; n \geq 0\}$ is a submartingale with respect to $(\Im_n)_{n\geq 0}$. The previous theorem implies in addition that every nonnegative martingale converges almost surely. The following example shows that, in general, there is no convergence in the mean.*

EXAMPLE 11.14

Suppose that $\{Y_n; n \geq 1\}$ is a sequence of i.i.d random variables with normal distribution each having mean 0 and variance σ^2. Let:

$$\begin{aligned} X_0 &:= 1 \\ X_n &:= \exp\left(\sum_{j=1}^{n} Y_j - \frac{n}{2}\sigma^2\right). \end{aligned}$$

It is easy to prove that $\{X_n; n \geq 0\}$ is a nonnegative martingale. By using the strong law of large numbers we obtain that $X_n \xrightarrow{a.s} 0$. Nevertheless, $X_n \overset{L_1}{\nrightarrow} 0$ since $E(X_n) = 1$ for all n. ▲

Now we present a theorem which gives a sufficient condition to ensure the almost sure convergence and convergence in the r-mean. Its proof is beyond the scope of this text, (refer to Williams, 2006).

Theorem 11.5 *If $\{X_n; n \geq 0\}$ is a martingale with respect to $(\Im_n)_{n\in\mathbb{N}}$ such that $\sup_n E(|X_n|^r) < \infty$ for some $r > 1$, then there is a random variable X such that*

$$X_n \longrightarrow X$$

converges almost surely and in the r-mean.

Next, we give a brief account of continuous-time martingales. Many of the properties of martingales in discrete time are also satisfied in the case of martingales in continuous time.

Definition 11.5 *Let $(\Omega, \Im, P)$ be a probability space. A filtration is a family of sub-σ-algebras $(\Im_t)_{t\in T}$ such that $\Im_s \subseteq \Im_t$ for all $s \leq t$.*

Definition 11.6 *A stochastic process $\{X_t; t \in T\}$ is said to be adapted to the filtration $(\Im_t)_{t\in T}$ if X_t is $\Im_t$-measurable for each $t \in T$.*

Definition 11.7 *Let $\emptyset \neq T \subseteq \mathbb{R}$. A process $\{X_t; t \in T\}$ is called a martingale with respect to the filtration $(\Im_t)_{t\in T}$ if:*

1. *$\{X_t; t \in T\}$ is adapted to the filtration $(\Im_t)_{t\in T}$.*
2. *$E(|X_t|) < \infty$ for all $t \in T$.*
3. *$E(X_t \mid \Im_s) = X_s$ a.s. for all $s \leq t$.*

Note 11.8

a. *If condition 3 is replaced by: $E(X_t \mid \Im_s) \geq X_s$ a.s. for all $s \leq t$, then the process is called a submartingale.*

b. *If condition 3 is replaced by:* $E(X_t \mid \Im_s) \leq X_s$ *a.s. for all* $s \leq t$, *then the process is called a supermartingale.*

Note 11.9 *Condition 3 in the previous definition is equivalent to:*

$$E(X_t - X_s \mid \Im_s) = 0 \text{ a.s. for all } s \leq t.$$

Note 11.10 *The sequence* $\{X_t; t \in T\}$ *is clearly adapted to the canonical filtration, that is, to the filtration* $(\Im_t)_{t\in T}$, *where* $\Im_t = \sigma(X_s, s \leq t)$ *is the smallest* σ*-algebra with respect to which the random variables* X_s *with* $s \leq t$ *are measurable.*

■ **EXAMPLE 11.15**

Let $\{X_t; t \geq 0\}$ be a process with stationary and independent increments. Assume $\Im_t = \sigma(X_s, s \leq t)$ and $E(X_t) = 0$ for all $t \geq 0$. Then:

$$\begin{aligned} E(X_t \mid \Im_s) &= E(X_t - X_s + X_s \mid \Im_s) \\ &= E(X_t - X_s \mid \Im_s) + E(X_s \mid \Im_s) \\ &= E(X_t - X_s) + X_s \\ &= E(X_{t-s}) + X_s \\ &= X_s. \end{aligned}$$

That is, $\{X_t; t \geq 0\}$ is a martingale with respect to $(\Im_t)_{t\geq 0}$. ▲

Note 11.11 *If in the above example we replace the condition "*$E(X_t) = 0$ *for all* $t \geq 0$*" by "*$E(X_t) \geq 0$ *for all* $t \geq 0$*" ["*$E(X_t) \leq 0$ *for all* $t \geq 0$*"] we find that the process is a submartingale (a supermartingale).*

■ **EXAMPLE 11.16**

Let $\{N_t; t \geq 0\}$ be a Poisson process with parameter $\lambda > 0$. The process $\{N_t; t \geq 0\}$ has independent and stationary increments and in addition $E(N_t) = \lambda t \geq 0$. Hence, $\{N_t; t \geq 0\}$ is a submartingale.

However, the process $\{N_t - \lambda t; t \geq 0\}$ is a martingale and is called a *compensated Poisson process.* ▲

11.2 BROWNIAN MOTION

The Brownian motion is named after the English botanist Robert Brown (1773–1858) who observed that pollen grains suspended in a liquid moved irregularly. Brown, as his contemporaries, assumed that the movement was due to the life of these grains. However, this idea was soon discarded as the observations remained unchanged by observing the same movement with inert particles. Later it was found that the movement was caused by continuous particle collisions with molecules of the liquid in which it was embedded. The first attempt to mathematically describe the Brownian motion was made by the Danish mathematician and astronomer Thorvald N. Thiele (1838–1910) in 1880. Then in the early twentieth century, Louis Bachelier (1900), Albert Einstein (1905) and Norbert Wiener (1923) initiated independently the development of the mathematical theory of Brownian motion. Louis Bachelier (1870–1946) used this movement to describe the behavior of stock prices in the Paris stock exchange. Albert Einstein (1879–1955) in 1905 published his paper "*Über die von der molekularischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen*" in which he showed that at time t, the erratic movement a particle can be modeled by a normal distribution. The American mathematician Norbert Wiener (1894–1964) was the first to perform a rigorous construction of Einstein's model of Brownian motion, which led to the definition of the so-called Wiener measure in the space of trajectories. In this section we introduce Brownian motion and present a few of its important properties.

Definition 11.8 *The stochastic process $B = \{B_t, t \geq 0\}$ is called a standard Brownian motion or simply a Brownian motion if it satisfies the following conditions:*

1. $B_0 = 0$.

2. *B has independent and stationary increments.*

3. *For $s < t$, every increment $\{B_t - B_s\}$ is normally distributed with mean 0 and variance $(t - s)$.*

4. *Sample paths are continuous with probability 1.*

Note 11.12

1. *The Brownian motion is a Gaussian process. This is because the distribution of a random vector of the form $(B_{t_1}, B_{t_2}, \cdots, B_{t_n})$ is a linear combination of the vector $\left(B_{t_1}, B_{t_2} - B_{t_1}, \cdots, B_{t_n} - B_{t_{n-1}}\right)$ which has normal distribution.*

2. *The Brownian motion is a Markov process with transition probability density function*

$$P\left(B_t \in dy \mid B_s = x\right) \equiv p_{xy}(t-s)dy = \frac{1}{\sqrt{2\pi(t-s)}} e^{-\frac{(x-y)^2}{2(t-s)}} dy$$

for any $x, y \in \mathbb{R}$ and $0 < s < t$.

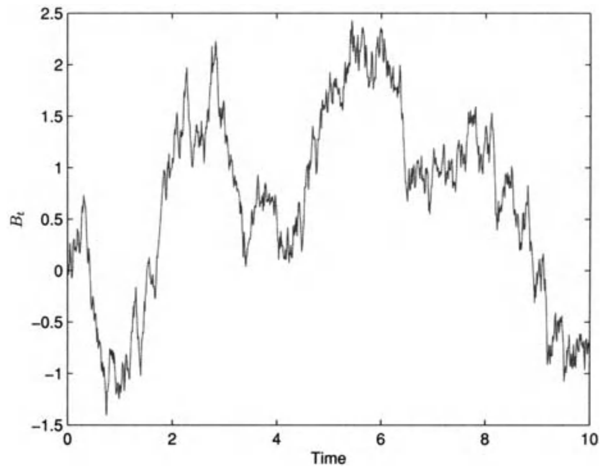


Figure 11.1 Sample path of Brownian motion

3. The probability density function of B_t is given by:

$$f_{B_t}(x) = \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} .$$

In the following algorithm, we simulate the sample path for the Brownian motion. This involves repeatedly generating independent standard normal random variables.

Algorithm 11.1

Input: T, N where T is the length of time interval and N is the time steps.

Output: $BM(k)$ for $k = 0(1)N$.

Initialization: $BM(0) := 0$

Iteration: For $k = 0(1)N - 1$ do:

$Z(k+1) = stdnormal(rand(0,1))$

$BM(k+1) = BM(K) + \sqrt{T/N} \times Z(k+1)$

where $stdnormal(rand(0,1))$ is the value of the standard normal random variable using the random number generated in the interval $(0,1)$. Using this algorithm, we obtain the sample path of Brownian motion as shown in Figure 11.1 for $T = 10$ and $N = 1000$.

Now we will discuss some simple and immediate properties of the Brownian motion:

1. $E(B_t) = 0$ for all $t \geq 0$.

2. $E(B_t^2) = t$ for all $t \geq 0$.

3. The covariance of Brownian motion $C(s,t) = \min(s,t)$. This is because, if $s \leq t$, then:

$$\begin{aligned} C(s,t) &= Cov(B_t, B_s) \\ &= E(B_t B_s) - E(B_t)E(B_s) \\ &= E([(B_t - B_s) + B_s] B_s) \\ &= E((B_t - B_s) B_s) + E(B_s^2) \\ &= 0 + s = \min(s,t). \end{aligned}$$

Similarly, if $t \leq s$, we get $C(s,t) = t$. Hence, the covariance of Brownian motion $C(s,t) = \min(s,t)$.

Theorem 11.6 *Let $\{B_t; t \geq 0\}$ be a Brownian motion. Then the following processes are also Brownian motions:*

1. *Shift Property: For any $s > 0$, $B_t^{(1)} = B_{t+s} - B_s$ is a Brownian motion.*

2. *Symmetry Property: $B_t^{(2)} = -B_t$ is a Brownian motion.*

3. *Scaling Property: For any constant $c > 0$, $B_t^{(3)} = \sqrt{c} B_{\frac{t}{c}}$ is a Brownian motion.*

4. *Time Reversal Property: $B_t^{(4)} = t B_{\frac{1}{t}}$ for $t > 0$ with $B_0 = 0$ is a Brownian motion.*

Proof: It is easy to check that $\{B_t^{(i)}; t \geq 0\}$ for $i = 1,2,3,4$ are processes with independent increments with $B_0^{(i)} = 0$. Also the increments are normally distributed with mean 0 and variance $(t - s)$. ∎

Brownian Motion as a Limit of Random Walks Let $\{X_t, t \geq 0\}$ be the stochastic process representing the position of a particle at time t. We assume that the particle performs a random walk such that in a small interval of time of duration Δt the particle moves forward a small distance Δx with probability p or moves backward by a small distance Δx with probability $q = 1 - p$, where p is independent of x and t. Suppose that the random variable Y_k denotes the length of the kth step taken by the particle in a small interval of time Δt and the Y_k's are independent and identically distributed random variables with $P(Y_k = +\Delta x) = p = 1 - P(Y_k = -\Delta x)$.

Suppose that the interval of length t is divided into n equal subintervals of length Δt. Then $n \cdot (\Delta t) = t$, and the total displacement X_t of the particle is the sum of n i.i.d. random variables Y_k, so that

$$X_t(\omega) := \sum_{i=1}^{n} Y_i$$

with $n \equiv [n(t)]$ and $n(t) = t/\Delta t$ for each $t \geq 0$. As a function of t, for each ω, X_t is a step function where steps occur every Δt units of time and steps are of magnitude Δx. We have:

$$E(Y_i) = (p-q)\Delta x \quad \text{and} \quad Var(Y_i) = 4pq(\Delta x)^2 \ .$$

Then:

$$E(X_t) = n(p-q)\Delta x \quad \text{and} \quad Var(X_t) = 4npq(\Delta x)^2 \ .$$

Substituting $n = \frac{t}{\Delta t}$, we have:

$$E(X_t) = t(p-q)\frac{\Delta x}{\Delta t} \quad \text{and} \quad Var(X_t) = 4pqt\frac{(\Delta x)^2}{\Delta t} \ .$$

When we allow $\Delta x \to 0$ and $\Delta t \to 0$, the corresponding steps n tend to ∞. We assume that the following expressions have finite limits:

$$E(X_t) = t(p-q)\frac{\Delta x}{\Delta t} \to \mu t \tag{11.3}$$

and

$$Var(X_t) = 4pqt\frac{(\Delta x)^2}{\Delta t} \to \sigma^2 t \tag{11.4}$$

where μ and σ are constants. Since the Y_k's are i.i.d. random variables, using the central limit theorem, for large $n = n(t)$ the sum $\sum_{i=1}^{n} Y_i = X_t$ is asymptotically normal with mean μt and variance $\sigma^2 t$. That is,

$$\frac{X_t - \mu t}{\sigma\sqrt{t}} \stackrel{d}{=} Z \tag{11.5}$$

where Z is a standard normal random variable.

Various Gaussian and non-Gaussian stochastic processes of practical relevance can be derived from Brownian motion. We introduce some of those processes which will find interesting applications in finance.

■ EXAMPLE 11.17

Let $\{B_t; t \geq 0\}$ be a Brownian motion. The stochastic process $\{R_t; t \geq 0\}$ defined by

$$R_t = |B_t| = \begin{cases} B_t & \text{if } B_t \geq 0 \\ -B_t & \text{if } B_t < 0 \end{cases}$$

is called a *Brownian motion reflected at the origin. The mean and variance of R_t are given by:*

$$\begin{aligned}
E(R_t) &= \int_{\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \\
&= 2\int_{0}^{\infty} x \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} \\
&= \sqrt{\frac{2t}{\pi}}
\end{aligned}$$

$$\begin{aligned}
Var(R_t) &= E\left(R_t^2\right) - [E(R_t)]^2 \\
&= E\left(B_t^2\right) - \frac{2t}{\pi} \\
&= \left(1 - \frac{2}{\pi}\right) t. \qquad \blacktriangle
\end{aligned}$$

■ EXAMPLE 11.18

Let $\{B_t; t \geq 0\}$ be a Brownian motion. The stochastic process $\{A_t; t \geq 0\}$ is defined by

$$A_t = \begin{cases} B_t & \text{if } t \leq T_0 \\ 0 & \text{if } t > T_0 \end{cases}$$

where $T_0 = inf\{t \geq 0 : B_t = 0\}$ is the hitting time at 0. Then A_t is called the *absorbed Brownian motion.* ▲

■ EXAMPLE 11.19

The stochastic process $\{U_t; 0 \leq t \leq 1\}$, defined as

$$U_t = B_t - tB_1 ,$$

is called a *Brownian bridge* or the *tied-down Brownian motion.*

The name Brownian bridge comes from the fact that it is tied down at both ends $t = 0$ and $t = 1$ since $U_0 = U_1 = 0$. In fact, the Brownian bridge $\{U_t; 0 \leq t \leq 1\}$ is characterized as being a Gaussian process with continuous sample paths and the covariance function

$$Cov(U_s, U_t) = s(1-t), \qquad 0 \leq s \leq t \leq 1 .$$

If $\{U_t; 0 \leq t \leq 1\}$ is a Brownian bridge, then it can be shown that the stochastic process

$$B_t = (1+t)U_{\frac{t}{1+t}}, \qquad t \geq 0 ,$$

is the standard Brownian motion. ▲

■ EXAMPLE 11.20

Let $\{B_t; t \geq 0\}$ be a Brownian motion. For $\mu \in \mathbb{R}$ and $\sigma > 0$, the process

$$B_t^{\mu} = \mu t + \sigma B_t, \qquad t \geq 0,$$

is called a *Brownian motion with drift* μ. *It is easy to check that* B_t^{μ} *is a Gaussian process with mean* μt *and covariance* $C(s,t) = \sigma^2 \min(s,t)$.
▲

■ EXAMPLE 11.21

Let $\{B_t; t \geq 0\}$ be a Brownian motion. For $\mu \in \mathbb{R}$ and $\sigma > 0$, the process

$$X_t = exp\left(\mu t + \sigma B_t\right), \qquad t \geq 0,$$

is called a *geometric Brownian motion.* ▲

This process has been used to describe stock price fluctuations (see next chapter for more details). It should be noted that X_t is not a Gaussian process. Now we will give the mean and covariance for the geometric Brownian motion.

Using the moment generating function of the normal random variable (4.2), we get:

$$E(X_t) = e^{\mu t} E\left(e^{\sigma B_t}\right) = e^{\mu t} E\left(e^{\sigma \sqrt{t} Z}\right) = e^{\left(\mu + \frac{1}{2}\sigma^2\right)t}. \tag{11.6}$$

Similarly we obtain the covariance of the geometric Brownian motion for $s < t$,

$$Cov(s,t) = e^{\left(\mu + \frac{1}{2}\sigma^2\right)(t+s)} \left(e^{\sigma^2 s} - 1\right), \tag{11.7}$$

and the variance is given by:

$$Var\left(X_t\right) = e^{\left(2\mu + \sigma^2\right)t} \left(e^{\sigma^2 t} - 1\right). \tag{11.8}$$

The previous section discussed continuous-time martingales. Presently we will see a Brownian motion as an example of a continuous-time martingale.

Theorem 11.7 *Suppose that* $\{B_t; t \geq 0\}$ *is a Brownian motion with respect to filtration* $\Im_t$, *where* $\Im_t := \sigma(B_s; s \leq t)$. *Then*

1. $\{B_t\}$ *is a martingale,*
2. $\{B_t^2 - t\}$ *is a martingale and*

3. *for $\sigma \in \mathbb{R}$, $\{exp(\sigma B_t - (\sigma^2/2)t)\}$ is a martingale (called an exponential martingale).*

Proof:

1. It is clear that, for every $t \geq 0$, B_t is adapted to the filtration $(\Im_t)_{t\geq 0}$ and $E(B_t)$ exists. For any $s, t \geq 0$ such that $s < t$:

$$\begin{aligned} E(B_t \mid \Im_s) &= E([B_t - B_s + B_s \mid \Im_s) \\ &= E(B_t - B_s \mid \Im_s) + E(B_s \mid \Im_s) \\ &= E(B_t - B_s) + B_s \\ &= B_s. \end{aligned}$$

2.

$$\begin{aligned} E\left(B_t^2 - B_s^2 \mid \Im_s\right) &= E\left((B_t - B_s)^2 + 2B_s(B_t - B_s) \mid \Im_s\right) \\ &= E\left((B_t - B_s)^2 \mid \Im_s\right) + 2B_s E\left((B_t - B_s) \mid \Im_s\right) \\ &= t - s. \end{aligned}$$

Thus:

$$\begin{aligned} E\left(B_t^2 - t \mid \Im_s\right) &= E\left(B_t^2 - {B_s}^2 + B_s^2 - (t-s) - s \mid \Im_s\right) \\ &= (t-s) + \ B_s^2 - (t-s) - s \\ &= B_s^2 \ - s. \end{aligned}$$

3. The moment generating function of $\{B_t; t \geq 0\}$ is given by:

$$\begin{aligned} m_B(\sigma) &= E\left(e^{\sigma B_t}\right) \\ &= \int_{-\infty}^{\infty} e^{\sigma x} f(x \mid 0, t) dx \\ &= \int_{-\infty}^{\infty} e^{\sigma x} \frac{1}{\sqrt{2\pi t}} e^{-\frac{x^2}{2t}} dx \\ &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(x-t\sigma)^2}{2t}} dx \cdot e^{\frac{\sigma^2}{2}t} \\ &= e^{\frac{\sigma^2}{2}t}. \end{aligned}$$

Therefore $E\left(e^{\sigma B_t - \frac{\sigma^2}{2}t}\right) = 1$ and $e^{\sigma B_t - \frac{\sigma^2}{2}t}$ is integrable. Now:

$$\begin{aligned} E\left(e^{\sigma B_{t+s} - \frac{\sigma^2}{2}(t+s)} \mid \Im_t\right) &= E\left(e^{\sigma B_t} \cdot e^{\sigma(B_{t+s} - B_t)} \mid \Im_t\right) e^{-\frac{\sigma^2}{2}(t+s)} \\ &= e^{\sigma B_t} E\left(e^{\sigma(B_{t+s} - B_t)}\right) e^{-\frac{\sigma^2}{2}(t+s)} \\ &= e^{\sigma B_t} E\left(e^{\sigma B_s}\right) e^{-\frac{\sigma^2}{2}(t+s)} \\ &= e^{\sigma B_t - \frac{\sigma^2}{2}t}. \end{aligned}$$

■

Note 11.13 *Let $\{X_t; t \geq 0\}$ be a stochastic process with respect to filtration $(\Im_t)_{t\geq 0}$. Then $\{X_t; t \geq 0\}$ is a Brownian motion if and only if it satisfies the following conditions:*

1. $X_0 = 0$ *a.s.*

2. $\{X_t; t \geq 0\}$ *is a martingale with respect to filtration* $\Im_t$.

3. $\{X_t^2 - t; t \geq 0\}$ *is a martingale with respect to filtration* $\Im_t$.

4. *With probability* 1, *the sample paths are continuous.*

The above result is known as Lévy's characterization of a Brownian motion (see Mikosh, 1998).

The possible realization of a sample path's structure and its properties play a crucial role and are the subject matter of deep study. Brownian motion has the continuity of the sample path by definition. Another important property is that it is nowhere differentiable with probability 1. The mathematical proof of this property is beyond the scope of this text. For rigorous mathematical proof, the reader may refer to Karatzas and Shreve (1991) or Breiman (1992).

Now we will see an important and interesting property of a Brownian motion called quadratic variation. In the following, we define the notion of quadratic variation for a real-valued function.

Definition 11.9 *Let $f(t)$ be a function defined on the interval $[0,T]$. The quadratic bounded variation of the function f is*

$$\lim_{\|\tau_n\|\to 0} \sum_{i=1}^{n} \left(f(t_i) - f(t_{i-1})\right)^2$$

where τ_n is a partition of the interval $[0,T]$,

$$\tau_n : 0 = t_0 < t_1 < \cdots < t_n = T$$

with:

$$\|\tau_n\| = \max_{1\leq i\leq n} (t_i - t_{i-1}) \to 0 \quad as\ n \to \infty.$$

Theorem 11.8 *The quadratic variation of the sample path of a Brownian motion over the interval $[0,T]$ converges in mean square to T.*

Proof: Let τ_n be a partition of the interval $[0,T]$:

$$\tau_n : 0 = t_0 < t_1 < \cdots < t_n = T.$$

Let

$$Q_n := \sum_{i=1}^{n} \left(B_{t_i} - B_{t_{i-1}}\right)^2.$$

Then for each n we have:

$$\begin{aligned} E(Q_n) &= \sum_{i=1}^{n} E\left(B_{t_i} - B_{t_{i-1}}\right)^2 \\ &= \sum_{i=1}^{n} (t_i - t_{i-1}) \\ &= T. \end{aligned}$$

Also:

$$\begin{aligned} Var(Q_n) &= \sum_{i=1}^{n} Var\left(B_{t_i} - B_{t_{i-1}}\right)^2 \\ &\leq 3\sum_{i=1}^{n} (t_i - t_{i-1})^2 \quad \text{since } E(B_t^4) = 3t^2 \\ &\leq 3T\left\|\tau_n\right\| \to 0 \quad \text{where } \left\|\tau_n\right\| \to 0. \end{aligned}$$

We conclude that:

$$\lim_{n\to\infty} E\left(\left(Q_n - T\right)\right)^2\Big) = 0.$$

Thus we have proved that Q_n converges to T in mean square. ∎

We can also prove Q_n converges to T with probability 1. This proof can be found in Breiman (1992) and Karatzas and Shreve (1991) (see Chapter 8 for different types of convergence of random variables).

As we have seen in this section, the sample path of Brownian motion is nowhere differentiable. Because the stochastic processes which are driven by Brownian motion are also not differentiable, we cannot apply classical calculus. In the following section we introduce the stochastic integral or Itô integral with respect to Brownian motion and its basic rules. We will do so using an intuitive approach which is based on classical calculus. For a mathematically rigorous approach on this integral see Karatzas and Shreve (1991) or Oksendal (2006).

11.3 ITÔ CALCULUS

The stochastic calculus or Itô calculus was developed during the year 1940 by Japanese mathematician K. Itô and is similar to the classical calculus of Newton which involves differentials and integrals of deterministic functions. In this section, we will study the stochastic integral of the process $\{X_t; t \geq 0\}$ with respect to a Brownian motion, that is, we adequately define the following expression:

$$I_t := I(X_t) = \int_0^t X_s dB_s. \tag{11.9}$$

In the classical calculus, the equations which consist of the expressions of the form dx are known as *differential equations.* If we replace the term dx by an expression of the form dX_t, the equations are known as *stochastic differential equations.* Formally, a stochastic differential equation has the form

$$dX_t = \mu(t, X_t)\, dt + \sigma(t, X_t) dB_t \tag{11.10}$$

where $\mu(x, t)$ and $\sigma(x, t)$ are given functions. Equation (11.10) can be written in integral form:

$$X_t(\omega) = X_0(\omega) + \int_0^t \mu(s, X_s(\omega))ds + \int_0^t \sigma(s, X_s(\omega))dB_s(\omega). \tag{11.11}$$

The first integral is a Riemann integral. How can we interpret the second integral? Initially we could take our inspiration from ordinary calculus in defining this integral as a limit of partial sums, such as

$$\sum_{i=1}^{n} \sigma(t_i^*, X_{t_i^*})(B_{t_i} - B_{t_{i-1}}) \qquad t_i^* \in [t_{i-1} - t_i]$$

provided the sum exists. Unlike the Riemann sums, the value of the sum here depends on the choice of the chosen points t_i's. In the case of stochastic integrals, the key idea is to consider the Riemann sums where the integrand is evaluated at the left endpoints of the subintervals. That is:

$$\sum_{i=1}^{n} \sigma(t_{i-1}, X_{t_{i-1}})(B_{t_i} - B_{t_{i-1}})$$

Observing that the sum of random variables will be another random variable, the problem is to show that the limit of the above sum exists in some suitable sense. The mean square convergence (see Chapter 8 for the definition) is used to define the stochastic integral. We establish the family of stochastic processes for which the Itô integral can be defined.

Definition 11.10 *Let L^2 be the set of all the stochastic processes $\{X_t; t \geq 0\}$ such that:*

(a.) *The process $X = \{X_t; t \geq 0\}$ is progressively measurable with respect to the given filtration $\Im = (\Im_t)_{t \geq 0}$. This means that, for every t, the mapping $(s, \omega) \to X_s(\omega)$ on every set $[0, t] \times \Omega$ is measurable.*

(b.) $E\left(\int_0^T X_t^2 dt\right) < \infty$ *for all $T > 0$.*

Now we give the definition of the Itô integral for any process $\{X_t; t \geq 0\} \in L^2$.

