WILEY SERIES IN PROBABILITY AND STATISTICS

INTRODUCTION TO PROBABILITY

MODELS AND APPLICATIONS

N. Balakrishnan, Markos V. Koutras, Konstadinos G. Politis

WILEY

INTRODUCTION TO PROBABILITY

WILEY SERIES IN PROBABILITY AND STATISTICS
Established by *Walter A. Shewhart and Samuel S. Wilks*

The ***Wiley Series in Probability and Statistics*** is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at
http://www.wiley.com/go/wsps

INTRODUCTION TO PROBABILITY

Models and Applications

N. Balakrishnan
McMaster University, Canada

Markos V. Koutras
University of Piraeus, Greece

Konstadinos G. Politis
University of Piraeus, Greece

WILEY

This edition first published 2020

Registered Office
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA

Editorial Office
111 River Street, Hoboken, NJ 07030, USA

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data

Names: Balakrishnan, N., 1956- author. | Koutras, Markos V., author. |
Politis, Konstadinos G., 1966- author.
Title: Introduction to probability : models and applications / N.
Balakrishnan (McMaster University, Canada), Markos V. Koutras (University
of Piraeus, Greece), Konstadinos G. Politis.
Description: Hoboken, NJ : Wiley, 2019. | Series: Wiley series in probability
and statistics | Includes bibliographical references and index. |
Identifiers: LCCN 2018060140 (print) | LCCN 2019000211 (ebook) | ISBN
9781118548714 (Adobe PDF) | ISBN 9781118548493 (ePub) | ISBN 9781118123348
(hardcover)
Subjects: LCSH: Probabilities–Textbooks.
Classification: LCC QA273 (ebook) | LCC QA273 .B254727 2019 (print) | DDC
519.2–dc23
LC record available at https://lccn.loc.gov/2018060140

Cover Design: Wiley
Cover Image: Photograph by Mike Jack/Moment/Getty Images

Set in 10/12.5pt NimbusRomNo9L by SPi Global, Chennai, India

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

To my mother, due to whom
I am who I am today

N. Balakrishnan

To my family

Markos V. Koutras

To my daughters

Konstadinos G. Politis

CONTENTS

PREFACE

Probability theory deals with phenomena whose outcome is affected by random events, and therefore they cannot be predicted with certainty. For example, the result of throwing a coin or a dice, the time of occurrence of a natural phenomenon or disaster (e.g. snowfall, earthquake, tsunami etc.) are some of the cases where “randomness” plays an important role and the use of probability theory is almost inevitable.

It is more than five centuries ago, when the Italians Luca Pacioli, Niccolo Tartaglia, Galileo Galilei and the French Pierre de Fermat and Blaise Pascal started setting the foundations of probability theory. Nowadays this area has been fully developed as an independent research area and offers valuable tools for almost all applied sciences. As a consequence, introductory concepts of Probability Theory are taught in the first years of most University and College programs.

This book is an introductory textbook in probability and can be used by majors in Mathematics, Statistics, Physics, Computer Science, Actuarial Science, Operations Research, Engineering etc. No prior knowledge of probability theory is required.

In most Universities and Colleges where an introductory Probability course, such as one that may be based on this textbook, is offered, it would normally follow a rigorous Calculus course. Consequently, the Probability course can make use of differential and integral calculus, and formal proofs for theorems and propositions may be presented to the students, thereof offering them a mathematically sound understanding of the field.

For this reason, we have taken a calculus-based approach in this textbook for teaching an introductory course on Probability. In doing so, we have also introduced some novelties hoping that these will be of benefit to both students and instructors.

In each chapter, we have included a section with a series of examples/problems for which the use of a computer is required. We demonstrate, through ample examples, how one can make effective use of computers for understanding probability concepts and carrying out various probability calculations. For these examples it is suggested to use a computer algebra software such as Mathematica, Maple, Derive, etc. Such programs provide excellent tools for creating graphs in an easy way as well as for performing mathematical operations such as derivation, summation, integration, etc; most importantly, one can handle symbols and variables without having to replace them with specific numerical values. In order to facilitate the reader, an example set of Mathematica commands is given each time (analogous commands can be assembled for the other programs mentioned above). These commands may be used to perform a specific task and then various similar tasks are requested in the form of exercises. No effort is made to present the most effective Mathematica program for tackling the suggested problem and no detailed description of the Mathematica syntax is provided; the interested reader is referred to the Mathematica Instruction Manual (Wolfram Research) to check the, virtually unlimited, commands available in this software (or alternative computer algebra software) and use them for creating several alternative instruction sets for the suggested exercises.

Moreover, a novel feature of the book is that, at the end of each chapter, we have included a section detailing a case study through which we demonstrate the usefulness of the results and concepts discussed in that chapter for a real-life problem; we also carry out the required computations through the use of Mathematica.

At the beginning of each chapter we provide a brief historical account of some pioneers in Probability who made exemplary contributions to the topic of discussion within that chapter. This is done so as to provide students with a sense of history and appreciation of the vital contributions made by some renowned probabilists. Apart from the books on the history of probability and statistics that can be found in the bibliography, we have used Wikipedia as a source for biographical details.

In most sections, the exercises have been classified into two groups, A and B. Group A exercises are usually routine extensions of the theory or involve simple calculations based on theoretical tools developed in the section and should be the vehicle for a self-control of the knowledge gained so far by the reader. Group B exercises are more advanced, require substantial critical thinking and quite often include fascinating applications of the corresponding theory.

In addition to regular exercises within each chapter, we have also provided a long list of True/False questions and another list of multiple choice questions. In our opinion, these will not only be useful for students to practice with (and assess their progress), but can also be helpful for instructors to give regular in-class quizzes.

Particular effort has been made to give the theoretical results in their simplest form, so that they can be understood easily by the reader. In an effort to offer the book user an additional means of understanding the concepts presented, intuitive approaches and illustrative graphical representations/figures are provided in several places.

The material of this book emerged from a similar book (*Introduction to Probability: Theory and Applications*, Stamoulis Publications) written by one of us (MVK) in Greek, which is being used as a textbook for many years in several Greek Universities. Of course, we have expanded and transformed this material to reach an international audience.

This is the first volume in a set of two for teaching probability theory. In this volume, we have detailed the basic rules and concepts of probability, combinatorial methods for probabilistic computations, discrete random variables, continuous random variables, and well-known discrete and continuous distributions. These form the core topics for an introduction to probability. More advanced topics such as joint distributions, measures of dependence, multivariate random variables, well-known multivariate discrete and continuous distributions, generating functions, Laws of Large Numbers and the Central Limit Theorem should come out as core topics for a second course on probability. The second volume of our set will expand on all these advanced topics and hence it can be used effectively as a textbook for a second course on probability; the form and structure of each chapter will be similar to those in the present volume.

We wish to thank our colleagues G. Psarrakos and V. Dermitzakis who read parts of the book and to our students who attended our classes and made several insightful remarks and suggestions through the years.

In a book of this size and content, it is inevitable that there are some typographical errors and mistakes (that have clearly escaped several pairs of eyes). If you do notice any of them, please inform us about them so that we can do suitable corrections in future editions of this book.

It is our sincere hope that instructors find this textbook to be easy-to-use for teaching an introductory course on probability, while the students find the book to be user-friendly with easy and logical explanations, plethora of examples, and numerous exercises (including computational ones) that they could practice with!

Finally, we would like to thank the Wiley production team for their help and patience during the preparation of this book!

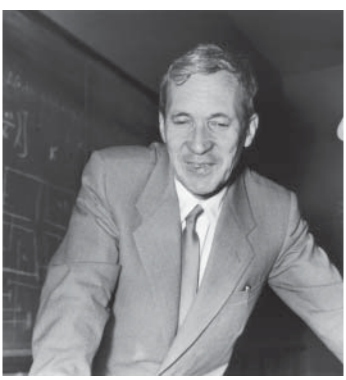
March, 2019

N. Balakrishnan
Markos V. Koutras
Konstadinos G. Politis

CHAPTER 1

THE CONCEPT OF PROBABILITY

Andrey Nikolaevich Kolmogorov
(Tambov, Russia 1903–Moscow 1987)
***Source*: Keystone-France /**
Getty Images.



Regarded as the founder of modern probability theory, Kolmogorov was a Soviet mathematician whose work was also influential in several other scientific areas, notably in topology, constructive logic, classical mechanics, mathematical ecology, and algorithmic information theory.

He earned his Doctor of Philosophy (PhD) degree from Moscow State University in 1929, and two years later, he was appointed a professor in that university. In his book, *Foundations of the Theory of Probability*, which was published in 1933 and which remains a classic text to this day, he built up probability theory from fundamental axioms in a rigorous manner, comparable to Euclid's axiomatic development of geometry.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

1.1 CHANCE EXPERIMENTS – SAMPLE SPACES

In this chapter, we present the main ideas and the theoretical background to understand what probability is and provide some illustrations of the way it is used to tackle problems in everyday life. It is rather difficult to try to answer the question “what is probability?” in a single sentence. However, from our experience and the use of this word in common language, we understand that it is a way to deal with uncertainty in our lives. In fact, probability theory has been referred to as “the science of uncertainty”; although intuitively most people associate probability with the degree of belief that something may happen, probability theory goes far beyond that as it attempts to formalize uncertainty in a way that is universally accepted and is also subject to rigorous mathematical treatment.

Since the idea of uncertainty is paramount when we discuss probability, we shall first introduce a concept that is broad enough to deal with uncertainties in a wide-ranging context when we consider practical applications. A **chance experiment** or a **random experiment** is any process which leads to an outcome that is not known beforehand. So tossing a coin, selecting a person at random and asking their age, or testing the lifetime of a new machine are all examples of random experiments.

Definition 1.1 A **sample space** Ω of a chance experiment is the set of all possible outcomes that may appear in a realization of this experiment. The elements of Ω are called sample points for this experiment. A subset of Ω is called an **event**.

An event $A = \{\omega\}$, consisting of a single sample point, i.e. a single outcome $\omega \in \Omega$, is called an **elementary event**. We use capital letters A, B, C, and so on to denote events.[1] If an event A consists of more than one outcome, then it is called a **compound event**.

The following simple examples illustrate the above concepts.

Example 1.1 Perhaps the simplest example of a chance experiment is tossing a coin. There are two possible outcomes – Heads (denoted by H) and Tails (denoted by T). In this notation, the sample space of the experiment is

$$\Omega = \{H, T\}.$$

If we toss two coins instead, there are four possible outcomes, represented by the pairs HH, HT, TH, TT. The sample space for this experiment is thus

$$\Omega = \{HH, HT, TH, TT\}.$$

Here, the symbol HH means that both coins land Heads, while HT means that the first coin lands Heads and the second lands Tails. Note in particular that we treat the two events HT and TH as distinguishable, rather than combining them into a single event. The main reason for this is that the events HT and TH are elementary events, while

[1] We shall avoid being too pedantic and, for typographical convenience, we shall often use the expression “the event ω” instead of “the event $\{\omega\}$” in the sequel.

the event "one coin lands Heads and the other lands Tails," which contains both HT and TH, is no longer an elementary event. As we will see later on, when we assign probabilities to the events of a sample space, it is much easier to work with elementary events, since in many cases such events are equally likely, and so it is reasonable the same probability to be assigned to each of them.

Consider now the experiment of tossing three coins. The sample space consists of triplets of the form HHT, HTH, TTH, and so on. Since for each coin toss there are two outcomes, for three coins the number of possible outcomes is $2^3 = 8$. More explicitly, the sample space for this experiment is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}. \tag{1.1}$$

Each of the eight elements of this set forms an elementary event. Note that for events which are not elementary, it is sometimes easier to express them in words, by describing a certain property shared by all elements of the event we consider, rather than by listing all its elements (which may become inconvenient if these elements are too many). For instance, let A be the event "exactly two Heads appear when we toss three coins." Then,

$$A = \{HHT, HTH, THH\}.$$

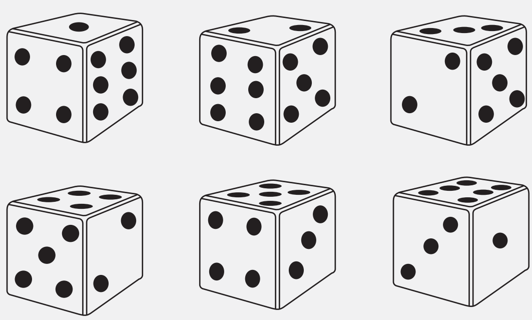
On the other hand, the event

$$B = \{HHH, TTT\}$$

could be described in words as "all three coin outcomes are the same."

Example 1.2 Another very simple experiment consists of throwing a single die. The die may land on any face with a number i on it, where i takes the values $1, 2, 3, 4, 5, 6$. Therefore, this experiment has sample space

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$



The elementary events are the sets

$$A_1 = \{1\}, \quad A_2 = \{2\}, \quad A_3 = \{3\}, \quad A_4 = \{4\}, \quad A_5 = \{5\}, \quad A_6 = \{6\}.$$

Any other event may again be described either by listing the sample points in it, such as

$$B = \{1, 3, 5, 6\}$$

or, in words, by expressing a certain property of its elements. For instance, the event

$$C = \{2, 4, 6\}$$

may be expressed as "the outcome of the die is an even integer."

For the experiments we considered in the above two examples, the number of sample points was finite in each case. For instance, in Example 1.2, the sample space has six elements, while in the experiment of throwing three coins there are eight sample points as given in (1.1). Such sample spaces which contain a finite number of elements (possible outcomes) are called **finite sample spaces**. It is obvious that any event, i.e. a subset of the sample space, in this case has also finitely many elements.

When dealing with finite sample spaces, the process of enumerating their elements, or the elements of events in such spaces, is often facilitated by the use of **tree diagrams**. Figure 1.1 depicts such a diagram which corresponds to the experiment of tossing three coins, as considered in Example 1.1.

From the "root" of the tree, two segments start, each representing an outcome (H and T, resp.) of the first coin toss. Thus, at the first stage, i.e. after the first throw, there are two nodes. From each of these, in turn, two further segments start corresponding to the two outcomes of the second toss. At the end of the second stage (after the second toss of the coin), there are four nodes. Finally, each of these is associated with two further

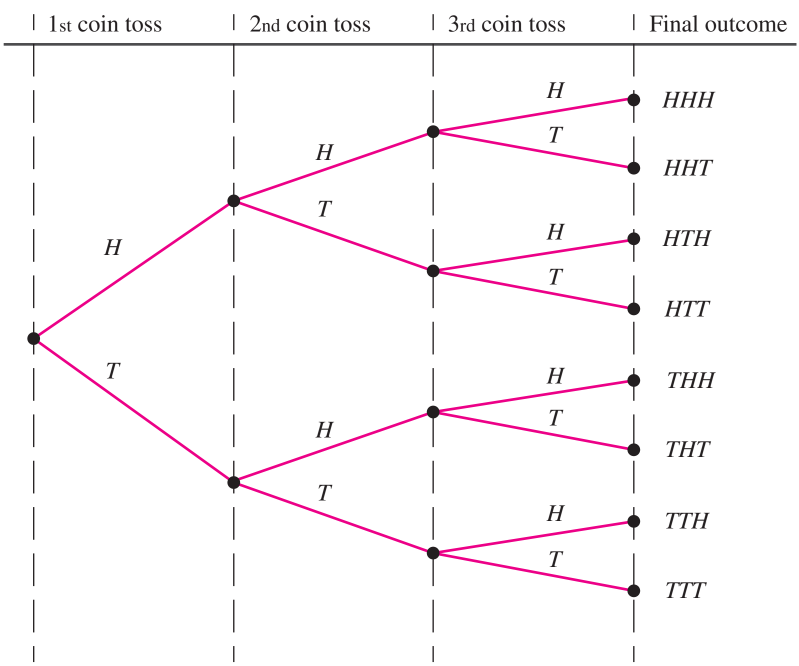


Figure 1.1 Tree diagram for the experiment of tossing three coins.

nodes, which are shown at the next level (end of the three coin tosses). The final column in Figure 1.1 shows the eight possible outcomes for this experiment, i.e. it contains all the elements of the sample space Ω. Each outcome can be traced by connecting the endpoint to the root and writing down the corresponding three-step tree route.

Example 1.3 (An unlimited sequence of coin tosses)
Let us consider the experiment of tossing a coin until "Tails" appear for the first time. In this case, our sample space consists of sequences like $T, HT, HHT, HHHT, \ldots$; that is, $\Omega = \{T, HT, HHT, HHHT, \ldots\}$. The event "Tails appear for the first time at the fifth trial" is then the elementary event

$$A = \{HHHHT\},$$

while the set

$$B = \{T, HT, HHT\}$$

has as its elements all outcomes where Tails appear in the first three tosses. So the event B can be described by saying "the experiment is terminated within the first three coin tosses." Finally, the event "there are at least four tosses until the experiment is terminated" corresponds to the set (event)

$$C = \{HHHT, HHHHT, HHHHHT, \ldots\}.$$

In the previous example, the sample space Ω has infinitely many points. In particular, and since these points can be enumerated, we speak of a **countably infinite sample space**. Examples of sets with countably[2] many points are the set of integers, the set of positive integers, the set of rationals, etc. When a sample space is countably infinite, the events of that sample space may have either finitely many elements (e.g. the event B in Example 1.3) or infinitely many elements (e.g. the event C in Example 1.3).

In contrast, a set whose points cannot be enumerated, is called an *uncountable set*; typical examples of such sets are intervals and unions of intervals on the real line. To illustrate this, we consider the following example.

Example 1.4 In order to monitor the quality of light bulbs that are produced by a manufacturing line, we select a bulb at random, plug it in and record the length of time (in hours) until it fails. In principle, this time length may take any nonnegative real value (however, this presupposes we can take an infinitely accurate measurement of time!). Therefore, the sample space for the experiment whose outcome is the life duration of the bulb is

$$\Omega = \{t : t \geq 0\} = [0, \infty).$$

The subset of Ω

$$A = \{t : 0 \leq t \leq 500\} = [0, 500]$$

---

[2]In mathematics, a countable set is a set which has the same cardinality (number of elements) as the set of natural numbers, $\mathbb{N} = \{1, 2, 3, \ldots\}$. This means that there exists a one-to-one correspondence between the elements of this set and the set $\mathbb{N}$.

describes the event "the life time of the light bulb does not exceed 500 hours," while the event "the light bulb works for at least 300 hours" corresponds to the set

$$B = \{t : t \geq 300\} = [300, \infty).$$

The sample space Ω in the last example is the half-line of nonnegative real numbers, which is an uncountable set. If Ω is uncountable, then it is usually referred to as a **continuous sample space**. Typically, the study of such sample spaces requires different treatment compared with sample spaces which are either finite or countably infinite. The latter two cases, however, present several similarities and in probability theory the techniques we use are very similar. As a consequence, there is a collective term for sample spaces which have either finitely many or a countably infinite number of elements, called **discrete sample spaces**.

At this point, it is worth noting the difference between an "ideal" continuous sample space and the one we use in practice. With reference to Example 1.4 regarding the lifetime of electric bulbs, such a lifetime does not, in practice, take values such as $\sqrt{12}$ or $(3 + \ln 20)/2$. Since time is measured in hours, it is customary to record a value rounded to the closest integer or, if more precision is required, keep two decimal places, say. In either case, and in contrast to the one used in the example above, the sample space is *countable*. Moreover, if we know that a lifetime of a bulb cannot exceed some (large) value a, the sample space becomes $\Omega = \{0, 0.01, 0.02, \ldots, a\}$ so that it is then in fact *finite*. However, the number of elements in that space is $100a + 1$, so that when a is a large integer, this can be very large. It is often the case that it is much simpler mathematically to assume that Ω is continuous although in practice we can only observe a finite (or infinitely countable) number of outcomes. This convention will be used frequently in the sequel, when we study, for example, weight, age, length, etc.

We conclude this section with an example which demonstrates that for the same experiment, we can define more than one sample space, depending on different aspects we might be interested in studying.

Example 1.5 (Different sample spaces for the same experiment)
Suppose that a store which sells cars has two salespersons. The store has in stock only two cars of a particular make. We are interested in the number of cars which will be sold by each of the two salespersons during next week. Then, a suitable sample space for this experiment is the set of pairs (i, j) for $i, j \in \{0, 1, 2\}$, where i stands for the number of cars sold by the first salesperson and j for the number of cars sold by the second one. However, since there are only two cars available for sale, it must also hold that $i + j \leq 2$, and we thus arrive at the following sample space

$$\Omega_1 = \{(0,0), (0,1), (0,2), (1,0), (1,1), (2,0)\}.$$

Notice that we treat again the pairs (i, j) and (j, i) as being distinguishable; if, however, the store owner is only interested in the `total number of cars` sold during next week, then we could use as a sample space the set

$$\Omega_2 = \{0, 1, 2\}.$$

In this case, an element $\omega \in \Omega_2$ denotes the total number of cars sold. It is worth noting that a specific event of interest is expressed in a completely different manner under these two different sample spaces. Consider, for example, the event

$$A: \text{ the number of cars that will be sold next week is 2.}$$

Viewed as a subset of Ω_1, the event A is a compound event which can be described as

$$A = \{(0, 2), (1, 1), (2, 0)\}.$$

However, when we consider the sample space Ω_2, the event A is an elementary event,

$$A = \{2\}.$$

EXERCISES

Group A

1. Provide suitable sample spaces for each of the following experiments. For each sample space, specify whether it is finite, infinitely countable or uncountable.
   (a) Two throws of a die
   (b) Gender of the children in a family that has three children
   (c) Random selection of a natural number less than 100
   (d) Random selection of a real number from the interval [0, 1]
   (e) Number of telephone calls that someone receives on a mobile phone during a day
   (f) Number of animals living in a certain forest area
   (g) Life duration of an electronic appliance
   (h) Change in the price of a stock during a day
   (i) Percentage change in the price of a stock during a day.
2. John throws a die and subsequently he tosses a coin.
   (i) Suggest a suitable sample space that describes the outcomes of this experiment.
   (ii) Let A be the event that "the outcome of the coin toss is Heads." Which elements of the sample space are included in the event A?
3. We toss a coin until either Heads appear for the first time or we have five tosses which all result in Tails. Give a suitable sample space for this experiment, and then write explicitly (i.e. by listing their elements) each of the following events:

   A: the experiment terminates at the third toss of the coin;

   B: the experiment terminates after the third toss of the coin;

   C: the experiment terminates before the fourth toss of the coin;

D: the experiment terminates with the appearance of H (Heads).

Which, if any, of the events A, B, C, D are elementary events in the sample space that you have considered?

4. A digital scale has an accuracy of two decimal places shown on its screen. Each time a person steps on the scale, we record his/her weight by rounding it to the closest integer (in kilograms). Thus, if the decimal part is 0.50 or greater, we round up to the next integer.

   (i) Give an appropriate sample space for the experiment whose outcome is the rounding error during the above procedure.

   (ii) Write explicitly each of the events

   A: the rounding error is at most 0.10;

   B: the absolute value of the rounding error is at least 0.20.

5. We throw a die twice. Give a suitable sample space for this experiment and then identify the elements each of the following events contains:

   A_1: the outcome of the first throw is 6;

   A_2: the outcome of the second throw is a multiple of 3;

   A_3: the outcome of the first throw is 6 and the outcome of the second throw is a multiple of 3;

   A_4: the sum of the two outcomes is 7;

   A_5: the sum of the two outcomes is at least 9;

   A_6: the two outcomes are identical.

6. A company salesman wants to visit the four cities a, b, c, d wherein his company has stores. If he plans to visit each city once, give a suitable sample space to describe the order in which he visits the cities. Then identify, by an appropriate listing of its elements, each of the following events:

   A_1: the salesman visits first city b;

   A_2: the salesman visits city b first and after that visits city d;

   A_3: the salesman visits city a before he visits city d;

   A_4: the salesman visits the cities a, b, c successively.

7. A box contains 3 red balls and 2 yellow balls. Give a suitable sample space to describe all possible outcomes for the experiment of selecting 4 balls at random, in each of the following schemes:

   (i) For each ball selected, we note its color and return it to the box so that it is available for the next selection (such a scheme is called *selection with replacement*).

   (ii) Every ball selected is subsequently removed from the box (which is called *selection without replacement*).

8. Irène has four books that she wants to put on a library shelf. Three of these books form a 3-volume set of a dictionary, so that they are marked as Volumes *I*, *II*, and *III*, respectively.

   (i) Find an appropriate sample space for all possible ways she can put the books on the shelf.

   (ii) Identify the elements of this sample space that each of the following three events contains:

   B_1: the three volumes of the dictionary are put next to one another;

   B_2: the three volumes of the dictionary are put in the right order (but not necessarily in adjacent places), so that Volume *I* is placed to the left of Volume *II*, which in turn is placed to the left of Volume *III*;

   B_3: the three volumes of the dictionary are placed next to one another and in the right order.

9. Mary has in her wallet three \$1 coins, one \$2 coin and four coins of 25 ¢. She selects four coins at random from her wallet.

   (i) Write down a sample space for the possible selections she can make.

   (ii) Express the following events as subsets of the above sample space:

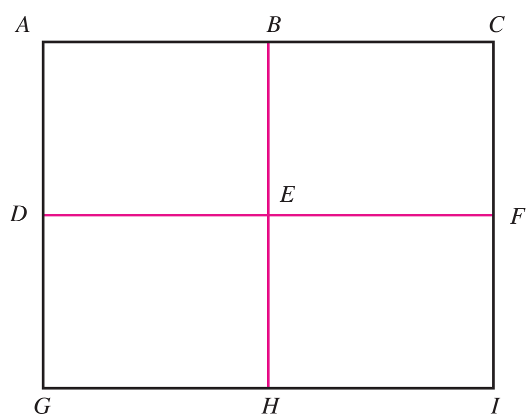
   C_1: exactly three 25 ¢ coins are selected;

   C_2: the total value of the coins selected is \$2.50;

   C_3: the total value of the coins selected is \$3.50.

Group B

10. Bill just visited a friend who lives in Place *A* of the graph below and he wants to return home, which is at Place *I* on the graph. In order to minimize the distance he has to walk, he moves either downwards (e.g. from Place *A* to Place *D*) or to the right (e.g. from Place *E* to Place *F* on the graph). At each time, he makes a choice for his next movement by tossing a coin.



(i) Give a sample space for the different routes Bill can follow to return home.

(ii) Write down explicitly the following events, representing them as subsets of the sample space given in (i):

A_1: he passes through Place E on his way back home;

A_2: he does not pass through Place D;

A_3: on his way back, he has to toss the coin only twice to decide on his next move.

11. A bus, which has a capacity of carrying 50 passengers, passes through a certain bus stop every day at some time point between 10:00 a.m. and 10:30 a.m. In order to study the time the bus arrives at this stop, as well as the number of passengers in the bus at the time of its arrival, we use the following sample space

$$\Omega = \{(k, t) : 0 \le k \le 50 \text{ and } 10 \le t \le 10.5\},$$

where k above denotes the number of passengers in the bus and t denotes the arrival time at the bus stop (in hours, expressed as a decimal number).

(i) Is the sample space Ω for this experiment countable or uncountable (continuous)?

(ii) Write down explicitly each of the events below:

A_1: the bus arrives at the stop at 10:10 a.m. carrying 25 passengers;

A_2: the bus arrives at the stop at 10:10 a.m. with less than 25 passengers;

A_3: the bus arrives at the stop between 10:10 a.m. and 10:15 a.m.;

A_4: the bus arrives at the stop between 10:10 a.m. and 10:15 a.m., carrying at most 40 passengers.

(iii) Which, if any, of the events in part (ii) is a countable set?

(iv) Suppose now that, in order to describe the experiment of the bus arrival at the bus stop, we use pairs (i, τ), where i is the number of passengers that may get on the bus when it arrives, while τ represents the time after 10:00 a.m. that the bus arrives at the stop. Write down a sample space Ω' for this experiment. Express each of the events in part (ii) as subsets of this new sample space.

12. At a car production line in a factory, each engine produced is tested to examine whether it is ready for use or has some fault. If two consecutive engines that are examined are found faulty, the production process stops and is revised (in such a case, the experiment is terminated). Otherwise, the process continues.

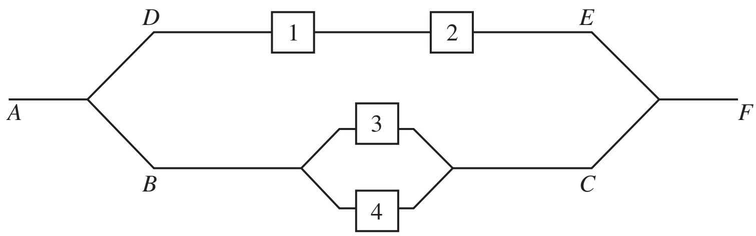
(i) Provide a suitable sample space to describe the inspection process of the engines.

(ii) Find an expression to describe each of the following events

A_i: the production line will be revised after i engines have been examined

for $i = 2, 3, 4, 5, 6$.

13. In a water supply network, depicted below, the water is transferred from point A to point F through water tubes. At the positions marked with the numbers $1, 2, 3$, and 4 on the graph, there are four switches which, if turned off, stop the water supply passing through the tube.



(i) Find a sample space for the experiment which describes the positions of the four switches (ON or OFF).

(ii) Identify each of the following events as a suitable subset of the sample space given in part (i):

A_1: there is water flow from point D to point E;

A_2: there is water flow from point B to point C;

A_3: there is water flow from point A to point F.

1.2 OPERATIONS BETWEEN EVENTS

In the preceding section, we made a distinction between discrete (i.e. finite or countably infinite) and continuous sample spaces. When a sample space Ω is discrete, then any subset of Ω is an event. For continuous sample spaces, however, some theoretical difficulties appear if we assume that all subsets of Ω are events. There are cases where certain[3] sets have to be excluded from the "family of events" related to a sample space Ω. The treatment of such technical difficulties is beyond the scope of the present book and in all applications we will consider, we assume that any subset of the sample space is an event.

Suppose that Ω is a sample space for a chance experiment and $A \subseteq \Omega$ is an event. If, in a realization of the experiment we are studying, we observe the outcome $\omega \in \Omega$ which belongs to A, then we say that A **has occurred** or that A has appeared. For example, if we toss a coin three times and we observe the outcome HTH, then (with reference to Example 1.1) we may say that

- the event $A = \{HHT, HTH, THH\}$ has occurred, but
- the event $B = \{HHH, TTT\}$ has not occurred.

For the study of events associated with a certain experiment, and the assignment of probabilities to these events later on, it is essential to consider various relations among the events of a sample space, as well as operations among them. Recall that each event is

[3]Such sets are studied in a specific branch of mathematics called *Measure Theory*. Subsets of Ω which cannot be considered as events for the sample space are called "nonmeasurable" sets.

mathematically represented by a set (a subset of Ω); thus, it is no surprise that the relations and operations we consider are borrowed from mathematical set theory.

To begin with, assume that A and B are events on the same sample space Ω. If every element (sample point) of A is also a member of B, then we use the standard notation for subsets and write $A \subseteq B$ (A is a subset of B). In words, this means that whenever A occurs, B occurs too. For instance, in a single throw of a die (see Example 1.2), consider the events:

$$A\colon \text{ the outcome of the die is 4,}$$

$$B\colon \text{ the outcome of the die is an even integer.}$$

Then, expressing A and B as sets, we have $A = \{4\}, B = \{2, 4, 6\}$ and it is clear that $A \subseteq B$. On the other hand, if we know that the outcome is 4 (A has occurred), then necessarily the outcome is even, so that B occurs, too.

If $A \subseteq B$ and $B \subseteq A$, then obviously A occurs iff (if and only if) B occurs in which case we have the following definition.

Definition 1.2 Two events A and B, defined on a sample space Ω, are called equivalent if when A appears, then B appears, and vice versa. In this case, we shall write $A = B$.

The entire sample space Ω itself is an event (we have trivially $\Omega \subseteq \Omega$) and, since Ω contains all possible experiment outcomes, it is called a **certain event**. On the other hand, if we are certain that an event cannot happen, then we call this an **impossible event** for which we use the empty set symbol, $\emptyset$.

Coming back to the experiment of throwing a die, consider again the event $A = \{4\}$ and, instead of B, the event

$$C\colon \text{ the outcome of the die is 5.}$$

Suppose now that Nick, who likes gambling, throws a die and wins a bet if the outcome of the die is *either* 4 *or* 5. Then, the event

$$D\colon \text{ Nick wins the bet}$$

occurs if and only if at least one of the events A and C occur. The event D is the same as the event "at least one of A and C occur" (more precisely, and according to Definition 1.2, these two events are equivalent), and so, using set notation, we can write $D = \{4, 5\}$. We thus see that, expressed as a set, D coincides with the union of the two sets A and C.

Definition 1.3 The **union** of two events A and B, denoted by $A \cup B$, is the event which occurs if and only if at least one of A and B occur.

The union operation can be easily extended when more than two sets are involved. More specifically, if $A_1, A_2, \ldots, A_n$ are events on a sample space, then the event "at least one of the A_i's occur" is called the union of the events $A_1, A_2, \ldots, A_n$ and is expressed in symbols as

$$A_1 \cup A_2 \cup \cdots \cup A_n = \bigcup_{i=1}^{n} A_i.$$

To illustrate the next concept, viz., the intersection between two events, we return to the example of throwing a die. Suppose we have two gamblers who play against an opponent.

In particular, each of them throws a die: the first player wins if the outcome of the die is greater than or equal to 4, while the other one wins if the outcome is an even integer. How do we describe the event that "both players win their bets?"

Let A be the event that the first gambler wins his bet, B the event that the second gambler wins, and C the event that they both win their bets. Then, clearly, C occurs if and only if both A and B occur. In order to find C explicitly, we resort to set notation again; A and B can be written as follows:

$$A = \{4, 5, 6\}, \quad B = \{2, 4, 6\}.$$

It is now apparent that the event "both A and B occur" contains exactly those elements of the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ which belong to both A and B, that is,

$$C = \{4, 6\}.$$

In set theory, the set which contains exactly those elements which are common to two other sets A and B is called the *intersection* between A and B. We thus arrive at the following definition.

Definition 1.4 The **intersection** between two events A and B is the event which occurs whenever both A and B occur. For the intersection between A and B, we write $A \cap B$ or simply AB.

It should be mentioned that in most probability books, in contrast to set theoretic books, the notation AB is more common than $A \cap B$ and this is the notation which will largely be followed in the present book.

As with the union before, we can extend the last definition to cover more than two sets; in particular, assume again that $A_1, A_2, \dots, A_n$ are events defined on a sample space. Then, the event "all the events A_i occur" is called the intersection of the events $A_1, A_2, \dots, A_n$ and we denote it by

$$A_1 \cap A_2 \cap \cdots \cap A_n = \bigcap_{i=1}^{n} A_i,$$

or, more commonly,

$$A_1 A_2 \cdots A_n = \prod_{i=1}^{n} A_i,$$

where the product symbol $\prod$ is used as a generic symbol for the intersection among the A_i's.

When we consider an infinite sequence of events $A_1, A_2, \dots$, their union and intersection are denoted by

$$\bigcup_{i=1}^{\infty} A_i \quad \text{and} \quad \bigcap_{i=1}^{\infty} A_i = \prod_{i=1}^{\infty} A_i,$$

respectively. When two events cannot occur simultaneously, then they are called *disjoint* or *mutually exclusive events*. An equivalent definition of this concept is the following.

Definition 1.5 Two events A and B, defined on a sample space Ω, are called **disjoint** or **mutually exclusive events** if their intersection is the empty set (impossible event), i.e. $AB = \emptyset$.

In the case where we have more than two events, $A_1, A_2, \dots, A_n$, and that

$$A_iA_j = \emptyset \quad \text{for any } i \neq j \ (i, j = 1, 2, \dots, n),$$

then the events A_i are said to be **pairwise disjoint**. A similar definition applies when the number of the A_i's is infinite.

Two further important concepts useful in probability theory, which arise again from similar concepts in set theory, are given in the next two definitions.

Definition 1.6 Let A be an event on a sample space. The **complement**, or the **complementary event** of A, is the event which occurs if and only if A does not occur.

An equivalent definition is that the complement of A contains exactly those elements of the sample space which are not elements of A.

For the complement of an event A, there are at least three different symbols which are used quite commonly, and they are

$$A' \quad \text{or} \quad \overline{A} \quad \text{or} \quad A^c.$$

In this book, we prefer to use the first of these three symbols.

Definition 1.7 The **difference** of an event B from an event A refers to the event which occurs when A occurs but B does not. The symbol we use for the difference of B from A is $A - B$.

Two useful expressions, which link the above concepts and follow easily from the definitions above, are the following:

$$A' = \Omega - A \quad \text{and} \quad A - B = AB'.$$

Example 1.6 (Emission of digital signals)

A digital signal is a stream of numbers, usually in binary form (0 or 1). Such signals are used, for example, in electrical engineering, telecommunications, biomedicine, seismology and so on. Assume that a source emits a sequence of four binary digits, such as

$$0010, \quad 0001, \quad 1010, \quad 1100, \quad \dots$$

As can be seen from the following tree diagram, for each 4-digit emission, the sample space consists of 16 elements, given explicitly in the final column (which has been labeled as "result") of the diagram (Figure 1.2).

Suppose we are interested in the following events:

A_i: the signal contains exactly i "0" digits, for $i = 0, 1, 2, 3, 4$;

B: the signal contains at least two "0" digits;

C: the signal contains exactly three digits which are the same;

D: the signal has at least one "0" digit and one "1" digit;

E: the signal contains exactly three successive digits which are the same.

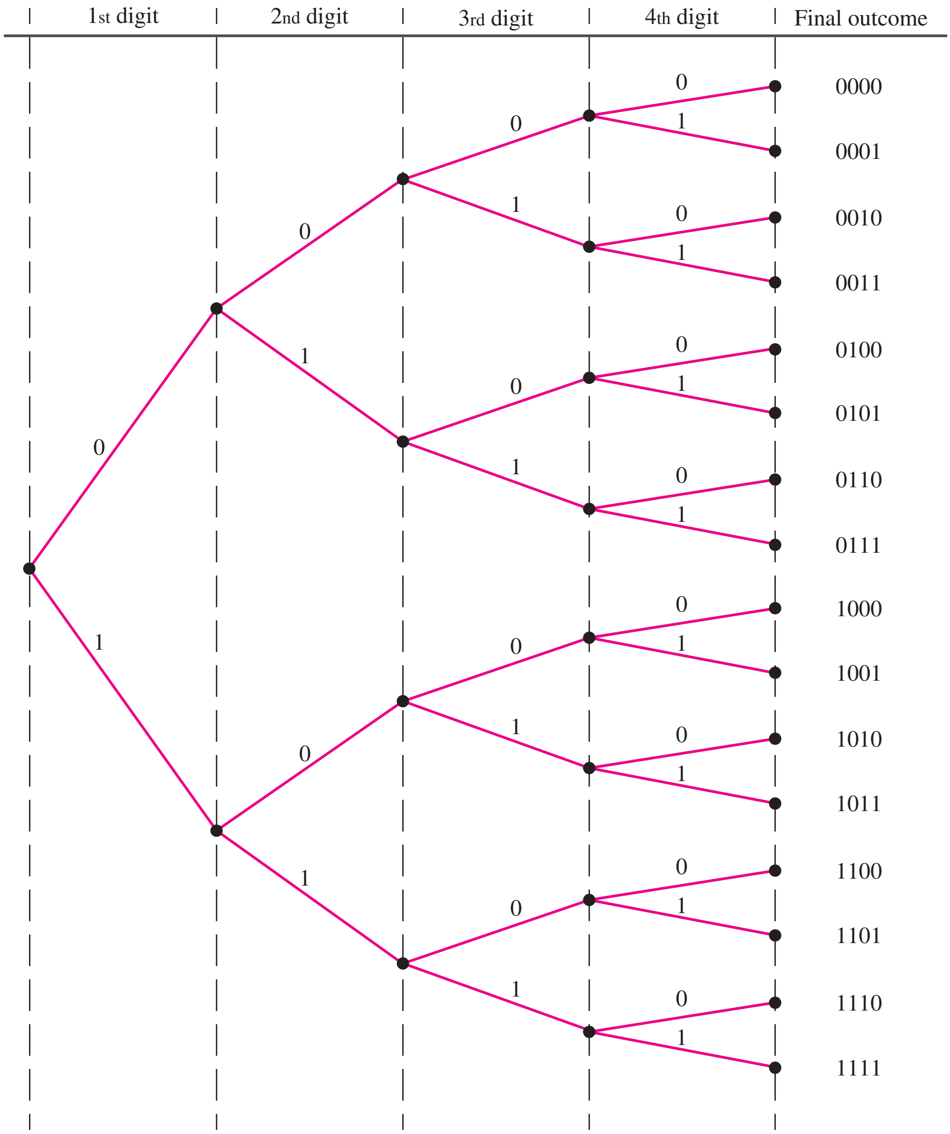


Figure 1.2 Tree diagram for the emission of a signal consisting of four binary digits.

After a careful look through the tree diagram, we conclude that the events A_i $(i = 0, 1, 2, 3, 4)$ can be written explicitly as follows:

$$A_0 = \{1111\}, \quad A_1 = \{0111, 1011, 1101, 1110\},$$
$$A_2 = \{0011, 0101, 0110, 1001, 1010, 1100\},$$
$$A_3 = \{0001, 0010, 0100, 1000\}, \quad A_4 = \{0000\}.$$

It is apparent that the A_i's are pairwise disjoint, i.e.

$$A_iA_j = \emptyset \quad \text{for any } i \neq j \ (i, j = 0, 1, 2, 3, 4).$$

Note that, in addition, we have

$$A_0 \cup A_1 \cup A_2 \cup A_3 \cup A_4 = \Omega$$

(a family of events with these two properties is called a **partition** of the sample space Ω). The events B, C, D can then be expressed in terms of the events A_i as follows:

$$B = A_2 \cup A_3 \cup A_4, \quad C = A_1 \cup A_3, \quad D = A_1 \cup A_2 \cup A_3 = A_0' A_4'.$$

Finally, from the description of the event E, it is obvious that

$$E \subseteq A_1 \cup A_3.$$

More precisely, the event E contains the following elements

$$E = \{0111, 1110, 0001, 1000\}$$

and there seems to be no simple way in which it can be expressed in terms of the events A_i, $i = 0, 1, 2, 3, 4$, by employing the usual operations between events.

Another tool from set theory that is useful often to describe and understand better relations between events is a Venn diagram. In such a diagram, the sample space Ω is represented by a rectangle and, in it, we plot circles or other geometrical shapes to represent certain events in that space (see Figure 1.3).

Figures 1.4–1.11 illustrate graphically the various concepts (relations and operations among events) which have been discussed so far.

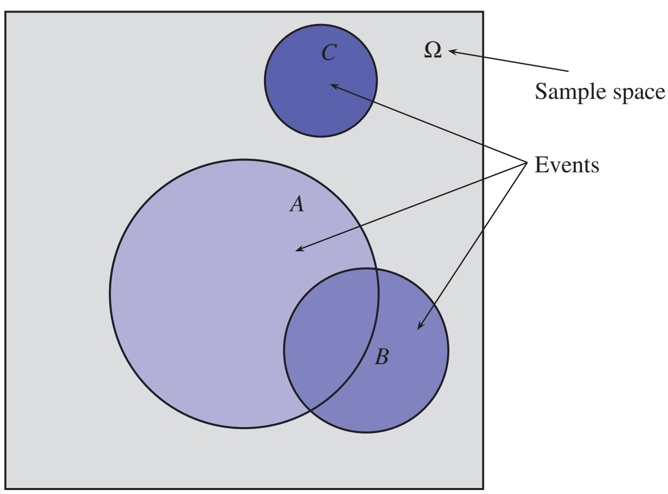


Figure 1.3 A Venn diagram.

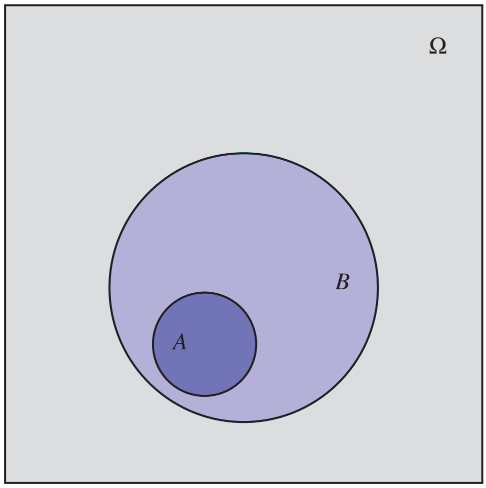


Figure 1.4 $A \subseteq B$.

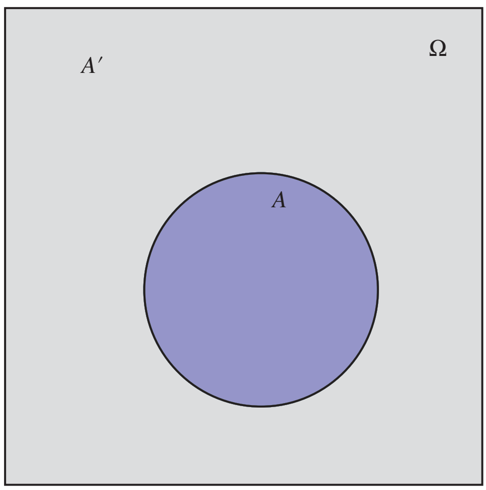


Figure 1.5 Complement of an event.

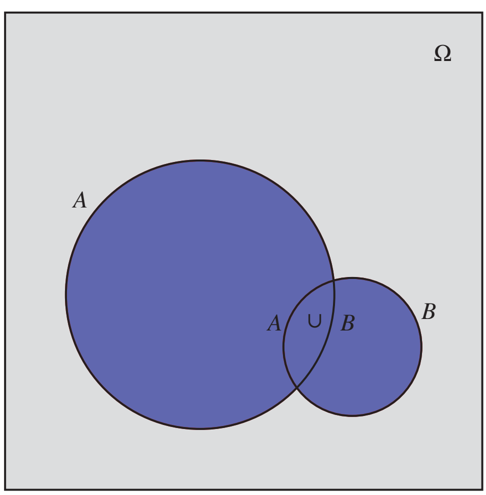


Figure 1.6 Union of events.

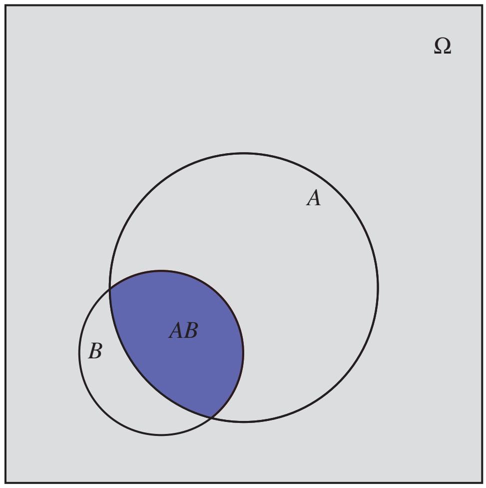


Figure 1.7 Intersection of events.

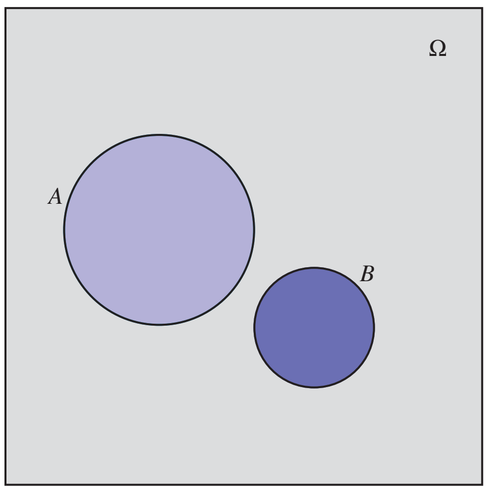


Figure 1.8 Disjoint events.

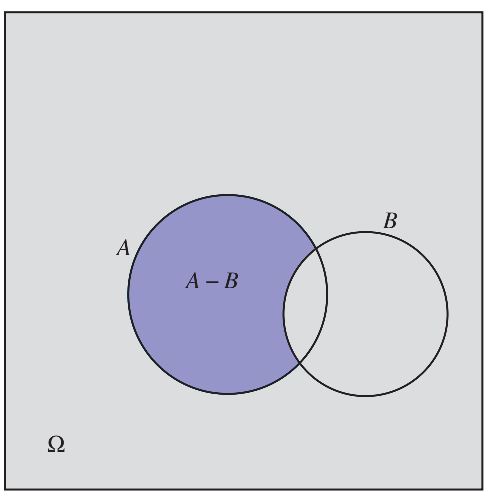


Figure 1.9 Difference between two events.

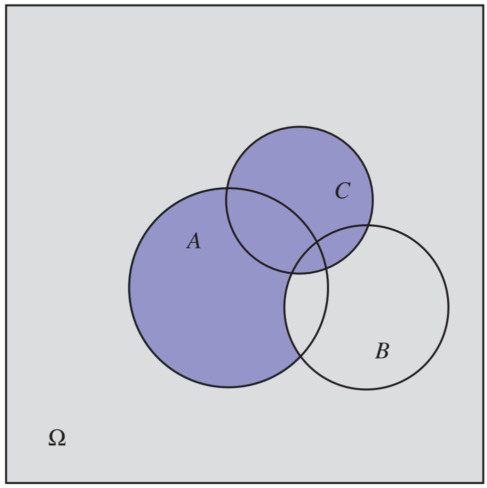


Figure 1.10 $(A - B) \cup C$.

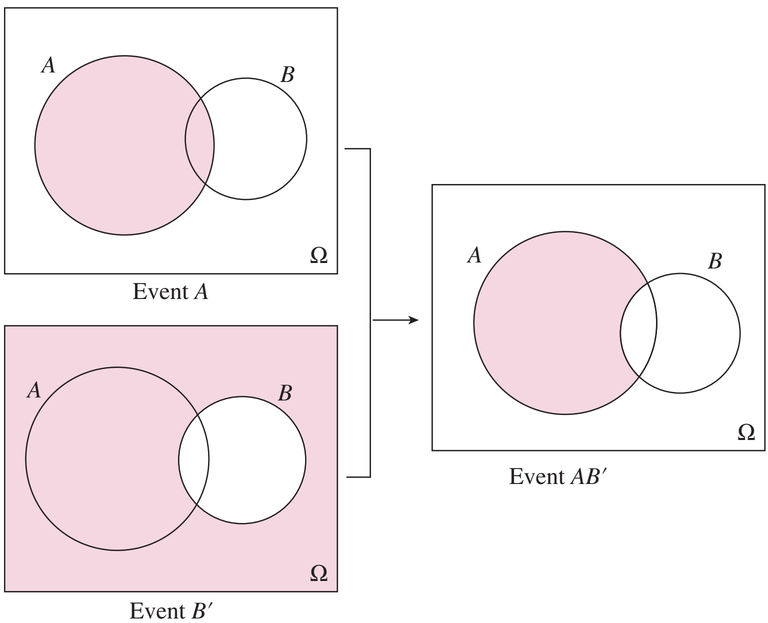


Figure 1.11 Venn diagram for the event AB'.

The use of such diagrams offers, typically quite useful, visualizations of the various relations among events, and enables us to depict graphically rather complicated events. For example, let A, B, C be three events on a sample space and consider the event

C: either A occurs but not B, or else C occurs.

From the definitions given earlier, we see that this event corresponds to the set $(A - B) \cup C$; this event is represented by the shadowed area in Figure 1.10.

Considering the various operations among events that have been introduced, we may state numerous properties that are particularly useful when we wish to simplify complicated expressions. Some of these properties are listed in the following proposition.

Proposition 1.1 *The following properties hold true for the operations among events:*

*SP*1. $A \cup A = A \qquad AA = A$

*SP*2. $A \cup \emptyset = A \qquad A\emptyset = \emptyset$

*SP*3. $A \cup \Omega = \Omega \qquad A\Omega = A$

*SP*4. $A \cup B = B \cup A \qquad AB = BA$

*SP*5. $A \cup (B \cup C) = (A \cup B) \cup C \qquad A(BC) = (AB)C$

*SP*6. $A \cup (BC) = (A \cup B)(A \cup C) \qquad A(B \cup C) = (AB) \cup (AC)$

*SP*7. $A \cup A' = \Omega \qquad AA' = \emptyset$

*SP*8. $(A')' = A$

*SP*9. *If* $A \subseteq B$ *and* $B \subseteq C$, *then* $A \subseteq C$

*SP*10. *If* $A \subseteq B$, *then* $B' \subseteq A'$ *and vice versa*

*SP*11. *If* $A \subseteq B$, *then* $AB = A$ *and* $A \cup B = B$.

Property *SP*5, known as the *associativity property,* allows us to omit the brackets and use a notation like $A \cup B \cup C$ or ABC, because the order in which the unions or intersections are performed is immaterial. Also, *SP*4 in the above proposition (commutative property) shows that when the union or intersection operators are used, we can change the order of the events considered.

We further note that, those properties in Proposition 1.1 which are not immediate, can be verified easily with the use of a Venn diagram; such diagrams can also be used to discover and study a number of additional relationships among events. As an illustration of this, which is of particular interest, we demonstrate the force of an identity concerning the intersection between the complements of two sets. Let A and B be these two sets (events). The shaded area in the two Venn diagrams on the left of Figure 1.12 present their complements, A' and B', respectively, while the graph on the right shows their intersection, $A'B'$.

From the diagram shown in Figure 1.13, it is apparent that the same result could be obtained if we consider the complement of the union, $A \cup B$.

It therefore seems from Figures 1.12 and 1.13 that the following identity holds:

$$(A \cup B)' = A'B'.$$

However, we stress that, no matter how illustrative they might be, Venn diagrams alone do not offer a rigorous mathematical proof for an identity like the one above. In order to prove such an identity formally, we must show that the sets on the two sides of an equation contain exactly the same elements. As an illustration of how this can be done, we present a detailed mathematical proof of the following proposition.

Proposition 1.2 ***(De Morgan formulas)*** *Suppose* A *and* B *are events on the same sample space* Ω. *Then, the following identities hold:*

$$(A \cup B)' = A'B', \quad (AB)' = A' \cup B'.$$

These identities are known as De Morgan identities or De Morgan formulas.

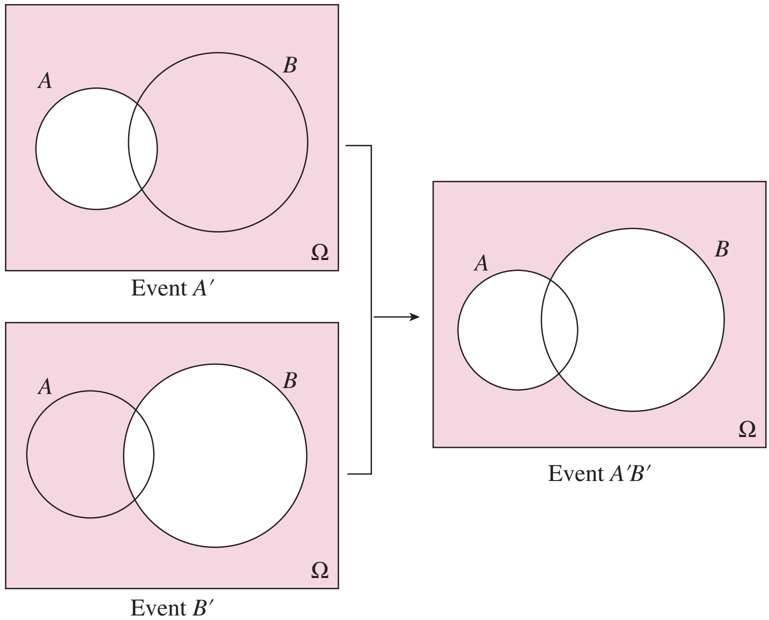


Figure 1.12 Venn diagram for the event $A'B'$.

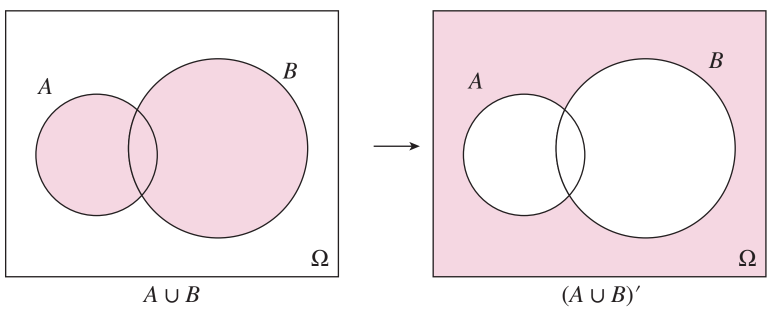


Figure 1.13 Venn diagram for the event $(A \cup B)'$.

Proof: To prove the first identity, it is sufficient to show that both $(A \cup B)' \subseteq A'B'$ and $A'B' \subseteq (A \cup B)'$ hold. To begin with, let ω be a sample point which belongs to the set $(A \cup B)'$. This means that $\omega \notin A \cup B$, and since the set $A \cup B$ contains all elements which are either in A or in B, we conclude that ω does not belong to any of them. Thus, ω belongs to both A' and B', and therefore it is also a member of their intersection, $A'B'$. The above arguments show that $(A \cup B)' \subseteq A'B'$.

Next, to establish the reverse relationship, suppose now that $\omega \in A'B'$. Since the set $A'B'$ contains exactly those elements of the sample space which belong to both A' and B', we see that ω is not a member of either A or B. Thus, it does not belong to their union, $A \cup B$, which shows immediately that $\omega \in (A \cup B)'$. We therefore see that $A'B' \subseteq (A \cup B)'$, and this completes the proof for the first assertion of the proposition.

The proof of the second identity is carried out in a similar fashion. □

De Morgan identities can be generalized for the case when more than two sets are involved. More specifically, if $A_1, A_2, \dots, A_n$ are events on a sample space, then we have the following identities:

$$(A_1 \cup A_2 \cup \cdots \cup A_n)' = A_1' A_2' \cdots A_n', \quad (A_1 A_2 \cdots A_n)' = A_1' \cup A_2' \cup \cdots \cup A_n'.$$

A similar result holds true if we have an infinite number of events.

We close this section with an example to demonstrate how we can use the various properties we have discussed so far in order to simplify complicated expressions involving sets and their operations.

Example 1.7 Suppose A, B, and C are three events on the same sample space Ω. Using the properties of the operations among events, simplify the expression

$$(A'B')' \cup (AB'C')'.$$

SOLUTION Using the second De Morgan identity from Proposition 1.2 (applied to the sets A' and B') and Property $SP8$, we obtain

$$(A'B')' = (A')' \cup (B')' = A \cup B.$$

In a similar fashion, we get

$$(AB'C')' = A' \cup (B')' \cup (C')' = A' \cup (B \cup C).$$

Inserting these two relations into the expression given in the statement of the example, we obtain

$$\begin{aligned}
(A'B')' \cup (AB'C')' &= (A \cup B) \cup (A' \cup (B \cup C)) \\
&= A \cup B \cup A' \cup B \cup C \\
&= (A \cup A') \cup (B \cup B) \cup C \quad \text{(Properties SP4, SP5)} \\
&= \Omega \cup B \cup C \quad \text{(Properties SP1, SP7)} \\
&= \Omega \cup (B \cup C) \\
&= \Omega. \quad \text{(Property SP3)}
\end{aligned}$$

EXERCISES

Group A

1. Let A, B, and C be three events in a sample space Ω. Express each of the following events by the use of the operators (unions, intersections, complements) among sets:

   (i) all three events occur;

   (ii) at least one of the three events occur;

   (iii) A occurs, but not A and B;

(iv) A and C occur, but not B;

(v) at least one of the three events occur, but C does not occur;

(vi) exactly two among the three events A, B, and C occur.

2. Consider the experiment of throwing a die twice, and define the following events:

   A: the sum of the two outcomes is 6;

   B: the two outcomes are equal;

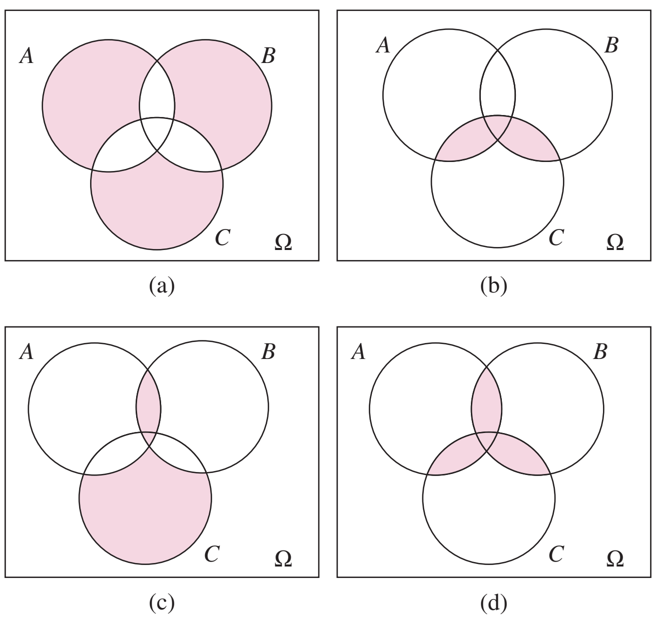
   C: the first outcome is an even integer;

   D: the first outcome is an odd integer.

   Write down the 36 sample elements for this experiment, and then identify which of these elements each of the sets A, B, C, D contains. Then describe the following events, both by expressing in words what each of them represents and by listing their elements:

$$AB, \quad AC', \quad AB'C', \quad AC \cup AD, \quad A \cup (CD), \quad BC' \cup BD, \quad (B \cup C)(B \cup D).$$

3. For each of the following graphs (a)–(d), express the event in the shaded area in terms of the events A, B, and C and state, in words, what this event represents.



4. What conclusions can we draw about the events A and B if the following relations hold?

   (i) $A \cup B = A$; (ii) $A - B = A$;

   (iii) $AB = A$; (iv) $A - B = B - A$.

5. Suppose for the events A, B, C, D, we have $A \subseteq B$ and $C \subseteq D$. Then, arguing as in the proof of Proposition 1.2, show that

$$A \cup C \subseteq B \cup D \quad \text{and} \quad AC \subseteq BD.$$

6. If, for the events A and B, we know that $A \subseteq B$ and $A \subseteq B'$, show that $A = \emptyset$. (*Hint*: Use the result of Exercise 5.)

7. Suppose the events A and B of a sample space Ω are such that $A \subseteq B$ and $A' \subseteq B$. Then prove that $B = \Omega$. (*Hint*: You may use the result of Exercise 5.)

8. Let A, B, and C be three events on a sample space Ω. Examine, possibly with the aid of Venn diagrams, which of the following results are always true:

   (i) $(A - AB)B = AB$;

   (ii) $(A \cup B)'C = A'B'C$;

   (iii) $(AB)'C' = (ABC)'$;

   (iv) $AB \subseteq A \subseteq A \cup B \cup C$;

   (v) $A'B \cup C'B = (A \cup C)'B$;

   (vi) $(A - B) - C = A - (B - C)$.

Group B

9. A box contains 15 balls numbered $1, 2, \dots, 15$. The balls numbered 1–5 are white and those numbered 6–15 are black. We select a ball at random, and record its color and the number on it.

   (i) Write down a suitable sample space for this experiment.

   (ii) We define the events

   A_i: the ball selected has a number less than or equal to i, for $1 \le i \le 15$;

   B_i: the ball selected has a number greater than or equal to i on it, for $1 \le i \le 15$;

   C: the selected ball is white;

   D: the selected ball is black.

   Examine which of the following assertions are correct and which are false:

   (a) $A_5 = C$;

   (b) $A_4 \subseteq C$;

   (c) $A_iB_i = \emptyset$ for any $i = 1, 2, \dots, 15$;

   (d) $A_{i-1}B_i = \emptyset$ for any $i = 2, 3, \dots, 15$;

   (e) $A_iB_{i+1} = \emptyset$ for any $i = 1, 2, \dots, 14$;

   (f) the events C and D are mutually exclusive;

   (g) $A_{10}B_5 \subseteq C$;

(h) $A_7D = \emptyset$;

(i) $A_5' = D$;

(j) $A_i \cup B_{i+1} = A_i$ for any $i = 1, 2, \dots, 14$;

(k) $A_1 \subseteq A_2 \subseteq \cdots \subseteq A_{15}$;

(l) $A_i \cup B_i = \Omega$ for any $i = 1, 2, \dots, 15$;

(m) $B_1 \subseteq B_2 \subseteq \cdots \subseteq B_{15}$;

(n) $A_i \cup B_{i+1} = \emptyset$ for any $i = 1, 2, \dots, 15$;

(o) $A_i' = B_{i+1}$ for $i = 1, 2, \dots, 14$;

(p) $(A_{10} - C)B_6 = \emptyset$;

(q) $(A_{12} - D) \subseteq B_5$;

(r) $D - B_{11} = A_{10} - A_5$.

10. Express each of the following events in terms of the events A_i, B_i, C, D defined in Exercise 9.

(i) The number on the ball selected is greater than 4 and less than 10.

(ii) The ball selected is either white or has a number greater than 10 on it.

(iii) The ball selected is black or has an even number on it.

(iv) The ball selected is black and has an odd number on it.

(v) The selected ball is either white with a number less than 3 on it, or else it is black with a number less than 10.

11. In order to describe a chance experiment, we have used the following (continuous) sample space

$$\Omega = \{(x, y) : -5 \le x \le 5 \text{ and } -3 \le y \le 7\}.$$

On this space, we define the following events:

$$\begin{aligned} A &= \{(x, y) \in \Omega : x = y\}, & B &= \{(x, y) \in \Omega : x^2 = y^2\}, \\ C &= \{(x, y) \in \Omega : x + y = 0\}, & D &= \{(x, y) \in \Omega : x \le y\}, \\ E &= \{(x, y) \in \Omega : x \ge y\}. \end{aligned}$$

Which of the following statements are correct and which are false?

(a) $B = AC$;

(b) $AC \subseteq B$;

(c) D and E are mutually exclusive events;

(d) $A = \{(x, x) : |x| \le 5\}$;

(e) $A = DE$;

(f) $D' = E - A$;

(g) $A = \{(x, x) : -3 \le x \le 3\}$.

12. Let $A_i, i = 1, 2, \ldots, n$, be events on the same sample space Ω. Express in words what conclusions can be drawn about these events in each of the following cases:

    (i) $A_1 \cup A_2 \cup \cdots \cup A_n = A_1$;

    (ii) $A_1 A_2 \cdots A_n = A_1$;

    (iii) $A_1 \cup A_2 \cup \cdots \cup A_n = A_1 \cup A_n$;

    (iv) $A_1 A_2 \cdots A_n = A_1 A_2$.

13. Simplify each of the expressions below by the use of properties among event operators:

    (i) $A'B'AB$;

    (ii) $(A' \cup B') \cup AB$;

    (iii) $(A \cup B)(A \cup B')(A' \cup B)$;

    (iv) $(A'B')(A \cup B)$;

    (v) $[(A \cup B)(A \cup B')] \cup [(A' \cup B)(A' \cup B')]$;

    (vi) $(A \cup B \cup C)(A \cup B \cup C')(A \cup B')$;

    (vii) $A \cup (B - AB) \cup (C - AC)$;

    (viii) $(A \cup B \cup C)(A \cup BC')$;

    (ix) $[ABC(A \cup B \cup C')] \cup (A \cup B')$.

14. From an ordinary card deck with 52 cards, we select n cards successively. We consider the events

    A_1: the first card drawn is an ace;

    A_n: in the first $n - 1$ selections ($n \geq 2$), neither an ace nor a face card is drawn and in the nth selection an ace is drawn,

    where $n = 2, 3, \ldots$ Write down an expression, in terms of $A_1, A_2, \ldots$, for the event that an ace is drawn before a face card.

15. Let A, B, and C be three events in a sample space Ω. In each case below, find an event X such that the union operator on the right hand side is applied between events which are disjoint (for example, in (i) we seek X such that AB and X are disjoint sets):

    (i) $B = (AB) \cup X$;

    (ii) $A \cup B = (A'B) \cup X$;

    (iii) $A \cup B \cup C = A \cup (A'B) \cup X$.

16. For any events A, B, and C in a sample space, verify the truth of the following relations:

    (i) $(A - B) - C = (A - C) - (B - C)$;

    (ii) $(B - A) \cup (C - A) = (B \cup C) - A$;

    (iii) $A - (B - C) = (A - B) \cup (AC)$;

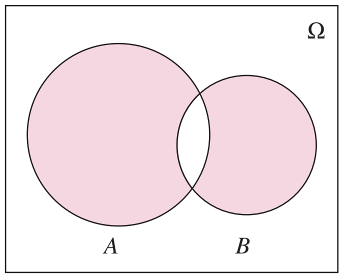
(iv) $(A - B) \cup B = A \cup B$;

(v) $A(B - C) = AB - AC$;

(vi) $A \cup B \cup C = (A - B) \cup (B - C) \cup (C - A) \cup (ABC)$.

17. For any events A and B in the same sample space Ω, we define the **symmetric difference** of A and B to be the event (see next figure)

$$A \Delta B = (A - B) \cup (B - A).$$



(i) Express in words what this event represents;

(ii) Show that the following hold:

$$A\Delta B = B\Delta A, \quad A\Delta B = (A \cup B) - AB, \quad (A\Delta B)\Delta B = A.$$

1.3 PROBABILITY AS RELATIVE FREQUENCY

As already mentioned, the main feature of a chance (random) experiment is that we do not know what its outcome will be in each realization of it. We therefore do not know whether a specific event, A, associated with this experiment will occur in a particular realization of the experiment. Yet, however, we very often want to know how likely it is for A to occur. For instance, what are the chances that it will rain tomorrow, or that I will pass a certain University exam, or that a political party will win the forthcoming elections? Today, hundreds of decisions are made on a daily basis, using the "degree of belief" we have that something will happen, or not, at some future point; just imagine plainly the persons who purchase stocks, or other financial products, hoping that their prices will go up. Probability theory emerged from the need to put this "degree of belief" in a proper mathematical framework, so that the concept of "how likely" an event is can be quantified on one hand, but on the other hand to be able to make logical deductions and draw conclusions about complicated events. More generally, probability theory enables us not only to make decisions but to deal with uncertainty in everyday life, and for this reason it has been referred to as "the science of uncertainty." From this viewpoint, it may seem paradoxical that the entire history of probability theory and its development, until the nineteenth century at least, is entangled and motivated by games of chance and gambling. Even so, it is curious that, although the origins of games of chance seem to have been lost in history,[4] the first systematic use of mathematics to deal with the uncertainty in games

[4] An archaeological excavation in south-eastern Iran unearthed a backgammon, which dates back to 3000BC.

occurs in fifteenth century, while the first worthwhile achievements were established in mid-seventeenth century. And, since then, it took nearly three centuries until a sound and coherent theory was developed so that probability theory became recognized as an independent branch of mathematics.

The first definition of probability to be widely used is attributed to the French mathematician Pierre Simon Laplace (1749–1827) and is usually referred to as the classical definition of probability. In his words, it is as follows:

> "The probability of an event is the ratio of the number of cases favorable to it, divided by the number of all cases possible when nothing leads us to expect that any one of these cases should occur more than any other, which renders them, for us, equally possible."

Thus, when we throw a die and consider the event

A: the outcome is an even integer,

it seems plausible to assume that all six faces of the die are "equally possible." The number of possible outcomes (the number of elements in the sample space, to conform with the terminology used in the previous sections) is 6, while the "favorable" outcomes are all those in which an even integer results in the throw. These are the elements of the set $\{2, 4, 6\}$ and we therefore obtain from the definition that the probability of the event A is $3/6$, that is, a half.

Although the definition seems to give plausible answers which agree with our intuition, it has at least two serious drawbacks:

(i) What happens if all the "cases" that may occur (that is, the elementary events in the sample space) can not be considered as being "equally possible?"

(ii) The definition assumes that the number of possible outcomes is finite. What happens if this is not the case for the experiment that we consider?

In view of the above problems, mathematicians searched for alternative ways to define probability. The first serious attempt to do this was made by R. von Mises in 1917. This attempt relies on the notion of a relative frequency which is discussed in the rest of this section. The approach taken by von Mises was rather empirical in nature, and this has been considered both as an advantage and a disadvantage; on the one hand, it is nice to link our "degree of belief" with our experience from the past. However, relying on experimental data to define probability has several drawbacks, as we shall see in the next section. Thus, and with the increasing recognition of the importance of probability theory in early twentieth century, scientists felt that there was a need to "axiomatize" probability theory, in the same way as was done with geometry for example. That is, to find a set of axioms which would seem universally acceptable and, from these, derive a chain of theoretical results and also the answer to practical problems encompassing uncertainty. In the next section, we present an approach which followed this path, due to the Russian probabilist A. N. Kolmogorov.

Let us agree, to begin with, that certain events are intuitively more likely than others; for instance,

- when two basketball teams, which are more or less at the same level, play against each other, we typically think that the home team has better chance to win the game;
- our experience suggests that it is "a lot more likely" to rain on a winter day in England than on a summer day in Las Vegas.

Both these assertions seem to be universally accepted, and this can be attributed to our common experience from what has happened in the past. This is more obvious in the second case above, wherein if we collect data say from the last 50 years, we may see for instance that the number of winter days with rain in England is 20 times as much as the number of rainy days in summer in Las Vegas. The same argument may work for the first case since if we select at random a large number of basketball games, it is very likely that home wins occur more often than away wins. For example, looking at the 2010–2011 NBA regular season, we see that among the 1230 matches which were played, there were 743 home wins and 487 away wins.

When the classical definition of probability cannot be used, it seems reasonable to use past experience to assign a degree of credibility to an event. To begin with, we may say that an event A is more likely than another event B, both associated with the same experiment, if in a large number of realizations of this experiment, A occurred more often than B. At a next level, we may attempt to quantify this difference, saying for instance that if A occurs twice as often as B does, then A is twice more likely to occur than B. However, we would wish to use a single number to tell us how likely an event is, rather than considering this always in relation to some other event. Following the above, it seems plausible that, if an experiment is conducted n times and A occurs n_A times out of them, then the ratio n_A/n can be used as a measure of our degree of belief for its appearance. Of course, for this to be valid, and consistent, all n repetitions have to be performed under the same conditions. We thus arrive at the following definition.

Definition 1.8 (Relative frequency) If in n repetitions of an experiment, under identical conditions, the event A occurs n_A times, the ratio

$$f_A = \frac{n_A}{n}$$

is called the *relative frequency* of A (in these n experiments).

From this definition, we have immediately the following properties for a relative frequency.

Proposition 1.3 ***(Properties of relative frequencies)*** *Let Ω be a sample space for a chance experiment. Then:*

(i) $f_A \geq 0$ *for any event A defined on Ω;*

(ii) $f_\Omega = 1$;

(iii) *for any disjoint events A and B, we have*

$$f_{A\cup B} = f_A + f_B.$$

Clearly, the third property can be generalized to more than two events. Thus, if $A_1, A_2, \dots, A_k$ are pairwise disjoint events, we have

$$f_{A_1 \cup A_2 \cup \cdots \cup A_k} = f_{A_1} + f_{A_2} + \cdots + f_{A_k} = \sum_{i=1}^{k} f_{A_i}.$$

Furthermore, if $A = \{\omega_1, \omega_2, \dots, \omega_k\} \subseteq \Omega$ is a (finite) event on Ω and we denote by

$$f_i = f_{\omega_i}, \quad i = 1, 2, \dots, k,$$

the relative frequencies of the elementary events $\{\omega_i\}$, $i = 1, 2, \dots, k$, then we may write

$$f_A = f_1 + f_2 + \cdots + f_k = \sum_{i=1}^{k} f_i.$$

Example 1.8 (Repetitions of a chance experiment)
Table 1.1 gives the results of 100 realizations for the random experiment of throwing a die twice. In the last three columns of this table, we have recorded the relative frequency (based on the first i trials, $i = 1, 2, \dots, 100$) for each of the events

A: at least one of the two outcomes is 6;

B: the outcome of the first die is 4;

C: the sum of the two outcomes is 7.

We see from Table 1.1 that the relative frequency for each of the three events A, B, C varies along with the number of throws and, in fact, changes at each step. However, it seems that as the number of throws gets larger, the fluctuations of the values for the three relative frequencies tend to become smaller. Further, throughout Table 1.1, it is apparent that event A occurs more often than the other two, which have about the same rate of occurrence.

Based on the total of 100 throws, we have the following relative frequencies for A, B, C:

$$f_A = 0.29, \quad f_B = 0.16, \quad f_C = 0.15.$$

Suppose now we define the events

$$B_i : \quad \text{the outcome of the first throw is } i$$

for $i = 5, 6$. Taking into account the result for the event B, it seems reasonable to state that

$$f_{B_5} = 0.16, \quad f_{B_6} = 0.16.$$

Further, for the event

$$E : \quad \text{the outcome of the first throw is greater than 4,}$$

since $E = B_5 \cup B_6$, we expect to have (by Proposition 1.3(iii)) that

$$f_E = f_{B_5} + f_{B_6} = 0.32.$$

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
1	5	4	0	0	0	0.0000	0.0000	0.0000
2	6	3	1	0	0	0.5000	0.0000	0.0000
3	4	4	1	1	0	0.3333	0.3333	0.0000
4	5	2	1	1	1	0.2500	0.2500	0.2500
5	3	3	1	1	1	0.2000	0.2000	0.2000
6	1	3	1	1	1	0.1667	0.1667	0.1667
7	5	6	2	1	1	0.2857	0.1429	0.1429
8	2	1	2	1	1	0.2500	0.1250	0.1250
9	2	4	2	1	1	0.2222	0.1111	0.1111
10	3	2	2	1	1	0.2000	0.1000	0.1000
11	2	3	2	1	1	0.1818	0.0909	0.0909
12	4	4	2	2	1	0.1667	0.1667	0.0833
13	5	4	2	2	1	0.1538	0.1538	0.0769
14	6	2	3	2	1	0.2143	0.1429	0.0714
15	2	5	3	2	2	0.2000	0.1333	0.1333
16	3	3	3	2	2	0.1875	0.1250	0.1250
17	6	3	4	2	2	0.2353	0.1176	0.1176
18	1	1	4	2	2	0.2222	0.1111	0.1111
19	4	3	4	3	3	0.2105	0.1579	0.1579
20	2	3	4	3	3	0.2000	0.1500	0.1500
21	1	4	4	3	3	0.1905	0.1429	0.1429
22	3	2	4	3	3	0.1818	0.1364	0.1364
23	6	5	5	3	3	0.2174	0.1304	0.1304
24	3	4	5	3	4	0.2083	0.1250	0.1667
25	4	5	5	4	4	0.2000	0.1600	0.1600
26	2	2	5	4	4	0.1923	0.1538	0.1538
27	4	5	5	5	4	0.1852	0.1852	0.1481
28	6	6	6	5	4	0.2143	0.1786	0.1429
29	6	3	7	5	4	0.2414	0.1724	0.1379
30	1	1	7	5	4	0.2333	0.1667	0.1333
31	5	3	7	5	4	0.2258	0.1613	0.1290
32	5	5	7	5	4	0.2188	0.1563	0.1250
33	4	3	7	6	4	0.2121	0.1818	0.1212
34	1	1	7	6	4	0.2059	0.1765	0.1176
35	4	3	7	7	5	0.2000	0.2000	0.1429
36	6	1	8	7	6	0.2222	0.1944	0.1667
37	3	5	8	7	6	0.2162	0.1892	0.1622
38	5	2	8	7	7	0.2105	0.1842	0.1842
39	6	3	9	7	7	0.2308	0.1795	0.1795
40	1	5	9	7	7	0.2250	0.1750	0.1750

Table 1.1 Dice outcomes of two throws (100 repetitions).

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
41	4	6	10	8	7	0.2439	0.1951	0.1707
42	5	4	10	8	7	0.2381	0.1905	0.1667
43	1	3	10	8	7	0.2326	0.1860	0.1628
44	1	3	10	8	7	0.2273	0.1818	0.1591
45	6	4	11	8	7	0.2444	0.1778	0.1556
46	2	2	11	8	7	0.2391	0.1739	0.1522
47	6	1	12	8	8	0.2553	0.1702	0.1702
48	1	3	12	8	8	0.2500	0.1667	0.1667
49	5	3	12	8	8	0.2449	0.1633	0.1633
50	4	2	12	9	8	0.2400	0.1800	0.1600
51	6	5	13	9	8	0.2549	0.1765	0.1569
52	2	6	14	9	8	0.2692	0.1731	0.1538
53	3	3	14	9	8	0.2642	0.1698	0.1509
54	1	4	14	9	8	0.2593	0.1667	0.1481
55	5	1	14	9	8	0.2545	0.1636	0.1455
56	3	6	15	9	8	0.2679	0.1607	0.1429
57	5	1	15	9	8	0.2632	0.1579	0.1404
58	3	4	15	9	9	0.2586	0.1552	0.1552
59	3	6	16	9	9	0.2712	0.1525	0.1525
60	6	1	17	9	10	0.2833	0.1500	0.1667
61	2	3	17	9	10	0.2787	0.1475	0.1639
62	1	4	17	9	10	0.2742	0.1452	0.1613
63	4	5	17	10	10	0.2698	0.1587	0.1587
64	6	4	18	10	10	0.2813	0.1563	0.1563
65	3	4	18	10	11	0.2769	0.1538	0.1692
66	2	6	18	10	11	0.2727	0.1515	0.1667
67	4	2	18	11	11	0.2687	0.1642	0.1642
68	4	1	18	12	11	0.2647	0.1765	0.1618
69	6	6	19	12	11	0.2754	0.1739	0.1594
70	6	3	20	12	11	0.2857	0.1714	0.1571
71	3	2	20	12	11	0.2817	0.1690	0.1549
72	2	5	20	12	12	0.2778	0.1667	0.1667
73	5	5	20	12	12	0.2740	0.1644	0.1644
74	5	1	20	12	12	0.2703	0.1622	0.1622
75	6	1	21	12	13	0.2800	0.1600	0.1733
76	4	4	21	13	13	0.2763	0.1711	0.1711
77	1	3	21	13	13	0.2727	0.1688	0.1688
78	5	6	22	13	13	02821.	0.1667	0.1667
79	6	5	23	13	13	0.2911	0.1646	0.1646
80	3	4	23	13	14	0.2875	0.1625	0.1750

Table 1.1 (*continued*)

Throw no.	Outcome		Frequency			Relative Frequency		
	First throw	Second throw	n_A	n_B	n_C	f_A	f_B	f_C
81	5	1	23	13	14	0.2840	0.1605	0.1728
82	6	5	24	13	14	0.2927	0.1585	0.1707
83	5	4	24	13	14	0.2892	0.1566	0.1687
84	4	4	24	14	14	0.2857	0.1667	0.1667
85	2	4	24	14	14	0.2824	0.1647	0.1647
86	4	1	24	15	14	0.2791	0.1744	0.1628
87	4	3	24	16	15	0.2759	0.1839	0.1724
88	1	2	24	16	15	0.2727	0.1818	0.1705
89	5	4	24	16	15	0.2697	0.1798	0.1685
90	3	2	24	16	15	0.2667	0.1778	0.1667
91	3	6	25	16	15	0.2747	0.1758	0.1648
92	2	4	25	16	15	0.2717	0.1739	0.1630
93	3	3	25	16	15	0.2688	0.1720	0.1613
94	4	2	25	16	15	0.2660	0.1702	0.1596
95	1	5	25	16	15	0.2632	0.1684	0.1579
96	1	1	25	16	15	0.2604	0.1667	0.1563
97	6	4	26	16	15	0.2680	0.1649	0.1546
98	5	6	27	16	15	0.2755	0.1633	0.1531
99	6	5	28	16	15	0.2828	0.1616	0.1515
100	3	6	29	16	15	0.2900	0.1600	0.1500

Table 1.1 (*continued*)

Looking at the relative frequencies of the events A, B, and C in Table 1.1, we observe that, although these tend to fluctuate initially a lot, as the number of trials performed grows, they seem to be concentrated around a fixed value. If we assume for now that the relative frequency converges to a certain limit, then this limit is called the **limiting relative frequency** of the event. It should be mentioned here that the fact that relative frequencies do converge to a limit is a consequence of one of the key theorems in probability theory, the *"law of large numbers"* (a visual illustration of this is given in Figure 1.14).

In the beginning of the twentieth century, the limit of the relative frequency was used by mathematicians for the definition of probability. In particular, Richard von Mises (1883–1953) gave the following definition of probability, known as the **statistical, or frequentist, definition of probability**.

Definition 1.9 (Frequentist definition of probability) Let Ω be a sample space for a chance experiment and A be an event on that space. If n_A denotes the number of times that A has appeared in n repetitions of the experiment ($0 \le n_A \le n$), then we define as the **probability of the event** A the limit

$$P(A) = \lim_{n\to\infty} \frac{n_A}{n} = \lim_{n\to\infty} f_A.$$

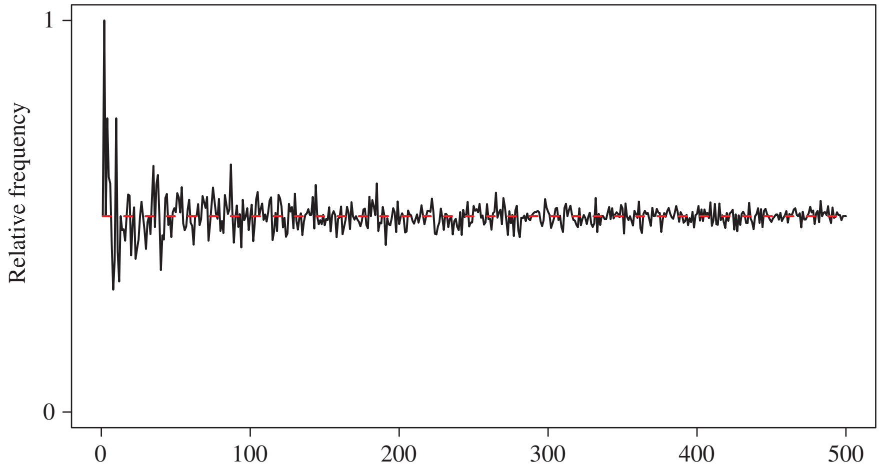


Figure 1.14 Long-run behavior of relative frequencies.

It is worth noting that, although not explicit in the notation, the quantities n_A in the above definition, depend on the number n of repetitions which have been carried out for the experiment that we consider.

We may assume for now that Definition 1.9 can be used to associate probabilities to events on a sample space. Then, by an appeal to the properties listed in Proposition 1.3, and in particular by taking the limit $n \to \infty$ there (that is, assuming that the number of experiments grows indefinitely), we obtain the following properties for probability:

FP1. $P(A) \geq 0$ for any event A on the sample space Ω;

FP2. $P(\Omega) = 1$;

FP3. For any pair of events A, B which are disjoint, we have

$$P(A \cup B) = P(A) + P(B).$$

As we will see in the next section, these three properties form the basis on which a rigorous mathematical framework can be built; this framework enables us to tackle both theoretical and practical problems associated with chance experiments.

Example 1.9 In a small city, there are just two local newspapers, called "The Weekly News" and the "News Herald." We asked 2000 inhabitants of this city which, if any, of these newspapers they read regularly. 460 persons answered they read "The Weekly News," 580 answered they read the "News Herald," while 140 answered that they read both. Assuming that the number of repetitions, $n = 2000$, is large enough so that the relative frequency of an event has approached its true value (limit), find the probability that if we select a person at random from this city, then he or she will

(i) be a reader of "The Weekly News";

(ii) be a reader of "News Herald";

(iii) read both newspapers;

(iv) read at least one newspaper;

(v) read the "The Weekly News" only;

(vi) be a reader of "News Herald" only;

(vii) will be a reader of exactly one local newspaper.

SOLUTION We consider the experiment of asking a person at random which newspaper(s) he/she reads. Let us define the events

A: a person is a reader of "The Weekly News"

B: a person is a reader of "News Herald."

The data in the example come from 2000 repetitions of the experiment. Based on this information, we have for the frequencies of the events A, B and their intersection

$$n_A = 460, \quad n_B = 580$$

and

$$n_{AB} = 140.$$

For the relative frequencies, we have

$$f_A = \frac{460}{2000} = 0.23 = 23\%, \quad f_B = \frac{580}{2000} = 0.29 = 29\%, \quad f_{AB} = \frac{140}{2000} = 0.07 = 7\%.$$

Since we have assumed that $n = 2000$ is a sufficiently large number of repetitions, we may estimate the various probabilities by the corresponding relative frequencies. We thus have the following results:

(i) The probability that someone reads the "The Weekly News" is $P(A) = 0.23$.

(ii) The probability that someone reads the "News Herald" is $P(B) = 0.29$.

(iii) The probability that someone reads both newspapers is $P(AB) = 0.07$.

(iv) Here we seek the probability of the event $A \cup B$. We use again the relative frequency as an estimate for this. The number of people who read exactly one newspaper, out of the 2000 persons we asked, is

$$n_{A\cup B} = 460 + 580 - 140 = 900$$

(because when we add $460 + 580$ we have counted the persons who read both newspapers twice). Thus,

$$P(A \cup B) = f_{A\cup B} = \frac{n_{A\cup B}}{n} = \frac{900}{2000} = 0.45;$$

(v) The number of persons who read “The Weekly News” only equals the number of persons who said that they read that newspaper minus the 140 persons who read both newspapers. Consequently,

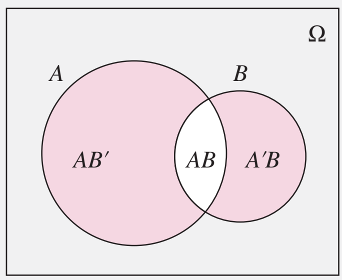
$$P(AB') = f_{AB'} = \frac{n_{AB'}}{n} = \frac{460 - 140}{2000} = 0.16.$$

(vi) Similarly, we have for the number of persons who read the “News Herald” only

$$P(A'B) = f_{A'B} = \frac{n_{A'B}}{n} = \frac{580 - 140}{2000} = 0.22;$$

(vii) The event “someone reads only one of the two newspapers” is expressed as $(AB') \cup (A'B)$. Since the events AB' and $A'B$ are disjoint, we obtain from Property *FP3* that

$$P(AB' \cup A'B) = 0.16 + 0.22 = 0.38.$$



EXERCISES

In the following exercises, assume that the number of repetitions is large enough to secure that the probability in question can be found using the associated relative frequency.

1. Use the first 50 throws from the data in Table 1.1 to calculate the probabilities of the events A, B, and C in Example 1.8. Then, repeat the same calculations using the second half of the throws (numbered 51–100 on the table). What do you observe?

2. For the experiment of throwing a die twice, use the results from Table 1.1 to calculate the probabilities of the following events:

   D: at least one of the two outcomes is 3;

   E: the outcome of the second throw is 5;

   F: the results of the two throws are the same.

   How do the probabilities of the events D and E compare to those of the events A and B in Example 1.8?

3. In a study of religious habits of persons living in a city, 500 persons were asked whether they go to church regularly. 240 of them were men, out of which 40 said that they attend a church service regularly, while among women, 80 said that they attend a church service regularly. Consider the following events, related to a randomly selected person among the persons who took part in this study:

   A: the person selected is male;

   B: the person selected goes to church.

   Express, in terms of the events A and B, each of the following events, and calculate their probabilities:

   (a) the person selected is male and attends church services regularly;

   (b) the person selected is female and attends church services regularly;

   (c) the person selected does not go to church.

4. In a survey conducted to examine whether there is correlation between gender and physical exercise, 700 persons were selected at random and their gender and whether they did some form of physical exercise at least twice a week or not were recorded. The outcomes are given in the following table:

	Physical exercise?	
Gender	Yes	No
Male	80	300
Female	75	245

   Consider the following events, associated with the experiment of selecting randomly a person among the 700 persons who took part in this survey:

   A: the person selected is a man;

   B: the person selected does physical exercise at least twice a week.

   Based on the data from the above table, calculate the probabilities of the events

$$A, B, AB, A', B', A'B', AB', A'B, A'B \cup AB', AB \cup A'B'.$$

5. In a University class, there are 200 students, who at the end of the last semester took three exams: Algebra (A), Calculus (C), and Probability (P). The numbers of students who obtained a first class mark were 24, 12, and 33 for A, C, and P, respectively. The numbers of students who obtained a first class mark in both A and C was 7, 10 students got a first class mark in A and P, and 6 students got a first class mark in C and P. If we select a student at random, find the probability that this student will have a first class mark

   (i) only in Algebra;

   (ii) only in Probability;

(iii) in both Probability and Algebra, but not in Calculus;

(iv) in both Calculus and Probability, but not in Algebra;

(v) in all three courses;

(vi) in none of these courses;

(vii) in exactly one of the three courses.

6. When Jenny returns home from her local supermarket, she has to pass through three sets of traffic lights. During the last 50 times she did that, she had to stop her car at the first set of traffic lights on 30 occasions, while 10 times the first traffic light was green but she had to stop at the second one. Furthermore, 6 times when both the first and the second lights were green, she had to stop at the third. Next time she returns from the supermarket, what is the probability that she has to stop in at least one of the three sets of traffic lights?

1.4 AXIOMATIC DEFINITION OF PROBABILITY

We have seen in the previous section that Definition 1.9 overcomes the difficulties encountered with Laplace's definition of probability. However, the approach by von Mises could not form the basis for a formal mathematical development of probability theory, as it is based on experimental data. Some of the problems which may arise if we use Definition 1.9 to define probability are as follows:

- On several occasions, the nature of the experiment is such that it may not be reasonable or even possible to repeat it a large number of times. Consider, for instance, the probability of the event that a political party wins the forthcoming election; or that a new spacecraft will land successfully on Mars. In some other occasions, when repetitions of an experiment are possible, it may be too costly or time consuming.
- When it is feasible to have a large number of repetitions, there appears to be no systematic way to control the error which arises when we use a relative frequency as an approximation for the probability of an event.
- In Definition 1.9, probability is defined as a limit as $n \to \infty$. How do we know that such a limit always exists? One way to get around this is to *assume* that the limit exists, as part of that definition; however, this is a rather strong assumption to be used as an axiom. Moreover, in practice, we can never have an infinite number of experiments and, especially when precision is essential, it may not be easy to "guess" what the value of the limit
$$\lim_{n\to\infty} f_A = \lim_{n\to\infty} \frac{n_A}{n}$$
is, when we have at hand only a finite number of repetitions n for the experiment.

The aforementioned concerns initiated a search for alternative definitions of probability. In fact, in a famous lecture delivered in 1900 at the *International Congress of Mathematicians* held in Paris, the German mathematician David Hilbert (1862–1943), who was perhaps the leading mathematician of his time, proposed 23 problems whose solution presented a major challenge and would form a breakthrough for mathematics of the twentieth century. One of these problems concerned the axiomatic foundation of probability theory. It was only in 1933 when the Russian mathematician Andrei Kolmogorov (1903–1989) developed a consistent framework within which any problem regarding probabilities could be, if not solved, at least formulated. In Kolmogorov's work, three "obvious" and "undoubtable" properties of probability are taken as axioms. Based on them, the entire theory of probability can be developed, making a series of mathematical and logical deductions. It is worth noting at this point that, using tools from Kolmogorov's axiomatic foundation of probability, we can prove the existence of the limit (in an appropriate sense) for the relative frequency of an event when the number of repetitions tends to infinity. Thus, Kolmogorov's theory does not contradict the statistical definition by von Mises; rather, it incorporates it by providing a formal reasoning for its validity.

We are now ready to give the definition of probability, based on Kolmogorov's approach, which will be used throughout this book.

Definition 1.10 Assume that Ω is a sample space for a chance experiment. Assume also that to each event A in Ω, there corresponds a real number, $P(A)$. If the function $P(\cdot)$ satisfies the following three axioms, then P will be called a probability on the sample space Ω, while the number $P(A)$ will be referred to as the probability of the event A:

*P*1. $P(A) \geq 0$ for any event A defined in the sample space Ω;

*P*2. $P(\Omega) = 1$;

*P*3. If $A_1, A_2, \ldots$ is a sequence of events in Ω which are pairwise disjoint (that is, $A_iA_j = \emptyset$ for any $i \neq j$), then

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \qquad (1.2)$$

According to this definition, the probability is considered as a **set function**, i.e. a function which associates any event A defined in Ω to a real number. Looking at Kolmogorov's axioms, and comparing them with Properties (*FP1–FP3*) stated in the previous section as direct consequences of von Mises' definition of probability, we see that they are not that different! In fact, the first two are the same, while $P3$ in Definition 1.10 includes *FP3* as a special case (just take $A_3 = A_4 = \cdots = \emptyset$).

Property $P3$ is known as the **additivity property of probability**. More precisely, as it involves an infinite summation (a sum of countably many terms), it is usually called the **countable additivity** axiom of probability. If we put $A_1 = A_2 = A_3 = \cdots = \emptyset$ in (1.2), we obtain immediately

$$P(\emptyset) = P(\emptyset) + P(\emptyset) + P(\emptyset) + \cdots ,$$

so that subtracting $P(\emptyset)$ on either side yields

$$0 = P(\emptyset) + P(\emptyset) + \cdots .$$

Since $P(\emptyset)$ is finite and uniquely defined, it has to be zero, that is,

$$P(\emptyset) = 0.$$

Next, if we put

$$A_{n+1} = A_{n+2} = \cdots = \emptyset$$

in (1.2), we arrive at the following result, known as the **finite additivity property**.

Proposition 1.4 *If $A_1, A_2, \ldots, A_n$ are n pairwise disjoint events in a sample space Ω (that is, $A_iA_j = \emptyset$ for any $i \neq j$), then*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i). \tag{1.3}$$

For $n = 2$, as already mentioned, we obtain Property *FP3* of the previous section; namely, if A and B are disjoint events, then

$$P(A \cup B) = P(A) + P(B). \tag{1.4}$$

We have already seen some examples of obtaining new results with the aid of the three axioms in Definition 1.10. However, the definition seems to offer no obvious way to *calculate* the probability $P(A)$ associated with a certain event A. This will be studied and illustrated in great detail in subsequent chapters of the book.

In the next propositions, we develop some useful properties for probabilities, which are easily deduced from the axioms of Definition 1.10.

Proposition 1.5 *Let $A = \{a_1, a_2, \ldots\}$ be a countably infinite subset of the sample space Ω. Then, the probability of the event A is*

$$P(A) = \sum_{i=1}^{\infty} P(A_i),$$

where $P(A_i)$ denotes the probability of the elementary event $\{a_i\}$. If A has finitely many elements, so that $A = \{a_1, a_2, \ldots, a_n\}$, then

$$P(A) = \sum_{i=1}^{n} P(A_i).$$

Proof: The first statement is obvious from (1.2) by putting $A_i = \{a_i\}$ for $i = 1, 2, \ldots,$ and observing that

$$\bigcup_{i=1}^{\infty} A_i = A \quad \text{and} \quad A_iA_j = \emptyset \quad \text{for } i \neq j.$$

For the second statement, put similarly $A_i = \{a_i\}$ for $i = 1, 2, \ldots, n$ in Proposition 1.4.

□

Applying the previous proposition to the case $A = \Omega$, when Ω is a finite or countably infinite sample space, we immediately obtain the following.

Corollary 1.1 *Let $\Omega = \{\omega_1, \omega_2, \ldots\}$ be a countably infinite sample space. Then the probabilities $p_i = P(\{\omega_i\})$ associated with the elementary events $\{\omega_i\}$ satisfy the condition*

$$\sum_{i=1}^{\infty} p_i = 1.$$

For the case when Ω is finite so that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, this reduces to

$$\sum_{i=1}^{n} p_i = 1.$$

There are certain cases where it is rather complicated to calculate the probability for an event A, but is easier to find the probability of its complement, A'. The following result is then useful.

Proposition 1.6 *Let A be an arbitrary event in a sample space Ω. Then, the probability of its complement, A', is given by*

$$P(A') = 1 - P(A). \tag{1.5}$$

Proof: This is obvious from (1.4) and the fact that

$$P(\Omega) = P(A \cup A') = P(A) + P(A'),$$

which holds since $A \cup A' = \Omega$ and A and A' are disjoint sets. □

Example 1.10 Suppose the probability of an event A is larger by 0.5 than the probability of its complement. Find $P(A)$.

SOLUTION We are given that

$$P(A) = P(A') + 0.5,$$

and replacing $P(A')$ by Proposition 1.5 gives

$$P(A) = 1 - P(A) + 0.5 = 1.5 - P(A).$$

Solving this, we immediately obtain $P(A) = 0.75$.

Example 1.11 Suppose the possible outcomes of an experiment are all the positive integers, while for $i = 1, 2, \ldots$, the probability that the outcome of the experiment is i is twice as much as the probability that the outcome is $i + 1$. Then, calculate

(i) the probabilities of all elementary events of this experiment;

(ii) the probability that the outcome is an even integer;

(iii) the probability that the outcome is an odd integer.

SOLUTION The sample space for this experiment is the set $\Omega = \{1, 2, \ldots\}$, and so an elementary event is a set $\{i\}$, where i is a positive integer. For any such i, we are given that

$$P(\{i\}) = 2P(\{i + 1\}).$$

Denoting $p_i = P(\{i\})$ for simplicity in notation, this reads

$$p_i = 2p_{i+1}, \quad i = 1, 2, \ldots .$$

Thus, we have, for example, $p_1 = 2p_2, p_2 = 2p_3$, and so on.

(i) Applying the last equation recursively, we can express the probability of any elementary event, p_i, in terms of p_1. More specifically,

$$p_1 = 2p_2 = 2(2p_3) = 4p_3 = 4(2p_4) = 8p_4 = \ldots ,$$

and in general $p_1 = 2^{i-1}p_i$ for $i = 1, 2, \ldots$, which can be rewritten as

$$p_i = 2^{-(i-1)}p_1 = \left(\frac{1}{2}\right)^{i-1} p_1, \quad i = 1, 2, \ldots .$$

Since Ω is countably infinite, we may use the first part of Corollary 1.1. Replacing the probabilities p_i there by the last expression, we deduce that

$$1 = \sum_{i=1}^{\infty} p_i = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} p_1 = p_1 \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} .$$

The quantity

$$\sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1}$$

is the sum of a geometric series,[5] and so we obtain

$$p_1 \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{i-1} = \left(1 + \frac{1}{2} + \frac{1}{2^2} + \frac{1}{2^3} + \cdots \right) p_1 = \frac{1}{1 - \frac{1}{2}} p_1 = 2p_1.$$

Since this must equal 1, we get $p_1 = 1/2$, and the probabilities of the elementary events are then given by

$$p_i = \left(\frac{1}{2}\right)^{i-1} p_1 = \left(\frac{1}{2}\right)^{i}, \quad i = 1, 2, \ldots .$$

[5]The main results for a geometric series, for readers not familiar with it, are listed in the Appendix A, along with some other formulas from calculus which are needed for the discussions in this book.

(ii) The probability that the outcome of the experiment is an even integer is given by the sum

$$\sum_{i=1}^{\infty} p_{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{2}\right)^{2i} = \sum_{i=1}^{\infty} \left(\frac{1}{4}\right)^{i}.$$

This is another sum of a geometric series, and so is equal to

$$\frac{1}{4} + \frac{1}{4^2} + \frac{1}{4^3} + \cdots = \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{1}{3}.$$

(iii) If A denotes the event considered in part (ii), viz., that the outcome is an even integer, then the event that the outcome is an odd integer is simply A', and so we immediately have

$$P(A') = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}.$$

Alternatively, the probability of the event A' could be obtained directly as follows:

$$\begin{aligned} P(A') &= p_1 + p_3 + p_5 + \cdots = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots \\ &= \frac{1}{2}\left(1 + \frac{1}{2^2} + \frac{1}{2^4} + \cdots\right) = \frac{1}{2}(1 + P(A)) \\ &= \frac{1}{2} \cdot \frac{4}{3} = \frac{2}{3}. \end{aligned}$$

We close this section by a brief reference to a rather subtle issue of the axiomatic foundation of probability, stemming from Definition 1.10. We have mentioned already that if the sample space Ω has at most countably many elements, then we can safely associate a probability to *any* subset A of Ω; for sample spaces with uncountably many elements, there may be subsets of Ω for which this is not possible. Such subsets are called *nonmeasurable* and are not considered here. Excluding such sets from the domain of the probability function $P(\cdot)$, we see that this set function is defined on a family of subsets of Ω, say $\mathcal{A}$, the members of which are events on the sample space. The minimum requirements that $\mathcal{A}$ must satisfy so that Definition 1.10 can be applied are the following:

$\Sigma 1$: $\Omega \in \mathcal{A}$;

$\Sigma 2$: If $A \in \mathcal{A}$, then $A' \in \mathcal{A}$;

$\Sigma 3$: If $A_i \in \mathcal{A}$ for $i = 1, 2, \ldots$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

A family of subsets of the sample space Ω with the above properties is called a σ-**field** or a σ-**algebra**. Many advanced books on probability theory begin by considering such families in order to introduce probability in a more rigorous manner from a mathematical viewpoint. Several such textbooks can be found in the list of references at the end of this book, and the interested reader may refer to any of these books for more details.

EXERCISES

Group A

1. If for the event A we have $2P(A) = P(A') + 0.5$, find the probability $P(A)$.

2. At 08:00 a.m., the probability that John is in bed is 0.2, while the probability of him having breakfast is 0.5. What is the probability that on a particular day he is neither in bed nor having breakfast at 08:00 a.m.?

3. The probability that the price of a certain stock increases during a day is 50% higher than the probability that the price of the stock decreases, while it is also three times as much as the probability that the price remains the same. Find the probability that, during a day, the price of the stock (i) increases; (ii) decreases; (iii) does not change.

4. When a salesman visits a certain town, he stays in one of three available hotels, H_1, H_2, H_3. Let A_i be the event that he stays in Hotel H_i, for $i = 1, 2, 3$. If it is known that

$$P(A_1 \cup A_2) = 2P(A_3) \quad \text{and} \quad 3P(A_2) = 1 - \frac{P(A_3)}{2},$$

find the probability that he stays in Hotel H_i, for $i = 1, 2, 3$.

5. Let $A_1, A_2, \dots, A_n$ be a finite collection of events on a sample space Ω such that $A_1 \cup A_2 \cup \cdots \cup A_n = \Omega$ and the A_i's are pairwise disjoint. If it is known that $P(A_i) = 3P(A_{i+1})$ for $i = 1, 2, \dots, n-1$, find $P(A_3)$ and $P(A_5)$ (assuming $n \geq 5$).

6. Let A and B be two disjoint events in a sample space Ω. Prove that

$$P(A'B') = P(A') + P(B') - 1.$$

7. Let A, B, and C be three events in a sample space Ω such that $AB = \emptyset$. Then:
   (i) Verify that the events AC and BC are disjoint.
   (ii) Calculate the probability of the event

$$(A' \cup C')(B' \cup C')$$

   in terms of the probabilities $P(AB)$ and $P(AC)$.

Group B

8. Consider the infinite countable sample space $\Omega = \{0, 1, 2, \dots\}$. On this space, assume that the probabilities of the elementary events $\{i\}$, for $i = 0, 1, \dots$, are given by

$$P(\{i\}) = \frac{3^i}{4^{i+1}}, \quad i = 0, 1, 2, \dots .$$

   (i) Show that the probability that the outcome of the experiment is at least k, for $k = 1, 2, \dots$, is given by $(3/4)^k$.
   (ii) What is the probability that we get an outcome less than k for $k = 1, 2, \dots$?

9. Let A_1, A_2, A_3, and A_4 be four events in a sample space Ω that are pairwise disjoint and such that

$$\bigcup_{i=1}^{4} A_i = \Omega.$$

If it is known that

$$P(A_2) = 2P(A_1), \quad P(A_4) = 3P(A_2), \quad P(A_2) = 4P(A_3),$$

calculate the probabilities of the union of any two and any three events among A_1, A_2, A_3, A_4.

10. For two disjoint events A and B in a sample space, suppose

$$P(A \cup B) = \frac{1}{2}, \quad 3P(A') + 2P(B) = 3.$$

Then, find the probabilities $P(A)$ and $P(B)$.

11. Consider the continuous sample space $\Omega = [1, 1000]$. For each event A in Ω ($A \subseteq \Omega$), we define the probability $P(A)$ to be $k/1000$, where k is the number of integers included in the set A. Then:

(i) Show that $P(\cdot)$ satisfies the three properties of Definition 1.10.

(ii) Calculate the probabilities of the events

$$A = [100, 200], \quad B = [1, 100] \cup [900, 1000], \quad C = \bigcup_{i=1}^{9} \left[100i, 100i + \frac{100}{2^{i+1}}\right].$$

1.5 PROPERTIES OF PROBABILITY

In this section, we look at various properties of probability. We have already seen, in the last section, results which follow from the conditions we set in Definition 1.10. Further results in this section, which are important for subsequent developments in this book, may illustrate the development of a chain of results from Kolmogorov's axioms. A major tool for the proofs in this section is the additive property of probability, given in Proposition 1.4. Note, however, that this applies to pairwise disjoint events. Therefore, in order to use that proposition, we must first express our event of interest as the union of two or more events which are pairwise disjoint. This is in fact the technique used in the proofs of the ensuing propositions.

Proposition 1.7 *For any events A and B in a sample space Ω, we have*

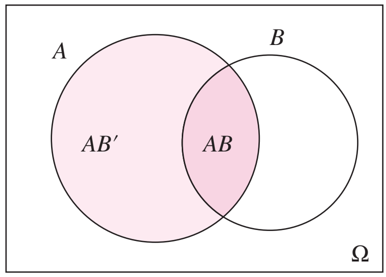
$$P(A - B) = P(AB') = P(A) - P(AB). \tag{1.6}$$

For the special case where $B \subseteq A$, we have that

$$P(A - B) = P(A) - P(B).$$

Proof: Recall that the set $A - B$ contains all elements of A which are not contained in B. This means that the events $A - B$ and AB' are identical, and so they must have the same probability, and this establishes the first part of (1.6). For the second part, note that the events AB and AB' are disjoint, while their union equals

$$(AB) \cup (AB') = A.$$



In view of the finite additivity property, we thus have

$$P(AB) + P(AB') = P(A),$$

and the second equation in (1.6) is now obvious.

Turning to the case where $B \subseteq A$, we then have $AB = B$, and so (1.6) immediately implies that $P(A - B) = P(A) - P(B)$. □

Proposition 1.8 ***(Monotonicity of probability)***

(i) *Let A and B be events in a sample space Ω such that $B \subseteq A$. Then,*

$$P(B) \leq P(A).$$

(ii) *For any event A on Ω, we have*

$$P(A) \leq 1.$$

Proof:

(i) When $B \subseteq A$, Proposition 1.7 yields

$$P(A - B) = P(A) - P(B).$$

Since the left-hand side is nonnegative, so must be the right-hand side. Thus $P(A) - P(B) \geq 0$, that is,

$$P(B) \leq P(A).$$

(ii) This is obvious from Part (i), the fact that for any event A on the sample space we have $A \subseteq \Omega$, and Property $P2$ of Definition 1.10. □

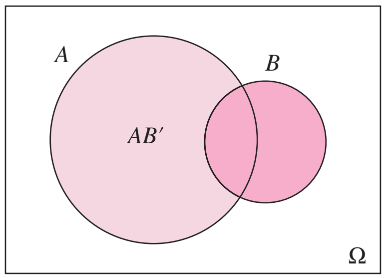
The following result is a key identity which generalizes (1.4), as it enables us to calculate the probability of a union of two events which are not necessarily disjoint.

Proposition 1.9 *For any events A and B on a sample space, we have*

$$P(A \cup B) = P(A) + P(B) - P(AB). \tag{1.7}$$

Proof: As stated above, we know how to deal with the probability of the union between two events only when these events are disjoint. We therefore try to express $A \cup B$ as the union of two such events. This is easily accomplished by considering the events B and AB'; these have no common element between them, while

$$A \cup B = AB' \cup B.$$



From the additivity property, we readily have

$$P(A \cup B) = P(AB' \cup B) = P(AB') + P(B).$$

Upon replacing $P(AB')$ from the result stated in Proposition 1.7, we get

$$P(A \cup B) = P(A) + P(B) - P(AB),$$

as required. □

Example 1.12 A store for women's clothes sells both shoes and handbags. The store manager has estimated that 20% of the customers who enter the shop buy a pair of shoes, 30% purchase a handbag while 10% of the customers buy both a pair of shoes and a handbag. Based on these assertions, estimate the proportions of customers who buy

(i) a pair of shoes only,
(ii) a handbag only,
(iii) at least one of these two items,
(iv) exactly one of these items.

SOLUTION Let us define the events

A: a person buys a pair of shoes from the store,

B: a person buys a handbag from the store.

Then, based on the manager's assertions, we have

$$P(A) = 0.20, \quad P(B) = 0.30, \quad P(AB) = 0.10.$$

The required probabilities are thus found as follows:

(i) $P(AB') = P(A) - P(AB) = 0.20 - 0.10 = 0.10 = 10\%$;

(ii) $P(A'B) = P(B) - P(AB) = 0.30 - 0.10 = 0.20 = 20\%$;

(iii) $P(A \cup B) = P(A) + P(B) - P(AB) = 0.20 + 0.30 - 0.10 = 0.40 = 40\%$;

(iv) $P(AB' \cup A'B) = P(AB') + P(A'B) = 0.10 + 0.20 = 0.30 = 30\%$.

Proposition 1.9 enables us to calculate the probability that at least one of two events occurs. Often, in practice, we have more than two events and we are interested in the probability that at least one of them occurs. When we have three events, say A_1, A_2, and A_3, we may apply Proposition 1.9 twice to obtain the corresponding result. More explicitly, we have

$$\begin{aligned}
P(A_1 \cup A_2 \cup A_3) &= P[A_1 \cup (A_2 \cup A_3)] \\
&= P(A_1) + P(A_2 \cup A_3) - P[A_1(A_2 \cup A_3)] \text{ by Proposition 1.9} \\
&= P(A_1) + P(A_2 \cup A_3) - P(A_1A_2 \cup A_1A_3) \text{ by } SP6 \text{ of Proposition 1.1} \\
&= P(A_1) + P(A_2) + P(A_3) - P(A_2A_3) \\
&\quad - [P(A_1A_2) + P(A_1A_3) - P(A_1A_2A_1A_3)] \text{ by Proposition 1.9} \\
&= P(A_1) + P(A_2) + P(A_3) - [P(A_1A_2) + P(A_1A_3) + P(A_2A_3)] \\
&\quad + P(A_1A_2A_3).
\end{aligned}$$

For the general case when we have n events, the following result, known as the **Poincaré formula**[6] or the **inclusion–exclusion formula**, holds.

Proposition 1.10 ***(Inclusion–exclusion formula)*** *Let $A_1, A_2, \dots, A_n$ be n events in a sample space Ω. Then, the probability that at least one of them occur is given by*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = S_1 - S_2 + S_3 - \cdots + (-1)^{n-1} S_n = \sum_{r=1}^{n} (-1)^{r-1} S_r,$$

where

$$S_1 = \sum_{i=1}^{n} P(A_i),$$

$$S_2 = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} P(A_iA_j),$$

[6]The formula is named after Jules Henri Poincaré (1854–1912), a French mathematician, physicist, engineer as well as a distinguished philosopher of science and mathematics.

$$S_3 = \sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} \sum_{k=j+1}^{n} P(A_i A_j A_k),$$

$$\vdots$$

$$S_r = \sum_{i_1=1}^{n-r+1} \sum_{i_2=i_1+1}^{n-r+2} \cdots \sum_{i_r=i_{r-1}+1}^{n} P(A_{i_1} A_{i_2} \cdots A_{i_r}),$$

$$\vdots$$

$$S_n = P(A_1 A_2 \cdots A_n).$$

Proof: The proof can be established by induction, using the result of Proposition 1.9. You may try proving it! □

Example 1.13 An insurance company offers its clients coverage against three different risks, labeled *I*, *II*, and *III*. The following table shows the percentage of insured clients who had opted to insure against one of these risks (diagonal elements in the table), as well as those who had insured against more than one of these risks (for instance, 17% of the clients had insured against risks *I* and *II*). We also know that the percentage of clients who had insured for all three risks is 5%.

	I	*II*	*III*
I	32%	17%	12%
II		23%	9%
III			18%

Calculate the percentage of clients who had insured

(i) for at least one of the three risks;

(ii) for none of the three risks;

(iii) for exactly two among the three risks.

SOLUTION We consider the events

A_i: a customer is insured against risk i,

for $i = 1, 2, 3$. According to the values in the table, we have

$$\begin{array}{lll} P(A_1) = 0.32, & P(A_2) = 0.23, & P(A_3) = 0.18, \\ P(A_1A_2) = 0.17, & P(A_1A_3) = 0.12, & P(A_2A_3) = 0.09. \end{array}$$

In addition, we have $P(A_1A_2A_3) = 0.05$.

(i) For this part, we use the inclusion–exclusion formula, since we seek the probability $P(A_1 \cup A_2 \cup A_3)$ (of course, this agrees with the formula for $n = 3$, given just before Proposition 1.10, so we could have used this formula instead). In this way, we get

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{r=1}^{3} (-1)^{r-1} S_r = S_1 - S_2 + S_3.$$

We now find S_r, for $r = 1, 2, 3$, as follows:

$$S_1 = \sum_{i=1}^{3} P(A_i) = P(A_1) + P(A_2) + P(A_3) = 0.32 + 0.23 + 0.18 = 0.73,$$

$$\begin{aligned} S_2 = \sum_{i=1}^{2} \sum_{j=i+1}^{3} P(A_i A_j) &= P(A_1A_2) + P(A_1A_3) + P(A_2A_3) \\ &= 0.17 + 0.12 + 0.09 = 0.38, \end{aligned}$$

and finally

$$S_3 = P(A_1A_2A_3) = 0.05.$$

We thus find the required probability to be

$$P\left(\bigcup_{i=1}^{n} A_i\right) = 0.73 - 0.38 + 0.05 = 0.40.$$

(ii) Here, we want to calculate the probability of the event $A_1'A_2'A_3'$. But, by De Morgan identity, this is the complement of the event $A_1 \cup A_2 \cup A_3$, and consequently

$$P(A_1'A_2'A_3') = 1 - P(A_1 \cup A_2 \cup A_3) = 1 - 0.40 = 0.60.$$

(iii) The event we are interested in can be expressed as

$$E = \underbrace{A_1A_2A_3'}_{\text{only } A_1 \text{ and } A_2 \text{ occur}} \cup \underbrace{A_1A_2'A_3}_{\text{only } A_1 \text{ and } A_3 \text{ occur}} \cup \underbrace{A_1'A_2A_3}_{\text{only } A_2 \text{ and } A_3 \text{ occur}}.$$

This is the union of three events which are pairwise disjoint; therefore,

$$P(E) = P(A_1A_2A_3') + P(A_1A_2'A_3) + P(A_1'A_2A_3).$$

From Proposition 1.7, we find the probabilities on the right hand side to be

$$P(A_1A_2A_3') = P((A_1A_2)A_3') = P(A_1A_2) - P(A_1A_2A_3) = 0.17 - 0.05 = 0.12,$$

$$P(A_1A_2'A_3) = P((A_1A_3)A_2') = P(A_1A_3) - P(A_1A_2A_3) = 0.12 - 0.05 = 0.07$$

and

$$P(A_1'A_2A_3) = P((A_2A_3)A_1') = P(A_2A_3) - P(A_1A_2A_3) = 0.09 - 0.05 = 0.04.$$

We thus obtain finally the required probability to be

$$0.12 + 0.07 + 0.04 = 0.23,$$

i.e. there is a 23% chance for a client to be insured against exactly two among the three risks.

EXERCISES

Group A

1. For a specific area, the probability that it is hit by a typhoon during a year is 0.02. For the same period, the probability that the area suffers severe damage due to excessive rain is 0.05, while the probability that both these events occur in a year is 0.001. Find the probability that over a particular year, the area
   (i) is hit by a typhoon but there is no excessive rain;
   (ii) experiences only excessive rain, but is not affected by a typhoon;
   (iii) experiences neither excessive rain nor the occurrence of a typhoon;
   (iv) has excessive rain or is hit by a typhoon.

2. When Sandra drives back home from work, she has to pass through two sets of traffic lights. The probability that she has to stop at the first is 0.35, while the probability that she has to stop at the second one is 0.40. Finally, the probability that she does not stop at a traffic light is 0.50. Calculate the probability that she has to stop
   (i) at both sets of traffic lights;
   (ii) at least at one set of traffic lights;
   (iii) at exactly one set of traffic lights.

3. Let A and B be two events in a sample space Ω. Establish the relationship

$$\begin{aligned} P(A')P(B) - P(A'B) = P(A)P(B') - P(AB') = P(A'B') - P(A')P(B') \\ = P(AB) - P(A)P(B). \end{aligned}$$

4. Suppose for the events A and B we have

$$2P(A) = 3P(B) = 4P(AB) \quad \text{and} \quad P(A'B) = 0.05.$$

   (i) Calculate the probabilities $P(A)$, $P(B)$, and $P(AB)$.
   (ii) Find the probabilities of the following events

$$A \cup B, \quad A'B', \quad AB', \quad AB' \cup A'B, \quad A \cup B', \quad A' \cup B.$$

5. Let A and B be two events in a sample space Ω with $P(A) = \alpha$ and $P(B) = \beta$. Confirm the validity for each of the following:

   (i) $P(A'B') = 1 - \alpha - \beta + P(AB)$;

   (ii) $\alpha + \beta - 1 \leq P(AB) \leq \min\{\alpha, \beta\}$;

   (iii) $\max\{\alpha, \beta\} \leq P(A \cup B) \leq \alpha + \beta$.

   Suppose $P(A) = 0.75$ and $P(B) = 0.80$. What is the range of possible values for each of $P(A \cup B)$ and $P(AB)$?