Definition 11.11 *Let $\{X_t; t \geq 0\}$ be a stochastic process in L^2 and $T > 0$ fixed. We define the stochastic integral or Itô integral of X_t with respect to Brownian motion B_t over the interval $[0, T]$ as*

$$\int_0^T X_t(\omega)dB_t(\omega) = \lim_{n\to\infty} \sum_{j=1}^{n} X_{t_{j-1}}(\omega)\left(B_{t_j}(\omega) - B_{t_{j-1}}(\omega)\right) \tag{11.12}$$

where τ_n is a partition of the interval $[0, T]$ such that

$$\tau_n : 0 = t_0 < t_1 < \cdots < t_n = T$$

with:

$$\|\tau_n\| = \max_{1\leq i\leq n} (t_i - t_{i-1}) \to 0 \quad as\ n \to \infty.$$

Notation:

$$I_T(X) = \int_0^T X_t(\omega)dB_t(\omega) \ .$$

■ **EXAMPLE 11.22**

Consider the stochastic integral

$$I_T(B) = \int_0^T B_t(\omega)dB_t(\omega)$$

where B_t is a Brownian motion. Let $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ be a partition of the interval $[0, T]$. From the definition of the stochastic integral, we have:

$$\int_0^T B_t(\omega)dB_t(\omega) = \lim_{n\to\infty} \sum_{i=1}^{n} B_{t_{i-1}}\left(B_{t_i} - B_{t_{i-1}}\right) \ .$$

By the use of the identity

$$a(b-a) = \frac{1}{2}(b^2 - a^2) - \frac{1}{2}(b-a)^2$$

we get:

$$\begin{aligned}\int_0^T B_t(\omega)dB_t(\omega) &= \lim_{n\to\infty} \sum_{i=1}^{n} \left[\frac{1}{2}\left(B_{t_i}^2 - B_{t_{i-1}}^2\right) - \frac{1}{2}\left(B_{t_i} - B_{t_{i-1}}\right)^2\right] \\ &= \frac{1}{2}\lim_{n\to\infty} \sum_{i=1}^{n} \left(B_{t_i}^2 - B_{t_{i-1}}^2\right) - \frac{1}{2}\lim_{n\to\infty} \sum_{i=1}^{n} \left(B_{t_i} - B_{t_{i-1}}\right)^2 \\ &= \frac{1}{2}B_T^2 - \frac{1}{2}T. \quad \blacktriangle\end{aligned}$$

The stochastic integral (11.12) for all $T > 0$ satisfies the following properties:

1. *Zero mean:*
$$E\left[\left(\int_0^T X_t dB_t\right)\right] = 0.$$

2. *Itô isometry*:
$$E\left[\left(\int_0^T X_t dB_t\right)^2\right] = E\left(\int_0^T X_t^2 dt\right).$$

3. *Martingale*: For $t \leq T$,
$$E\left[\int_0^T X_s dB_s \mid \Im_t\right] = \int_0^t X_s dB_s.$$

4. *Linearity:* For $\{X_t; t \geq 0\}, \{Y_t; t \geq 0\} \in L^2$,
$$\int_0^T (\alpha X_t + \beta Y_t)\, dB_t = \alpha \int_0^T X_t dB_t + \beta \int_0^T Y_t dB_t.$$

Proof: We now prove only the martingale property of the Itô integral. For proofs of the remaining properties, the reader may refer to Karatzas and Shreve (1991). Consider

$$\begin{aligned}
E\left(\int_0^T X_s dB_s \mid \Im_t\right) &= E\left(\int_0^t X_s dB_s + \int_t^T X_s dB_s \mid \Im_t\right) \\
&= E\left(\int_0^t X_s dB_s \mid \Im_t\right) + E\left(\int_t^T X_s dB_s \mid \Im_t\right) \\
&= \int_0^t X_s dB_s + E\left(\int_t^T X_s dB_s\right) \\
&= \int_0^t X_s dB_s
\end{aligned}$$

where the above equality follows by the zero mean property. ■

■ EXAMPLE 11.23

Let $X_t = \int_0^t e^{B_t} dB_t$ be an Itô integral. We have $E(X_t) = 0$ by property (11.22). The variance is calculated by use of the *mgf* of Brownian motion and Itô isometry. We have:

$$E\left(\left[\int_0^t e^{B_s} dB_s\right]^2\right) = \int_0^t E\left(e^{2B_s}\right) ds = \int_0^t e^{2s} ds = \frac{1}{2}\left(e^{2t} - 1\right). \quad \blacktriangle$$

In the context of ordinary calculus, the Itô formula is also known as the change of variable or chain rule for the stochastic calculus.

Theorem 11.9 (Itô's Formula) *Let $f : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable function and let $B = \{B_t; t \geq 0\}$ be a Brownian motion that starts at x_0, that is, $B_0 = x_0$. Then*

$$f(B_t) = f(x_0) + \int_0^t f'(B_s)\, dB_s + \frac{1}{2}\int_0^t f''(B_s)\, ds$$

or in the differential form:

$$df(B_t) = f'(B_t)dB_t + \frac{1}{2}f''(B_t)dt.$$

Proof: Fix $t > 0$. Let $\tau_n : 0 = t_0 < t_1 < \cdots < t_n = t$ be a partition of $[0, t]$. By Taylor's theorem, we have:

$$\begin{aligned} f(B_t) - f(B_0) &= \sum_{i=1}^{n} \left[f(B_{t_i}) - f\left(B_{t_{i-1}}\right)\right] \\ &= \sum_{j=1}^{n} \Big[f'(B_{t_{j-1}})\left(B_{t_j} - B_{t_{j-1}}\right) \\ &\qquad + \frac{1}{2}f''(B_{t_{j-1}})\left(B_{t_j} - B_{t_{j-1}}\right)^2 + \cdots\Big]. \end{aligned}$$

Taking the limit $n \to \infty$ when $\Delta t \to 0$, we find that the first sum of the right-hand side converges to the Itô integral and the second sum on the right-hand side converges to $\frac{1}{2}\int_0^t f''(B_s)ds$ because of mean square convergence. We get:

$$f(B_t) - f(B_0) = \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds.$$

Thus:

$$f(B_t) = f(x_0) + \int_0^t f'(B_s)dB_s + \frac{1}{2}\int_0^t f''(B_s)ds$$

. ■

■ EXAMPLE 11.24

Let $f(x) = x^2$ and $B = \{B_t; t \geq 0\}$ be a standard Brownian motion. The Itô formula establishes that:

$$B_t^2 = \int_0^t 2B_s dB_s + \frac{1}{2}\int_0^t 2ds.$$

That is:

$$\int_0^t B_s dB_s = \frac{1}{2}B_t^2 - \frac{1}{2}t. \quad \blacktriangle$$

■ **EXAMPLE 11.25**

Let $f(x) = x^3$ and $B = \{B_t; t \geq 0\}$ be a standard Brownian motion. The Itô formula establishes that:

$$B_t^3 = \int_0^t 3B_s^2 dB_s + 3\int_0^t B_s ds.$$

That is:

$$\int_0^t B_s^2 dB_s = \frac{1}{3}B_t^3 - \int_0^t B_s ds. \quad \blacktriangle$$

■ **EXAMPLE 11.26**

Let $\beta_n(t) = E(B_t^n)$ for a Brownian motion $\{B_t; t \geq 0\}$ with $B_0 = 0$. Prove that

$$\beta_n(t) = \frac{1}{2}n(n-1)\int_0^t \beta_{n-2}(s)ds, \qquad n \geq 2\,,$$

and hence find $E(B_t^4)$ and $E(B_t^6)$.

Solution. By the Itô's formula, we have:

$$B_t^n = n\int_0^t B_s^{n-1}dB_s + \frac{1}{2}n(n-1)\int_0^t B_s^{n-2}ds$$

Taking expectation we have:

$$\beta_n(t) = \frac{1}{2}n(n-1)\int_0^t \beta_{n-2}(s)ds$$

Since $\beta_2(t) = t$, we get:

$$\begin{aligned} \beta_4(t) &= \frac{1}{2}\cdot 4\cdot 3\cdot \int_0^t s\,ds = 3t^2 \\ \beta_6(t) &= \frac{1}{2}\cdot 6\cdot 5\cdot \int_0^t 3s^2\,ds = 15t^3. \quad \blacktriangle \end{aligned}$$

Definition 11.12 *For a fixed $T > 0$, the stochastic process $\{X_t; 0 \leq t \leq T\}$ is called an Itô process if it has the form*

$$X_t = X_0 + \int_0^t Y_s ds + \int_0^t Z_s dB_s, \qquad 0 \leq t \leq T, \tag{11.13}$$

where X_0 is $\Im_0$-measurable and the processes Y_t and Z_t are $\Im_t$-adapted such that, for all $t \geq 0$, $E(|Y_t|) < \infty$ and $E(|Z_t|^2) < \infty$. An Itô process has the differential form

$$dX_t = Y_t dt + Z_t dB_t. \tag{11.14}$$

We now give the Itô formula for an Itô process.

Theorem 11.10 (Itô's Formula for the General Case) *Let $\{X_t; t \geq 0\}$ be an Itô process given in (11.14). Suppose that $f(t,x)$ is a twice continuously differentiable function with respect to x and t. Then $f(t, X_t)$ is also an Itô process and:*

$$\begin{aligned} f(t, X_t) &= f(0, X_0) + \int_0^t Z_s \frac{\partial f(s, X_s)}{\partial x} dB_s \\ &\quad + \int_0^t \left[\frac{\partial f(s, X_s)}{\partial t} + Y_s \frac{\partial f(s, X_s)}{\partial x} + \frac{1}{2} Z_s^2 \frac{\partial^2 f(s, X_s)}{\partial x^2} \right] ds. \end{aligned}$$

Proof: See Oksendal (2006). ■

Note 11.14 *We introduce the notation*

$$(dX_t)^2 = (Z_t^2)\, dt$$

which is computed using the following multiplication rules:

$\bullet$	dt	dB_t
dt	0	0
dB_t	0	dt

The Itô formula then can be expressed in the following form:

$$\begin{aligned} df(t, X_t) &= Z_t \frac{\partial f(t, X_t)}{\partial x} dB_t \\ &\quad + \left[\frac{\partial f(t, X_t)}{\partial t} + Y_t \frac{\partial f(t, X_t)}{\partial x} + \frac{1}{2} Z_t^2 \frac{\partial^2 f(t, X_t)}{\partial x^2} \right] dt. \end{aligned}$$

Note 11.15 *Itô's formula can also be expressed in differentials as:*

$$df(t, X_t) = \frac{\partial f(t, X_t)}{\partial t} dt + \frac{\partial f(t, X_t)}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 f(t, X_t)}{\partial x^2} (dX_t)^2.$$

■ EXAMPLE 11.27

Let $X_t = t$ and $f(t,x) = g(x)$ be a twice-differentiable function. It is easy to see that:

$$X_t = 0 + \int_0^t ds + \int_0^t 0 dB_s.$$

Thus, applying Itô's formula, we get:

$$g(t) - g(0) = \int_0^t g'(s) ds.$$

That is, the fundamental theorem of calculus is a particular case of Itô's formula. ▲

■ EXAMPLE 11.28

Let $X_t = h(t)$ where h is a differentiable function and let $f(t,x) = g(x)$ be a twice-differentiable function. It is easy to check that:

$$X_t = h(0) + \int_0^t h'(s)ds + \int_0^t 0dB_s.$$

Applying Itô's formula, we obtain :

$$g(h(t)) - g(h(0)) = \int_0^t h'(s)g'(h(s))\,ds.$$

In this case also, the substitution theorem of calculus is a particular case of Itô's formula. ▲

■ EXAMPLE 11.29

Let $\{B_t; t \geq 0\}$ be a Brownian motion and consider the following differential equation:

$$dY_t = \mu Y_t dt + \sigma Y_t dB_t. \tag{11.15}$$

Let $Z_t = \log(Y_t)$. Then, by Itô's formula, we have:

$$dZ_t = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t.$$

Thus:

$$d\log(Y_t) = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma dB_t\ .$$

Integrating we get

$$\log(Y_t) - \log(Y_0) = \left(\mu - \frac{1}{2}\sigma^2\right) t + \sigma B_t$$

so that the solution of equation (11.15) is:

$$Y_t = Y_0 \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right) + \sigma B_t\right). \quad ▲$$

■ **EXAMPLE 11.30**

Consider the Langevin equation

$$dX_t = -\beta X_t dt + \alpha dB_t$$

where $\alpha \in \mathbb{R}$ and $\beta > 0$. The process $\{X_t; t \geq 0\}$ with $X_0 = x_0$ can be written as:

$$X_t = x_0 + \alpha B_t - \beta \int_0^t X_s ds.$$

Let $f(t, x) = e^{\beta t} x$. Applying Itô's formula, we get:

$$\begin{aligned} d\left(e^{\beta t} X_t\right) &= \beta e^{\beta t} X_t dt + e^{\beta t} dX_t \\ &= \beta e^{\beta t} X_t dt + e^{\beta t}\left(-\beta X_t dt + \alpha dB_t\right) \\ &= \alpha e^{\beta t} dB_t. \end{aligned}$$

Integration of the above equation gives for $s \leq t$:

$$X_t = e^{-\beta t} X_0 + \alpha \int_0^t e^{-\beta(t-s)} dB_s.$$

The solution of the Langevin equation with initial condition $X_0 = x_0$ is called an Ornstein-Uhlenbeck process. ▲

We complete this chapter with the Itô formula for functions of two or more variables.

Multidimensional Itô Formula

We now give the Itô formula for functions of two variables. Consider a two-dimensional process

$$dX_t = \mu_t dt + \sigma_t dB_t^{(1)} \tag{11.16}$$

$$dY_t = \alpha_t dt + \beta_t dB_t^{(2)} \tag{11.17}$$

where $\{B_t^{(1)}; t \geq 0\}$ and $\{B_t^{(2)}; t \geq 0\}$ are two Brownian motions with their covariances given by

$$Cov\left(B_t^{(1)}, B_t^{(2)}\right) = E\left(B_t^{(1)} \cdot B_t^{(2)}\right) = \rho t \tag{11.18}$$

where ρ is the correlation coefficient of the two Brownian motions. Let $g(t, x, y)$ be a twice-differentiable function and let $Z_t = g(t, X_t, Y_t)$. Then Z_t is also an Itô process and satisfies:

$$\begin{aligned} dZ_t = &\ \frac{\partial g}{\partial t}(t, X_t, Y_t)dt + \frac{\partial g}{\partial x}(t, X_t, Y_t)dX_t + \frac{\partial g}{\partial y}(t, X_t, Y_t)dY_t \\ &+ \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(t, X_t, Y_t)(dX_t)^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}(t, X_t, Y_t)(dY_t)^2 \\ &+ \frac{\partial^2 g}{\partial x \partial y}(t, X_t, Y_t)(dX_t)(dY_t). \end{aligned}$$

For the proof, the reader may refer to Karatzas and Shreve (1991).

Note 11.16 *For any two Itô processes $\{X_t; t \geq 0\}$ and $\{Y_t; t \geq 0\}$, we have the following product rule for the differention:*

$$d(X_tY_t) = X_t dY_t + Y_t dX_t + (dX_t)(dY_t). \tag{11.19}$$

Theorem 11.11 *Let X_t and Y_t be two Itô processes such that $E\left(\int_0^T X_t^2 dt\right) < \infty$ and $E\left(\int_0^T Y_t^2 dt\right) < \infty$. Then:*

$$E\left(\int_0^T X_t dB_t \int_0^T Y_t dB_t\right) = E\left(\int_0^T X_t Y_t dt\right) .$$

Proof: Let $I_1 = \int_0^T X_t dB_t$ and $I_2 = \int_0^T Y_t dB_t$.
By using the identity

$$I_1 I_2 = \frac{1}{2}\left((I_1 + I_2)^2 - I_1^2 - I_2^2\right)$$

and taking expectation, we get:

$$E(I_1 I_2) = \frac{1}{2}\left(E(I_1 + I_2)^2 - E(I_1^2) - E(I_2^2)\right) .$$

By use of Itô's isometry property we get the desired result. ■

■ EXAMPLE 11.31

Suppose that $X_t = tB_t$. Use of product rule (11.19) gives us:

$$dX_t = t dB_t + B_t dt. \quad \blacktriangle$$

■ EXAMPLE 11.32

Suppose that $X_t = tB_t$ and Y_t satisfies the stochastic differential equation

$$dY_t = \frac{1}{2} Y_t dt + Y_t dB_t, \qquad Y_0 = 1.$$

We know that $Y_t = e^{B_t}$ is a geometric Brownian motion. Then the use of product rule (11.19) gives us:

$$d(X_tY_t) = X_t dY_t + Y_t dX_t + tY_t dt.$$

This is because:

$$(dX_t)(dY_t) = (tdB_t + B_t dt)\left(\frac{1}{2}Y_t dt + Y_t dB_t\right) = tY_t dt. \qquad \blacktriangle$$

■ EXAMPLE 11.33

Suppose that

$$dX_t = \alpha dB_t + \beta dW_t \tag{11.20}$$

with $X_0 = 0$, $\alpha, \beta \in \mathbb{R}$ and $\{B_t; t \geq 0\}$ and $\{W_t; t \geq 0\}$ are two Brownian motions. Let $f(t, x) = x^2$. Then, from Itô's formula,

$$dX_t^2 = (\alpha^2 + \beta^2)dt + 2\alpha X_t dB_t + 2\beta X_t dW_t \tag{11.21}$$

with $X_0^2 = 0$. Note that $X_t = \alpha B_t + \beta W_t$ and:

$$\int_0^t dX_s^2 = X_t^2 - X_0^2 = X_t^2. \tag{11.22}$$

From equations (11.21) and (11.22), we get:

$$(\alpha B_t + \beta W_t)^2 = (\alpha^2 + \beta^2)t + 2\int_0^t \alpha X_s dB_s + 2\beta \int_0^t X_s dW_s.$$

Using the relation

$$\int_0^t B_s dB_s = \frac{B_t^2}{2} - \frac{t}{2},$$

we have the following interesting result:

$$2B_t W_t = \int_0^t B_s dW_s + \int_0^t W_s dB_s. \qquad \blacktriangle$$

Without recourse to measure theory, we have presented various tools necessary in dealing with financial models with the use of stochastic calculus. This chapter does not make a full-fledged analysis and is intended as a motivation for the further study. For a more rigorous treatment, the reader may refer to Grimmett and Stirzaker (2001), Oksendal (2005), Mikosch (2002), Shreve (2004), and Karatzas and Shreve (1991).

EXERCISES

11.1 In Example 11.11 verify

$$Y_n^{(1)} = \sum_{j=1}^{n} \xi_j$$

and

$$Y_n^{(2)} = \left(\sum_{j=1}^{n} (\xi_j - E(\xi_j)) \right)^2 - nVar(\xi_1).$$

11.2 Let $\{X_n; n \geq 0\}$ be a martingale (supermartingale) with respect to the filtration $(\Im_n)_{n \geq 0}$. Prove that

$$E(X_{n+k} \mid \Im_n) = X_n \quad (\leq \text{ for supermantingale})$$

for all $k \geq 0$.

11.3 Let $\{X_n; n \geq 0\}$ be a martingale (supermartingale) with respect to the filtration $(\Im_n)_{n \geq 0}$. Prove that:

$$E(X_n) = E(X_k) \quad (\leq \text{ for supermantingale})$$

for all $0 \leq k \leq n$

11.4 Let $\{X_n; n \geq 0\}$ be a martingale with respect to the filtration $(\Im_n)_{n \geq 0}$ and assume f to be a convex function. Prove that $\{f(X_n); n \geq 0\}$ is a submartingale with respect to the filtration $(\Im_n)_{n \geq 0}$.

11.5 If $\{X_t; t \geq 0\}$ is a martingale with respect to $(\Im_t)_{t \geq 0}$ and if $h : \mathbb{R} \longrightarrow \mathbb{R}$ is a convex function such that $E(|h(X_t)|) < \infty$ for all $t \geq 0$, show that $\{h(X_t); t \geq 0\}$ is a submartingale with respect to $(\Im_t)_{t \geq 0}$.

11.6 Let $\xi_1, \xi_2, \cdots$ be i.i.d. random variables, such that $P(\xi_n = 1) = p$ and $P(\xi_n = -1) = 1 - p$ for some p in $(0, 1)$. Prove that $\{M_n; n \geq 0\}$ with

$$\begin{aligned} M_0 &: = 1 \\ M_n &: = \left(\frac{1-p}{p} \right)^{\sum_{i=1}^{n} \xi_i} \end{aligned}$$

is a martingale with respect to $(\Im_n)_{n \geq 0}$, where $\Im_0 = \{\emptyset, \Omega\}$ and $\Im_n = \sigma(\xi_1, \xi_2, \cdots, \xi_n)$ for $n \geq 1$.

11.7 Let $X_1, X_2, \cdots$ be a sequence of i.i.d. random variables satisfying

$$P\left(X_1 = \frac{3}{2} \right) = P\left(X_1 = \frac{1}{2} \right) = \frac{1}{2}.$$

Let $M_0 := 0$, $M_n := X_1 X_2 \cdots X_n$ and $\Im_n = \sigma(X_1, X_2, \cdots, X_n)$. Is $\{M_n; n \geq 0\}$ a martingale with respect to $(\Im_n)_{n \geq 0}$? Explain.

11.8 Let $X_1, X_2, \cdots$ be a sequence of random variables such that $E(X_n) = 0$ for all $n = 1, 2, \cdots$ and suppose $E(e^{X_n})$ exists for all $n = 1, 2, \cdots$.

a) Is the sequence $\{Y_n; n \geq 1\}$ with $Y_n := \exp\left(\sum_{i=1}^{n} X_i \right)$ a submartingale with respect to $(\Im_n)_{n \geq 1}$, where $\Im_n = \sigma(X_1, X_2, \cdots, X_n)$ for $n \geq 1$? Explain.

b) Find (if possible) constants α_n such that the sequence $\{Z_n; n \geq 1\}$ with $Z_n := \exp\left(\sum_{i=1}^{n} X_i - \alpha_n\right)$ is a martingale with respect to $(\Im_n)_{n\geq 1}$, where $\Im_n = \sigma(X_1, X_2, \cdots, X_n)$ for $n \geq 1$.

11.9 **(Doob's descomposition)** Let $\{Y_n; n \geq 0\}$ be a submartingale with respect to the filtration $(\Im_n)_{n\geq 0}$. Show that

$$M_n = Y_0 + \sum_{k=1}^{n} (Y_n - E(Y_n \mid \Im_{n-1}))$$

for $n = 1, 2, \cdots$ is a martingale with respect to $(\Im_n)_{n\geq 0}$ and that the sequence $A_n := Y_n - M_n$, $n = 1, 2, \cdots$, satisfies $0 \leq A_1 \leq A_2 \leq \cdots$. Is A_n measurable with respect to $\Im_{n-1}$? Explain.

11.10 Let $X_1, X_2, \cdots$ be a sequence of independent random variables such that $E(X_n^2)$ exists for all $n = 1, 2, \cdots$ and suppose $S_n := X_1 + \cdots + X_n$, $n = 1, 2, \cdots$. Is $\{S_n^2; n \geq 1\}$ a submartingale? If it is so, then determine the process $\{A_n; n \geq 1\}$ as in the exercise above.

11.11 Let $\{X_n; n \geq 1\}$ be a sequence of random variables adapted to the filtration $(\Im_n)_{n\geq 1}$. Suppose that τ_1 is the time at which the process $\{X_n; n \geq 1\}$ reaches for the first time the set A and let:

$$\tau_2 := \min\{n \geq \tau_1 : X_n \in A\}.$$

Show that τ_2 is a stopping time. What does τ_2 represent?

11.12 Let τ be a stopping time with respect to the filtration $(\Im_n)_{n\geq 1}$ and k be a fixed positive integer. Show that the following random variables are stopping times: $\tau \wedge k$, $\tau \vee k$, $\tau + k$.

11.13 Let $\{X_n; n \geq 1\}$ be the independent random variables with $E[X_n] = 0$ and $Var(X_n) = \sigma^2$ for all $n \geq 1$. Set $M_0 = 0$ and $M_n = S_n^2 - n\sigma^2$, where $S_n = X_1 + X_2 + \cdots + X_n$. Is $\{M_n; n \geq 1\}$ a martingale with respect to the sequence X_n?

11.14 Let $\{N_t; t \geq 0\}$ be a Poisson process with rate λ and $\{\Im_t; t \geq 0\}$ is a filtration associated with N_t. Write down the conditional distribution of $N_{t+s} - N_t$ given $\Im_t$, where $s > 0$, and use your answer to find $E[\theta^{N_t - s} \mid \Im_s]$.

11.15 **(Lawler, 1996)** Consider the simple symmetric random walk model $Y_n = X_1 + X_2 + \ldots + X_n$ with $Y_0 = 0$, where the steps X_i's are independent and identically distributed with $P[X_k = 1] = 1/2$ and $P[X_k = -1] = 1/2$ for all k. Let $T := inf\{n : Y_n = -1\}$ denote the hitting time of -1. We know that $P[T < \infty] = 1$. Show that if $s > 0$, then $M_n := \frac{s^{Y_n}}{[\phi(s)]^n}$ with $M_0 = 1$ is a martingale, where $\phi(s) := (s^2 + 1)/2s$.

11.16 Let $X_1, X_2, \cdots$ be independent random variables such that

$$X_n = \begin{cases} a_n & \text{with probability } \frac{1}{2}n^{-2} \\ 0 & \text{with probability } 1 - n^{-2} \\ -a_n & \text{with probability } \frac{1}{2}n^{-2}, \end{cases}$$

where $a_1 = 2$ and $a_n = \sum_{i=1}^{n-1} a_i$. Is $Y_n = \sum_{i=1}^{n} X_i$ a martingale?

11.17 Let B_t be a Brownian motion. Find $E\left((B_t - B_s)^4\right)$.

11.18 Let $\{B_t; t \geq 0\}$ and $\{B_t'; t \geq 0\}$ be two independent Brownian motions. Show that

$$X_t = \frac{B_t + B_t'}{\sqrt{2}}$$

is also a Brownian motion. Find the correlation between B_t and X_t.

11.19 Let B_t be a Brownian motion. Find the distribution of $B_1 + B_2 + B_3 + B_4$.

11.20 Let $\{B_t; t \geq 0\}$ be a Brownian motion. Show that $e^{-\alpha t} B_{e^{2\alpha t}}$ is a Gaussian process. Find its mean and covariance functions.

11.21 Let $\{B_t; t \geq 0\}$ be a Brownian motion. Find the distribution for the integral

$$\int_0^t s dB_s.$$

11.22 S_t has the following differential equations:

$$dS_t = \mu S_t dt + \sigma S_t dB_t.$$

Find the equation for the process $Y_t = S_t^{-1}$.

11.23 Use the Itô formula to write down the stochastic differential equations for the following equations. $\{B_t; t \geq 0\}$ is a Brownian motion process.

a) $X_t = B_t^3$.

b) $Y_t = tB_t$.

c) $Z_t = \exp(ct + \alpha B_t)$.

11.24 Let:

$$I_t(B) = \int_0^2 e^{B_t} dB_t \ .$$

Find $E(I_t(B))$ and $E(I_t(B)^2)$.

11.25 Evaluate the integral

$$\int_0^t \frac{1}{1+B_s^2} dB_s \ .$$

11.26 Evaluate the integral

$$\int_0^t e^{B_s - \frac{1}{2}s} dB_s.$$

11.27 Suppose that X_t satisfies:

$$X_t = 2\int_0^t (3s + e^s)ds + \int_0^t \cos(s)dB_s \ .$$

Let $Y_t = f(t, X_t) = (2t+3)X_t + 4t^2$. Find Y_t.

11.28 Use Itô's formula to show that:

$$e^{B_t} - 1 = \int_0^t \frac{1}{2} e^{B_s} ds + \int_0^t e^{B_s} dB_s \ .$$

11.29 Consider the stochastic differential equation

$$dX_t = -\frac{1}{8}dt + \frac{1}{2}dB_t$$

with $X_0 = 0$.

a) Find X_t.

b) Let $Z_t = e^{X_t}$. Find the stochastic differential equation for Z_t using Itô's formula

11.30 Find the solution of the stochastic differential equation

$$dZ_t = Z_t dt + 2Z_t dB_t \ .$$

11.31 Solve the following stochastic differential equation for the spot rate of interest:

$$dr_t = (b - r_t)dt + \sigma dB_t$$

where r_t is an interest rate, $b \in \mathbb{R}$ and $\sigma \geq 0$.

11.32 Suppose that X_t follows the process $dX_t = 0.05X_t dt + 0.25X_t dB_t$. Using Itô's lemma find the equation for process $Y_t = \log X_t$ $Z_t = X_t^3$.

CHAPTER 12

INTRODUCTION TO MATHEMATICAL FINANCE

Mathematical finance is the study of financial markets and is one of the rapidly growing subjects in applied mathematics. This is due to the fact that in recent years mathematical finance has become an indispensable tool for risk managers and investors. The fundamental problem in the mathematics of financial derivatives is the pricing and hedging. During the years 1950–1960, the research focus was basically on the resolution of problems in economics and statistics. However, many researchers were concerned with the problem that was initiated in the early 20th century by Bachelier, the father of modern mathematical finance: what is the fair price for an option on a particular stock? The answer to this question was given in 1973 by researchers Fisher Black and Myron Scholes. They argued that a rational investor does not wait passively until the expiry of the contract but, in contrast, invests in a portfolio consisting of a risk-free asset and a risky stock, so that the value of this portfolio will be equal to the value of the option. Therefore, the fair price of the option is the present value of the portfolio. To obtain this value they generated a partial differential equation whose solution is known as the "Black-Scholes formula".