6. Let A and B be two events in a sample space Ω such that $P(A) = 1$ and $P(B) = 1$. Use the results of Exercise 5 to show that

$$P(A \cup B) = 1, \quad P(AB) = 1, \quad P(A'B') = 0.$$

7. Let A and B be two events in a sample space Ω such that $P(A) = 0$ and $P(B) = 0$. Use the results of Exercise 5 to show that

$$P(A \cup B) = 0, \quad P(AB) = 0, \quad P(A'B') = 1.$$

8. Let A_1, A_2, and A_3 be three events in a sample space Ω, which are not necessarily pairwise disjoint.

   (i) Prove that

$$P(A_1 \cup A_2 \cup A_3) \leq P(A_1) + P(A_2) + P(A_3).$$

   (ii) Show that the equality

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3)$$

   holds if and only if

$$P(A_1A_2) + P(A_1A_3) + P(A_2A_3) = P(A_1A_2A_3).$$

Group B

9. With reference to the out-patient visits to a hospital, we consider the following events:

   A: a patient is over 50 years of age;

   B: a patient is under 25 years of age;

   C: the condition of a patient upon arrival is considered to be serious;

   D: a patient needs to stay at the hospital for at least one night.

   We are given the following probabilities associated with the above events:

$$P(A) = 0.52, \quad P(B) = 0.22, \quad P(C) = 0.25, \quad P(D) = 0.24,$$
$$P(AC) = 0.17, \quad P(AD) = 0.12, \quad P(BC) = P(BD) = 0.06,$$
$$P(CD) = 0.18, \quad P(ACD) = 0.08.$$

Express in words what each of the following events represents and find the probability of occurrence for them:

$$A'B', \quad (A \cup B)C, \quad AC' \cup D, \quad (A \cup B)D'.$$

10. Suppose A_1, A_2, and A_3 are three events in a sample space. Let us consider the sums

$$S_1 = P(A_1) + P(A_2) + P(A_3),$$
$$S_2 = P(A_1A_2) + P(A_1A_3) + P(A_2A_3),$$
$$S_3 = P(A_1A_2A_3).$$

Express mathematically, in terms of S_1, S_2, and S_3, the probability for the event that

(i) exactly one of the A_i occurs;
(ii) exactly two of the A_i occur;
(iii) none of the A_i occur.

Application: Among the customers of an insurance company, the percentages of those who have home insurance, liability insurance and car insurance are respectively, 35%, 15%, and 40%. Suppose 8% have both home and liability insurances, 25% have both home and car insurances, and 10% have both liability and car insurances. Finally, 3% of the customers have all three types of insurance.

For $i = 0, 1, 2$, calculate the percentage of persons who have exactly i types of insurance with the company.

11. Let A, B, and C be three events such that $C \subseteq B \subseteq A$. Let $a = P(A), b = P(B)$, and $c = P(C)$.

(i) Calculate the probabilities

$$P(A - B), \quad P(A - (B - C)), \quad P((A - B) - C)$$

in terms of a, b, c. What do you observe?

(ii) Obtain a numerical answer to the result in (i) for $a = 1/2, b = 1/4, c = 1/6$.

12. For the events A, B, and C of a sample space, we are given

$$P(ABC) = a, \quad P(A'BC) = b_1, \quad P(AB'C) = b_2, \quad P(ABC') = b_3,$$
$$P(AB'C') = c_1, \quad P(A'BC') = c_2, \quad P(A'B'C) = c_3.$$

Calculate the probabilities of the events A, B, and C in terms of a, b_i, and c_i. Similarly find for the events

$$A - B, \quad B - C, \quad (A - B) - C, \quad A - (B - C), \quad (A \cup BC)' - AC.$$

13. Consider the events $A_1, A_2, \ldots, A_n$ defined in a sample space Ω. Show that the probability that none of them appear is equal to

$$P(A_1' A_2' \cdots A_n') = \sum_{r=0}^{n} (-1)^r S_r,$$

where $S_1, S_2, \ldots, S_n$ are the sums defined earlier in Proposition 1.10, and $S_0 = 1$.

Application: For $n = 3$ assume that $P(A_i) = 1/3$ for $i = 1, 2, 3$. Moreover, assume that $P(A_i A_j) = 1/9$ for any $i \neq j$, and $P(A_1 A_2 A_3) = 1/27$. Then, show that

$$P(A_1' A_2' A_3') = P(A_1')P(A_2')P(A_3').$$

14. For the events A_1, A_2, and A_3 in the same sample space, it is known that

$$P(A_i) = \frac{1}{2^i}, \quad i = 1, 2, 3,$$

and

$$P(A_i A_{i+1}) = \frac{1}{2^{i+2}}, \quad i = 1, 2.$$

If, in addition, we know that the events A_1 and A_3 are mutually exclusive, calculate the probability that at least one of A_1, A_2, A_3 occur.

1.6 THE CONTINUITY PROPERTY OF PROBABILITY

We know from calculus that a real-valued function is continuous if it preserves limits. More precisely, a real function $f : \mathbb{R} \mapsto \mathbb{R}$ is continuous if and only if for each converging sequence of real numbers $(x_n)_{n \geq 1}$, we have

$$\lim_{n \to \infty} f(x_n) = f(\lim_{n \to \infty} x_n).$$

A similar property holds for the probability as a function. Note, however, that probability is not defined on the set of real numbers, but according to Definition 1.10, it is a set function. In order to consider continuity for such a function, we need to introduce the following concepts:

- A sequence of events $A_1, A_2, \ldots,$ defined in a sample space Ω, is said to be an **increasing sequence** if

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq A_n \subseteq A_{n+1} \subseteq \cdots.$$

(see Figure 1.15);

- A sequence of events $A_1, A_2, \ldots,$ defined in Ω, is said to be a **decreasing sequence** if

$$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots \supseteq A_n \supseteq A_{n+1} \supseteq \cdots.$$

(see Figure 1.16).

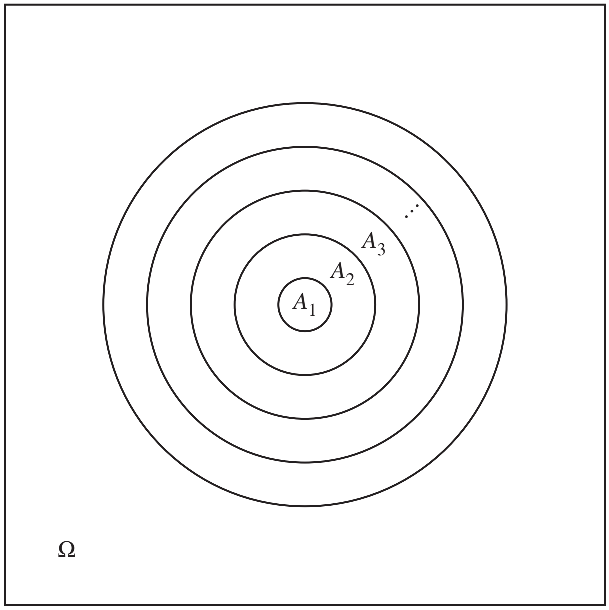


Figure 1.15 An increasing sequence of events.

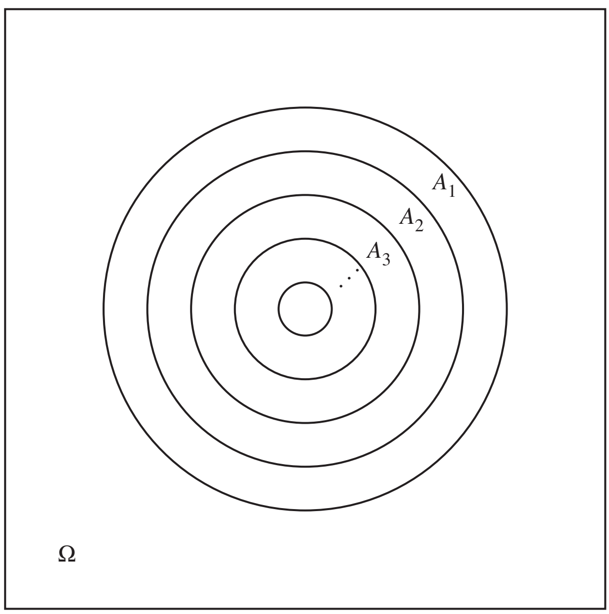


Figure 1.16 A decreasing sequence of events.

It is evident that for an increasing sequence $\{A_n\}_{n\geq 1}$, we have

$$A_n = \bigcup_{i=1}^{n} A_i, \quad n = 1, 2, \dots .$$

Similarly, for a decreasing sequence $\{A_n\}_{n\geq 1}$, we have

$$A_n = \bigcap_{i=1}^{n} A_i = A_1 A_2 \cdots A_n, \quad n = 1, 2, \dots .$$

If a sequence of sets is either increasing or decreasing, it is said to be a **monotone sequence**. For such a sequence, we define its **limit**, as follows:

$$\lim_{n\to\infty} A_n = \begin{cases} \bigcup_{i=1}^{\infty} A_i, & \text{if } \{A_n\}_{n\geq 1} \text{ is an increasing sequence} \\ \bigcap_{i=1}^{\infty} A_i & \text{if } \{A_n\}_{n\geq 1} \text{ is a decreasing sequence.} \end{cases}$$

The following example illustrates the above concepts.

Example 1.14 Consider the sample space $\Omega = \mathbb{R}$, the set of real numbers. In it, we define the sequences of events

$$A_n = \left[2 + \frac{1}{n}, 6 - \frac{1}{n}\right], \quad n = 1, 2, 3, \dots ,$$

and

$$B_n = \left[4 - \frac{1}{n}, 4 + \frac{1}{n}\right], \quad n = 1, 2, 3, \dots .$$

It is easy to see (see also Figure 1.17) that the sequence $\{A_n\}_{n\geq 1}$ is increasing, while the sequence $\{B_n\}_{n\geq 1}$ is decreasing. Further, as $n \to \infty$, the lower endpoint of the set A_n approaches 2, while the upper endpoint converges to 6. Thus,

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = [2, 6].$$

In an analogous manner, we obtain

$$\lim_{n\to\infty} B_n = \bigcap_{n=1}^{\infty} B_i = \{4\}.$$

We are now ready to prove the following proposition which shows that, when applied to a monotone sequence, probability preserves limits. Apart from its own interest, this is important in proving a number of other important results, as we shall see in later chapters.

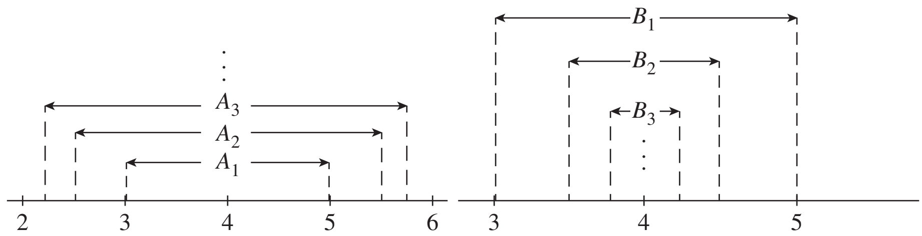


Figure 1.17 The sequences A_n, B_n, $n = 1, 2, \ldots$.

Proposition 1.11 *For any monotone sequence of events $\{A_n\}_{n\geq 1}$ of a sample space Ω, we have*

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

Proof: Suppose first that the sequence $\{A_n\}_{n\geq 1}$ is increasing. Then, the sets

$$B_1 = A_1, \quad B_2 = A_2 - A_1, \quad B_3 = A_3 - A_2, \ldots, \quad B_n = A_n - A_{n-1}, \ldots$$

are pairwise disjoint, and that they further satisfy the following equalities:

$$\bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n, \quad n = 1, 2, \ldots,$$

and

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i.$$

In words, each set B_n contains those elements of A_n which are not contained in any of the A_i, for $i = 1, 2, \ldots, n-1$. In view of Property *P3* of Definition 1.10, we now obtain

$$P(\lim_{n\to\infty} A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\bigcup_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} P(B_n).$$

However, for the sequence of partial sums in the series on the right hand side, we observe that

$$\sum_{i=1}^{n} P(B_i) = P\left(\bigcup_{i=1}^{n} B_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) = P(A_n), \quad n = 1, 2, \ldots,$$

and, taking the limit as $n \to \infty$, we deduce

$$\sum_{n=1}^{\infty} P(B_n) = \lim_{n\to\infty} \sum_{i=1}^{n} P(B_i) = \lim_{n\to\infty} P(A_n).$$

This shows that the result of the proposition is true for an increasing sequence. Assume now that the sequence $\{A_n\}_{n\geq 1}$ is decreasing. Then, the sequence of their complements, $\{A_i'\}$, is increasing, so that the previous result applies to them, namely,

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

But,

$$\lim_{n\to\infty} P(A'_n) = \lim_{n\to\infty}(1 - P(A_n))$$

and so

$$P\left(\lim_{n\to\infty} A'_n\right) = P\left(\bigcap_{i=1}^{\infty} A'_i\right) = P\left(\left(\bigcup_{i=1}^{\infty} A_i\right)'\right) = 1 - P\left(\bigcup_{i=1}^{\infty} A_i\right) = 1 - P\left(\lim_{n\to\infty} A_n\right).$$

It now follows from the above that

$$1 - \lim_{n\to\infty} P(A_n) = 1 - P\left(\lim_{n\to\infty} A_n\right),$$

which completes the proof. □

Example 1.15 Suppose we have a population (1st generation) whose individuals may produce descendants, thus forming the 2nd generation, and so on. We are given that the probability this population becomes extinct by the n-th generation (since all individuals of that generation die before they give birth to any descendants), equals $\exp[-(3n+1)/(4n)]$. What is the probability that the population will never become extinct?

SOLUTION Let us denote by A_n, for $n = 1, 2, \dots$, the events

A_n: the population becomes extinct by the n-th generation of individuals.

Then, it is apparent that the sequence of events $\{A_n\}_{n\geq 1}$ is increasing, that is

$$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots \subseteq A_n \subseteq A_{n+1} \subseteq \cdots .$$

The event that the population becomes extinct eventually is described by the union

$$\bigcup_{n=1}^{\infty} A_n = \lim_{n\to\infty} A_n,$$

and the probability of this event is given by

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} \exp\left(-\frac{3n+1}{4n}\right) = e^{-3/4}.$$

As a consequence, the required probability equals

$$1 - P\left(\bigcup_{n=1}^{\infty} A_n\right) = 1 - e^{-3/4} = 0.52763.$$

EXERCISES

Group A

1. Let the sample space for an experiment be the set of positive integers, and consider the sequence of events $\{A_n\}_{n\geq 1}$ defined in it, where A_n is given by

$$A_n = \{n, n+1, n+2, \dots\}, \quad \text{for } n = 1, 2, \dots .$$

Show that $\{A_n\}_{n\geq 1}$ is a monotone sequence and find its limit as $n \to \infty$.

2. This exercise shows that an arbitrary sequence of events in a sample space can be written as the union of increasing events. Let $\{A_n\}_{n\geq 1}$ be a sequence of events in a sample space (not necessarily monotone), and define a new sequence $\{C_n\}_{n\geq 1}$ by $C_n = \cup_{i=1}^{n} A_i$, $n = 1, 2, \dots$.

   (i) Show that $\{C_n\}_{n\geq 1}$ is an increasing sequence of events in Ω.

   (ii) Verify that

$$\lim_{n\to\infty} C_n = \bigcup_{i=1}^{\infty} A_i.$$

   (iii) Show that

$$\lim_{n\to\infty} P(C_n) = P\left(\bigcup_{i=1}^{\infty} A_i\right).$$

3. Let the sample space Ω for an experiment be the set of real numbers, a be a real number, and let further the probabilities of the events (in this case, the open intervals)

$$A_n = \left(a - \frac{n}{2}, a + \frac{n}{2}\right), \quad n = 1, 2, \dots ,$$

be given by

$$P(A_n) = \frac{1}{n}, \quad n = 1, 2, \dots .$$

Calculate the probability of the elementary event $\{a\}$.

4. Let the sample space Ω for an experiment be the set of real numbers. In this space, we define the events

$$A_n = \left[2 + \frac{1}{2^n}, 4 - \frac{1}{2^n}\right], \quad n = 1, 2, \dots .$$

Assuming that for $n = 1, 2, \dots$, the probability of the event A_n is

$$P(A_n) = \frac{2^n - 1}{2^{n+1}},$$

calculate the probability of the open interval $(2, 4)$.

Group B

5. Let $\{A_n\}$ be a decreasing sequence of events on a sample space, for which we have

$$P(A_1) = q, \quad P(A_n - A_{n+1}) = q^n p, \qquad n = 1, 2, \dots,$$

where $0 < q < 1$ and $p = 1 - q$.

(i) Show that $P(A_n) = q^n$ for each $n = 1, 2, \dots$.

(ii) Calculate the probability $P(A)$ for the limiting event

$$A = \lim_{n\to\infty} A_n.$$

6. (**Borel–Cantelli lemma**). Let $\{A_n\}$ be a sequence of events in a sample space Ω, such that the sequence of real numbers

$$a_n = \sum_{i=1}^{n} P(A_i), \quad n = 1, 2, \dots,$$

converges to a real number. Prove that

$$P\left(\bigcap_{r=1}^{\infty} \bigcup_{n=r}^{\infty} A_n\right) = 0,$$

that is, the probability that infinitely many, among the events $A_1, A_2, \dots$, occur is zero.

(*Hint*: Apply Proposition 1.11 to the sequence of events $\{B_r\}_{r\geq 1}$, where $B_r = \cup_{n=r}^{\infty} A_n$ for $r = 1, 2, \dots$.)

1.7 BASIC CONCEPTS AND FORMULAS

Sample space Ω	The set of all possible outcomes which may appear in a realization of a chance experiment
An event A in a sample space Ω	A subset of the sample space, $A \subseteq \Omega$
Intersection AB between two events	The event which occurs when both A and B occur
Union $A \cup B$ between two events	The event which occurs when at least one of the events A and B occur
Disjoint events A and B	Two events which cannot happen together, $AB = \emptyset$
Complement A' of an event A	The event which occurs if and only if A does not occur ($A \cup A' = \Omega$ and A, A' disjoint)
Difference $A - B$ between two events	The event which occurs when A does but B does not occur

Relative frequency f_A	n_A/n, where n is the total number of (identical) repetitions of an experiment and n_A is the number of occurrences of A in these repetitions
Frequentist definition of probability	$P(A) = \lim_{n\to\infty} \frac{n_A}{n}$
Axiomatic definition of probability	$P1.$ $P(A) \geq 0$ for any event A on the sample space Ω; $P2.$ $P(\Omega) = 1$; $P3.$ If $A_1, A_2, \dots$ are pairwise disjoint events, $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$
Properties of probability	$P(\emptyset) = 0$, $P(\Omega) = 1$; $0 \leq P(A) \leq 1$ for any event A; $P(A') = 1 - P(A)$; $P(A) \leq P(B)$ for any events $A \subseteq B$; $P(A - B) = P(AB') = P(A) - P(AB)$; $P(A) = \sum_{i=1}^{\infty} P(a_i)$, for $A = \{a_1, a_2, \dots\}$; $P(A \cup B) = P(A) + P(B) - P(AB)$, so that if A and B are disjoint, $P(A \cup B) = P(A) + P(B)$.
Continuity property of probability	$\lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n)$, for any monotone sequence of events $\{A_n\}_{n\geq 1}$.

1.8 COMPUTATIONAL EXERCISES

The die outcomes, as well as the other entries in Table 1.1, were created using the algebraic package Mathematica. Specifically, with the following sequence of commands, we can simulate 100 repetitions for the experiment of throwing two dice, and then calculate the frequencies and relative frequencies for the events of interest.

```
n=100
a=Table[Random[Integer,{1,6}],{n}];
b=Table[Random[Integer,{1,6}],{n}];
s1=0;s2=0;s3=0;
Print["Throw  Outcome    Frequency for  Relative Frequency for"];
Print["nr                  A    B    C       A        B         C"];
Print["-----------------------------------------------------------"]
Do[
If[(a[[i]]==6 || b[[i]]==6),s1=s1+1];
If [ a[[i]]==4 , s2=s2+1];
If [ a[[i]]+b[[i]]==7, s3=s3+1];
f1=N[Round[10000*s1/i] /10000,3];
```

```
f2=N[Round[10000*s2/i] /10000,3];
f3=N[Round[10000*s3/i] /10000,3];

Print [i,"       ",a[[i]],"  ",b[[i]],"        ",s1,"   ",s2,"
",s3,"      ",f1,"         ",f2,"         ",f3]
,{i,1,n}]
```

Working in a similar way, use Mathematica (or any other appropriate software, such as Maple, R or a programming language) to obtain simulated repetitions of the experiments described in Exercises 1–7 below, and to find the probabilities of the events $A, B, C, \ldots$ defined in these exercises.

1. We toss a coin four times:

   A: the first and the third outcomes are "Heads";

   B: exactly two heads appear;

   C: the number of heads is 3 or more.

2. We roll a die twice:

   A: the first outcome is 5 and the second one is greater than 2;

   B: the first outcome is at least 4 and the second outcome is an even integer;

   C: the sum of the two outcomes is 5;

   D: the sum of the two outcomes is at least 9;

   E: the difference between the two outcomes is at most 3.

3. We throw a die until a six appears for the first time:

   A: the experiment is completed after exactly 3 throws;

   B: the experiment is completed after at least 3 throws;

   C: the experiment is completed after at most 3 throws.

4. We throw a die until we get three consecutive identical outcomes (e.g. 111, 222 and so on).

   A: the experiment finishes after more than 20 throws;

   B: the experiment terminates at the 12th throw;

   C: the experiment finishes at or before the 20th throw.

5. Each of two boxes contain 20 red balls and 30 black balls. We select a ball from the first box randomly and place it into the second one. Then, we select a ball from the second box and place it into the first one.

   A: both balls selected were red;

   B: the ball selected from the second box was black;

   C: at the end of the above experiment, each box contains the same number of balls from each color as in the beginning.

6. Each of five boxes contain 15 red balls and 25 black balls. We select a ball from the first box randomly and place it into the second one. Then, we select a ball from the second box and place it into the third one, and so on, until finally we select a ball from the 5th box:

   A: the ball selected from the 5th box is red;

   B: the ball selected from the 5th box has the same color with that selected from the 4th box;

   C: all five balls selected are of the same color.

7. We play the following game 50 times. We throw a die and, if the outcome is $1, 2, 3$ or 4 we win, otherwise we lose:

   A: we win at least 40 times;

   B: we win at most 25 times.

1.9 SELF-ASSESSMENT EXERCISES

1.9.1 True–False Questions

1. When throwing a die, the event "the outcome is a multiple of 3" is an elementary outcome.
2. For two events A and B in a sample space, the probability that at least one of them occurs is $P(A \cup B)$.
3. The sample space for an experiment is the closed interval $[0, 4]$ of the real line. Then, Ω is a finite sample space.
4. If A, B are disjoint events in a sample space and we know that A occurs, then B does not occur.
5. If an event A satisfies the condition $P(A) = 1 - 3P(A')$, then $P(A) = 1/4$.
6. For any events A and B in a sample space, we have $(A \cup B)' = A'B'$.
7. Let A be an event in a sample space such that $P(A) = 1$, and B be another event on that space. Then, $P(A \cup B') = 0$.
8. Let A, B, and C be three events in a sample space. If $A \subseteq B$ and the events A and C are disjoint, then the events B and C are also disjoint.
9. Assume that A, B, and C are three events on a sample space. If the events A and C are disjoint, and the events A and B are disjoint, then B and C are also disjoint events.
10. We toss a coin 300 times and observe that in 160 of these tosses the outcome is "Heads." Based on this, the relative frequency of the event "Heads occur in a single toss of a coin" is 160.

11. For any events A, B, and C in the same sample space, we have

$$A(BC) = (AB) \cup (AC).$$

12. If $P(A) = 1/4$ and $P(A') = 5P(B) - 1$, then the probability of the event B is also $1/4$.

13. Let A and B be two events in a sample space Ω such that $A \subseteq B$, and C be another event in Ω. If the events B and C are disjoint, then A and C will also be disjoint events.

14. Let A and B be two events in a sample space Ω with $A \subseteq B$. Then, $A' \cup B = \Omega$.

15. If for the events A and B we know that $P(A) - P(B) = 1/3$, then $P(A - B) \leq 1/3$.

16. If two events A and B, defined in the same sample space Ω, are mutually exclusive, then $A' \cup B = \Omega$.

17. Let $A_1, A_2, \ldots$ be a sequence of events defined in a sample space Ω, and let a new sequence $\{B_n\}_{n \geq 1}$ be defined by

$$B_n = A_1 A_2 \cdots A_n, \quad n = 1, 2, \ldots .$$

Then, $\{B_n\}_{n \geq 1}$ is a decreasing sequence of events in Ω.

18. Let $A_1, A_2, \ldots$ be a sequence of events defined in a sample space Ω, and let a new sequence $\{C_n\}_{n \geq 1}$ be defined by

$$C_n = \bigcup_{i=n+1}^{\infty} A_i, \quad n = 1, 2, \ldots .$$

Then, $\{C_n\}_{n \geq 1}$ is a decreasing sequence of events in Ω.

19. Let $A_1, A_2, \ldots$ be a monotone sequence of events defined in a sample space Ω. If it is known that

$$P(A_i) = \frac{1}{5}\left(\frac{2}{3}\right)^i, \quad \text{for } i = 1, 2, \ldots,$$

then the probability of the event that at least one of the A_i's occur is equal to zero.

1.9.2 Multiple Choice Questions

1. The event that exactly one of the three events A, B, and C in a sample space Ω occurs is

   (a) $(ABC)'$
   (b) $(A \cup B \cup C)'$
   (c) $AB'C'$
   (d) $(AB'C') \cup (A'BC') \cup (A'B'C)$
   (e) $AB \cup C$

2. For the events A and B in a sample space Ω, we know that

$$P(A) = (1 + \lambda)/3, \quad P(B) = 1 - \lambda^2$$

for some real number λ. The admissible range of values for λ is

(a) $\lambda \geq 0$ (b) $0 \leq \lambda \leq 1$ (c) $-1 \leq \lambda \leq 1$

(d) $0 \leq \lambda \leq 2/3$ (e) $-1/3 \leq \lambda \leq 2/3$.

3. Let A, B, and C be three events in a sample space Ω, such that $P(A) = 2P(B)$ and $P(B) = 2P(C)$. If in addition we have $A = (B \cup C)'$, then the probability of the event A is

   (a) $1/7$ (b) $2/7$ (c) $4/7$ (d) $1/2$ (e) $3/5$.

4. Marc tosses a coin until "Heads" appear for the second time. He is interested in the number of "Tails" which appear before the second appearance of "Heads." Then, a suitable sample space for this experiment is

   (a) $\{0, 1, 2, \ldots\}$ (b) $\{0, 1, 2\}$ (c) $\{1, 2\}$ (d) $\{2, 3, \ldots\}$ (e) $\{1, 2, \ldots\}$.

5. In a single throw of a die, consider the events $A = \{1, 2, 3\}$ and $B = \{2, 4, 6\}$. The event $(B - A)'$ is equal to

   (a) $\{4\}$ (b) $\{2, 4, 5, 6\}$ (c) $\{1, 3\}$

   (d) $\{4, 6\}$ (e) $\{1, 2, 3, 5\}$.

6. For the events A and B in a sample space Ω, we know that $P(A \cup B) = 1/6$. Then, the value of $P(A') - P(B)$ is

   (a) always equal to $5/6$ (b) equal to $5/6$ provided that $AB = \emptyset$

   (c) always equal to $1/6$ (d) equal to $1/6$ provided that $AB = \emptyset$

   (e) none of the above.

7. Maria is waiting at a bus stop for her friend Sarah so that they meet and go to a concert together. Maria wants to know how long she will have to wait until the next bus arrives at the bus stop and whether Sarah will be on it. A suitable sample space for this experiment is (a "0" below indicates that Sarah will not be on the next bus, while a "1" indicates she will be; further, t stands for Maria's waiting time until the arrival of the next bus):

   (a) $\Omega_1 = \{0, 1\}$ (b) $\Omega_2 = \{(1, t) : t \geq 0\}$

   (c) $\Omega_3 = \{(i, t) : i = 0, 1 \text{ and } t \geq 0\}$ (d) $\Omega_4 = \{0, 1\} \cup [0, \infty)$

   (e) $\Omega_5 = \{0, 1\} \cap [0, \infty)$.

8. Which of the following statements is correct with reference to the sample spaces Ω_i, for $i = 1, 2, 3, 4, 5$, defined in the last problem?

   (a) Ω_1 and Ω_5 are the only sample spaces which are finite

   (b) Ω_2 is a discrete sample space, but Ω_4 is not

   (c) Ω_4 is a discrete sample space, but Ω_2 is not

   (d) the set Ω_5 has infinitely many elements

   (e) Ω_3 is an infinitely countable sample space.

9. Let A, B, and C be three events in a sample space Ω, such that $A \cup B \cup C = A$. Then, the following is always true:

   (a) $B \cup C = A$ (b) $B \cup C \subseteq A$ (c) $A \subseteq B \cup C$

   (d) $BC = A$ (e) $A \subseteq BC$.

10. We throw a die until a six appears for the first time, at which point the experiment stops. Let A_i be the event that the outcome of the ith throw is a six. The event "a six appears for the first time in the 3rd throw" can be expressed in terms of the sets A_i as

    (a) $A_1A_2A_3$ (b) $A_1A_2A_3'$ (c) $A_1'A_2'A_3$

    (d) $A_1 \cup A_2 \cup A_3$ (e) $A_1 \cup A_2 \cup A_3'$.

11. We toss a coin successively. Let B_i be the event that the outcome of the ith toss is Heads. Then, the event "Heads occur for the first time at, or after, the second toss" can be written, in terms of the sets B_i, as follows:

    (a) B_1 (b) B_1B_2 (c) $B_1'B_2$ (d) $B_1'B_2B_3B_4\cdots$ (e) B_1'.

12. For the disjoint events A and B in a sample space Ω, we know that

$$P(A \cup B) = 5/6, \quad 4P(A) + P(B') = 1.$$

    Then, the probabilities of the events A and B are, respectively, equal to

    (a) $P(A) = 1/6, \quad P(B) = 2/3$ (b) $P(A) = 2/3, \quad P(B) = 1/6$

    (c) $P(A) = 1/3, \quad P(B) = 1/2$ (d) $P(A) = 1/2, \quad P(B) = 1/3$

    (e) $P(A) = 1/18, \quad P(B) = 7/9$.

13. At a large University class, there are 140 male students, 40 of whom own a car, and 160 female students, 20 of whom own a car. Let A be the event that "a randomly selected student is female" while B represents the event "a student owns a car." Using the relative frequency as an estimate for the probability of an event, the probability of the event $A'B'$ equals

    (a) $7/15$ (b) $1/15$ (c) $7/8$ (d) $5/7$ (e) $1/3$.

14. When Carol visits the local supermarket she buys a pack of crisps with probability 0.3, a chocolate with probability 0.4 and her favorite fruit juice with probability 0.6. The probability that she buys both crisps and chocolate is 0.2, that she buys both crisps and the fruit juice is 0.45, and that she buys chocolate and the fruit juice is 0.15. The probability that she will buy at least one of these three products in her next visit is

    (a) equal to 0.3 (b) equal to 0.7 (c) equal to 0.5

    (d) at least 0.5 (e) at most 0.5.

1.10 REVIEW PROBLEMS

1. In a major athletics competition, an athlete has three attempts to clear a certain height. If he succeeds in his first or second attempt, he makes no other attempts on this height.

   (i) Write down a suitable sample space Ω for this experiment and identify which elements of that space are included in the event

   E: the athlete clears the height in one of his three attempts.

   (ii) Let E_i be the event that "the athlete makes a total of i attempts in this height," for $i = 1, 2, 3$. Identify the elements of each event E_i.

   (iii) Express in terms of the events E_1, E_2, and E_3 each of the following events:

   A: the athlete fails in his first attempt;

   B: the athlete fails in his first two attempts;

   C: the athlete clears the height in his second attempt.

   Can you do the same for the event E in Part (i)? If yes, explain how; otherwise, explain why not.

2. Lena, Nick and Tom, who work for the same company, when they arrive for work one morning, meet at the ground floor elevator of the company building, which has three floors above the ground floor. Write down a sample space for the experiment which describes at which floor each person is going to. Then, identify the following events as suitable subsets of this sample space:

   A_1: Lena will leave the elevator on the first floor.

   A_2: Nick and Tom will leave the elevator on the second floor.

   A_3: none of the three will get off on the first floor.

   A_4: Lena and Nick will not go beyond the second floor.

   A_5: Nick is the only person to leave the elevator on the first floor.

   A_6: all three are going to different floors of the building.

   A_7: at least two persons are going to the same floor of the building.

3. The claims arriving at an insurance company are classified by the company as either Large (L) or Small (S) according to their size. The company wants to study the number of claims arriving prior to the first large claim.

   (i) Give an appropriate sample space Ω which can be used for the above study. Is Ω a finite sample space, a countably infinite or a continuous one?

   (ii) Write down explicitly each of the following events:

   A: the first large claim is the ith claim to arrive at the company, for $i = 1, 3, 5, 10$;

   B: the number of claims until the arrival of the first large claim is an even integer;

   C: the number of claims until the arrival of the first large claim is an odd integer;

D: the number of claims until the arrival of the first large claim is at most 5;

E: the number of claims until the arrival of the first large claim is greater than 5.

Which of these subsets of the sample space Ω are finite sets and which ones are countably infinite?

4. Let A and B be events in a sample space Ω for which it is known that $P(A) \geq a$ and $P(B) \geq b$, where a and b are given real numbers.

   (i) Show that
   $$P(AB) \geq a + b - 1.$$

   (ii) If $a + b > 1$, what do you conclude about the events A and B?

5. Assume that A, B, and C are three events in a sample space Ω. Show that the following relations hold:

   (i) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A'BC) - P(AB'C)$
   $-P(ABC') - 2P(ABC)$;

   (ii) $P(ABC) \geq P(A) + P(B) + P(C) - 2$;

   (iii) $P(A'BC) + P(AB'C) + P(ABC') \leq 2P(A' \cup B' \cup C')$;

   (iv) $P(AB'C') + P(A'BC') + P(A'B'C) \leq 2P(A \cup B \cup C)$.

6. Assume that the probability of each elementary event $\{i\}$ defined in the sample space $\Omega = \{1, 2, 3, \ldots\}$ is given by
   $$P(\{i\}) = 5^{i-1}/7^i, \quad i = 1, 2, 3, \ldots.$$
   Let us define the events
   $$A_n = \{n, n+1, n+2, n+3, n+4\}, \quad n = 1, 2, 3, \ldots.$$

   (i) After establishing that the relationship
   $$P(A_n) = \left(\frac{5}{7}\right)^{n-1} P(A_1), \quad n = 1, 2, 3, \ldots,$$
   holds, verify that $\lim_{n\to\infty} P(A_n) = 0$.

   (ii) Can you use the result of Proposition 1.11 to conclude that
   $$P\left(\lim_{n\to\infty} A_n\right) = 0 \quad ?$$

7. In the experiment of throwing a die twice, consider the following events:

   A: the first outcome is 6;

   B: the second outcome is 4;

   C: the sum of the two outcomes is 9;

   D: the first outcome is greater than the second.

Provide a suitable sample space for this experiment and then find which elements of this sample space each of the events below contains:

$$AB, AB', (AB') \cup D, BD, BCD, BC'D, A \cup (C'D'),$$
$$ABCD, ABC'D, A \cup B \cup C \cup D.$$

8. For the experiment of throwing a die twice, we consider again the events A, B, C, and D from the last exercise and denote by E the event "exactly two among the events A, B, C, and D occur":

   (i) Express the event E in terms of the events A, B, C and D.

   (ii) Find which elements of the sample space are contained in E.

9. On a particular day, a restaurant has a special three-course menu with the following choices:

Course	Choices
Starter	Caesar salad or shrimp cocktail or stuffed mushrooms
Main course	Roast beef or chicken or fish or vegetarian lasagna
Dessert	Ice cream or profiterole or fresh strawberries

   Poppy, who is visiting this restaurant with her friends, is to choose one course from each category above.

   (i) How many outcomes does the sample space for this experiment have?

   (ii) Let A be the event that she chooses stuffed mushrooms for a starter. How many outcomes are in the event A?

   (iii) Suppose B is the event she decides to have vegetarian lasagna as her main course. List the outcomes of the sample space which the event B contains.

   (iv) With A and B as defined above, how many outcomes are there in the event AB?

   (v) How many outcomes are there in the event $A \cup B$?

10. We consider the events $A_1, A_2, \ldots, A_n$ of a sample space Ω and, from these events, we form n new events $B_1, B_2, \ldots, B_n$ defined as follows: $B_1 = A_1$, while for $i = 2, 3, \ldots, n$,

$$B_i = A_i - \left(\bigcup_{j=1}^{i-1} A_j \right).$$

   (i) Verify that the events $B_1, B_2, \ldots, B_n$ are pairwise disjoint and that they satisfy the relations

$$\bigcup_{i=1}^{n} A_i = \bigcup_{i=1}^{n} B_i$$

   and

$$P(B_i) \le P(A_i), \quad i = 1, 2, \ldots, n.$$

(ii) Express the probability of the event

$$A = \bigcup_{i=1}^{n} A_i$$

in terms of the probabilities of the events B_i, $i = 1, 2, \ldots, n$;

(iii) Show that the following inequality holds:

$$P\left(\bigcup_{i=1}^{n} A_n\right) \le \sum_{i=1}^{n} P(A_i).$$

This is known as the **subadditive property** of probability (or as **Boole's inequality**), and it is also true for an infinite collection of events $A_1, A_2, \ldots$.

11. Let $A_1, A_2, \ldots, A_n$ be an arbitrary collection of n events in a sample space Ω. Show that

$$P(A_1 A_2 \cdots A_n) \ge 1 - \sum_{i=1}^{n} P(A_i') = \sum_{i=1}^{n} P(A_i) - (n-1).$$

This is known as **Bonferroni's inequality**.

(*Hint*: Apply Boole's inequality from the last exercise to the events A_i', $i = 1, 2, \ldots, n$.)

12. Let Ω be a sample space and suppose we have defined a set function, $P(\cdot)$, which satisfies the properties $P1$–$P3$ of Definition 1.10, on that space. Examine whether each of the set functions defined below can be used as a probability on that space; here, A is an arbitrary event on Ω.

(i) $P_1(A) = P(A)/2$ (ii) $P_2(A) = 1 - P(A)$

(iii) $P_3(A) = |P(A)|$ (iv) $P_4(A) = [P(A)]^2$

(v) $P_5(A) = (P(A) + 1)/2$ (vi) $P_6(A) = (1 - P(A))/2$.

13. A water network has three connections, C_1, C_2, C_3, as shown in Figure 1.18. For each connection, at the places marked $1, 2, 3$, some switches have been put and at a particular instant, any switch can be either ON (thus allowing water to pass through) or OFF. A connection is considered to be working if water can pass from point A to point B.

(i) Give a suitable sample space to describe the possible switch positions in the network.

(ii) For each connection, identify the events that describe the working status of the connection.

(iii) Let us denote by p_i, $i = 1, 2, 3$, the probability that the switch i is in the ON position and by p_{ij} for $i \ne j$ the probability that both switches i and j are ON ($i = 1, 2, 3$ and $j = 1, 2, 3$). Finally, p_{123} denotes the probability that all three switches are in the ON position. Express the probabilities of the events found in Part (ii) in terms of $p_1, p_2, p_3, p_{12}, p_{13}, p_{23}$, and p_{123}.

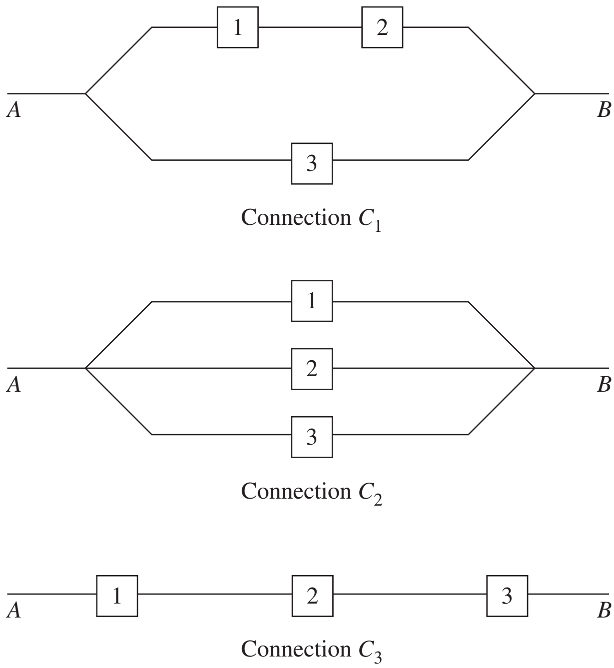


Figure 1.18 Water network with three connections.

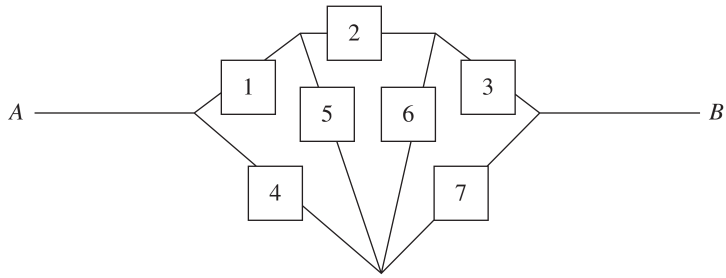
1.11 APPLICATIONS

1.11.1 System Reliability

Reliability engineering is a field that deals with the study of the performance of units/components or systems, i.e. sets of interacting or interdependent components that have been assembled as an integrated whole to carry out a specific task.

In general, the **reliability** of a system (or a unit) is defined as the ability of the system to perform and maintain the function(s) it has been specifically designed for. In probability theory, the term reliability is traditionally used to describe the probability that a system is capable of completing its expected function during an interval of time. Reliability evaluation is of crucial interest in reliability engineering, since it can be used effectively during the design stage so that its performance is optimized. The theoretical results of reliability theory have a lot of applications in a wide spectrum of areas such as airplane industry, nuclear accelerators safety, telecommunications networks, fluid transportation systems, and so on. It is worth mentioning that the failure of mechanical devices such as pumps, network relay stations, ships, airplane engines, etc. is similar in many ways to the death of biological organisms. In this case, the death of the organism should be interpreted as failure of the unit/system and the reliability corresponds to the death rate.

As an illustration, let us consider the following pipeline system, used for transferring water from point A to point B (arcs 1-7 represent seven pump stations). This system consists of 7 units (the 7 pump stations) and each of them can be either working (permits flow of the water through it) or not working:



The system as a whole is working if water flow from point A to point B is feasible through working pumps.

It is clear that if components 1, 2, and 3 work, the system will be functioning no matter what the other components' states are (working or not working). If we denote by $I = \{1, 2, \ldots, 7\}$ the set of all components of the system, then the subset $P_1 = \{1, 2, 3\}$ will be called a **path set** of the system in the sense that it describes a path of working components that guarantees the functioning of the system. On the other hand, there exists specific subsets of I with the following property: If all components of the subset are down, the system will not work no matter what the states of the rest of the components are. As an example, we mention $C_1 = \{1, 4\}$ or $C_2 = \{3, 4, 5, 6\}$. These sets are usually referred as **cut sets** of the system.

Should one be able to spot out all the path sets (or all the cut sets) of a system, then he/she could describe its operation (or breakdown) as a union of events and subsequently calculate its reliability by applying the inclusion–exclusion formula (see Proposition 1.10).

Now, we shall present an illustrative example of reliability evaluation for a specific structure.

Let n components be linearly connected in such a way that the system fails if and only if at least k consecutive components fail. This type of structure is usually called **consecutive-k-out-of-n: F system** and has a lot of applications in several areas such as network telecommunication systems and, fluid transportation networks. Let us consider, for example, a sequence of n microwave stations designed to transmit information from place A to place B. Assume that the microwave stations are equidistantly spaced between places A and B and each station is able to transmit information a distance up to k microwave stations. Then, it can be readily verified that the system fails if and only if at least k consecutive microwave stations fail.

From the definition, it is clear that, if we denote by $I = \{1, 2, \ldots, n\}$ the set of system components, the cut sets of the system can be described as follows.

$$C_i = \{i, i+1, \ldots, i+k-1\}, \quad i = 1, 2, \ldots, n-k+1.$$

Denoting by A_i, $i = 1, 2, \ldots, n-k+1$, the event

$$A_i : \text{all the components contained in } C_i \text{ are not working,}$$

the probability that the system does not work equals $P\left(\bigcup_{i=1}^{n-k+1} A_i\right)$ and consequently the reliability R of the system can be expressed as

$$
\begin{aligned}
R &= P(\text{systems works}) = 1 - P(\text{system does not work}) \\
&= 1 - P\left(\bigcup_{i=1}^{n-k+1} A_i\right).
\end{aligned}
$$

As an example, let us consider the special case of $n = 5$ and $k = 3$. Such a system consists of 5 components and fails if and only if 3 consecutive components fail. Therefore,

$$C_1 = \{1, 2, 3\}, \quad C_2 = \{2, 3, 4\}, \quad C_3 = \{3, 4, 5\}$$

and A_1, A_2, and A_3 are the events

$$
\begin{aligned}
A_1 &: \ \text{components } 1, 2, 3 \text{ do not work,} \\
A_2 &: \ \text{components } 2, 3, 4 \text{ do not work,} \\
A_3 &: \ \text{components } 3, 4, 5 \text{ do not work,}
\end{aligned}
$$

while the system reliability is given by

$$R = 1 - P(A_1 \cup A_2 \cup A_3).$$

It should be stressed that, besides the three cut sets given above, one might verify that there exist some additional cut sets that are supersets of C_1, C_2, and C_3, such as $C_4 = \{1, 2, 3, 4\}$, $C_5 = \{2, 3, 4, 5\}$, etc. However, these cut sets are not necessary for our analysis since the inclusion of the associated events in the union appearing in the RHS of the last expression does not affect the value of the resulting probability.

Exploiting the inclusion–exclusion formula, we get

$$R = 1 - S_1 + S_2 - S_3,$$

where

$$
\begin{aligned}
S_1 &= P(A_1) + P(A_2) + P(A_3), \\
S_2 &= P(A_1A_2) + P(A_2A_3) + P(A_1A_3), \\
S_3 &= P(A_1A_2A_3).
\end{aligned}
$$

Let us denote by q_i the probability that component i is not working, $q_{i,j}$ the probability that both components i and j are not working ($i \neq j$), $q_{i,j,k}$ the probability that all three components i, j, k are not working ($i \neq j \neq k$) and so on. Then, it is clear that

$$S_1 = q_{123} + q_{234} + q_{345}.$$

Since the event A_1A_2 occurs if and only if all components $1, 2, 3, 4$ are down, A_2A_3 occurs if and only if all components $2, 3, 4, 5$ are down, and A_1A_3 occurs if and only if all components $1, 2, 3, 4, 5$ are down, we conclude that

$$S_2 = q_{1234} + q_{2345} + q_{12345}.$$

Likewise, the event $A_1A_2A_3$ occurs if and only if all components $1, 2, 3, 4, 5$ are down, and so $S_3 = q_{12345}$. Therefore, the system reliability is given by

$$R = 1 - (q_{123} + q_{234} + q_{345}) + q_{1234} + q_{2345}. \tag{1.8}$$

Let us next consider the case when the components of the system are of the same quality, i.e. each one of them has the same probability, say p $(0 < p < 1)$, to be functioning (working) at a specific time point. If, in addition, we assume that the components work independently of each other, as we shall see later on in Chapter 3, the quantities appearing in (1.8) can be expressed as $q_{ijk} = (1-p)^3$ for all $i \neq j \neq k$, and $q_{ijk\ell} = (1-p)^4$ for $i \neq j \neq k \neq \ell$. In this case, the reliability of the system in (1.8) becomes

$$R = 1 - 3(1-p)^3 + 2(1-q)^4 = p + 3p^2 - 5p^3 + 2p^4. \tag{1.9}$$

In Figure 1.19, we present a plot of R as a function of p. Clearly, $r = R(p)$ is an increasing function of p (for $0 \leq p \leq 1$) attaining the value 0 for $p = 0$ and 1 for $p = 1$. In Table 1.2, we provide the values of $R = R(p)$ for several values of p.

Based on the above results, one may easily answer the question: What is the quality p of the components that guarantees a specific reliability level, say R_0, for the system? More explicitly, we see that we have to solve the inequality

$$p + 3p^2 - 5p^3 + 2p^4 \geq R_0,$$

or make use of Table 1.2 and look for the smallest value of p that satisfies the inequality $R(p) \geq R_0$. Therefore, if we wish to maintain a system reliability (at least) 95%, we should use components with functioning probability $p = 0.73$; instead, if we are satisfied with a system reliability 90% we can use components with $p = 0.65$, and so on.

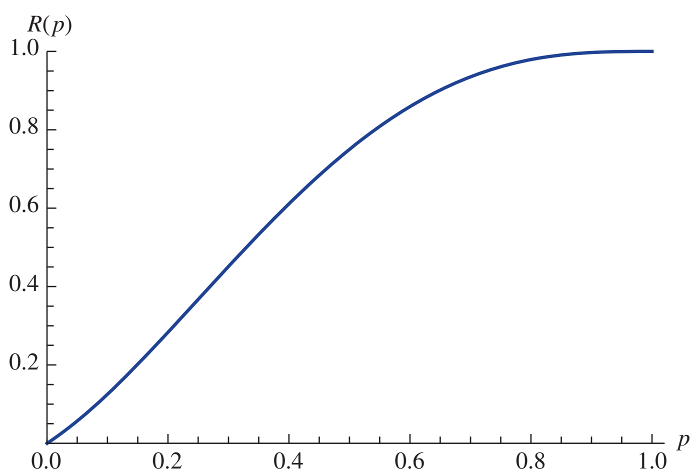


Figure 1.19 Plot of the system reliability, $R = R(p)$, as a function of p.

p	$R(p)$	p	$R(p)$	p	$R(p)$	p	$R(p)$
0.01	0.0103	0.26	0.3841	0.51	0.7623	0.76	0.9652
0.02	0.0212	0.27	0.4009	0.52	0.7744	0.77	0.9691
0.03	0.0326	0.28	0.4177	0.53	0.7861	0.78	0.9727
0.04	0.0445	0.29	0.4345	0.54	0.7975	0.79	0.9761
0.05	0.0569	0.30	0.4512	0.55	0.8086	0.80	0.9792
0.06	0.0697	0.31	0.4678	0.56	0.8194	0.81	0.9820
0.07	0.0830	0.32	0.4843	0.57	0.8299	0.82	0.9846
0.08	0.0967	0.33	0.5007	0.58	0.8400	0.83	0.9869
0.09	0.1108	0.34	0.5170	0.59	0.8498	0.84	0.9890
0.10	0.1252	0.35	0.5331	0.60	0.8592	0.85	0.9909
0.11	0.1399	0.36	0.5491	0.61	0.8683	0.86	0.9925
0.12	0.1550	0.37	0.5649	0.62	0.8771	0.87	0.9940
0.13	0.1703	0.38	0.5805	0.63	0.8855	0.88	0.9952
0.14	0.1858	0.39	0.5960	0.64	0.8936	0.89	0.9963
0.15	0.2016	0.40	0.6112	0.65	0.9014	0.90	0.9972
0.16	0.2176	0.41	0.6262	0.66	0.9088	0.91	0.9979
0.17	0.2338	0.42	0.6410	0.67	0.9159	0.92	0.9985
0.18	0.2501	0.43	0.6555	0.68	0.9227	0.93	0.9990
0.19	0.2666	0.44	0.6698	0.69	0.9291	0.94	0.9994
0.20	0.2832	0.45	0.6839	0.70	0.9352	0.95	0.9996
0.21	0.2999	0.46	0.6977	0.71	0.9410	0.96	0.9998
0.22	0.3166	0.47	0.7112	0.72	0.9464	0.97	0.9999
0.23	0.3335	0.48	0.7244	0.73	0.9516	0.98	1.0000
0.24	0.3503	0.49	0.7374	0.74	0.9564	0.99	1.0000
0.25	0.3672	0.50	0.7500	0.75	0.9609	1.00	1.0000

Table 1.2 The system reliability for several values of functioning probability p.

In closing, we mention that the analysis of the system could also be carried out by working with path sets instead of cut sets. It is not difficult to verify that the path sets of the system are

$$P_1 = \{3\}, \quad P_2 = \{1,4\}, \quad P_3 = \{2,5\}, \quad P_4 = \{2,4\}$$

(we have again excluded the path sets that contain – as a subset – any of P_1, P_2, P_3, P_4) and, introducing the events,

$$B_i \,:\, \text{all components in the path set } i \text{ are working,}$$

for $i = 1, 2, 3, 4$, we may express the system reliability as

$$R = P(B_1 \cup B_2 \cup B_3 \cup B_4) = S_1 - S_2 + S_3 - S_4,$$

where

$$
\begin{aligned}
S_1 &= P(B_1) + P(B_2) + P(B_3) + P(B_4),\\
S_2 &= P(B_1B_2) + P(B_1B_3) + P(B_1B_4) + P(B_2B_3) + P(B_2B_4) + P(B_3B_4),\\
S_3 &= P(B_1B_2B_3) + P(B_1B_2B_4) + P(B_1B_3B_4) + P(B_2B_3B_4),\\
S_4 &= P(B_1B_2B_3B_4).
\end{aligned}
$$

The probabilities involved in S_1, S_2, S_3, S_4 can be expressed in terms of

$$
\begin{aligned}
&p_i \,:\, \text{component } i \text{ is working},\\
&p_{ij} \,:\, \text{both components } i \text{ and } j \text{ are working } (i \neq j),\\
&p_{ijk} \,:\, \text{all three components } i, j, \text{ and } k \text{ are working } (i \neq j \neq k),
\end{aligned}
$$

and so on. For example,

$$P(B_1) = p_3, \quad P(B_1B_2) = p_{134}, \quad P(B_3B_4) = p_{245}, \text{ and } P(B_2B_3B_4) = p_{1245}.$$

In the special case of independent components with equal working probability p, we can make use of the formulas (the proof of which can be easily verified from the theory presented in Chapter 3)

$$p_i = p, \quad p_{ij} = p^2, \quad p_{ijk} = p^3, \quad p_{ijk\ell} = p^4, \quad p_{12345} = p^5,$$

and arrive at the expression (1.9) once again.

KEY TERMS

axiom of countable additivity
certain (sure) event
complement of an event
continuity property of probability
continuous sample space
difference of events
discrete sample space
disjoint (or mutually exclusive) events
event
finite additivity property
finite sample space
impossible event
intersection of events
law of large numbers
monotonically decreasing sequence of events
monotonically increasing sequence of events
nondeterministic experiment
random (or stochastic or nondeterministic) experiment
relative frequency
sample space
statistical regularity
tree diagram
union of events

CHAPTER 2

FINITE SAMPLE SPACES – COMBINATORIAL METHODS

Blaise Pascal
(Clermont-Ferrand 1623–Paris 1662)

French mathematician, physicist, and philosopher. From an early age, he showed exceptional skills in mathematics and physics; at the age of 11, he wrote an essay on sound waves, while his first scientific endeavor, *Essai pour les coniques* (Essay on Conics), written at the age of 17, drew the attention of René Descartes, one of the leading mathematicians of that time. While still a teenager, Blaise Pascal became obsessed with calculating machines. In 1642, he completed the design of the first computer, which was, however, too expensive to be put into production.

In 1654, motivated by a gambling problem posed to him by the French nobleman Chevalier de Méré, he started a correspondence with another renowned French mathematician, Pierre de Fermat. The exchange of ideas between the two Frenchmen is now considered by many historians to have laid the foundations not only of probability theory but also of modern scientific thinking.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

2.1 FINITE SAMPLE SPACES WITH EVENTS OF EQUAL PROBABILITY

In this chapter, we deal exclusively with sample spaces possessing a finite number of elements. Let us assume that $\omega_1, \omega_2, \dots, \omega_N$ are the elements of our sample space Ω, that is

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}.$$

Then, every nonempty subset of Ω must be of the form

$$A = \{\omega_{j_1}, \omega_{j_2}, \dots, \omega_{j_k}\},$$

where $j_1, j_2, \dots, j_k$ are k distinct integers in the set $\{1, 2, \dots, N\}$. We shall also use the simpler notation

$$p_i = P(\omega_i) = P(\{\omega_i\}), \quad i = 1, 2, \dots, N, \tag{2.1}$$

for the probabilities of the simple (elementary) events $\{\omega_i\}, i = 1, 2, \dots, N$.

Applying Proposition 1.5 for the entire sample space Ω (or using Corollary 1.1), we have

$$1 = P(\Omega) = \sum_{i=1}^{N} P(\omega_i) = \sum_{i=1}^{N} p_i.$$

Thus, the probabilities p_i, $i = 1, 2, \dots, N$, satisfy the conditions

(a) $p_i \geq 0, \quad i = 1, 2, \dots, N;$

(b) $\sum_{i=1}^{N} p_i = 1.$

Moreover, the probability that the event A occurs is easily expressed in terms of p_i's since, by employing Proposition 1.5 once again (this time for the set A), we readily find

$$P(A) = \sum_{i=1}^{k} P(\omega_{j_i}) = p_{j_1} + p_{j_2} + \cdots + p_{j_k}.$$

As a result, we see that in a finite sample space, it is relatively easy to find the probability of an event occurring provided we have first obtained the probabilities of the elementary events of that space. However, for the latter, we cannot just use the conditions (a) and (b) above, but we need some further information about their relative chances of occurrence.

Example 2.1 An insurance company classifies the arriving claims into four categories, depending on the size of these claims. From past data, it has been estimated that the first two categories, associated with the largest claims, arrive with the same probability, a claim from the third category is three times as much probable, and a claim from the fourth category is five times as much probable compared to a claim from the first or the second group. Then:

(i) Find the probability that a claim belongs to category i, for $i = 1, 2, 3, 4$.

(ii) What is the probability that the next claim to arrive at the company will be large (i.e. it will belong either to the first or to the second category)?

(iii) What is the probability that the next claim to arrive at the company will belong either to the second or to the fourth category?

(iv) What is the probability that the next claim to arrive at the company will *not* be of the first category?

SOLUTION A sample space that is appropriate for the analysis in the present example is

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\},$$

where $\{\omega_i\}$, for $i = 1, 2, 3, 4$, denotes the elementary event that a claim that arrives to the company belongs to category i.

(i) For the probabilities

$$p_i = P(\omega_i) = P(\{\omega_i\}), \quad i = 1, 2, 3, 4,$$

we note first that we must have $p_i \geq 0$, for $i = 1, 2, 3, 4$, and

$$p_1 + p_2 + p_3 + p_4 = 1. \tag{2.2}$$

Moreover, from the given relations among these probabilities, we have

$$p_1 = p_2, \quad p_3 = 3p_1, \quad p_4 = 5p_1.$$

Substituting each of these relations into (2.2), we obtain that

$$p_1 + p_1 + 3p_1 + 5p_1 = 1,$$

which gives $p_1 = 1/10$. Thus, the probabilities for the four categories of the claims are

$$p_1 = \frac{1}{10}, \quad p_2 = \frac{1}{10}, \quad p_3 = 3p_1 = \frac{3}{10}, \quad p_4 = 5p_1 = \frac{5}{10}.$$

(ii) The probability that, when a claim arrives, it belongs to one of the first two categories, associated with the large claims, is

$$P(\{\omega_1, \omega_2\}) = \frac{1}{10} + \frac{1}{10} = \frac{2}{10}.$$

(iii) Similarly, we find

$$P(\{\omega_2, \omega_4\}) = \frac{1}{10} + \frac{5}{10} = \frac{6}{10}.$$

(iv) Here, the event that a claim is not of the first category is $\Omega - \{\omega_1\} = \{\omega_2, \omega_3, \omega_4\}$, and this has probability

$$P(\{\omega_2, \omega_3, \omega_4\}) = \frac{1}{10} + \frac{3}{10} + \frac{5}{10} = \frac{9}{10}.$$

In the above example, the sample space had just four elements, and the associated probabilities were not all equal. This can make manipulations with probabilities awkward, or very tedious, if the sample space is large. Fortunately, in finite sample spaces, it is very common that all the elementary events have the same probability of occurrence. This is typically a consequence of a symmetry reasoning or it is simply because we have no reason to suspect that any particular event is more (or less) probable than any other. The following simple situations illustrate this point:

- When someone tosses a coin, it is natural to assume that $P(\text{“Heads”}) = P(\text{“Tails”}) = 1/2$ (unless someone has special skills).
- In the experiment of throwing a die, we assume that all six faces are equally likely.
- When we consider births of children, it is customary to assume that in each birth, both sexes have the same probability.
- When we pick a card at random from a pack of 52 cards, each of these 52 different outcomes represents an elementary event and the assumption is that any such event has probability $1/52$.

When two or more events are equally likely, these are often referred to as **equiprobable events**. Finite sample spaces with all elementary events being equiprobable are discussed at length in this chapter.

In such a space, recall from (2.1) that p_i stands for the probability of the elementary event ω_i. When all elementary events have the same probability, then all the p_i's are equal, that is

$$p_1 = p_2 = \cdots p_N = p$$

and since $p_1 + p_2 + \cdots + p_N = 1$, we obtain

$$p + p + \cdots + p = Np = 1,$$

so that

$$p_i = \frac{1}{N} \quad \text{for all } i = 1, 2, \ldots, N.$$

This readily yields that the probability the event

$$A = \{\omega_{j_1}, \omega_{j_2}, \dots, \omega_{j_k}\}$$

occurs to be

$$P(A) = p_{j_1} + p_{j_2} + \cdots + p_{j_k} = p + p + \cdots + p = kp = \frac{k}{N}.$$

We have thus reached the conclusion that, for finite sample spaces with equiprobable elementary events, the probability that an arbitrary event A occurs depends only on the number of elements,[1] $k = |A|$, that A contains and not on which elements it contains. More precisely, we have the following.

Proposition 2.1 ***(Classical definition of probability)*** *If the sample space Ω of an experiment is finite and all its elementary events are equiprobable, the probability of appearance of an event A is given by the formula*

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\textit{number of elements in } A}{\textit{number of elements in } \Omega}. \tag{2.3}$$

It is worth noting that, starting from the axiomatic foundation, the last expression is a logical deduction of the properties that the set function $P(\cdot)$ satisfies. However, long before this axiomatic foundation was given by Kolmogorov, Laplace had suggested in 1812 the use of that formula as *the definition of probability* (see Section 1.3), and for that reason we often call this "the classical definition of probability." Moreover, a very similar expression for the definition of probability was given earlier by Thomas Bayes in 1763, while it was also used implicitly even earlier than that by the French mathematicians Blaise Pascal, Pierre Fermat, Abraham de Moivre, and some others. Often, the elements of an event A are called **favorable cases** or **favorable outcomes** for the event A, while the elements of Ω are called **possible cases** or **possible outcomes**. Then, the definition of probability of A becomes

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{number of favorable outcomes for the event } A}{\text{number of possible outcomes (for the experiment)}}.$$

Example 2.2 Zoë, who is four years old, plays with three cubes. Each of these cubes has one of the letters J, O, Y written on it.

(i) How many three-letter words (most of them meaningless) can Zoë make, if she puts one cube next to the other?

(ii) If she puts the three cubes in some order completely at random, what is the probability that

  (a) the word she makes starts with J?

  (b) she produces a word that has a meaning?

[1]For a set A, we denote by $|A|$ the **cardinality** of A, which is simply the number of elements in A.

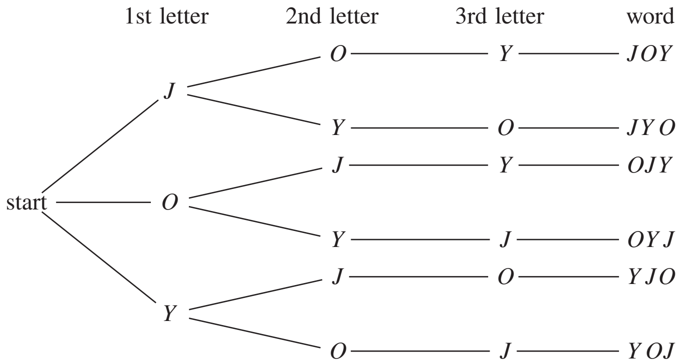


Figure 2.1 Tree diagram for the words that can be formed.

SOLUTION

(i) The set of all possible outcomes, i.e. the sample space for this experiment, as can be seen from the tree diagram above, consists of six elements. More specifically (see Figure 2.1),

$$\Omega = \{JOY, JYO, OJY, OYJ, YJO, YOJ\}.$$

(ii) Since each of these six outcomes has the same probability (equiprobable events), we can use the classical definition of probability.

(a) Let A be the event that the word she makes starts with J. Then, mathematically, the set A may be written as

$$A = \{JOY, JYO\}$$

which yields

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{6} = \frac{1}{3}.$$

(b) Next, let B be the event that the word that Zoë makes has a meaning. Out of the six three-letter words in Ω, only one word (JOY) has a meaning, so $B = \{JOY\}$ and clearly $P(B) = 1/6$.

Example 2.3 A large New York store that sells toys is going to hold a draw and the winner will receive a free one-week holiday. The store has 6000 tickets for sale to its customers during a week in the pre-Christmas season. If the tickets are numbered from 1 to 6000, what is the probability that the number on the winning ticket is a multiple of 2 or 5?

SOLUTION The sample space for this experiment (the draw) is the set $\{1, 2, \ldots, 6000\}$. Let us define the events

A: the winning number is a multiple of 2 (an even integer),
B: the winning number is a multiple of 5.

Clearly,

$$|\Omega| = 6000, \quad |A| = \frac{6000}{2} = 3000, \quad |B| = \frac{6000}{5} = 1200.$$

The intersection of the events A and B, i.e. the event AB, has those elements of Ω that are divisible by both 2 and 5. But these are precisely the integers that are divisible by 10, and so

$$|AB| = \frac{6000}{10} = 600.$$

We want the probability of the event $A \cup B$, and by using Proposition 1.9, we get

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(AB) \\ &= \frac{3000}{6000} + \frac{1200}{6000} - \frac{600}{6000} = \frac{1}{2} + \frac{1}{5} - \frac{1}{10} = \frac{6}{10} = 0.6. \end{aligned}$$

We should point out that, in order to use the classical definition, we must first ensure that both conditions for its use are met, namely,

- the sample space Ω of the experiment has finitely many elements;
- all elementary events have the same probability (they are equiprobable).

If one, or both, of these conditions is violated, using the formula from the classical definition may lead to wrong results. This is illustrated in the following two examples.

Example 2.4 Jean le Rond D'Alembert (1717–1783), one of the mathematicians who dealt with probability theory in its early stages, suggested the following calculation for the probability that, in a toss of two coins, heads (H) appears at least once. He considered as the sample space of this experiment the set

$$\Omega = \{0, 1, 2\},$$

where each of the elementary events $\{i\}$, for $i = 0, 1, 2$, describes the number of times H appears in the experiment. Since we are interested in the probability of the event

$$A = \{1, 2\},$$

D'Alembert claimed that

$$P(A) = \frac{|A|}{|\Omega|} = \frac{2}{3}.$$

However, one could argue as follows: the outcomes of the experiment are HH, HT, TH, TT (here, T stands for tails). Hence, we may take as sample space the set

$$\Omega_1 = \{HH, HT, TH, TT\}$$

and the event that at least one H appears is represented by the set

$$A_1 = \{HH, HT, TH\}.$$

In this way, we have three favorable outcomes out of a set of four possible outcomes, that is

$$P(A_1) = \frac{|A_1|}{|\Omega_1|} = \frac{3}{4}.$$

It is quite clear that between those two solutions, the second is the correct one. Can you see what went wrong with D'Alembert's argument?

Example 2.5 Suppose that we want to find the probability of selecting an even integer from the set of positive integers, $\Omega = \{1, 2, \ldots\}$, which is our sample space. In this case, Ω has infinitely many elements, and so we may think of using the following trick. We consider a finite subset Ω_n of Ω, consisting of just the first n positive integers. We then calculate the probability of the event

A_n: an even integer is selected from the set Ω_n

and, finally, we take the limit $\lim_{n\to\infty} P(A_n)$. However, this trick *does not work*, since, as we will see, it leads to contradictory results.

Suppose that first we use the sets

$$\Omega_{2n} = \{1, 2, 3, \ldots, 2n\}, \quad n = 1, 2, \ldots$$

Then we get $A_{2n} = \{2, 4, 6, \ldots, 2n\}$, so that

$$P(A_{2n}) = \frac{|A_{2n}|}{|\Omega_{2n}|} = \frac{n}{2n} = \frac{1}{2} \quad \text{and} \quad \lim_{n\to\infty} P(A_{2n}) = \frac{1}{2}.$$

Consider now as sample space a set of the form

$$\Omega_{2n-1} = \{1, 2, 3, \ldots, 2n-1\}, \quad n = 1, 2, \ldots$$

In this case, we have $A_{2n-1} = \{2, 4, \ldots, 2n-2\}$, and so we get

$$P(A_{2n-1}) = \frac{|A_{2n-1}|}{|\Omega_{2n-1}|} = \frac{n-1}{2n-1} \quad \text{and} \quad \lim_{n\to\infty} P(A_{2n-1}) = \frac{1}{2}.$$

However, suppose that we rearrange the positive integers as

$$2, 4, 1, 6, 8, 3, 10, 12, 5, \ldots$$

(in this way, we put two even integers, then an odd one, and so on). Then, by considering the subsets $A_{3n}, A_{3n-1}, A_{3n-2}$, for $n = 1, 2, \ldots$, it is easy to check that

$$\lim_{n\to\infty} P(A_{3n-1}) = \lim_{n\to\infty} P(A_{3n-2}) = \lim_{n\to\infty} P(A_{3n}) = \frac{2}{3},$$

(e.g. A_{3n} has $3n$ elements, $2n$ of which are even integers), which does not agree with the result found previously!

From these examples, it is clear that both a finite sample space as well as equiprobable elementary events are essential for making use of the classical definition of probability to obtain the right result.

Moreover, when our sample space is infinite, then applying the classical definition to a suitable finite space and then passing to the limit is inappropriate, as Example 2.5 demonstrates.

EXERCISES

Group A

1. For the experiment of tossing two coins, find the probability that at least one head appears using the sample space $\Omega = \{0, 1, 2\}$, suggested by D'Alembert (see Example 2.4).

   (Hint: Keep in mind that the events $\{0\}$ and $\{2\}$ have the same probability, while the probability of the event $\{1\}$ is twice as much.)

2. When throwing a die once, find the probability that the outcome is

   (i) an even integer;

   (ii) at least 5;

   (iii) at most 3;

   (iv) divisible by 3.

3. The sample space of an experiment is $\Omega = \{5, 6, 7, \ldots\}$. If the probability that the event $\{i\}$ occurs is three times the probability that the event $\{i + 1\}$ occurs, calculate

   (i) the probability of any elementary event, $\{i\}$, in this experiment;

   (ii) the probability that the outcome is an even integer;

   (iii) the probability that the outcome is an odd integer;

   (iv) the probability that the outcome is a multiple of 3.

4. The cholesterol level of a person, after a blood test, can be classified as normal, acceptable, or high. From the data of previous blood tests in a hospital, it has been estimated that the probability a patient has acceptable cholesterol level is twice as much as the probability of having normal cholesterol level and three times as much as the probability of having high cholesterol level. Find the probability that a randomly selected patient in this hospital belongs to each of these three groups according to their levels of cholesterol in their blood.

5. When tossing a coin three times, find the probability of each elementary event and then calculate the probability that
   (i) all three outcomes are the same (i.e. either three heads or three tails);
   (ii) exactly two heads appear;
   (iii) at least two heads appear;
   (iv) at least two successive outcomes are the same.

6. Consider the situation in Example 1.6 with the emission of digital signals. In this case, calculate the probabilities of the events

$$A_i, \quad \text{for } i = 0, 1, 2, 3, 4,$$

   as well as those of the events B, C, D, and E. For this calculation, assume that all 16 possible outcomes are equiprobable.

7. In the experiment of throwing a die twice,
   (i) write down all 36 possible outcomes that may appear;
   (ii) assuming that these outcomes are equiprobable, calculate the probability that
      (a) at least one outcome is a 4,
      (b) the first outcome is a 4,
      (c) the sum of the two outcomes is 6.

      Compare your results above with those found empirically in Table 1.1, using the frequentist definition of probability.

8. A building has two elevators: one for the first two floors only, and the other for floors three to five. Two persons enter the building and they both take the second elevator.
   (i) Write down all nine possible outcomes in the form (i, j), where i denotes the floor that the first person will exit the elevator and j denotes the floor that the second person will exit the elevator.
   (ii) Assuming that all nine outcomes are equiprobable, what is the probability that the two persons will exit the elevator in different floors?

Group B

9. An urn contains 2000 lottery tickets, numbered from 1 to 2000. If we select a ticket at random, find the probability that the number on the ticket
   (i) is a multiple of 6;
   (ii) ends in 3 or 7;
   (iii) does not end in $0, 2, 4, 6$, or 8;
   (iv) is divisible by at least one of the numbers $3, 5, 6$.

10. We consider the random experiment of selecting a real number from the interval $\Omega = (0, 1)$. Such a selection corresponds to a nonterminating process during which we successively select the first, second, third, … decimal digit of the number.

For instance, the number $\pi/4 = 0.785\ 498\ldots$ is produced by selecting the digits $7, 8, 5, 4, 9, 8, \ldots$, while the number 0.4 arises from the selection of the digits $4, 0, 0, 0, 0, \ldots$

We define the events

A_n: the first n decimal digits of the number drawn are equal to 1,

for $n = 1, 2, \ldots$

(i) Show that the sequence of events $\{A_n\}_{n\geq 1}$ is decreasing, i.e. $A_{n+1} \subseteq A_n$, for any $n = 1, 2, \ldots$.

(ii) Assuming that all digits $0, 1, 2, \ldots, 9$ have the same probability of being selected at each step, verify that

$$P(A_n) = \frac{1}{10^n}, \quad n = 1, 2, \ldots .$$

(iii) Using Proposition 1.6 (continuity of probability) show that the probability of selecting the number $1/7 = 0.142\ 857\ldots$ equals zero. Observe that the same is true for *every* real number in the sample space $\Omega = (0, 1)$.

2.2 MAIN PRINCIPLES OF COUNTING

According to the description in the last section, in order to find the probability of an event A, which is defined on a finite sample space Ω, it is sufficient to enumerate the elements of A and those of Ω. At first sight, this seems to be a straightforward task; we record each element of A and Ω and then simply count them. In practical terms, however, this can be cumbersome if the number of elements in Ω is very large.

When considering real-life problems, it is often the case that Ω is a very large set, and recording each element in Ω may be impractical. Suppose that, for example, we toss a coin (with just two outcomes in a single toss) 32 times, and we want to write down all possible realizations of these experiments. Even if we record possible outcomes at a speed of one per second, we would require more than a century to write down all of them! It is therefore imperative for such cases to find a systematic way of counting possible outcomes of experiments, exploiting the particular structure of the problem, and then conveniently decomposing the experiment in question into simpler ones. In this way, there will be *no need to record* each possible outcome but simply count their number, along with the number of favorable outcomes for the event of interest. A general term for this approach is *counting principles*, and such principles are the object of a particular branch of mathematics, known as **combinatorial methods** or simply **combinatorics**.

In the ensuing sections of this chapter, we describe some results and tools from combinatorics. The reader should bear in mind that the exposition in this chapter is only introductory and so it only serves as a vehicle for tackling problems in probability theory. Further results and more advanced techniques of combinatorial analysis may be found in specialized texts on the subject, such as those given in the bibliography of this book.

The most important principle of counting is the so-called *multiplicative principle* (or basic principle of counting). To introduce this informally, suppose that we have two experiments, the first of which has m possible outcomes, while the second one has n

different outcomes. Then the total number of outcomes, for the combination of these two experiments, is mn.

More generally, assume that the following conditions are satisfied for the enumeration of the elements of a certain set:

- the enumeration procedure may be split into k steps, which can be executed successively;
- the number of possible choices (outcomes) at each step is completely determined once the results of the previous steps are known.

Then the enumeration can be carried out by exploiting the following **multiplicative principle** (or **multiplicative law**).

Proposition 2.2 *Suppose that an element (object) a_1 can be chosen with n_1 different ways, and for each choice of a_1, the element a_2 can be chosen in n_2 different ways, and so on, so that for any selection of the elements $a_1, a_2, \ldots, a_{k-1}$, the element a_k can be chosen in n_k different ways. Then, the k-tuple $(a_1, a_2, \ldots, a_k)$ can be chosen successively and in that order in $n_1 n_2 \cdots n_k$ different ways.*

The above result is fundamental and is used throughout without further reference. An equivalent, and somewhat simpler, formulation of the multiplicative principle is the following:

Assume that $E_1, E_2, \ldots, E_k$ are sets so that E_i contains n_i elements. Then, there are $n_1 n_2 \cdots n_k$ different ways to choose first an element of E_1, then an element of E_2, and so on, until finally an element of E_k is chosen.

Example 2.6 Lucy has in her wardrobe five handbags, four dresses, three pairs of gloves, and six pairs of shoes. How many different outfits are possible for Lucy to wear for a ball that she has been invited to?

SOLUTION Assuming that each handbag may be combined with any dress, pair of gloves, and pair of shoes (although Lucy may not agree with that!), we see from the multiplicative law that there are

$$5 \cdot 4 \cdot 3 \cdot 6 = 360$$

different ways for her to be dressed for the ball.

Example 2.7 A city A is connected to city B through two different routes, while another city, C, can be reached from city B through four different routes.

(i) In how many different ways one may reach city C from city A?

(ii) If the choice of routes from city A to city C is completely at random, what is the probability that someone uses a specific route?

SOLUTION

(i) We denote by r_1 and r_2 the two routes from city A to city B and by R_1, R_2, R_3, and R_4 the four routes from B to C, as shown in Figure 2.2.

The choice of the route from A to B (element a_1) can be made in $n_1 = 2$ different ways. When someone has reached city B, there are $n_2 = 4$ ways to select the route (element a_2) to travel to city C. Therefore, there are $n_1 n_2 = 2 \cdot 4 = 8$ ways to select a route (with the above notation, this amounts to selecting the elements a_1, a_2 and in that order) from A to C. The above description can be seen diagrammatically in Figure 2.3.

(ii) Since the choice of the route from A to C is made completely at random, and there are eight possible choices (the sample space Ω of that experiment has eight elements), each route, which corresponds to an elementary event of Ω, has probability $1/8$.

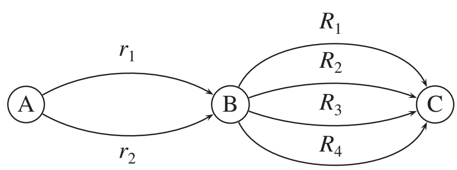


Figure 2.2 Diagram with the routes from city A to city C.

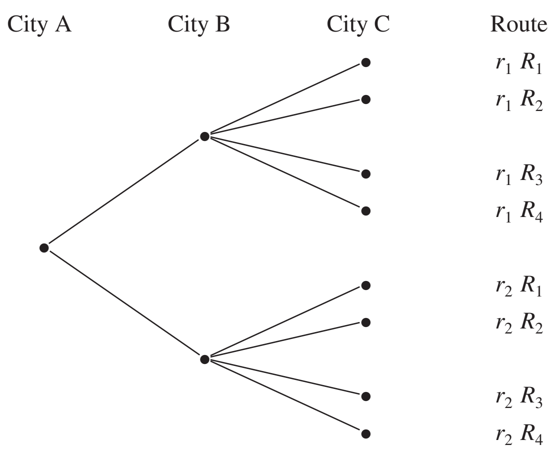


Figure 2.3 Tree diagram showing all possible routes from City A to City C.

Example 2.8 In a certain country, license plates in cars have seven places, the first three of which are letters and the other four are digits.

(i) How many different plate numbers are possible?

(ii) How many different plate numbers are possible so that no digit appears twice?

SOLUTION

(i) For each of the three letters there are 26 different choices, while for each of the four digits there are 10 choices. Thus, the total number of different plate numbers that can be formed is

$$26 \cdot 26 \cdot 26 \cdot 10 \cdot 10 \cdot 10 \cdot 10 = 175\ 760\ 000.$$

(ii) This time, there are 10 choices for the first among the four digits, but 9 choices for the second digit, since the digit which appears first needs to be excluded from selection in the following places. Similarly, there are eight choices for the third one and, finally, seven choices for the fourth. Consequently, the multiplicative law yields now the number of different plate numbers to be

$$26 \cdot 26 \cdot 26 \cdot 10 \cdot 9 \cdot 8 \cdot 7 = 88\ 583\ 040.$$

Example 2.9 At the end of an academic year, the best two students in a class consisting of 25 students will receive a prize from their school. Find how many different choices are possible for the prize recipients.

SOLUTION Here, we have 25 choices for the first student to be selected as a prize recipient, and for any such choice, there are 24 students remaining (so that one of them wins the second prize), thus giving a total of

$$25 \cdot 24 = 600$$

different pairs of students.

It is important to note, however, that all these pairs are distinguishable *only* if we treat the two prizes offered by the school as being distinguishable (i.e. the best student wins the first prize, which is more important than the second one). If we treat the two prizes as being identical, so that it does not matter who wins the first and who wins the second prize, then *the answer to the example is different*. Essentially, this means that it does not matter whether a student is chosen first or second in the two stages of this experiment. Therefore, if the order within each pair of winners is irrelevant, in the total of 600 pairs that we found above, we have counted each pair twice; for example, if Anna (A) and Laura (L) are the prize winners, the two pairs (AL) and (LA) are in this context identical and both pairs may not enter the sample space together. If we are interested in the pair of prize winners but not in their respective order, then the answer to the example is that there are

$$\frac{25 \cdot 24}{2} = 300$$

different pairs of students.

The point here is that when we apply the multiplicative law, we treat each possible outcome as being *ordered*. The distinction between ordered and unordered outcomes in an experiment involving successive selections is one of the main themes of this chapter, and we shall discuss it in detail in the following sections.

Finally, we note that a special case wherein the multiplicative law applies is for the calculation of the cardinality (i.e. the number of elements) for the Cartesian product among k sets $A_1, A_2, \dots, A_k$. More specifically, given the sets $A_1, A_2, \dots, A_k$, their Cartesian product is defined by

$$A_1 \times A_2 \times \cdots \times A_k = \{(a_1, a_2, \dots, a_k) : a_1 \in A_1, a_2 \in A_2, \dots, a_k \in A_k\}.$$

Proposition 2.3 *For $k \geq 2$, let $A_1, A_2, \dots, A_k$ be k finite sets and denote by $A_1 \times A_2 \times \dots \times A_k$ their Cartesian product. Then,*

$$|A_1 \times A_2 \times \cdots \times A_k| = |A_1| \cdot |A_2| \cdots |A_k| = \prod_{i=1}^{k} |A_i|.$$

In particular, when $A_i = A$ for $i = 1, 2, \dots, k$, denoting by A^k the Cartesian product of A with itself k times, we readily have the formula

$$|A^k| = |A|^k.$$

EXERCISES

Group A

1. For the council of the International Mathematical Society, there are 6 candidates for President, among whom 2 are women, 5 nominations for Vice President, 3 of whom are women, and 10 candidates are for Executive Secretary, with 5 women among them. Assuming that for each position all candidates have equal probabilities of being elected, what is the probability that all three members of the council elected will be women?

2. A University exam consists of n multiple choice questions. For the first question there are k_1 possible answers, and for the second one there are k_2 answers and so on, and finally for the nth question there are k_n answers to choose from.

   (i) If a student does not know the answer to any question and chooses her answers completely at random, how many different selections she has in total? What is the probability that she answers all questions correctly?

   (ii) Assuming again that the student gives her answers at random, what is the probability that she does not answer any questions correctly?

3. Peter and Wendy have applied for a job in two different posts that have been advertised; including them, for the first post there are 12 applicants, while for the second there are 9 applicants. Assuming that all applicants in each post are equally likely to be offered the position, what is the probability that both Peter and Wendy

get a job? We assume that if both posts are offered to the same person, she/he will accept only one of them while the other will remain vacant.

4. A car manufacturer offers three versions for a certain type of car: normal (N), luxury (L), and executive (E). Each of the three versions of this type may be equipped with any of the four engine capacities (in liter): $1.8, 2.0, 2.5$, and 3.0. Finally, the interested buyer may choose among six different colors that are available for this car.

   (i) How many different versions are available in total for the car in question?

   (ii) Assuming that all different versions have the same probability of being chosen, what proportion of cars that will be sold will have an engine capacity of *at least* 2.5 liter?

5. In a certain country, vehicle registration plates consist of three letters and four digits (e.g. ABC1234). What is the probability that a randomly chosen license plate

   (i) starts with A, E, or I?

   (ii) ends with a 4 or 5?

   (iii) starts with A, E, or I *and* ends with a 4 or 5?

6. A telecommunications channel transmits digits that are either 0 or 1. A sequence of four digits is transmitted.

   (i) Identify the sample space Ω for this experiment.

   (ii) Assuming that the elements of the sample space are equiprobable, what is the probability that

      (a) all transmitted digits are the same?

      (b) no two consecutive digits are the same?

      (c) there are three successive 1's transmitted?

      (d) there are exactly three successive 1's transmitted?

Group B

7. A psychologist is carrying out a memory test using three-letter "words" (either meaningful or not). For the first letter, she chooses among the letters G, H, D, L, the second letter can be one of A, O, I, E, while the third can be one of the letters D, X, W, P, R.

   (i) What is the total number of three-letter words available for this experiment?

   (ii) If all words are equiprobable, what is the probability that the word chosen

      (a) begins with G?

      (b) ends with either D or X?

      (c) begins and ends with the same letter?

      (d) includes the letter D?

8. Nick throws a die four times in succession.
   (i) What is the probability that the first three outcomes are $3, 2$, and 5 (in that order)?
   (ii) What is the probability that at least two of the four outcomes are the same?

9. Maria wants to select a number that has four digits and all these digits belong to the set $\{1, 2, 3, 4, 5, 6\}$. Assuming that all selections are equally likely, what is the probability that the number she selects
   (i) is the number 6543?
   (ii) has at least two digits that are the same?

10. Suppose that we wish to make a three-letter word in the following way. The first letter is chosen among the first nine letters of the Latin alphabet (A, B, …, H, I), the second letter among the next nine letters (J, K, …, Q, R), and the third letter among the last 8 letters of the alphabet (S, T, …, Y, Z).
    (i) How many different words (not necessarily meaningful) are possible?
    (ii) If all three letter words are equiprobable, what is the probability that the three-letter word contains at least one of the letters A, E, I, O, U, Y?

11. Consider the sample space

$$\Omega = \{(x, y) : x, y \in \{1, 2, \dots, k\}\},$$

where k is a positive integer. We define the events

$$A_i = \{(i, y) \in \Omega : y < i\}, \quad i = 2, 3, \dots, k,$$

and

$$A = \{(x, y) \in \Omega : y < x\}.$$

   (i) Using either the multiplicative law or the result of Proposition 2.3, find the number of elements of Ω.
   (ii) Verify that the events A_i are pairwise disjoint and that

$$A = \bigcup_{i=2}^{k} A_i;$$

   (iii) Assuming that all elements of Ω have the same probability, calculate the probability of each event A_i and consequently the probability of the event A.

12. Consider now the sample space

$$\Omega = \{(x, y, z) : x, y, z \in \{1, 2, \dots, k\}\},$$

where k is a positive integer. We define the events

$$A_i = \{(i, y, z) \in \Omega : y < i \text{ and } z < i \}, \quad i = 2, 3, \ldots, k,$$

and

$$A = \{(x, y, z) \in \Omega : y < x \text{ and } z < x\}.$$

(i) Arguing as in part (i) of the previous exercise, find the number of elements of Ω.

(ii) Verify that the events A_i are pairwise disjoint and that

$$A = \bigcup_{i=2}^{k} A_i;$$

(iii) Assuming that all elements of Ω are equiprobable, calculate the probability of each event A_i and, hence, find the probability of the event A.

2.3 PERMUTATIONS

The multiplicative law, introduced in the previous section, enables us to enumerate ordered k-tuples, say $(a_1, a_2, \ldots, a_k)$, where the element a_i is chosen from a set A_i, that is $a_1 \in A_1, a_2 \in A_2, \ldots, a_k \in A_k$. In many practical applications of the multiplicative law, all the elements a_i are chosen from the same set with or without further restrictions. For instance,

- when we throw a die k times, the results of the successive throws always belong to the set $X = \{1, 2, 3, 4, 5, 6\}$;
- from a lottery with 49 balls numbered from 1 to 49, in each draw 6 balls (numbers) are selected from the set $\{1, 2, 3, \ldots, 49\}$.

Although the two examples above look similar, they differ in an important aspect that becomes crucial in statistics, notably as to how a sample is selected from a population. In the first example above, if the first outcome is 3, then 3 might very well be the outcome of the second and the third throws as well. This means that the successive throws of a die are *identical repetitions* of the first throw, i.e. of the same experiment, and thus the result of the first throw has no impact on the subsequent ones. We therefore see that the single experiment of throwing a die k times can be considered to be the same as *k independent* repetitions of an experiment involving just throwing a die once. But, the situation differs in the second example above since if the first number drawn is 17, this number cannot turn up in any of the five remaining selections. Thus, although conceptually all six numbers are drawn from the set $\{1, 2, 3, \ldots, 49\}$, as mentioned above, in fact only in the first selection there are 49 possibilities, since the second number drawn would be selected from the set $\{1, 2, 3, \ldots, 16, 18, 19, \ldots, 49\}$. Similarly if the second number that appears is 23, for the third selection there are only 47 choices left, all natural numbers up to 49 except 17 and 23.

The difference between the two situations is perhaps best understood when phrased in terms of sampling: in the first case, we sample (from the set $X = \{1, 2, 3, 4, 5, 6\}$) *with replacement*, while the lottery example is typical of a sampling plan *without replacement*. Generally, when we have a (finite) population of size n, and we wish to draw a sample of size k, we speak of sampling with replacement when each unit selected for inclusion in the sample is put back for the subsequent selection and may appear as a sample unit again. Otherwise, when each unit selected is removed from the next selection, we have sampling without replacement. Note that in the second example above, the sample size k cannot exceed the population size n (49, in this case). Probabilistic arguments for the case wherein the sampling is without replacement are typically more subtle and are discussed in more detail in this chapter, since in this case the experiment of choosing k units from a population cannot be thought of as (probabilistically, identical) experiments involving the selection of a single unit; moreover, these single selections are not independent.

Another distinction that is important and we must be aware of when selecting k units successively from a population of size n is whether *the order* these units are chosen matters or not. To illustrate the difference between these two cases, we present the following example.

Example 2.10 In a cooking competition, candidates for the prizes are eliminated successively until four contestants remain: Nicky, Liz, Chris, and Tony. Three of them will win the prizes. Assuming that all arrangements for the final classification of the four contestants are equally likely, how many different arrangements are possible for the three winners if

(i) The three prizes of the competition are in descending order, so that the first prize wins the largest amount of money, followed by the second which in turn is followed by the third prize?

(ii) All prizes are of equal value (so that it does not matter who finishes first, second, or third, but only which contestant is eliminated from the prizes)?

SOLUTION Initially, we observe that both parts (i) and (ii) of the example involve sampling without replacement from a population of size $n = 4$ (the same contestant cannot win more than one prize). Next, it is clear from the statement that in part (i) the actual positions of the three prize winners are important while in part (ii) they are not. For instance, the arrangement (N, C, L), which means that Nicky wins the cooking contest, Chris finishes second, Liz third, and Tom is fourth is identical, under the assumptions of part (ii), with the arrangement (L, C, N–with the obvious interpretation for the letters) whereas for part (i) these two arrangements are not identical.

(i) Here, we have an application of the multiplicative law (Proposition 2.2). There are clearly four choices for the first prize of the competition. For each of these choices, there are three choices for the second prize (for example, if L wins the first place, the candidates for the second place are N, C, T). Similarly, if the first two places have been decided, then there are two options available for the third place, and finally, when the three prize winners have been selected, there remains only one contestant who necessarily will be placed fourth. As a consequence, the multiplicative law immediately yields the number of different arrangements to be

$$4 \cdot 3 \cdot 2 \cdot 1 = 24.$$

(ii) Under the assumptions for this part, we simply have to select three out of the four contestants, irrespectively of their order. But this can be done in four ways, because each selection of three persons is the same as selecting one who will *not* win any prize. Thus, the four choices available for the three prize winners are

$$\{N, C, L\}, \ \{N, C, T\}, \ \{N, L, T\}, \ \{C, L, T\}.$$

The arrangements that arise when we select a set of k units from a larger set of n units are called

- **permutations** if the ordering in which these k units are drawn is important, as in part (i) of the example above;
- **combinations** if the ordering is irrelevant, as in part (ii) of the example.[2]

In the remainder of this section, we deal with permutation arrangements, while in the following section we consider combinations.

Definition 2.1 Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set and k be a positive integer such that $k \le n$. Then, any ordered k-tuple $(a_1, a_2, \ldots, a_k)$, where $a_i \in X$ for $i = 1, 2, \ldots, n$, and $a_i \neq a_j$ whenever $i \neq j$, is said to be a **k-element permutation** of the n elements. When $k = n$, then we simply say that $(a_1, a_2, \ldots, a_n)$ is a permutation of the n elements.

The number of k-element permutations for a given set of n elements will be denoted by $(n)_k$. To present a simple example, consider the set $X = \{a, b, c\}$ consisting of $n = 3$ elements. Then, the two-element permutations of this set are

$$(a, b), (a, c), (b, a), (b, c), (c, a), (c, b),$$

while the (three-element) permutations of the three elements a, b, c are

$$(a, b, c), (a, c, b), (b, a, c), (b, c, a), (c, a, b), (c, b, a),$$

so that

$$(3)_2 = (3)_3 = 6.$$

The next proposition gives a general formula for the number of k-element permutations of a set of n elements.

[2]In statistics, when a sample is drawn from a population, to distinguish between these two cases, we speak of an *ordered sample* when ordering is important, as opposed to the plain use of the term *sample*, which means that ordering of units drawn does not matter.

Proposition 2.4

(i) *The number of k-element permutations among n elements is given by*

$$(n)_k = n(n-1)(n-2)\cdots(n-k+1), \quad 1 \le k \le n. \tag{2.4}$$

(ii) *The number of permutations among n elements is given by*

$$(n)_n = n(n-1)(n-2)\cdots 2\cdot 1, \quad n \ge 1.$$

Proof:

(i) The argument in the proof is essentially the same as the one we used for Part (i) of Example 2.10.

Suppose that we select, one after the other, k elements from a set of n elements. For the first element to be chosen, there are n choices. Once we have chosen this, the remaining $n-1$ elements of the set are candidates for the second element that we choose. For each selection of the first two elements, there are $n-2$ ways to choose the third element, and so on until we have chosen $k-1$ elements, so that any of the remaining $n-(k-1) = n-k+1$ elements may be picked as the final item in our sample. The required result now follows readily from the multiplicative law (Proposition 2.2).

(ii) This is a special case of Part (i) when $k = n$. □

The quantity $n(n-1)(n-2)\cdots 2\cdot 1$, which appears in Part (ii) of the above proposition, is a key quantity not only in combinatorial analysis but also in several other areas of mathematics and a special symbol is used for it. Specifically, we shall use from now on the symbol $n!$, where n is a positive integer, to denote

$$n! = 1\cdot 2\cdot 3\cdots n$$

(the symbol $n!$ is pronounced as "n factorial"). In fact, $n!$ can be used even for a noninteger n, but this will not be considered in the present book.[3]

Following the above, we can see that the number $(n)_k$ of k-element permutations among n elements can be expressed using factorials. Specifically, multiplying $(n)_k$ by $(n-k)!$ yields

$$\begin{aligned}(n)_k\cdot(n-k)! &= [n(n-1)\cdots(n-k+1)][(n-k)(n-k-1)\cdots 2\cdot 1]\\ &= n(n-1)\dots 2\cdot 1 = n!,\end{aligned}$$

so that we have $(n)_k \cdot (n-k)! = n!$ and, consequently,

$$(n)_k = \frac{n!}{(n-k)!}.$$

In view of this, we can now restate Part (i) of Proposition 2.5 as follows.

[3]Instead, we shall meet in the last chapter the so-called *Gamma function*, which can be thought of as a generalization of the concept of a factorial.

Corollary 2.1 *The number of k-element permutations among n elements is given by*

$$(n)_k = \frac{n!}{(n-k)!}, \quad 1 \leq k \leq n.$$

The symbol $n!$ is extended for $n = 0$ with the convention that $0! = 1$. Moreover, the symbol $(n)_k$ is also defined for $k = 0$ by setting $(n)_0 = 1$.

Similarly, we mention that the symbol $(n)_k$ may be extended for negative integers as follows. Let n be a positive integer. Then, replacing n by $-n$ in formula (2.4), we get

$$\begin{aligned}(-n)_k &= -n(-n-1)(-n-2)\cdots(-n-k+1)\\ &= (-1)^k\, n(n+1)(n+2)\cdots(n+k-1).\end{aligned}$$

The product $n(n+1)(n+2)\cdots(n+k-1)$ is called the **ascending factorial** (or the rising factorial) of order k and is denoted by $[n]_k$. For $k = 0$, we once again have the convention $[n]_0 = 1$.

In the arrangements we have considered so far, we have excluded the possibility that an element can be selected more than once. In the case where this is allowed, so that any element of the set X may be chosen $0, 1, 2, \ldots$ times, we speak of an arrangement with repetitions.

Definition 2.2 Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set and k be a positive integer such that $k \leq n$. Then, any ordered k-tuple $(a_1, a_2, \ldots, a_k)$, where $a_i \in X$ for $i = 1, 2, \ldots, n$ (with the a_i's not necessarily distinct), is called a ***k*-element permutation with repetitions** of the n elements.

In the following result, we calculate the number of k-element permutations with repetitions for a set of n elements.

Proposition 2.5 *The number of k-element permutations with repetitions of a set with n elements equals n^k.*

Proof: From the definition of the Cartesian product, we see that the k-element permutations with repetitions of the elements from a set $X = \{x_1, x_2, \ldots, x_n\}$ are precisely the elements of the Cartesian product $A_1 \times A_2 \times \cdots \times A_k$, where $A_1 = A_2 = \cdots = A_k = X$. The result is then obvious from Proposition 2.3.

Alternatively, we may regard the process of forming an ordered k-tuple as consisting of k steps. In the first step, we select randomly an element, say a_1, among the elements of X. Since X has n elements, there are $n_1 = n$ ways to select a_1. For each such selection, we then (in the second step) select a second element of X, say a_2. Since replacement is allowed, so that a_1 may be selected again, there are $n_2 = n$ ways to select a_2. Continuing in this way, we reach the kth and final step wherein we select an element, say a_k, of X, and for this there are $n_k = n$ ways. Then, from the multiplicative law, the ways to perform all k steps are

$$n_1 \cdot n_2 \cdots n_k = \underbrace{n \cdot n \cdots n}_{k \text{ terms}} = n^k.$$

□

Example 2.11 We ask four friends, who are art lovers, to choose their favorite painter, among Cézanne, Monet, Picasso, and Dali. What is the probability that each of the four friends chooses a different painter?

SOLUTION Let A be the event of interest, namely, that each chooses a different artist. The sample space for this experiment consists of all ordered quadruples (a_1, a_2, a_3, a_4), where a_i stands for the selection of the ith person and can be any of the four painters that they may choose from (e.g. a_2 = "P" if the second person prefers Picasso). Since two or more of the a_i can be the same, Ω consists of all four-element permutations of a set with four elements, when repetitions are allowed. Thus, by Proposition 2.5, we have

$$|\Omega| = 4^4.$$

On the other hand, A contains all quadruples (a_1, a_2, a_3, a_4) in Ω with $a_i \neq a_j$ for any $i, j \in \{1, 2, 3, 4\}$, $i \neq j$. In other words, A contains all permutations (without repetitions) of a set of elements, so that

$$|A| = 4!,$$

and thus the required probability is

$$P(A) = \frac{|A|}{|\Omega|} = \frac{4!}{4^4} = \frac{24}{256} = \frac{3}{32}.$$

The implicit assumption is here that each painter has the same probability of being selected by any of the four art lovers.

Example 2.12 (The birthday problem.)
Suppose that a class contains k students. Assuming that a year has 365 days (i.e. excluding the possibility that someone is born on 29th February), show that the probability that no two students share the same birthday equals

$$p_k = \frac{(365)_k}{365^k}. \tag{2.5}$$

Verify that for $k = 23$, this probability is approximately a half. The value of $k = 23$ seems to be much smaller than what most people would have guessed, and this is the reason that this problem (apart from the birthday problem) is also known as the *birthday paradox*.

SOLUTION Suppose that $X = \{1, 2, 3, \ldots, 365\}$ represents the set of all days within a year, so that 17 represents 17th January, 45 corresponds to 14th February, and so on. For k students in a class, the sample space for all possible outcomes (birthdays of all students) is

$$\Omega = \{(a_1, a_2, \ldots, a_k) : a_i \in X, \quad i = 1, 2, \ldots, k\} = X \times X \times \cdots \times X.$$

With the above notation, a k-tuple $(a_1, a_2, \ldots, a_k)$ represents the event that the birthday of the first student is on day a_1, that of the second student on day a_2, and so on, until the kth student who has his/her birthday on day a_k. This means that the order of $a_1, a_2, \ldots, a_k$ matters, but some of the a_i's can be identical, which suggests that Ω consists in fact of all permutations of k elements among the 365 *with repetitions*. Therefore,

$$|\Omega| = 365^k. \tag{2.6}$$

Alternatively, we may see this as follows: the birthday of the first student may be on any of the 365 days of the year, and for each of these selections, the birthday of the second student might also be chosen in 365 different ways, as there are no restrictions for multiple occurrences when considering the set of possible outcomes. Continuing in this way until the kth student, and applying the multiplicative law, we obtain again (2.6).

We now define the event

A: no two students share the same birthday

and let us now count the number of elements that are contained in the set A. There are 365 ways to choose the birthday of the first student and, for each such choice, there are 364 ways to choose the birthday of the second student. For each choice of these two birthdays, there are 363 ways to choose the day of the year that the third student was born, and so on. Finally, for each selection $(a_1, a_2, \ldots, a_{k-1})$ of birthdays for the first $k-1$ students, such that $a_i \neq a_j$, there are $365 - k + 1$ days left and any of these may be chosen for the birthday of the kth student. Thus, we see that

$$|A| = 365 \cdot 364 \cdot 363 \cdots (365 - k + 1) = (365)_k.$$

This, along with (2.6), yields immediately that

$$P(A) = p_k = \frac{(365)_k}{365^k} = \frac{365 \cdot 364 \cdot 363 \cdots (365 - k + 1)}{365 \cdot 365 \cdot 365 \cdots 365}, \quad 1 \leq k \leq 365.$$

Then, the probability that at least two students share the same birthday is

$$P(A') = 1 - P(A) = 1 - \frac{(365)_k}{365^k}, \quad 1 \leq k \leq 365.$$

It is worth mentioning that this probability grows unexpectedly rapidly, as a function of the number k of students, and this is illustrated in the following table. For example, we see that with 70 students, it is highly unlikely that we do not have a common birthday. From the table below, we also see that for $k = 23$, the corresponding probability $1 - p_k$ is approximately a half.

k	2	5	10	23	25	50	70
$1 - p_k$	0.003	0.027	0.117	0.507	0.569	0.970	0.999

EXERCISES

Group A

1. A mother of three young children buys three presents for them for Christmas. She then asks her children to write down, in a piece of paper, which of the three presents they prefer, so that each one does not know the choices of the other two. What is the probability that

   (i) no two children make the same choice?

   (ii) at least two children make the same choice?

   (iii) all three children make the same choice?

2. We throw a die three times. What is the probability that

   (i) a six does not appear in any of the three throws?

   (ii) six appears at least once in the three throws?

3. What is the probability that three randomly chosen persons had their last birthday on different days of the week?

4. We ask a person to select a nonnegative integer less than 1 000 000. If all numbers have the same probability to be chosen, find the probability that the number selected does not contain the digit 5.

5. Let k and n be positive integers such that $1 \leq k \leq n$. Verify that the following relations are true:

   (i) $(n+1)! = (n+1)n!$;

   (ii) $(n)_1 = n$;

   (iii) $(n)_{n-1} = n!$;

   (iv) $(n)_k = n \cdot (n-1)_{k-1}, \quad k \geq 2$;

   (v) $(n)_k = (n-k+1) \cdot (n)_{k-1}, \quad 2 \leq k \leq n+1$;

   (vi) $(n)_k = (n)_r \cdot (n-r)_{k-r}, \quad 1 \leq r \leq k$;

   (vii) $(n)_k = \dfrac{n}{n-k} \cdot (n-1)_k, \quad 1 \leq k \leq n-1$.

6. For any nonnegative integer, n, show that $(3n)!$ is divisible by $2^n 3^n$.

7. In a company of k students with $k \leq 12$, what is the probability that at least two of them have their birthday in the same month of the year? (Assume that all months are equally likely for a birthday.)

8. A father of five children has bought seven gifts. How many distinct gift allocations are possible if each child is to receive exactly one gift (so that 2 gifts will not be given to anyone)?

9. A bus starts its route with k persons. The number of stops in this route, including the final one where all remaining passengers get off, is n.

(i) Calculate the number of different ways the k persons might leave the bus in the n stops.

(ii) For $k \leq n$, find the probability that in at least one bus stop, more than one passenger gets off the bus.

10. A bowl contains nine stones, numbered 1–9. We select successively six stones, without replacement, and write down the six-digit number that is produced from the numbers on the selected stones (in the order that they were selected).

(i) Find the probability that this six-digit number does not include the numbers 2 and 5.

(ii) What would be the answer to Part (i) if the sampling of the stones from the ball was *with replacement*?

Group B

11. We have k envelopes numbered 1 to k and n typed letters that are numbered 1 to n. In each envelope, we may put from 0 up to n letters.

(i) Under the above assumptions, in how many different ways the letters may be put into the envelopes?

(ii) For $1 \leq k \leq n$, and assuming that the allocation of the letters to the envelopes is completely random, what is the probability that an envelope will contain two or more letters?

12. A high-school class is attended by 10 boys $B_1, B_2, \ldots, B_{10}$ and 5 girls $G_1, G_2, \ldots, G_5$. After their final exams, all students are graded and ranked from 1 to 15.

(i) How many different rankings are possible?

(ii) What is the probability that all 10 boys have consecutive ranks (e.g. from rank 2 to 11)?

(iii) What is the probability that at least one girl is ranked immediately after another girl?

13. Suppose that Carol belongs to a class of k students. What is the probability that at least one of Carol's classmates has his/her birthday the same day as Carol? Compare the result with that found in Example 2.12 (the birthday problem).

14. Find the integer k that satisfies the equality

$$\frac{1}{\binom{6}{k}} - \frac{1}{\binom{7}{k}} = \frac{16}{3\binom{8}{k}}.$$

15. Suppose that n and k are two positive integers with $k \leq n$. Verify the truth of the following identities:

(i) $(n+1)_k = (n)_k + k(n)_{k-1}$;

(ii) $(n+1)_k = k! + k[(n)_{k-1} + (n-1)_{k-1} + \cdots + (k)_{k-1}]$.

Using the result of Part (ii) for $k = 2$, prove the well-known formula

$$1 + 2 + \cdots + n = \frac{n(n+1)}{2},$$

which gives the sum of the first n positive integers.

16. (The Chevalier de Méré problem[4]) Which of the following two events is more likely to occur?
    (i) we get at least one six when we throw a die four times;
    (ii) we get at least one *double six* when we have two dice and we throw them 24 times.
17. In relation to the second event of the previous problem, suppose now that the number of throws is not fixed and consider more generally the event

    A_k: we get at least one *double six* when we throw two dice k times.

    (i) Verify that the probabilities $P(A_k)$ increase as k increases.
    (ii) Find the smallest value of k such that $P(A_k) \geq 1/2$.
18. In the birthday problem (Example 2.12), use the approximation $e^x \cong 1 + x$ (valid when x is small) to obtain an approximation for the probability p_k in (2.5) as

$$p_k \cong e^{-k(k-1)/(2\times 365)}.$$

Check that with $k = 23$, the required probability is still about a half.

2.4 COMBINATIONS

In the permutations that we discussed in the preceding section, when selecting a number of elements from a set, the order in which these elements are selected is important. As already mentioned in the discussion following Example 2.10, when the order is immaterial, the resulting arrangement is called a combination. We have the following specific definition.

Definition 2.3 Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set having n (distinct) elements and k an integer less than or equal to n. Then, a combination of the n elements per k is any (unordered) collection of k distinct elements $a_1, a_2, \ldots, a_k$ from the set X.

With the terminology of set theory, we can say that each combination of the n elements per k is a subset $A = \{a_1, a_2, \ldots, a_k\}$ of X with cardinality k.

[4]Chevalier de Méré (1607–1684) was a French nobleman with a keen interest in gambling. He used to gamble on the occurrence of both the abovementioned events thinking that they both had the same probability. In fact, he thought that this probability was $2/3$, which is wrong in both cases! De Mére then posed this problem to the French mathematician Blaise Pascal and this has led to a number of early developments in probability theory. Although this problem is known as the "de Méré problem," W. Feller mentions in his book (1968) that the problem was in fact treated earlier by the Italian mathematician G. Cardano (1501–1576).

The number of all different combinations of the n elements per k is denoted by $\binom{n}{k}$. We have already seen an example wherein we enumerated the number of combinations from a set of units (Example 2.10). To give another example, consider the set $X = \{a, b, c, d, e\}$ and suppose that we want to calculate the number of different combinations of pairs from the set X. There are five choices for the first element of the pair and, for each of these, there are four possible selections for the second element to be selected. But in this way, each pair is counted twice; for example, the pairs (a, c) and (c, a) are identical. We thus find the number of combinations to be

$$\binom{5}{2} = \frac{5 \cdot 4}{2} = 10.$$

Suppose that we want to find all possible k-element permutations of n elements. One possible procedure for doing this is the following:

Step 1 We select k distinct elements out of the n elements (i.e. we create a combination of n units per k).

Step 2 For each combination created in Step 1, we consider all possible arrangements of the k units in different order.

Following this plan, in Step 1, we create all $\binom{n}{k}$ combinations of the n units per k. But for each combination, we know from the previous section that there are $k!$ different permutations of these k elements. Thus, from the multiplicative law, we see that the total number of k-unit permutations of the n elements is

$$\binom{n}{k} k!$$

Further, we also know from the last section that the number of k-element permutations among n elements equals

$$(n)_k = n(n-1)\cdots(n-k+1),$$

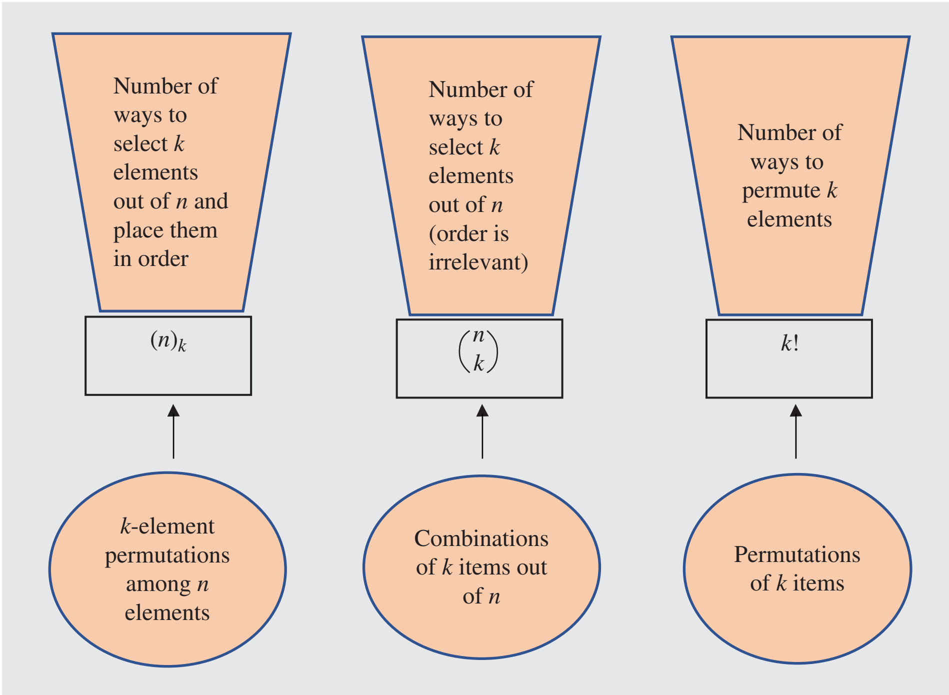
and so these two quantities must be equal. In other words,

$$\binom{n}{k} = \frac{(n)_k}{k!}.$$

We therefore have the following proposition.

Proposition 2.6 *The number* $\binom{n}{k}$ *of combinations of k elements from a set of n elements equals*

$$\binom{n}{k} = \frac{(n)_k}{k!} = \frac{n(n-1)\cdots(n-k+1)}{k!} = \frac{n!}{k!(n-k)!}, \quad 1 \le k \le n.$$



For $k = 0$, the symbol $\binom{n}{k} = \binom{n}{0}$ does not have a combinatorial interpretation. However, by convention we agree that

$$\binom{n}{0} = \frac{(n)_0}{0!} = \frac{1}{1} = 1.$$

Also, in the case $k > n$, it is not possible to create a combination of k distinct elements, since the number of elements, n, to choose from is less than k. So, we define

$$\binom{n}{k} = 0 \quad \text{for } k > n.$$

Moreover, in the special cases $n = 1, n = k, n = k - 1$, we may easily establish that

$$\binom{n}{1} = \frac{n!}{1!(n-1)!} = n,$$

$$\binom{n}{n} = \frac{n!}{0!n!} = 1,$$

$$\binom{n}{n-1} = \frac{n!}{(n-1)!(n-(n-1))!} = \frac{n!}{(n-1)!1!} = n.$$

In Table 2.1, we present the values of $\binom{n}{k}$ for $n = 0, 1, 2, \ldots, 10$ and $0 \le k \le n$.

The following result contains two formulas that, among other things, are useful for calculating numerical values of the quantities $\binom{n}{k}$ for $1 \le k \le n$.

$n \backslash k$	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6	1	6	15	20	15	6	1				
7	1	7	21	35	35	21	7	1			
8	1	8	28	56	70	56	28	8	1		
9	1	9	36	84	126	126	84	36	9	1	
10	1	10	45	120	210	252	210	120	45	10	1

Table 2.1 Values of $\binom{n}{k}$ for $n = 0, 1, 2 \ldots, 10$ and $0 \leq k \leq n$.

Proposition 2.7 *For the combinations* $\binom{n}{k}$, $1 \leq k \leq n$, *the following equalities hold:*

$$\binom{n}{k} = \binom{n}{n-k}$$

and

$$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}. \tag{2.7}$$

Proof: The first result is immediate since

$$\binom{n}{n-k} = \frac{n!}{(n-k)!(n-(n-k))!} = \frac{n!}{k!(n-k)!} = \binom{n}{k}.$$

The result is also obvious from a combinatorial viewpoint, since each selection of k distinct elements out of n corresponds uniquely to a selection of the remaining $n-k$ elements (this argument has in fact already been used in Part (ii) of Example 2.10).

In order to prove (2.7), assume first that $1 < k < n$. Then, manipulating the right-hand side of the equality, we get successively

$$\begin{aligned}
\binom{n-1}{k-1} + \binom{n-1}{k} &= \frac{(n-1)!}{(k-1)!(n-k)!} + \frac{(n-1)!}{k!((n-1)-k)!} \\
&= \frac{k(n-1)!}{k!(n-k)!} + \frac{(n-k)(n-1)!}{k!(n-k)!} \\
&= \frac{(n-1)![k+(n-k)]}{k!(n-k)!} = \frac{(n-1)!n}{k!(n-k)!} \\
&= \frac{n!}{k!(n-k)!} = \binom{n}{k}.
\end{aligned}$$

The result is also true for $k = 1$ and $k = n$, since for these two cases we have

$$\binom{n}{1} = \binom{n-1}{0} + \binom{n-1}{1} \quad \text{(equivalently, } n = 1 + (n-1)\text{)}$$

and

$$\binom{n}{n} = \binom{n-1}{n-1} + \binom{n-1}{n} \quad \text{(equivalently, } 1 = 1 + 0\text{)}$$

and the proof of the proposition is thus complete. □

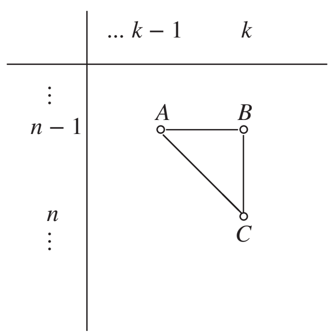
The first part of the last proposition is frequently employed when we want to calculate the number of ways of choosing k elements out of n, when n is large and the value of k is close to that of n. For instance, instead of writing

$$\binom{100}{96} = \frac{(100)_{96}}{96!} = \frac{100 \cdot 99 \cdot 98 \cdots 5}{96 \cdot 95 \cdots 2 \cdot 1}$$

and then cancelling out the terms $96 \cdot 95 \cdots 5$, we can write directly

$$\binom{100}{96} = \binom{100}{100-96} = \binom{100}{4} = \frac{(100)_4}{4!} = \frac{100 \cdot 99 \cdot 98 \cdot 97}{4 \cdot 3 \cdot 2 \cdot 1}.$$

Incidentally, Equation (2.7) is known as **Pascal's triangle**. The term triangle is due to the fact that, when presenting the values of $\binom{n}{k}$ in an array, the quantities $\binom{n-1}{k-1}, \binom{n-1}{k}, \binom{n}{k}$ appear as vertices A, B, C of a triangle from a horizontal, a vertical, and a diagonal side (see also Figure 2.4).



From (2.7), the value that corresponds to point C of the triangle is found by summing up the values at the points A and B. Pascal suggested the point C to be placed between points A and B, as shown in Figure 2.4. In fact, the set of numbers that form Pascal's triangle was well-known before Pascal (the first explicit reference seems to be in India during the tenth century AD in commentaries on the *Chandas Shastra*, an ancient Indian book written by Pingala between the fifth and second centuries BC). However, Pascal considered various applications of it and was the first one to organize all the information together in his treatise, *Traité du triangle arithmétique* (1653).

$$
\begin{array}{ccccccccccccc}
 & & & & & & 1 & & & & & & \\
 & & & & & 1 & & 1 & & & & & \\
 & & & & 1 & & 2 & & 1 & & & & \\
 & & & 1 & & 3 & & 3 & & 1 & & & \\
 & & 1 & & 4 & & 6 & & 4 & & 1 & & \\
 & 1 & & 5 & & 10 & & 10 & & 5 & & 1 & \\
1 & & 6 & & 15 & & 20 & & 15 & & 6 & & 1
\end{array}
$$

Figure 2.4 Pascal's triangle.

Example 2.13 A class teacher gives her class a set of 20 problems and she tells the students that 7 among these will form the exam. By the day of the exam, Patty has managed to solve 12 of the problems, but she has no idea as to how to do the rest. What is the probability that

(i) she will answer correctly five exam questions?

(ii) she will answer correctly at least five exam questions?

SOLUTION First, we consider the sample space for this experiment and find how many elements it contains. The teacher may choose any set of 7 problems among the 20, and since it is clear that no repetitions are allowed, the number of elements of the sample space Ω is $\binom{20}{7}$.

(i) The sample space consists of all 7-tuples among the set of 20 problems. The favorable outcomes for this part of the example are those seven-tuples that contain five problems for which Patty knows the answer and two problems that she does not know. For simplicity, and without loss of generality, assume that all possible problems for the exam are numbered $1, 2, \dots, 20$, and those that Patty has worked out are problems $1, 2, \dots, 12$. So, any favorable outcome is any choice of 7 numbers from 1 to 20 such that

$$\underbrace{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12}_{5 \text{ among these}} \quad \underbrace{13, 14, 15, 16, 17, 18, 19, 20}_{2 \text{ among these}}.$$

This entails making two selections: choosing 5 numbers out of 12, which can be done in $\binom{12}{5}$ different ways, and choosing 2 numbers out of 8, which can be done in $\binom{8}{2}$ different ways. By the multiplicative law, we thus obtain that there are

$$\binom{12}{5} \cdot \binom{8}{2} = \frac{12!}{5!7!} \cdot \frac{8!}{2!6!} = \frac{8 \cdot 9 \cdot 10 \cdot 11 \cdot 12}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} \cdot \frac{7 \cdot 8}{1 \cdot 2} = 792 \cdot 28 = 22\ 176$$

favorable outcomes. As a result, the required event has probability

$$\frac{22\ 176}{\binom{20}{7}} = \frac{22\ 176}{\frac{20!}{13!7!}} = \frac{22\ 176}{77\ 520} = 0.286.$$

(ii) We define the events

A_1: Patty answers correctly five exam questions,
A_2: Patty answers correctly six exam questions,
A_3: Patty answers correctly seven exam questions.