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

The initial focus on the valuations of financial derivatives was to describe the appropriate movements of stock prices and solving partial differential equations. However, another purely probabilistic approach was developed by Cox et al. (1979) among others, who considered the discrete case. Harrison and Kreps (1979) observed that for a market model to be sensible from the economic point of view, the discounted stock price process should be a martingale under an appropriate change of measure. This viewpoint was further expounded by Harrison and Pliska (1981) who observed that by making use of the Itô representation theorem, the present value of the price of an option can be characterized by a stochastic integral and the discounted price of the stock can be transformed into a martingale by a suitable change of measure. This approach is also known as the martingale pricing method or risk-neutral valuation. The objective of this chapter is to introduce risk-neutral valuation within the framework of financial derivatives. We will keep this chapter as simple as possible to present the basic modeling approach for the financial derivatives. In the next section we review the technical terminology used in the financial derivatives. Interested reader may refer to Hull (2009) for more details.

12.1 FINANCIAL DERIVATIVES

Definition 12.1 *A financial derivative, or contingent claim, is a financial contract whose value at expiration date T is determined exactly by the price of the underlying financial asset at time T.*

The financial derivatives or financial securities are financial contracts whose value is derived from some underlying assets. The assets could be stocks, currencies, equity indices, bonds, interest rates, exchange rates, and commodities. There are different types of financial markets, namely stock markets, bond markets, currency markets, and commodity markets. In general, financial derivatives can be grouped into three groups: options, forwards, and futures. The options constitute an important building block for pricing financial derivatives. Now we give a formal definition for an option.

Definition 12.2 *An option is a contract that gives the holder the right but not the obligation to undertake a transaction at a specified price at a future date.*

Note 12.1 *A specific or prescribed price is called a strike price, denoted by K. A specific or a future time is called an expiry date, denoted by T.*

There are two main types of option contracts: call options and put options.

1. A *call option* gives one the right to buy an asset at a specific price within a specific time period.

2. A *put option* gives the right to sell an asset at a specific price within a specific time period.

A call option is the right to buy (not an obligation). When a call is exercised, the buyer pays the writer (the other party of the contract, who does have a potential obligation that he must sell the asset if the holder chooses to buy) the strike price and gives up the call in exchange for the asset.

The contract for a call (put) option specifies:

1. The company whose shares are to be bought (sold)
2. The number of shares that can be bought (sold)
3. The purchase (selling) price
4. The date when the right to buy (sell) expires

Based on how they are exercised, options can be divided into two types:

1. *European option* that can be exercised only at the time of expiry of the contract.
2. An *American option* that can be exercised at any time up to the expiry date.

The call or put option described above is also called a simple or vanilla option. An option which is not a vanilla option is called an exotic option. The well-known exotic option is a path-dependent option where the payoff depends on the past and present values of an underlying asset. The following exotic options are widely used.

1. An *Asian option* is a type of option whose payoff function is determined by the average price of an underlying asset over some period of time before expiry.
2. A *barrier option* is a type of option whose payoff depends on whether or not the underlying asset has reached or exceeded a predetermined price
3. A *lookback option* is an option whose payoff is determined by the maximum or minimum of the underlying asset.
4. A *perpetual option* has no expiry date (that is, it has an infinite time horizon).

In the following sections some financial terms are used. The practical meaning of these terms are as follows:

1. *Intrinsic value*: At time $t \leq T$, the intrinsic value of a call option with strike price K is equal to $\max(S_t - K, 0)$. Similarly for a put option, the intrinsic value is $\max(K - S_t, 0)$.

2. *Time value*: The time value of an option is the difference between the price of the option and its intrinsic value. That is, for a European call option, the time value is $C(S_t, t) = \max(S_t - K, 0)$.

3. *Long position*: A party assumes a long position by agreeing to buy an asset for the strike price K at expiry date T because he anticipates the price to go up. That is, one can have positive amounts of an asset by buying assets that one does own.

4. *Short position*: The opposite of assuming a long position in which one party agrees to sell the asset because he anticipates the price to decline. That is, one can have negative amounts of an asset by selling assets that one does not own.

5. *Arbitrage*: This refers to buying an asset at one place and selling it for a profit at another place without taking any risk. We will give a precise mathematical definition in the next section.

6. *Hedging*: This is the concept of elimination of risk by taking opposite positions in two assets whose prices are correlated.

Forwards and Futures

Forwards and futures are both contracts to deliver a specific asset at an agreed date in the future at a fixed price. Unlike an option, there is no choice involved as to whether the contract is completed.

Definition 12.3 *A forward contract is an agreement to buy or sell an asset S at certain future date T for a certain price K.*

A forward contract is usually between large and sophisticated financial agents (banks, institutional investors, large corporations and brokerage firms) and is not traded in an exchange. The agent who agrees to buy the underlying asset is said to have a long position while the other agent assumes a short position. The settlement date is called delivery date and the specified price is referred to as the delivery price.

Two persons or parties can enter into a forward contract to buy or sell Colombian coffee at any time with any delivery date and delivery price they choose. When a forward contract is created at time 0 with delivery date T for delivery price K, this delivery price is also referred as the forward price at time 0 for delivery date T. At the time t between 0 and T, it would be possible to create a new forward contract with the same delivery date T. The delivery price of this new contract is also called the forward price at the time t with delivery date T and might not be the same as the delivery price K which was the delivery price with delivery date T on the contract that was set up at time $t = 0$.

The disadvantage of forwards is that they are hardly traded at exchanges, the reason being the individual, nonstandardized character of the contract, and that they are exposed to the risk of default by their counter parties. Therefore, there are significant costs in finding a partner for the contract. These problems lead to standardization with the aim of making trading at an exchange possible.

A *futures* contract, although like a forward contract, differs from it as shown below.

1. A forward is an over-the-counter agreement between two individuals. In contrast, a future is a trade organized by an exchange.

2. A forward is settled on the delivery date. That is, there is a single payment made when the contract is delivered. The profit/loss on a future is settled on a daily basis to avoid default. The exchange acts as an intermediary between the long and short parties and insists on maintained margins.

A basket of forwards and/or options with different delivery or expiry dates is called a *swap*, which is a more complex financial derivative. A swap is an agreement where two parties undertake to exchange, at known dates in the future, various financial assets (or cash flows) according to a prearranged formula that depends on the value of one or more underlying assets. The two commonly used swaps are interest rate swaps and currency swaps.

The focus of this chapter will be on the pricing and hedging of options. We will restrict ourselves to European and American call and put options. The option buyer has the right but not the obligation, which puts him in a privileged position. The option seller, that is, the option writer, provides the privilege which the buyer holds. We have seen two types of options, namely, option to buy, a call option, and option to sell, a put option. Using asymmetry in rights, there are four possibilities:

1. Buy an option to buy, i.e., buy call

2. Buy an option to sell, i.e., buy put

3. Sell an option to buy, i.e., write call

4. Sell an option to sell, i.e., write put

The *payoff*, or the value, of an option depends on the value of an underlying asset at a future time T. Consider a European call option with strike price K. Let S_T be the price of the underlying asset at the time of expiry T. If $S_T < K$, that is the option is expiring *out-of-money*, then the holder can buy the asset in the market at S_T, a cost less than the strike price K. The terminal payoff of the long position in a European call option is 0. But if, at time T, $S_T > K$, that is the option expires *in-the-money*, then the holder of the call option will choose to exercise the option. The holder can buy the

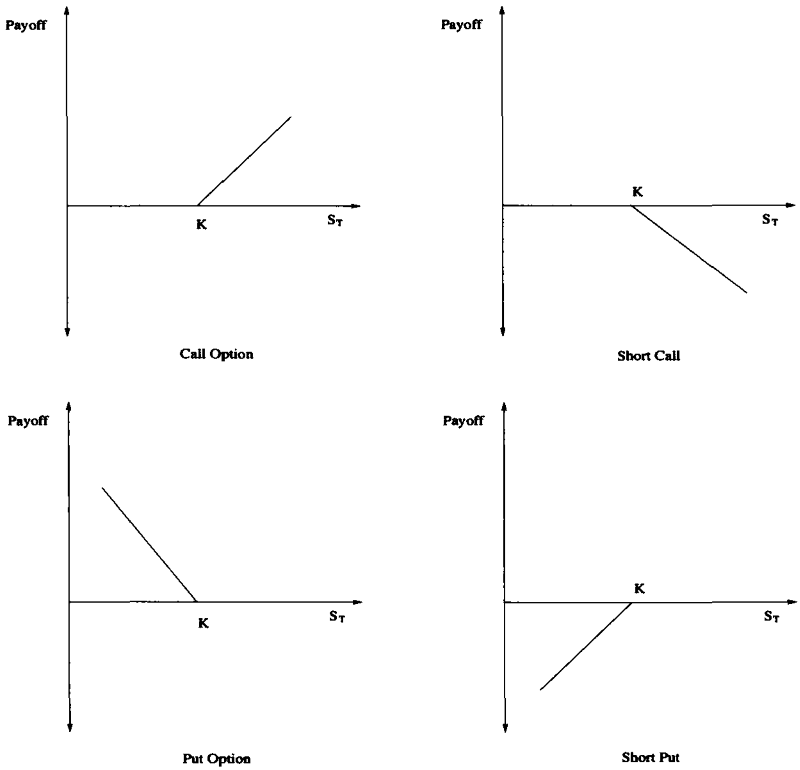


Figure 12.1 Payoff diagram of call and put options

asset at a price K and sell it at a higher price S_T. The net profit will be $S_T - K$. Both cases can be summarized mathematically as:

$$C(S_t, T) = \max(S_T - K, 0) \quad \text{or} \quad C(S_t, T) = (S_T - K)^+,$$

where $(x)^+ = \max(x, 0)$. An argument similar to the value of a call option leads to the payoff for a put option at expiry date T and is given by:

$$P(S_t, T) = \max(K - S_T, 0) \quad \text{or} \quad P(S_t, T) = (K - S_T)^+.$$

A chart of the profits and losses for a particular option is called a *payoff diagram* and is a plot of the expected profit or loss against the price of the underlying asset on the same future date. The payoff diagrams for call and put options are given in Figure 12.1.

For European options, there is a simple theoretical relationship between the prices of the corresponding call and put options. This important relationship is known as *put-call parity* and may be expressed as:

$$C(S_t, T) - S_t = P(S_t, T) - Ke^{-r(T-t)},$$

where K is the strike price of the underlying stock S, T is the expiry date and r is the constant interest rate; sometimes we use $C = C(S_t, T)$ and $P = P(S_t, T)$

■ **EXAMPLE 12.1**

Suppose that a share of a certain stock is worth \$100 today for a strike price of \$110 with one year expiry date. The difference between the costs of put and call options is \$5. The interest rate is calculated by using put-call parity. We have:

$$\begin{aligned} S + P - C &= Ke^{-rt} \\ 100 + 5 &= 110e^{-r} \\ r &= -\ln\left(\frac{105}{110}\right) \\ r &= 0.0465 = 4.65\%. \quad \blacktriangle \end{aligned}$$

The main assumptions of the options are:

1. There are some market participants.
2. There are no transaction costs.
3. All trading profits (net of trading losses) are subject to the same tax rate.
4. Borrowing and lending at the risk-free rate are possible.

We now establish an interesting result between European and American call options.

Theorem 12.1 *With a nonnegative interest rate r, it is not optimal to exercise early an American call option on a non-dividend paying asset with the same strike price and expiry date.*

Proof: Suppose that it is optimal to exercise the American call option before expiry date.

Let C_A and C_E be the value of the European and American options, respectively. It is easily seen that the payoff value of the European call options:

$$C_E = S - Ke^{-r(T-t)}.$$

We know that:

$$C_A > C_E$$

This is because an American option has the benefits of having the right of early exercise. Also, we have:

$$C_A > S - Ke^{-r(T-t)}.$$

Since $r > 0$, it follows that $C_A > S - K$. An American option will always be worth at least its payoff $S - K$. However, the previous equation shows that $C_A > S - K$. Thus, we have a contradiction based on the assumption that it is optimal to exercise early. Hence the value of the American option must be equal to the value of the European call option. However, it can be advantageous to exercise an American put option on a non-dividend paying asset. ∎

In the continuation we present a discrete-time model and a continuous-time model for the valuation of financial derivatives. We closely follow the approach presented by Lamberton and Lapeyre (1996), Bingham and Kiesel (2004), and Williams (2006).

12.2 DISCRETE-TIME MODELS

In this section we give a brief introduction to discrete-time models of a *finite market*, $\mathcal{M}$, that is, models with a finite number of trading dates in which all asset prices take a finite number of values. The finite market model is considered on a finite probability space $(\Omega, \Im, P)$, where $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, $\Im$ is a σ-algebra consisting of all subsets of Ω and P is a probability measure on $(\Omega, \Im)$ such that $P(\{\omega\}) > 0$, $\forall \omega \in \Omega$. We assume that there is a finite number of trading dates $t = 0, 1, 2, \cdots, N$ where $N < \infty$ at which trades occur. The time N is the terminal date or maturity of the financial contract considered. We now use a finite filtration $(\Im_n)_{n\geq 0}^N = \{\Im_n; n = 0, \cdots, N\}$ to model the information flow over time. Further, $\Im_0 = \{\emptyset, \Omega\}$ and $\Im_N = \Im_n = \mathcal{P}(\Omega)$.

We assume that the market consists of $k+1$ assets whose prices at time n are given by the nonnegative random variables $R_n, S_n^1, \cdots, S_n^k$ which are measurable with respect to $(\Im_n)_{n\geq 0}^N$. We assume that there exists one risk-free asset on the money market, known as a bond, R, and k risky assets or stocks on the financial market, $S = (S^1, S^2, \cdots, S^k)$.

Definition 12.4 *A numéraire is a price process $\{X_n; n = 0, \cdots, N\}$ which is strictly positive for all $n = 0, 1, \cdots, N$.*

For example, suppose that $R_0 = 1$. If the return of a bond over a period is a constant r, then $R_n = (1+r)^n$. The coefficient $\frac{1}{R_n} = (1+r)^{-n}$ is interpreted as the discount factor. For $i = 1, 2, \cdots, k$, we define:

$$\widetilde{S_n^i} = \frac{S_n^i}{R_n} \qquad \text{for } n = 0, 1, \cdots, N$$

Then $\widetilde{S_n} = \left(\widetilde{S_n^1}, \widetilde{S_n^2}, \cdots, \widetilde{S_n^k}\right)$ is the vector of discounted stock prices at time n. The investors are interested in creating a dynamic portfolio which involves trading strategies and a portfolio process. Hence we now give the definition of the trading strategies and portfolio or wealth processes.

Definition 12.5 *A trading strategy in the finite market model is a collection of $k+1$ random variables $\varphi = (\varphi_n^0, \varphi_n^1, \cdots, \varphi_n^k)$ where φ_n^i is a real-valued $\Im_{n-1}$ measurable random variable for $i = 0, 1, \cdots, k$. We interpret φ_n^i as the number of shares of asset i to be held over the time $(n-1, n]$.*

■ EXAMPLE 12.2

For a financial market model with a single bond and a single stock, the trading strategy is a collection of pairs of random variables

$$\varphi = \{(\alpha_n, \beta_n); n = 0, 1, \cdots, N\}.$$

The random variable α_n represents the number of shares of stock to be held over the time interval $(n-1, n]$, and the random variable β_n represents the number of units of the bond to be held over the same time interval. Here both α_n and β_n are $\Im_{n-1}$ measurable and hence are called predictable random variables. ▲

Definition 12.6 (Value of Portfolio or Wealth Process) *The value of the portfolio or wealth process $V_n(\varphi)$ of a trading strategy φ at time n equals:*

$$V_n = V_n(\varphi) = \varphi_n^0 R_n + \sum_{i=1}^{k} \varphi_n^i S_n^i.$$

The discounted values are given by $\widetilde{V}_n = \frac{1}{R_n} V_n(\varphi)$.

For simplicity we consider the case of two assets in our model, a risk-free money market account, a bond R, and a risky asset, the stock S. We assume that the stock price is adapted to the filtration $(\Im_n)$ for the trading dates $n = 0, 1, 2, \cdots, N$.

Definition 12.7 *A trading strategy is called self-financing if for all $n = 1, \cdots, N$:*

$$V_n = \beta_n R_n + \alpha_n S_n = \beta_{n+1} R_n + \alpha_{n+1} S_n. \tag{12.1}$$

This implies that no funds are withdrawn from or added the strategy.

Theorem 12.2 *A trading strategy φ is self-financing if and only if we have for all $n = 1, \cdots, N$:*

$$V_n(\varphi) = V_0(\varphi) + \sum_{j=1}^{n} \beta_j (R_j - R_{j-1}) + \sum_{j=1}^{n} \alpha_j (S_j - S_{j-1}) \ . \tag{12.2}$$

Hence the value of a self-financing strategy consists of the initial investment V_0 and the gains (or losses) from the trade stock and money market account.

Proof: From the definition of the value of a portfolio we get that:

$$V_{n+1}(\varphi) - V_n(\varphi) = \beta_{n+1}R_{n+1} + \alpha_{n+1}S_{n+1} - \beta_n R_n - \alpha_n S_n. \tag{12.3}$$

Now φ is self-financing if and only if $\beta_{n+1}R_n + \alpha_{n+1}S_n = \beta_n R_n + \alpha_n S_n$ for all $n = 1, \cdots, N$. Substituting this into (12.3) gives:

$$V_{n+1}(\varphi) - V_n(\varphi) = \beta_{n+1}(R_{n+1} - R_n) + \alpha_{n+1}(S_{n+1} - S_n). \tag{12.4}$$

As $V_{n+1}(\varphi) = V_0(\varphi) + \sum_{i=0}^{n} V_{i+1}(\varphi) - V_i(\varphi)$, the lemma follows by summing over (12.4). ∎

We give a similar characterization of a self-financing strategy in terms of the discounted value process.

Theorem 12.3 *A trading strategy is self-financing if and only if we have for all $n = 1, \cdots, N$:*

$$\widetilde{V}_n(\varphi) = \widetilde{V}_0(\varphi) + \sum_{j=1}^{n} \alpha_j \left(\widetilde{S}_j - \widetilde{S}_{j-1}\right). \tag{12.5}$$

Proof: The proof of this theorem is omitted as it is similar to the proof of the previous theorem. ∎

Definition 12.8 *A trading strategy $\varphi = \{\varphi_n; n = 1, \cdots, N\}$ is called admissible if it is self-financing with $V_n(\varphi) > 0$ for $n = 1, 2, \cdots, N$.*

Definition 12.9 (Arbitrage Opportunity) *An arbitrage opportunity in the financial markets is a trading strategy φ such that $V_0(\varphi) = 0$, $V_N(\varphi) \geq 0$ and $E\left(V_N(\varphi)\right) > 0$.*

Definition 12.10 *A finite market model $\mathcal{M}$ is called viable or arbitrage-free, if there are no arbitrage opportunities.*

Note 12.2 (European Contingent Claim) *The value of the European contingent claim (ECC) $H \geq 0$ with expiry date N is an $\Im_N$ measurable random variable. The random variable H is a known payoff of the claim. For example, a European call option with strike price K and expiry date N is:*

$$H = \max(S_N - K, 0).$$

Definition 12.11 (Attainable or Replicating Strategy) *Given a finite market model $\mathcal{M}$, a contingent claim H with maturity N is called attainable if there is an admissible, self-financing strategy $\varphi_n = (\alpha_n, \beta_n)$ such that $V_N(\varphi) = H$; φ is called a replicating strategy for the derivative.*

Definition 12.12 *A finite market $\mathcal{M}$ is called complete if every contingent claim is attainable or equivalently, if for every $\Im_N$-measurable random variable H there exists at least one trading strategy φ such that $V_N(\varphi) = H$.*

The completeness of the finite market model is a highly desirable property. Under market completeness, any European or American contingent claim can be priced by an arbitrage-free and self-financing strategy. If the market is not complete, then it is called an *incomplete market model*.

In order to characterize arbitrage-free markets, we use the concept of equivalent martingale measures.

Definition 12.13 *Let $\mathcal{M}$ be a finite market model. A probability measure Q on $(\Omega, \Im)$ such that*

1. *Q is equivalent to P, i.e., for all $A \in \Im$ we have $Q(A) = 0 \Leftrightarrow P(A) = 0$, and*
2. *the discounted stock price $\widetilde{S}$ is a martingale under measure Q*

is called an equivalent martingale measure or a risk-neutral measure for $\mathcal{M}$.

Note 12.3

1. *The measure Q is said to be absolutely continuous with respect to P, i.e., for all $A \in \Im$ we have $P(A) = 0 \Rightarrow Q(A) = 0$.*
2. *A martingale under measure Q is referred as a Q-martingale.*

Theorem 12.4 *Let Q be an equivalent martingale measure for the market $\mathcal{M}$. Consider a self-financing, admissible trading strategy φ. Then the discounted process $\widetilde{V}_n(\varphi)$ is a Q-martingale.*

Proof: As φ is self-financing, we get from Theorem 12.3:

$$\widetilde{V}(\varphi) = \widetilde{V}_0(\varphi) + \sum_{j=1}^{n} \alpha_j \left(\widetilde{S}_j - \widetilde{S}_{j-1}\right) = \widetilde{V}_n(\varphi) + \alpha_{n+1} \left(\widetilde{S}_{n+1} - \widetilde{S}_n\right) .$$

As φ is admissible, φ_{n+1} is $\Im_n$-measurable and $\widetilde{S}$ is a Q-martingale:

$$\begin{aligned} E_Q\left(\widetilde{V}_{n+1}(\varphi) - \widetilde{V}_n(\varphi) \mid \Im_n\right) &= E_Q\left(\alpha_{n+1}(\widetilde{S}_{n+1} - \widetilde{S}_n) \mid \Im_n\right) \\ &= \alpha_{n+1} E_Q\left(\widetilde{S}_{n+1} - \widetilde{S}_n \mid \Im_n\right) = 0. \end{aligned}$$

■

Theorem 12.5 *If an equivalent martingale measure exists for the finite market model $\mathcal{M}$, then the model $\mathcal{M}$ is arbitrage-free.*

Proof: Consider a self-financing strategy φ with $V_N(\varphi) \geq 0$, $E\left(V_N(\varphi)\right) > 0$. We will show that the existence of an equivalent martingale measure Q gives $V_0(\varphi) > 0$, which implies that the finite market model $\mathcal{M}$ is arbitrage-free. If $V_N(\varphi)$ and $\widetilde{V}_N(\varphi)$ have the same sign, then it follows that $\widetilde{V}_N(\varphi) > 0$ and

$E\left(\widetilde{V}_N(\varphi)\right) > 0$.

We have that the measure Q is equivalent to P, which implies that $Q\left(\widetilde{V}_N(\varphi) > 0\right) > 0$ and this gives $E_Q(\widetilde{V}_N(\varphi)) > 0$. We know that $\{\widetilde{V}_n(\varphi)); n = 1, \cdots, N\}$ is a Q-martingale, which shows that $\widetilde{V}_0(\varphi) = E_Q\left(\widetilde{V}_N(\varphi)\right) > 0$, and therefore, we conclude that $V_0(\varphi) > 0$. ■

We now state the first fundamental theorem of asset pricing .

Theorem 12.6 *A finite market $\mathcal{M}$ is arbitrage-free if and only if there is a probability measure Q equivalent to P such that the discounted asset price processes are Q-martingales.*

Proof: For the proof, see Williams (2006). ■

Theorem 12.7 *Let $\mathcal{M}$ be an arbitrage-free market with an attainable contingent claim H and replicating strategy φ. If Q is an equivalent martingale measure for $\mathcal{M}$, then the fair price of the claim H at time $n \leq N$ is given by*

$$V_n(\varphi) = E_Q\left((1+r)^{-(N-n)}H \mid \Im_n\right) \tag{12.6}$$

and, in particular, we have:

$$V_0(\varphi) = E_Q\left((1+r)^{-N}H\right). \tag{12.7}$$

Proof: From the replicating strategy φ of the claim, we have $V_N(\varphi) = H$ and hence $(1+r)^{-N}H = \widetilde{V}_N(\varphi)$. We know that $\{\widetilde{V}_n(\varphi); n = 0, \cdots, N\}$ is a Q-martingale, and hence by Theorem 12.4, we have:

$$E_Q\left((1+r)^{-N}H \mid \Im_n\right) = E_Q\left(\widetilde{V}_N(\varphi) \mid \Im_n\right) = \widetilde{V}_n(\varphi) = (1+r)^{-n}V_n(\varphi). \tag{12.8}$$

Therefore, $V_n(\varphi) = E_Q\left((1+r)^{-(N-n)}H \mid \Im_n\right)$. ■

We have seen in the first fundamental theorem of asset pricing that the existence of an equivalent martingale measure implies that the finite market model is arbitrage-free. We will now state without proof a second fundamental theorem of asset pricing which gives us the necessary and sufficient condition for the existance of complete market[see Harrison and Kreps (1979), and Harrison and Pliska (1981)].

Theorem 12.8 *An arbitrage-free market $\mathcal{M}$ is complete if and only if there exists a unique equivalent martingale measure Q.*

Proof: For a proof we refer to Williams (2006). ■

As an example we now present the binomial model of Cox et al. (1979). This simple model provides a useful computing method for pricing financial derivatives.

12.2.1 The Binomial Model

This section considers a single discrete-time financial market model known as the *binomial model* or *Cox-Ross-Rubinstein (CRR) model*. In this model, we assume that there are a finite number of trading times $t = 1, 2, \cdots, N$, where $N < \infty$. At each of these time instants, the values of two assets are observed. The risky asset is called a *stock* and the risk-free asset a *bond*. The bond is assumed to yield a deterministic rate of return r over each time period $(n-1, n]$. We assume that the bond is valued at \$1 at time $n = 0$, i.e., $R_0 = 1$. The value of the bond at any time n is given by

$$R_n = (1+r)^n \qquad \text{for } n = 0, 1, 2, \cdots, N \tag{12.9}$$

We suppose that the stock prices in each time interval are independent and identically distributed random variables each of which goes up by a factor u or down by a factor d. We assume the stock price as a random walk such that

$$S_n = S_{n-1}\zeta_n \qquad \text{for } n = 1, 2, \cdots, N \tag{12.10}$$

where $\{\zeta_j; j = 1, \cdots, N\}$ is a sequence of independent and identically distributed random variables with

$$\begin{aligned} P(\zeta_j = u) &= p \\ P(\zeta_j = d) &= q \end{aligned}$$

such that $p + q = 1$. The binomial tree for the stock price process $\{S_n; n = 0, 1, \cdots, N\}$ for $N = 4$ is represented in Figure 12.2.

We assume that there exists a money market so that one can invest or borrow money with a fixed interest rate r such that $0 < d < 1+r < u$. This condition is necessary to avoid arbitrage opportunities in the market model. Equation (12.10) can be written as:

$$S_n = S_0 \prod_{i=1}^{n} \zeta_i \qquad \text{for } n = 1, 2, \cdots, N. \tag{12.11}$$

For the binomial model, one can easily enumerate the probability space $(\Omega, \Im, P)$. The sample space consists of 2^N possible outcomes for the values of the sequence $\{\zeta_1, \zeta_2, \cdots, \zeta_N\}$. The σ-algebra $\Im$ contains all possible subsets of Ω and P is the probability measure associated with N independent Bernoulli trials, each with probability p. In the context of the previous chapter, $\Im_n := \sigma(\zeta_1, \zeta_2, \cdots, \zeta_n)$ for $1 \le n \le N$ and $(\Im_n)_{n=1}^N$ is the filtration.

We use a simple model to explain the important results in pricing financial derivatives. Consider a European contingent claim H. Suppose that we have a single stock with price S_0 at time $n = 0$. Assume that there is a single period to the expiry of the European contingent claim H.

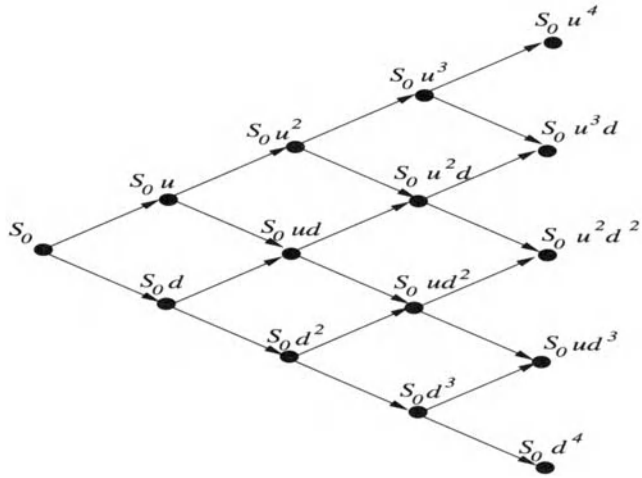


Figure 12.2 Binomial tree for $N = 4$

The fundamental assumption of the binomial model is that the price of the stock may take one of the two values at the expiration date. Either it will be uS_0 with probability p or it will be dS_0 with probability $q = 1 - p$ where $0 < d < u$. If there is a risk-free asset with interest rate r, as we remarked earlier, we assume that $0 < d < 1 + r < u$ to avoid arbitrage opportunities in the market model.