Then, we seek the probability

$$P(A_1 \cup A_2 \cup A_3).$$

As it is clear that A_1, A_2, A_3 are disjoint events, we have

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3). \tag{2.8}$$

The probability of the event A_1 was found in Part (i) above. Working in a similar way, we obtain for the event A_2

$$P(A_2) = \frac{|A_2|}{|\Omega|} = \frac{\binom{12}{6}\binom{8}{1}}{\binom{20}{7}} = \frac{\frac{12!}{6!6!} \cdot \frac{8!}{7!1!}}{\frac{20!}{13!7!}} = \frac{924 \cdot 8}{77\ 520} = 0.0954$$

and for the event A_3

$$P(A_3) = \frac{|A_3|}{|\Omega|} = \frac{\binom{12}{7}\binom{8}{0}}{\binom{20}{7}} = \frac{\frac{12!}{7!5!} \cdot \frac{8!}{8!0!}}{\frac{20!}{13!7!}} = \frac{792 \cdot 1}{77\ 520} = 0.0102.$$

Upon substituting the above results, along with that of Part (i), into (2.8), we finally obtain the required probability to be

$$P(A_1 \cup A_2 \cup A_3) = 0.286 + 0.0954 + 0.0102 = 0.3916.$$

One of the main themes of the present chapter is that the number of arrangements that can be formed, when selecting a number of elements from a given set, differs depending on whether the order in which these elements are drawn is important or not. However, since this applies equally to the numerator and the denominator in the classical definition of probability, when it comes to probability questions, it is often the case that the results are the same, either we treat the order as relevant or not. The next example illustrates this point.

Example 2.14 In the US parliament, suppose that a three-member committee needs to be formed for deciding on an important issue. There are 40 parliament members who are eligible to take part in this committee, 22 of whom are Democrats and 18 are Republicans. If the three members to participate in the committee are chosen at random among these 40 persons, what is the probability that the committee will contain exactly two Democrats?

SOLUTION Suppose that the order in which the three members are chosen for the committee is important (e.g. the person who is chosen first is the committee's chairperson, etc.). In this case, the number of possible outcomes (the cardinality of the sample space) is $40 \cdot 39 \cdot 38 = 59\,280$. For the number of favorable outcomes, we consider three distinct cases:

- the first two persons chosen are Democrats followed by a Republican;
- the first and the third persons chosen are Democrats, while the second one is a Republican;
- the first person chosen is a Republican, followed by two Democrats.

As these events are mutually exclusive, the total number of favorable outcomes is the sum of the number of ways that each of these may occur. For the first event, there are $22 \cdot 21 \cdot 18$ distinct ways; for the second, there are $22 \cdot 18 \cdot 21$ ways, and for the third there are $18 \cdot 22 \cdot 21$ ways that it may occur (notice that the number is the same for all three events). Therefore, the number of favorable outcomes is

$$3 \cdot 22 \cdot 21 \cdot 18 = 24\,948.$$

Hence, the required probability becomes

$$\frac{24\,948}{59\,280} = \frac{2079}{4940} \cong 0.420\,85.$$

Suppose we now regard the order in which the three members are selected for the committee to be irrelevant. In this case, the sample space contains $\binom{40}{3}$ elements, while there are $\binom{22}{2}$ ways for the two Democrats and $\binom{18}{1}$ ways for the Republican to be chosen. Consequently, the probability that there are exactly two Democrats in the committee in this case becomes

$$\frac{\binom{22}{2}\binom{18}{1}}{\binom{40}{3}} = \frac{\frac{22!}{20!2!} \cdot \frac{18!}{17!1!}}{\frac{40!}{37!3!}} = \frac{\frac{22 \cdot 21}{2} \cdot 18}{\frac{40 \cdot 39 \cdot 38}{3 \cdot 2 \cdot 1}}$$

$$= \frac{4158}{9880} = \frac{2079}{4940} \cong 0.420\,85,$$

which is exactly the same answer as the one obtained for the other case.

In the last section, we have seen that, when considering the k-element permutations of k among n elements, it makes sense to consider such permutations with repetitions. In analogy with that, we now consider combinations with repetitions. For instance, suppose we have the set

$$\{x_1, x_2, x_3, x_4\}.$$

Then we can form $\binom{4}{3} = 4$ usual combinations by choosing three elements from the set X. Specifically, these selections are

$$\{x_1, x_2, x_3\}, \quad \{x_1, x_2, x_4\}, \quad \{x_1, x_3, x_4\}, \quad \{x_2, x_3, x_4\}.$$

When allowing an element of X to be chosen more than once, then we have additionally the selections given in Table 2.2, corresponding to three elements from the set X.

In total, using the four combinations with distinct elements and those in Table 2.2, we see that the number of choosing $k = 3$ elements from a set of $n = 4$ elements is 20.

Definition 2.4 Let $X = \{x_1, x_2, \dots, x_n\}$ be a finite set consisting of n elements and k be a positive integer. Then a **k-combination with repetitions** (sometimes also called a k-multicombination) is a finite collection of k, not necessarily distinct, elements of the set X.

Element of the set X	Combinations using an element more than once			
x_1	x_1, x_1, x_2	x_1, x_1, x_3	x_1, x_1, x_4	x_1, x_1, x_1
x_2	x_2, x_2, x_1	x_2, x_2, x_3	x_2, x_2, x_4	x_2, x_2, x_2
x_3	x_3, x_3, x_1	x_3, x_3, x_2	x_3, x_3, x_4	x_3, x_3, x_3
x_4	x_4, x_4, x_1	x_4, x_4, x_1	x_4, x_4, x_1	x_4, x_4, x_1

Table 2.2 Choosing three elements with repetitions from a set of four elements.

Suppose

$$a_1, a_2, \ldots, a_k$$

is a k-combination with repetitions from the set X. Then, since the a_i's are not necessarily distinct, we shall not use the notation $\{a_1, a_2, \ldots, a_k\}$ of a set; in set theory, the elements of a set are necessarily distinct. The notation we shall use for the number of k-combinations with repetitions from a set that has n elements is $\begin{bmatrix} n \\ k \end{bmatrix}$.

Our next target is to give a general formula for the number of k-combinations with repetitions. Before this result, however, we present an example to help understanding the enumeration process in this case. Note in particular that, when combinations are made with repetitions, k may be greater than n.

Example 2.15 A small car sales store wants to place an order for five cars of a new type to the manufacturer. The car is available in three colors: white (W), silver-gray (S), and red (R). How many different orders for the five cars could be placed?

SOLUTION Here, we have $n = 3$ different elements to choose from, namely,

$$X = \{W, S, R\}$$

and we want to form combinations with $k = 5$ elements. Some typical such choices are

$$W, W, W, W, W \quad W, W, S, S, W \quad R, W, S, R, S.$$

Note that, since we are considering combinations (rather than permutations, as in the previous section), the order in which the elements appear in a combination is irrelevant; so, the combinations R, W, S, R, S and W, S, R, S, R are considered to be identical. It is only *the number of cars from each color* present in each combination that we have to consider. Thus, the different selections can be summarized as in Table 2.3.

Number of cars ordered			
White	Silver-gray	Red	Total
5	0	0	5
4	1	0	5
3	2	0	5
3	1	1	5
3	0	2	5
.	.	.	.
.	.	.	.
.	.	.	.

Table 2.3 Different selections for the five cars ordered.

W	S	R
● ● ● ● ●		
● ● ● ●	●	
● ● ●	● ●	
● ● ●	●	●
● ● ●		● ●
.	.	.
.	.	.
.	.	.

Table 2.4 A different way of presenting combinations with repetitions.

Representing the presence of a car with a specific color in the sample of five cars by a dot, an alternative presentation is given by Table 2.4.

If in Table 2.4, we replace vertical lines (separating the different colors) by the number 1 and dots by 0, each row (combination) corresponds to a seven-unit arrangement of five 0's and two 1's; more explicitly, we will have

$$
\begin{gathered}
0000011\\
0000101\\
0001001\\
0001010\\
0001100\\
\dots
\end{gathered}
$$

Conversely, each row of $7 = 3 + 5 - 1$ binary (0 or 1) digits of which $k = 5$ digits are 0, determines a possible order of cars. For instance, the row

$$0001010$$

indicates that the order of five cars to the manufacturer consists of

- three white cars (there are as many zeros before the appearance of the first 1);
- one silver-gray car (there is one zero between the first and the second appearance of 1);
- one red car (there is one zero after the second appearance of the digit 1).

Notice that, although there are three different colors, we only need two 1's in each row, each of them acting as a separator between two consecutive colors.

From the above, we see that the problem of finding the total number of k-combinations from a set of n elements reduces to that of enumerating the ways in which we can put $n + k - 1 = 7$ binary digits in a row, with the restriction that five of these digits must be zeros. In view of this, yet another alternative (but equivalent) formulation of the problem is as follows:

- Consider that we have seven places

$$
\begin{array}{ccccccc}
\sqcup & \sqcup & \sqcup & \sqcup & \sqcup & \sqcup & \sqcup \\
0 & 0 & 0 & 0 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0
\end{array}
$$

- We choose two of these seven places to put the ones (and this can be done in $\binom{7}{2}$ ways);
- We fill in the remaining five places with five zeros.

Therefore, we see that the total number of different five-car color orders that the store can make is equal to

$$\binom{7}{2} = \frac{7!}{2!5!} = \frac{7 \cdot 6}{1 \cdot 2} = 21.$$

By generalizing the procedure in the above example, we arrive at the following proposition.

Proposition 2.8 *The number,* $\begin{bmatrix} n \\ k \end{bmatrix}$, *of ways of selecting* k *elements from a set of* n *elements, with repetitions allowed, is given by*

$$
\begin{aligned}
\begin{bmatrix} n \\ k \end{bmatrix} &= \binom{n+k-1}{k} = \binom{n+k-1}{n-1} = \frac{(n+k-1)!}{k!(n-1)!} \\
&= \frac{n(n+1)(n+2)\cdots(n+k-1)}{k!}, \qquad n \geq 1,\ k \geq 1.
\end{aligned}
$$

Proof: Let $X = \{x_1, x_2, \ldots, x_n\}$ be the set of n elements to choose from. A typical combination with repetitions will have the form given in Figure 2.5 wherein the number of $\bullet$ under each x_i, for $i = 1, 2, \ldots, n$, indicates how many times the element x_i appears in the combination.

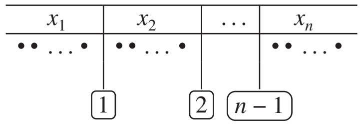


Figure 2.5 Combinations with repetitions.

Replacing vertical lines by 1's and each symbol • by a zero gives a row of the form

$$\underbrace{00\ldots0100\ldots01\ldots100\ldots0}_{\text{Number of 1's is } n-1\text{, number of 0's is } k}.$$

Each such row has a total of $n+k-1$ elements and is uniquely characterized by the places that the k zeros occur (or, equivalently, the places where the $n-1$ ones occur). Thus, there are exactly

$$\binom{n+k-1}{k} = \binom{n+k-1}{n-1}$$

different combinations with repetitions and this completes the proof of the proposition. □

Finally, for the combinations with repetitions wherein we select k elements from n distinct elements, we have the following simple results: first, for $k=0$,

$$\left[\begin{matrix} n \\ 0 \end{matrix}\right] = \binom{n+0-1}{0} = 1, \quad n \geq 1,$$

while for $k=1$ and $k=n$, we have respectively

$$\left[\begin{matrix} n \\ 1 \end{matrix}\right] = \binom{n+1-1}{1} = \binom{n}{1} = n$$

and

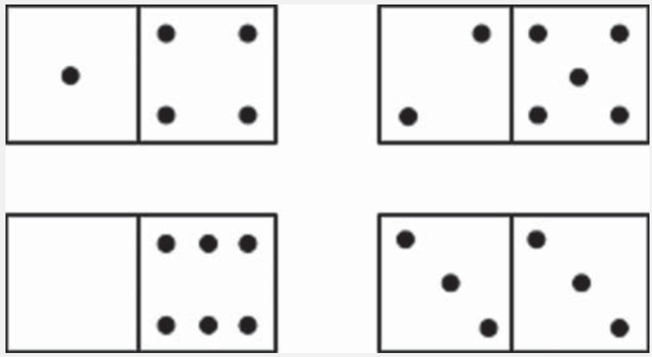
$$\left[\begin{matrix} n \\ n \end{matrix}\right] = \binom{n+n-1}{n} = \binom{2n-1}{n} = \frac{(2n-1)!}{n!(n-1)!}.$$

Another result is that for $n=1$,

$$\left[\begin{matrix} 1 \\ k \end{matrix}\right] = \binom{1+k-1}{k} = \binom{k}{k} = 1$$

(If there is only one element to choose from, all k elements we select must be equal to that and so there is just one combination with repetitions.)

Example 2.16 Each of the two sides on a tile of a domino gaming set has a number of k spots (or pips), where k is an integer in the set $\{0,1,2,3,4,5,6\}$. Thus, we can have a double blank tile, a 3–4 tile, a 5–1 tile, etc. We do not distinguish between the two sides so that a 5–1 tile is identical with a 1–5 tile in the game.



(i) How many different tiles can be formed under the above assumptions?

(ii) If we select a tile at random, what is the probability that the numbers of spots in the two sides

(a) are equal?

(b) differ by two?

SOLUTION

(i) The sample space Ω of the experiment consists of the combinations of $k = 2$ elements from the set $\{0, 1, 2, 3, 4, 5, 6\}$, which has seven elements, when repetitions are allowed. So, the number of different domino tiles that can be formed is

$$|\Omega| = \left[\begin{matrix} 7 \\ 2 \end{matrix} \right] = \binom{7+2-1}{2} = \frac{8!}{2!6!} = 28.$$

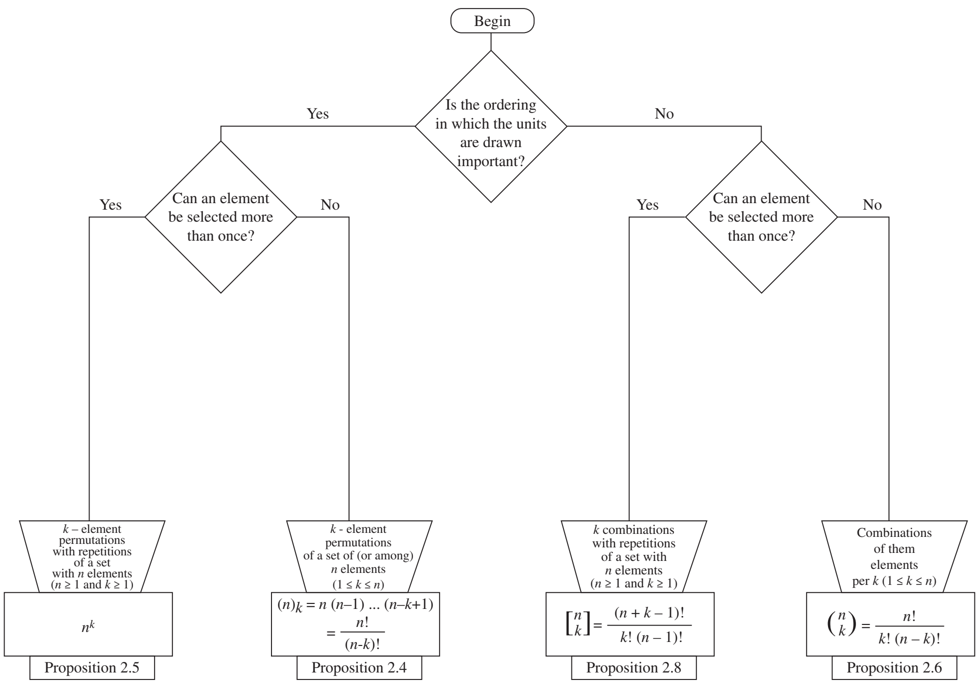
(ii) (a) The tiles where the two sides are equal are the double blank (or a 0–0 tile), a tile with two 1's, two 2's, and so on, until the two 6's. It is clear that there are seven such tiles, and so the event A that both sides of the selected tile are equal has probability

$$P(A) = \frac{|A|}{|\Omega|} = \frac{7}{28} = \frac{1}{4}.$$

(b) Let B be the event that the two sides of the domino piece selected differ by two. The favorable outcomes are the pairs 0–2, 1–3, 2–4, 3–5, 4–6, and this gives the required probability to be

$$P(B) = \frac{|B|}{|\Omega|} = \frac{5}{28}.$$

The following diagram summarizes the main results of this and the previous section.


Begin
Is the ordering in which the units are drawn important?
Yes
No
Can an element be selected more than once?
Yes
No
Can an element be selected more than once?
Yes
No
k – element permutations with repetitions of a set with n elements ($n \geq 1$ and $k \geq 1$)
n^k
Proposition 2.5
k - element permutations of a set of (or among) n elements ($1 \leq k \leq n$)
$(n)_k = n\,(n-1)\ldots(n-k+1) = \frac{n!}{(n-k)!}$
Proposition 2.4
k combinations with repetitions of a set with n elements ($n \geq 1$ and $k \geq 1$)
$\left[{n \atop k}\right] = \frac{(n+k-1)!}{k!\,(n-1)!}$
Proposition 2.8
Combinations of them elements per k ($1 \leq k \leq n$)
$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$
Proposition 2.6

EXERCISES

Group A

1. In a volleyball tournament, six teams take part and each team has to play every other team exactly once. How many matches will be played in total?

2. At the repair department of a store that sells electric appliances, there are currently six TV sets and eight DVD players waiting to be repaired. The staff of the store can repair a total of six appliances on any particular day. If the ones to be repaired today are chosen randomly, what is the probability that the following items will get repaired?

   (i) Three TV sets and three DVD players;

   (ii) at most two DVD players.

3. At the National Basketball Association (NBA) championships, 30 teams participate, 15 of which belong to the Eastern Conference and 15 to the Western Conference. At the end of the regular season, the best eight teams from each conference qualify for the playoffs. How many different selections are possible for the 16 teams to qualify for the playoffs?

4. A company has 12 senior and 20 junior employees. The members of staff in the company want to form a five-member committee. Find how many different ways are possible to select the committee, under the restriction that at least two senior and one junior staff members must take part in it.

5. A box has 25 balls numbered 1–25. Suppose four balls are selected randomly without replacement. What is the probability that the smallest number in the balls chosen is at least six?

6. If n and k are positive integers, reduce the following sums to an expression involving just one combinatorial term:

$$\binom{2n}{n+1} + 2\binom{2n}{n} + \binom{2n}{n-1}$$

   and

$$\binom{n}{k-1} + \binom{n-1}{k} + \binom{n-1}{k-1}.$$

7. For nonnegative integers n, k, and r, prove the following identities:

   (i) $k\binom{n}{k} = n\binom{n-1}{k-1}, \quad 1 \le k \le n;$

   (ii) $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}, \quad 1 \le k \le n;$

   (iii) $\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}, \quad 0 \le k < n;$

   (iv) $\binom{n}{k}\binom{k}{r} = \binom{n}{r}\binom{n-r}{k-r}, \quad 0 \le r \le k \le n;$

   (v) $(k)_r\binom{n}{k} = (n)_r\binom{n-r}{k-r}, \quad 0 \le r \le k \le n;$

   (vi) $n\binom{n}{k} = (k+1)\binom{n}{k+1} + k\binom{n}{k} = k\binom{n+1}{k+1} + \binom{n}{k+1}.$

8. (i) For any positive integer n, show that

$$\binom{\binom{n}{2}}{2} = 3\binom{n+1}{4}.$$

(ii) For $n \geq 6$, verify that the following inequality holds:

$$\binom{\binom{n}{2}}{2} < \binom{\binom{n}{3}}{2}.$$

9. In the Mathematics Department of a University, there are 80 freshmen, 65 sophomore, 70 junior, and 90 senior students. If five students have been chosen for the chess team, what is the probability that in the chess team there are

(i) exactly two juniors?

(ii) five students from the same year in their studies?

10. In a forested area, it is known that there exist 300 animals from a protected species. A scientific team selects 100 of these animals, marks them and sets them free. After a certain period, so that the marked animals are well mixed in the forest along with other animals, the scientists selects another set of 100 animals. Find the probability that exactly 10 of them have been previously marked.

11. Mary has 10 coins in her pocket, of which 4 are gold ones and 6 are silver ones. She selects four coins at random. Find the probability that exactly two of them are gold if

(i) the selection of the coins is without replacement,

(ii) the selection of the coins is with replacement,

and compare your results.

12. A company has 25 trucks, among which 5 have fuel emissions above a certain level. If one of the company's technicians selects six trucks at random to check for their fuel emissions, what is the probability that among them there are

(i) exactly three trucks with high emissions?

(ii) at most two trucks with high emissions?

(iii) at least one truck with high emissions and at least one truck with emissions at normal level?

13. From an usual pack of 52 cards, we select 5 cards. What is the probability that at least one King is selected if

(i) after each card is chosen, the card is put back for the next selection?

(ii) when a card is chosen, it is removed from the pack for the next selection?

14. An usual pack of cards has 26 black cards and 26 red ones. If we split the pack into two halves with 26 each, what is the probability that each of these parts has exactly 13 black cards?

15. For positive integers n and k, prove the following identities:

    (i) $\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} k+1 \\ n-1 \end{bmatrix}$;

    (ii) $\begin{bmatrix} n \\ k \end{bmatrix} = \begin{bmatrix} n-1 \\ k \end{bmatrix} + \begin{bmatrix} n \\ k-1 \end{bmatrix}, \quad n \geq 2$;

    (iii) $k \begin{bmatrix} n \\ k \end{bmatrix} = (n+k-1) \begin{bmatrix} n \\ k-1 \end{bmatrix}$;

    (iv) $\begin{bmatrix} n \\ k \end{bmatrix} = \dfrac{n}{k} \begin{bmatrix} n+1 \\ k-1 \end{bmatrix}$;

    (v) $\begin{bmatrix} n \\ k \end{bmatrix} = \dfrac{n+k-1}{n-1} \begin{bmatrix} n-1 \\ k \end{bmatrix}$;

    (vi) $\begin{bmatrix} n \\ n \end{bmatrix} = \dfrac{(2n)!}{(n!\sqrt{2})^2} = \dfrac{1}{2}\dbinom{2n}{n}$.

16. In a lottery that has 49 numbers, we select six numbers at random and buy a ticket. What is the probability that we get

    (i) six winning numbers?

    (ii) five winning numbers?

    (iii) four winning numbers?

    (iv) at least four winning numbers?

Group B

17. Twenty couples attend a Halloween party and enter a competition for the best outfit of the night. There are three prizes for the best outfits and prizes are won individually. What is the probability that among the three winners,

    (i) there are only women?

    (ii) there are two men and a woman?

    (iii) there is no couple?

18. A committee of 5 people is to be chosen from a group of 12 people. How many different selections are available if among the 12 persons there are two who do not agree to serve in the committee together?

19. In the warehouse of a car tyre factory, there are 1200 tyres of a certain type. Among these, there are 25 tyres that have a defect, 30 tyres for which the sell-by date has expired, while the remaining 1145 are suitable for sale. A car tyre retail store orders 20 tyres of this type from the factory. If the 20 tyres are randomly selected (from the 1200 tyres) and sent to the retail store, what is the probability that among them there are

    (i) exactly five tyres suitable for use?

    (ii) exactly 3 tyres with a defect and 15 tyres suitable for use?

    (iii) exactly 15 tyres suitable for use and at least 3 tyres for which the sell-by date has expired?

20. Suppose that each side of every piece in the domino game (see Example 2.16) is marked with a number of spots, where this number is chosen from the set $\{0, 1, 2, \dots, n-1\}$. For which value of n, 66 different tiles can be produced (at least)? In that case, i.e. in a domino with 66 different tiles, what is the probability if we select a tile at random that the numbers of spots on the two sides of the tiles

    (i) are equal (e.g. a double 4)?

    (ii) differ by one (e.g. 4–5)?

    (iii) differ by two (e.g. 3–5)?

2.5 THE BINOMIAL THEOREM

The combinations $\binom{n}{k}$, which denote the number of ways of choosing k elements among n identical elements, are perhaps the most important quantities in combinatorial analysis and they are called **binomial coefficients**. This term is due to their appearance as coefficients of powers of a and b in the binomial expansion of the nth power of the sum $a + b$. To be specific, we have the following result.

Proposition 2.9 *Let a and b be any real numbers and n be a positive integer. Then, the nth power of the sum a + b is given by the expansion*

$$\begin{aligned}(a+b)^n &= \binom{n}{0}a^n + \binom{n}{1}a^{n-1}b + \cdots + \binom{n}{n-1}ab^{n-1} + \binom{n}{n}b^n \\ &= \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k} = \sum_{k=0}^{n}\binom{n}{k}a^{n-k}b^k. \end{aligned} \tag{2.9}$$

Proof: We shall present two different proofs for the result, one using combinatorial methods and the other based on the principle of mathematical induction.

Combinatorial proof: Since

$$(a+b)^n = \underbrace{(a+b)\cdot(a+b)\cdots(a+b)}_{n \text{ times}},$$

in order to establish the expansion (i.e. to write $(a+b)^n$ as a sum of powers of a and b), we must choose one of two numbers, a or b, from each term $(a + b)$ above and multiply them together. If from the n terms that are all equal to $a + b$, a is chosen k times and b is chosen $n - k$ times, we get the product $a^k b^{n-k}$ for any $k = 0, 1, 2, \dots, n$.

But, choosing k times the number a among the n terms can be done in $\binom{n}{k}$ ways, which means that the coefficient of $a^k b^{n-k}$ in the expansion is $\binom{n}{k}$, and this completes the combinatorial proof.

Proof by induction: For $n = 0$, the formula in the proposition is trivially true, since

$$(a+b)^0 = 1 = \sum_{k=0}^{0}\binom{0}{k}a^k b^{0-k} = \binom{0}{0}a^0 b^0.$$

Assume now that the result is true for $n = r$, i.e.

$$(a+b)^r = \sum_{k=0}^{r} \binom{r}{k} a^k b^{r-k}. \tag{2.10}$$

We shall show that it is also true for $n = r + 1$, namely,

$$(a+b)^{r+1} = \sum_{k=0}^{r+1} \binom{r+1}{k} a^k b^{r+1-k}.$$

Using (2.10), we may write

$$\begin{aligned}
(a+b)^{r+1} &= (a+b)(a+b)^r = (a+b)\sum_{k=0}^{r} \binom{r}{k} a^k b^{r-k} \\
&= \sum_{k=0}^{r} \binom{r}{k} a^{k+1} b^{r-k} + \sum_{k=0}^{r} \binom{r}{k} a^k b^{r-k+1} \\
&= \binom{r}{0} ab^r + \binom{r}{1} a^2 b^{r-1} + \cdots + \binom{r}{r-1} a^r b + \binom{r}{r} a^{r+1} \\
&\quad + \binom{r}{0} b^{r+1} + \binom{r}{1} ab^r + \cdots + \binom{r}{r-1} a^{r-1} b^2 + \binom{r}{r} a^r b.
\end{aligned}$$

Now, upon recalling Pascal's triangle (see (2.7))

$$\binom{r}{j-1} + \binom{r}{j} = \binom{r+1}{j}, \quad j = 1, 2, \ldots, r,$$

and the fact that

$$\binom{r}{0} = \binom{r}{r} = 1,$$

we obtain

$$(a+b)^{r+1} = b^{r+1} + \binom{r+1}{1} ab^r + \binom{r+1}{2} a^2 b^{r-1} + \cdots + \binom{r+1}{r} a^r b + \binom{r+1}{r+1} a^{r+1},$$

or, more compactly,

$$(a+b)^{r+1} = \sum_{k=0}^{r+1} \binom{r+1}{k} a^k b^{r+1-k}.$$

The last expression shows that the desired formula holds for $r + 1$, and by the principle of induction, this shows that the formula is valid for all nonnegative integer values of n.

By the way, the last equality in (2.9) follows easily since, by a change of variable $j = n - k$, we get

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{n-k} a^k b^{n-k} = \sum_{j=0}^{n} \binom{n}{j} a^{n-j} b^j. \qquad \square$$

In Proposition 2.9, we assumed that n is an integer. However, the binomial theorem can be extended to the case when n is not an integer, but an arbitrary real number. This was done originally by Sir Isaac Newton around 1665. The proof of the general case, however, is rather involved and it is therefore not presented here. Instead, we shall give a proof of the binomial expansion of $(a+b)^n$ when n is a negative integer and $a = 1$, $|b| < 1$. Note that in this case the summation extends over an infinite range of values.

Proposition 2.10 ***(Binomial expansion for negative integer powers)*** *Let t be a real number such that $|t| < 1$ and n be a positive integer. Then, the following formula holds:*

$$(1-t)^{-n} = \frac{1}{(1-t)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} t^k.$$

Proof: Using the identity

$$\frac{1}{1-t} = 1 + t + t^2 + \cdots \quad \text{for } -1 < t < 1,$$

we can write

$$\frac{1}{(1-t)^n} = \underbrace{(1+t+t^2+\cdots)(1+t+t^2+\cdots)\cdots(1+t+t^2+\cdots)}_{n \text{ terms}} \quad . \tag{2.11}$$

Multiplying term by term the quantities on the right-hand side above, we obtain an expression of the form

$$\frac{1}{(1-t)^n} = \sum_{k=0}^{\infty} a_{n,k} t^k$$

for some suitable positive integers $a_{n,k}$. The power t^k, $k \geq 0$, may arise from (2.11) in the following way: among the n factor terms in the brackets there, we select k terms with replacement and, if a bracket is chosen i times (for $i \geq 0$), then from this bracket we choose the term t^i.

It therefore follows that the number of ways in which the term t^k arises on the right side of (2.11) is equal to the number of combinations with repetitions when we choose k elements (bracket terms) among n such elements. This shows that the coefficient, $a_{n,k}$ of the power t^k, equals

$$a_{n,k} = \left[\begin{matrix} n \\ k \end{matrix} \right] = \binom{n+k-1}{k},$$

which completes the proof. □

The binomial theorem can be used to establish several combinatorial identities, mostly involving binomial coefficients. We shall give a few typical examples of such identities, while some others are presented in the exercises of this section.

The methods we use to obtain these identities can be classified as being one of the following three kinds. In particular, identities that may arise:

I. directly from Propositions 2.9 and 2.10 as special cases for a and b or the variable t there;

II. after one or more transformations in the part which contains the binomial coefficients, so that the summation is easier to handle;

III. by differentiating or integrating both sides of a binomial expansion or other identities that stem from that.

In the following two examples, we deal with typical situations that best illustrate the above techniques.

Example 2.17 Calculate the sums

$$S_1 = \sum_{k=0}^{n} \binom{n}{k}, \quad S_2 = \sum_{k=0}^{n} (-1)^k \binom{n}{k},$$

as well as the following ones:

$$S_3 = \sum_{k=0}^{\left[\frac{n-1}{2}\right]} \binom{n}{2k+1}, \quad S_3' = \sum_{k=0}^{\left[\frac{n}{2}\right]} \binom{n}{2k}.$$

SOLUTION Putting $a = b = 1$ in the formula of the binomial expansion (Proposition 2.9), we find

$$(1+1)^n = \sum_{k=0}^{n} \binom{n}{k} 1^k = \sum_{k=0}^{n} \binom{n}{k},$$

which means that

$$S_1 = \sum_{k=0}^{n} \binom{n}{k} = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{n} = 2^n. \tag{2.12}$$

Similarly, for $a = -1$ and $b = 1$, the binomial formula gives

$$(-1+1)^n = \sum_{k=0}^{n} (-1)^k \binom{n}{k}$$

so that, for $n = 1, 2, \ldots$, we have

$$S_2 = \sum_{k=0}^{n} (-1)^k \binom{n}{k} = \binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \cdots + (-1)^n \binom{n}{n} = 0. \tag{2.13}$$

Adding Equations (2.12) and (2.13) side by side, we see that the binomial coefficients cancel except those that are in the form $\binom{n}{2j}$, for $j = 0, 1, \ldots$, such that $2j \le n$. In fact, each of the remaining binomial coefficients appears twice, so that

$$S_1 + S_2 = 2 \sum_{j=0}^{\left[\frac{n}{2}\right]} \binom{n}{2j} = 2S_3',$$

and using the results found above for S_1 and S_2 we conclude that

$$S_3' = \frac{1}{2}(S_1 + S_2) = \frac{1}{2}(2^n + 0) = 2^{n-1}.$$

Further, subtracting each side of (2.13) from (2.12), we see that the binomial coefficients that *do not* cancel are precisely those of the form $\binom{n}{2j+1}$, with $2j + 1 \le n$. Consequently, we obtain

$$S_1 - S_2 = 2 \sum_{j=0}^{\left[\frac{n-1}{2}\right]} \binom{n}{2j+1} = 2S_3,$$

and so we finally deduce that

$$S_3 = \frac{1}{2}(S_1 - S_2) = \frac{1}{2}(2^n - 0) = 2^{n-1}.$$

Example 2.18 Suppose n and r are two nonnegative integers such that $r \le n$ and $x \in \mathbb{R}$. Calculate the sum

$$S = \sum_{k=r}^{n} \binom{n}{k} \binom{k}{r} x^k.$$

Use this result to simplify the quantities

$$S_4 = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k}, \quad S_5 = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k q^{n-k},$$

where p and q are positive numbers such that $p + q = 1$.

SOLUTION We can easily check (see Part (iv) in Exercise 7 of Section 2.4) that

$$\binom{n}{k}\binom{k}{r} = \binom{n}{r}\binom{n-r}{k-r},$$

so that the sum S we wish to calculate takes the form

$$S = \sum_{k=r}^{n} \binom{n}{r}\binom{n-r}{k-r} x^k = \binom{n}{r} \sum_{k=r}^{n} \binom{n-r}{k-r} x^k.$$

Now, by making the change of variable $k - r = j$, we get

$$\binom{n}{r} \sum_{j=0}^{n-r} \binom{n-r}{j} x^{r+j} = x^r \binom{n}{r} \sum_{j=0}^{n-r} \binom{n-r}{j} x^j.$$

Upon employing Proposition 2.9, we obtain the result that

$$S = \binom{n}{r} x^r (1+x)^{n-r}. \tag{2.14}$$

It is worth noting that the same result may also be obtained if we start with the identity

$$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$$

which is a special case of Proposition 2.9 for $a = 1$ and $b = x$, and then differentiate both sides of it $1, 2, \ldots, r$ times, so that we obtain the following expressions:

$$n(1+x)^{n-1} = \sum_{k=0}^{n} \binom{n}{k} (x^k)' = \sum_{k=1}^{n} \binom{n}{k} kx^{k-1},$$

$$n(n-1)(1+x)^{n-2} = \sum_{k=1}^{n} \binom{n}{k} k(x^{k-1})' = \sum_{k=2}^{n} \binom{n}{k} k(k-1)x^{k-2},$$

$$\cdots$$

$$n(n-1)\cdots(n-r+1)(1+x)^{n-r} = \sum_{k=r-1}^{n} \binom{n}{k} k(k-1)\cdots(k-r+2)(x^{k-r+1})'$$

$$= \sum_{k=r}^{n} \binom{n}{k} k(k-1)\cdots(k-r+2)(k-r+1)x^{k-r}.$$

As a consequence, we obtain

$$(n)_r(1+x)^{n-r} = \sum_{k=r}^{n} (k)_r \binom{n}{k} x^{k-r} \tag{2.15}$$

and by multiplying both sides of Eq. (2.15) by $x^r/r!$, we arrive again at formula (2.14).

Next, in order to calculate the sums S_4 and S_5, it suffices to put $x = p/q$ into (2.14) and consider the special cases $r = 1$ and $r = 2$. In this way we obtain, first for $r = 1$,

$$\sum_{k=1}^{n} \binom{n}{k} k \cdot \frac{p^k}{q^k} = n \cdot \frac{p}{q} \cdot \left(\frac{p+q}{q}\right)^{n-1}$$

and for $r = 2$,

$$\sum_{k=2}^{n} \binom{n}{k} \frac{k(k-1)}{2} \cdot \frac{p^k}{q^k} = \frac{n(n-1)}{2} \cdot \left(\frac{p}{q}\right)^2 \cdot \left(\frac{p+q}{q}\right)^{n-2}.$$

Using the fact that $p + q = 1$, it is easy to see from the last two expressions that we get the desired simplifications for the sums S_4 and S_5, namely,

$$S_4 = \sum_{k=0}^{n} k \binom{n}{k} p^k q^{n-k} = \sum_{k=1}^{n} k \binom{n}{k} p^k q^{n-k} = np$$

and

$$S_5 = \sum_{k=0}^{n} k(k-1) \binom{n}{k} p^k q^{n-k} = \sum_{k=2}^{n} k(k-1) \binom{n}{k} p^k q^{n-k} = n(n-1)p^2.$$

We will revisit the sums S_4 and S_5 in Chapter 5 when they will be particularly useful in relation to one of the most important probability distributions.

We now close this section by presenting a result that is known as the **Cauchy formula**.

Proposition 2.11 *Let m, n, and r be nonnegative integers such that $r \leq m + n$. Then, the following formula holds:*

$$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}. \tag{2.16}$$

Proof: The result can be proved by considering all the distinct ways in which r elements can be chosen from a set with $m + n$ elements. More explicitly, we define the following three sets

$$Y = \{y_1, y_2 \dots, y_m\}, \quad Z = \{z_1, z_2 \dots, z_n\}, \quad X = Y \cup Z,$$

where $y_1, y_2 \dots, y_m, z_1, z_2 \dots, z_n$ are $m + n$ distinct elements.

The quantity

$$\binom{m+n}{r}$$

indicates the number of ways r elements can be chosen among the $m + n$ elements of the set X. This can be done by choosing k elements from the set Y and $r - k$ elements from the set Z, for $k = 0, 1, 2, \dots$ The following table illustrates this point.

Number of elements chosen from set		Number of selections of these elements from set		Number of total selections from Y and Z
Y	Z	Y	Z	
0	r	$\binom{m}{0}$	$\binom{n}{r}$	$\binom{m}{0} \cdot \binom{n}{r}$
1	$r-1$	$\binom{m}{1}$	$\binom{n}{r-1}$	$\binom{m}{1} \cdot \binom{n}{r-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	$r-k$	$\binom{m}{k}$	$\binom{n}{r-k}$	$\binom{m}{k} \cdot \binom{n}{r-k}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
r	0	$\binom{m}{r}$	$\binom{n}{0}$	$\binom{m}{r} \cdot \binom{n}{0}$
Total number of selections from the set $X = Y \cup Z$				$\sum_{k=0}^{r} \binom{m}{k} \cdot \binom{n}{r-k}$

As can be seen from the table, the number of combinations for which k elements are chosen from the set Y and $r-k$ elements from the set Z is given by

$$\binom{m}{k} \cdot \binom{n}{r-k},$$

and the required result then follows by summing these terms over k. □

EXERCISES

Group A

1. How many different subsets does a set of n elements (such as the set $\{1, 2, \ldots, n\}$) have?

2. In the binomial expansion of the following quantities, identify the coefficient of the constant term, i.e. the term which does not involve x:

   (i) $\left(2x^2 - \dfrac{4}{x^3}\right)^{15}$;

   (ii) $(4x^3 - 3x^{-2})^{15}$.

3. The formula by A.T. Vandermonde (1737–1796) expresses the number of permutations of $m+n$ elements in terms of the numbers of permutations of m elements and that of n elements, as follows:

$$(m+n)_r = \sum_{k=0}^{r} \binom{r}{k} (m)_k (n)_{r-k}.$$

   Show that this formula can be deduced immediately from the Cauchy formula (Proposition 2.11).

4. By choosing appropriate values for a, b, and t in the binomial expansions of Propositions 2.9 and 2.10, calculate the following sums (n here is a nonnegative integer):

   (i) $\displaystyle\sum_{k=0}^{n} \binom{n}{k} 3^{n-k}$;

   (ii) $\displaystyle\sum_{k=0}^{\infty} \binom{n+k-1}{k} \frac{1}{2^k}$;

   (iii) $\displaystyle\sum_{k=0}^{\infty} (-1)^k \binom{n+k-1}{k} \frac{1}{2^{n+k}}$;

   (iv) $\displaystyle\sum_{k=0}^{\infty} n^{-k} \binom{n+k-1}{k}, \quad n \geq 1.$

5. Assume that n is a positive integer. Then, use the Cauchy formula to calculate the following sums:

   (i) $\displaystyle\sum_{k=1}^{n}\binom{n}{k}\binom{n-1}{k-1}$;

   (ii) $\displaystyle\sum_{k=0}^{n}\binom{n}{k}^2$.

6. Using the identity in (2.15), calculate the following sums:

$$\sum_{k=0}^{n} k\binom{n}{k}, \quad \sum_{k=0}^{n} k(k-1)\binom{n}{k}$$

   and

$$\sum_{k=0}^{n} k(k-1)(k-2)\binom{n}{k}.$$

Group B

7. Let r and n be two nonnegative integers such that $r \le n$ and $x \in \mathbb{R}$. Using (2.15), calculate the following sums:

$$\sum_{k=0}^{n}(-1)^k k\binom{n}{k}, \quad \sum_{k=0}^{n}(-1)^k k(k-1)\binom{n}{k}$$

   and

$$\sum_{k=0}^{n}(-1)^k k(k-1)(k-2)\binom{n}{k}.$$

   Use these results and those of the previous exercise to calculate each of the following sums:

   (i) $\displaystyle\sum_{k=0}^{n}(3k+2)k\binom{n}{k}$; (ii) $\displaystyle\sum_{k=0}^{n}(-1)^k(3k+2)\binom{n}{k}$;

   (iii) $\displaystyle\sum_{k=0}^{n}k^2\binom{n}{k}$; (iv) $\displaystyle\sum_{k=0}^{n}(-1)^k k^2\binom{n}{k}$;

   (v) $\displaystyle\sum_{k=0}^{n}k(k-4)\binom{n}{k}$; (vi) $\displaystyle\sum_{k=0}^{n}(-1)^k k(k-4)\binom{n}{k}$.

8. Let r and n be two nonnegative integers and p and q be two positive real numbers such that $p+q=1$. Then, show that

$$\sum_{k=r}^{\infty}(k)_r\binom{n+k-1}{k}p^k q^{n+k} = n(n+1)(n+2)\cdots(n+k-1)p^r.$$

   (*Hint*: Use Proposition 2.9 and then proceed as in Example 2.18.)

2.6 BASIC CONCEPTS AND FORMULAS

Equiprobable events A, B	$P(A) = P(B)$
Classical definition of probability	$P(A) = \frac{\lvert A\rvert}{\lvert\Omega\rvert} = \frac{\text{number of elements of } A}{\text{number of elements of } \Omega}$ $= \frac{\text{number of favorable outcomes for the event } A}{\text{number of possible outcomes for the experiment}}$
Multiplicative law	Suppose that an element (object) a_1 can be chosen with n_1 different ways and for each choice of a_1, the element a_2 can be chosen in n_2 different ways, and so on, so that for any selection of the elements $a_1, a_2, \ldots, a_{k-1}$, the element a_k can be chosen in n_k different ways. Then, the k-tuple $(a_1, a_2, \ldots, a_k)$ can be chosen successively, and in that order, in $n_1 n_2 \cdots n_k$ different ways. In the above definition, the word "element" can be replaced by the words "object," "action," etc.
Cardinality of a Cartesian product	$\lvert A_1 \times A_2 \times \cdots \times A_k\rvert = \lvert A_1\rvert \cdot \lvert A_2\rvert \cdots \lvert A_k\rvert$.
A k-element permutation of n elements	An ordered k-tuple consisting of k distinct elements chosen among the n elements
Number of k-element permutations of n elements	$(n)_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$, for $1 \le k \le n$
Number of n-element permutations of n elements	$n!$
A k-element permutation of n elements with repetitions	An ordered k-tuple consisting of k elements chosen among the n elements
Number of k-element permutations of n elements with repetitions	n^k
Combination of k elements among n	An unordered k-tuple consisting of k distinct elements chosen among the n elements
Number of combinations of k elements among n	$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad 1 \le k \le n$
Properties of combinations of k elements among n	• $\binom{n}{k} = \binom{n}{n-k}$ • $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$ (Pascal's triangle)
Combination of k elements among n with repetitions	An unordered k-tuple consisting of k elements chosen among the n elements (repetitions are allowed)

Number of combinations of k elements among n with repetitions	$\begin{bmatrix} n \\ k \end{bmatrix} = \binom{n+k-1}{k} = \binom{n+k-1}{n-1}$, $k \geq 1, n \geq 1$
Binomial theorem	$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k} = \sum_{k=0}^{n} \binom{n}{k} a^{n-k} b^k$, where $a, b \in \mathbb{R}$ and $n \geq 0$ is an integer
Binomial expansion for negative integer powers	$(1-t)^{-n} = \frac{1}{(1-t)^n} = \sum_{k=0}^{\infty} \binom{n+k-1}{k} t^k$, where $t \in \mathbb{R}$ such that $\|t\| < 1$ and $n \geq 1$ is an integer
Cauchy formula	$\binom{m+n}{r} = \sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}$, where m, n, and r are three nonnegative integers such that $r \leq m+n$

2.7 COMPUTATIONAL EXERCISES

1. It is apparent that, as n grows large, the factorial $n!$ can be astronomical. For instance, 20! is about 2.43×10^{18}, while 100! is around 9.3×10^{157}. Mathematica can be used to find the binomial coefficients $\binom{n}{k}$. For the number of combinations of k elements out of n (i.e. for the binomial coefficients), the command *Binomial* can be used directly. For instance,

```
In[1]:=  25!
Out[1]= 15511210043330985984000000
In[2]:= 50!/40!
Out[2]=  37276043023296000
In[3]:=  Binomial[30,10]
Out[3]= 30045015
```

2. The following sequence of commands

```
Do[ a=Binomial[Binomial[n,2],2]; b=Binomial[(n+1),4];
       Print[n,"  ",a,"  ",b,"  ",a-3*b],
       {n,3,10}]
```

can be used to verify numerically the validity of the identity

$$\binom{\binom{n}{2}}{2} = 3\binom{n+1}{4}$$

(see Part (i) Exercise 8 of Section 2.4). Write a set of commands to verify the second part of the same exercise.

3. Counting of all possible permutations of a set of elements (with or without repetitions) can be done with the following sequence of commands:

```
In[1]:=  perm=Permutations[{a,b,c}];
Print["The permutations of {a,b,c} are : ",perm];
Print["The number of permutations of {a,b,c} equals ",
       Length[perm]]

Out[2]=  The permutations of {a,b,c} are :
        {{a,b,c},{a,c,b},{b,a,c},{b,c,a},{c,a,b},{c,b,a}}
The number of permutations of {a,b,c} equals 6

In[1]:=  perm=Permutations[{a,b,b}];
Print["The permutations of {a,b,b} are : ",perm];
Print["The number of permutations of {a,b,b} equals ",
       Length[perm]];

Out[2]=  The permutations of {a,b,b} are :
  {{a,b,b},{b,a,b},{b,b,a}}
The number of permutations of {a,b,b} equals 3
```

Consider now the following problem: three men and two women wait in a queue at a bank. If the order that these five persons arrived in the bank can be considered to be completely random, what is the probability that

(i) men and women alternate in the queue?

(ii) the three men are standing next to each other?

(iii) the two women are in adjacent positions in the queue?

Solve each of the three parts above by a complete enumeration of the elements of the sample space for the experiment in question and another enumeration of the favorable outcomes in each case.

4. The next program finds all possible outcomes in the experiment of throwing two dice. Then, it calculates the number of outcomes in which the sum of the dice equals six as well as the probability of that event.

```
In[1]:= m=6;n=6;
twodice=Flatten[Table[{i,j}, {i,1,m}, {j,1,n}],1]
total=Length[twodice];
s=0;
Do[y=twodice[[i]];
If[y[[1]]+y[[2]]==6,s=s+1],
{i,1,total}]
Print["Probability=",s/total]

Out[2]= {{1,1},{1,2},{1,3},{1,4},{1,5},{1,6},{2,1},{2,2},
        {2,3},{2,4},{2,5},{2,6},{3,1},{3,2},{3,3},{3,4},
        {3,5},{3,6},{4,1},{4,2},{4,3},{4,4},{4,5},{4,6},
        {5,1},{5,2},{5,3},{5,4},{5,5},{5,6},{6,1},{6,2},
        {6,3},{6,4},{6,5},{6,6}}
Probability=5/36
```

Make suitable modifications to the above program so that the following questions, relating to the same experiment, can be answered using Mathematica: what is the probability that when throwing two dice,

(i) the sum of the two outcomes is less than or equal to six?

(ii) the difference between the two outcomes is exactly three?

(iii) the sum of the squares of the two outcomes is less than or equal to 36?

(iv) the largest of the two outcomes equals five?

(v) the smallest of the two outcomes equals three?

(vi) at least one of the outcomes equals six?

5. Working as in the previous exercise, answer the following questions:

   (i) An urn contains 10 balls numbered 0–9. We select a ball at random, note the number on it, and return the ball into the urn. Then, we take out a second ball and also make a note of the number on that ball. What is the probability that
      (a) the two numbers are the same?
      (b) the sum of the two numbers equals 10?
      (c) the difference between the two numbers is at least 6?
      (d) the sum of the cubes of the two numbers exceeds 20?
      (e) at least one of the two numbers is greater than or equal to 5?

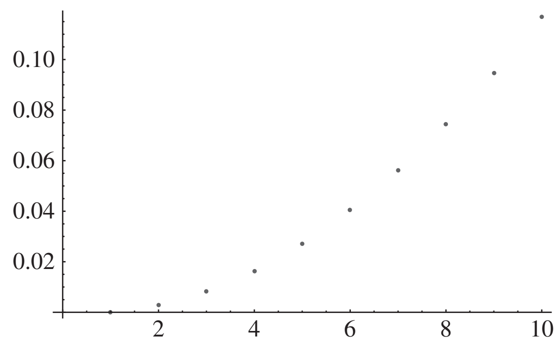
   (ii) In the experiment of tossing a coin five times, find the probability that
      (a) at least three heads appear;
      (b) heads turn up at least once;
      (c) the first three outcomes of the coin are all heads;
      (d) the first three outcomes of the coin are the same;
      (e) heads appear in at least two successive tosses of the coin.

6. Recall the birthday problem (see Example 2.12). Let $1 - p_k$ denote the probability that in a group of k students, at least two share the same birthday. This probability is found below using Mathematica.

```
In[1]:= k=5;
        birth=   1-Product[(366-r)/365,{r,1,k}]
        birth1=  N[birth]

Out[2]=  481626601/17748900625
Out[3]=  0.0271356
```

(Notice that, if we do not use the operator $N[\ ]$, which gives the result in decimal form, the answer is given as a fraction.) The result above was found for $k = 4$. If we want to plot $1 - p_k$, as a function of k, we have the following series of commands:

```
In[4]:= n=10;
        birthtable=  Table[{k,1-Product[(365-j+1)/365,
                               {j,1,k}]},{k,1,n}]//N
        ListPlot[birthtable]
Out[5]=  {{1.,0.},{2.,0.00273973},{3.,0.00820417},
          {4.,0.0163559},{5.,0.0271356},{6.,0.0404625},
          {7.,0.0562357},{8.,0.0743353},{9.,0.0946238},
          {10.,0.116948}}
```



(i) Check empirically that, as k grows large, $1 - p_k$ approaches unity.

(ii) Find the least value of k such that the corresponding probability $1 - p_k$ is at least

(a) 50%;

(b) 90%;

(c) 95%;

(d) 99%.

7. Generalizing the Chevalier de Méré problem (see Exercise 16 in Section 2.3), suppose p_k denotes the probability that we get at least one six when we throw a die k times, and q_k denotes the probability that we get at least one double six when we throw two dice $6k$ times.

(i) Verify that

$$p_k = 1 - \left(\frac{5}{6}\right)^k, \quad q_k = 1 - \left(\frac{35}{36}\right)^{6k}, \quad k = 1, 2, \ldots .$$

(ii) Use a series of commands similar to those given in the previous exercise

(a) to draw plots of p_k and q_k for different values of k,

(b) to find the smallest value of k such that p_k is at least I. 50% II. 90% III. 95%.

Find also for which values of k the probability q_k exceeds each of these three values above.

(iii) Find the value of k that minimizes the difference between p_k and q_k.

8. The creation and enumeration of all possible combinations from a set can be done in Mathematica with the command `KSubsets`. More specifically, the command `KSubsets[A, k]` gives all subsets of a set A containing exactly k elements,

ordered lexicographically (for old versions of Mathematica, notice that we have to load the package *"Combinatorica"* first, as shown below. As of Version 10, much of the functionality covered by Combinatorica has been implemented in the Wolfram System.)

The following commands illustrate the use of the command `KSubsets`:

```
In[1]:= <<Combinatorica`
        comb=     KSubsets[{a,b,c}, 2]
        Print["number of combinations = ",Length[comb]]
Out[2]=   {{a,b},{a,c},{b,c}}

In[4]:= comb=KSubsets[{a1,a2,a3,k1,k2}, 3]
        Print["number of combinations = ",Length[comb]]
Out[4]=    {{a1,a2,a3},{a1,a2,k1},{a1,a2,k2},{a1,a3,k1},
           {a1,a3,k2},{a1,k1,k2},{a2,a3,k1},{a2,a3,k2},
           {a2,k1,k2},{a3,k1,k2}}

In[6]:= comb=KSubsets[{a,a,a,k,k,k}, 3]
        Print["number of combinations = ",Length[comb]]
Out[6]=    {{a,a,a},{a,a,k},{a,a,k},{a,a,k},{a,a,k},
           {a,a,k},{a,a,k},{a,k,k},{a,k,k},{a,k,k},
           {a,a,k},{a,a,k},{a,a,k},{a,k,k},{a,k,k},
           {a,k,k},{a,k,k},{a,k,k},{a,k,k},{k,k,k}}
```

Using a similar set of commands and the classical definition of probability (i.e. recording both the favorable outcomes for an event and all possible outcomes for the experiment), solve the following problem. The 20 students in a class are listed alphabetically, according to their surname, and a number from 1 to 20 is given to each of them. Then, these numbers are placed in a box and for selecting a class committee, five numbers are drawn at random. Find the probability that

(i) one of the numbers selected is 4;

(ii) $2, 4, 6$ are all among the selected numbers;

(iii) none of the numbers $1, 3, 5, 7$ is selected;

(iv) all five integers selected are even;

(v) the smallest number selected is 7;

(vi) the smallest number selected is greater than 7;

(vii) the difference between the smallest and the largest numbers selected is at least 5.

9. A drawer contains six pairs of gloves, each of a different color. A person selects randomly 4 gloves among the 12. By recording all possible and all favorable events in each case (as in the previous exercise), find the probability that among the four gloves selected

   (i) there are two pairs;

   (ii) there is exactly one pair;

   (iii) there is no pair of gloves.

10. Among 15 couples who attend a party, we choose randomly 4 persons. What is the probability that among them

    (i) there are only men?

    (ii) there are only women?

    (iii) there is at least one couple?

    (Remember that if you want to enumerate all possible combinations from a set and work with old versions of Mathematica, you have to load the package *"Combinatorica"* first; see the note on Exercise 8.)

11. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set and k be a positive integer such that $k \le n$. We denote

$$p_n = \frac{(n)_k}{n^k}$$

for the ratio of the number of k-element permutations among the n elements without repetitions to the number of k-element permutations with repetitions. We also define correspondingly

$$q_n = \frac{\binom{n}{k}}{\left[\begin{matrix} n \\ k \end{matrix}\right]}$$

for the ratio of combinations instead of permutations. Examine numerically how p_n, q_n vary as n increases and draw a graph of each of these ratios as a function of n.

12. For large values of a positive integer n, an approximation which is often used for its factorial, $n!$, is the following, known as **Stirling's formula**:

$$n! \sim \sqrt{2\pi n}\left(\frac{n}{e}\right)^n.$$

Here, the symbol $\sim$ means that the ratio of the two sides tends to 1 as $n \to \infty$, so that the formula can also be written in the form

$$\lim_{n\to\infty} \frac{n!\mathrm{e}^n}{n^{n+1/2}} = \sqrt{2\pi}.$$

Calculate the values of the quantity

$$\frac{n!\mathrm{e}^n}{n^{n+1/2}}$$

for $n = 20, 30, 40, 50, 60, 80, 100, 120, 150$ and plot these points against n on a graph. Check that even for modest values of n, this gets very close to $\sqrt{2\pi}$.

2.8 SELF-ASSESSMENT EXERCISES

2.8.1 True–False Questions

1. When we toss a coin three times, the sample space with all different outcomes of the three tosses has eight elements.

2. When we throw a die, the probability that the outcome is greater than or equal to 3 is $1/2$.

3. A box contains 50 lottery tickets, numbered $1, 2, \dots, 50$. We select one ticket at random. The probability that the number on it is a multiple of 3 is $1/3$.

4. Nick rolls a die twice. The probability that the second outcome is higher than the first is $1/2$.

5. The number of different 5-member committees, consisting of 3 men and 2 women, that can be formed among 18 men and 12 women is $\binom{18}{3}\binom{12}{2}$.

6. We select three digits from one to nine with replacement. The number of all possible outcomes, in order to form a three-digit number, is $(9)_3$.

7. We select three digits from one to nine without replacement. The number of all possible outcomes, in order to form a three-digit number, is 9^3.

8. Five students are ranked in terms of their academic performance. The total number of possible rankings is 5!

9. For any nonnegative integer n, we have
$$\binom{n+1}{n} = n+1.$$

10. For nonnegative integers n and r with $r \le n$, it is always true that
$$\binom{n}{r} = \binom{n}{n-r}.$$

11. For nonnegative integers n and r with $r \le n$, it is always true that
$$\begin{bmatrix} n \\ r \end{bmatrix} \ge \binom{n}{r}.$$

12. Let n and r be two integers with $r > n \ge 1$. Then both $\begin{bmatrix} n \\ r \end{bmatrix}$ and $\binom{n}{r}$ are equal to zero.

13. A bookshelf contains 30 books. Among them, 12 are fiction and the remaining are nonfiction books. If we select five books at random, the probability that there will be exactly four fiction books among them is

$$\frac{4 \cdot 18}{\binom{20}{5}}.$$

14. Three friends throw one die each. The probability that the sum of the three outcomes is 17 equals $1/72$.

15. The coefficient of x^2 in the binomial expansion of $(x+3)^6$ is $\binom{6}{2}$.

16. The coefficient of x^4 in the binomial expansion of $(1-x)^{-6}$ is $\binom{9}{4}$.

17. For nonnegative integers n and k with $k \leq n$, the sum

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

equals 2^n.

18. For nonnegative integers m, n, and r with $r \leq n+m$, the sum

$$\sum_{k=0}^{r} \binom{m}{k} \binom{n}{r-k}$$

equals 2^{m+n}.