The objective is to find a strategy $\varphi = (\alpha_1, \beta_1)$ where $\alpha_1, \beta_1 \in \mathbb{R}$ such that

$$V_1(\varphi) = \alpha_1 S_1 + \beta_1 R_1 = H$$

where $R_1 = (1 + r)$ and $S_1 = S_0\zeta_1$ with

$$\begin{aligned} P(\zeta_1 = u) &= p \\ P(\zeta_1 = d) &= q \end{aligned}$$

such that $p + q = 1$.

The European contingent claim H takes two positive values H^u and H^d respectively when $\zeta_1 = u$ and $\zeta_1 = d$:

$$\begin{aligned} H^u &= \alpha_1 S_0 u + \beta_1 (1 + r) \\ H^d &= \alpha_1 S_0 d + \beta_1 (1 + r) \, . \end{aligned} \tag{12.12}$$

Solving the above system of equations for the two unknowns (α_1, β_1), we get:

$$\begin{aligned} \alpha_1 &= \frac{H^u - H^d}{(u - d)S_0} \\ \beta_1 &= \frac{1}{1 + r}\left(\frac{uH^d - dH^u}{u - d}\right) \, . \end{aligned}$$

Set:

$$p^* = \frac{1+r-d}{u-d} \tag{12.13}$$

$$q^* = \frac{u-(1+r)}{u-d} \tag{12.14}$$

$$H^* = \frac{H}{1+r}\,. \tag{12.15}$$

Then the initial wealth required to finance a strategy which results in values H^u and H^d is given by

$$\begin{aligned} V_0(\varphi) &= \alpha_1 S_0 + \beta_1 R_0 \\ &= \frac{H^u - H^d}{(u-d)S_0} S_0 + \frac{1}{1+r}\left(\frac{uH^d - dH^u}{u-d}\right) \\ &= \frac{1}{1+r}\left(\frac{1+r-d}{(u-d)} H^u + \frac{u-(1+r)}{u-d} H^d\right) \\ &= \frac{1}{1+r}\left(p^* H^u + q^* H^d\right) \\ V_0 &= E_Q(H^*) \end{aligned}$$

where $E_Q(\cdot)$ denotes the expectation with respect to the measure $Q = (p^*, q^*)$. The initial wealth V_0 is also known as the *manufacturing* cost of the contingent claim. The measure $Q = (p^*, q^*)$ is called the *risk-neutral probability*. The above equation suggests that the value of the European contingent claim is a discounted expectation of the final value under the probability measure Q. Hence it is called a risk-neutral valuation. Here we wish to note that the actual probabilities of the stock price do not appear in the valuation formula of the European contingent claim.

Note 12.4 *An investor could be classified according to his preference for risk taking:*

1. *An investor who prefers to take the expected value of a payoff to the random payoff is said to be risk averse.*
2. *An investor who prefers the random payoff to the expected payoff is said to be risk preferring.*
3. *An investor who is neither risk averse nor risk preferring is said to be risk-neutral.*

Theorem 12.9 *The discounted stock price process* $\{(1+r)^{-n}S_n; n = 0, 1, \cdots, N\}$ *is a martingale under measure* Q.

Proof: Since

$$E_Q\left(S_{n+1} \mid \Im_n\right) = (p^* u + q^* d) S_n, \quad p^* = \frac{1+r-d}{u-d}, \quad q^* = \frac{u-(1+r)}{u-d},$$

we have:

$$\begin{aligned} E_Q\left((1+r)^{-n}S_n \mid \Im_{n-1}\right) &= (1+r)^{-n}(p^*u+q^*d)S_{n-1} \\ &= (1+r)^{-n}\left(\frac{1+r-d}{u-d}u+\frac{u-(1+r)}{u-d}d\right)S_{n-1} \\ &= (1+r)^{-n}\left(\frac{(u-d)(1+r)}{u-d}\right)S_{n-1} \\ &= (1+r)^{-(n-1)}S_{n-1}\,. \end{aligned}$$

Thus we proved that $\{(1+r)^{-n}S_n; n=0,1,\cdots,N\}$ is a martingale under measure Q. ∎

12.2.2 Multi-Period Binomial Model

We now extend a one-period binomial model to an N-period model. The state space $\Omega = \{u,d\}^N$ is such that the elements of Ω are N-tuples with $\{u,d\}$.

For $1 \le n < N$, we now show that there is a replication strategy for a European contingent claim H. Let $\varphi = \{(\alpha_n,\beta_n) : n=1,\cdots,N\}$ such that:

$$V_N(\varphi) = \alpha_N S_N + \beta_N R_N = H. \tag{12.16}$$

Let $V_N = H$ where H is an $\Im_N$-measurable random variable. Consider the time period $(N-1,N]$ by using a single period model as before. We obtain

$$\alpha_N = \frac{V_N^u - V_N^d}{(u-d)S_{N-1}} \tag{12.17}$$

$$\beta_N = \frac{1}{1+r}\left(\frac{uV_N^d - dV_N^u}{(u-d)}\right) \tag{12.18}$$

and

$$V_{N-1} = \frac{1}{1+r}E_Q\left(V_N \mid \Im_{N-1}\right) \tag{12.19}$$

where Q is the risk-neutral measure.

We can find a trading strategy $\varphi = \{(\alpha_n,\beta_n) : n=1,\cdots,N\}$ with value process $V_n(\varphi)$ such that $V_N = H$. We do this by working backward through the binomial tree.

For $n=1,2,\cdots,N-1$, we assume that the self-financing strategies $\{(\alpha_n,\beta_n) : n=1,\cdots,N\}$ have been determined during the time interval $(n-1,n]$ with value process $V_1, V_2, \cdots, V_{N-1}$ such that

$$V_m = \frac{1}{(1+r)^{N-m}}E_Q\left(H \mid \Im_m\right)$$

for $m = n, n+1, \cdots, N-1$.

Let V_n^u and V_n^d be the two possible values of V_n. We define

$$\alpha_n = \frac{V_n^u - V_n^d}{(u-d)S_{n-1}} \tag{12.20}$$

$$\beta_n = \frac{1}{(1+r)^n}\left(\frac{uV_n^d - dV_n^u}{(u-d)}\right) \tag{12.21}$$

and

$$\begin{aligned}
V_{n-1} &= \frac{1}{1+r}E_Q\left(V_n \mid \Im_{n-1}\right) \\
&= \frac{1}{1+r}E_Q\left(E_Q\left(V_{n+1} \mid \Im_n\right) \mid \Im_{n-1}\right) \\
&= \frac{1}{(1+r)^2}E_Q\left(V_{n+1} \mid \Im_{n-1}\right) \\
&\vdots \\
&= \frac{1}{(1+r)^{N-n+1}}E_Q\left(V_N \mid \Im_{n-1}\right) \\
V_{n-1} &= \frac{1}{(1+r)^{N-n+1}}E_Q\left(H \mid \Im_{n-1}\right) \qquad \text{(since } V_N = H).
\end{aligned}$$

In particular:

$$V_0 = \frac{1}{(1+r)^N}E_Q(H). \tag{12.22}$$

■ **EXAMPLE 12.3**

We now give an example of a multiperiod pricing formula for a European call option. We suppose as before that the stock prices in each time interval are independent and identically distributed random variables each of which goes up by a factor u with probability p or down by a factor d with probability q. We know now how to price a contingent claim in a one-period model.

A European call option has a payoff at the time of expiry N as $H = \max(S_N - K, 0)$. The value of the European call option $C_n = C(S_n, n)$ at time $0 \leq n < N$ is given by:

$$V_n = \frac{1}{(1+r)^{N-n}}E_Q\left(H \mid \Im_n\right)$$

As we illustrated earlier in Figure 12.2 for $N = 4$, the pricing formula for $n = 0, 1, \cdots, N-1$, we have

$$C_n = (1+r)^{-N+n}\sum_{i \in E(n)}\left(S_n u^i d^{N-n-i} - K\right)\binom{N-n}{i}p^{*i}q^{*N-n-i},$$

where $E(n)$ is the set of indices i for which $S_n u^i d^{N-n-i} > K$. Notice that $E(n)$ is possibly empty in $\{0, \cdots, N-n\}$. Set $a_n = \min E(n)$ to get:

$$
\begin{aligned}
C_n &= (1+r)^{-N+n} \sum_{i=a_n}^{N-n} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} p^{*i} q^{*N-n-i} \\
&= (1+r)^{-N+n} \sum_{i=a_n}^{N-n} (S_n u^i d^{N-n-i} - K) \binom{N-n}{i} p^{*i} q^{*N-n-i} \\
&= S_n \sum_{i=a_n}^{N-n} \binom{N-n}{i} \left(\frac{up^*}{1+r}\right)^i \left(\frac{dq^*}{1+r}\right)^{N-n-i} \\
&\quad - K(1+r)^{n-N} \sum_{i=a_n}^{N-n} \binom{N-n}{i} p^{*i} q^{*N-n-i}. \qquad (12.23)
\end{aligned}
$$

The above formula is known as the *CRR formula* for binomial pricing models. We observe that both sums in equation (12.23) can be expressed in terms of binomial probabilities. Let $\Pi(k, p_u, a)$ be the probability that the cumulative binomial probability $\mathcal{B}(k, p_u)$ is larger than or equal to a with $p_u = \frac{up^*}{1+r}$. Then equation (12.23) can be rewritten as:

$$C_n = S_n \Pi(N-n, p_u, a_n) - K(1+r)^{n-N} \Pi(N-n, q_d, a_n). \quad \blacktriangle \qquad (12.24)$$

Note 12.5 *Let us assume that trading takes place during the interval* $[0, T]$. *We consider the stock price process introduced earlier for the fixed time interval* $[0, T]$ *divided into* N *subintervals of equal length* Δ_N *with* $\Delta_N = \frac{T}{N}$. *We let the parameters depend on* N *as follows. For a given volatility of the stock,* $\sigma > 0$, *the step interest rate up and down factors are given by:*

$$
\begin{aligned}
r_N &= \exp(r\Delta_N) - 1 && (12.25) \\
u_N &= \exp(\sigma\sqrt{\Delta_N}) && (12.26) \\
d_N &= \exp(-\sigma\sqrt{\Delta_N}). && (12.27)
\end{aligned}
$$

Assume $N \to \infty$. *At any time, the fair price of a European call option with payoff* $\max(S^N(T) - K, 0)$ *has the limiting value*

$$S\Phi(d_1(t)) - Ke^{-r(T-t)}\Phi(d_2(t)), \qquad (12.28)$$

where:

$$
\begin{aligned}
d_1(t) &= \frac{\log(S/K) + (r + \frac{1}{2}\sigma^2)(T-t)}{\sigma\sqrt{T-t}} \\
d_2(t) &= d_1(t) - \sigma\sqrt{T-t}\,.
\end{aligned}
$$

Equation (12.28) is the famous Black-Scholes formula. For more details see Bingham and Kiesel (2004).

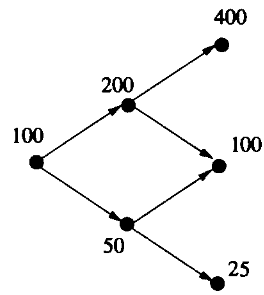


Figure 12.3 Binary tree

■ EXAMPLE 12.4

[Williams (2006), page 28]. Consider a European call option in the CRR model with $N = 2$, $S_0 = \$100$ and $S_1 = \$200$ or $S_1 = \$50$ with strike price $K = \$80$ and expiry date $N = 2$. Assume the risk-free interest rate $r = 0.1$.

(a) Calculate the arbitrage-free price for the European call option at time $n = 0$.

(b) Find the hedging strategy for this option.

(c) Suppose that the call option is initially priced \$2 below the arbitrage-free price. Describe a strategy that gives an arbitrage.

(d) Using put-call parity, find the value of the European put option at time $n = 0$.

Solution: We have $N = 2$, $S_0 = \$100$, $u = 2$, $d = \frac{1}{2}$, $r = 0.1$ and $K = \$80$. We calculate the risk-neutral probability

$$p^* = \frac{1 + r - d}{u - d} = \frac{1.1 - 0.5}{2 - 0.5} = \frac{2}{5}.$$

The binary tree for the stock price is given in Figure 12.3.

(a) The possible values for European call option $H = \max(S_2 - K, 0)$ are \$0, \$20, \$320 respectively. By using the CRR formula for the European call option, we have:

$$\begin{aligned} C_0 &= \frac{1}{(1+r)^2} E_Q(H) \\ &= \frac{1}{(1.1)^2}\left(p^{*2}(320) + 2p^*q^*(20) + q^{*2}(0)\right) \\ &= \frac{1}{(1.1)^2}\left(\left(\frac{2}{5}\right)^2 320 + 2\left(\frac{2}{5}\right)\left(\frac{3}{5}\right)20 + \left(\frac{3}{5}\right)^2(0)\right) \\ &= \frac{6080}{121} = \$50.25\ . \end{aligned}$$

(b) Using expressions (12.20) and (12.21) we calculate

$$\alpha_2^u = 1, \alpha_2^d = \frac{4}{15}, \beta_2^u = -\frac{8000}{121}, \beta_2^d = -\frac{2000}{363}, \alpha_1 = \frac{4}{5}$$

and

$$\beta_1 = -\frac{10800}{363} = -29.752\ .$$

Hence:

$$C_0 = \alpha_1 S_0 + \beta_1 R_0 = \frac{4}{5}(100) + (-29.752)(1) = \$50.25\ .$$

(c) We suppose that the value of the call option is priced at \$48.25. Then one can buy the option at \$48.25 and invest \$2 in the bond. At time $N = 2$, he earns the profit of $\$2(1.1)^2 = \$2.42 > 0$, which gives an arbitrage opportunity.

(d) The value of the put option is calculated from (a) using put-call parity:

$$\begin{aligned} C_0 - P_0 &= S_0 - \frac{K}{(1+r)^2} \\ P_0 &= C_0 - S_0 + \frac{K}{(1+r)^2} \\ &= 50.25 - 100 + 80\frac{1}{(1.1)^2} \\ &\approx 16.37. \end{aligned}$$

▲

Note 12.6 *Now we will explain briefly how to use a CRR binomial tree to value American options. The interested reader may refer to Bingham and Kiesel (2004) or Williams (2006).*

An American contingent claim X is represented by a sequence of random variables $X = \{X_n; n = 0, 1, \cdots, N\}$ such that X_n is $\Im_{n-1}$-measurable for

$n = 0, 1, \cdots, N$. For the American call option on the stock S with strike price K, $X_n = \max(S_n - K, 0)$. To define the price of the contingent claim associated with $\{X_n; n = 0, 1, \cdots, N\}$, we proceed with backward induction starting at time N. Let U_m be the minimum amount of money required for the writer of the American contingent claim at time m to cover the payoff at time N if a buyer decides to claim at time $0 \leq m \leq N$. The time m is a random variable and is also known as a stopping time.

A self-financing strategy is called a super hedging for the writer if $\varphi = \{(\alpha_n, \beta_n); n = 1, 2, \cdots, N\}$ with value $V_n(\varphi)$ at time n such that $U_n \leq V_n(\varphi)$ for $n = 0, 1, \cdots, N$. We assume that $U_N = X_N$. The minimum amount of money required at time $N-1$ to cover the payoff of the American contingent claim at time N is:

$$\frac{1}{1+r} E_Q \left(U_N \mid \Im_{N-1} \right).$$

But in the case of the holder exercising his right at time $N-1$, he will earn X_{N-1}. Hence the value of the American contingent claim at time $N-1$ is:

$$U_{N-1} = \max \left\{ X_{N-1}, \frac{1}{1+r} E_Q \left(U_N \mid \Im_{N-1} \right) \right\}.$$

By induction, we can see that the value of the American contingent claim for $n = 1, 2, \cdots, N$ is:

$$U_{n-1} = \max \left\{ X_{n-1}, \frac{1}{1+r} E_Q \left(U_n \mid \Im_{n-1} \right) \right\}. \qquad (12.29)$$

12.3 CONTINUOUS-TIME MODELS

We will discuss the continuous-time modeling of financial derivatives. We consider a finite time interval $[0, T]$ for $0 < T < \infty$. The continuous-time market modeled by a probability space $(\Omega, \Im, P)$ and a filtration $(\Im_t)_{t \leq T} = \{\Im_t; 0 \leq t \leq T\}$ is the standard filtration generated by a Brownian motion $\{B_t; 0 \leq t \leq T\}$. As in the case of discrete-time, we consider a financial market consisting of a risky asset, namely a stock with price process $S = \{S_t; 0 \leq t \leq T\}$ and a risk-free asset with price process $R = \{R_t; 0 \leq t \leq T\}$. As suggested by Black-Scholes (1973), we assume that the behavior of the strike price is determined by the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dB_t \qquad (12.30)$$

where μ and σ are two constants and B_t is a Brownian motion. We also assume that a bond is risk-free, which yields interest at rate r, where the constant r is the continuously compounding rate of return. The price dynamics of the bond takes the form

$$dR_t = rR_t dt. \qquad (12.31)$$

The solution of equations (12.30) and (12.31) are easily computed to obtain

$$S_t = S_0 e^{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t} \tag{12.32}$$

and

$$R_t = e^{rt} \tag{12.33}$$

for $0 \le t \le T$.

Note 12.7 *The stock price process described in equation (12.32) implies that the log-returns*

$$log(S_T) - log(S_t) = \sigma\left(B_T - B_t\right) + \left(\mu - \frac{1}{2}\sigma^2\right)(T-t)$$

are normally distributed with mean $\left(\mu - \frac{1}{2}\sigma^2\right)(T-t)$ *and variance* $\sigma^2(T-t)$. *The parameter* σ *is also called the volatility of the stock. Hence, the stock price* S_t *has a log-normal distribution.*

We now define the trading strategy and value of the portfolio process for the continuous-time models.

Definition 12.14 *A trading strategy is a stochastic process* $\varphi = \{\varphi_t = (\alpha_t, \beta_t) : 0 \le t \le T\}$ *satisfying the following conditions:*

1. $\varphi : [0,T] \times \Omega \to \mathbb{R}^2$ *is* $(\mathcal{B}_T \times \Im_T)$*-measurable where* $\varphi(t,\omega) = \varphi_t(\omega)$ *for each* $0 \le t \le T$ *and* $\omega \in \Omega$.

2. φ_t *is* $\Im_t$*-adapted for each* $0 \le t \le T$.

3. $\int_0^T \alpha_t^2 dt < \infty$ *and* $\int_0^T |\beta_t| dt < \infty$ *a.s.*

As before α_t *and* β_t *represent the number of units of shares and bonds respectively.*

The value process or value of the portfolio at time t is given by:

$$V_t(\varphi) = \alpha_t S_t + \beta_t R_t.$$

In the discrete-time models we have seen self-financing strategies. Similarly the trading strategy φ is called a *self-financing* strategy if

$$V_t(\varphi) = V_0(\varphi) + \int_0^t \alpha_u dS_u + \int_0^t \beta_u dR_u$$

for all $0 \le t \le T$, or equivalently

$$dV_t(\varphi) = \alpha_t dS_t + \beta_t dR_t$$

for all $0 \le t \le T$.

Now we define discounted price processes for the stock and value processes. The *discounted stock price process* is given by

$$\widetilde{S}_t = e^{-rt} S_t \qquad \text{for } 0 \le t \le T$$

and the *discounted value process* is given by

$$\widetilde{V}_t(\varphi) = e^{-rt} V_t(\varphi) \qquad \text{for } 0 \le t \le T\ .$$

By use of Itô's formula:

$$\begin{aligned} d\widetilde{S}_t &= d\left(S_t e^{-rt}\right) \\ &= -rS_t e^{-rt} dt + e^{-rt}\left(\mu S_t dt + \sigma S_t dB_t\right) \\ &= (\mu - r) e^{-rt} S_t dt + \sigma e^{-rt} S_t dB_t. \end{aligned}$$

Thus:

$$d\widetilde{S}_t = (\mu - r)\widetilde{S}_t dt + \sigma \widetilde{S}_t dB_t.$$

In integral form this is

$$\widetilde{S}_t = S_0 + \int_0^t \left((\mu - r) - \frac{1}{2}\sigma^2\right) \widetilde{S}_u du + \int_0^t \sigma \widetilde{S}_u dB_u$$

since $\widetilde{S}_0 = S_0$.

Theorem 12.10 *A trading strategy φ is self-financing if and only if the discounted process $\widetilde{V}_t(\varphi)$ can be expressed for all $t \in [0, T]$ as:*

$$\widetilde{V}_t(\varphi) = V_0(\varphi) + \int_0^t \alpha_u d\widetilde{S}_u\ . \tag{12.34}$$

Proof: We know that

$$dV_t(\varphi) = \alpha_t dS_t + \beta_t dR_t$$

so that:

$$\begin{aligned} d\left(\widetilde{V}_t(\varphi)\right) &= d\left(e^{-rt} V_t(\varphi)\right) \\ &= -re^{-rt} V_t(\varphi) + e^{-rt} dV_t(\varphi) \\ &= -re^{-rt}\left(\alpha_t S_t + \beta_t R_t\right) + e^{-rt}\left(\alpha_t dS_t + \beta_t dR_t\right) \\ &= e^{-rt}(\mu - r)\alpha_t S_t dt + e^{-rt}\sigma\alpha_t S_t dB_t \\ &= \alpha_t d\widetilde{S}_t. \end{aligned}$$

Hence

$$\widetilde{V}_t(\varphi) = V_0(\varphi) + \int_0^t \alpha_u d\widetilde{S}_u$$

for $0 \le t \le T$. Conversely suppose that (12.34) holds. Then

$$\begin{aligned} d\left(\widetilde{V}_t(\varphi)\right) &= \alpha_t d\widetilde{S}_t \\ &= e^{-rt}\alpha_t\left(-rS_t dt + dS_t\right) \\ &= e^{-rt}\left(-rV_t(\varphi)dt + \alpha_t dS_t + \beta_t dR_t\right). \end{aligned}$$

Again by Itô's formula

$$\begin{aligned} d\left(\widetilde{V}_t(\varphi)\right) &= d\left(e^{-rt}V_t(\varphi)\right) \\ &= -re^{-rt}V_t(\varphi)dt + e^{-rt}dV_t(\varphi) \\ &= \alpha_t dS_t + \beta_t dR_t \end{aligned}$$

and hence φ is a self-financing strategy. ∎

Definition 12.15 *A self-financing strategy φ is called an arbitrage opportunity if*

$$V_0(\varphi) = 0, \quad V_T(\varphi) \ge 0 \quad \textit{and} \quad E\left(V_T(\varphi)\right) > 0.$$

We now introduce an equivalent martingale measure.

Definition 12.16 *A probability measure Q is called an equivalent martingale measure or risk-neutral measure if:*

1. *Q is equivalent to measure P.*
2. *The discounted stock price process $\widetilde{S}_t = \{e^{-rt}S_t, 0 \le t \le T\}$ is a martingale under measure Q, i.e., $E_Q\left(\widetilde{S}_t \mid \Im_{t-1}\right) = \widetilde{S}_{t-1}$ for all $0 \le t \le T$.*

We now give the valuation formula for a European contingent claim under a risk-neutral measure.

Theorem 12.11 *Consider a European contingent claim H which is a $\Im_T$-measurable random variable such that $E_Q(|H|) < \infty$. Then the fair price or arbitrage-free price for the European contingent claim H at time t equals:*

$$V_t = E_Q\left(e^{-r(T-t)}H \mid \Im_t\right).$$

Proof: Let us assume that there exists an admissible trading strategy $\varphi = \{(\alpha_t, \beta_t); 0 \le t \le T\}$ that replicates a European contingent claim. Then the value of the replicating portfolio at time t is:

$$V_t = \alpha_t S_t + \beta_t R_t.$$

The discounted value process at time t is:

$$\begin{aligned} \widetilde{V}_t &= e^{-rt}V_t \\ \widetilde{V}_t &= \alpha_t \widetilde{S}_t + \beta_t. \end{aligned}$$

Since no funds are added or removed from the replicating portfolio and the portfolio is self-financing, by Theorem 12.10 we can write the portfolio as:

$$\widetilde{V}_t(\varphi) = V_0(\varphi) + \int_0^t \alpha_u d\widetilde{S}_u.$$

Also H can be replicated by self-financing strategy φ and the value process $\{V_t; t \geq 0\}$.

We have $H = V_T$ and hence:

$$\widetilde{H} = e^{-rT} H = \widetilde{V}_T.$$

Now let $M_t = \int_0^t \alpha_u d\widetilde{S}_u$ be a martingale under measure Q. We have:

$$E_Q\left(\int_t^T \alpha_u d\widetilde{S}_u \mid \Im_t\right) = E_Q\left(M_T - M_t \mid \Im_t\right) = 0 .$$

Hence:

$$E_Q\left(e^{-rt} H \mid \Im_t\right) = \widetilde{V}_t = e^{-rt} V_t.$$

Therefore:

$$V_t = E_Q\left(e^{-r(T-t)} H \mid \Im_t\right).$$

■

12.3.1 Black-Scholes Formula European Call Option

In this section we derive Black-Scholes formula for European call option using two different approaches. In the first approach we apply the results of the risk-neutral measure for a European call option.

Let us consider a European call option on a stock with strike price K. The claim has a payoff function $H = \max(S_T - K, 0)$ at expiry time T. The strike price follows the geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$.

Theorem 12.12 *Suppose that the stock follows the geometric Brownian motion described above with strike price K and expiry date T. Then the value of the European call option at time t is given by*

$$C = S_t \Phi(d_1) - K e^{-r(T-t)} \Phi(d_2) \tag{12.35}$$

where

$$d_1 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r + \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} \tag{12.36}$$

$$d_2 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}} = d_1 - \sigma\sqrt{T-t}, \tag{12.37}$$

where $\Phi(.)$ is the cumulative normal distribution and r is the risk-neutral interest rate.

Proof: We know that S_t satisfies the equation

$$dS_t = \mu S_t dt + \sigma S_t dB_t$$

where B_t is a Brownian motion. Thus S_t is of the form

$$S_T = S_0 e^{\left(r - \frac{1}{2}\sigma^2\right)T + \sigma B_T}$$

where S_0 is the price of the stock at time $t = 0$.

Under a risk-neutral measure the fair price of the call option is:

$$C_t = E_Q(e^{-r(T-t)} \max(S_T - K, 0) \mid \Im_t) \ .$$

Without loss of generality, we assume that $t = 0$. We have:

$$\begin{aligned} C_0 &= E_Q\left(e^{-rT}\max(S_T - K, 0)\right) \\ &= E_Q\left(e^{-rT} S_T \mathcal{X}_{\{S_T > K\}}\right) - Ke^{-rT} E_Q\left(\mathcal{X}_{\{S_T > K\}}\right) \\ &= (I) - (II). \end{aligned}$$

First we evaluate the term (II). We have:

$$\begin{aligned} E_Q\left(\mathcal{X}_{\{S_T > K\}}\right) &= Q(S_T > K) \\ &= Q(\ln S_T > \ln K) \\ &= Q\left(B_t > \frac{\log K - \log S_0 - \left(r + \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right). \end{aligned}$$

But we know that, under measure Q, $B_t \stackrel{d}{=} Z = \mathcal{N}(0, t)$ and $Q(Z > x) = Q(Z < -x)$. Hence

$$\begin{aligned} Q(S_T > K) &= Q\left(Z < \frac{\log\left(\frac{S_0}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)T}{\sigma\sqrt{T}}\right) \\ &= \Phi(d_2) \end{aligned}$$

where

$$d_2 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)(T - t)}{\sigma\sqrt{T - t}}.$$

Now we compute the term (I):

$$\begin{aligned}
E_Q\left(e^{-rT}S_T\mathcal{X}_{\{S_T>K\}}\right) &= E_Q\left(S_0e^{-\frac{1}{2}\sigma^2T+\sigma B_T}\mathcal{X}_{\{S_T>K\}}\right) \\
&= \frac{1}{\sqrt{2\pi T}}\int_{-d_2}^{\infty}\left(S_0e^{\sigma\sqrt{T}y-\frac{1}{2}\sigma^2T}\right)e^{-\frac{y^2}{2}}dy \\
&= S_0\frac{1}{\sqrt{2\pi T}}\int_{-d_2}^{\infty}e^{-\frac{1}{2}(y-\sigma\sqrt{T})^2}dy \\
&= S_0\frac{1}{\sqrt{2\pi T}}\int_{-d_2-\sigma\sqrt{T}}^{\infty}e^{-\frac{x^2}{2}} \qquad \text{(since } x = y-\sigma\sqrt{T}\text{)} \\
&= \Phi\left(d_2+\sigma\sqrt{T}\right) \\
&= \Phi(d_1)
\end{aligned}$$

where:

$$d_1 = d_2 + \sigma\sqrt{T}.$$

Thus for any $0 \le t < T$ we have

$$C_t = S_t\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where d_1 and d_2 are given in equations (12.36) and (12.37). ∎

We will prove the previous theorem by the partial differential equation approach proposed by Black-Scholes (1973). We consider the stock price process that follows geometric Brownian motion

$$dS_t = \mu S_t dt + \sigma S_t dB_t \tag{12.38}$$

where $\mu \in \mathbb{R}$ and $\sigma > 0$ and B_t is a Brownian motion. Consider a portfolio at time t which consists of α_t shares on stock and B_t units of bonds with value R_t. The bond is assumed to be risk-free with deterministic equation

$$dR_t = rR_t dt \tag{12.39}$$

where $r > 0$ is the interest rate. At time t, the value of the portfolio is:

$$V_t = \alpha_t S_t + \beta_t R_t. \tag{12.40}$$

The portfolio is supposed to hedge a European call option with value $C = C_t = C(S_t, t)$. The European call option with strike price K and expiry date T has pay-off value $C_t = \max(S_T - K, 0)$. From Itô's formula for the call option C, we have:

$$\begin{aligned}
dC &= \frac{\partial C}{\partial t}dt + \frac{\partial C}{\partial S}dS + \frac{\partial^2 C}{\partial S^2}(dS)^2 \\
&= \left(\frac{\partial C}{\partial t} + \mu S\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2S^2\frac{\partial^2 C}{\partial S^2}\right)dt + \sigma S\frac{\partial C}{\partial S}dB_t\,. \qquad (12.41)
\end{aligned}$$

The value of the portfolio satisfies:

$$\begin{aligned} dV_t &= \alpha_t dS_t + \beta_t dR_t \\ &= \alpha_t(\mu S_t dt + \sigma S_t dB_t) + B_t r R_t dt \\ &= (\alpha\mu S_t + \beta r R_t)dt + \alpha_t \sigma S_t dB_t. \end{aligned} \tag{12.42}$$

Under a self-financing strategy, equation (12.42) must coincide with equation (12.41). Since terms dt and dB_t are independent, the respective coefficients must be equal. Otherwise there shall be an arbitrage opportunity. Therefore:

$$\alpha_t S_t + \beta_t R_t = C_t \tag{12.43}$$

$$\alpha\mu S + \beta r R = \frac{\partial C}{\partial t} + \mu S \frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} \tag{12.44}$$

$$\alpha\sigma S = \sigma S \frac{\partial C}{\partial S}. \tag{12.45}$$

From (12.45), we have:

$$\alpha = \frac{\partial C}{\partial S}. \tag{12.46}$$

Substituting (12.46) in (12.43), we have:

$$\beta = \frac{1}{R}\left(C - S\frac{\partial C}{\partial S}\right). \tag{12.47}$$

Again substituting equations (12.46) and (12.47) in (12.44), we get:

$$\frac{\partial C}{\partial t} + rS\frac{\partial C}{\partial S} + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 C}{\partial S^2} = rC. \tag{12.48}$$

Thus we have shown that the value of a call option satisfies the above partial differential equation. This equation is also known as the *Black-Scholes equation*, which can be solved under appropriate conditions. Equation (12.48) can be solved for the final condition

$$C_T = C(S_T, T) = \max(S_T - K, 0)$$

along with the boundary conditions

$$C(0,t) = 0 \text{ and } C(S,t) \to S \text{ as } S \to \infty.$$

Finding the solution of equation (12.48) is a long computational procedure and the reader may refer to Wilmott et al. (1995). The solution of (12.48) for the value of a European call option at time T with strike price K, underlying stock price S, risk-free interest rate r and volatility σ starting from time t is given as

$$C(S_t, t) = S\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2)$$

where:

$$\begin{aligned} d_1 &= \frac{\log(S/K) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \\ d_2 &= \frac{\log(S/K) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} \, . \end{aligned}$$

■ **EXAMPLE 12.5**

Consider a European call option with strike price \$105 and three months to expiry. The stock price is \$110 and risk-free interest rate is 8% per year, and the volatility is 25% per year.

Thus we have $S_0 = 110$, $K = 105$, $T = 0.25$, $r = 0.08$ and $\sigma = 0.25$. From equations (12.36) and (12.37) we get:

$$d_1 = \frac{\log\left(\frac{60}{65}\right) + \left(0.08 + \frac{1}{2}(0.3)^2\right)(0.25)}{0.3\sqrt{0.25}} = 0.5947$$

$$d_2 = d_1 - 0.3\sqrt{0.25} = 0.4697$$

The value of the European call option calculated by using the Black-Scholes formula (12.35) is:

$$\begin{aligned} C_0 &= S\Phi(d_1) - Ke^{-rT}\Phi(d_2) \\ &= 110\Phi(0.5947) - 105e^{-0.08(0.25)}\Phi(0.4697) \\ &= 9.5778. \quad \blacktriangle \end{aligned}$$

12.3.2 Properties of Black-Scholes Formula

The option pricing formula depends on five parameters, namely S, K, T, r and σ. It is important to analyze the change or variations of option price with respect to these parameters. These variations are known as Greek-letter measures or simply *Greeks*. We now give a brief description of these measures.

Delta

The *delta* of a European call option is the rate of change of its value with respect to the underlying asset price:

$$\Delta = \frac{\partial C}{\partial S} = \Phi(d_1) \, .$$

Since $0 < \Phi(d_1) < 1$, it follows that $\Delta > 0$, and hence the value of a European call option is always increasing as the underlying asset price increases.

Using the information from Example 12.5, $\Delta = 0.724$, which means that the call price will increase (decrease) with about \$0.724 if the underlying asset increases (decreases) by one dollar. The delta of the put option is also given by the option's first derivative with respect to the underlying asset price. The delta of the put option is given by:

$$\Delta_P = \Phi(d_1) - 1 < 0.$$

Gamma: The Convexity Factor

The *gamma* (Γ) of a derivative is the sensitivity of Δ with respect to S:

$$\Gamma = \frac{\partial^2 C}{\partial S^2} .$$

The concept of gamma is important when the hedged portfolio cannot be adjusted continuously in time according to $\Delta(S(t))$. If gamma is small, then delta changes only slowly and adjustments in the hedge ratio need only be made infrequently. However, if gamma is large, then the hedge ratio delta is highly sensitive to changes in the price of the underlying security. According to the Black-Scholes formula, we have:

$$\Gamma = \frac{1}{S\sqrt{2\pi}\sigma\sqrt{T-t}} \exp(-d_1^2/2) .$$

Notice that $\Gamma > 0$ so that the Black-Scholes formula is always concave up with respect to S. Note the relation of a delta hedged portfolio to the option price due to concavity. The call and put options have the same gamma. For Example 12.5, the gamma for the European call option is $\Gamma = 0.0243$, which means that the rise (fall) in the underlying asset price by one dollar yields a change in the delta from 0.724 to 0.724+0.0243 = 0.7476 (0.6990).

Theta: The Time Decay Factor

The *theta* (Θ) of a European claim with value function $C(S_t, t)$ is defined as:

$$\Theta = \frac{\partial C}{\partial t}.$$

By defining the rate of change with respect to the real time, the theta of a claim is sometimes referred to as the time decay of the claim. For a European call option on a non-dividend-paying stock:

$$\Theta = -\frac{S\exp(-d_1^2/2)\sigma}{2\sqrt{2\pi}\sqrt{T-t}} + rK\exp(-r(T-t))\Phi(d_2).$$

Note that Θ for a European call option is negative, so that the value of a European call option is a decreasing function of time. Theta does not act like a hedging parameter as is the case for delta and gamma. This is because although there is some uncertainty about the future stock price, there is no uncertainty about the passage of time. It does not make sense to hedge against the passage of time on an option.

Rho: The Interest Rate Factor

The *rho* (ρ) of a financial derivative is the rate of change of the value of the financial derivative with respect to the interest rate. It measures the sensitivity of the value of the financial derivative to interest rates. For a European call option on a non–dividend paying stock,

$$\rho = K(T-t)\exp(-r(T-t))\Phi(d_2) \ .$$

We see that ρ is always positive. An increase in the risk-free interest rate means a corresponding increase in the derivative.

Vega: The Volatility Factor

The *Vega* (ν) of a financial derivative is the rate of change of value of the derivative with respect to the volatility of the underlying asset. Here we wish to note that vega is not the name of any Greek letters, while the names of other options sensitivities have corresponding Greek letters. For a European call option on a non-dividend-paying stock,

$$\nu = S\sqrt{T-t}\frac{1}{\sqrt{2\pi}}\exp(-d_1^2/2)$$

so that the vega is always positive and is identical for call and put options. An increase in the volatility will lead to an increase in the call option value. See Hull (2009) for more details.

Note 12.8 *We wish to note that the Black-Scholes partial differential equation can also be written as:*

$$\Theta + rS\Delta + \frac{1}{2}\sigma^2 S^2\Gamma = rC.$$

12.4 VOLATILITY

We have seen that the Black-Scholes formula does not depend on the drift parameter μ of the stock and depends only on the parameter σ, the volatility that appears in the formula. In order to use the Black-Scholes formula, one must know the value of the parameter σ of the underlying stock. We end this chapter by giving a brief description of how to estimate the volatility of stock. The volatility is a crucial component for pricing finance derivatives and estimating the value σ is the subject of study in financial statistics. One method to estimate volatility is the *historical volatility* whose estimates require the use of appropriate statistical estimators, usually an estimator of variance. One of the main problems in this regard is to select the sample size, or the number of observations that will be used to estimate σ. Different observations tend to give different volatility estimates.

To estimate the volatility of a stock price empirically, the stock price is observed at regular intervals, such as every day, every week or every month. Define:

1. The number of observations $n+1$
2. S_i, $i = 0, 1, 2, 3, \cdots, n$, the stock price at the end of the ith interval
3. τ, the length of each time interval (in years).

Let

$$u_i = \ln(S_i) - \ln(S_{i-1}) = \ln\left(\frac{S_i}{S_{i-1}}\right)$$

for $i = 1, 2, 3, \cdots$ be the increment of the logarithms of the stock prices. We are assuming that the stock price acts as a geometric Brownian motion, so that $\ln(S_i) - \ln(S_{i-1}) \sim N(r\tau, \sigma^2\tau)$.

Since $S_i = S_{i-1}e^{u_i}$, u_i is the continuously compounded return (not annualized) in the ith interval. Then the usual estimate s of the standard deviation of the u_i's is given by

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2}$$

where $\bar{u}$ is the mean of the u_i's. Sometimes it is more convenient to use the equivalent formula

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}u_i^2 - \frac{1}{n(n-1)}\left(\sum_{i=1}^{n}u_i\right)^2}.$$

As usual, we assume the stock price varies as a geometric Brownian motion. That means that the logarithm of the stock price is a Brownian motion with the same drift and in the period of time τ would have a variance $\sigma^2\tau$. Therefore, s is an estimate of $\sigma\sqrt{t}$. It follows that σ can be estimated as:

$$\sigma \approx \frac{s}{\sqrt{\tau}}.$$

Choosing an appropriate value for n is not easy because σ does change over time and data that are too old may not be relevant for the present or the future. A compromise that seems to work reasonably well is to use closing prices from daily data over the most recent 90 to 180 days. Empirical research indicates that only trading days should be used, so days when the exchange is closed should be ignored for the purposes of the volatility calculation.

Another method is known as *implied volatility*, which is the numerical value of the volatility parameter that makes the market price of an option equal to the value from the Black-Scholes formula. We have known that it is not possible to "invert" the Black-Scholes formula to explicitly express σ as a function of the other parameters. Therefore, one can use numerical techniques such as the bisection method or Newton-Raphson method to solve for σ. The efficient method is to use the Newton-Raphson method, which is an iterative method to solve the equation

$$f(\sigma, S, K, r, T-t) - C = 0,$$

where C is the market price of an option and $f(.)$ is a pricing model that depends on σ. From an initial guess of σ_0, the iteration function is:

$$\sigma_{i+1} = \sigma_i - f(\sigma_i)/(df(\sigma_i)/d\sigma).$$

This means that one has to differentiate the Black-Scholes formula with respect to σ. This derivative is known as *vega*. A formula for nu for a European call option is

$$\frac{df}{d\sigma} = S\sqrt{T-t}\Phi'(d_1)\exp(-r(T-t)).$$

With the help of software such as MATlab or Matematica one can solve the above equation. Implied volatility is a "forward-looking" estimation technique, in contrast to the "backward-looking" historical volatility. That is, it incorporates the market's expectations about the prices of securities and their derivatives. In general volatility need not be a constant and could depend on the price of the underlying stock. These models are known as stochastic volatility models. As we stated in the introduction, we presented only a simple modeling approach for mathematical finance. Interested readers are advised to take a proper course on stochastic analysis which will take them to a more formal theory of mathematical finance. (see Williams (2006), Bingham and Kiesel (2004), and Lamberton and Lapeyre (1996).

EXERCISES

12.1 If a stock sells for \$100, the value of the call option is \$6, the value of the put option is \$4 and both options have the same strike price, \$100, what is the risk-free interest rate.

12.2 Suppose that an investor buys a European call option on certain underlying stock ABC with a \$75 strike price and sells a European put option on ABC with the same strike price. Both options will expire in 3 months. Describe the investor position.

12.3 Suppose that a stock price is \$120 and in the next year it will either be up by 10% or fall by 20%. The risk-free interest rate is 6% per year.

A European call option on this stock has a strike price of \$130. Find the probability that the stock price will rise and also the value of a call option.

12.4 Consider a European call option with 6 months to expiration. The underlying stock price is \$100, strike price is \$100, the risk-free interest rate is 6% and volatility is 30% . Find the risk-neutral probability and the value of the European call option using the CRR binomial model for $N = 4$.

12.5 Consider an American put option with 6 months to expiration. The underlying stock price is \$100, the strike price is \$95, the risk-free interest rate is 8% and the volatility is 30%. Find the value of the American put option at $n = 0$ using the CRR binomial model for $N = 5$.

12.6 Consider a one-period binomial model with $S_0 = \$40$, $S_1 = \$44$ or $S_1 = \$36$, $K = 42$. The risk-free interest rate $r = 12\%$. Find the risk-neutral probability and calculate the value of the three-month European call option

12.7 (Williams, 2006) Consider the CRR binomial model for $N = 3$ with $S_0 = \$50$, $u = 2$, $d = \frac{1}{2}$ and risk-free interest rate $r = 7\%$. The value of the European contingent claim at time $N = 3$ is given by:

$$K = \max(S_0, S_1, S_2, S_3).$$

a) Find the arbitrage-free initial price of the European call option at time $n = 0$.

b) Determine the hedging strategy for this option.

c) Suppose that the option in (12.7) is initially priced \$2 below the arbitrage-free price. Describe a strategy that gives an arbitrage.

12.8 As one increases the strike price K (keeping all other parameters fixed), does the value of a call option increase? Briefly justify your answer.

12.9 Find the expression for the delta of the put option.

12.10 Assume single-period binomial model with $S_0 = \$10$, $S_u = 12$, $S_d = 9$ and $r = 0.5$. A call option on this stock has an expiry date 5 months from today and a strike price of \$10. Find the value of this option.

12.11 Consider a two-period CRR model with $N = 2$, $S_0 = \$40$ and strike price $K = \$42$. In each of the next 3-month periods, the stock is expected to go up by 10% or down by 10%. The risk-free interest rate is 10% per year.

a) What is the value of a 6-month European put option?

b) What is the value of a 6-month American put option?

12.12 Consider a European call option with expiry date $t = n$ and strike price K. Let C_0 denote its price at time $t = 0$. Argue that $C_0 \leq S_0$; the price of any such option is always less than or equal to the price of the stock at time $t = 0$.

12.13 Use the Black-Scholes formula to price a European call option for a stock whose price today is \$20 with expiry date 6 months from now, strike price \$22 and volatility 20%. The risk-free interest is 5% per year. Find the value of the call option halfway to expiry if the stock price at that time is \$21.

12.14 Use the Black-Scholes formula to price a European call option for a stock whose price today is \$75 with expiry date 3 months from now, strike price \$70 and volatility 20%. The risk-free interest is 7% per year.

a) Find the value of the European call option and compute delta and gamma for this option.

b) Find the value of the European put option and compute delta and gamma for this option.

12.15 Let r denote the risk-free (annual nominal) interest rate. Suppose that we consider the stock under its risk-neutral measure, that is, when μ is replaced by $\mu^* = r - \sigma^2/2$. Show that the expected rate of return now is r, the same as the risk-free interest rate. Explain why this makes sense.

12.16 Suppose that the observations on a stock price (in dollars) at the end of each of 15 consecutive weeks are as follows: 100.50, 102, 101.25, 101, 100.25, 100.75, 100.65, 105, 104.75, 102, 103.5, 102.5, 103.25, 104.5, 104.25. Estimate the stock price volatility.

12.17 Suppose that a call option on an underlying stock $S_0 = 36.12$ that pays no dividends for 6 months has a strike price of \$35 with market price of \$2.15 and expiry date of 7 weeks. The risk-free interest rate $r = 7\%$. Find the implied volatility of this stock?

12.18 A call option on a non-dividend paying stock has a market price of \$2.50. The stock price is \$15, the exercise price is \$13, the time to maturity is 3 months and the risk-free interest rate is 5% per annum. What is the implied volatility?

APPENDIX A

BASIC CONCEPTS ON SET THEORY

In this appendix we present some basic concepts and results from set theory that have been used throughout the book. Intuitively, a set is a well-defined grouping of objects. The objects belonging to the set are called elements. Sets are represented with uppercase Latin letters: A, B, C, M, X, $\cdots$ while the elements of a set are usually represented with lowercase Latin letters: a, b, c, m, x, $\cdots$. To indicate that an element x belongs to the set A, we write $x \in A$, and if x is not an element of the set A, then we write $x \notin A$.

Sets can be described by enumerating all of their elements or by enunciating properties that those elements must have. In the first case we say that the set is determined by extension and in the second case we say the set is determined by comprehension. Thus, we have, for example, that the set

$$A = \{1, 3, 5, 9\}$$

is described by extension, while the set

$$B = \{x : \ x \text{ is a rational number less than or equal to } 5\}$$

has been defined by comprehension.

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

A set whose elements correspond to nonnegative integers is called finite. A set that does not have any elements is called an empty set and is notated $\emptyset$. A set is said to be infinite if it is not finite.

Two sets are said to be equal if and only if they have exactly the same elements. If all the elements of a set A are also elements of a set B we say that A is contained in B (alternatively, B contains A or A is a subset of B) and write $A \subseteq B$ (or $B \supseteq A$). From this definition, it is clear that every set is a subset of itself and $\varnothing$ is a subset of every set.

If A is a subset of B and there is at least an element in B that does not belong to A, we say that A is a proper subset of B and write $A \subset B$.

Applications of set theory usually consider all sets to be subsets of one single set called the universe, represented as $\mathcal{U}$.

The union of two sets A and B, written $A \cup B$, is the set of all elements belonging to either A or B or both. That is:

$$A \cup B := \{x : \ x \in A \ \vee \ x \in B\} .$$

In a similar way, if $\{A_i\}_{i \in I}$ is a family of sets and $J \subseteq I$, then:

$$\bigcup_{i \in J} A_i := \{x : \ x \in A_i \ \text{ for at least one } \ i \in J\} .$$

The intersection of two sets A and B, represented as $A \cap B$, is the set of elements belonging to both A and B. That is:

$$A \cap B := \{x : \ x \in A \ \wedge \ x \in B\} .$$

In a similar way, if $\{A_i\}_{i \in I}$ is a family of sets and $J \subseteq I$, then:

$$\bigcap_{i \in J} A_i := \{x : \ x \in A_i \ \text{ for all } \ i \in J\} .$$

If A and B do not have elements in common, that is, if $A \cap B = \varnothing$, it is said that they are mutually exclusive or disjoint.

The difference of two sets A and B, notated as $A - B$, is the set of all elements belonging to A but not to B; in other words:

$$A - B := \{x : \ x \in A \ \wedge \ x \notin B\} .$$

The complement of a set A, notated as A^c, is the difference between the universe $\mathcal{U}$ and A; in other words:

$$A^c := \{x : \ x \in \mathcal{U} \ \wedge \ x \notin A\} .$$

Next, we present, without demonstration, the basic properties of the set operations defined above. The interested reader can find the proofs in Muñoz (2002).

Theorem A.1 *Let A, B and C be subsets of a universe $\mathcal{U}$. Then:*

1. *(Commutative Laws)*
   (a) $A \cup B = B \cup A$
   (b) $A \cap B = B \cap A$
2. *(Associative Laws)*
   (a) $A \cup (B \cup C) = (A \cup B) \cup C$
   (b) $A \cap (B \cap C) = (A \cap B) \cap C$
3. *(Distributive Laws)*
   (a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   (b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
4. *(Complement Laws)*
   (a) $A \cup A^c = \mathcal{U}$
   (b) $A \cap A^c = \varnothing$
   (c) $A \cup \mathcal{U} = \mathcal{U}$
   (d) $A \cap \mathcal{U} = A$
   (e) $A \cup \varnothing = A$
   (f) $A \cap \varnothing = \varnothing$
5. *(Difference Laws)*
   (a) $A - B = A \cap B^c$
   (b) $A - B = A - (A \cap B) = (A \cup B) - B$
   (c) $A - (B - C) = (A - B) \cup (A \cap C)$
   (d) $(A \cup B) - C = (A - C) \cup (B - C)$
   (e) $A - (B \cup C) = (A - B) \cap (A - C)$
   (f) $(A \cap B) \cup (A - B) = A$
   (g) $(A \cap B) \cap (A - B) = \varnothing$
6. *(De Morgan Laws)*
   (a) $(A \cap B)^c = (A^c \cup B^c)$
   (b) $(A \cup B)^c = (A^c \cap B^c)$
7. *(Involutive Law)* $(A^c)^c = A$
8. *(Idempotence Laws)*

(a) $A \cup A = A$

(b) $A \cap A = A$

Next, we present one of the most important concepts in mathematics: If for each element of a set A we establish a well-defined association with a unique element of a set B then, that association is a function from A to B. Notating such association with f, we write:

$$f : A \to B .$$

The set A is called the domain of the function and the set B is called the codomain of f. If $a \in A$, then the element of B that f associates with a is notated as $f(a)$ and is called the image of a.

If f is a function from A to B and $b \in B$, we define the preimage (or inverse image) of b as the set of all elements belonging to A that have b as their image; in other words:

$$f^{-1}(b) := \{a \in A : \ f(a) = b\} .$$

Furthermore, if C is a subset of B, then the set of all elements in A whose images under f are in C is called the preimage (or inverse image) of C under f and is notated as $f^{-1}(C)$. That is:

$$f^{-1}(C) := \{a \in A : \ f(a) \in C\} .$$

Analogously, if D is a subset of A, then the subset of B whose elements are images of elements belonging to D under f is called the image (or direct image) of D under f and is commonly notated as $f(D)$. That is:

$$\begin{aligned} f(D) &:= \{b \in B : \ b = f(a) \ \text{for a} \ a \in D\} \\ &= \{f(a) : \ a \in D\} . \end{aligned}$$

Some of the most important properties of the inverse and direct images of sets under a given function are summarized in the next theorem, whose proof can be checked in Muñoz (2002).

Theorem A.2 *If $f : A \to B$ is a function and N_1, N_2 and N are subsets of B and M is a subset of A, then:*

1. $f^{-1}(N_1 \cup N_2) = f^{-1}(N_1) \cup f^{-1}(N_2)$.
2. $N_1 \subseteq N_2 \Longrightarrow f^{-1}(N_1) \subseteq f^{-1}(N_2)$.
3. $f^{-1}(N_1 \cap N_2) = f^{-1}(N_1) \cap f^{-1}(N_2)$.
4. $f(f^{-1}(N)) = N$.
5. $M \subseteq f^{-1}(f(M))$.

A set A is said to be countable if and only if there is a function f from A to the set of natural numbers $\mathbb{N}$ satisfying:

1. f is one to one or injective, that is, for any $x, y \in A$ with $x \neq y$, $f(x) \neq f(y)$.
2. f is onto or surjective, that is, $f(A) = \mathbb{N}$.

APPENDIX B

INTRODUCTION TO COMBINATORICS

In Chapter 1 we saw that probability in Laplace spaces reduces itself to counting the number of elements of a finite set. The mathematical theory dealing with counting problems of that sort is formally known as combinatorial analysis. All the counting techniques are based on the following fundamental principle.

Theorem B.1 (The Basic Principle of Counting) *Suppose that two experiments are carried out. The first one has m possible results and for each result of this experiment there are n possible results of the second experiment. Then, the total number of possible results of the two experiments, when carried out in the indicated order, is mn.*

Proof: The basic principle of counting can be proved by enumerating all possible results in the following way:

$$\begin{array}{cccc}
(1,1) & (1,2) & \cdots & (1,n) \\
(2,1) & (2,2) & \cdots & (2,n) \\
\vdots & \vdots & \ddots & \vdots \\
(m,1) & (m,2) & \cdots & (m,n)
\end{array}$$

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

This array consists of m rows and n columns and therefore has mn entries. ■

■ EXAMPLE B.1

There are 10 professors in the statistics department of a University, each with 15 graduate students under their tutelage. If one professor and one of his students are going to be chosen to represent the department at an academic event, how many different ways can the selection be made?

In this case there are 10 possible ways to pick the professor and, once chosen, 15 ways of choosing the student. By the basic principle of counting there are 150 ways of selecting the pair that will represent the department. ▲

The basic principle of counting can be generalized as follows:

Theorem B.2 (Generalization of the Basic Principle of Counting) *If r experiments are carried out in such a way that the first one has n_1 possible outcomes, for each of those n_1 results there are n_2 possible results for the second experiment, for each outcome of the second experiment there are n_3 possible results of the third experiment, and so on, then the total number of results when the r experiments are carried put in the indicated order is $n_1 \times n_2 \times \cdots \times n_r$.*

Proof: Left as an exercise for the reader. ■

■ EXAMPLE B.2

The total number of automobile plate numbers that can be made if each plate number consists of three different letters and five numbers equals $26 \times 25 \times 24 \times 10^5 = 1.56 \times 10^9$. ▲

■ EXAMPLE B.3

If a fair dice is rolled four consecutive times, then the total number of possible results of this experiment is $6^4 = 1296$. ▲

■ EXAMPLE B.4

Suppose that there are n distinguishable balls and r distinguishable urns. Then the number of ways in which the balls can be put inside the urns equals r^n. Indeed, the first ball can be placed in any of the r urns, the second ball can be placed in any of the r urns, and so on; therefore, there are

$$\underbrace{r \times r \times \cdots \times r}_{n \text{ times}} = r^n$$

ways to put the balls inside the urns. ▲

■ EXAMPLE B.5 Permutations

A permutation is an arrangement of objects from a given set in a particular order. For example, the permutations of the letters a, b, and c are abc, acb, bac, bca, cab, and cba. That is, there are six permutations of the three letters.

The total number of permutations of the objects belonging to a given set can be found without having to explicitly write down each possible permutation by reasoning as follows: Suppose that a set containing n elements is given, then the first position can be filled with any of the n numbers, the second position with any of the remaining $n-1$ elements, the third position with any of the $n-2$ remaining elements, and so on. Therefore, the total number $P(n,n)$ of permutations of the n elements is:

$$P(n,n) = n(n-1)(n-2)\cdots 1.$$

Since the product of a positive integer n with all the positive integers preceding is notated as $n!$ and called n factorial, we can write:

$$P(n,n) = n!$$

Remember that, by definition, $0! := 1$.