2.8.2 Multiple Choice Questions

1. When we throw a die three times, the probability that no six appears is

(a) $\frac{1^3}{6^3}$ (b) $\frac{1^3}{5^3}$ (c) $\frac{5^3}{6^3}$ (d) $1-\frac{1^3}{6^3}$ (e) $1-\frac{5^3}{6^3}$

2. In two throws of a die, the probability that the sum of the two outcomes equals three is

(a) $\frac{3}{6}$ (b) $\frac{1}{6}$ (c) $\frac{1}{36}$ (d) $\frac{1}{18}$ (e) $\frac{4}{36}$

3. In three throws of a die, the probability that no two outcomes are the same equals

(a) $\frac{2}{9}$ (b) $\frac{20}{6^3}$ (c) $\frac{1}{3}$ (d) $\frac{5}{9}$ (e) $\frac{4}{9}$

4. A high-school offers four language courses: French, Spanish, German, and Italian. Each student has to select exactly two of these courses. Assuming that all choices are equally likely, the probability that a student chooses French and Italian is

(a) $\frac{1}{3}$ (b) $\frac{1}{4}$ (c) $\frac{1}{6}$ (d) $\frac{2}{3}$ (e) $\frac{3}{8}$

5. If we select 3 cards without replacement from an usual pack of 52 cards, the number of possible outcomes is

(a) $\binom{49}{3}$ (b) $\frac{49!}{3!}$ (c) 52^3 (d) $\frac{52^3}{3!}$ (e) $\binom{52}{3}$

6. If we select 2 cards without replacement from an usual pack of 52 cards, the probability that neither of them is an ace is

(a) $\frac{\binom{48}{2}}{\binom{52}{2}}$ (b) $\frac{48!2!}{\binom{52}{2}}$ (c) $\frac{52!}{48!4!}$ (d) $\frac{48!4!}{52!}$ (e) $\frac{48^2}{52^2}$

7. For nonnegative integers n and k with $k \leq n$, the ratio

$$\frac{(n)_k}{\binom{n}{k}}$$

equals

(a) $k!$ (b) $\frac{1}{k!}$ (c) $\frac{n^k}{k!}$ (d) $\frac{k!}{n^k}$ (e) $\frac{k}{n^k}$

8. Suppose that there are five red balls and seven black balls in a box. If we select three balls at random and without replacement, the probability that none of them is black equals

(a) $\frac{\binom{7}{0}}{\binom{12}{3}}$ (b) $\frac{\binom{7}{0}}{\binom{5}{3}}$ (c) $\frac{\binom{5}{3}+\binom{7}{0}}{\binom{12}{3}}$ (d) $\frac{\binom{5}{3}}{\binom{12}{3}}$ (e) $\frac{\binom{5}{3}}{\binom{5}{3}+\binom{7}{0}}$

9. Ariadne and Athena are schoolmates. The probability that they will celebrate their next birthday on the same day of the week is

(a) $\frac{6}{7}$ (b) $\frac{1}{7}$ (c) $\frac{2}{7}$ (d) $\frac{\binom{6}{2}}{\binom{7}{2}}$ (e) $\frac{6 \cdot 7}{\binom{7}{2}}$

10. A mathematics teacher has four ties and each day he selects at random one of them to wear to work. In a period of three days, the probability that he does not wear the same tie more than once equals

(a) $\frac{3}{4}$ (b) $\frac{1}{2}$ (c) $\frac{1}{4}$ (d) $\frac{3}{8}$ (e) $\frac{3}{16}$

11. Marie, as she opens her email, finds out that she has five new messages, of which three are spam messages. The probability that the two nonspam messages were received one after the other, without any spam messages between them, is

(a) $\frac{1}{5}$ (b) $\frac{1}{4}$ (c) $\frac{2}{5}$ (d) $\frac{3}{10}$ (e) $\frac{1}{30}$

12. During a hand of poker, a player receives 5 cards out of 52 cards from an ordinary deck. The probability that, at a given hand, he receives at least one ace is

(a) $\frac{1}{13}$ (b) $\frac{5}{52}$ (c) $\frac{\binom{4}{1}\binom{48}{4}}{\binom{52}{5}}$ (d) $\frac{\binom{5}{1}\binom{47}{4}}{\binom{52}{5}}$ (e) $1-\frac{\binom{48}{5}}{\binom{52}{5}}$

13. There are six red balls and three black balls in a box. All nine balls are identical except their color. Nick claims that he can pull out six balls (without replacement) and that all six will be red. Assuming that he has no magical skills and he chooses at random, let x be the probability that he will succeed. He then pulls out a ball and it turns out to be red. At that stage, the probability that he will succeed has been increased, compared with the original probability x, by

(a) 10% (b) 15% (c) 16.67% (d) 25% (e) 50%

14. In the parking area of a supermarket, there are 40 parking places, numbered 1–40. At a particular instant, 15 places are occupied. Assuming that all places are equally likely to be occupied, the probability that exactly one of the parking places numbered 1 and 2 is occupied equals

(a) $\frac{2\binom{40}{13}}{\binom{40}{15}}$ (b) $\frac{\binom{38}{13}}{\binom{38}{15}}$ (c) $\frac{\binom{38}{14}}{\binom{40}{15}}$ (d) $\frac{\binom{2}{1}\binom{38}{14}}{\binom{40}{15}}$ (e) $\frac{\binom{38}{13}+\binom{38}{14}}{\binom{40}{15}}$

15. The value of the sum

$$\sum_{k=0}^{5}\binom{10}{k}\binom{6}{5-k}$$

is

(a) $\binom{16}{5}$ (b) $\binom{10}{5}+\binom{6}{5}$ (c) $\binom{10}{5}\cdot\binom{6}{5}$ (d) 2^5 (e) 2^{16}

16. The value of the sum

$$\sum_{k=1}^{n} k^2 \binom{n}{k}$$

is

(a) $n(n-1)2^{n-2}$ (b) $n(n+1)2^{n-2}$ (c) $n^2 2^{n-2}$
(d) $n^2 2^{n-2} + 1$ (e) $n^2 2^{n-1}$

17. The number of different domino tiles that can be formed if the number on each side of a tile is chosen from the set $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ is

(a) 45 (b) 55 (c) 63 (d) 36 (e) 72

18. When we throw simultaneously three identical dice, the number of *distinguishable outcomes* is

(a) $\binom{8}{5}$ (b) $\binom{6}{5}$ (c) $\binom{6}{3}$ (d) 6^3 (e) 6^2

2.9 REVIEW PROBLEMS

1. A sprinter is about to run a 100 m race on lane 3 of a track that has 8 lanes. If two other runners from the same country take part in this race, what is the probability that neither of them will run next to him?

2. We want to place n mathematics books and k physics books on a book-shelf.
   (i) How many ways are there for placing the books on the shelf?
   (ii) If $k \leq n + 1$, find the probability that no two physics books are put next to one another.

3. We ask 15 people about the day of the week they were born. What is the probability that no one was born on a Sunday?

4. A company of n students decides to have lunch at the Student Union on a particular day. Each of the students will go alone as they attend different classes in the morning. However, there are k restaurants at the Student Union and they have not specified at which one they are supposed to meet.
   (i) How many different selections are possible for the restaurants that the students will go for lunch?
   (ii) Assuming that each student chooses a restaurant at random with the same probability, what is the probability that they all choose the same restaurant?

5. A safe box requires a four-digit number to be opened.
   (i) Find the number n of different numbers that are possible codes for the safe box.
   (ii) If a burglar chooses at random four-digit numbers trying to open the box, find the probability that he succeeds
      (a) at the kth attempt $(1 \leq k < n)$;
      (b) during one of his first r attempts $(1 \leq r < n)$;
      (c) at the nth (and final) attempt.

6. A computer machine selects at random an even integer which is greater than or equal to 10 000 and less than 70 000. What is the probability that the number selected has no two digits which are the same?

7. Each student, in a class of 50 students, has to choose two optional courses among the four optional courses offered by their Department, so that there are $\binom{4}{2}$ alternative choices (course combinations) for any of them. What is the probability that there is exactly one course combination not chosen by any of the students?

8. In a certain town, the percentage of people with blood type A is approximately the same to that with blood type O. Moreover, the percentage of inhabitants with blood type B is about $1/5$ compared to that of type A and three times as much compared to the percentage of persons with type AB. Estimate the probability that the next child to be born in this town
   (i) will have a blood type AB;
   (ii) will have a blood type either A or B;
   (iii) will not have blood type O.

9. The coefficients of the quadratic polynomial $(a + 1)x^2 + bx + c$ are decided from the outcomes of a fair coin, which is tossed three times. Specifically, if at the first toss the coin lands heads we put $a = 1$, otherwise we put $a = 0$. Similarly, $b = 1$ if at the second toss the coin lands heads ($b = 0$ otherwise) and $c = 1$ if at the third toss the coin lands heads ($c = 0$ otherwise). Find the probability that the polynomial has no real roots.

10. Consider the following experiment: we ask a computer to select an integer at random from the set $\{0, 1, 2, \ldots, 99\,999\}$. What is the probability that the number selected
    (i) does not contain the digit 7?
    (ii) contains both digits 5 and 7?
    (iii) contains, at least once, each of the digits $1, 3, 5, 7$?

11. Nick has six mathematics books, five physics books, three chemistry books, seven French books, and two dictionaries. He buys a new bookshelf and his mother places the books at random on the shelf. What is the probability that all books of the same subject are placed together?

12. Suppose that n students, say $S_1, S_2, \ldots, S_n$, apply for a postgraduate program, and they all receive an invitation to attend an interview. They all arrive in time and they are called upon one after the other.

    (i) Find the number of different ways for the order they will be called for the interview.

    (ii) What is the probability that students S_1 and S_2 will be called first and second, respectively, for the interview?

13. A bowl contains n lottery tickets numbered $1, 2, \ldots, n$. We select a ticket at random, record the number on it, and put it back in the bowl. The same procedure is followed a total of k times.

    Find the probability that

    (i) the number 1 is selected at least once;

    (ii) both numbers 1 and 2 are selected at least once.

    (*Hint*: Define the events

    $$A_i \ : \text{ number } i \text{ does not appear in any of the } k \text{ selections}$$

    for $i = 1, 2$. Then, the required probabilities are

    $$P(A_1') \quad \text{and} \quad P(A_1' A_2') = 1 - P(A_1 \cup A_2).)$$

14. Let $X = \{x_1, x_2, \ldots, x_n\}$ be a finite set with $n \geq 2$ (distinct) elements. Suppose that we want to use

    - r_1 times the element x_1,
    - r_2 times the element x_2,

      $\ldots$
    - r_n times the element x_n

    in order to form a collection $(a_1, a_2, \ldots, a_r)$ of r units, where $r = r_1 + r_2 + \cdots + r_n$. Such an arrangement is called an *r-permutation of n* (*distinct*) *elements*. Show that the number of all possible permutations of this type is given by the formula

    $$\binom{r}{r_1, r_2, \ldots, r_n} = \frac{r!}{r_1! r_2! \cdots r_n!}.$$

    Application: A part of the human DNA chain is represented as a series with elements A, C, G, T (the letters stand for the nucleobases adenine, cytosine, guanine, and thymine, respectively). How many different compositions (sequences) are possible for a segment of length r, such that r_1 elements are of type A, r_2 elements are of type C, r_3 elements are of type G, and r_4 elements are of type T $(r = r_1 + r_2 + r_3 + r_4)$? Under the assumption that all such compositions have the same probability of appearing, what is the probability that a randomly selected sequence has the elements corresponding to each of the four bases being adjacent, i.e. we have compositions as in the following examples

$$\underbrace{AA\ldots A}_{r_1}\ \underbrace{CC\ldots C}_{r_2}\ \underbrace{TT\ldots T}_{r_4}\ \underbrace{GG\ldots G}_{r_3}$$

or

$$\underbrace{GG\ldots G}_{r_3}\ \underbrace{AA\ldots A}_{r_1}\ \underbrace{CC\ldots C}_{r_2}\ \underbrace{TT\ldots T}_{r_4}$$

and so on?

15. Let $X = \{x_1, x_2, \ldots, x_n\}$ and k be a positive integer with $k \le n$. Denote by

$$p_n = \frac{(n)_k}{n^k}$$

the ratio of the number of k-element permutations among the n elements without repetitions to the number of k-element permutations with repetitions. We also define

$$q_n = \frac{\binom{n}{k}}{\left[\begin{matrix} n \\ k \end{matrix}\right]},$$

which is the corresponding ratio for combinations instead of permutations. What is the limit of the sequences p_n and q_n as $n \to \infty$? Check if your results agree with the numerical evidence found in Exercise 11 of Section 2.7.

16. Peter tosses a die k times for some $2 \le k \le 6$. Find the probability that the results of this die tosses

(i) are all the same;

(ii) contain at least two outcomes which are the same.

17. In a city with $n + 1$ inhabitants, a person tells a rumor to a second person, who then repeats it to a third person, and so on. At each step, the person who tells the rumor chooses the next recipient at random among the n people available.

Find the probability that the rumor will be told r times without

(i) returning to the first person;

(ii) being repeated to any person.

18. There are n couples competing in a dancing competition. Suppose there are n prizes available in total. What is the probability that exactly one person from each couple wins a prize?

19. In Jenny's shoe rack, there are seven pairs of shoes. She picks six shoes at random. What is the probability that among these

(i) there is no pair of shoes?

(ii) there is exactly one pair of shoes?

(iii) there are exactly two pairs?

(iv) there are exactly three pairs?

20. When we throw three dice, which one is more likely: that the sum of the three outcomes equals 10 or that it equals 9?

    (This problem is historically associated with the name of Galileo Galilei (1564–1642), who calculated the probabilities of both these events and showed that they were *not equal*, as it was thought to be so at that time.)

21. In the World Cup of football, there are 32 teams that participate. After a draw is made, these teams are divided into eight groups with four teams each. How many different ways there are to form the eight groups?

22. A train has three coaches and when it stops at the first station during its journey, n passengers embark it, for $n > 3$. Assuming that passengers embark on a particular coach of the train independently of one another, what is the probability that at least one passenger embarks on each coach?

    (*Hint*: Define the events

    A_i : no passenger embarks the ith coach

    for $i = 1, 2, 3$, and express the event of interest in terms of A_1, A_2, A_3.)

23. In a tennis tournament, 16 female players will compete for the title. In the first round of the tournament, the players will form randomly eight pairs and each pair will play a match. The eight winners of these matches will then form four new pairs. Each of these new pairs will play a match and the four winners will enter the semi-final round. There will be another draw to make two semi-final pairs; the two winners of the semi-final round will enter the final.

    Amelia and Irène are two players who enter the tournament. Find the probability that they will meet

    (i) in the first round;

    (ii) in the second round;

    (iii) in the final.

24. (The Mathematics of poetry[5]) The Indian writer Acharya Hemachandra (c. 1150 AD) studied the rhythms of Sanskrit poetry. Syllables in Sanskrit are either long (L) or short (S). Long syllables have twice the length of short syllables. The question he asked is *how many rhythm patterns with a given total length can be formed by a sequence of short and long syllables?*

[5]This is cited from the article *The Mathematics of Poetry* by Rachel Hall, which can be found at http://www.sju.edu/~rhall/Multi/rhythm2.pdf.

For instance, the number of patterns that have the length of four short syllables is 4: SSSS, SSL, SLS, LSS.

(i) For $1 \le k \le 7$, find the number r_k of patterns with total length k.

(ii) Prove that, for $k = 2, 3, \dots$, the following identity holds:[6]

$$r_k = r_{k-1} + r_{k-2}.$$

25. Considering the special case of Pascal's triangle ($k = 5$)

$$\binom{j}{5} - \binom{j-1}{5} = \binom{j-1}{4}, \quad j \ge 5,$$

and adding side-by-side the resulting equations for $j = 5, 6, \dots, 10$, check that the following identity holds true:

$$\binom{10}{5} = \binom{4}{4} + \binom{5}{4} + \cdots + \binom{9}{4} = \sum_{j=5}^{10} \binom{j-1}{4}.$$

Using a similar method, generalize the previous result to show that for any positive integers n and k with $k \le n$, we have

$$\binom{n+1}{k+1} = \sum_{j=k}^{n} \binom{j}{k}.$$

Apply this result to reestablish the well-known identities

$$1 + 2 + \cdots + n = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}$$

and

$$1^2 + 2^2 + \cdots + n^2 = \sum_{i=1}^{n} i^2 = \frac{n(n+1)(2n+1)}{6}.$$

26. The chess clubs of two schools compete against each other every year. Each school team has n players and on the day of the contest, each member of the first team is drawn to play against a member of the second team. In two consecutive contests between the two teams, each of the two teams has the same members and two separate draws, one in each occasion, take place. We want to calculate the probability of the event

B_n: at least one member of the first team plays against the same opponent of the second team in the two contests.

[6]In fact, the sequence $\{r_k : k = 1, 2, \dots\}$ is known in Mathematics as the *Fibonacci sequence*, after the Italian mathematician Fibonacci, although his work was published about 70 years after Hemachandra's.

Let us define the events

A_i: the ith member of the first team plays against the same opponent in the two contests

for $i = 1, 2, \ldots, n$.

(i) For the special case $n = 3$, verify that

$$P(A_1) = P(A_2) = P(A_3) = \frac{2!}{3!},$$

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = \frac{1!}{3!},$$

$$P(A_1A_2A_3) = \frac{1}{3!},$$

and conclude that

$$P(B_3) = 1 - \frac{1}{2!} + \frac{1}{3!}.$$

(ii) Using the Poincaré formula (see Proposition 1.10), show that, in the general case,

$$P(B_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \cdots + \frac{(-1)^{n-1}}{n!}.$$

(iii) Verify that, as $n \to \infty$, we have

$$\lim_{n\to\infty} P(B_n) = 1 - \frac{1}{\mathrm{e}} \cong 63\%.$$

27. Consider the identity

$$\frac{1}{1-t} = 1 + t + t^2 + \cdots = \sum_{j=0}^{\infty} t^j, \quad |t| < 1. \tag{2.17}$$

(i) Verify that the derivative, of order $n - 1$, of the function

$$f(t) = \frac{1}{1-t}$$

is given by

$$f^{(n-1)}(t) = \frac{(n-1)!}{(1-t)^n}.$$

(ii) By differentiating both sides of (2.17) $n - 1$ times establish that

$$\frac{1}{(1-t)^n} = \sum_{j}^{\infty} \frac{(j)_{n-1}}{(n-1)!} t^{j-n+1}, \quad |t| < 1.$$

(iii) Making the change of variable $j - n + 1 = k$ on the right-hand side of the last expression, obtain an alternative proof of Proposition 2.10.

2.10 APPLICATIONS

2.10.1 Estimation of Population Size: Capture–Recapture Method

We are often interested in estimating the number of inhabitants in a specific area (e.g. number of animals living in a forest, number of fish in a lake, etc.). If performing a census is not feasible, like in the two aforementioned examples, a popular technique to proceed at the estimation of the population size is the method of *capture–recapture*.

Let us assume that our population consists of N individuals. We randomly select r of them, place a mark on them, and set them free so that they mix with the rest of the population. For example, in the case of fish living in a lake, we fish r of them, mark them by a nonremovable mark, and put them back in the lake. After some time, say a couple of days, so that the marked fish have mixed well with other living ones in the lake, we catch randomly n fish from the lake (this is usually referred to as the recapture phase) and count the number of marked ones among the n fish.

To start with, let us first calculate the probability that in the recapture phase x out of the n individuals have a mark on them. Apparently, the sample space of our experiment contains $\binom{N}{n}$ elements (number of ways to select the n individuals from a population consisting of N individuals). The favorable events are formed by selecting x marked individuals (out of the r marked ones in the population) and $n-x$ unmarked (among the $N-r$ unmarked). There exist $\binom{r}{x}$ ways to select the marked individuals and $\binom{N-r}{n-x}$ ways to select the unmarked individuals. Consequently, the probability that the sample of n individuals contains exactly x marked ones equals

$$f(N,r,n;x) = \frac{\binom{r}{x}\binom{N-r}{n-x}}{\binom{N}{n}}$$

for $\max(0, n-N+r) \le x \le \min(n, r)$ (these restrictions are necessary so that the probability does not vanish).

As an illustration, in the next table, we present the values of $f(N,r,n;x)$ for $r = 20$, $n = 10$, and several values of N (the entries with vanishing probability indicate that the probability is less than 10^{-3}).

Let us next assume that N is not known and we address the problem of estimating N with the aid of available data. Suppose we have marked $r = 20$ individuals from N (N is now unknown) and, after some time, we select at random $n = 10$ individuals, among whom we observe that $x = 5$ have been marked. If the size of the population was $N = 25$, the probability of observing this specific outcome equals 0.005; if $N = 30$, this probability increases to 0.130, while if $N = 35$ it increases to 0.254, etc. It seems reasonable that, if we were to choose one of the tabulated values of N, based on the evidence provided by the experimental outcome (five marked individuals during the recapture phase), the natural choice is the one that maximizes the probability we are looking at, i.e. $N = 40$.

The above discussion leads to the following question, which highlights the estimation method for obtaining a reasonable value for the unknown N: for given values of n, r, and x, find the value of N that maximizes the probability $f(N) = f(N,r,n;x)$. In statistical terminology, the last probability is usually referred to as *likelihood* and the estimator of N obtained by maximizing it as *maximum likelihood estimator* (MLE).

	N					
x	25	30	35	40	45	50
0	0.000	0.000	0.000	0.000	0.001	0.003
1	0.000	0.000	0.001	0.004	0.013	0.028
2	0.000	0.000	0.007	0.028	0.064	0.108
3	0.000	0.005	0.040	0.104	0.172	0.226
4	0.000	0.034	0.132	0.222	0.269	0.280
5	0.005	0.130	0.254	0.284	0.258	0.215
6	0.059	0.271	0.288	0.222	0.154	0.103
7	0.237	0.310	0.192	0.104	0.056	0.031
8	0.385	0.189	0.072	0.028	0.012	0.005
9	0.257	0.056	0.014	0.004	0.001	0.000
10	0.057	0.006	0.001	0.000	0.000	0.000

Table 2.5 Values of the probabilities $f(N, r, n; x)$ for $r = 20$, $n = 10$.

In order to establish a general formula for the MLE of N, we observe that

$$\frac{f(N)}{f(N-1)} = \frac{(N-r)(N-n)}{N(N-r-n+x)} = \frac{1 - \dfrac{n}{N}}{1 - \dfrac{n-x}{N-r}},$$

which yields

$$f(N) > f(N-1) \iff \frac{n}{N} < \frac{n-x}{N-r} \iff x < \frac{nr}{N}.$$

Consequently, the probability $f(N)$ increases for $x < nr/N$ and decreases for $x > nr/N$. Taking into account that x takes on only positive integer values, we conclude that the value of N that maximizes the probability of $f(N)$ equals

$$\hat{N} = \left[\frac{rn}{x}\right]$$

($[a]$ denoting the integer part of a).

For $r = 20$, $n = 10$, and $x = 5$, we thus obtain

$$\widehat{N} = \left[\frac{20 \cdot 10}{5}\right] = 40,$$

which is the same value we chose based on the probabilities presented in Table 2.5.

KEY TERMS

binomial theorem (or binomial expansion)
classical definition of probability
combination
combination with repetition
combinatorial analysis
enumeration methods (or counting methods)
equally likely (or equally probable) events
factorial
favorable outcomes (or favorable results)
k-permutation
k-permutation with repetitions
multiplication principle
Pascal triangle
permutation
possible outcomes (or possible results)

CHAPTER 3

CONDITIONAL PROBABILITY – INDEPENDENT EVENTS

Thomas Bayes
(London c. 1701 – Tunbridge Wells, Kent, England 1761)

An English mathematician and Presbyterian minister, best known for the theorem that bears his name. He studied theology at Edinburgh University (1719–1722) and, following his father, became a Presbyterian minister at Tunbridge Wells, southeast of London. Although he received no formal mathematical training and published only one piece of mathematical work in his lifetime, in which he defended the foundations of Newton's calculus (published anonymously in 1736), this work showed him as an able mathematician and presumably on this account he was elected a Fellow of the Royal Society in 1742.

In his later years, he developed a keen interest in probability, apparently after reading a book written by de Moivre. After his death, his relatives asked another Presbyterian minister to examine the mathematical papers, written by him, which they had found. One of these papers contained a special case of what is today famously known as Bayes' theorem.

In his work, Bayes treated probabilities as quantities expressing a degree of belief. This interpretation of probability, which is also known as *subjective probability* and was popularized in the nineteenth century by Laplace, has become quite popular in statistics and decision theory. The term *Bayesian statistics* (or *Bayesian inference*) refers to the statistical field in which the true state of things can be expressed in terms of *degree of confidence* (Bayesian probabilities) we have on different statements.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

3.1 CONDITIONAL PROBABILITY

In the situations we have encountered so far, when we assigned probabilities to events, i.e. to possible outcomes of a future experiment, we assumed that we have no knowledge whatsoever about the outcome that will occur. There are many instances, however, when we have *some* knowledge about what will happen or what has already happened, if we are speculating in the middle of the realization of the experiment.

For example, consider the experiment of throwing two dice one after the other, and we are interested in the probability of the event

$$A\text{: both throws result in a six.}$$

It is helpful to define also the events

$$B\text{: the outcome of the first die is a six}$$

and

$$C\text{: the outcome of the second die is a six,}$$

so that $A = BC$. Since there are 36 equiprobable events in the sample space, and A contains just one of them, we have

$$P(A) = P(BC) = \frac{1}{36}.$$

Now, suppose that we throw the first die and the outcome is a six. What is *at this point* the probability of the event A? Common sense dictates that this can no longer be equal to $1/36$, as the event "two sixes in the two rolls of the die" is now more likely than it was at the beginning of the experiment. Moreover, most people would think (and it is true) that the probability of A after the first trial resulted in a six equals $1/6$, since for A to occur we only require the outcome of a single roll of a die (the second one) to be a six.

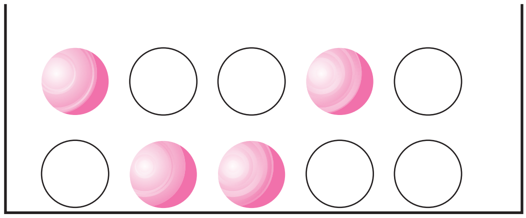
Things might become clearer when we are pushed to the extreme: suppose that we roll the die once and the outcome of the first trial *is not a six*. What is the probability we assign to A when we have that knowledge? It is obvious that A is then impossible to occur, and so this probability must be zero. However, we cannot write $P(A) = 0$, since we argued above that the probability of A (without having any partial knowledge about the outcome of the experiment) is $1/36$, and we cannot assign more than one probabilities to the same event.

In both the above cases, we tried to make probability statements about A when we have some partial information available on the outcome of the experiment that is being conducted. Thus, we came up with two answers that are both different from the original probability we assigned to A. To formalize this concept mathematically, we say that, if some knowledge about the outcome of an experiment is available, then the probability we assign to an event *given that knowledge* is a *conditional probability*.

Therefore, if we know in the example above that the first throw of the die resulted in a six, then the conditional probability of two sixes in two throws is $1/6$. On the other hand, if the first result is not a six, then the conditional probability of two sixes is zero.

In order to clarify further the above ideas, let us consider another example. Suppose we have a box that contains four blue balls and six white balls. We select randomly one out of the 10 balls and then, without returning this to the box,[1] we take another one.

[1] So, we are using here sampling without replacement.



Let us define the events:

A_i : At the ith selection of a ball, the color of the ball taken from the box is white

and

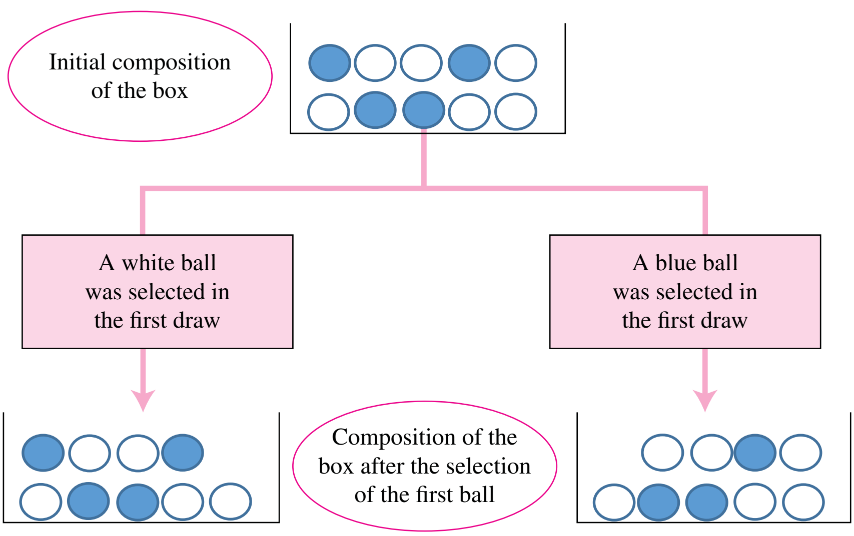
B_i : At the ith selection of a ball, the color of the ball taken from the box is blue

for $i = 1, 2$.

Before the first selection, there are 10 balls in the box, 6 of which are white. It is then evident that

$$P(A_1) = \frac{6}{10}, \quad P(B_1) = \frac{4}{10}.$$

Turning to find the probabilities $P(A_2), P(B_2)$, we face the following problem. After the first ball is drawn from the box, it is not returned to the box; so, although we know that there will be nine balls in the box when the second selection is made, we do not know exactly how many of these will be white. It is clear that the outcome of the second experiment depends on the outcome of the first; we can, however, consider two possibilities for the outcome of the first selection and, for each of these, we can find the probabilities of the events A_2 and B_2. Specifically, we argue as follows:



- If the first ball selected was white, the probability that the second ball drawn is white is $5/9$, and the probability that the second ball drawn is blue is $4/9$.
- If the first ball selected was blue, the probability that the second ball drawn is white is $6/9$, and the probability that the second ball drawn is blue is $3/9$.

Probabilities of this kind, i.e. probabilities of an event assuming that another event has occurred, are called conditional probabilities. We then write for the above probabilities

$$P(A_2|A_1) = \frac{5}{9}, \quad P(B_2|A_1) = \frac{4}{9},$$

and, similarly,

$$P(A_2|B_1) = \frac{6}{9}, \quad P(B_2|B_1) = \frac{3}{9}.$$

When we write, for instance, $P(A_2|A_1) = 5/9$, we say that the probability of A_2 *given* A_1 (has occurred) is $5/9$, and so on.

As we will see in one of the following sections, the four conditional probabilities given above can be used to calculate the (unconditional) probabilities of the events A_2 and B_2 as well.

We now return to the previous example with two dice being thrown and recall the definition of the events A, B, C there. We argued that the probability that both dice result in a six *if we know* that the outcome of the first die is a six equals $1/6$. With the notation just introduced, we can then write

$$P(A|B) = \frac{1}{6}.$$

In the previous chapter, we have seen that the probability (of an event) is equal to the ratio of the number of favorable outcomes to the number of all possible outcomes. Let us see how this translates in the case of a *conditional event*, such as the event $A|B$, i.e. the event that something happens if we know that something else has already happened.

For the combined experiment of throwing two dice, the sample space consists of all pairs of outcomes

$$(1, 1), (1, 2), (1, 3), \dots, (6, 5), (6, 6),$$

with the number of such pairs being 36. If we have the information that the outcome of the first throw was a six, essentially our sample space reduces to only the following six pairs:

$$(6, 1)(6, 2), (6, 3), (6, 4), (6, 5), (6, 6),$$

with all other events (pairs) being impossible. Thus, if we know that B has occurred, the sample space has only six elements, and only one of them is favorable for A to occur, namely, the outcome $(6, 6)$. In that sense, conditional probabilities can also be thought of as being equal to the number of favorable events divided by the number of possible events (in the reduced sample space).

Another way to view this is to notice that, out of the possible pairs for the two throws, only six are favorable for the event B and, among these, just one is favorable for both A and B. In other words, we observe that

$$P(A|B) = \frac{1}{6} = \frac{|AB|}{|B|} = \frac{\frac{|AB|}{|\Omega|}}{\frac{|B|}{|\Omega|}},$$

so that we have

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

This is not a coincidence, but a fact that can be verified generally in cases where we can calculate directly both the conditional probability $P(A|B)$ and the unconditional probabilities $P(AB)$ and $P(B)$. The above answer we get for the probability that A occurs if we know that B has already occurred agrees with our intuition and it is in fact taken as the definition of the conditional probability, as given next.

Definition 3.1 Assume that Ω is a sample space and $B \subseteq \Omega$ an event in that space such that $P(B) > 0$. Then, for any event A defined on that space, the **conditional probability of A given B** is defined as

$$P(A|B) = \frac{P(AB)}{P(B)}. \tag{3.1}$$

It should be clear that the probability of B in the definition must be strictly positive, i.e. nonzero, since we cannot assume that an event has happened if that event has a zero probability. This is reflected mathematically in the fact that the denominator on the right-hand side of (3.1) cannot be zero.

It is worth noting that formula (3.1) is neither an axiom nor a theorem. It is a definition for conditional probability and its use as such, at least at an intuitive level, might be justified by a number of examples like the one given before Definition 3.1.

Historically, we have seen that the notion and foundations of probability theory arose from games of chance. In fact, people interested in such games were tacitly using this formula before the concept of conditional probability was formalized by Abraham de Moivre in his book *The Doctrine of Chances* in 1738.

In analogy with the conditional probability of the event A given B, we can define the conditional probability of B given A (provided that $P(A) > 0$) as

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

A common misconception, even today, is that $P(A|B)$ and $P(B|A)$ are equal. However, a quick glance at the last expression along with (3.1) reveals that this is generally not true (it is only true when $P(A) = P(B)$).

Example 3.1 In a certain country, the probability that a woman lives at least 70 years is 0.87, while the probability that she lives 75 years or more is 0.80. Suppose we select randomly a 70-year-old woman from that country. What is the probability that she will survive for the next five years, so that she reaches the age of 75?

SOLUTION Let A and B be the events that a randomly selected woman from the population lives at least 75 and 70 years, respectively. Then, it is known that

$$P(A) = 0.80, \quad P(B) = 0.87,$$

and we also observe that $AB = A$, since $A \subseteq B$. Then the required probability is the conditional probability $P(A|B)$, which can be obtained as

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(A)}{P(B)} = \frac{0.80}{0.87} \cong 0.92.$$

Proposition 3.1 *Let Ω be a sample space and $B \subseteq \Omega$ be an event with $P(B) > 0$. Then, the following properties hold:*

PC1. $P(A|B) \geq 0$ *for any event A in the sample space Ω.*

PC2. $P(\Omega|B) = 1$.

PC3. If $A_1, A_2, \ldots$ is a sequence of pairwise disjoint events in Ω, then

$$P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \sum_{i=1}^{\infty} P(A_i|B).$$

Proof: Properties PC1 and PC2 are immediate consequences from the definition of conditional probability and the obvious facts

$$P(AB) \geq 0, \quad \Omega B = B.$$

For PC3, we observe that

$$P\left(\bigcup_{i=1}^{\infty} A_i | B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) B\right)}{P(B)} = \frac{P\left(\bigcup_{i=1}^{\infty} A_i B\right)}{P(B)}$$

$$= \frac{\sum_{i=1}^{\infty} P(A_i B)}{P(B)} = \sum_{i=1}^{\infty} \frac{P(A_i B)}{P(B)} = \sum_{i=1}^{\infty} P(A_i|B),$$

where, in the second step, we have used the associative rule for the union–intersection of events, while in the third step, we have used the fact that the events $A_1B, A_2B, A_3B, \ldots$ are pairwise disjoint; this follows since for $i \neq j$,

$$(A_jB)(A_iB) = (A_jA_i)B = \emptyset B = \emptyset. \qquad \square$$

The above proposition shows that the set function $Q_B(\cdot)$, which is defined by the relationship

$$Q_B(A) = P(A|B) = \frac{P(AB)}{P(B)},$$

for all events A in Ω, satisfies properties P1, P2, P3 of Kolmogorov's axiomatic definition (see Definition 1.10). When we considered various properties of probability, all results we showed emerged from these three axioms; we thus see that the same properties are in fact valid for conditional probabilities as well.

More specifically, we have the following proposition.

Proposition 3.2 *Let Ω be a sample space and $B \subseteq \Omega$ be an event in that space with $P(B) > 0$. Then, the conditional probability $P(\cdot|B)$ satisfies the following properties:*

(a) $P(\emptyset|B) = 0$;

(b) For any event A, with complementary event A', we have $P(A'|B) = 1 - P(A|B)$;

(c) For any events A and C, we have $P(AC'|B) = P(A - C|B) = P(A|B) - P(AC|B)$;

(d) If $C \subseteq A$, then $P(C|B) \leq P(A|B)$;

(e) For any events A and C, we have $P(A \cup C|B) = P(A|B) + P(C|B) - P(AC|B)$;

(f) For any monotone sequence of events $\{A_n\}_{n \geq 1}$ in Ω, we have

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

In summary, we note that there are generally two ways of calculating a conditional probability of the form $P(A|B)$:

1. directly, by limiting the sample space of the experiment to those elements that are contained in the set (event) B and studying, in that space, the favorable events associated with the event A;

2. calculating the probabilities $P(AB)$ and $P(B)$ and then employing formula in (3.1).

It is advisable that one is aware of these two possibilities when dealing with conditional events and their associated probabilities. The choice of the best option for a particular problem depends inevitably on the nature of the problem and the ease with which the quantities involved can be calculated.

The next example illustrates the above and highlights the importance of identifying the correct sample space when the original space is limited due to conditioning upon some event.

Example 3.2

(i) Suppose Mr. and Mrs. Smith tell you they have two children and that one of them is a girl. What is the probability that the other is a girl too?

(ii) Suppose instead the Smiths tell you that their eldest child is a girl. What is the probability that the youngest one is a girl too?

SOLUTION

(i) The answer most people would give intuitively is $1/2$, since the other child is either a boy or a girl, and we assume equal probabilities for either.

There are, however, four possible gender outcomes: BB, BG, GB, and GG, where B and G denote boy and girl, respectively, and the letters are arranged in order of birth. Each combination is equally likely and so has a probability of 1/4. In exactly three cases, BG, GB, and GG, the family includes a girl; in just one of this group, GG, the other child is also a girl. So the probability of two girls, given that there is at least one girl, is actually 1/3.

The key to understand the solution here is that the information which the Smiths gave us is not that a specific child between the two they have is a girl, but that *at least one of their children is a girl*. With that knowledge, the restricted sample space of this experiment becomes $\{BG, GB, GG\}$.

Notice that with the above reasoning, no calculations were carried out to obtain the answer.[2] For those not convinced yet, we can alternatively employ (3.1) to obtain the same result as follows. We want the probability that "given that one child is a girl, the other is also a girl." Let A be the event that at least one of the Smith children is a girl. Define B as the event that the other child (the one that the Smiths did not tell us about) is also a girl. Then, the required probability is $P(B|A)$ and this is

$$P(B|A) = \frac{P(AB)}{P(A)}.$$

But clearly, assuming that both sexes are equally likely, we have

$$P(A) = \frac{3}{4} \quad \text{and} \quad P(AB) = \frac{1}{4},$$

since AB is the event that both children are girls. The answer is now obvious from the last two relations.

(ii) This time, the possible gender distributions are GB and GG, and the younger child is a girl only for GG. So, the probability becomes 1/2.

[2]This example is cited from the article "Mathematical Recreations: The interrogator's fallacy" by Ian Stewart in the *Scientific American*, September 1996. For a far-reaching discussion, stressing the importance of the *context* (loosely speaking, choosing the right sample space) when calculating probabilities, see the original article.

Example 3.3 In a hand of bridge, each of the four players receives 13 cards from a standard deck of 52 cards. Players are divided in pairs so that two players sitting opposite to one another are partners and play against the other two. Suppose now that Nick and Leopold are partners. At a given hand, Leopold has no ace. What is the probability that Nick does not have an ace either?

SOLUTION Let A be the event that Nick has no ace and B be the event that Leopold has no ace among his cards for the given hand. Then, the required probability is $P(A|B)$, which by the definition of conditional probability, equals

$$P(A|B) = \frac{P(AB)}{P(B)}.$$

The event B occurs if none of the 13 cards that Leopold has is an ace. The number of different ways in which this can occur is $\binom{48}{13}$, since his cards must be chosen from the 48 cards in the deck (excluding the four aces). Since there are $\binom{52}{13}$ different ways for Leopold's cards to be selected at any hand, we have

$$P(B) = \frac{\binom{48}{13}}{\binom{52}{13}} = \frac{\dfrac{48!}{13!35!}}{\dfrac{52!}{13!39!}} = \frac{48!39!}{52!35!}.$$

Consider now the event AB. This occurs if neither Leopold nor Nick have any aces, and so their combined set of 26 cards is selected entirely from the 48 cards in the deck excluding the four aces. Therefore, by a similar reasoning as above,

$$P(AB) = \frac{\binom{48}{26}}{\binom{52}{26}} = \frac{\dfrac{48!}{26!22!}}{\dfrac{52!}{26!26!}} = \frac{48!26!}{52!22!}.$$

The conditional probability $P(A|B)$ then equals

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{\dfrac{48!26!}{52!22!}}{\dfrac{48!39!}{52!35!}}$$
$$= \frac{26!35!}{22!39!},$$

which is about 18.2%.

The same answer can be obtained with less effort, however, by the direct argument suggested before Example 3.2. Arguing in this way, we limit the sample space of the experiment taking into account the knowledge that the event B has occurred. In this case, the solution can be formulated as follows: suppose we know *exactly* which 13 cards Leopold has (obviously, no ace is included in them). The remaining cards are

$52 - 13 = 39$ and there are $\binom{39}{13}$ ways for Nick's cards to be chosen. Among them, there are exactly $\binom{35}{13}$ different ways in which Nick's 13 cards contain no ace. Thus, if we know precisely which cards Leopold has, the probability that Nick has no ace equals

$$\frac{\binom{35}{13}}{\binom{39}{13}} = \frac{\frac{35!}{13!22!}}{\frac{39!}{13!26!}} = \frac{35!26!}{22!39!}, \tag{3.2}$$

which is the same result as the one obtained above. Note, however, that in order to obtain the result, we assumed that we know exactly the cards that Leopold has (not just that no ace is included in them). But, for any of the $\binom{39}{13}$ combinations of cards Leopold has without an ace, the answer is the same, and therefore, by reasons of symmetry, it must be the same if we do not know the exact cards, but we simply know that no ace is included in them. As a consequence, we find again that the probability of the event $P(A|B)$ is given by (3.2).

EXERCISES

Group A

1. In the examination of a Probability I course at a University, 180 students participated in the exam. Among these students, 80 study for a Mathematics degree, 60 for a Statistics degree, and 40 are in a joint degree in Mathematics and Economics.

   The course examiner selects a script with answers at random, and announces that this does not belong to a Mathematics degree student. What is the probability that the script belongs to a

   (a) Statistics student?

   (b) Mathematics and Economics student?

2. The percentages of people with each of the four blood types (O, A, B, and AB) in Iceland are as follows:

   type O: 56%; type A: 31%; type B: 11%; type AB: 2%.

   For a certain person in Iceland, we know that his red blood cells express A antigen, so that his blood type is either A or AB. What is the probability that he has blood type AB?

3. Nicky throws a die three times in succession. Consider the events

   A: the outcome of the second throw is a four;

   B: two throws out of the three resulted in a four.

   Calculate the probabilities $P(B|A)$ and $P(A|B)$. Are these equal?

4. Andrew tosses three coins. Find the probability that all three coins land heads if we know that

   (i) the first of the three coins landed heads;

   (ii) at least one coin landed heads.

5. Paul selects a card at random from a pack of 52 cards, and then selects a second one among the 51 remaining cards (i.e. without replacement).

   What is the probability that the second card drawn is an ace if we know that the first card

   (i) is a king?

   (ii) is an ace?

6. Suppose that Paul selects three cards at random without replacement. Find the probability that the third card drawn is a spade given that the first two cards included k spades. Give your answer for $k = 0, 1, 2$. (If, unlike Paul, you do not like card games at all, there are 13 spades in a regular 52-card pack!).

7. Henry throws two dice simultaneously. He observes the outcomes of the two throws and tells us that the two dice showed different faces.

   What is the probability that the sum of the two outcomes is

   (i) a six?

   (ii) either a two or a twelve?

8. The percentage of unemployed women in a population is 14%, while the general unemployment rate in the population is 11%. Assuming that the two sexes to be equally likely, we select a person at random and this person turns out to be unemployed. What is the probability that this person is

   (i) male?

   (ii) female?

9. Mary selects three cards at random from a regular 52-card pack without replacement. Let A_i be the event that the ith card drawn is a Queen for i = 1, 2, 3. Calculate the probabilities

   (i) $P(A_2|A_1)$;

   (ii) $P(A_3|A_1A_2)$;

   (iii) $P(A_2A_3|A_1)$.

10. Let A, B, and C be three events in a sample space Ω. Assuming that the following inequalities hold

$$P(A|B) \geq P(C|B) \quad \text{and} \quad P(A|B') \geq P(C|B'),$$

   verify that $P(A) \geq P(C)$.

11. In a large company, there are 500 electronic systems installed. Each of them is either connected to a network (N) or functions as a separate unit (U). Also, some

of them have incorporated a new high-speed device (H) while the remaining ones operate on an older, low-speed device (L). The frequencies for each of the four combinations above in the set of 500 systems are as follows:

	N	U
H	280	70
L	130	20

An engineer selects a system at random for inspection.

(i) If she finds out that the system is connected to the network, what is the probability that it operates on the high-speed device?

(ii) If she finds out that the system operates as a unit, what is the probability that the system operates on the low-speed device?

(iii) If she discovers that the system operates on the high-speed device, what is the probability that it is connected to the network?

(iv) If she discovers that the system operates on the low-speed device, what is the probability that it is connected to the network?

Using your results from the above parts, verify that the following **is not true** for two events A and B on a sample space:

$$P(A|B') = 1 - P(A|B).$$

Give an intuitive explanation for this fact.

12. Stephie, who is a theater-lover, attends a theater performance every week in one of the 25 theaters in her city. This year, 11 of these performances are comedies, while the remaining 14 are dramas. Every week, she selects a theater to visit at random among those she has not attended. After Stephie has been to five theater performances, her friend Nancy asks her what kind of shows she has been to. Stephie replies that the first three shows she went were comedies. What is the probability that the other two performances she attended were both dramas?

13. Suppose A and B are two events in a sample space Ω such that

$$P(A) = \alpha, \quad P(B) = \beta.$$

Show that the conditional probability $P(A|B)$ satisfies the inequality

$$\frac{\alpha + \beta - 1}{\beta} \leq P(A|B) \leq \frac{\alpha}{\beta}.$$

14. Let A and B be two events in a sample space Ω. Prove that $P(A|B) > P(A)$ holds if and only if $P(B|A) > P(B)$. In such a case, the two events A and B are said to be **positively correlated** since the knowledge that one has appeared increases the probability that the other appears, too.

Verify also the dual statement: $P(A|B) < P(A)$ holds if and only if $P(B|A) < P(B)$ (in which case, we say that A and B are **negatively correlated**).

Application: When selecting a card twice in succession, we define the events

A: at least one ace turns up;

B: the first two outcomes are different;

C: the first outcome *is not* an ace.

Examine whether each of the two pairs of events (A, B) and (A, C) are either positively or negatively correlated.

Group B

15. A large department store wants to study the purchasing habits of its customers for three specific products, a, b, c. The data given in the following table come from a market research that the store conducted and show the proportions of customers who purchase one or more among the three products during a specific period.

Product	Percentage of buyers
a	0.25
b	0.40
c	0.12
a & b	0.17
a & c	0.10
b & c	0.07
a, b & c	0.05

Assuming that these proportions reflect the probability for any customer to buy these product(s), calculate the probability that a customer

(i) buys product a given (s)he has bought at least one of products b and c;

(ii) buys at least one of products b and c if (s)he has bought product a;

(iii) buys product a if we know that (s)he has bought at least one of the three products a, b, and c.

16. Paul selects 6 cards from a pack of 52 cards and announces that three of them are spades. What is the probability that all six cards selected are spades?

17. A large PC manufacturing unit has 1000 CPU (central processing units) with speed 2.6 GHz. Each unit has been labeled with a number from 1 to 1000. The same manufacturer has also 1750 CPU with speed 3.0 GHz. Each of those units has been labeled with a number from 1 to 1750. We choose randomly a CPU, without knowing its speed, and observe that the number on its label is divisible by 6. What is the probability that it is also divisible by 8?

18. Prove Property (f) of Proposition 3.2 directly using the definition of conditional probability (Definition 3.1).

19. Suppose A_1, A_2, and A_3 are three events on a sample space Ω and let B be another event such that $P(B) > 0$. Show that

$$P(A_1 \cup A_2 \cup A_3) = S_1 - S_2 + S_3,$$

where

$$S_1 = P(A_1|B) + P(A_2|B) + P(A_3|B),$$
$$S_2 = P(A_1A_2|B) + P(A_1A_3|B) + P(A_2A_3|B),$$

and

$$S_3 = P(A_1A_2A_3|B).$$

Then, generalize this result for the case of n events $A_1, A_2, \dots, A_n$.

20. (The prisoner's dilemma) Three prisoners, A, B, and C, are sentenced to death and they have been put in separate cells. All three have equally good grounds to apply for parole and the parole board has selected one of them at random to be pardoned. The warden knows which one is pardoned, but is not allowed to tell. Knowing this, Prisoner A asks the warden to let him know the identity of one of the others who is going to be executed. "If B is to be pardoned, give me C's name. If C is to be pardoned, give me B's name. And if I'm to be pardoned, toss a coin to decide whether to name B or C."

    The warden tells A that B is to be executed. Prisoner A is pleased because he believes that his probability of surviving has gone up from 1/3 to 1/2, as it is now between him and C. Prisoner A secretly tells C the news, who when hearing the news believes that A still has a chance of 1/3 to be the pardoned one, but his chance has gone up to 2/3.

    What is the correct answer? Prisoner C is right, A's probability of surviving is still 1/3, but prisoner C's probability of receiving the pardon is 2/3. Explain why.

3.2 THE MULTIPLICATIVE LAW OF PROBABILITY

In many real life problems, we have to study families of events that can be put in some order such as logical, chronological, and so on. If such an ordering is possible, it is often easy to calculate conditional probabilities directly, upon conditioning each time on the events we have observed. To illustrate this, we follow on the example from the previous section with the box that contains four blue balls and six white balls. Sampling here is without replacement and we select successively two balls. Suppose we are interested in the probability that *both balls* drawn are white.

Define the events A_1, A_2, B_1, B_2 as in the previous section, so that for $i = 1, 2$, A_i is the event that the ith chosen ball is white and B_i is the event that the ith chosen ball is blue. Then, we observe that the required probability is $P(A_1A_2)$.

We have already seen in Section 3.1 that

$$P(A_1) = \frac{6}{10}$$

and

$$P(A_2|A_1) = \frac{5}{9}.$$

So far, we have used formula (3.1) to calculate the conditional probability $P(A|B)$ when both the (unconditional) probabilities $P(B)$ and $P(AB)$ are known. However, a second use of (3.1) is (as in the present example) to move in the other direction so that, using conditional probabilities, we can calculate the probability of the simultaneous occurrence of two events. More explicitly, applying (3.1) to the events A_1 and A_2 above, we have

$$P(A_2|A_1) = \frac{P(A_1A_2)}{P(A_1)},$$

which yields

$$P(A_1A_2) = P(A_1)P(A_2|A_1).$$

So, we obtain immediately that

$$P(A_1A_2) = P(A_1)P(A_2|A_1) = \frac{6}{10} \cdot \frac{5}{9} = \frac{1}{3}.$$

This example illustrates that sometimes it is easier to calculate a conditional probability $P(A|B)$ rather than the probability $P(AB)$ for the intersection of two events. In this case, formula (3.1) is used in the form

$$P(AB) = P(A)P(B|A).$$

Moreover, if $P(B) > 0$, the same argument applies when we interchange A and B, so that we have

$$P(AB) = P(B)P(A|B).$$

Therefore, we see that, in order to calculate the probability for the intersection of two events, it suffices to know the probability of one of them and the conditional probability of the other, given that the first has occurred. Depending on which of $P(A|B)$ and $P(B|A)$ is easier to find, we may use either of the last two formulas to obtain $P(AB)$.

Example 3.4 In a University class in Tokyo, 12 of the students own a car. Among these, seven have a Japanese car while the other five have a foreign car. If we ask two of these students about the origin of their car, what is the probability that

(i) the first person asked has a Japanese car and the second one has a foreign car?

(ii) the first person has a foreign car and the second one has a Japanese car?

(iii) both persons own a Japanese car?

(iv) both persons own a foreign car?

SOLUTION Consider the events

A_i: the ith person asked owns a Japanese car,

B_i: the ith person asked owns a foreign car.

The event in Part (i), namely, that the first person has a Japanese car and the second one has a foreign car, is the event A_1B_2, and similarly the events in Parts (ii), (iii), and (iv) are A_2B_1, A_1A_2, and B_1B_2, respectively.

Then, for the first part we see that

$$P(A_1B_2) = P(A_1)P(B_2|A_1) = \frac{7}{12}\frac{5}{11} = \frac{35}{132},$$

since, if the first person selected owns a Japanese car, among the remaining 11 students 5 own a foreign car, so that $P(B_2|A_1) = 5/11$.

Similarly, we find the probabilities for the other three parts to be

$$P(B_1A_2) = P(B_1)P(A_2|B_1) = \frac{5}{12}\frac{7}{11} = \frac{35}{132},$$

$$P(A_1A_2) = P(A_1)P(A_2|A_1) = \frac{7}{12}\frac{6}{11} = \frac{42}{132},$$

$$P(B_1B_2) = P(B_1)P(B_2|B_1) = \frac{5}{12}\frac{4}{11} = \frac{20}{132}.$$

Clearly, the sum of these four probabilities is one since one of the associated four events must occur (these events are mutually exclusive and collectively exhaustive). Thus, the last probability could have also been calculated by the formula

$$P(B_1B_2) = 1 - P(A_1B_2) - P(B_1A_2) - P(A_1A_2).$$

It is also useful to note that all four probabilities may be calculated alternatively by using the classical definition of probability. For convenience, we assign a number from 1 to 7 to each of the 7 students who have a Japanese car while, to each of the students who have a foreign car, we assign one of the numbers 8, 9, 10, 11, 12. Then, the sample space for this experiment (the selection of the two students among the 12) consists of all pairs of numbers between 1 and 12, so that it can be written in the form

$$\Omega = \{(i,j) : i = 1, 2, \dots, 12,\ j = 1, 2, \dots, 12 \text{ and } i \neq j\}.$$

The number of elements in Ω, which is the number of all possible outcomes for the experiment, is $|\Omega| = 12 \cdot 11 = 132$ and since the two students were selected completely at random, each of these 132 (elementary) events has the same probability, $1/132$. It is now easy to describe each of the events A_iB_j for $i, j \in \{1, 2\}$ and, for each for them, to find the number of favorable outcomes. For instance, the event A_1B_2 is described by the subset

$$B_1A_2 = \{(i,j) : i = 8, 9, \dots, 12 \text{ and } j = 1, 2, \dots, 7\},$$

while the event B_1B_2 by the subset

$$B_1B_2 = \{(i,j) : i, j = 8, 9, \dots, 12 \text{ and } i \neq j\}.$$

Consequently, we obtain the probabilities of these two events to be

$$P(B_1A_2) = \frac{|B_1A_2|}{|\Omega|} = \frac{5 \cdot 7}{132} = \frac{35}{132}$$

and

$$P(B_1B_2) = \frac{|B_1B_2|}{|\Omega|} = \frac{5 \cdot 4}{132} = \frac{20}{132}.$$

We can similarly find the probabilities for the other two parts of the example.

We have seen above that the formula

$$P(AB) = P(B)P(A|B) = P(A)P(B|A)$$

gives us a very convenient way to find probabilities for intersections of events, and we may use either of the two equalities above, typically depending on which of $P(A|B)$ and $P(B|A)$ is easier to get. The same argument applies when we have more than two events; suppose A, B, and C are three events defined on a sample space Ω and assume that $P(AB) > 0$ (this implies that $P(A) > 0$). Then, we can write

$$P(B|A) = \frac{P(AB)}{P(A)}, \quad P(C|AB) = \frac{P(ABC)}{P(AB)}$$

and these two expressions now yield

$$P(ABC) = P(C|AB)P(AB) = P(C|AB)P(B|A)P(A).$$

In general, we have the following result, known as the **multiplicative law** (or the multiplicative rule) of probability.

Proposition 3.3 *Assume that $A_1, A_2, \dots, A_n$ are events in a sample space Ω for which we have $P(A_1A_2 \cdots A_{n-1}) > 0$. Then, we have*

$$P(A_1A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1}).$$

Proof: We observe first that

$$P(A_1) \geq P(A_1A_2) \geq \cdots \geq P(A_1A_2 \cdots A_{n-1}) > 0,$$

so that all the conditional probabilities in the formula we have to prove are properly defined. Thus, using Definition 3.1 for each of these, we obtain

$$\begin{aligned} &P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1}) \\ &= P(A_1) \cdot \frac{P(A_1A_2)}{P(A_1)} \cdot \frac{P(A_1A_2A_3)}{P(A_1A_2)} \cdots \frac{P(A_1A_2 \cdots A_n)}{P(A_1A_2 \cdots A_{n-1})} \end{aligned}$$

and noticing that all terms but $P(A_1A_2 \cdots A_n)$ on the right-hand side cancel, the result follows. □

Example 3.5 A box contains five red (R) balls and eight blue (B) balls. If we select four balls at random from the box, find the probability that we observe the sequence of colors *RBRB* if

(i) the ball selected at each step is not put back in the box (sampling without replacement);

(ii) when we select a ball from the box, we observe its color and then put it back in the box (sampling with replacement).

SOLUTION Let us define the events

$$A_i : \text{the } i\text{th ball selected is red}$$

and

$$B_i : \text{the } i\text{th ball selected is blue,}$$

for $i = 1, 2, 3, 4$. Then, we require the probability $P(A_1B_2A_3B_4)$ that, according to Proposition 3.3, can be expressed in the form

$$P(A_1B_2A_3B_4) = P(A_1)P(B_2|A_1)P(A_3|A_1B_2)P(B_4|A_1B_2A_3).$$

(i) We calculate each of the probabilities on the right-hand side of the last expression. Specifically, we have

$$P(A_1) = \frac{5}{13}.$$

Next, given that A_1 has occurred, after the first selection, there are 12 balls in the box, 8 of which are blue; thus

$$P(B_2|A_1) = \frac{8}{12} = \frac{2}{3}.$$

Conditioning upon the intersection of events A_1 and B_2, meaning that both A_1 and B_2 have occurred (that is, the first ball we selected was red and the second one was blue), there are 11 remaining balls of which 4 are red. This means that the probability to draw a red ball in the third selection is

$$P(A_3|A_1B_2) = \frac{4}{11}.$$

Finally, arguing as above we see that

$$P(B_4|A_1B_2A_3) = \frac{7}{10}.$$

Combining all the above expressions, we find the required probability to be

$$P(A_1B_2A_3B_4) = \frac{5}{13} \cdot \frac{2}{3} \cdot \frac{4}{11} \cdot \frac{7}{10} = \frac{28}{429} \approx 0.065\,27.$$

(ii) In this case, since we have replacement, the number of red and blue balls in the box remains the same before each selection, and this is the same as it is in the beginning, i.e. five red and eight blue balls. Consequently, we get

$$P(A_1) = P(A_3|A_1B_2) = \frac{5}{13}$$

and

$$P(B_2|A_1) = P(B_4|A_1B_2A_3) = \frac{8}{13}.$$

By employing the multiplicative law in Proposition 3.3, we get the required probability to be

$$P(A_1B_2A_3B_4) = \left(\frac{5}{13}\right)^2 \cdot \left(\frac{8}{13}\right)^2 = \frac{1600}{28\ 561} \approx 0.056\ 02,$$

and we see that the answers for the two parts of the question are quite different, as we might have expected.

Example 3.6 Each pack from a certain brand of cereals has a coupon. There is a total of six different coupons and a person who collects one from each type wins a prize. Jimmy has just started collecting coupons and bought three packs of cereals. What is the probability that all three have different types of coupons?

SOLUTION This problem can obviously be solved by combinatorial methods, but it is probably easier to think about it in terms of conditional probabilities. We define the events

A: The first two coupons are of a different type.

B: All three coupons are of a different type.

The probability we want is $P(B)$ but, rather than attacking this directly, it is much easier to find $P(A)$ and the conditional probability $P(B|A)$. More specifically, the probability of the event A is simply $5/6$; assuming that all coupons are equally likely, there are five favorable choices for A, which are all types of coupons except that on the first pack of cereals. Similarly, once we know that the first two packs have different coupons, then for B to occur we require that the third pack contains one from the four remaining types out of the possible six, and so

$$P(B|A) = \frac{4}{6}.$$

An appeal to the multiplicative law now yields the answer to the question readily as

$$P(B) = P(A)P(B|A) = \frac{5}{6} \cdot \frac{4}{6} = \frac{20}{36} = \frac{5}{9}.$$

Example 3.7 After a bank robbery, the police have arrested 12 people, 4 of whom are actually responsible for the robbery. The police investigator selects one person among the 12 for questioning, then a second and finally a third. What is the probability that among these three persons,

(i) all three are innocent;

(ii) they are all innocent or all guilty.

SOLUTION We define the events

A_i: The ith person chosen for questioning is innocent,

B_i: The ith person chosen for questioning is guilty,

for $i = 1, 2, 3$.

(i) We require the probability of the event $P(A_1A_2A_3)$ and, by the multiplicative law, this is

$$\begin{aligned} P(A_1A_2A_3) &= P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \\ &= \frac{8}{12} \cdot \frac{7}{11} \cdot \frac{6}{10} = \frac{336}{1320} = \frac{42}{165} \cong 0.2545. \end{aligned}$$

(ii) For this part, we seek the probability of the event $A_1A_2A_3 \cup B_1B_2B_3$ and, since the events $A_1A_2A_3$ and $B_1B_2B_3$ are disjoint, we have $P(A_1A_2A_3 \cup B_1B_2B_3) = P(A_1A_2A_3) + P(B_1B_2B_3)$. The first of these probabilities was found above, while for the second, we similarly get

$$\begin{aligned} P(B_1B_2B_3) &= P(B_1)P(B_2|B_1)P(B_3|B_1B_2) \\ &= \frac{4}{12} \cdot \frac{3}{11} \cdot \frac{2}{10} = \frac{24}{1320} = \frac{3}{165} \cong 0.0182. \end{aligned}$$

Therefore the required probability is

$$P(A_1A_2A_3 \cup B_1B_2B_3) = 0.2545 + 0.0182 = 0.2727.$$

EXERCISES

Group A

1. Kate is in the final year of her studies and she has to choose exactly one of two optional courses offered this semester. She would prefer to take Course I, which she likes best, but she feels that this is difficult and estimates that the probability of getting an A grade in this course is 25%, while in Course II she estimates that probability to be 40%. She decides to leave the decision in the hands of chance and throws a die. If the outcome is at least 3, she will choose Course I, otherwise she will choose Course II.

(i) What is the probability that Kate chooses Course II and gets an A grade in the exam?

(ii) What is the probability that Kate gets an A grade in the exam?

2. With reference to Example 3.6, suppose that there are m different types of coupons and Jimmy buys r packs of cereals (with $r < m$).

   (i) What is the probability that the coupons contained in these r packs are all of a different type?

   (ii) Assuming $r > 2$, find the probability that the r coupons belong to exactly 2 different types.

3. From an usual pack of 52 cards, we select cards without replacement until the first diamond is drawn. What is the probability that this happens with the 4th card drawn? (*Hint*: Let E_i be the event that the ith card selected is *not* a diamond. The event we seek is then $E_1 E_2 E_3 E_4'$.)