A permutation of n objects taken $r \leq n$ at a time is defined as an arrangement of r of the n objects in a particular order. Thus, the permutations of the letters a, b, c, and d taken two at a time are ab, ba, ac, ca, ad, da, bc, cb, bd, db, cd, and dc. That is, there are 12 permutations of the letters a, b, c and d taken two at a time. The total number $P(n,r)$ of permutations of n objects taken r at a time can be found as follows: There are n objects that can be chosen for the first position, there are $n-1$ objects that can be chosen for the second position, and so on, until we reach the rth position, for which there will be $n-r+1$ possible objects to choose from. In other words:

$$\begin{aligned} P(n,r) &= n(n-1)\cdots(n-r+1) \\ &= \frac{n!}{(n-r)!}; \quad r \leq n \end{aligned}$$ ▲

■ **EXAMPLE B.6**

Suppose that four girls and three boys have to be seated in a row. If the boys and girls can be seated in any order, then we would have $7! = 5040$ ways to do so. If we wish that the boys and girls are alternated in the row, then there would be

$$4 \times 3 \times 3 \times 2 \times 2 \times 1 \times 1 = 144$$

ways to seat them. If we wish that the boys and the girls are seated together (boys with boys, girls with girls) then, there would be $2 \times 4! \times 3! = 288$ ways to seat them. ▲

■ **EXAMPLE B.7**

A student wishes to put 4 calculus, 2 physics, 5 probability, and 3 algebra books on a shelf in such a way that all books belonging to the same subject are grouped together. There are $4! \times 2! \times 5! \times 3!$ ways to place the books if the first ones are calculus books, then the physics books followed by the probability books and the algebra books. Since there are $4!$ ways to organize the subjects, there is a total of

$$4! \times 4! \times 2! \times 5! \times 3! = 8.2944 \times 10^5$$

ways to put the books on the shelf. ▲

■ **EXAMPLE B.8**

We wish to calculate the number of ways in which 3 Mexican, 4 Egyptian, 3 English and 5 Chinese people can be seated at a round table if individuals of the same nationality insist on sitting together. In this case we have four groups of people: Mexicans, Egyptians, Englishmen, and Chinese. The number of ways to place this groups around the table is $3!$. The Mexicans can be seated in $3!$ ways, the Egyptians in $4!$ ways, the English in $3!$ ways and the Chinese in $5!$ ways. Therefore, there are

$$3! \times 3! \times 4! \times 3! \times 5! = 6.2208 \times 10^5$$

ways to seat the people around the table. ▲

■ **EXAMPLE B.9**

Let N be the number of different permutations of the letters in the word "*experiment*". If all the letters were different, then the total number of permutations would be 10!, but since the three "e" can be permuted between them in 3! ways, then $3!N = 10!$. That is:

$$N = \frac{10!}{3!} \ .$$

In general we have that the total number N of different permutations of n objects, of which $n_1, n_2, \cdots, n_r$ are equal between them, is:

$$N = \frac{n!}{n_1!n_2!\cdots n_r!} \ . \qquad \blacktriangle$$

■ **EXAMPLE B.10 Combinations**

Suppose that there are n different objects. Each possible election of $r \leq n$ of the objects is called a combination of order r. In other words, a combination of order r from a set S with n elements is a subset of S having exactly r elements. For example, the combinations of order 2 of the letters a, b, c and d are $\{a, b\}$, $\{a, c\}$, $\{a, d\}$, $\{b, c\}$, $\{b, d\}$ and $\{c, d\}$; that is, there are six possible combinations of order 2 of the letters a, b, c and d. To determine the number of combinations $C(n, r)$ of order r taken from n objects, we observe that if we took the elements in order there would be $P(n, r)$ ways of choosing the r objects, and since r objects can be permuted in $r!$ ways, then:

$$\begin{aligned} r!C(n,r) &= P(n,r) \\ C(n,r) &= \frac{n!}{(n-r)!r!} \ . \end{aligned}$$

The number $C(n, r)$ is called "n choose r" and is written as $\binom{n}{r}$. It is further defined that:

$$\binom{n}{r} := 0 \ \text{ if } \ r < 0 \ \text{ or } \ r > n. \qquad \blacktriangle$$

■ **EXAMPLE B.11**

Five couples, each couple a man and a woman, must be chosen for a dance from a group consisting of 10 women and 12 men. We wish to determine the number of possible selections.

We have that there are $\binom{12}{5}$ ways to pick the men, and after they are chosen we select the women, which can be done in $\binom{10}{5}$ different ways. Therefore, there are

$$\binom{12}{5}\binom{10}{5} = 199,584$$

ways to pick the five couples. ▲

■ EXAMPLE B.12

A committee of 3 people must be formed by selecting its members from a group of 5 men and 3 women. How many possible committees are there? How many if the committee must have at least a woman? How many if two men do not get along and therefore cannot both be part of the committee? How many if there is a man-woman couple that would only accept to join the committee if they both belong to it?

First, we have that there are $\binom{8}{3} = 56$ different ways of selecting the committee members.

Now, if the committee must have at least a woman, then there are

$$\binom{3}{1}\binom{5}{2} + \binom{3}{2}\binom{5}{1} + \binom{3}{3}\binom{5}{0} = 46$$

ways of selecting the committee.

In the case where two men do not get along, there are two options: either one of them is included or both are excluded, which means that there are

$$\binom{2}{1}\binom{6}{2} + \binom{6}{3} = 50$$

ways of selecting the committee.

Finally, in the last case we have two options, either both members of the couple are included or both are excluded, yielding

$$\binom{6}{1} + \binom{6}{3} = 26$$

ways of selecting the committee. ▲

■ EXAMPLE B.13

A set of n elements is to be partitioned in m different groups of sizes $r_1, r_2, \cdots, r_m$, where:

$$r_1 + r_2 + \cdots + r_m = n.$$

We wish to find the number of different possible ways to accomplish this task. We observe that there are $\binom{n}{r_1}$ ways of selecting the first group, for every possible choice of the first group there are $\binom{n-r_1}{r_2}$ ways of selecting the second group, and so on. Therefore, there are

$$\binom{n}{r_1}\binom{n-r_1}{r_2}\cdots\binom{n-r_1-r_2-\cdots r_{m-1}}{r_m} = \frac{n!}{r_1!\times r_2!\times\cdots\times r_m!}$$

ways of selecting the groups. This number is called the multinomial coefficient and is notated as:

$$\binom{n}{r_1, r_2, \cdots, r_m}. \quad \blacktriangle$$

■ EXAMPLE B.14

Thirty students in a third-grade class must be split in 5 groups, each having 12, 5, 3, 6 and 4 students, respectively. How many possible partitions are there? According to the previous example, we have that there are

$$\binom{30}{12, 5, 3, 6, 4} = \frac{30!}{12!\times 5!\times 3!\times 6!\times 4!} = 4.4509\times 10^{16}$$

possible ways to split the class. ▲

■ EXAMPLE B.15

Suppose that there are n indistinguishable balls. How many ways are there to distribute the n balls in r urns?

The result of this experiment can be described by a vector $(x_1, \cdots, x_r)$ where x_i represents the number of balls placed on the ith urn. This allows us to reduce the problem to finding all the vectors $(x_1, \cdots, x_r)$ with nonnegative integer entries such that $x_1 + \cdots + x_r = n$.

To solve this problem, suppose that the objects are put in a horizontal line and then divided in r groups by proceeding as follows: from the $n-1$ spaces separating the objects we choose $r-1$ and trace dividing lines there; by doing this we get r nonempty groups. That is, there are $\binom{n-1}{r-1}$ vectors $(x_1, \cdots, x_r)$ with positive components satisfying $x_1+\cdots+x_r = n$.

Since the number of nonnegative solutions of $x_1 + \cdots + x_r = n$ equals the number of positive solutions of $y_1 + \cdots + y_r = n + r$ with $y_i = x_i + 1$ for $i = 1, 2, \cdots, r$, then the total number of ways to distribute the n indistinguishable balls inside the r urns is:

$$\binom{n+r-1}{r-1}. \quad \blacktriangle$$

■ **EXAMPLE B.16**

Twelve gifts are divided between 7 children. How many different distributions are possible? How many if each kid must get at least one present?

According to the previous example, there are $\binom{12+7-1}{7-1} = \binom{18}{6} = 18,564$ ways of distributing the gifts between the children.

If, furthermore, each kid must receive at least one present, then, the number of possible distributions reduces to $\binom{12-1}{7-1} = \binom{11}{6} = 462$. ▲

EXERCISES

B.1 How many three-digit numbers less than 500 with all digits different from each other can be formed with the digits $1, 2, 3, 4, 5, 6$ and 7?

B.2 How many three-digit numbers can be formed with the digits $1, 4, 8$ and 5 if:

1. The three digits are different?
2. The numbers must be odd?
3. The numbers must be divisible by 5?

B.3 How many ways are there to seat in a row four boys and four girls if they must be seated alternated? How many if the boys and the girls are seated together? How many if only the girls are seated together?

B.4 An inspector checks six different machines during the day. In order to avoid the operators knowing when those inspections are going to be made, he varies the order each time. In how many ways can he do this?

B.5 A group of 5 German, 6 Australian, 4 Japanese and 6 Colombian people must be seated at a round table. How many ways are there to do this? How many if people having the same nationality must be together? How many if the Colombians must stay together?

B.6 How many different arrangements can be formed with the letters in the word "successes"?

B.7 In a probability exam, a student must answer 10 out of 13 questions. How many ways of answering the exam does the student have? How many if he must answer at least 3 from the first 5 questions? How many if he must answer exactly 3 of the first 5 questions?

B.8 The effects of two medications A and B are going to be compared in a pharmaceutical study involving 50 volunteers. Twenty volunteers receive medication A, another 20 receive medication B and the remaining 10 take a placebo. How many different ways of distributing the medications and placebos between the volunteers are there?

B.9 The statistics department of a university has 33 professors who must be divided into four groups of 15, 8, 7 and 3 members. How many possible partitions are there? How many if the 5 marked members of the advisor committee must be in the first group?

B.10 In how many ways can 7 gifts be divided between three children if one child must receive 3 of them and the other two kids must get 2 each?

B.11 The sciences faculty of a university has received 30 identical computers to be split between the 7 departments of the faculty. How many different distributions are possible? How many if each department must get at least 2 computers?

B.12 An investor has \$600,000 to invest in six possible bonds. Each investment has to be made in thousands of dollars. How many investment strategies are possible?

APPENDIX C

TOPICS ON LINEAR ALGEBRA

This appendix presents some of the results from linear algebra used throughout the book. It is assumed that the reader has some knowledge of basic matrix theory. The reader interested in further reading on this topics is referred to Searle (1982).

1. The vectors in $\mathbb{R}^n$ are considered to be row-vectors. If $\mathbf{x} = (x_1, \cdots, x_n)$ and $\mathbf{y} = (y_1, \cdots, y_n)$ are vectors in $\mathbb{R}^n$, then their inner product is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^{n} x_i y_i.$$

The Euclidean norm of a vector $\mathbf{x} = (x_1, \cdots, x_n)$ of $\mathbb{R}^n$ is the real number $\|\mathbf{x}\| := \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$.

2. The determinant of a square matrix A is notated $det\,(A)$.

3. The trace of a square matrix A is written $tr\,(A)$.

4. The transpose of a matrix A is notated A^T.

Introduction to Probability and Stochastic Processes with Applications, First Edition. By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

5. A square matrix is said to be diagonal if all the off-diagonal elements are equal to zero.

6. A diagonal matrix having only 1's in its main diagonal is called the identity of order n (n being the size of the matrix) and is written as I_n.

7. A square matrix A is said to be symmetric if it is equal to its transpose.

8. A square matrix of order n is said to be orthogonal if $AA^T = A^TA = I_n$.

9. The eigenvalues of a matrix A of order n are the solutions to $det\,(A - \lambda I_n) = 0$ where λ is a real or complex number

10. An eigenvector of the square matrix A of order n associated with the eigenvalue λ is a vector $\mathbf{0} \neq \mathbf{v} \in \mathbb{R}^n$ such that $\mathbf{v}A = \lambda\mathbf{v}$.

11. A square matrix A is said to be singular if and only if it has 0 as an eigenvalue. A square matrix that is not singular is called nonsingular.

12. A square matrix of order n is said to be positive-definite if for any vector $\mathbf{x}$ in $\mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}$ we have that $\mathbf{xAx}^T > 0$.

13. A square matrix of order n is said to be positive-semidefinite if for any vector $\mathbf{x}$ in $\mathbb{R}^n$ with $\mathbf{x} \neq \mathbf{0}$ we have that $\mathbf{xAx}^T \geq 0$.

The following theorem presents some of the interesting properties of symmetric and positive-definite matrices:

Theorem C.1 *Let A be a symmetric and positive-definite matrix. Then:*

1. *There is a matrix W such that $A = WW^T$.*

2. *The eigenvalues of A are all positive (and therefore the matrix A is nonsingular).*

3. *The inverse matrix of A is symmetric and positive-definite.*

4. *There exist square matrices Λ and V such that $A = V\Lambda V^T$, where Λ is a diagonal matrix whose entries correspond to the eigenvalues of A and V is an orthogonal matrix whose columns are the vectors $\frac{\mathbf{v}}{\|\mathbf{v}\|}$, with $\mathbf{v}$ an eigenvector of A. (This still holds if A is just symmetric.)*

APPENDIX D

STATISTICAL TABLES

D.1 BINOMIAL PROBABILITIES

$$P(X \leq t) = \sum_{x=1}^{[t]} \binom{n}{x} p^x (1-p)^{n-x}$$

		p						
n	$[t]$	**0.05**	**0.1**	**0.15**	**0.20**	**0.30**	**0.40**	**0.50**
2	0	0.9025	0.8100	0.7225	0.6400	0.4900	0.3600	0.2500
	1	0.9975	0.9900	0.9775	0.9600	0.9100	0.8400	0.7500
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.8574	0.7290	0.6141	0.5120	0.3430	0.2160	0.1250
	1	0.9927	0.9720	0.9392	0.8960	0.7840	0.6480	0.5000
	2	0.9999	0.9990	0.9966	0.9920	0.9730	0.9360	0.8750
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

(continued)

Introduction to Probability and Stochastic Processes with Applications, First Edition.
By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

(continued)

n	$[t]$	0.05	0.1	0.15	0.20	0.30	0.40	0.50
4	0	0.8145	0.6561	0.5220	0.4096	0.2401	0.1296	0.0625
	1	0.9860	0.9477	0.8905	0.8192	0.6517	0.4752	0.3125
	2	0.9995	0.9963	0.9880	0.9728	0.9163	0.8208	0.6875
	3	1.0000	0.9999	0.9995	0.9984	0.9919	0.9744	0.9375
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
5	0	0.7738	0.5905	0.4437	0.3277	0.1681	0.0778	0.0313
	1	0.9774	0.9185	0.8352	0.7373	0.5282	0.3370	0.1875
	2	0.9988	0.9914	0.9734	0.9421	0.8369	0.6826	0.5000
	3	1.0000	0.9995	0.9978	0.9933	0.9692	0.9130	0.8125
	4	1.0000	1.0000	0.9999	0.9997	0.9976	0.9898	0.9687
	5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
6	0	0.7351	0.5314	0.3771	0.2621	0.1176	0.0467	0.0156
	1	0.9672	0.8857	0.7765	0.6554	0.4202	0.2333	0.1094
	2	0.9978	0.9842	0.9527	0.9011	0.7443	0.5443	0.3438
	3	0.9999	0.9987	0.9941	0.9830	0.9295	0.8208	0.6563
	4	1.0000	0.9999	0.9996	0.9984	0.9891	0.9590	0.8906
	5	1.0000	1.0000	1.0000	0.9999	0.9993	0.9959	0.9844
	6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
7	0	0.6983	0.4783	0.3206	0.2097	0.0824	0.0280	0.0078
	1	0.9556	0.8503	0.7166	0.5767	0.3294	0.1586	0.0625
	2	0.9962	0.9743	0.9262	0.8520	0.6471	0.4199	0.2266
	3	0.9998	0.9973	0.9879	0.9667	0.8740	0.7102	0.5000
	4	1.0000	0.9998	0.9988	0.9953	0.9712	0.9037	0.7734
	5	1.0000	1.0000	0.9999	0.9996	0.9962	0.9812	0.9375
	6	1.0000	1.0000	1.0000	1.0000	0.9998	0.9984	0.9922
	7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
8	0	0.6634	0.4305	0.2725	0.1678	0.0576	0.0168	0.0039
	1	0.9428	0.8131	0.6572	0.5033	0.2553	0.1064	0.0352
	2	0.9942	0.9619	0.8948	0.7969	0.5518	0.3154	0.1445
	3	0.9996	0.9950	0.9786	0.9437	0.8059	0.5941	0.3633
	4	1.0000	0.9996	0.9971	0.9896	0.9420	0.8263	0.6367
	5	1.0000	1.0000	0.9998	0.9988	0.9887	0.9502	0.8555
	6	1.0000	1.0000	1.0000	0.9999	0.9987	0.9915	0.9648
	7	1.0000	1.0000	1.0000	1.0000	0.9999	0.9993	0.9961
	8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
9	0	0.6302	0.3874	0.2316	0.1342	0.0404	0.0101	0.0020
	1	0.9288	0.7748	0.5995	0.4362	0.1960	0.0705	0.0195
	2	0.9916	0.9470	0.8591	0.7382	0.4628	0.2318	0.0898
	3	0.9994	0.9917	0.9661	0.9144	0.7297	0.4826	0.2539
	4	1.0000	0.9991	0.9944	0.9804	0.9012	0.7334	0.5000
	5	1.0000	0.9999	0.9994	0.9969	0.9747	0.9006	0.7461
	6	1.0000	1.0000	1.0000	0.9997	0.9957	0.9750	0.9102

(continued)

(continued)

n	$[t]$	0.05	0.1	0.15	0.20	0.30	0.40	0.50
	7	1.0000	1.0000	1.0000	1.0000	0.9996	0.9962	0.9805
	8	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9980
	9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
10	0	0.5987	0.3487	0.1969	0.1074	0.0282	0.0060	0.0010
	1	0.9139	0.7361	0.5443	0.3758	0.1493	0.0464	0.0107
	2	0.9885	0.9298	0.8202	0.6778	0.3828	0.1673	0.0547
	3	0.9990	0.9872	0.9500	0.8791	0.6496	0.3823	0.1719
	4	0.9999	0.9984	0.9901	0.9672	0.8497	0.6331	0.3770
	5	1.0000	0.9999	0.9986	0.9936	0.9527	0.8338	0.6230
	6	1.0000	1.0000	0.9999	0.9991	0.9894	0.9452	0.8281
	7	1.0000	1.0000	1.0000	0.9999	0.9984	0.9877	0.9453
	8	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983	0.9893
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9990
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
11	0	0.5688	0.3138	0.1673	0.0859	0.0198	0.0036	0.0005
	1	0.8981	0.6974	0.4922	0.3221	0.1130	0.0302	0.0059
	2	0.9848	0.9104	0.7788	0.6174	0.3127	0.1189	0.0327
	3	0.9984	0.9815	0.9306	0.8389	0.5696	0.2963	0.1133
	4	0.9999	0.9972	0.9841	0.9496	0.7897	0.5328	0.2744
	5	1.0000	0.9997	0.9973	0.9883	0.9218	0.7535	0.5000
	6	1.0000	1.0000	0.9997	0.9980	0.9784	0.9006	0.7256
	7	1.0000	1.0000	1.0000	0.9998	0.9957	0.9707	0.8867
	8	1.0000	1.0000	1.0000	1.0000	0.9994	0.9941	0.9673
	9	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993	0.9941
	10	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
12	0	0.5404	0.2824	0.1422	0.0687	0.0138	0.0022	0.0002
	1	0.8816	0.6590	0.4435	0.2749	0.0850	0.0196	0.0032
	2	0.9804	0.8891	0.7358	0.5583	0.2528	0.0834	0.0193
	3	0.9978	0.9744	0.9078	0.7946	0.4925	0.2253	0.0730
	4	0.9998	0.9957	0.9761	0.9274	0.7237	0.4382	0.1938
	5	1.0000	0.9995	0.9954	0.9806	0.8822	0.6652	0.3872
	6	1.0000	0.9999	0.9993	0.9961	0.9614	0.8418	0.6128
	7	1.0000	1.0000	0.9999	0.9994	0.9905	0.9427	0.8062
	8	1.0000	1.0000	1.0000	0.9999	0.9983	0.9847	0.9270
	9	1.0000	1.0000	1.0000	1.0000	0.9998	0.9972	0.9807
	10	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9968
	11	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
13	0	0.5133	0.2542	0.1209	0.0550	0.0097	0.0013	0.0001
	1	0.8646	0.6213	0.3983	0.2336	0.0637	0.0126	0.0017
	2	0.9755	0.8661	0.6920	0.5017	0.2025	0.0579	0.0112

(continued)

(continued)

n	$[t]$	0.05	0.1	0.15	0.20	0.30	0.40	0.50
	3	0.9969	0.9658	0.8820	0.7473	0.4206	0.1686	0.0461
	4	0.9997	0.9935	0.9658	0.9009	0.6543	0.3530	0.1334
	5	1.0000	0.9991	0.9925	0.9700	0.8346	0.5744	0.2905
	6	1.0000	0.9999	0.9987	0.9930	0.9376	0.7712	0.5000
	7	1.0000	1.0000	0.9998	0.9988	0.9818	0.9023	0.7095
	8	1.0000	1.0000	1.0000	0.9998	0.9960	0.9679	0.8666
	9	1.0000	1.0000	1.0000	1.0000	0.9993	0.9922	0.9539
	10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9987	0.9888
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9983
	12	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
14	0	0.4877	0.2288	0.1028	0.0440	0.0068	0.0008	0.0001
	1	0.8470	0.5846	0.3567	0.1979	0.0475	0.0081	0.0009
	2	0.9699	0.8416	0.6479	0.4481	0.1608	0.0398	0.0065
	3	0.9958	0.9559	0.8535	0.6982	0.3552	0.1243	0.0287
	4	0.9996	0.9908	0.9533	0.8702	0.5842	0.2793	0.0898
	5	1.0000	0.9985	0.9885	0.9561	0.7805	0.4859	0.2120
	6	1.0000	0.9998	0.9978	0.9884	0.9067	0.6925	0.3953
	7	1.0000	1.0000	0.9997	0.9976	0.9685	0.8499	0.6047
	8	1.0000	1.0000	1.0000	0.9996	0.9917	0.9417	0.7880
	9	1.0000	1.0000	1.0000	1.0000	0.9983	0.9825	0.9102
	10	1.0000	1.0000	1.0000	1.0000	0.9998	0.9961	0.9713
	11	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9935
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9991
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
15	0	0.4633	0.2059	0.0874	0.0352	0.0047	0.0005	0.0000
	1	0.8290	0.5490	0.3186	0.1671	0.0353	0.0052	0.0005
	2	0.9638	0.8159	0.6042	0.3980	0.1268	0.0271	0.0037
	3	0.9945	0.9444	0.8227	0.6482	0.2969	0.0905	0.0176
	4	0.9994	0.9873	0.9383	0.8358	0.5155	0.2173	0.0592
	5	0.9999	0.9978	0.9832	0.9389	0.7216	0.4032	0.1509
	6	1.0000	0.9997	0.9964	0.9819	0.8689	0.6098	0.3036
	7	1.0000	1.0000	0.9994	0.9958	0.9500	0.7869	0.5000
	8	1.0000	1.0000	0.9999	0.9992	0.9848	0.9050	0.6964
	9	1.0000	1.0000	1.0000	0.9999	0.9963	0.9662	0.8491
	10	1.0000	1.0000	1.0000	1.0000	0.9993	0.9907	0.9408
	11	1.0000	1.0000	1.0000	1.0000	0.9999	0.9981	0.9824
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9963
	13	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

(continued)

(continued)

n	$[t]$	0.05	0.1	0.15	0.20	0.30	0.40	0.50
16	0	0.4401	0.1853	0.0743	0.0281	0.0033	0.0003	0.0000
	1	0.8108	0.5147	0.2839	0.1407	0.0261	0.0033	0.0003
	2	0.9571	0.7892	0.5614	0.3518	0.0994	0.0183	0.0021
	3	0.9930	0.9316	0.7899	0.5981	0.2459	0.0651	0.0106
	4	0.9991	0.9830	0.9209	0.7982	0.4499	0.1666	0.0384
	5	0.9999	0.9967	0.9765	0.9183	0.6598	0.3288	0.1051
	6	1.0000	0.9995	0.9944	0.9733	0.8247	0.5272	0.2272
	7	1.0000	0.9999	0.9989	0.9930	0.9256	0.7161	0.4018
	8	1.0000	1.0000	0.9998	0.9985	0.9743	0.8577	0.5982
	9	1.0000	1.0000	1.0000	0.9998	0.9929	0.9417	0.7728
	10	1.0000	1.0000	1.0000	1.0000	0.9984	0.9809	0.8949
	11	1.0000	1.0000	1.0000	1.0000	0.9997	0.9951	0.9616
	12	1.0000	1.0000	1.0000	1.0000	1.0000	0.9991	0.9894
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9979
	14	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
17	0	0.4181	0.1668	0.0631	0.0225	0.0023	0.0002	0.0000
	1	0.7922	0.4818	0.2525	0.1182	0.0193	0.0021	0.0001
	2	0.9497	0.7618	0.5198	0.3096	0.0774	0.0123	0.0012
	3	0.9912	0.9174	0.7556	0.5489	0.2019	0.0464	0.0064
	4	0.9988	0.9779	0.9013	0.7582	0.3887	0.1260	0.0245
	5	0.9999	0.9953	0.9681	0.8943	0.5968	0.2639	0.0717
	6	1.0000	0.9992	0.9917	0.9623	0.7752	0.4478	0.1662
	7	1.0000	0.9999	0.9983	0.9891	0.8954	0.6405	0.3145
	8	1.0000	1.0000	0.9997	0.9974	0.9597	0.8011	0.5000
	9	1.0000	1.0000	1.0000	0.9995	0.9873	0.9081	0.6855
	10	1.0000	1.0000	1.0000	0.9999	0.9968	0.9652	0.8338
	11	1.0000	1.0000	1.0000	1.0000	0.9993	0.9894	0.9283
	12	1.0000	1.0000	1.0000	1.0000	0.9999	0.9975	0.9755
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9995	0.9936
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9988
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
18	0	0.3972	0.1501	0.0536	0.0180	0.0016	0.0001	0.0000
	1	0.7735	0.4503	0.2241	0.0991	0.0142	0.0013	0.0001
	2	0.9419	0.7338	0.4797	0.2713	0.0600	0.0082	0.0007
	3	0.9891	0.9018	0.7202	0.5010	0.1646	0.0328	0.0038
	4	0.9985	0.9718	0.8794	0.7164	0.3327	0.0942	0.0154
	5	0.9998	0.9936	0.9581	0.8671	0.5344	0.2088	0.0481
	6	1.0000	0.9988	0.9882	0.9487	0.7217	0.3743	0.1189

(continued)

(continued)

n	[t]	0.05	0.1	0.15	0.20	0.30	0.40	0.50
	7	1.0000	0.9998	0.9973	0.9837	0.8593	0.5634	0.2403
	8	1.0000	1.0000	0.9995	0.9957	0.9404	0.7368	0.4073
	9	1.0000	1.0000	0.9999	0.9991	0.9790	0.8653	0.5927
	10	1.0000	1.0000	1.0000	0.9998	0.9939	0.9424	0.7597
	11	1.0000	1.0000	1.0000	1.0000	0.9986	0.9797	0.8811
	12	1.0000	1.0000	1.0000	1.0000	0.9997	0.9942	0.9519
	13	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987	0.9846
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	0.9962
	15	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9993
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
19	0	0.3774	0.1351	0.0456	0.0144	0.0011	0.0001	0.0000
	1	0.7547	0.4203	0.1985	0.0829	0.0104	0.0008	0.0000
	2	0.9335	0.7054	0.4413	0.2369	0.0462	0.0055	0.0004
	3	0.9868	0.8850	0.6841	0.4551	0.1332	0.0230	0.0022
	4	0.9980	0.9648	0.8556	0.6733	0.2822	0.0696	0.0096
	5	0.9998	0.9914	0.9463	0.8369	0.4739	0.1629	0.0318
	6	1.0000	0.9983	0.9837	0.9324	0.6655	0.3081	0.0835
	7	1.0000	0.9997	0.9959	0.9767	0.8180	0.4878	0.1796
	8	1.0000	1.0000	0.9992	0.9933	0.9161	0.6675	0.3238
	9	1.0000	1.0000	0.9999	0.9984	0.9674	0.8139	0.5000
	10	1.0000	1.0000	1.0000	0.9997	0.9895	0.9115	0.6762
	11	1.0000	1.0000	1.0000	1.0000	0.9972	0.9648	0.8204
	12	1.0000	1.0000	1.0000	1.0000	0.9994	0.9884	0.9165
	13	1.0000	1.0000	1.0000	1.0000	0.9999	0.9969	0.9682
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9994	0.9904
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9999	0.9978
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9996
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
20	0	0.3585	0.1216	0.0388	0.0115	0.0008	0.0000	0.0000
	1	0.7358	0.3917	0.1756	0.0692	0.0076	0.0005	0.0000
	2	0.9245	0.6769	0.4049	0.2061	0.0355	0.0036	0.0002
	3	0.9841	0.8670	0.6477	0.4114	0.1071	0.0160	0.0013
	4	0.9974	0.9568	0.8298	0.6296	0.2375	0.0510	0.0059
	5	0.9997	0.9887	0.9327	0.8042	0.4164	0.1256	0.0207
	6	1.0000	0.9976	0.9781	0.9133	0.6080	0.2500	0.0577
	7	1.0000	0.9996	0.9941	0.9679	0.7723	0.4159	0.1316
	8	1.0000	0.9999	0.9987	0.9900	0.8867	0.5956	0.2517
	9	1.0000	1.0000	0.9998	0.9974	0.9520	0.7553	0.4119