4. Tom has a bowl that contains four white balls and three red balls. He selects balls successively from the bowl (at random and without replacement) and puts them one next to the other.

   The following day, Tom has the same bowl with four white balls and three red ones, but he thinks the previous game was boring, so he tries a variation which he finds funnier. When he selects a ball from the bowl, if it is white he returns it into the bowl along with *two more white balls*. If the ball selected is red, he returns it into the bowl and puts *three more red balls* in it.

   For which of the two variations of Tom's game above there is a higher probability that the first three balls selected are all red?

5. Maria has bought a toy which contains a bag with the 26 letters of the alphabet in it.

   (i) Maria selects five letters at random. What is the probability that the letters she chose can be rearranged so that the word MATHS is produced?

   (ii) If she selects seven letters rather than five, what is the probability that five of the letters chosen produce the word MATHS?

6. During a football season in the English Premier League in football, Manchester United won 25 games, had 9 draws, and lost 4 games. If we do not know the order that United faced their opponents, so that the probability that United wins, loses or draw a game is the same for all 38 matches, find the probability that they drew the first three games they played.

Group B

7. An insurance company classifies the claims arriving as being either low (L) or high (H). On a certain day, 21 claims arrived, 12 of which were L. At the end of the day, a company employee registers the claims in a file without knowing the order in which they were received during the day. What is the probability that, in the first four claims registered, there is

(i) the sequence *LHLH* in the claims?

(ii) at least one high claim?

(iii) at least one high claim and at least one low claim?

8. A pharmaceutical company produces boxes of tablets for a particular disease. Each box contains 20 tablets. The quality control unit of the company selects a box at random and examines the tablets to see if any of them are defective. If a particular box contains two defective tablets, what is the probability that

   (i) the two defective ones are found in the first three tablets examined?

   (ii) the third tablet examined is defective?

   (iii) the third tablet checked is the second one that is found defective?

9. In an oral exam at a University, the course lecturer has to examine r female and s male students. The order in which the students are examined is assumed to be random.

   Consider the events

   A: all female students are examined before the first male student;

   B: no two students of the same gender are examined one after the other.

   (i) If $s = r + 1$, show that $P(A) = P(B)$;

   (ii) If $s = r$, prove that $P(A) = 2P(B)$.

10. An urn contains a red and b green balls. We select k balls without replacement with $k \leq \min\{a, b\}$. Show that the probability all selected balls are of the same color equals
$$\frac{(a)_k + (b)_k}{(a+b)_k}.$$

    Application: In a lottery with 49 numbers, 6 are selected in each draw. Find the probability that, in a particular draw, all six winning numbers are

    (i) even;

    (ii) odd.

11. A University degree program enrolled this year r female and s male students. If students are registered at the University in a completely random order, what is the probability that, for $k \leq \min(r, s)$, during the first $2k$ registrations no two students of the same sex are registered successively?

3.3 THE LAW OF TOTAL PROBABILITY

Let A and B be two events on a probability space and suppose we want to find the probability, $P(A)$, that A occurs. There are many instances where this is difficult to work out directly, but it is much easier to find the conditional probability of A given that B occurs, i.e. $P(A|B)$. If we can also find $P(B)$ and $P(A|B')$, where, as usual, B' is the complement of B, then the following result, known as the *law of total probability*, enables us to find $P(A)$.

Proposition 3.4 *Let B be an event defined on a sample space* Ω *such that* $0 < P(B) < 1$. *Then, for any event A defined on the same sample space, we have*

$$P(A) = P(A|B)P(B) + P(A|B')P(B'). \tag{3.3}$$

Proof: It is apparent that

$$A = A\Omega = A(B \cup B') = AB \cup AB'.$$

In addition, we note that AB and AB' are disjoint events. Consequently,

$$P(A) = P(AB \cup AB') = P(AB) + P(AB').$$

The condition $0 < P(B) < 1$ implies that both $P(B)$ and $P(B') = 1 - P(B)$ are strictly positive, and so the conditional probabilities $P(A|B)$ and $P(A|B')$ are both properly defined. Using the multiplicative law, we can then write

$$P(AB) = P(A|B)P(B), \quad P(AB') = P(A|B')P(B')$$

and the result of the proposition follows readily by combining the above equations. □

The above proposition tells us that the (unconditional) probability of an event A can be expressed as a weighted average of the probabilities $P(B)$ and $P(B')$ for some other event B, where the weights are, respectively, the conditional probability that A occurs given that B occurs and the conditional probability that A occurs given that B does not occur.

Example 3.8 An insurance company classifies its customers as being of type I (high risk) or type II (low risk). From the company's records, it has been estimated that 20% of its customers are type I while 80% are type II. Also, the company estimates that a type-II customer has a probability 0.25 of making at least one claim in any given year, while the corresponding probability for a type-I customer is twice as much.

Calculate the probability that a new customer will make at least one claim for the first year of being insured.

SOLUTION Let us define the events

A: the newly arrived customer makes at least one claim during the first year of being insured,

B: the new customer belongs to type I, i.e. he/she is a high-risk customer.

We then have been given that

$$P(B) = 0.2, \quad P(B') = 0.8, \quad P(A|B) = 0.5, \quad P(A|B') = 0.25.$$

We seek the probability that the customer makes at least one claim, which is $P(A)$, and from the law of the total probability, this is given by

$$\begin{aligned} P(A) &= P(A|B)P(B) + P(A|B')P(B') \\ &= (0.5)\cdot(0.2) + (0.25)\cdot(0.8) = 0.1 + 0.2 = 0.3, \end{aligned}$$

that is, there is a 30% chance that the new customer will make at least one claim.

The tree diagram in Figure 3.1 illustrates the calculations necessary in order to apply the law of total probability for this example.

As is apparent from the proof of Proposition 3.4, the key step in establishing that proposition was writing the event A in the form

$$A = AB \cup AB',$$

which is a union of two disjoint events. This expression and the argument underlying Proposition 3.4 can be extended to cover the case wherein the sample space Ω is written as the union of more than two disjoint events. To be specific, let us assume that the events $B_1, B_2, \ldots, B_n$ are such that

$$B_1 \cup B_2 \cup \cdots \cup B_n = \Omega, \qquad B_i \cap B_j = \emptyset \quad \text{for } i \neq j.$$

Under the above assumptions, the family $\{B_1, B_2, \ldots, B_n\}$ is called a *partition* of the sample space (see Figure 3.2). To illustrate this idea, we present a few simple examples:

- In a single throw of a die, if A_i represents the event that the outcome of the die is i, for $1 \leq i \leq 6$, it is clear that the events A_i are pairwise disjoint and their union covers the sample space $\Omega = \{1, 2, 3, 4, 5, 6\}$ for this experiment.

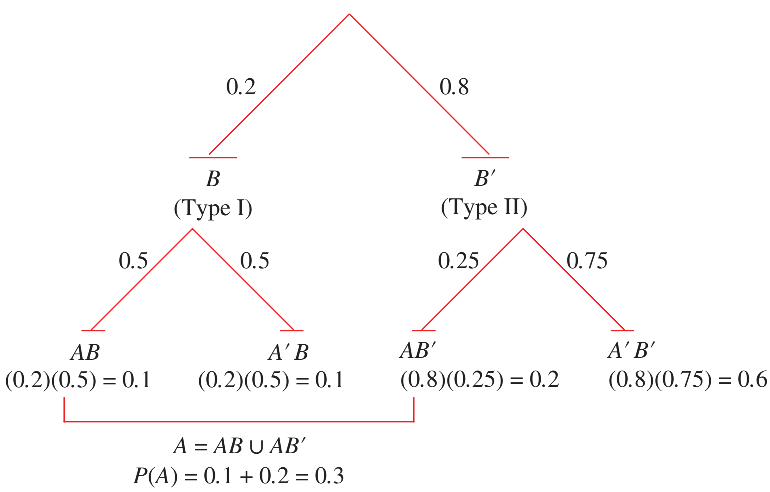


Figure 3.1 A tree diagram for the law of total probability for Example 3.8.

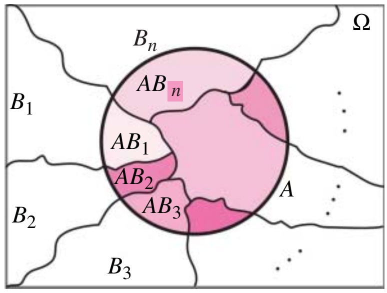


Figure 3.2 A partition of the sample space Ω.

However, there are many more partitions of Ω for the same experiment. For instance, if C_1 denotes the event that the outcome is an odd integer and C_2 the event that the outcome is an even integer, then $C_1 \cup C_2 = \Omega$ and $C_1 \cap C_2 = \emptyset$, and so $\{C_1, C_2\}$ is again a partition of Ω.

- In the experiment of tossing a coin three times in succession, the sample space has the form

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}.$$

Let now B_i, for $i = 0, 1, 2, 3$, be the event that heads turn up i times in this experiment. Then, $\{B_0, B_1, B_2, B_3\}$ forms a partition of Ω, where the set B_0 has one element (TTT), B_1 has three elements (HTT, THT, TTH), B_2 has three elements (THH, HHT, HTH), and B_3 has one element (HHH).

- Suppose that the price of a stock at close on a particular day has a known value, say x. We want to know whether this stock price will be higher, lower or the same at close on the following day. Let E_1 be the event that the price increases, E_2 that it remains unaltered and E_3 that it goes down. Then, $\{E_1, E_2, E_3\}$ is a partition of the sample space for this experiment.

The following proposition uses the concept of a partition of a sample space to generalize the result of Proposition 3.4. The idea is that if $\{B_1, B_2, \ldots, B_n\}$ is such a partition, and we know the conditional probabilities $P(A|B_i)$ as well as the unconditional probabilities $P(B_i)$, then we are in a position to find $P(A)$.

Proposition 3.5 *Given a sample space Ω, let $\{B_1, B_2, \ldots, B_n\}$ be a partition of Ω such that $P(B_i) > 0$, for all $i = 1, 2, \ldots, n$. Then, for any $A \subseteq \Omega$, we have*

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n) \\ &= \sum_{i=1}^{n} P(A|B_i)P(B_i). \end{aligned}$$

Proof: We first write, in analogy with Proposition 3.4,

$$A = A\Omega = A(B_1 \cup B_2 \cup \cdots \cup B_n) = AB_1 \cup AB_2 \cup \cdots \cup AB_n$$

and since for the events AB_i, AB_j, we have, for any $i \neq j$,

$$(AB_i)(AB_j) = A(B_iB_j) = A\emptyset = \emptyset,$$

it follows from the additive property of probabilities that

$$P(A) = P\left(\bigcup_{i=1}^{n} AB_i\right) = \sum_{i=1}^{n} P(AB_i).$$

The result of the proposition becomes now obvious since from the multiplicative law, for each summand above, we have $P(AB_i) = P(A|B_i)P(B_i)$; note that we have also used the fact that $P(B_i) > 0$ for all i, which shows that each conditional probability $P(A|B_i)$ is properly defined. □

When $n = 2$, it is obvious that Proposition 3.5 reduces to Proposition 3.4. It is also worth noting that Proposition 3.4 can be generalized for the case in which the partition $\{B_i\}$ contains not a finite number but countably many events $\{B_1, B_2, \dots\}$. In this case, the following formula holds:

$$P(A) = \sum_{i=1}^{\infty} P(A|B_i)P(B_i).$$

Finally, we mention that the condition $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$ in Proposition 3.5 could be replaced by the weaker condition

$$A \subseteq B_1 \cup B_2 \cup \cdots \cup B_n$$

(can you explain why Proposition 3.5 still holds in this case?).

Example 3.9 About 14% of men and 2% of women are colorblind. Find the probability that, if we select randomly a person from a large auditorium attending a lecture, this person will be colorblind, assuming that in the auditorium there are

(i) 80 men and 60 women;

(ii) an equal number of men and women;

(iii) three times more women than men.

SOLUTION We define the events

A: the person selected is colorblind,
B: the person selected is female,

so that B' is the event that the selected person is male. Then, we are given that

$$P(A|B) = \frac{2}{100}, \quad P(A|B') = \frac{14}{100}.$$

(i) Since there are 60 women and 80 men in the auditorium, we have

$$P(B) = \frac{60}{60+80} = \frac{3}{7}, \quad P(B') = \frac{80}{60+80} = \frac{4}{7}$$

(for the latter probability, we could simply use $P(B') = 1 - P(B)$). The law of total probability (Proposition 3.4) now yields readily

$$\begin{aligned} P(A) &= P(A|B) \cdot P(B) + P(A|B') \cdot P(B') \\ &= \frac{2}{100} \cdot \frac{3}{7} + \frac{14}{100} \cdot \frac{4}{7} = \frac{62}{700}, \end{aligned}$$

or about 8.86%.

(ii) In this case we have $P(B) = P(B') = 1/2$, and so upon using Proposition 3.4, we obtain

$$\begin{aligned} P(A) &= P(A|B) \cdot P(B) + P(A|B') \cdot P(B') \\ &= \frac{2}{100} \cdot \frac{1}{2} + \frac{14}{100} \cdot \frac{1}{2} = \frac{16}{200}, \end{aligned}$$

i.e. there is a 8% chance that the person chosen will be colorblind (since there are as many males as females, this percentage is simply the average of the two probabilities for males and females, respectively, that is, 14% and 2%).

(iii) Here, we are given that there are three times more females than males, and so

$$P(B) = \frac{3}{4}, \quad P(B') = \frac{1}{4}.$$

Upon using Proposition 3.4, we see that

$$\begin{aligned} P(A) &= P(A|B) \cdot P(B) + P(A|B') \cdot P(B') \\ &= \frac{2}{100} \cdot \frac{3}{4} + \frac{14}{100} \cdot \frac{1}{4} = \frac{20}{400} = \frac{1}{20}, \end{aligned}$$

so that the required probability is now 5%.

Example 3.10 A computer equipment store sells USB sticks of a certain type. Currently, there are 45 USB sticks in the shelves of the store of which, due to a manufacturer's problem, four are defective. The salesperson, who does not know the existence of the defective items, each time a customer wants to buy a USB stick, selects one at random from the shelves and hands it to the customer. Find the probability that

(i) the third stick which will be sold will be defective;

(ii) both the second and the third USB sticks that will be sold will be defective.

SOLUTION We begin by defining the events

D_i: the ith USB stick that will be sold is defective,

N_i: the ith USB stick that will be sold is not defective

for $i = 1, 2, 3$.

(i) We want to find the probability of the event D_3. Since we do not have any information about the first two USB sticks sold, we would expect intuitively that $P(D_3) = 4/45$, that is, the probability that the third stick is defective is the same as the probability that the first stick is defective.

To verify this formally, we have to consider how many (and which, if any) of the first two items sold are defective. We observe first that the events $D_2N_1, D_2D_1, N_2D_1, N_2N_1$ form a partition of the sample space for the selection of the first two items sold. Thus, the law of the total probability gives that

$$\begin{aligned} P(D_3) &= P(D_3|D_2N_1)P(D_2N_1) + P(D_3|D_2D_1)P(D_2D_1) \\ &\quad + P(D_3|N_2D_1)P(N_2D_1) + P(D_3|N_2N_1)P(N_2N_1) \\ &= \frac{3}{43}P(D_2N_1) + \frac{2}{43}P(D_2D_1) + \frac{3}{43}P(N_2D_1) + \frac{4}{43}P(N_2N_1). \end{aligned} \tag{3.4}$$

By an application of the multiplicative law, we find that

$$P(D_2N_1) = P(D_2|N_1)P(N_1) = \frac{4}{44} \cdot \frac{41}{45},$$

and similarly

$$\begin{aligned} P(D_2D_1) &= P(D_2|D_1)P(D_1) = \frac{3}{44} \cdot \frac{4}{45}, \\ P(N_2D_1) &= P(N_2|D_1)P(D_1) = \frac{41}{44} \cdot \frac{4}{45}, \\ P(N_2N_1) &= P(N_2|N_1)P(N_1) = \frac{40}{44} \cdot \frac{41}{45}. \end{aligned}$$

Upon substituting all the above values into (3.4), we obtain $P(D_3) = 4/45$, which agrees with the initial guess that we made, before doing these calculations.

(ii) In this case, we want the probability $P(D_2D_3)$. Another application of the law of total probability now yields this to be

$$P(D_2D_3) = P(D_2D_3|N_1)P(N_1) + P(D_2D_3|D_1)P(D_1). \tag{3.5}$$

If we know that the first stick that the store sells is nondefective, the remaining sticks will be 4 defective and 40 nondefective ones. Therefore

$$P(D_2D_3|N_1) = \frac{4}{44} \cdot \frac{3}{43}.$$

In a similar fashion, we find

$$P(D_2D_3|D_1) = \frac{3}{44} \cdot \frac{2}{43}.$$

Since $P(N_1) = 41/45$, $P(D_1) = 4/45$, putting all this information into (3.5) we get

$$\begin{aligned} P(D_2D_3) &= \frac{4}{44} \cdot \frac{3}{43} \cdot \frac{41}{45} + \frac{3}{44} \cdot \frac{2}{43} \cdot \frac{4}{45} \\ &= \frac{1}{165}, \end{aligned}$$

after some straightforward calculations.

EXERCISES

Group A

1. Sixty percent of the students in a University class are females. If, among the female students, 25% have joined the University Sports Club to do at least one sport, and the corresponding percentage for male students is 35%, calculate the proportion of students who do at least one sport at the University.

2. A bowl contains six white and five red balls, while a second bowl contains three white and seven red balls. We select randomly a ball from the first bowl and place it in the second. Then, we choose at random a ball from the second bowl. What is the probability that the ball selected is red?

3. From an usual pack of 52 cards, we select a card at random. Then we select another card from the remaining 51 cards. What is the probability that *the second card chosen* is

   (a) an ace?

   (b) a diamond?

   (c) the ace of diamonds?

4. Among the drivers insured with an insurance company, 45% made no claims during a year, 35% made one claim, and 20% made at least two claims. The probabilities that a driver will make more than one claim during a year if during the previous year the driver had 0, 1, and 2 or more claims, are $0.1, 0.3$, and 0.6, respectively. Find the probability that a randomly selected driver will make at least two claims during the following year.

5. A factory has three production lines that produce 50%, 30%, and 20%, respectively, of the items made in the factory during a day. It has been found that 0.7% of the items produced in the first line of production are defective, while in the second and third lines the corresponding proportions are 1% and 1.2%, respectively. Calculate the proportion of defective items produced in the factory during a day.

6. John has a red and a blue die and throws them simultaneously.
   (i) What is the probability that the outcome of the blue die is larger than that of the red die?
   (ii) Find the probability that the difference between the two outcomes is equal to 2.

Group B

7. We throw a die and, if the outcome is i, then we toss a coin i times. What is the probability that in these coin tosses,
   (a) no heads appear?
   (b) only one face of the coin appears, that is if we toss the coin i times, then we get either i heads or i tails?

8. We throw a die and if the outcome is k $(1 \le k \le 6)$, then we select a ball from an urn that contains $2k$ white balls and $14 - 2k$ black ones. Show that the probability of selecting a white ball is equal to the probability of selecting a black one from the urn.

9. (A generalization of the law of total probability) Let $B_1, B_2, \ldots, B_n$ be disjoint events on a sample space Ω such that $P(B_i) > 0$, for all $i = 1, 2, \ldots, n$. Prove that for any event A on this sample space, the following holds:

$$P(A|B) = \frac{1}{P(B)} \sum_{i=1}^{n} P(B_i)P(A|B_i),$$

   where

$$B = \bigcup_{i=1}^{n} B_i.$$

   Explain how Proposition 3.5 can be deduced as a special case of this result.

10. Let Ω be a sample space, A an event of that space and $B_1, B_2, \ldots, B_n$ be pairwise disjoint events in Ω with $P(B_i) > 0$ for all i. Assume further that

$$P(A|B_i) = p \quad \text{for all } i = 1, 2, \ldots, n.$$

    If the event B is defined by

$$B = \bigcup_{i=1}^{n} B_i,$$

    show that the conditional probability of A given B is also equal to p, that is $P(A|B) = p$. (*Hint*: You may find the result of the previous exercise useful.)

11. Assume that in a lottery, 20 balls numbered 1 to 20 are put in a large bowl and then 3 balls are selected, one after the other, at random and without replacement. What is the probability that the *third ball* drawn has the largest number on it?

    (*Hint*: Let E be the event that the third ball has the largest number, and B_i be the event that the maximum of the first two numbers is equal to i for $1 \le i \le 20$. Find the probabilities $P(B_i)$ for each i and apply the law of the total probability.)

3.4 BAYES' FORMULA

A very common problem in probability theory is the calculation of the *a posteriori* probabilities $P(B_i|A)$, based on the a priori (or unconditional) probabilities $P(B_i)$ and the conditional probabilities $P(A|B_i)$. The general expression for that calculation is in fact an application of the multiplicative law and the law of total probability and is attributed to Reverend Thomas Bayes.

Proposition 3.6 *Let $\{B_1, B_2, \ldots, B_n\}$ be a partition of a sample space Ω such that $P(B_i) > 0$, for all $i = 1, 2, \ldots, n$. Then, for any event A of the same sample space such that $P(A) > 0$, we have*

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)}$$

$$= \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}, \quad i = 1, 2, \ldots, n. \tag{3.6}$$

Proof: The proof is immediate if in the expression

$$P(B_i|A) = \frac{P(AB_i)}{P(A)}$$

we replace the numerator by using the multiplicative law, namely,

$$P(AB_i) = P(A|B_i)P(B_i) \quad \text{for } i = 1, 2, \ldots, n$$

and the denominator from the law of the total probability (Proposition 3.4). □

In the form presented above, Proposition 3.6 is due to Pierre-Simon Laplace (1749–1827), but he called it *Bayes' theorem*, to honor the English philosopher and priest Thomas Bayes (1701–1761). The latter had attempted a systematic study for calculating conditional probabilities of the form $P(B|A)$ via the conditional probability $P(A|B)$. Bayes' work was followed up by famous mathematicians such as Laplace, Gauss, and others. Laplace, in particular, used the above proposition more than two hundred years ago, to tackle a number of problems in diverse areas such as medicine, celestial mechanics, and jurisprudence.

Bayes' theorem has found a large variety of applications in statistics and for this reason it has attracted a lot of interest (but has also been the subject of debate) among statisticians. Nowadays, the statistical theory, named Bayesian statistics, has become quite prominent! In this approach, probability is considered as a degree of belief in something to occur, i.e. probabilities are treated in a subjective manner. Bayesian statistics has grown enormously in recent decades and is currently applied to every scientific area which uses statistics.

In the special case of $n = 2, B_1 = B$, and $B_2 = B'$, Bayes' theorem (sometimes also called Bayes' rule or Bayes' formula) takes on the simpler form

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')} \tag{3.7}$$

and

$$P(B'|A) = \frac{P(A|B')P(B')}{P(A|B)P(B) + P(A|B')P(B')}.$$

These formulas are used when we know, or we can easily find, each of the quantities $P(A|B), P(A|B')$, and $P(B)$ and we are interested in one (or both) of the probabilities $P(B|A), P(B'|A)$, where the conditioning has been reversed.

In the most common applications of Bayes' rule, the event A may occur (logically or in time) *after* the event B. Therefore, it is typically easier to find the conditional probability that the most recent event A occurred, given that the (earlier) event B occurred, and then seek the probability that B occurred given that the later event A occurred.

The following examples aim to clarify the above ideas.

Example 3.11 A contestant on a television show has to answer multiple choice questions with four possible answers. The probability that the contestant knows the answer to a question is 75%. If the contestant does not know the answer to a particular question, she gives an answer at random. If she has answered the first question correctly, what is the probability that she knew the answer?

SOLUTION We define the events

A: the contestant answers the question correctly,

B: the contestant knows the answer to the question.

This is a case in which we want to find $P(B|A)$ and we know both the unconditional probabilities of the events B and B' as well as the conditional probabilities of A given that either B or B' has occurred. Specifically, from what is given, we know that

$$P(B) = \frac{3}{4}, \quad \text{and so } P(B') = 1 - P(B) = \frac{1}{4},$$

and

$$P(A|B) = 1, \quad P(A|B') = \frac{1}{4}.$$

It is now clear that (3.7) applies to yield

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B')P(B')} = \frac{12}{13}.$$

In this example, we know that the event B occurs (or not) before we know whether A occurs; once the contestant hears the question, she either knows the answer or not and, after she chooses an answer, we find out whether this is correct. The answer to this exercise, which is $12/13$, shows that it is quite unlikely the contestant has replied at random if the answer to the question she has given is correct.

Example 3.12 A factory has three production lines associated with the production of an item. The percentage of items produced in each of these three lines is 50% for the first, 30% for the second and 20% for the third. Assume further that 0.7% of the items produced in line A are defective, while the percentage of defective items in the second line is 0.9%, and in the third line it is 1.3%. A sample of the items produced is then examined and suppose it is found to be defective. Find the probability that it was produced in

(i) the first line;

(ii) the second line;

(iii) the third line.

SOLUTION For any particular item produced in the factory, we define the events

A: the item is defective

and

B_i: the item was produced in line i (for $i = 1, 2, 3$).

The required probability for Part (i) is then $P(B_1|A)$ and, by Bayes' theorem, this is

$$
\begin{aligned}
P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\
&= \frac{(0.50)(0.007)}{(0.50)(0.007) + (0.30)(0.009) + (0.20)(0.013)} \\
&\cong 0.398 = 39.8\%.
\end{aligned}
$$

The probabilities for the second and third lines are, respectively,

$$
\begin{aligned}
P(B_2|A) &= \frac{P(A|B_2)P(B_2)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\
&= \frac{(0.30)(0.009)}{(0.50)(0.007) + (0.30)(0.009) + (0.20)(0.013)} \\
&\cong 0.307 = 30.7\%
\end{aligned}
$$

and

$$
\begin{aligned}
P(B_3|A) &= \frac{P(A|B_3)P(B_3)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + P(A|B_3)P(B_3)} \\
&= \frac{(0.20)(0.013)}{(0.50)(0.007) + (0.30)(0.009) + (0.20)(0.013)} \\
&\cong 0.295 = 29.5\%
\end{aligned}
$$

(the last probability might also have been found by subtracting from one the sum of the other two).

Example 3.13 It has been estimated that about 0.25% of the general population suffer from a disease. To diagnose whether someone suffers from this disease, there is a medical examination which has a probability 1% of giving a false result if someone has the disease and 2% if someone does not have the disease. If we select at random a person from the general population and he/she tests positive for the disease, what is the probability that this person actually suffers from this disease?

SOLUTION We define the events

B_1: the person selected suffers from the disease,

B_2: the person selected does not suffer from the disease,

A: the result of the test is positive.

Then, we are given that

$$P(B_1) = 0.0025, \quad P(B_2) = 1 - 0.0025 = 0.9975$$

and

$$P(A|B_1) = 0.99, \quad P(A|B_2) = 0.02.$$

Then, by an appeal to Bayes' formula, we get

$$\begin{aligned} P(B_1|A) &= \frac{P(A|B_1)P(B_1)}{P(A|B_1)P(B_1) + P(A|B_2)P(B_2)} \\ &= \frac{(0.99)(0.0025)}{(0.99)(0.0025) + (0.02)(0.9975)} = \frac{0.002\,475}{0.022\,425} \cong 0.1104. \end{aligned}$$

Notice that, although the medical examination is quite accurate (it gives the wrong answer only in 1% or 2% of the cases, depending on the patient's actual status), if a person is detected as suffering from the disease, the probability that (s)he is actually ill is just 11%. This might seem paradoxical at first sight; however, it can explained if one bears in mind that the disease is *very rare* (only 0.25% of the population suffer from it). So, when screening a patient with no a priori knowledge of his/her health status, we should be very cautious even if the result of the test is positive. As the above calculations show, most of the positive diagnoses are due to the inaccuracy of the medical examination rather than due to patients actually suffering from the disease.

In statistics, especially when applied to medical sciences, it is important to differentiate between the two types of error associated with a patient's diagnosis. The case when the patient does not have the disease but the result of the test is positive is known as a *false positive* result, while if the result is negative but the patient does in fact suffer from the disease is called a *false negative* result. In the present example, the probability of a false positive result is

$$P(B_2|A) = 1 - 0.1104 = 0.8896,$$

i.e. about 89%, which is obviously very large.

EXERCISES

Group A

1. Students at a University take a Probability exam in three classrooms. The number of students who are well-prepared (W) and poorly-prepared (P) for the exam in each of the three classrooms are as follows:

   Classroom I: 60 W, 20 P;

   Classroom II: 50 W, 30 P;

   Classroom III: 65 W, 15 P.

   Students who are well prepared pass the exam with a probability of 0.85, while students who are poorly prepared pass the exam with a probability of 0.5.

   (i) We select a classroom randomly and then from this room we select a student at random. What is the probability that he/she passes the exam?

   (ii) If the student selected has passed the exam, what is the probability that he/she took the exam in classroom II?

2. A motor insurance company classifies its customers as good drivers (G) and bad drivers (B). 65% of the company's customers are classified as G. The probability that a good customer makes a claim in any particular month is 0.02, while the same probability for a bad driver is 0.07. What is the probability that the last claim made to the company came from a good customer?

3. A box B_1 contains 3 red and 6 blue balls, a second box B_2 contains 7 red and 7 blue balls, while a third box B_3 has 5 red and 9 blue balls. We select a box at random and then from this box we pick a ball at random.

   If the ball selected is red,

   (i) what is the probability that the ball came from B_1?

   (ii) what is the probability that the ball came from either B_1 or B_3?

4. A box B_1 contains four red and six black balls, while a second box B_2 contains seven red and three black balls. We select a ball from B_1 and place it in B_2. Then, we pick up a ball from B_2 at random and find out that it is red. What is the probability that the ball we selected from B_1 and placed into B_2 was red?

5. Electric bulbs manufactured in a production unit are packaged in boxes, with each box containing 120 bulbs. The probability that a box has i defective bulbs is $1/5$, for each $i = 0, 1, 2, \ldots, 4$. If we choose 10 bulbs from a box and none is defective, what is the probability that this box contains

   (i) no defective bulbs?

   (ii) at least two defective bulbs?

6. Suppose that in a painting exhibition, 96% of the exhibits are genuine, while the remaining 4% are fake. A painting collector can identify a genuine painting as such with a probability 90%, while if the painting is fake the probability that the collector

finds this out is 80%. If she exits the exhibition having just bought a painting, which she obviously thinks is genuine, what is the probability that it is not?

7. In a certain company, there are three secretaries responsible for typing the mail of the manager. When she types a letter, Secretary A has a probability of 0.04 for making at least one misprint, while this probability for Secretary B is 0.06 and for Secretary C is 0.02. The probability that a letter is typed either by Secretary A or Secretary B is the same, while Secretary C types three times more letters than any of the other two secretaries.

   This morning, the manager has left a hand-written letter on the secretaries' box and when he returned, he found that there was a misprint on it.

   (i) What is the probability that the letter was typed by Secretary A?

   (ii) What would this probability be if the manager knew that Secretary C is on leave this week?

8. Diana is about to go out with her friends and her mother asks her how much money she has in her purse. Diana says she has either a \$10 note or a \$20 note, but she can't remember. Her mother puts in her purse a \$20 note without Diana noticing it. Later on, when she visits the local cinema, Diana puts her hand in the purse and takes out a \$20 note. What is the probability that the one left in her purse is also a \$20 note?

9. A telecommunications system transmits binary signals (0 or 1). The system includes a transmitter that emits the signals and a receiver which receives those signals. The probability that the receiver registers a signal 1 when the transmitter has sent a signal 1 is 99.5%, while the probability that the receiver registers a signal 0 when the transmitter has sent a signal 0 is 98%. Signals are transmitted every second, and so 60 signals are sent during one minute. If the last signal has been registered as 1, find the probability that the original signal transmitted was also 1, if it is known that

   (a) equal numbers of 0's and 1's are transmitted every minute;

   (b) the number of 0's transmitted during a minute is four times the number of 1's.

Group B

10. Among male smokers, the lifetime risk of developing lung cancer is 17%; among female smokers, this risk is 12%. For nonsmokers, this risk is significantly lower: 1.3% for men and 1.5% for women. Assuming equal numbers of men and women in the general population, find the probability that

    (i) a man who develops lung cancer is a nonsmoker;

    (ii) a woman who develops lung cancer is a smoker;

    (iii) a *person* who develops lung cancer is a smoker.

11. We have n chips numbered $1, 2, \dots, n$. Tom, who likes fancy experiments, selects a chip at random and if the number on that chip is i, he tosses a coin i times $(1 \le i \le n)$. Tom has just completed this experiment and he tells us that no heads

appeared in the coin tossing. What is the probability that the number of coin tosses was k (for $k = 1, 2, \ldots, n$)?

12. Pat takes part in a quiz show with multiple choice questions. There are three possible answers to each question. The probability that she knows the answer to a question is 80%. If Pat does not know the answer to a particular question, she gives an answer at random. If she has answered the first two questions correctly, what is the probability that she guessed both answers?

3.5 INDEPENDENT EVENTS

If A and B are two disjoint events of a sample space Ω such that $P(B) > 0$, we have $AB = \emptyset$, and so we have

$$P(A|B) = \frac{P(AB)}{P(B)} = 0.$$

On the other hand, if two events A and B are such that $B \subseteq A$, then

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Both these results are easily interpreted in an intuitive way:

- In the first case, since it is impossible for A and B to occur simultaneously, knowledge that B has occurred means that A cannot occur;
- If B is a subset of A, this means that B may only occur in conjunction with A; thus, knowing that B has occurred, we are certain that A will occur too.

In the above situations, knowing that an event B occurred makes us certain that another event, A, has or has not occurred. In the present section, we look at the other extreme; knowing that B occurs gives us no information about the chances of A occurring. Intuitively this means that, knowing that B occurs does not change the probability for A to occur, i.e. the unconditional and the conditional (given B) probabilities of an event A are the same.

To illustrate this idea, we first present an example. We throw two dice and then consider the events

- A: the outcome of the first die is an even number;
- B: the outcome of the second die is an odd number.

One expects here that, if we know that B has (or has not) occurred, this does not have any impact on whether A occurs or not. In order to verify this formally, we calculate first the two conditional probabilities, $P(A|B)$ and $P(A|B')$.

It is clear that the sample space for this experiment (throwing two dice) encompasses all 36 pairs of possible outcomes. Then, it is very easy to see that A appears in exactly 18 among these possible outcomes, and the same is true for B. Further, there are 9 combinations of pairs of outcomes in which both A and B occur (all pairs (i, j), where i is 2, 4 or 6 and j is 1, 3 or 5), so that

$$|A| = 18, \quad |B| = 18, \quad |AB| = 9$$

and, consequently,

$$P(A) = P(B) = P(A') = P(B') = \frac{1}{2}, \quad P(AB) = \frac{9}{36} = \frac{1}{4}.$$

Thus, we obtain from the above that

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{1/4}{1/2} = \frac{1}{2}$$

and

$$P(A|B') = \frac{P(AB')}{P(B')} = \frac{P(A) - P(AB)}{1 - P(B)} = \frac{1}{2},$$

so that the two conditional probabilities are equal.

More generally, if for two events A and B, we have

$$P(A|B) = P(A|B') = p,$$

then it follows from the law of total probability that

$$P(A) = P(A|B)P(B) + P(A|B')P(B') = pP(B) + p(1 - P(B)) = p,$$

so that

$$P(A) = P(A|B') = P(A|B') = p.$$

If for two events A and B with $P(B) > 0$ we have

$$P(A) = P(A|B),$$

then we say that event A is *independent* of event B. In the case where we have also $P(A) > 0$, from the last expression we successively obtain

$$\begin{aligned} P(A) = P(A|B) &\Longleftrightarrow \frac{P(AB)}{P(B)} = P(A) \Longleftrightarrow P(AB) = P(A)P(B) \\ &\Longleftrightarrow \frac{P(AB)}{P(A)} = P(B) \Longleftrightarrow P(B|A) = P(B), \end{aligned}$$

which means that B is also independent of A. Thus, we see that the relationship of independence between events is symmetric. For this reason, instead of saying "the event A is independent of B" or that "the event B is independent of A", we say that "the events A and B are independent." Technically, in order to avoid using the conditions $P(A) > 0$ and $P(B) > 0$ (which are needed so that the conditional probabilities $P(B|A)$ and $P(A|B)$ are meaningful), we use the third of the equivalent conditions above, i.e. the condition $P(AB) = P(A)P(B)$ in the definition of independence between events, which is as follows.

Definition 3.2 Let A and B be two events on the same sample space. Then, A and B are said to be **independent** if

$$P(AB) = P(A)P(B).$$

If, on the other hand, we have

$$P(AB) \neq P(A)P(B),$$

then we say that A and B are not independent, or that they are **dependent**.

From the definition above it follows that, if for an event A we have $P(A) = 0$ or $P(A) = 1$, then A is independent of *any* other event B defined on the same sample space (see Exercise 3 at the end of this section).

Example 3.14 In the experiment of throwing a die twice, we define the events

$$A: \text{the outcome of the first die is 4,}$$

$$B: \text{the sum of the two outcomes is 6.}$$

Considering all 36 possible outcomes of this experiment and counting how many of these belong to each of the sets A, B, and AB, we find

$$P(A) = \frac{6}{36} = \frac{1}{6}, \quad P(B) = \frac{5}{36}, \quad P(AB) = \frac{1}{36}.$$

It is then clear that $P(AB) \neq P(A)P(B)$, and so the events A and B are not independent. A simple intuitive interpretation of this fact is obtained as follows: if the outcome of the first die is 6 (in which case A does not occur), then we are certain that B will not occur; if, on the other hand, the first die lands 4, so that A occurs, the probability that B occurs is $1/6$. Thus, whether or not B occurs depends on the outcome of the first die which means that A and B are not independent.

Consider, however, a third event C defined as

$$C: \text{ the sum of the two outcomes is 7.}$$

Then, we can easily check that $P(AC) = 1/36$ and

$$P(A)P(C) = \frac{1}{6} \cdot \frac{6}{36} = \frac{1}{36},$$

and so in this case the events A and C *are* independent. Can you explain intuitively why this is so?

Example 3.15 In order to succeed in a computer science course at a University, students have to pass two exams: a class test and a computer lab exam. The course teacher estimates that 87% of the students pass the first exam, 78% of the students pass the second exam, while the percentage of those who pass both exams, so that they complete the course successfully, is 72%. Is the performance of a student in the first exam independent of that in the second one?

SOLUTION Let us define the events

$$A: \text{the student passes the first exam,}$$

$$B: \text{the student passes the second exam.}$$

Then, we are given that $P(A) = 0.87$ and $P(B) = 0.78$. Further, AB represents the event that the student passes both exams and we are also given that $P(AB) = 0.72$. Since

$$P(AB) = 0.72 \neq (0.87)(0.78) = P(A)P(B),$$

we see that the events A and B are *not* independent.

The next result shows that the property of independence between two events is preserved if one, or both, of these events are replaced by their complements.

Proposition 3.7 *Let A and B be two events on a sample space Ω and assume that A and B are independent. Then, the same is true for each of the following pairs of events:*

(i) *A and B';*

(ii) *A' and B;*

(iii) *A' and B'.*

Proof: This follows immediately from the definition of independence and the result of Exercise 3 in Section 1.5. □

It is worth noting at this point that if an event A is independent of two other events B and C, then *it is not* always independent of their union $B \cup C$ or their intersection BC. A counterexample for this is given in Exercise 13.

The concept of independence can be generalized when we consider more than two events. Let us start with the case of three events, A_1, A_2, A_3. A natural generalization of the concept of independence considered above is to suggest that three events A_1, A_2, A_3 are independent if the occurrence of any of these three events, *or any pair of these events*, does not change the probability that the remaining event(s) occur. In this line of thinking, A_1, A_2, A_3 would be independent if and only if each of the pairs

$$\{A_1, A_2\}, \{A_1, A_3\}, \{A_2, A_3\}, \{A_1, A_2A_3\}, \{A_2, A_1A_3\}, \text{ and } \{A_3, A_1A_2\}$$

consist of independent events. This amounts to saying that the following relations all hold:

$$P(A_1A_2) = P(A_1)P(A_2), \quad P(A_1A_3) = P(A_1)P(A_3), \quad P(A_2A_3) = P(A_2)P(A_3)$$

and

$$P(A_1(A_2A_3)) = P(A_1)P(A_2A_3),$$
$$P(A_2(A_1A_3)) = P(A_2)P(A_1A_3),$$
$$P(A_3(A_1A_2)) = P(A_3)P(A_1A_2).$$

However, the last three relations could be replaced (why?) by the following simpler formula

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3).$$

We thus arrive at the following definition.

Definition 3.3 Let A_1, A_2, and A_3 be three events on the same sample space Ω. Then, A_1, A_2, and A_3 are said to be (completely) independent if the following relations hold:

$$P(A_1A_2) = P(A_1)P(A_2), \quad P(A_1A_3) = P(A_1)P(A_3), \quad P(A_2A_3) = P(A_2)P(A_3) \tag{3.8}$$

and

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3). \tag{3.9}$$

In view of the above definition, one might wonder whether the condition (3.9) is redundant, i.e. if it can be inferred from those in (3.8). The answer is negative, as the following example demonstrates.

Example 3.16 During a television competition show Peggy, who is tonight's winner of the show, enters the final stage where she will win (at least) one of three possible prizes, say, $\alpha_1, \alpha_2, \alpha_3$. In particular, she is presented with four boxes; three of them contain one of the prizes $\alpha_1, \alpha_2, \alpha_3$, respectively, while the fourth contains all three prizes. Let us define the events

$$A_i: \text{ Peggy wins the prize } \alpha_i,$$

for $i = 1, 2, 3$. Since she may open each of the four boxes with equal probability, it is clear that $P(A_i) = 2/4$; further, the chance that she wins any pair of the prizes, or all three of them is $1/4$, that is

$$P(A_1A_2) = P(A_1A_3) = P(A_2A_3) = \frac{1}{4}$$

and

$$P(A_1A_2A_3) = \frac{1}{4}.$$

It is therefore immediate that

$$P(A_iA_j) = P(A_i)P(A_j), \quad 1 \le i < j \le 3, \tag{3.10}$$

but

$$P(A_1A_2A_3) = \frac{1}{4} \neq P(A_1)P(A_2)P(A_3).$$

Following the above example, three events A_1, A_2, and A_3 such that (3.10) holds are said to be **pairwise independent**, so that we distinguish them from completely independent events, the latter being as defined in Definition 3.3. To make the distinction between these two concepts clearer, we mention that in the case of events which are pairwise independent, the simultaneous occurrence of two of them may affect the probability that the third event occurs, something which cannot happen in the case of completely independent events.

Once we have considered independence among two and three events, we now introduce the general notion of independence for a finite collection of events. Let $A_1, A_2, \ldots, A_n$ be n such events on a sample space. Then, we say that these are (completely) independent if the (simultaneous) occurrence of any number among them does not affect the probability that any other event (or combination of the remaining events) occurs.

For this case, if we write down all conditions which must hold for the n events to be completely independent, we arrive at a rather cumbersome set of equations. As in the case of three events above, it turns out that this set is equivalent to a neater set of expressions, which in fact can be written in a simple formula as given in the following definition.

Definition 3.4 Assume that $A_1, A_2, \ldots, A_n$, $n \geq 2$, are events on the same sample space Ω. Then, these events are said to be (completely) independent if

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$$

holds for any choice of k different indices $i_1, i_2, \ldots, i_k$ in the set of integers $\{1, 2, \ldots, n\}$ and for any $k = 2, 3, \ldots, n$.

According to Definition 3.4, in order to verify that the events $A_1, A_2, \ldots, A_n$ are independent we have to check, for all combinations of indices $1 \leq i < j < k < \ldots \leq n$, that the following relations are true:

$$\begin{aligned}
P(A_iA_j) &= P(A_i)P(A_j),\\
P(A_iA_jA_k) &= P(A_i)P(A_j)P(A_k),\\
&\vdots\\
P(A_1A_2\cdots A_n) &= P(A_1)P(A_2)\cdots P(A_n).
\end{aligned}$$

It is clear that there are $\binom{n}{2}$ equalities with two indices i and j, $\binom{n}{3}$ equalities with three indices i, j, and k – and so on – and finally $\binom{n}{n} = 1$ equality with all n indices. Thus, we have to examine the validity of

$$\binom{n}{2} + \binom{n}{3} + \cdots + \binom{n}{n} = \sum_{r=2}^{n}\binom{n}{r} = \sum_{r=0}^{n}\binom{n}{r} - \binom{n}{0} - \binom{n}{1} = 2^n - n - 1$$

equations. Obviously, this number grows very quickly with n, and so it can in principle be very laborious to check whether n events are independent. Fortunately, what happens usually in practice, is that **we do not examine** whether n given events are independent, but **we use** the independence conditions to calculate probabilities which refer to intersections among these events. The reason for this is that in many cases, from the nature of the experiment and the very definition of the events, we are in a position to determine whether two or more events are independent or not. For example,

- in an experiment when we throw two dice, any event associated **only** with the outcome of the first die is independent from any event associated **only** with the outcome of the second die;
- when the experiment involves successive draws of a lottery (e.g. draws in successive weeks), it is reasonable to assume that any event associated with the draw taking place in week i is independent from any event associated with the draw taking place in week j when $i \neq j$;
- in experiments relating to the measurement of total lifetimes, or other performance indices of various industrial products, such as electrical appliances and equipment, computers, cars, airplanes, and so on, we typically make the assumption that events associated with different items are independent.

We also note at this point that, in view of Proposition 3.7 and the definition of independence for more than two events, it is easy to see that the proposition remains valid when an arbitrary number of independent events are involved. Specifically, we have the following property.

Assume that the events $A_1, A_2, \dots, A_n$ are completely independent and let, for $j = 1, 2, \dots, n$, B_j be either the event A_j or its complement, A'_j. Then, we have

$$P(B_{i_1} B_{i_2} \cdots B_{i_k}) = P(B_{i_1}) P(B_{i_2}) \cdots P(B_{i_k})$$

for any choice of k different indices $i_1, i_2, \dots, i_k$ in the set $\{1, 2, \dots, n\}$ and for any $k = 2, 3, \dots, n$.

Example 3.17 (System reliability)
An electrical system comprises four parts (units), which function independently of one another and are connected as shown in Figure 3.3. In order that the system works, we must have either units 1 and 2 working simultaneously or units 3 and 4 working simultaneously. The probabilities that the four component units of the system work (known in engineering as *"unit reliability"*) are, respectively, 0.98, 0.90, 0.95, and 0.85, for units 1, 2, 3, and 4. What is the probability that at a particular instant the *system works?* (As explained in Section 1.11, this is known as *"system reliability"* and is usually denoted by R.)

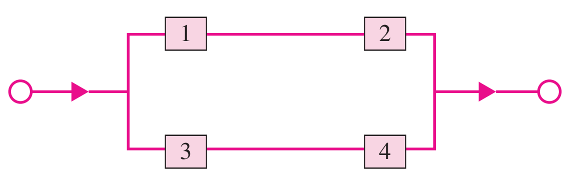


Figure 3.3 An electrical system.

SOLUTION For the particular point in time we are interested in, let us define the events

$$A_i: \quad \text{the } i\text{th unit of the system works}$$

for $i = 1, 2, 3, 4$. Then, the system works if

- both units 1 and 2 are functioning properly, i.e. the event A_1A_2 occurs, or
- both units 3 and 4 are functioning properly, i.e. the event A_3A_4 occurs.

Consequently, the system reliability R is given by

$$R = P(A_1A_2 \cup A_3A_4),$$

which gives

$$R = P(A_1A_2) + P(A_3A_4) - P(A_1A_2A_3A_4).$$

Since we know that the units operate independently, the events A_1, A_2, A_3, A_4 are (completely) independent, which yields

$$P(A_1A_2) = P(A_1)P(A_2) = (0.98)(0.90) = 0.882,$$

$$P(A_3A_4) = P(A_3)P(A_4) = (0.95)(0.85) = 0.8075,$$

$$P(A_1A_2A_3A_4) = P(A_1)P(A_2)P(A_3)P(A_4) = (0.882)(0.8075) = 0.712\,215.$$

These expressions finally give

$$R = 0.882 + 0.8075 - 0.712\,215 = 0.977\,285,$$

that is, the reliability of the system is almost 98%.

Example 3.18 (Parallel and serial connections)
Let $A_1, A_2, \dots, A_n$ be independent events on a sample space Ω with $P(A_i) = p_i$, $i = 1, 2, \dots, n$. Assuming that the probabilities p_i are known, find an expression for the probability of the following events:

(i) none of the events $A_1, A_2, \dots, A_n$ appear;

(ii) at least one of the events $A_1, A_2, \dots, A_n$ appear.

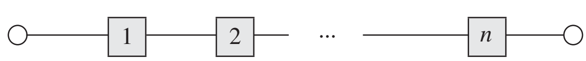


Figure 3.4 Serial connection.

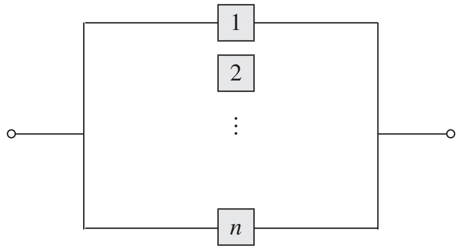


Figure 3.5 Parallel connection.

Application: In engineering, a connection of n units is called a **serial connection** when the system works if and only if all its component units work; on the other hand, a **parallel connection** is when a system works if and only if at least one of its units work. Find the reliability of a serial and a parallel system consisting of n independent units, when each unit has reliability p (Figures 3.4 and 3.5).

SOLUTION

(i) The event we are interested in here is

$$A = A_1' A_2' \cdots A_n'$$

and by using the independence assumption of the events A_i (see also the comment before Example 3.17), we find that

$$P(A) = P(A_1' A_2' \cdots A_n') = \prod_{i=1}^{n} P(A_i') = \prod_{i=1}^{n} (1 - P(A_i)) = \prod_{i=1}^{n} (1 - p_i).$$

(ii) Here, we seek the probability of the event

$$B = A_1 \cup A_2 \cup \cdots \cup A_n = (A_1' A_2' \cdots A_n')' = A',$$

and from Part (i) this is

$$P(B) = 1 - P(A) = 1 - \prod_{i=1}^{n} (1 - p_i).$$

Application. Let for $i = 1, 2, \ldots, n$, the event A_i be

$$A_i: \text{ the } i\text{th unit of the system works.}$$

Then, we are given that the events $A_1, A_2, \ldots, A_n$ are independent and $P(A_i) = p$, for all i. First we find that the reliability R_s of a serial system is

$$R_s = P(A_1 A_2 \cdots A_n) = \prod_{i=1}^{n} P(A_i) = \prod_{i=1}^{n} p = p^n,$$

while in the case of a parallel system, applying the results above we see that its reliability R_p is

$$R_p = P(A_1 \cup A_2 \cup \cdots \cup A_n) = P(B) = 1 - \prod_{i=1}^{n}(1-p) = 1-(1-p)^n.$$

Many random experiments can be analyzed if we decompose them into simpler ones, or if we express them as a repetition of the same experiment, say n times. In such a case, let $\Omega_1, \Omega_2, \dots, \Omega_n$ be the sample spaces of these simpler experiments; then, the events of the whole experiment can be expressed as an ordered n-tuple of events in the spaces Ω_i. As a result, it is reasonable that for the original experiment, we take its sample space to be the Cartesian product

$$\Omega = \Omega_1 \times \Omega_2 \times \cdots \times \Omega_n. \tag{3.11}$$

If we now have $A_i \subseteq \Omega_i$, for $i = 1, 2, \dots, n$, as events in the respective sample spaces Ω_i, the set

$$A = A_1 \times A_2 \times \cdots \times A_n \tag{3.12}$$

is an event for the original experiment, and this event occurs if

- the outcome of the first simple experiment belongs to A_1, and
- the outcome of the second simple experiment belongs to A_2, and

$$\vdots$$

- the outcome of the nth simple experiment belongs to A_n.

Assuming that the n individual experiments (or the n repetitions of the same experiment) are independent, we obtain the probability of event A in (3.12) as

$$P(A) = P(A_1)P(A_2)\cdots P(A_n) = \prod_{i=1}^{n} P(A_i).$$

It is worth noting, however, that there are events defined on Ω which cannot take on the form of a Cartesian product, as above. For instance, the event

$$A' = \Omega - (A_1 \times A_2 \times \cdots \times A_n)$$

cannot always be written as a Cartesian product of subsets of the spaces $\Omega_1, \Omega_2, \dots, \Omega_n$. For such events, we usually calculate their probabilities using the standard properties of probability (see Section 1.4). It is also important to note that not all subsets of the Cartesian product in (3.11) are events for the original experiment, but only those that can be written using the set operations of union, intersection and taking complements among sets of the form $A_1 \times A_2 \times \cdots \times A_n$. For our purposes in this book, we shall not pursue this topic further but refer the interested reader to more advanced texts on probability theory.[3]

[3]In such texts, spaces of the form (3.11) are called product spaces.

We also note that the above discussion can be extended to the case where an infinite number of experiments is considered.

Independent experiments have a central role in probability theory. Beyond cases wherein the nature of the problem leads naturally to the consideration of **successive independent experiments**, it is often useful to adopt this setup although the description of the problem may not immediately point to this. For example,

- tossing a coin n times, or tossing n coins simultaneously, may be considered equivalent to an experiment consisting of n independent repetitions of an experiment which entails tossing a coin *once*;
- throwing two dice can be regarded, for probability calculations, the same as throwing a die twice;
- studying the genders of the children in a family with n children, the random "experiment," which consists of n births, is essentially the same as n independent repetitions of the experiment, which involves the single birth of a child.

Proposition 3.8 *In a chance experiment with sample space* Ω*, let* A *and* B *be two disjoint subsets of* Ω*.*

(a) In n *independent repetitions of this experiment,*

(i) the probability that the event A *occurs in each of the* n *repetitions is* $[P(A)]^n$*;*

(ii) the probability that the event A *does not occur in any of the* n *repetitions is* $[1 - P(A)]^n$*.*

(b) In successive repetitions of this experiment, the probability that A *appears before* B *equals*

$$\frac{P(A)}{P(A) + P(B)}.$$

Proof:

(a) For $i = 1, 2, \ldots, n$, we define the events

A_i: the event A occurs in the ith repetition of the experiment;

B_i: the event B occurs in the ith repetition of the experiment.

Then, we have

$$P(A_i) = P(A), \quad P(B_i) = P(B), \qquad i = 1, 2, \ldots, n.$$

The result in (i) is an immediate consequence of the formula

$$P(A_1 \times A_2 \times \cdots \times A_n) = \prod_{i=1}^{n} P(A_i) = [P(A)]^n,$$

while the result in (ii) is established using that

$$P(A_1' \times A_2' \times \cdots \times A_n') = \prod_{i=1}^{n} P(A_i') = \prod_{i=1}^{n} [1 - P(A_i)] = [1 - P(A)]^n.$$

(b) We now define the events

$$C_1: \text{the event } A \text{ appears in the 1st repetition of the experiment,}$$

and for $n = 2, 3, \ldots,$

C_n: the event $A \cup B$ does not occur in the first $n - 1$ repetitions of the experiment and A occurs in the nth repetition of the experiment.

Then, we have

$$P(C_1) = P(A),$$

$$P(C_n) = [1 - P(A \cup B)]^{n-1} P(A), \quad n = 2, 3, \ldots .$$

The event we are interested in can then be expressed as the union of disjoint events of the form $\bigcup_{n=1}^{\infty} C_n$, so that its probability can be determined as

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} C_n\right) &= \sum_{n=1}^{\infty} P(C_n) = P(C_1) + \sum_{n=2}^{\infty} P(C_n) \\
&= P(A) + \sum_{n=2}^{\infty} [1 - P(A \cup B)]^{n-1} P(A) \\
&= \sum_{n=1}^{\infty} [1 - P(A \cup B)]^{n-1} P(A) \\
&= \frac{P(A)}{1 - [1 - P(A \cup B)]} = \frac{P(A)}{P(A \cup B)} \\
&= \frac{P(A)}{P(A) + P(B)},
\end{aligned}$$

where in the last step we have used the fact that A and B are disjoint sets. □

Example 3.19 Peter throws two dice successively. He wins if he gets an outcome where the sum of the two dice is at least 10 *before* an outcome of 1 or 2 appears on any of the two dice. What is the probability that Peter wins this game?

SOLUTION We shall try to use the result of the previous proposition, since we want the probability that, in a series of independent trials, "something happens before something else happens." We define the events

A: the sum of the two dice outcomes (in a single throw of the dice) is $10, 11$ or 12,

and

B: 1 or 2 appears on any of the two dice.

First, we note that A and B cannot occur on the same trial, and so they are disjoint events. It is then easy to see that

$$P(A) = \frac{6}{36} = \frac{1}{6}, \quad P(B) = 1 - P(B') = 1 - \frac{4^2}{6^2} = \frac{20}{36} = \frac{5}{9},$$

so that, using the result of Proposition 3.8, we obtain Peter's chance of winning this game to be

$$\frac{P(A)}{P(A) + P(B)} = \frac{3}{13},$$

that is, he has about a 23% chance.

EXERCISES

Group A

1. If A and B are two events on a sample space such that $0 < P(A), P(B) < 1$, verify that each of the following conditions is equivalent to the independence of A and B:

   (i) $P(A|B) + P(A') = 1$;

   (ii) $P(B|A) + P(B') = 1$;

   (iii) $P(B|A') = P(B|A)$;

   (iv) $P(A|B) = P(A|B')$;

   (v) $P(B'|A') = P(B'|A)$;

   (vi) $P(A'|B') = P(A'|B)$.

2. If the events A and B are independent and $A \subseteq B$, show that $P(A) = 0$ or $P(B) = 1$.

3. Let A and B be two events on a sample space. Then, show the following:

   (i) If $P(A) = 0$, this implies that $P(AB) = 0$.

   (ii) If $P(A) = 1$, then we have $P(A \cup B) = 1$; use this to establish that $P(AB) = P(B)$.

   Use the above results to prove that if $P(A) = 0$ or $P(A) = 1$, then the event A is independent of any other event B on the same sample space.

4. If for three independent events A, B, and C of a sample space, we know that

$$P(AB) = 0.3, \quad P(AC) = 0.48, \quad P(BC) = 0.1,$$

   find the probability of the event $A \cup B \cup C$.

5. If an event A is independent of an event B, and B is independent of another event C, is it always true that A and C are independent? Prove or disprove (using a counterexample in the latter case) this assertion.

6. A manufacturing item exhibits two types of faults, say α and β, which occur independently of one another. The probability of a fault, which is type-α, is 12%, while a type-β fault has a probability of 16% of occurring. Find the probability

   (i) that an item has both types of faults;

   (ii) that an item has at least one type of these faults;

   (iii) that the item has a fault which is of type-β, if we know that it also has a fault which is type-α.

7. A company sells laundry machines and refrigerators made by a certain manufacturer. It is known from this manufacturer that 5% of the laundry machines and 3% of the refrigerators need servicing before the end of the guarantee period. If someone buys today both a fridge and a laundry machine, what is the probability that

   (i) both need servicing before the end of their guarantee period?

   (ii) at least one of these appliances would require servicing before the end of their guarantee period?

   (iii) exactly one of these appliances would require servicing before the end of their guarantee period?

8. For two medical diseases a and b, it is known that the percentage of people in the general population who suffer from a only is 20%, the percentage of those who suffer from b only is 12%, while the percentage of people suffering from both diseases is 4%. Are the occurrences of the two diseases independent? If not, what is the percentage increase or decrease of the probability that someone suffers from a once it becomes known that he/she suffers also from b?

9. We select at random an integer between 1 and 100 (so that our sample space is $\Omega = \{1, 2, 3, \ldots, 100\}$). We consider the events A_1, A_2, and A_3 which consist of all integers divisible by 2, 5, and 7, respectively.

   (i) Show that the events A_1 and A_3 are independent, but the events A_2 and A_3 are not.

   (ii) If the sample space changes to the set $\Omega' = \{1, 2, 3, \ldots, 140\}$, then show that A_1, A_2, and A_3 are pairwise independent. Are they completely independent?

10. John has a red die and a blue die and he throws them together. Consider the events

    A: the outcome of the red die is an odd integer;

    B: the outcome of the blue die is an odd integer;

    C: the sum of the two outcomes is an odd integer.

    Prove that the events A, B, and C are pairwise independent. Are they completely independent?

11. John has again a red and blue dice and throws them once. For $i = 1, 2, 3$, let the events A_i be

    A_1: the outcome of the red die is 1, 2, or 5;

A_2: the outcome of the red die is 4, 5, or 6;

A_3: the sum of the two outcomes is equal to 9.

Prove that

$$P(A_1A_2A_3) = P(A_1)P(A_2)P(A_3)$$

while each of the pairs $\{A_1, A_2\}$, $\{A_1, A_3\}$, and $\{A_2, A_3\}$ consist of dependent events.