(continued)

(continued)

n	$[t]$	0.05	0.1	0.15	0.20	0.30	0.40	0.50
	10	1.0000	1.0000	1.0000	0.9994	0.9829	0.8725	0.5881
	11	1.0000	1.0000	1.0000	0.9999	0.9949	0.9435	0.7483
	12	1.0000	1.0000	1.0000	1.0000	0.9987	0.9790	0.8684
	13	1.0000	1.0000	1.0000	1.0000	0.9997	0.9935	0.9423
	14	1.0000	1.0000	1.0000	1.0000	1.0000	0.9984	0.9793
	15	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9941
	16	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9987
	17	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998
	18	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	19	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
	20	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

D.2 POISSON PROBABILITIES

$$P(X \leq t) = \sum_{k=0}^{[t]} e^{-\lambda} \frac{\lambda^k}{k!}$$

$\lambda \diagdown [t]$	0	1	2	3	4	5	6
0.1	0.9048	0.9953	0.9998	1.0000	1.0000	1.0000	1.0000
0.2	0.8187	0.9825	0.9989	0.9999	1.0000	1.0000	1.0000
0.3	0.7408	0.9631	0.9964	0.9997	1.0000	1.0000	1.0000
0.4	0.6703	0.9384	0.9921	0.9992	0.9999	1.0000	1.0000
0.5	0.6065	0.9098	0.9856	0.9982	0.9998	1.0000	1.0000
0.6	0.5488	0.8781	0.9769	0.9966	0.9996	1.0000	1.0000
0.7	0.4966	0.8442	0.9659	0.9942	0.9992	0.9999	1.0000
0.8	0.4493	0.8088	0.9526	0.9909	0.9986	0.9998	1.0000
0.9	0.4066	0.7725	0.9371	0.9865	0.9977	0.9997	1.0000
1.0	0.3679	0.7358	0.9197	0.9810	0.9963	0.9994	0.9999
1.2	0.3012	0.6626	0.8795	0.9662	0.9923	0.9985	0.9997
1.4	0.2466	0.5918	0.8335	0.9463	0.9857	0.9968	0.9994
1.6	0.2019	0.5249	0.7834	0.9212	0.9763	0.9940	0.9987
1.8	0.1653	0.4628	0.7306	0.8913	0.9636	0.9896	0.9974
2.0	0.1353	0.4060	0.6767	0.8571	0.9473	0.9834	0.9955
2.5	0.0821	0.2873	0.5438	0.7576	0.8912	0.9580	0.9858
3.0	0.0498	0.1991	0.4232	0.6472	0.8153	0.9161	0.9665
3.5	0.0302	0.1359	0.3208	0.5366	0.7254	0.8576	0.9347
4.0	0.0183	0.0916	0.2381	0.4335	0.6288	0.7851	0.8893
5.0	0.0067	0.0404	0.1247	0.2650	0.4405	0.6160	0.7622

(continued)

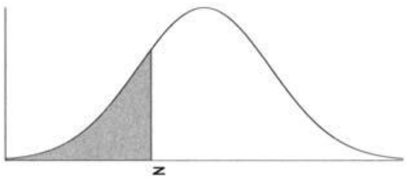
(continued)

$\lambda \setminus [t]$	0	1	2	3	4	5	6
6.0	0.0025	0.0174	0.0620	0.1512	0.2851	0.4457	0.6063
7.0	0.0009	0.0073	0.0296	0.0818	0.1730	0.3007	0.4497
8.0	0.0003	0.0030	0.0138	0.0424	0.0996	0.1912	0.3134
9.0	0.0001	0.0012	0.0062	0.0212	0.0550	0.1157	0.2068
10.0	0.0000	0.0005	0.0028	0.0103	0.0293	0.0671	0.1301
15.0	0.0000	0.0000	0.0000	0.0002	0.0009	0.0028	0.0076
20.0	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003

$\lambda \setminus [t]$	7	8	9	10	11	12
0.1	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.5	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.7	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.8	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
0.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.0	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
1.4	0.9999	1.0000	1.0000	1.0000	1.0000	1.0000
1.6	0.9997	1.0000	1.0000	1.0000	1.0000	1.0000
1.8	0.9994	0.9999	1.0000	1.0000	1.0000	1.0000
2.0	0.9989	0.9998	1.0000	1.0000	1.0000	1.0000
2.5	0.9958	0.9989	0.9997	0.9999	1.0000	1.0000
3.0	0.9881	0.9962	0.9989	0.9997	0.9999	1.0000
3.5	0.9733	0.9901	0.9967	0.9990	0.9997	0.9999
4.0	0.9489	0.9786	0.9919	0.9972	0.9991	0.9997
5.0	0.8666	0.9319	0.9682	0.9863	0.9945	0.9980
6.0	0.7440	0.8472	0.9161	0.9574	0.9799	0.9912
7.0	0.5987	0.7291	0.8305	0.9015	0.9467	0.9730
8.0	0.4530	0.5925	0.7166	0.8159	0.8881	0.9362
9.0	0.3239	0.4557	0.5874	0.7060	0.8030	0.8758
10.0	0.2202	0.3328	0.4579	0.5830	0.6968	0.7916
15.0	0.0180	0.0374	0.0699	0.1185	0.1848	0.2676
20.0	0.0008	0.0021	0.0050	0.0108	0.0214	0.0390

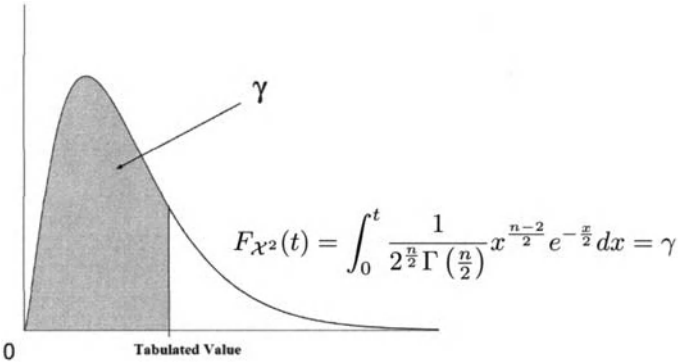
D.3 STANDARD NORMAL DISTRIBUTION FUNCTION

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}} dt$$



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990

D.4 CHI-SQUARE DISTRIBUTION FUNCTION



$n \backslash \gamma$	0.995	0.990	0.975	0.950	0.900	0.750
1	7.8794	6.6349	5.0239	3.8415	2.7055	1.3233
2	10.5966	9.2103	7.3778	5.9915	4.6052	2.7726
3	12.8382	11.3449	9.3484	7.8147	6.2514	4.1083
4	14.8603	13.2767	11.1433	9.4877	7.7794	5.3853
5	16.7496	15.0863	12.8325	11.0705	9.2364	6.6257
6	18.5476	16.8119	14.4494	12.5916	10.6446	7.8408
7	20.2777	18.4753	16.0128	14.0671	12.0170	9.0371
8	21.9550	20.0902	17.5345	15.5073	13.3616	10.2189
9	23.5894	21.6660	19.0228	16.9190	14.6837	11.3888
10	25.1882	23.2093	20.4832	18.3070	15.9872	12.5489
11	26.7568	24.7250	21.9200	19.6751	17.2750	13.7007
12	28.2995	26.2170	23.3367	21.0261	18.5493	14.8454
13	29.8195	27.6882	24.7356	22.3620	19.8119	15.9839
14	31.3193	29.1412	26.1189	23.6848	21.0641	17.1169
15	32.8013	30.5779	27.4884	24.9958	22.3071	18.2451
16	34.2672	31.9999	28.8454	26.2962	23.5418	19.3689
17	35.7185	33.4087	30.1910	27.5871	24.7690	20.4887
18	37.1565	34.8053	31.5264	28.8693	25.9894	21.6049
19	38.5823	36.1909	32.8523	30.1435	27.2036	22.7178
20	39.9968	37.5662	34.1696	31.4104	28.4120	23.8277
21	41.4011	38.9322	35.4789	32.6706	29.6151	24.9348
22	42.7957	40.2894	36.7807	33.9244	30.8133	26.0393
23	44.1813	41.6384	38.0756	35.1725	32.0069	27.1413
24	45.5585	42.9798	39.3641	36.4150	33.1962	28.2412
25	46.9279	44.3141	40.6465	37.6525	34.3816	29.3389
26	48.2899	45.6417	41.9232	38.8851	35.5632	30.4346
27	49.6449	46.9629	43.1945	40.1133	36.7412	31.5284
28	50.9934	48.2782	44.4608	41.3371	37.9159	32.6205
29	52.3356	49.5879	45.7223	42.5570	39.0875	33.7109
30	53.6720	50.8922	46.9792	43.7730	40.2560	34.7997
40	66.7660	63.6907	59.3417	55.7585	51.8051	45.6160
50	79.4900	76.1539	71.4202	67.5048	63.1671	56.3336
60	91.9517	88.3794	83.2977	79.0819	74.3970	66.9815
70	104.2149	100.4252	95.0232	90.5312	85.5270	77.5767
80	116.3211	112.3288	106.6286	101.8795	96.5782	88.1303
90	128.2989	124.1163	118.1359	113.1453	107.5650	98.6499

(continued)

(continued)

$n \diagdown \gamma$	**0.995**	**0.990**	**0.975**	**0.950**	**0.900**	**0.750**
100	140.1695	135.8067	129.5612	124.3421	118.4980	109.1412
200	255.2642	249.4451	241.0579	233.9943	226.0210	213.1022
500	585.2066	576.4928	563.8515	553.1268	540.9303	520.9505
1000	1118.9481	1106.9690	1089.5309	1074.6794	1057.7239	1029.7898

$n \diagdown \gamma$	**0.500**	**0.250**	**0.100**	**0.050**	**0.025**	**0.010**	**0.005**
1	0.455	0.102	0.016	0.004	0.001	0.000	0.000
2	1.386	0.575	0.211	0.103	0.051	0.020	0.010
3	2.366	1.213	0.584	0.352	0.216	0.115	0.072
4	3.357	1.923	1.064	0.711	0.484	0.297	0.207
5	4.351	2.675	1.610	1.145	0.831	0.554	0.412
6	5.348	3.455	2.204	1.635	1.237	0.872	0.676
7	6.346	4.255	2.833	2.167	1.690	1.239	0.989
8	7.344	5.071	3.490	2.733	2.180	1.646	1.344
9	8.343	5.899	4.168	3.325	2.700	2.088	1.735
10	9.342	6.737	4.865	3.940	3.247	2.558	2.156
11	10.341	7.584	5.578	4.575	3.816	3.053	2.603
12	11.340	8.438	6.304	5.226	4.404	3.571	3.074
13	12.340	9.299	7.042	5.892	5.009	4.107	3.565
14	13.339	10.165	7.790	6.571	5.629	4.660	4.075
15	14.339	11.037	8.547	7.261	6.262	5.229	4.601
16	15.338	11.912	9.312	7.962	6.908	5.812	5.142
17	16.338	12.792	10.085	8.672	7.564	6.408	5.697
18	17.338	13.675	10.865	9.390	8.231	7.015	6.265
19	18.338	14.562	11.651	10.117	8.907	7.633	6.844
20	19.337	15.452	12.443	10.851	9.591	8.260	7.434
21	20.337	16.344	13.240	11.591	10.283	8.897	8.034
22	21.337	17.240	14.041	12.338	10.982	9.542	8.643
23	22.337	18.137	14.848	13.091	11.689	10.196	9.260
24	23.337	19.037	15.659	13.848	12.401	10.856	9.886
25	24.337	19.939	16.473	14.611	13.120	11.524	10.520
26	25.336	20.843	17.292	15.379	13.844	12.198	11.160
27	26.336	21.749	18.114	16.151	14.573	12.879	11.808
28	27.336	22.657	18.939	16.928	15.308	13.565	12.461
29	28.336	23.567	19.768	17.708	16.047	14.256	13.121
30	29.336	24.478	20.599	18.493	16.791	14.953	13.787
40	39.335	33.660	29.051	26.509	24.433	22.164	20.707
50	49.335	42.942	37.689	34.764	32.357	29.707	27.991
60	59.335	52.294	46.459	43.188	40.482	37.485	35.534
70	69.334	61.698	55.329	51.739	48.758	45.442	43.275
80	79.334	71.145	64.278	60.391	57.153	53.540	51.172
90	89.334	80.625	73.291	69.126	65.647	61.754	59.196
100	99.334	90.133	82.358	77.929	74.222	70.065	67.328
200	199.334	186.172	174.835	168.279	162.728	156.432	152.241
500	499.333	478.323	459.926	449.147	439.936	429.388	422.303
1000	999.333	969.484	943.133	927.594	914.257	898.912	888.564

SELECTED PROBLEM SOLUTIONS

SOLUTIONS FOR CHAPTER 1

1.1 $A \cap B = \{(H,H,T),(H,T,T)\}$; $A \cup B = \{(H,H,H),(H,T,H),(T,H,T),(T,T,T),(H,H,T),(H,T,T)\}$; $A^c = \{(T,H,H),(T,T,H),(T,H,T),(T,T,T)\}$; $A^c \cap B^c = \{(T,H,H),(T,T,H)\}$; $A \cap B^c = \{(H,T,H),(H,H,H)\}$

1.2 a) $A \cap B \cap C^c$ b) $A \cap B \cap C$ c) $A \cap B^c \cap C^c$ d) $A \cup B \cup C$ e) $E = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \cup (A^c \cap B^c \cap C^c)$ f) $D = E \cup (A \cap B \cap C^c) \cup (A^c \cap B \cap C) \cup (A \cap B^c \cap C)$

1.3 a) $\{(2,2),(2,3),(4,2)\}$ c) $\emptyset$

1.5 a) $B = \cup_{i=1}^n A_i$ b) $C = \cap_{i=1}^n A_i^c$ c) $D = \cup_{i=1}^n \cap_{\substack{j=1 \\ i \neq j}}^n (A_i \cap A_j^c)$ d) $E = C \cup D$

1.6 a) $E_1 = A \cup B \cup C$ b) $E_2 = (A \cap B \cap C^c) \cup (A \cap B^c \cap C) \cup (A^c \cap B \cap C)$ c) $E_3 = E_2 \cup (A \cap B \cap C)$ d) $E_4 = (A \cap B^c \cap C^c) \cup (A^c \cap B \cap C^c) \cup (A^c \cap B^c \cap C) \cup (A^c \cap B^c \cap C^c)$

1.7 33%; 60%

1.9 $\Im_1 = \{\emptyset, \Omega\}$; $\Im_2 = \{\emptyset, \{1\}, \{2,3,4\}, \Omega\}$; $\Im_3 = \{\emptyset, \{1\}, \{2\}, \{1,2\}, \{3,4\}, \{2,3,4\}, \{1,3,4\}, \Omega\}$; $\Im_4 = \{\emptyset, \{1\}, \{2\}, \{3\}, \{4\}, \{1,2\}, \{1,3\}, \{1,4\}, \{2,3\}, \{2,4\}, \{3,4\}, \{2,3,4\}, \{1,3,4\}, \{1,2,4\}, \{1,2,3\}, \Omega\}$

1.13 a) $\sigma(\{2\},\{3\}) = \{\emptyset, \Omega, \{1\}, \{2\}, \{3\}, \{2,3\}, \{1,3\}, \{1,2\}\}$ b) Yes

1.14 a) $\alpha = \frac{17}{63}$ b) $\frac{46}{63}$; $\frac{45}{63}$; $\frac{35}{63}$

1.18 No

1.19 $\frac{2}{3}$; $\frac{4}{9}$; $\frac{2}{9}$

1.20 a) $\frac{8}{100}$ b) $\frac{4}{100}$ c) $\frac{149}{198}$

1.21 No

1.22 $P(A_k) = \frac{1}{2^k}$ for $k = 1, 2, \cdots, n$

1.23 $1 - \sum_{j=1}^{10} (-1)^{j-1} \binom{7}{j} \left(1 - \frac{j}{7}\right)^{10}$; 0.37937

1.24 0.24763

1.26 $\frac{43}{216}$; $\frac{173}{216}$

1.28 $\frac{43}{84}$

1.29 $0.58\overline{3}$

1.30 8.3479×10^{-5}

1.31 $\frac{5}{84}$

1.33 a) $1 - p^5(2 - p^2)$ b) $1 - p^4(2 - p^2)$ where $p = \frac{1}{5}$

1.34 0.5687

1.35 $1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \cdots + \frac{(-1)^n}{n!} \simeq e^{-1}$

1.37 $\frac{6}{15} \times \frac{8}{17} \times \frac{10}{19}$

1.38 9.3778×10^{-2}; 0.43925

1.39 $\frac{1}{3}$; $\frac{4}{7}$; 1; $\frac{2}{7}$; $\frac{1}{7}$

1.40 $\frac{5}{6}$; $\frac{1}{3}$

1.41 a) 1 b) 0.8748 c) 0.9318

1.42 $\frac{1}{13}$; 0.7159

1.43 $\frac{2}{6} \times \frac{8}{13} \times \frac{2}{6} + \frac{7}{12} \times \frac{3}{7} \times \frac{4}{6}$

1.44 0.205

1.45 $\frac{8}{13}$

1.46 $\frac{1}{4}$

1.47 0.43; $\frac{35}{43}$

1.51 $\frac{5}{9}$

1.52 $\{\emptyset, \{a,b\}, \{a,c\}, \{b,c\}, \Omega\}$

1.54 $\frac{13}{84}$; $\frac{5}{28}$

1.63 a) 0.703704 b) 0.25926 c) $0.\overline{2}$

1.64 a) 6.12×10^{-3} b) 0.9938 c) 3.1874×10^{-2}

1.66 $\frac{\binom{n}{k}\binom{N-n}{m-k}}{\binom{N}{m}}$

1.67 a) $\frac{1}{n}\left(\left[\frac{n}{3}\right] + \left[\frac{n}{4}\right] - \left[\frac{n}{12}\right]\right)$ b) $\frac{1}{2}$

SOLUTIONS FOR CHAPTER 2

2.1 No

2.2 $\wp(\Omega)$

2.3 $E(X) = \frac{9}{7}$

2.4 a) $\frac{1}{3}$ b) $\frac{5}{6}$ c) $\frac{2}{3}$

2.6 $$F_Y(y) = \begin{cases} 0 & \text{if } y < -2 \\ \frac{1}{16} & \text{if } -2 \le y < -1 \\ \frac{5}{16} & \text{if } -1 \le y < 0 \\ \frac{11}{16} & \text{if } 0 \le y < 1 \\ \frac{15}{16} & \text{if } 1 \le y < 2 \\ 1 & \text{if } y \ge 2 \end{cases}$$

2.7 a) $S = \{1,2,3\}$
b) $P(X_3 = 3 \mid X_2 \in \{1,2\},\ X_1 = 3) = 0$ and $P(X_3 = 3 \mid X_2 \in \{1,2\}) = \frac{1}{2}$

2.8 7.5758×10^{-2}; 4.8351×10^{-2}

2.9 a) $c = 1.2$ c) 0.25 d) 0.710

2.10 a) $c = \frac{2}{3}$ b) $F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{4}{3}x - \frac{x^2}{3} & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$ c) 0.74667

2.11 a) 0; 0.9375 b) $f_X(x) = \begin{cases} 2 - 2x & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$

2.12 a) X is discrete; Y is mixed and Z is continuous b) 0.5; 0.3; 0.2 c) 0.5; 0.75; 0.5 d) 0; 0.5; 0

2.13 $P(X = k) = \frac{2(n-k)+1}{n^2}, \quad k = 1, 2, \cdots, n$

2.14 a) $c = 1$ b) $c^{-1} = \ln 2$ c) $c^{-1} = \frac{\pi^2}{6}$ d) $c^{-1} = e^2 - 1$

2.15 a) $c = \pi^{-1}$ b) $c = 1$

2.16 a) $c^{-1} = 8$ b) 0.70313

2.18 a) $F_X(x) = \sum_{k=0}^{[x]} pq^{k-1}$ b) $F_X(x) = \frac{[x]([x]+1)(2[x]+1)}{n(n+1)(2n+1)}$

2.19 a) $P(X = i) = \frac{1}{7}, \quad i = 1, 2, \cdots, 7$ b) $P(X \leq 2) = \frac{2}{7}; P(X = 5) = \frac{1}{7}$

2.20 0.617

2.21 $P(X = -6000) = 0.23333$, $P(X = -3000) = 0.2$, $P(X = 0) = 0.025$, $P(X = 2000) = 0.33333$, $P(X = 5000) = 0.125$, $P(X = 10000) = 0.08334$

2.22 $P(X = 0) = 0.18$, $P(X = 1000) = 0.27$, $P(X = 1800) = 0.27$, $P(X = 2000) = 0.07$, $P(X = 2800) = 0.14$, $P(X = 3600) = 0.07$

2.24 Yes; No

2.26

$$F_d(x) = \begin{cases} 0 & \text{si} & x < \frac{1}{5} \\ \frac{1}{20} & \text{si} & \frac{1}{5} \leq x < \frac{2}{5} \\ \frac{1}{5} & \text{si} & \frac{2}{5} \leq x < \frac{3}{5} \\ 1 & \text{si} & x \geq \frac{3}{5} \end{cases}$$

$$F_c(x) = \begin{cases} 0 & \text{si} & x < \frac{2}{5} \\ 5x^2 - \frac{4}{5} & \text{si} & \frac{2}{5} \leq x < \frac{3}{5} \\ 1 & \text{si} & x \geq \frac{3}{5} \end{cases}$$

2.31 $E(X) = 0.4$

2.33 a) $C^{-1} = \pi$ b) 0.5 c) $E(X)$ and $Var(X)$ does not exist d) $F_X(x) = \frac{1}{\pi}\left(\tan^{-1} x + \frac{\pi}{2}\right)$

2.34 $\frac{161}{36}$; $\frac{91}{36}$

2.35 a) $\mu = 1$, $\sigma^2 = 0.0667$ b) 0.94536

2.36 a) $\mu = \frac{29}{60}$, $\sigma^2 = 6.6389 \times 10^{-2}$ b) $c = 0.5$

2.37 a) $c = \frac{\pi}{10}$ b) $E(X) = \frac{5}{2}$ and $Var(X) = 50\left(\frac{1}{8} - \frac{1}{\pi^2}\right)$

2.40 $P(X = 0) = 0.08546$, $P(X = 1) = 0.37821$, $P(X = 2) = 0.40812$, $P(X = 3) = 0.12821$

2.45 a) $C = 2$ b) 1.5

2.46 $E(X)$ exist, $Var(X)$ does not exist

2.48 b) $P(0 \leq X \leq 40) \geq \frac{19}{20}$

2.49 a) $\frac{16}{25}$ b) $k = 10$

2.50 b) $E(Y) \simeq 1.82$, $Var(Y) \simeq 0.003489$

2.51 $\varphi_X(t) = \frac{2}{3}e^{it} + \frac{3}{5}$

2.52 $\varphi_X(t) = \frac{1}{(b-a)it}\left(e^{itb} - e^{ita}\right)$ if $t \neq 0$, $\varphi_X(t) = 1$ if $t = 0$

2.53 $\varphi_X(t) = 1$ if $t = 0$, $\varphi_X(t) = \frac{2-2\cos t}{t^2}$ if $t \neq 0$

2.54 a) Lower quartile =2; upper quartile =3; median lies in [2,3] c) $E(X)$ does not exist, median = 5

2.55 a) Mode 1 b) $r = 7, \lambda = \frac{1}{4}$

SOLUTIONS FOR CHAPTER 3

3.1 $\binom{n}{\frac{n}{2}}\left(\frac{1}{2}\right)^n$

3.2 0.608; 2

3.3 0.2248

3.4 0.87842; 0.12158

3.9 $E(X) = Var(X) = 3.9$

3.10 0.98752

3.11 0.4335

3.12 3.8812×10^{-4}

3.13 $X \stackrel{d}{=} \mathcal{G}\left(\frac{9}{19}\right)$

3.15 a) 3.8305×10^{-8} b) 4.4179×10^{-2} c) 1.2379×10^{-3}

3.16 e^{-2}

3.17 5

3.18 15,000

3.20 a) 2.861×10^{-5} b) 5 c) 4 d) 9.8877×10^{-2}

3.21 0.9945

3.22 $P(X^2 = j) = pq^{\sqrt{j}-1}$, $j = 1, 4, 9, \cdots$; $P(X + 3 = j) = pq^{j-4}$, $j = 4, 5, 6, \cdots$

3.23 a) 0.674% b) 73.5%

3.24 a) 5.06×10^{-4} b) It is not correct

3.25 0.0671

3.26 0.96801

3.27 76

3.28 0.43; 0.57

3.29 816; 169; 15

3.30 0.78343

3.31 2.07×10^{-5}

3.33 20; 120

3.34 $E(X) = -0.27154$

3.35 13

3.36 a) 2 b) Does not exist

3.37 240

3.38 $\lambda = k$

3.43 $p = \frac{k}{n}$

3.47 $-\frac{p \ln p}{1-p}$

3.49 $\varphi_X(t) = \left(q + pe^{it}\right)^n$

3.50 $\varphi_X(t) = \exp(\lambda(e^{it} - 1))$

3.51 6

SOLUTIONS FOR CHAPTER 4

4.1 a) $\frac{1}{2}$; $\frac{1}{3}$; $\frac{\sqrt{3}}{3}$ b) $x = 1.3$

4.2 a) 0.1 b) 0.1 c) 0.1

4.4 $\frac{3}{4}$; $\frac{1}{2}$

4.10 0.6830

4.11 a) $\frac{1}{6}$ b) $f_Y(y) = \frac{1}{2}\mathcal{X}_{(0,2)}(y)$

4.12 a) $f_Y(y) = \exp(y)\mathcal{X}_{(-\infty,0)}(y)$ b) $f_Z(z) = \frac{1}{3}(z-2)^{-\frac{2}{3}}\mathcal{X}_{(2,3)}(z)$ c) $f_W(w) = \frac{1}{w^2}\mathcal{X}_{(1,\infty)}(w)$

4.13 5.2

4.14 a) 0.7265 b) 0.1337 c) 0.98163 d) 0.07636 e) 0.24173

4.15 a) 3 b) 5.8896 c) 6.3556

4.16 a) 18.307 b) 0.68268 c) 6.6257 d) 44.3141 e) 0.072; 0.989; 11.808

4.17 a) 92 b) 57

4.18 89.97%; 5.82%; 84.2%

4.19 Normal distribution with mean $5\mu - 1$ and variance $25\sigma^2$

4.21 No, calculate, for example, $P(4 \leq X \leq 9)$

4.22 0.17619

4.23 $\alpha = 3.87$

4.27 $54,000$

4.29 0.28347

4.30 0.60653

4.32 a) e^{-3} b) e^{-1}

4.34 a) $f_Y(y) = \lambda \exp(y - \lambda \exp(y)), \ y \in \mathbb{R}$

4.35 a) $F(t) = \left[1 - e^{-\frac{\alpha}{\ln \beta}(\beta^t - 1)}\right] \mathcal{X}_{[0,\infty)}(t)$
b) $F(t) = \left[1 - e^{-\gamma t - \frac{\alpha}{\ln \beta}(\beta^t - 1)}\right] \mathcal{X}_{[0,\infty)}(t)$

4.38 $F(t) = \left(1 - \frac{1}{1+t}\right) \mathcal{X}_{[0,\infty)}(t)$

4.39 $F(t) = \left[1 - \exp\left(-\frac{\alpha}{\beta+1}t^{\beta+1}\right)\right] \mathcal{X}_{[0,\infty)}(t)$

4.41 0.275

4.46 $\sqrt{\frac{2}{\pi}}$

4.48 g is the inverse of the standard normal probability density function

4.53 Standard Cauchy distribution

SOLUTIONS FOR CHAPTER 5

5.9 a) 0.3 b) 0.9 d) $P(Z = -2) = 0.1, P(Z = -1) = 0.2, P(Z = 0) = 0.3, P(Z = 1) = 0.4$

5.11 $\alpha = 0.01,\ \beta = 0.3,\ \gamma = 0.1,\ \eta = 0.03,\ \kappa = 00.04,\ \delta = 0.02$

5.12 a) $P(X = 0, Y = 0) = \frac{1}{55}$, $P(X = 0, Y = 1) = \frac{2}{11}$, $P(X = 0, Y = 2) = \frac{2}{11}$, $P(X = 1, Y = 0) = \frac{8}{55}$, $P(X = 1, Y = 1) = \frac{4}{11}$, $P(X = 2, Y = 0) = \frac{6}{55}$ b) $E(X) = \frac{8}{11}$, $E(Y) = \frac{10}{11}$

5.13 b) $E(X) = \frac{88}{121}$, $E(Y) = \frac{110}{121}$

5.14 a) 6 b) 6

5.16 a) 3.7308×10^{-2} b) 9

5.17 $a = b = k = h = \frac{1}{18}$, $d = \frac{4}{5}$, $e = f = \frac{4}{45}$; $E(XY) = 0$

5.18 a) $\mathcal{P}(9.4)$ b) 0.29

5.19 $P(Z = -1, W = 1) = \frac{1}{16}$, $P(Z = 0, W = 0) = \frac{1}{4}$, $P(Z = 0, W = 2) = \frac{3}{16}$, $P(Z = 1, W = 1) = \frac{1}{8}$, $P(Z = 1, W = 3) = \frac{1}{16}$, $P(Z = 2, W = 0) = \frac{3}{16}$, $P(Z = 2, W = 2) = \frac{1}{16}$, $P(Z = 3, W = 1) = \frac{1}{16}$

5.20 $\frac{41}{324}$; 0.1251

5.22 -1

5.23 0.35185

5.25 No

5.27 $\rho = \frac{-\sigma_2^2}{\sqrt{\sigma_2^2+\sigma_3^2}\sqrt{\sigma_1^2+\sigma_2^2}}$

5.28 $9 - 2\sqrt{2}$

5.32 $E(X) = m\left(\frac{n+1}{2}\right)$

5.33 $E(X) = \binom{r}{k}\left(\frac{1}{365}\right)^{k-1}\left(\frac{364}{365}\right)^{r-k}$

5.34 $E(X) = 365 - \left(\frac{1}{365}\right)^{-1}\left(\frac{364}{365}\right)^{r} - r\left(\frac{364}{365}\right)^{r-1}$

5.41 Yes, $f(x,y) = \frac{e^{-x}}{\pi(1+y^2)},\ x \geq 0, y \in \mathbb{R}$

5.42 a) $\frac{1}{2}$ b) $f_X(x) = f_Y(x) = \frac{1}{2}(\cos x + \sin x),\ x \in (0, \frac{\pi}{2})$

5.43 b) $\frac{1}{25}$

5.44 b) 0.25 c) 0.30556

5.46 0.5

5.47 $\frac{1}{8}$

5.48 a) $\frac{3}{4}$ b) $\frac{2}{3}$

5.52 a) $f_Z(z) = 2\left[z - \frac{1}{\lambda}\left(1 - e^{-\lambda z}\right)\right]\mathcal{X}_{(0,1)}(z) + 2\lambda e^{-\lambda z}\left[\frac{1}{\lambda}e^{\lambda} - \frac{1}{\lambda^2}(e^{\lambda} - 1)\right]\mathcal{X}_{[1,\infty)}(z)$ b) $\frac{1}{\lambda^2}$

5.55 a) $f_Z(z) = \lambda\exp(-\lambda \mid z \mid)\mathcal{X}_{(0,\infty)}(z)$ b) $f_W(w) = \lambda\left(1 + \frac{1}{3\sqrt[3]{w^2}}\right)\exp(-\lambda\sqrt[3]{w} - \lambda w)\mathcal{X}_{(0,\infty)}(w)$

5.56 b) $\frac{4}{9}$

5.57 Cauchy

5.59 $f(y_1, y_2) = \mathcal{X}_{(0,1)}(y_1)\chi_{(0,1)}(y_2)$

5.61 $f_U(x) = \frac{3}{2}(3x)^2\exp(-3x)\mathcal{X}_{(0,\infty)}(x)$

5.62 $f_Z(x) = \frac{3}{4}\left(1 - \frac{x^2}{4}\right)\mathcal{X}_{(0,2)}(x)$

5.63 a) 27.7 b) 0.32 c) 4.26

5.64 F_n^1

5.71 $\mathcal{X}_n^2$

5.72 $\mathcal{X}_{n-1}^2$

5.73 0.5464

5.74 $t_{(n-1)}$

SOLUTIONS FOR CHAPTER 6

6.2 a) $f(x,y) = \begin{cases} \frac{1}{x} & \text{if } 0 < y < x < 1 \\ 0 & \text{other cases} \end{cases}$ b) $f_Y(y) = -(\ln x)\chi_{(0,1)}(y)$

6.3 $E(Z) = p\lambda = Var\,(Z)$

6.4 a) $\frac{2}{3}$ b) $\frac{1}{4}$ c) $\frac{1}{2}$

6.5 $f_{Y|X}\,(y|x) = \begin{cases} xe^{-xy} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{other cases} \end{cases}$

$f_{X|Y}\,(x|y) = \begin{cases} (y+1)^2 xe^{-x(y+1)} & \text{if } x > 0 \text{ and } y > 0 \\ 0 & \text{other cases} \end{cases}$

6.6 a) $X \stackrel{d}{=} \mathcal{N}\,(3,1)$; $Y \stackrel{d}{=} \mathcal{N}\,(3,1)$ b) $f_{Y|X}(y|2) = \frac{1}{2\pi \times (0.6)} \exp\left\{-\frac{1}{2\times 0.36}(y-2.2)^2\right\}$ c) $c = 3.187$

6.7 b) $E(X \mid Y = 1) = 1$; $E(Y \mid X = 1) = \frac{5}{2}$

6.9 $\frac{5}{6}$

6.11 1.909

6.12 a) $E\,(X \mid Y = 0) = -\frac{1}{8}$; $E\,(X \mid Y = 1) = \frac{5}{2}$ b) $E(X) = \frac{4}{3}$

6.15 a) $p_{X,Y}(x,y) = \binom{x}{y}\left(\frac{1}{2}\right)^x \frac{x}{3}$ for $y = 0, \cdots, x$ and $x = 1, 2$ b) $E\,(X \mid Y = 0) = \frac{3}{2}$; $E\,(X \mid Y = 1) = \frac{5}{3}$; $E\,(X \mid Y = 2) = 2$

6.16 $1 + p$

6.17 a) 8.8492×10^{-2} b) $\frac{y}{2}$ for $y > 0$ c) $(x+1)$ for $x > 0$

6.20 1.3137

6.22 $Var\,(Y \mid X = 0) = \frac{2}{3}$; $Var\,(Y \mid X = 1) = \frac{1}{4}$; $Var\,(Y \mid X = 2) = 0$

6.23 $\frac{e}{4}$

6.25 $E(X|Y = y) = \frac{3+2y}{4+3y}$

6.26 $E\,(Y \mid X = x) = \frac{2-x}{2}$

6.27 a) $E(X \mid Y = y) = \frac{2(1+y+y^2)}{3(1+y)}$ b) $E(X^2 \mid Y = y) = \frac{1}{2}(1+y^2)$ c) $Var(X \mid Y = y) = \frac{9(1+y^2)(1+y)^2 - 8(1+y+y^2)^2}{18(1+y)^2}$

6.28 7

6.29 a) $\frac{4}{7}$ b) $\frac{26}{21}$

6.30 2

6.31 a) 6 b) 7 c) 5.8192

6.32 $\frac{12}{5}$; $\frac{9}{5}$; $\frac{6}{5}$; $\frac{3}{5}$; 0

SOLUTIONS FOR CHAPTER 7

7.2 $\mathbf{Y} \stackrel{d}{=} \mathcal{N}(\mu, \mathbf{\Sigma})$ where $\mu = (-2, 1)$ and $\mathbf{\Sigma} = \begin{pmatrix} 10 & -5 \\ -5 & 5 \end{pmatrix}$

7.3 $\alpha = -2$

7.6 a) $\mathbf{Y} \stackrel{d}{=} \mathcal{N}(\mu, \mathbf{\Sigma})$ where $\mu = (0, 0)$ and $\mathbf{\Sigma} = \begin{pmatrix} 1 & -1 \\ -1 & 9 \end{pmatrix}$

b) $f_Y(y_1, y_2) = \frac{1}{2\pi\sqrt{8}} \exp\left\{-\frac{1}{2}\left(\frac{6y_1^2}{8} + \frac{2y_1y_2}{8} + \frac{y_2^2}{8}\right)\right\}$

7.8 $P(X > 0, Y > 0, Z > 0) = \frac{1}{8} + \frac{1}{4\pi}\left\{\sin^{-1}\rho_1 + \sin^{-1}\rho_2 + \sin^{-1}\rho_3\right\}$

7.9 $f(x, y, z) = \frac{1}{2\pi\sqrt{230\pi}} \exp\left\{-\frac{1}{230}\left(39x^2 + 36y^2 + 26z^2 - 44xy + 36xz - 38yz\right)\right\}$

7.12 $\mathcal{N}(0, 3)$

7.13 $\mathcal{N}(2, 1)$

7.15 $\mathcal{N}(0, 1)$

SOLUTIONS FOR CHAPTER 8

8.6 $\frac{3}{4}$

8.10 $E(X^3) \geq 8$; $E(\ln X) \leq \ln 2$

8.15 0.8475

8.16 0.99324

8.17 0.9922

8.18 0.003

8.19 66564

8.20 4096

SOLUTIONS FOR CHAPTER 9

9.1 a) $Cov(X_s, X_t) = E\left(\frac{A^2}{2}\cos\eta(t-s)\right)$ b) Yes

9.2 Yes

9.7 a) 0 - absorbing, {1,2} - positive recurrent, {3,4} - transient b) $(\pi_1 \pi_2) = (\frac{2}{5} \frac{3}{5})$ d) $\frac{9}{16}$

9.8 a) 0.212 b) 0.0032

9.9 a) $\{S_1, S_2\}$ - Transient, class of positive recurrent $\{S_3, S_4, S_5\}$, $\lambda(S_3) = \lambda(S_4) = \lambda(S_5) = 1$ b) $(0, 0, \frac{10}{45}, \frac{20}{45}, \frac{15}{45})$

9.10 a) $f_{j0}^{(n)} = p^{n-j} q^j$ for $j \leq n$ b) $\frac{q(p^n - q^n)}{(p-q)(10pq)} + \frac{nq^{n+1}}{p(1-pq)}$

9.11 a) $C(0) = \{0, 1, 2\}$ is a class with positive recurrent states b) $P^n = \begin{pmatrix} \frac{1}{2} & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$ if n is even, $P^n = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix}$ if n is odd c) Does not exist

9.12 a) $C(0) = \{1\}$ - transient, $C(1) = \{0, 2\}$ - class of positive recurrents, $C(2) = \{3, 4\}$ - class of positive recurrents b) $(\frac{1}{2}, 0, \frac{1}{2}, 0, 0)$; $(0, 0, 0, \frac{1}{2}, \frac{1}{2})$

9.13 a) All states are positive recurrent, aperiodic b) $(\frac{12}{37}, \frac{6}{37}, \frac{4}{37}, \frac{3}{37}, \frac{12}{37})$

9.14 a) $P = \begin{pmatrix} (1-\alpha)^2 & 2\alpha(1-\alpha) & \alpha^2 \\ \beta(1-\alpha) & (1-\alpha)(1-\beta) + \alpha\beta & \alpha(1-\beta) \\ 0 & \beta & 1-\beta \end{pmatrix}$ b) $(1-\alpha)^2[(1-\alpha)^2 + 2\alpha\beta]$

9.16 b) $\pi_i = \frac{1-2p}{1-p}\left(\frac{p}{1-p}\right)^i$, $i = 0, 1, \cdots$

9.20 c) $(\frac{7}{24}, \frac{3}{24}, \frac{5}{24}, \frac{9}{24})$

9.21 a) $e^{-20\lambda} \frac{(20\lambda)^2}{2!}$ b) $1 - e^{-20\lambda}$

9.25 a) $Q = \begin{pmatrix} -2 & 2 & 0 \\ 1 & -3 & 2 \\ 0 & 2 & -2 \end{pmatrix}$ c) $(\frac{1}{5}, \frac{2}{5}, \frac{2}{5})$

9.26 a) 0.909 b) $\frac{3}{4}$ c) $\frac{1}{160}$

9.28 a) $\binom{n}{k}(1 - e^{-\lambda t})^k e^{-(n-k)\lambda t}$, $k = 0, 1, \cdots, n$ b) $ne^{-\lambda t}$

9.31 a) e^{-5} b) 20 c) 50

9.33 b) $\alpha_1, \alpha_2 = \frac{-(3\lambda+\mu) \pm \sqrt{(3\lambda+\mu)^2 - 8\lambda^2}}{2}$; reliability $= 1 - \frac{2\lambda^2}{\alpha_1 \alpha_2} - \frac{2\lambda^2}{\alpha_1(\alpha_1 - \alpha_2)} e^{\alpha_1 t} - \frac{2\lambda^2}{\alpha_2(\alpha_2 - \alpha_1)} e^{\alpha_2 t}$

SOLUTIONS FOR CHAPTER 10

10.1 4.047

10.2 a) 0.0223 b) 0.1705

10.3 4 minutes

10.4 a) 2.81 b) 2.5371

10.5 $P_n = \frac{\rho^n/n!}{\sum_{i=0}^{c} \rho^i/i!}, \quad n = 0, 1, \cdots, c$ where $\rho = \frac{\lambda}{\mu}$

10.6 a) 6.06 cars b) 12.3 minutes

10.8 $\frac{8}{\mu}$

10.9 b) $\pi_n = \frac{\rho^n/n!}{\sum_{i=0}^{c} \rho^i/i!}, \quad n = 0, 1, \cdots, c$ where $\rho = \frac{\lambda}{\mu}$

SOLUTIONS FOR CHAPTER 11

11.13 Yes

11.16 Yes

11.17 $3(t-s)^2$

11.19 $\mathcal{N}(0, 30)$

11.21 $\mathcal{N}(0, \frac{t^3}{3})$

11.23 a) $dX_t = 3B_t^2 dB_t + \frac{1}{2} B_t dt$ b) $dY_t = B_t dt + t dB_t$ c) $dZ_t = \left(c + \frac{1}{2}\alpha^2\right) Z_t dt + \alpha Z_t dB_t$

11.24 $E(I_t(B)) = 0$ and $E(I_t(B)^2) = \frac{1}{2}\left(e^4 - 1\right)$

11.25 $\arctan B_t + \int_0^t \frac{B_s}{(1+B_s^2)^2} ds$

11.26 $e^{B_t - \frac{1}{2}t} - 1$

11.28 $Y_t = Y_0 + 4t^2 + \int_0^t B_s ds + \int_0^t (2s+3) dB_s$

11.29 a) $X_t = -\frac{1}{8}t + \frac{1}{2}B_t$, b) $dZ_t = \frac{1}{2} Z_t dB_t$

11.30 $Z_t = Z_0 e^{-t+2B_t}$

11.31 $r_t = r_0 e^{-t} + b\left(1 - e^{-t}\right) + \sigma \int_0^t e^{-(t-s)} dB_s$

SOLUTIONS FOR CHAPTER 12

12.1 $r = 8\%$

12.3 0.8667 and \$1.64

12.4 $p^* = 0.541$ and $C_0 = 7.44$

12.5 $P_0 = 4.92$

12.6 $p^* = 0.6523$ and $C_0 = 1.266$

12.7 a) $V_0 = 142.15$ b) $\alpha_1 = 1.09$ and $\beta_1 = 36.37$

12.11 a) 2.1183 b) 2.5372

12.13 0.5373

12.14 a) Call option value is \$7.5133, $\Delta = 0.7747$ and $\Gamma = 0.0320$
b) Put option value is \$1.3, $\Delta = -0.2253$ and $\Gamma = 0.0320$

12.17 0.2509

REFERENCES

1. Ash, R. (1972), *Real Analysis and Probability*, Academic Press, New York.
2. Bauer, H. (1991), *Wahrscheinlichkeitstheorie*, Walter de Gruyter, Berlin.
3. Bhat, U.N. (2008), *An introduction to Queueing Theory: Modeling and Analysis in Applications*, Springer, Boston.
4. Bhat and Miller (2002), *Elements of Applied Stochastic Processes*, Wiley, 3rd edition, New Jersey.
5. Bingham, N.H. and R. Kiesel (2004), *Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives*, Springer, New York.
6. Black, F. and M. Scholes (1973), The pricing of options and corporate liabilities, Journal of Political Economy, **81**, 637–654.
7. Bolch, G., S. Greiner, H. Meer and K.S. Trivedi (2006), *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications*, John Wiley & Sons, 2nd edition, New Jersey.
8. Brieman, J. (1992), *Probability*, SIAM, Philadelphia, PA.
9. Cinlar, E. (1975), *Introduction to Stochastic Processes*, Prentice Hall, Englewood Cliffs, New Jersey.
10. Cox, J.C., S.A. Ross and M. Robinstein (1979), Option pricing: A simplified approach, *Journal of Financial Economics*, **7**,229–263.
11. Feller, W. (1968), *An Introduction to Probability Theory and Its Applications*, Volume I, John Wiley & Sons, 3rd edition, New York.

By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

12. Gorostiza, L.G. (2001), *La probabilidad del siglo XX*, Miscelánea Matemática, No. 33, Sociedad Matemática Mexicana, México.
13. Grimmett, G. and D. Stirzaker (2001), *Probability and Random Processes*, Oxford University Press, New York.
14. Gross, D. and C.M. Harris (1998), *Fundamentals of Queueing Theory*, Wiley, 2nd edition, New York.
15. Harrison, J.M. and D.M.Kreps (1979), Martingales and arbitrage in multiperiod securities markets, Journal of Economic Theory,**20** , 381-408.
16. Harrison, J.M. and D.M.Kreps (1981) Martingales and stochastic integrals in the theory of continuous trading, stochastic processes and their applications, **11**, 215-260
17. Hernández, A. and O. Hernández (2003), *Elementos de Probabilidad y Estadística,* Textos 21, Sociedad Matemática Mexicana, Mexico.
18. Hernández, F.M. (2003), *Cálculo de Probabilidades,* Textos 25, Nivel Elemental, Sociedad Matemática Mexicana, México.
19. Hoel, P., S.C. Port and C.J. Stone (1971) *Introduction to Probability Theory*, Houghton Mifflin Company, Boston, MA.
20. Hoel, P., S.C. Port and C.J. Stone (1972), *Introduction to Stochastic Processes*, Waveland Press, Prospect Heights, IL.
21. Hull, J.C. (2009), *Options, Futures, and Other Derivatives*, Prentice Hall, seventh edition, New Jersey.
22. Jacod, J. and P. Protter (2004), *Probability Essentials*, Springer-Verlag, 2nd edition, New York.
23. Kahane, J.A. (1997), A century of interplay between Taylor series, Fourier series and Brownian motion, *Bulletin of the London Mathematical Society*, **29**, 257–279.
24. Karatzas, I. and S. Shreve (1991), *Brownian Motion and Stochastic Calculus*, Springer, New York.
25. Karlin, S. and H.M. Taylor (1975), *A First Course in Stochastic Processes*, Academic Press, 2nd edition, New York.
26. Korn, R. and E. Korn (2000), *Option Pricing and Portafolio Optimization*, American Mathematical Society Providence, RI.
27. Krengel, U. (2000), *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, Vieweg Verlag, Braunschweig.
28. Lamberton, D. and B. Lapeyre (1996), *Introduction to Stochastic Calculus Applied to Finance*, CRC Press, London,
29. Lawler, G.F. (2006), *Introduction to Stochastic Processes*, Chapman & Hall, Boca Raton, FL.
30. Little, J. (1961) A proof for the queueing formula: $L = \lambda W$, *Operations Research*, **9**, 383–387.
31. Medhi, J. (1994), *Stochastic Processes*, New Age International, New Delhi.
32. Medhi, J. (2002), *Stochastic Models in Queueing Theory*, Academic Press, 2nd edition, San Diego, CA.

33. Meyer, P.L. (1970), *Introductory probability and statistical applications*, Addison Wesley, Boston, MA.

34. Mikosch, T. (1998), *Elementary Stochastic Calculus*, World Scientific, Singapore.

35. Muñoz, J.M. (2002), *Introducción a la teoría de conjuntos*, Facultad de Ciencias, Universidad Nacional de Colombia, Bogotá, Colombia.

36. Muñoz, M. and L. Blanco (2002), *Introducción a la teoría avanzada de la probabilidad*, Colección Textos, Unibiblos, Universidad Nacional de Colombia, Bogotá, Colombia.

37. Nelson, E. (1967), *Dinamical Theories of Brownian Motions*, Princeton University Press, Princeton, NJ.

38. Oksendal, B. (2006), *Stochastic Differential Equations: An Introduction with Applications*, Springer, 6th edition, Berlin, Heidelberg.

39. Ospina, D. (2001), *Introducción al muestreo*, Facultad de Ciencias, Universidad Nacional de Colombia. Bogotá, Colombia.

40. Rao, C.R. (1973), *Linear Statistical Inference and Its Applications*, John Wiley & Sons, New York.

41. Resnick, S.I. (1994), *Adventures in Stochastic Processes*, Birkhuser, Boston, MA.

42. Rincón, L. (2011), *Introducción a los procesos estocásticos*, Facultad de Ciencias, UNAM, México.

43. Ross, S.M. (1996), *Stochastic Processes*, John Wiley & Sons, New York.

44. Ross, S.M. (1998), *A First Course in Probability*, Prentice Hall, New Jersey.

45. Ross, S.M. (2007), *Introduction to Probability Models*, Academic Press, San Diego, CA.

46. Royden, H.L. (1968), *Real Analysis*, Macmillan, New York.

47. Schuh, H.J. (1986–1987) *Stochastik I und II*, Vorlesungsskript, Johannes Gutenberg-Universität Mainz.

48. Searle, S.R. (1982), *Matrix Algebra Useful for Statistics*, John Wiley & Sons, New York.

49. Shreve, S.E. (2004), *Stochastic Calculus for Finance II: Continuous-Time Models*, Springer, New York.

50. Wackerly, D.D., W. Mendenhall and R. L. Scheaffer (2008), *Mathematical Statistics with Applications*, Cengage Learning, Belmont, CA.

51. Williams, R.J. (2006), *Introduction to the Mathematics of Finance*, American Mathematical Society, Providence, RI.

52. Willmott. P., S. Howison and J. Dewynne. (1995), *The Mathematics of Financial Derivatives: A Student Introduction*, Cambridge University Press, New York.

53. Wolff, R.W. (1982), Poisson arrivals see time averages, *Operations Research*, **30**, 223–231.

GLOSSARY

$\|\alpha\|$	Absolute value of the number α.
$a \approx b$	a is approximately equal to b.
$A \setminus B$	Difference of the sets A and B.
$A \cup B$	Union of the sets A and B.
$A \cap B$	Intersection of the sets A and B.
A^c	Complement of the set A.
$\|A\|$	Number of elements of the set A.
A^T	Transpose of the matrix A.
$(a_{ij})_{m \times n}$	Matrix of size $m \times n$ whose element in the ith row and jth column is a_{ij}.
$\mathcal{B}$	Borel σ-algebra over $\mathbb{R}$.
$\mathcal{B}_n$	Borel σ-algebra over $\mathbb{R}^n$.
$\binom{n}{r}$	n choose r (or combinatory n, r).
$\mathbb{C}$	Set of complex numbers.
EX	Expected value of the random variable X.
$E(X \mid Y)$	Expected value of the random variable X given the random variable Y.
$E(X \mid B)$	Expected value of the random variable X given the event B.
$E(X \mid \mathcal{G})$	Expected value of the random variable X given the σ-algebra $\mathcal{G}$.
$\det(A)$	Determinant of the matrix A.
$\frac{d^r}{dx^r} f(x)$	Derivative of order r of the function f.

By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja

$f(\cdot)$	Function of a real variable.
$f_{X\mid Y}(\cdot \mid y)$	Density function of the random variable X given $Y = y$.
$F_{X\mid Y}(\cdot \mid y)$	Conditional distribution function of the random variable X given $Y = y$.
$f(x^+)$	Right-hand limit of $f(x)$.
$f(x^-)$	Left-hand limit of $f(x)$.
$\phi_X(\cdot)$	Characteristic function of the random variable X.
$\Phi(\cdot)$	Standard normal distribution function.
$\phi(\cdot)$	Standard normal density function.
$\mathcal{X}_A(\cdot)$	Characteristic function of the set A.
$\ln(x)$	Natural logarithm of x.
$m_X(\cdot)$	Moment generating function of the random variable X.
μ_r'	rth central moment around *zero*.
μ_r	rth central moment around the mean.
$\mathbb{N}$	Set of natural numbers: $\{0, 1, 2, \cdots\}$.
$P(A)$	Probability of the event A.
$P(A \mid B)$	Conditional probability of the event A given the event B.
$\langle \mathbf{x}, \mathbf{y} \rangle$	Inner product of the vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
$\mathbb{Q}$	Set of rational numbers.
$\mathbb{R}$	Set of real numbers.
$rf(A)$	Relative frequency of the event A.
σ_X^2	Variance of the random variable X.
σ_X	Standard deviation of the random variable X.
$\sigma(L)$	Smallest σ-algebra containing the collection L.
$[t]$	integer part, or floor, of t.
$tr(A)$	Trace of the matrix A.
$\lvert X \rvert$	Absolute value of the random variable X.
$(x_n)_{n\in\mathbb{N}}$	Sequence of real numbers.
$X \stackrel{d}{=} Y$	Indicates that the random variables X and Y both have the same distribution.
$X \stackrel{d}{=} \mathcal{B}(n, p)$	Indicates that the random variable X has a binomial distribution with parameters n and p.
$X \stackrel{d}{=} Hg(n, R, N)$	Indicates that the random variable X has a hypergeometric distribution with parameters n, R and N.
$X \stackrel{d}{=} \mathcal{P}(\lambda)$	Indicates that the random variable X has a Poisson distribution with parameter λ.
$X \stackrel{d}{=} \mathcal{B}_N(k, p)$	Indicates that the random variable X has a negative binomial distribution with parameters k and p.
$X \stackrel{d}{=} \mathcal{G}(p)$	Indicates that the random variable X has a geometric distribution with parameter p.
$X \stackrel{d}{=} \mathcal{U}[a, b]$	Indicates that the random variable X has a uniform distribution over the interval $[a, b]$.

$X \stackrel{d}{=} \mathcal{N}\left(\mu, \sigma^2\right)$	Indicates that the random variable X has a normal distribution with mean μ and variance σ^2.
$X \stackrel{d}{=} \Gamma(r, \lambda)$	Indicates that the random variable X has a gamma distribution with parameters r and λ.
$X \stackrel{d}{=} Exp(\lambda)$	Indicates that the random variable X has an exponential distribution with parameter λ.
$X \stackrel{d}{=} \mathcal{X}^2_{(k)}$	Indicates that the random variable X has a chi-squared distribution with k degrees of freedom.
$X \stackrel{d}{=} \beta(a, b)$	Indicates that the random variable X has a beta distribution with parameters a and b.
$X \stackrel{d}{=} Weibull(\alpha, \beta)$	Indicates that the random variable X has a Weibull distribution with parameters α and β.
$X \stackrel{d}{=} t_{(k)}$	Indicates that the random variable X has a t-Student distribution with k degrees of freedom.
$X \stackrel{d}{=} F_n^m$	Indicates that the random variable X has an F distribution with m degrees of freedom on the numerator and n degrees of freedom on the denominator.
$\mathbf{X} \stackrel{d}{=} \mathcal{N}(\mu, \Sigma)$	Indicates that the random vector X has a multivariate normal distribution with mean vector μ and variance-covariance matrix Σ.
$\lVert\mathbf{X}\rVert$	Euclidean norm of the vector $\mathbf{X}$.
$X_n \xrightarrow[n\to\infty]{a.s.} X$	Almost sure (or with probability 1) convergence of the sequence of random variables $(X_n)_n$ to the random variable X.
$X_n \xrightarrow[n\to\infty]{d} X$	Convergence in distribution of the sequence of random variables $(X_n)_n$ to the random variable X.
$X_n \xrightarrow[n\to\infty]{P} X$	Convergence in probability of the sequence of random variables $(X_n)_n$ to the random variable X.
$\mathbb{Z}$	Set of integer numbers.
$\mathbb{Z}_+$	Set of positive integer numbers $\{1, 2, 3, \cdots\}$.
$:=$	Replaces the symbol "=" in assignations or definitions.

Index

By Liliana Blanco Castañeda, Viswanathan Arunachalam and Selvamuthu Dharmaraja