12. Suppose that A, B, and C are (completely) independent events. Show that each of the following pairs consists of independent events:

    (i) $\{A, B \cup C\}$;

    (ii) $\{A, B - C\}$.

13. We select at random an integer from the sample space $\Omega = \{1, 2, \dots, 20\}$. Consider the following three events in that sample space:

$$A = \{4, 5, \dots, 13\}, \quad B = \{9, 10, \dots, 18\},$$

and

$$C = \{4, 5, 6, 7, 8\} \cup \{14, 15, \dots, 18\}.$$

    (i) Show that the event A is independent from each of the events B and C.

    (ii) Prove that the events A and BC are *not* independent.

    (iii) Prove that the events A and $B \cup C$ are dependent.

14. Suppose that an event A of a sample space Ω is independent of B_i, $i = 1, 2, \dots$, where $\{B_i\}$ is a sequence of pairwise disjoint events of the same space. Show that the events A and $B = \bigcup_{i=1}^{\infty} B_i$ are independent.

15. When tossing a coin n times, let us consider the events

    A: heads appears at most once in the n tosses,

    B: both sides of the coin appear *at least once* in the n tosses.

    For $n = 2$ and $n = 3$, examine whether A and B are independent. Can you interpret the results?

16. Let A and B be two independent events on a sample space Ω. If $P(A) = 0.7$ and $P(A'B') = 0.18$, find the probabilities of the events

    (i) AB';

    (ii) $A'B$;

    (iii) $A \cup B$.

17. We throw a die six times. If at the ith throw, $1 \le i \le 6$, we get an outcome i, we say that there is a *concordance* (between the serial number of throws and the outcome of the die). What is the probability to have at least one concordance in the six throws?

18. An urn contains n balls numbered $1, 2, \ldots, n$. We select balls successively *with replacement*. If the ith ball selected from the urn has the number i on it, we say that we have a *concordance* (between the serial number of selections and the number on the ball). What is the probability to have at least one concordance in the first n selections? What is the limit of this probability as $n \to \infty$?

Group B

19. (Generalization of the Chevalier de Mére problem) In the experiment of throwing two dice n times, for $n \geq 2$, let us consider the event

$$B_n: \text{ a double six appears at least once in the } n \text{ throws.}$$

   (i) Using the result of Part (ii) of Example 3.18, calculate the probability of the event B_n.

   (ii) How many throws are needed so that the probability of B_n is at least a half?

20. Suppose that $A_1, A_2, \ldots, A_n$ are completely independent events of a sample space Ω with $P(A_i) = p_i$, with $p_i < 1$ for $i = 1, 2, \ldots, n$.

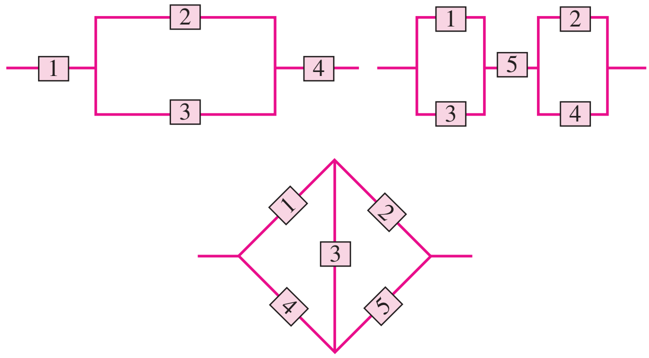
   (i) Verify that the probability that *exactly one* of the events $A_1, A_2, \ldots, A_n$ appear equals

$$(1 - p_1)(1 - p_2) \cdots (1 - p_n) \sum_{i=1}^{n} \frac{p_i}{1 - p_i}.$$

   (ii) If we throw two dice n times, what is the probability that a double four appears exactly once? What can we say about this probability as n grows larger $(n \to \infty)$?

21. An electrical system consists of two components that work independently of one another. We have estimated that the probability both components work properly is 0.756, while the probability that neither of them works is 0.016. For the proper functioning of the system, it is required that at least one of its components work properly. Find the reliability for each of the two components, as well as the overall reliability of the system (that is, the probability that the system works properly).

22. The following figures show some electrical circuits with switches on them (labeled $1, 2$, and so on). Every switch works independently of others and can be at the position OFF (allowing the transmission of electricity) with a probability p_i, $i = 1, 2, \ldots$, or at the position ON, in which case the electricity cannot pass through. If at the ends of each of the three circuits, we apply some voltage find, as a function of p_i, the probability that electricity is transmitted through the circuit.



Application: Find numerical results for the case $p_i = 0.7 + 0.05i$, for $i = 1, 2, 3, 4, 5$.

23. In the first circuit of the last exercise, assume that the probabilities p_i are all equal. What value should each p_i be so that the probability of electricity transmission through the circuit is 99%?

24. There are $n \geq 2$ shooters who are going to shoot against the same target, independently of one another. If the probability that the ith shooter finds the target is p_i, for $i = 1, 2, \ldots, n$, find the probability that

    (i) none of them find the target;

    (ii) exactly one finds the target;

    (iii) at least two find the target.

    (*Hint*: Use the results of Example 3.18 and Part (i) in Exercise 20.)

25. We throw two dice simultaneously until either a sum of 7 or a sum of 9 occurs. What is the probability that

    (i) exactly n throws will be needed?

    (ii) the experiment stops with the last throw having an outcome 9?

    (iii) the experiment stops with the last throw having an outcome 7?

    (iv) the experiment carries on forever?

26. (**The Huygens problem**) Tom and Daniel play a game in which they throw alternatively a pair of dice. Daniel wins if he gets a sum of 6 before Tom gets a sum of 7. What is the probability that Daniel wins if he plays first?

27. Three contestants in a quiz show have to answer, in turn, a series of questions. Let p_1, p_2, and p_3 be the probabilities that the first, second, and third player answer a question correctly. The game ends when the first correct answer is given. For $n = 1, 2, \ldots$, define the events

$$A_n: \text{ the first player wins the game at the } n\text{th round of questions.}$$

(a) Show that

$$P(A_n) = [(1 - p_1)(1 - p_2)(1 - p_3)]^{n-1} p_1;$$

(b) Find the probability that the first player wins the game;

(c) By finding the winning probabilities of the other two players, examine whether it is possible that the game goes on forever.

28. Bill, Greg, and John are friends attending the same course in French translation at University and they prepare for their exam. The course lecturer has recommended two textbooks, say A and B, to prepare for the exams, and the text they will have to translate will be either from textbook A or from B. Since none of them are too keen on spending a lot of time preparing for the exam, they decide to study only textbook A. The day before the exam they feel that if the exam is based on textbook A, each of them has a chance of 85% of passing the exam, independently of the other two, while if the exam is based on textbook B, each of them has a 30% chance of passing (again, independently of the other two). Further, they think the lecturer likes textbook A a bit more, and so there is a 60% chance that the exam is based on that book.

(a) What is the probability that Bill passes the exam, but the other two students fail?

(b) What is the probability that Bill passes the exam if we know that the other two students have failed?

(*Hint*: Let A_1, A_2, and A_3 be the events that Bill, Greg, and John pass the exam, respectively. Define also the events E that the exam is based on textbook A and F that it is based on textbook B. In Part (i), for instance, we want the probability of the event $A_1 A_2' A_3'$ and, in order to find this, we need to condition upon whether E or F occurs. In fact, this example illustrates the idea of **conditional independence** between events. **If we know** which of F or E has occurred, then the events A_1, A_2, A_3 are independent, but if we do not have that information, they are not independent.)

3.6 BASIC CONCEPTS AND FORMULAS

Conditional probability of the event A, given that the event B has occurred	$P(A\|B) = \dfrac{P(AB)}{P(B)}$, for $P(B) > 0$
Main properties of conditional probability	*PC1.* $P(A\|B) \geq 0$; *PC2.* $P(\Omega\|B) = 1$; *PC3.* $P\left(\bigcup_{i=1}^{\infty} A_i\|B\right) = \sum_{i=1}^{\infty} P(A_i\|B)$ for pairwise disjoint events $A_1, A_2, \ldots$.

Other properties of conditional probability	$P(\emptyset\|B) = 0$; $P(A'\|B) = 1 - P(A\|B)$; $P(A - C\|B) = P(AC'\|B) = P(A\|B) - P(AC\|B)$; If $C \subseteq A$ then $P(C\|B) \leq P(A\|B)$; $P(A \bigcup C\|B) = P(A\|B) + P(C\|B) - P(AC\|B)$; $\lim_{n\to\infty} P(A_n\|B) = P(\lim_{n\to\infty} A_n\|B)$ for any monotone sequence of events $\{A_n\}_{n\geq 1}$
Multiplicative law for probabilities	$P(A_1A_2\cdots A_n) =$ $P(A_1)P(A_2\|A_1)\cdots P(A_n\|A_1A_2\cdots A_{n-1})$, whenever $P(A_1A_2\cdots A_{n-1}) > 0$
Law of total probability	$P(A) = P(A\|B)P(B) + P(A\|B')P(B')$, for $0 < P(B) < 1$. More generally, $P(A) = \sum_{i=1}^{n} P(A\|B_i)P(B_i)$ for any partition $B_1, B_2, \ldots, B_n$ of Ω such that $P(B_i) > 0, i = 1, 2, \ldots, n$
Bayes' formula (Bayes theorem)	$P(B_i\|A) = \dfrac{P(A\|B_i)P(B_i)}{\sum_{j=1}^{n} P(A\|B_j)P(B_j)}, \quad i = 1, 2, \ldots, n,$ for any partition $B_1, B_2, \ldots, B_n$ of Ω such that $P(B_i) > 0, i = 1, 2, \ldots, n$
Independent events A and B	$P(AB) = P(A)P(B)$
Dependent events A and B	$P(AB) \neq P(A)P(B)$
Independence and complementary events	If A and B are independent, then each of the pairs $(A, B'), (A', B), (A', B')$ consists of independent events
(Completely) independent events $A_1, A_2, \ldots, A_n$	$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$ for any choice of k different indices $i_1, i_2, \ldots, i_k$ in the set of integers $\{1, 2, \ldots, n\}$ and for any $k = 2, 3, \ldots, n$
Probability that the event A occurs before the event B (in independent repetitions of an experiment)	$\dfrac{P(A)}{P(A) + P(B)}$, for disjoint events A and B

3.7 COMPUTATIONAL EXERCISES

1. Consider the following problem with conditional probabilities: we toss a coin four times and we seek the probability that we get exactly two heads, if we know that

   (i) the first outcome is heads;

   (ii) the first two outcomes are both heads.

With the set of commands below, we obtain the desired probabilities. The way to tackle this in Mathematica is as follows: first, we create the complete sample space for the problem, then we find the restricted sample space (as it is defined by the event we condition upon) and, finally, we find the elements of the event of interest in this restricted sample space. As usual, the required probability is the ratio of the number of favorable events to the total number of possible events.

```
In[1]:=
Print["Complete sample space"];
omeg=Flatten[Table[{i,j,k,l}, {i,0,1}, {j,0,1}, {k,0,1},
      {l,0,1}],3]
Print["Restricted sample space (first outcome heads)"];
omeg1=Flatten[Table[{i,j,k,l}, {i,1,1}, {j,0,1}, {k,0,1},
      {l,0,1}],3]
Print["Number of heads in each outcome of the restricted
      sample space"];
val1=Apply[Plus,omeg1,1]
Print["Restricted sample space (first two outcomes heads)"];
omeg11=Flatten[Table[{i,j,k,l}, {i,1,1}, {j,1,1}, {k,0,1},
      {l,0,1}],3]
Print["Number of heads in each outcome of the restricted
      sample space"];
val11=Apply[Plus,omeg11,1]
Print["Probability of event (i) :" ,Count[val1,2]/Length[omeg1]];
Print["Probability of event (ii) :" ,Count[val11,2]/Length[omeg11]];
```

The output we obtain is given below.

```
Complete sample space
Out[2]= {{0,0,0,0},{0,0,0,1},{0,0,1,0},{0,0,1,1},{0,1,0,0},
         {0,1,0,1},{0,1,1,0},{0,1,1,1},{1,0,0,0},{1,0,0,1},
         {1,0,1,0},{1,0,1,1},{1,1,0,0},{1,1,0,1},{1,1,1,0},
         {1,1,1,1}}

Restricted sample space (first outcome heads)
Out[3]= {{1,0,0,0},{1,0,0,1},{1,0,1,0},{1,0,1,1},{1,1,0,0},
        {1,1,0,1},{1,1,1,0},{1,1,1,1}}

Number of heads in each outcome of the restricted sample space
Out[4]= {1,2,2,3,2,3,3,4}

Restricted sample space (first two outcomes heads)
Out[5]= {{1,1,0,0},{1,1,0,1},{1,1,1,0},{1,1,1,1}}

Number of heads in each outcome of the restricted sample space
Out[6]= {2,3,3,4}

Probability of event (i) :3/8
Probability of event (ii) :1/4
```

Proceeding in an analogous manner, obtain the conditional probabilities $P(A|B_i)$ and $P(B_i|A)$ in each of the following problems:

(a) We throw a die twice, and consider the events

A: the sum of the two outcomes is 4,

B_1: the outcomes of the two throws are equal,

B_2: the outcome of the first throw is an even integer,

B_3: the outcome of the second throw is an odd integer,

B_4: the first outcome of the die is different from the second one.

(b) From a pack of 52 cards, we select two cards at random, and define the events

A: both cards are aces,

B_1: the first card drawn is a diamond, and the second one is an ace,

B_2: the first card drawn is a spade,

B_3: both cards drawn are spades,

B_4: the two cards are not of the same suit.

(c) From a box with 10 balls numbered from 1 to 10, we select two balls without replacement, and let

A: the numbers on the two balls drawn differ by at least 2 (in absolute value),

B_1: the numbers on the two balls drawn have a sum of squares greater than 10,

B_2: both numbers selected are even,

B_3: the numbers on the two balls have a sum less than 10.

2. With reference to Exercise 22 of Section 3.5, suppose that the probabilities $p_i, i = 1, 2, \dots,$ are all equal, that is,

$$p_1 = p_2 = \cdots = p.$$

Let $R_1(p), R_2(p)$, and $R_3(p)$ be the probabilities that electricity passes through the circuits (a), (b), and (c) of that exercise, respectively.

(i) Draw a graph of $R_j(p), j = 1, 2, 3$, for $0 \le p \le 1$. What do you observe for $R_j(0)$ and $R_j(1)$ for $j = 1, 2, 3$?

(ii) For each $j = 1, 2, 3$, find the smallest value of p for which

(a) $R_j(p) \ge 0.90$;

(b) $R_j(p) \ge 0.95$;

(c) $R_j(p) \ge 0.99$.

3. Assume that $n \ge 2$ shooters shoot against the same target, independently of each another. The probability that each shooter hits the target is the same for all shooters and is equal to p. Let $R(p, n)$ be the probability that at least two shooters hit the target.

(a) Verify that

$$R(p, n) = 1 - (1 - p)^n - np(1 - p)^{n-1}.$$

(b) Draw, on the same graph, the function $R(p, n)$ as p varies inside the interval $[0, 1]$ and for different values of n. What do you observe?

(c) For $n = 5, 10$, and 15, find the smallest value of p for which the probability $R(p, n)$ is greater than

- 0.90;
- 0.95;
- 0.99.

3.8 SELF-ASSESSMENT EXERCISES

3.8.1 True–False Questions

1. Let A and B be two events on a sample space Ω such that $A \subset B$, with $P(B) < 1$. Then $P(A|B') = 0$.

2. If A and B are two events on a sample space Ω with $P(B) < 1$, then we have

$$P(A'|B') = 1 - P(A|B').$$

3. If A and B are two events on a sample space such that $P(A) \neq 0$, $P(B) \neq 0$, then we have

$$P(A|B)P(B|A) = 1.$$

4. A, B, and C are three events on a sample space Ω. If $P(AB) \neq 0$, then

$$P(ABC) = P(B)P(A|B)P(C|AB).$$

5. If A and B are two mutually exclusive events such that $P(A) = P(B) = 0.3$, then A and B are independent.

6. Suppose that A_1, A_2, and B are three events on a sample space and $P(B) > 0$. If A_1 and A_2 are independent, then

$$P(A_1A_2|B) = P(A_2|B)P(A_2|B).$$

7. The probability for the intersection of two events equals the product of the probabilities for each event.

8. For the events A and B of a sample space Ω, it is known that $P(A) \neq 0$ and $P(B) \neq 0$. If A and B are independent, then they are also disjoint (mutually exclusive).

9. If A and B are independent events, then the events A' and B are also independent.

10. If A and B are mutually exclusive events such that $P(A) > 0$ and $P(B) > 0$, then A and B cannot be independent.

11. Let A_1 and A_2 be two disjoint events on a sample space and assume that B is another event with $P(B) > 0$. Then, we have

$$P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B).$$

12. Assume that, for the events A and B of a sample space Ω, we have $0 < P(A) < 1$ and $P(B'|A) = P(B'|A')$. Then A and B are independent events.

13. Suppose that $\{B_1, B_2\}$ is a partition of a sample space Ω such that $P(B_i) > 0$, for $i = 1, 2$. Then, for any event A in Ω, we have

$$P(A) = P(A|B_1')P(B_1') + P(A|B_2')P(B_2').$$

14. For two independent events A and B on a sample space Ω, we have

$$P(A'B') = P(A') + P(B') - 1.$$

15. Suppose that A_1, A_2, A_3, and B are events on a sample space with $P(B) > 0$. If the events A_1, A_2, and A_3 are pairwise disjoint, then

$$P(A_1 \cup A_2 \cup A_3|B) = P(A_1|B) + P(A_2|B) + P(A_3|B).$$

16. When throwing a die twice, let A be the event that the first outcome is even and B be the event that the product of the two outcomes is 6. Then, A and B are independent.

17. Let A, B, and C be three events on a sample space such that $P(B) > 0$. Then,

$$P(AC'|B) = P(A|B) - P(AC|B).$$

18. Let $A_1, A_2, \ldots, A_n$ be completely independent events on a sample space Ω with $P(A_i) = p_i$, for $i = 1, 2, \ldots, n$. The probability that none of the events A_i occur is equal to

$$\prod_{i=1}^{n}(1 - p_i).$$

3.8.2 Multiple Choice Questions

1. An exam contains two multiple choice questions. The first question has four possible answers, while the second one has five. If a student taking the exam answers both questions completely at random, the probability that he/she gives at least one wrong answer is

   (a) $12/20$ (b) $19/20$ (c) $1/15$ (d) $1/20$ (e) $14/15$

2. Let A and B be two events of a sample space such that $P(A) = 0.5$ and $P(A \cup B) = 0.8$. If it is known that $P(A|B) = 0.25$, then the probability of the event B is

   (a) 0.9 (b) 0.3 (c) 0.2 (d) 0.4 (e) 0.1

3. If the events A and B are independent, each having a positive probability of occurrence, then

   (a) $AB = \emptyset$ (b) $P(B'|A) = P(B)$ (c) $P(A \cup B) = P(A) + P(B)$
   (d) $P(A|B) = P(B|A)$ (e) $P(A - B) = (1 - P(B))P(A)$

4. Assume that A_1, A_2, and B are three events on a sample space such that $P(B) > 0$. Then,

   (a) $P(A_1|B) = P(A_1A_2 \cup A_1A_2'|B)$
   (b) $P(A_1 - A_2|B) = P(A_1'A_2'|B)$
   (c) $P(A_1A_2|B) = 1 - P(A_1'A_2'|B)$
   (d) $P(A_1A_2|B) = 1 - P(A_1' \cup A_2'|B)$
   (e) $P(A_1' \cup A_2'|B) = P(A_1'|B) + P(A_2'|B)$

5. Let A and B be two independent events on a sample space Ω. If $P(A) = 3/4$ and $P(AB') = 21/40$, then the probability of event B is

   (a) $3/10$ (b) $7/10$ (c) $2/5$ (d) $3/5$ (e) $9/40$

6. If A and B are independent events with $P(A) = 1/5$ and $P(B) = 1/2$, then the conditional probability $P(A|A \cup B)$ is equal to

   (a) $4/10$ (b) $3/10$ (c) $1/2$ (d) $1/10$ (e) $1/3$

7. For the events A and B of a sample space it is known that $P(B) = 0.5$ and $P(A'B) = 0.2$. Then, the conditional probability $P(A|B)$ equals

   (a) 0.6 (b) 0.3 (c) 0.5 (d) 0.1 (e) 0.15

8. From a box that contains 10 balls numbered $1, 2, \dots, 10$, we select randomly 2 balls with replacement. For the events

   $$A\colon \text{ the number on the first ball selected is odd}$$

   and

   $$B\colon \text{ the number on each of the two balls selected is even,}$$

   which of the following statements is true?

   (a) They have the same probability;
   (b) They are independent;
   (c) They are mutually exclusive;
   (d) They form a partition of the sample space;
   (e) They are complementary.

9. John is going to take two exams for his degree course next week. The first one is more difficult and he feels that the probability he will pass this exam is $3/4$, while he is more confident about the second one giving a success probability to himself of $9/10$. The probability that John passes at least one of these two exams is

   (a) $4/5$ (b) $39/40$ (c) $19/20$ (d) 1 (e) none of the above

10. With the assumptions of the previous question, the probability that John passes both exams if he knows that he passed at least one is

    (a) $27/40$ (b) $39/40$ (c) $19/20$ (d) $9/13$ (e) $13/40$

11. We select (without replacement) two cards from a pack of 52 cards. The probability that both cards selected are spades is

    (a) $\dfrac{1}{51 \cdot 52}$ (b) $\dfrac{13}{51 \cdot 52}$ (c) $\dfrac{12 \cdot 13}{51 \cdot 52}$ (d) $\dfrac{13^2}{51 \cdot 52}$ (e) $\dfrac{2 \cdot 12 \cdot 13}{51 \cdot 52}$

12. We select (without replacement) two cards from a pack of 52 cards. The probability that the two cards selected belong to different suits is

    (a) $\dfrac{3 \cdot 4}{51 \cdot 52}$ (b) $\dfrac{3}{4}$ (c) $\dfrac{39}{51}$ (d) $\dfrac{3}{51}$ (e) $\dfrac{12}{51}$

13. An urn contains six red balls and five blue balls. We select one ball at random and then, without replacing it, we select another ball. Then, the probability that the **second ball drawn** is blue is

    (a) $\dfrac{5}{11}$ (b) $\dfrac{4}{11}$ (c) $\dfrac{2 \cdot 6 \cdot 5}{\binom{11}{2}}$ (d) $\dfrac{1}{2}$ (e) $\dfrac{5}{\binom{11}{2}}$

14. At a sports club of a University, there are 15 students who play basketball and 13 students who play baseball. We choose successively two students. The probability that the first student selected plays baseball and the second one plays basketball equals

    (a) $\dfrac{\binom{13}{1}\binom{15}{1}}{\binom{28}{2}}$ (b) $\dfrac{13}{28} \cdot \dfrac{15}{28}$ (c) $\dfrac{13+15}{\binom{28}{2}}$ (d) $\dfrac{2(13+15)}{\binom{28}{2}}$ (e) $\dfrac{13}{28} \cdot \dfrac{15}{27}$

15. In a hand of bridge, each of the four players receives 13 cards from a standard pack of 52 cards. Nick and Leopold are partners. What is the probability that, in a given hand, Nick has at most one spade given that Leopold has exactly three spades?

    (a) $\dfrac{\binom{39}{10} + \binom{39}{9}\binom{13}{1}}{\binom{52}{13}}$

    (b) $\dfrac{\binom{29}{13} + \binom{29}{12}\binom{10}{1}}{\binom{39}{13}}$

    (c) $\dfrac{\binom{26}{10} + \binom{26}{9}\binom{13}{1}}{\binom{39}{10}}$

    (d) $\dfrac{\binom{26}{10} + \binom{26}{9}\binom{13}{1}}{\binom{39}{13}}$

    (e) $\dfrac{\binom{49}{13} + \binom{48}{12}\binom{10}{1}}{\binom{52}{13}}$

16. Pat takes part in a quiz show with multiple choice questions. There are three possible answers to each question. The probability that she knows the answer to a question is p. If Pat does not know the answer to a particular question, she gives an

answer at random. If she answered the first question correctly, the probability that she knew the answer is

(a) $\dfrac{3p}{1+2p}$ (b) p (c) $\dfrac{p}{p+1/3}$ (d) $\dfrac{p}{p+1/2}$ (e) $\dfrac{3p}{2}$

17. In a telecommunications channel, transmitted bits are either 0 or 1 with respective probabilities $5/8$ and $3/8$. Due to noise, a 0-transmitted bit is received as 1 with probability $1/5$, while a 1-transmitted bit is received as 0 with probability $1/10$.

    If the last bit received was 0, the probability that it has been received correctly equals

    (a) $43/80$ (b) $4/5$ (c) $40/43$ (d) $5/8$ (e) $7/8$

3.9 REVIEW PROBLEMS

1. Let A and B be two events on a sample space such that

$$P(A) = 0.5, \quad P(A \cup B) = 0.75.$$

   Find the probability of the event B in each of the following cases:

   (i) A and B are disjoint events;

   (ii) A and B are independent events;

   (iii) We have $P(A|B) = 0.3$.

2. Let A and B be two disjoint events of the same sample space with $P(A) \neq 0$ and $P(B) \neq 0$. Show that

$$\frac{P(A|A \cup B)}{P(A)} = \frac{P(B|A \cup B)}{P(B)} = \frac{1}{P(A) + P(B)}.$$

3. For the events A and B of a sample space, it is given that

$$P(A|B) = \alpha, \quad P(B|A) = \beta, \quad P(A|B') = \gamma.$$

   Find the unconditional probability of event A in terms of α, β, and γ.

4. The percentage of the population in a city who suffer from a serious disease is 3%. A person has just taken two medical examinations to see whether he suffers from the disease. Each of the two tests make the correct diagnosis (whether someone suffers from the disease or not) with a probability 98% and the results of the two tests are independent. Calculate the probability that he has the disease

   (i) given that at least one of the two outcomes of the tests is positive;

   (ii) given that both tests give a positive result.

5. Peter throws three dice and he announces to Mary that the three outcomes are different. What is the probability that at least one outcome is 4?

6. A box contains n lottery tickets, k of which are winning tickets. Frank and Geoffrey select tickets from the box successively, one after the other and without replacement. Whoever picks a winning ticket first wins the game. Who has the higher probability of winning, Frank who chooses first or Geoffrey?

7. A store that sells alcoholic drinks has n bottles of white wine and k bottles of red wine on its shelves. We select a bottle randomly from the shelves and then a second, without replacing the first. What is the probability that the second bottle contains white wine if the first one was also a white wine bottle? What is the probability that the first bottle contains white wine if we know that the second one was also a white wine bottle? Are these two probabilities equal?

8. Among the students who enter a University degree program, 82% of the females and 71% of the males obtain their degree within the scheduled time. If 52% of the students in this program are males, what is the percentage of students who finish in time? What proportion among those who finish in time are women?

9. A military aircraft that carries n bombs shoots against a target until the target is destroyed or the aircraft runs out of bombs. The probability a bomb finds the target is p_1, while the probability that a bomb destroys the target when it has already been hit by a bomb is p_2. If successive bomb shots can be considered as independent "experiments," find the probability that

   (i) the aircraft destroys its target without using all the bombs;

   (ii) when the target is destroyed, there are two unused bombs;

   (iii) For the destruction of the target, at most two bombs are needed.

10. Peter throws 10 dice, and he announces to Mary that at least one six appeared. What is the probability that at least two sixes were observed in total?

11. A driver responsible for a car accident disappears after the accident. After investigating the case, police believe with 70% probability that the accident was caused by a car with plate number XYZ 1867. At this point, an eye witness appears and tells the police that he is certain that the last digit in the number plate of the car was 7, but he did not see or remember any other digits or letters from the plate. Using this piece of information, what is now the probability that the driver of the car with plate number XYZ 1867 caused the accident?

    Suppose alternatively the eye witness says he is only 90% certain that the last digit of the plate is a 7; what would be the probability that the accident was caused by a car with plate number XYZ 1867 in this case?

    In both cases, assume that every combination of three letters and four digits has the same probability of appearing in a plate number.

    (*Hint*: Use Bayes' theorem.)

12. If Steve tosses a coin n times, for $n \geq 2$, consider the events

A: tails appear at most once in the n tosses,

B: each of the two sides of the coin appears at least once.

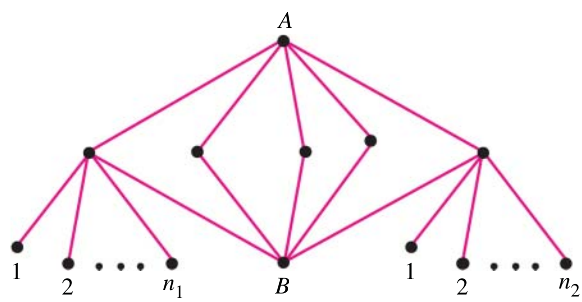
(i) Find the probabilities of the events A, B, and AB;

(ii) Hence, show that A and B are independent only when $n = 3$.

13. Jenny starts from point A of the diagram below and, at each node, she selects her route with equal probabilities among the possible options.

(i) What is the probability that she arrives at point B?

(ii) If it is known that she reached point B, what is the probability that she came through one of the three middle nodes?



14. We have n boxes and for $i = 1, 2, \dots, n$, the ith box contains n_i balls of which r_i are white, and the remaining are black. We first select a box and then choose randomly one of the balls in that box. What is the probability that a white ball is selected if

(i) the selection of the box is completely random?

(ii) the probability to select the ith box is p_i, for $i = 1, 2, \dots, n$, where $p_i > 0$ and $p_1 + p_2 + \cdots + p_n = 1$?

15. We assume that the probability of a family having n children is ap_n for $n = 1, 2, \dots$ (where a is a positive real number such that $ap_n < 1$ for all n), and the probability that a family has no children is

$$1 - \sum_{n=1}^{\infty} ap_n.$$

We also assume that boys and girls are equally likely to be born.

(i) Show that the probability a family with n children to have exactly k girls is

$$\binom{n}{k} \frac{1}{2^n};$$

(ii) Establish that the probability a family has k girls, for $k \geq 1$, is $2ap^k/(2-p)^{k+1}$;

(iii) Suppose it is known that a certain family has at least one girl. What is the probability that this family has two or more girls?

$\Bigg($*Hint*: For Part (ii), you may use the identity

$$\sum_{n=k}^{\infty} \binom{n}{k} t^n = \frac{t^k}{(1-t)^{k+1}}. \Bigg)$$

16. Three darts players have respective probabilities p_1, p_2, p_3 of hitting the center of the dartboard. Each of them shoots against the target once and then we examine how many of them hit the center. Find the probability that the third player found the center if the number of darts found there is (i) one; (ii) two.

    Give a numerical answer in each case assuming that $p_1 = 0.3, p_2 = 0.4$, and $p_3 = 0.5$.

17. Two urns labeled I and II contain n_1 and n_2 balls, of which r_1 and r_2, respectively, are green. We select a ball at random from Urn I and place it in the second urn, and then we take a ball from Urn II. What is the probability that this ball is green?

18. Three boxes contain b blue and r red chips each. We select a chip at random from the first box and place it in the second box. Then we select a chip from the second box and place it in the third. If we now select at random a chip from the third box, what is the probability that this is a red one?

19. For a certain make of a car, spare parts are produced by two manufacturers, A and B. The total production of the manufacturer B is n times as large as that of A. The proportions of defective spare parts produced by the manufacturers A and B are p_1 and p_2, respectively.

    (i) If we purchase a spare part and find out that it is defective, what is the probability that it came from A?

    (ii) What relation must hold between p_1 and p_2 so that the probability a defective item would belong to either of the manufacturers A and B is the same?

20. An urn contains $a > 3$ white and $k > 1$ red balls. Suddenly, a ball disappears from the urn. We select two balls from the urn and check their color, to find out that they are both white. What is the probability that the missing ball is red?

21. A student has to take a multiple-choice exam with n possible answers in each question (only one being correct in each case). The probability that the student knows the correct answer to a question is p $(0 < p < 1)$. If the student does not know the answer to a particular question, he chooses completely at random among the n possible options.

    (i) What is the proportion of correct answers given by the student in the exam?

    (ii) If the student gives the correct answer to a particular question, what is the probability that he knew the answer?

(iii) The course lecturer wants to ensure that a student who has studied carefully with p at least 80%, when he/she answers a question correctly, the probability that he/she knew the answer is at least 95%. Find the smallest number n of possible answers the lecturer should assign for each question.

22. Two boxes contain exactly the same total number of balls, and in each box some balls are white and the rest are black. Let x and y be the number of white and black balls in the first box, respectively, and let z be the number of white balls in the second box. From each box, we pick up n balls with replacement. Find the conditions under which the following two quantities are equal:

    (i) The probability all balls selected from the second ball are white.

    (ii) The probability that the balls chosen from the first box are all of the same color, i.e. they are all white or all black.

    Show that finding positive integers n, x, y, z such that the condition for the above two probabilities to be equal amounts to finding positive integers that satisfy the equation $x^n + y^n = z^n$ (the solution to this problem is very well-known[4] as "**Fermat's last theorem**."

23. A company classifies its customers into k classes according to the frequency by which they place their orders. The probability that a customer of class j makes no orders in a period of one month is $(n-j)/n$, for $j = 1, 2, \ldots, k$, where $1 \leq k < n$. The proportion of customers who are classified into class j is p_j, so that $p_1 + p_2 + \cdots + p_k = 1$.

    If a certain customer of the company has placed at least one order during the last month, find the probability that he belongs to class j, for $j = 1, 2, \ldots, k$.

    Application: Consider a motor insurance company which classifies car drivers into $k = 5$ classes. Suppose the proportion of insured car drivers that fall into class j is $p_j = j/15$, for $j = 1, 2, \ldots, 5$, and assume further that a driver that belongs to class j has a probability $1 - j/100$ not to make a claim in a given time period. Find the probability that a driver who makes at least one claim during that period belongs to class j, for $j = 1, 2, \ldots, 5$.

24. If we throw a die 6 times, what is the probability that the number of sixes minus the numbers of ones we get is equal to 3?

25. We carry out the following experiment successively: we toss a coin and throw two dice simultaneously. What is the probability that an outcome of tails occurs before we get a sum of 3 in a throw of the two dice?

26. A primary school teacher asks four children what is their favorite season of the year. If the first two children gave different answers (e.g. summer and autumn), show that the probability that a season is chosen by exactly three children is $1/8$.

[4]This is one of the most famous problems of all times in mathematics. For $n = 2$, this is obviously the Pythagorean theorem that has infinitely many integer solutions. For $n \geq 3$, however, establishing whether integer solutions x, y, z exist presented unimaginable difficulties when Fermat originally posed the problem and said that there is no solution in 1637, but gave no proof(!). Numerous great mathematicians attempted to provide a formal proof since then until this was accomplished by Andrew Wiles in 1995.

27. Recall the definition of a parallel system connection from Example 3.18. Consider now a parallel system that has n components and assume that each of these components works, independently of others, with probability $1/2$. Find the probability that the first component works given that the system is functioning.

28. Isaac and Mary enter a quiz show and reach the final stage in which they will win the grand prize if they answer a true–false question correctly. Assume that each of them has a probability p of giving the correct answer. Before the question is posed, Isaac tells Mary that they should choose one person between them and let that person give the answer to the question. Mary, on the contrary, thinks that they should both think about the question and, if their answers agree, they should give that common answer; otherwise, they should flip a coin to see who answers the question. Which of the two strategies do you think maximizes their probability of winning the prize?

29. Two groups of children decide to make a draw so that a three-member committee is selected. The first group consists of 7 boys and 3 girls, while the second group has 1 boy and 5 girls. The selection takes place as follows: first, one person is chosen randomly from each group. In order to select the third person, the names of all the remaining children (from both groups) are put in a ballot and one name is picked at random. Find the probability that the person selected in this second draw is a girl.

30. An amateur meteorologist uses the following, rather primitive, weather forecasting system. Each day is classified as dry or wet and he then assumes that the following day will be of the same type (dry or wet, resp.) with probability p $(0 < p < 1)$. Using statistical data from previous years, he decides that the probability December 31st of a particular year to be a dry day is p_0 for some $p_0 \in (0, 1)$. Let A_n be the event that the nth day of the following year is a dry day and

$$\alpha_n = P(A_n), \quad n = 0, 1, 2, \ldots$$

(i) Verify that for $n = 1, 2, \ldots$, we have

$$\alpha_n = (2p - 1)\alpha_{n-1} + (1 - p);$$

(ii) Deduce that the probabilities α_n are given by the formula

$$\alpha_n = \frac{1}{2} + (2p - 1)^n \left(p_0 - \frac{1}{2}\right), \quad n = 0, 1, 2, \ldots;$$

(iii) Calculate the probability that the nth day of the year is dry when n is very large $(n \to \infty)$.

(*Hint*: Use induction for Part (ii).)

31. A coin that does not have equal probabilities of landing heads and tails is said to be a *biased coin*. Suppose that we have two biased coins; the probability of landing heads for the first of these (coin A) is p_1, while for the second one (coin B) it is p_2. We select initially one of these two coins at random and we toss it until it lands

"tails" for the first time. Then we start tossing the other coin, and we switch again when it lands "tails" for the first time, and so on.

(i) Find the probability that, in the third toss from the beginning of the experiment, coin A is used.

(ii) If $p_2 = 1 - p_1$, what can we say for the probability that coin A is used in the nth toss of this experiment, for $n = 3, 4, \ldots$?

32. A certain University exam involves True–False questions. Students who take this exam can be classified into three classes: skilled, who answer each question correctly with probability 0.95, well-prepared, who answer each question correctly with probability 0.75, and guessers who simply take a stab at the answers. One of the students taking the exam has answered the first three questions correctly. What is the probability that she will give a correct answer to the fourth one, too?

3.10 APPLICATIONS

3.10.1 Diagnostic and Screening Tests

A diagnostic test is any kind of medical test that is utilized in the diagnosis or detection of a specific disease. On the other hand, screening tests are typically carried out in a population to identify a disease among individuals without signs or symptoms. Despite the fact that screening tests may lead to an early diagnosis, their outcomes should be treated very carefully, especially when carried out for checking the presence of rare diseases.

Let us assume that we are screening people for a specific disease, say tuberculosis, using a test which is quite accurate, in the sense that it will give a positive result with probability $p_1 = 0.95$ when the disease is present and a negative result with probability $p_2 = 0.99$ when tuberculosis is not present. If a person tests positive, what is the probability that he/she has tuberculosis? The naive answer that is usually given by (a strikingly high proportion of) medical students is 95%. As we shall shortly explain by a careful probabilistic analysis, this answer is wrong!

Let us first introduce the necessary notation that will enable us to arrive at the required answer. Denote by B the event that a randomly selected person from the population suffers from tuberculosis and A the event that a person who is subject to the screening test will give a positive result. Then, the aforementioned probabilities p_1 and p_2 are in fact referring to the conditional events $A|B$ and $A'|B'$, respectively. Therefore, we may write

$$P(A|B) = p_1, \quad P(A'|B) = 1 - p_1, \quad P(A'|B') = p_2, \quad P(A|B') = 1 - p_2.$$

In order to answer the question whether a person that tests positive suffers from tuberculosis or not, we need to compute the probability $P(B|A)$ that is clearly different from both $P(A|B)$ and $P(B)$.

A straightforward application of Bayes' theorem (see formula (3.7)) yields

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + (P(A|B')P(B')} = \frac{p_1 P(B)}{p_1 P(B) + (1 - p_2)(1 - P(B))},$$

and it is clear that our answer to the question addressed above depends not only on the probabilities p_1 and p_2 but also on the percentage $r = P(B)$ of the people in the population who suffer from the disease. For example, if we assume that $r = 0.1\%$, we get

$$P(B|A) = \frac{(0.95)(0.001)}{(0.95)(0.001) + (0.01)(0.999)} = 0.0868,$$

while for $r = 10\%$ (which, of course, may be unrealistic for the specific disease we are referring to), the required conditional probability becomes

$$P(B|A) = \frac{(0.95)(0.1)}{(0.95)(0.1) + (0.01)(0.9)} = 0.913.$$

It is noteworthy that, despite the fact that the test used is very accurate (it guarantees a correct diagnosis for the presence of the disease with probability 95% and for the absence of it 99%), a person that is tested positive has a probability of only 8.7% to actually suffer from the disease if the disease is very rare ($r = 0.1\%$).

The above remarks can be illustrated graphically by creating some plots of the conditional probability of interest, $P(A|B)$, in terms of the parameters p_1, p_2, and r. To make our task easier, let us assume that our test has the same accuracy for both cases of correctly identifying the presence and nonpresence of the disease, i.e. $p_1 = p_2 = p$. Then, the quantity of interest becomes

$$P(A|B) = \frac{pr}{pr + (1-p)(1-r)}.$$

Figure 3.6 shows how the conditional probability varies with respect to r when $p = 0.95$. Evidently, this probability decreases rapidly as r gets close to 0 (which means that the

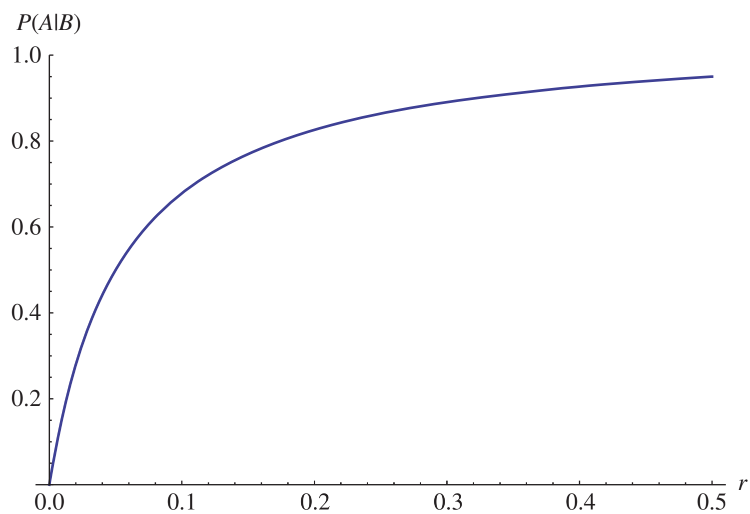


Figure 3.6 Plot of the probability $P(A|B)$ as a function of r for $p = 0.95$.

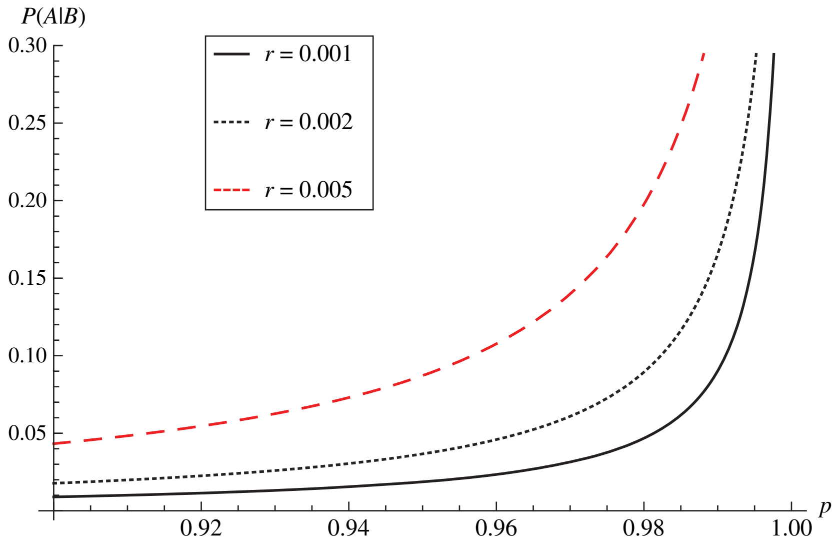


Figure 3.7 Plot of the probability $P(A|B)$ as a function of p for several choices of r.

screening test is carried out for a very rare disease). Figure 3.7 plots the same probability as a function of p for several choices of r. From this figure, it is also clear how sensitive the quantity we are looking at (correct diagnosis of a disease) is to the incidence rate of the disease (percentage of the people in the population suffering from it).

KEY TERMS

Bayes' theorem
conditional probability
dependent events
independent events
independent experiments
law of total probability
multiplication formula (or multiplication law)
pairwise independent events
parallel system
partition of a set
posterior probability (or a posteriori probability)
prior probability (or a priori probability)
reliability of a system
sampling with replacement
sampling without replacement
series system

CHAPTER 4

DISCRETE RANDOM VARIABLES AND DISTRIBUTIONS

Andrey Markov
(Ryazan 1856–St. Petersburg 1922)

A Russian mathematician best known for his work on stochastic processes (families of random variables, typically evolving over time). In particular, he introduced a family of stochastic processes exhibiting a certain "memoryless" property, nowadays known as the Markov property.

One of the results that bears his name is Markov's inequality, although it was apparently first proved by P. Chebychev, who was Markov's teacher. An example of an application of Markov's inequality is the fact that (assuming incomes are nonnegative) no more than one-fifth of the population can have more than five times the average income of the population.

In the March to April 2013 issue of the *American Scientist* (Volume 101, Number 2), Brian Hayes presented an article entitled "First links in the Markov Chain: Probability and Poetry were Unlikely Partners in the Creation of a Computational Tool," which starts as follows:

One hundred years ago the Russian mathematician A. A. Markov founded a new branch of probability theory by applying mathematics to poetry. Delving into the text of Alexander Pushkin's novel in verse Eugene Onegin, Markov spent hours sifting through patterns of vowels and consonants. On January 23, 1913, he summarized his findings in an address to the Imperial Academy of Sciences in St. Petersburg. His analysis did not alter the understanding or appreciation of Pushkin's poem, but the technique he developed, now known as a Markov chain, extended the theory of probability in a new direction. Markov's methodology went beyond coin-flipping and dice-rolling situations (where each event is independent of all others) to chains of linked events (where what happens next depends on the current state of the system).

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

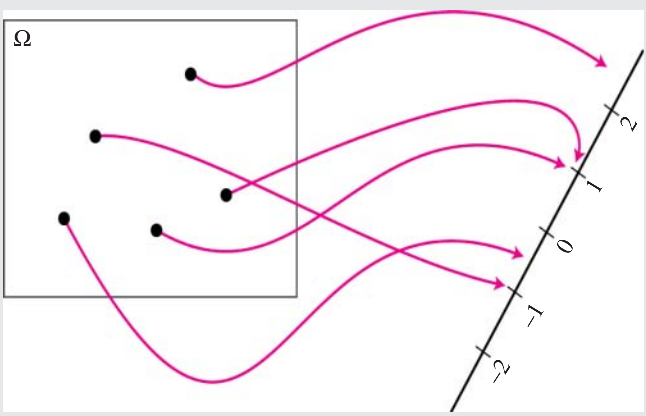
4.1 RANDOM VARIABLES

In the discussions until now, we have assigned probabilities to events. An event is a set that represents a possible realization of an experiment and, as we have seen, assigning a probability to any such realization offers a suitable formulation of the problem. Furthermore, it enables us, in many situations, to answer almost any conceivable question about how likely is something to happen in relation to that particular experiment. However, although there is a number of operations admissible for sets (unions, intersections, taking the complement, etc.), which offer the opportunity to define and study complex outcomes in terms of simpler ones, mathematically it is less convenient to deal with sets rather than real numbers. It therefore seems highly fruitful to consider a concept that obeys the usual operations of arithmetic and use this concept to express probabilities of future outcomes. Moreover, when we pose a question to express our uncertainty about the unknown outcome of an experiment, the answer to that question is typically a real number. For instance,

- How many children a couple will have?
- What is the annual salary of a school teacher in the United States?
- How many points will a certain basketball player score in his next game?
- How many burglaries occur during a day in the city of London?
- How much does Jim weigh? What is the weight of Nick? What is the combined weight between the two of them?

The concept of a *random variable*, which is given next and is fundamental in probability theory, offers a straightforward formulation to any of the questions above. It is also a more natural tool compared to the use of sets that we have favored in the preceding chapters. Further, this concept provides the passage from the sample space of an experiment to the set of real numbers.

Definition 4.1 Suppose Ω is the sample space of an experiment. A function that takes real values, $X : \Omega \mapsto \mathbb{R}$, is called a **random variable**, or simply a variable (for this experiment) if for any interval $I \subset \mathbb{R}$, the set $\{\omega \in \Omega : X(\omega) \in I\}$ is an event of the sample space Ω. The probability that X takes values on I will be simply denoted by $P(X \in I)$.



We note that, when the sample space for an experiment is discrete (i.e. it is either finite or countably infinite), then every subset of Ω is an event and therefore any function $X : \Omega \mapsto \mathbb{R}$ is a random variable. Even in continuous (uncountably infinite) sample spaces, however, discussed later in Chapter 6, in the problems and all applications we shall consider the technical difficulty mentioned in the end of Section 1.4 does not show up. As a consequence, from now on, we shall freely use the term *random variable* for a function that maps a sample space to the real line.

When dealing with probabilities associated with random variables, it is often useful to have the following representation in mind:

$$\omega \xrightarrow{X} X(\omega) \xrightarrow{P} P(X(\omega)).$$

Example 4.1 (Number of tails in three tosses of a coin)
Suppose we toss a coin three times and let X be the number of times that tails appear. The sample space for this experiment is

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

while for the function $X : \Omega \mapsto \mathbb{R}$ we can write

$$X(HHH) = 0, X(HHT) = 1, X(HTH) = 1, \dots, X(TTT) = 3.$$

In a tabular form, the mapping of the elementary events ω to $X(\omega)$ is summarized below:

$\omega \in \Omega$	HHH	HHT	HTH	HTT	THH	THT	TTH	TTT
$X(\omega)$	0	1	1	2	1	2	2	3

Further, by taking into account that the eight elements of Ω are equiprobable, we get

$$P(X = 0) = P(\{\omega \in \Omega : X(\omega) = 0\}) = P(\{HHH\}) = \frac{1}{8},$$

$$P(X = 1) = P(\{\omega \in \Omega : X(\omega) = 1\}) = P(\{HHT, HTH, THH\}) = \frac{3}{8},$$

and so on.

Example 4.2 Consider a random experiment when someone is waiting for a taxi in the street. Let X be the random variable that stands for the number of occupied (O) taxis passing by before the first free (F) one arrives. Since the experiment stops with the appearance of the first free taxi, the sample space for this experiment is

$$\{F, OF, OOF, OOOF, \dots\}$$

while for the random variable X defined above, we can write

$$X(F) = 0, \quad X(OF) = 1, \quad X(OOF) = 2, \quad X(OOOF) = 3, \quad \dots,$$

Notice that, in contrast to the previous example, here we have a one-to-one correspondence between the sample space elements $\omega \in \Omega$ and the values of the random variable X.

Example 4.3 (Random numbers in an interval)
John asks his friend Matt to tell him a (real) number in the interval $[0, 10]$. Without telling him, John has decided that if the number that Matt selects is x, he would give him $[x]$ dollars, where $[x]$ denotes the *integer part* of x, i.e. the largest integer not exceeding x. Let us define the quantities

X: the number that Matt picks,
Y: the amount, in dollars, he receives from John.

In this way, we have two random variables

$$X : [0, 10] \mapsto \mathbb{R} \text{ and } Y : [0, 10] \mapsto \mathbb{R},$$

which are defined for $\omega \in \Omega = [0, 10]$, as

$$X(\omega) = \omega, \quad Y(\omega) = [\omega],$$

where, since ω is real, $[\omega]$ is its integer part.

We see from above that a random variable depicts the sample space Ω to a subset of the set of real numbers. Specifically, this subset is

$$\{x \in \mathbb{R} : X(\omega) = x \text{ for some } \omega \in \Omega\}.$$

This set is called the **range of values** (or simply the range) of the random variable X and we shall denote it by R_X. We may therefore view the range of values of X as a *transformed sample space* for the experiment we are studying. While the original sample space Ω contains the (potentially, nonnumerical) outcomes of the experiment, the elements of the set R_X are numerical values associated with these outcomes. In the special case $X(\omega) = \omega$, so that the elementary events in Ω are also real numbers, we see that the original and the transformed sample spaces coincide.

More generally, the sample space Ω can be

- finite,
- countably infinite, or
- uncountably infinite.

However, the "size" of R_X may not always be of the same nature. For instance, in Example 4.1, we have

$$R_X = \{0, 1, 2, 3\}$$

while $|\Omega| = 8$ (so that both sets are finite); in Example 4.2, we saw that

$$R_X = \{0, 1, 2, \ldots\}$$

and Ω is also countably infinite; and finally, in Example 4.4, Ω is an interval of the real line (which is uncountable) while

$$R_X = \Omega = [0, 10] \quad \text{(continuous sample space)},$$

$$R_Y = \{0, 1, 2, \ldots, 10\} \quad \text{(finite sample space)}.$$

Example 4.4 Suppose we throw a die twice and consider the random variables

X_1: the outcome in the first throw,

X_2: the outcome in the second throw,

Y: the sum of the two outcomes,

Z: the difference of the second outcome from the first one,

W: the smallest of the two outcomes.

The sample space consists of the 36 pairs

$$\Omega = \{(1, 1), (1, 2), \ldots, (1, 6), (2, 1), \ldots, (6, 6)\}$$

so that, if the outcome of the experiment is, say $\omega = (4, 2)$, the values of the above five random variables are

$$X_1((4, 2)) = 4, \quad X_2((4, 2)) = 2, \quad Y((4, 2)) = 6, \quad Z((4, 2)) = 2, \quad W((4, 2)) = 2.$$

Further, the ranges of these variables are

$$R_{X_1} = R_{X_2} = R_W = \{1, 2, 3, 4, 5, 6\},$$

$$R_Y = \{2, 3, \ldots, 12\}, \quad R_Z = \{-5, -4, \ldots, 4, 5\}.$$

Observe that in Example 4.4, the random variables Y, Z, and W can be expressed in terms of the variables X_1 and X_2 as follows:

$$Y = X_1 + X_2, \; Z = X_1 - X_2, \; W = \min\{X_1, X_2\}.$$

It turns out that in many practical situations, there is a need to build new random variables using existing ones, and then associate probabilities to, or study the properties of, these new random variables. More formally, if $X_1 : \Omega \mapsto \mathbb{R}$ and $X_2 : \Omega \mapsto \mathbb{R}$ are two variables defined on a sample space Ω, the usual operations between functions, such as $aX_1 + bX_2$ $(a, b \in \mathbb{R})$, $X_1X_2, X_1/X_2$ (assuming $X_2 \neq 0$), and so on, define new random variables on the same sample space. Similarly, if $f : \mathbb{R} \mapsto \mathbb{R}$ is a real function, the composition $f \circ X_1 : \Omega \mapsto \mathbb{R}$ defines a new random variable, provided that the condition we imposed in Definition 4.1 is satisfied. For example, let $f(x) = x^3$. Then, the composition $f \circ X_1$ defines the variable X_1^3. Combining the above, we may create more elaborate forms of random variables, such as

$$3 \sin X_1 + 2 \cos X_2, \quad -5 \ln Y + \mathrm{e}^Z, \quad (X_1^2 + X_2^2) \cdot (X_1 + \ln X_2),$$

and so on.

Functions of random variables arise in real-life applications in a completely natural way. Suppose that a darts player aims at a target, and let X and Y represent the coordinates of a darts throw in a Cartesian system, having its origin at the center of the target. Then the random variable

$$R = \sqrt{X^2 + Y^2}$$

expresses the distance of the point hit by the darts to the center.

EXERCISES

1. Nick throws a die four times in succession. Define the random variable X: the number of times that the outcome is heads.

    (i) Find the sample space Ω for this experiment.

    (ii) What is the range of values for X?

    (iii) Identify the elements that each of the following sets (events) contains:

    $$A_1 = \{\omega \in \Omega : X(\omega) = 2\},$$
    $$A_2 = \{\omega \in \Omega : X(\omega) \geq 3\}.$$

    (iv) Calculate the probabilities of each event in (iii).

2. Find the range of values for each of the following random variables:

    (i) the number of successful hits in n throws of a dart against a target;

    (ii) the age of a person randomly selected from a population;

    (iii) the number of throws of a coin until two successive outcomes are the same;

    (iv) the number of throws of a coin until two outcomes (not necessarily successive) are the same;

    (v) the weight of a student selected at random from a class;

    (vi) the measurement error when weighing an item on a scale that has a margin of ± 0.1 g.

In each case, state whether the range of the variable in question is finite, countably infinite, or uncountable.

3. When an oil tanker sinks, the radius of the contaminated area by the oil spot is represented by a random variable X.
   (i) What is the range of values for X?
   (ii) Explain how we can define a new variable Y that gives the area of the contaminated region in the sea. You may assume that the contaminated area forms a circle whose center is at the ship that sunk. What is the range of values for Y?

4. A building plot has the shape of a rectangle whose length and width are represented by two variables X and Y. The ranges of these two variables are the intervals $R_X = [5, 8]$ and $R_Y = [4, 6]$. Let E be the area of the building plot.
   (i) Find the range of values, R_E, for this variable.
   (ii) Explain whether each of the sets R_X, R_Y and R_E is countable.

5. A gas station has four pumps with unleaded petrol and six pumps with diesel. For an arbitrary time instant during a day, find the range of values for each of the following variables:

   X: the number of pumps with unleaded petrol which are in use;
   Y: the total number of pumps in use;
   Z: the larger number of pumps, between the two types, which are in use.

6. We select at random a point A in the interval $[-1, 1]$. We then select a second point B, also at random, but now from the interval $[-1, A]$. Identify the sets representing the ranges of each of the following variables:
   (i) X: the position of the point A in the interval $[-1, 1]$;
   (ii) Y: the position of the point B in the interval $[-1, 1]$;
   (iii) Z: the distance between the two points A and B;
   (iv) $W = [10X]$;
   (v) $V = [Z]$;
   (vi) $U = |X - Y|$,

   where, as usual, $|\cdot|$ denotes the absolute value and $[\cdot]$ the integer part.

7. A bullet is shot against a target whose center is located at the point O with coordinates $(0, 0)$ in a Cartesian system. The position of the point that the bullet hits the target is described by its coordinates (X, Y), which are random variables. The ranges of values for X and Y are $R_X = [-2, 2]$ and $R_Y = [-3, 4]$, respectively.
   (i) Identify the range of values for each of the random variables $|X|, |Y|, X + Y$, and $X - Y$.
   (ii) If R stands for the distance between the bullet's position and the center, find the range of values for R.

4.2 DISTRIBUTION FUNCTIONS

In the last section, we introduced the concept of a random variable. Using this concept, we can formulate questions of interest associated with events on a sample space and then answer these questions by calculating the corresponding event probabilities. Some typical examples are as follows:

- When throwing two dice, we may be interested in the probability of the event that the sum of the two outcomes is 7. In such a case, we introduce a random variable X that represents the sum of the two dice outcomes and then try to calculate the probability $P(X = 7)$. If, instead, we are interested in the probability that the sum of the outcomes is *at least* 7, this can be expressed as $P(X \geq 7)$;
- Suppose we wait for the next train to arrive at a metro station. Since we do not want to wait for too long we wish to know, for example the probability that our waiting time at the platform is at most three minutes. The way to formulate this is by defining a random variable X that represents the time we wait until the next train arrives. Then, the probability of interest is $P(X \leq 3)$, with X expressed in minutes;
- Suppose we invest in a stock and we want to estimate our financial earnings (or losses) from this investment. We define a random variable X that represents our profit from the investment, noting that this profit can be negative if we suffer a loss. Then we can calculate probabilities such as $P(X > 0)$, corresponding to the probability that the investment is profitable, $P(X \geq 0)$ for the probability that we do not lose from this investment, or $P(10 \leq X \leq 20)$ for the probability that our profit is at least 10 and at most 20 monetary units;
- In a family with n children, assume that we want to find the probability that there is at least one boy and at least one girl in that family. At first sight, this seems to involve two random variables, but simplicity is maintained by the observation that if X stands for the number of girls in the family, then the number of boys is $n - X$. Thus, the required probability that the family has children of both genders can be expressed as $P(1 \leq X \leq n - 1)$.

More generally, when we are interested in calculating a probability associated with a random variable X, this typically reduces to the determination of a probability in one of the following forms:

$$\begin{gathered} P(X = a), \quad P(X > a), \quad P(X \geq a), \quad P(X < b), \quad P(X \leq b), \\ P(a \leq X \leq b), \quad P(a < X \leq b), \quad P(a \leq X < b), \quad P(a < X < b). \end{gathered} \tag{4.1}$$

As we will see, all these probabilities can be found once we have an expression for the probabilities $P(X \leq t)$ for any $t \in \mathbb{R}$. We thus consider the *function* $F(t) = P(X \leq t)$, which contains essentially all the information we need for the random variable X.

Definition 4.2 Let X be a random variable defined on the sample space Ω. The real-valued function $F : \mathbb{R} \mapsto \mathbb{R}$ defined by

$$F(t) = P(X \leq t) = P(\omega \in \Omega : X(\omega) \leq t)$$

is called the cumulative distribution function, or simply the **distribution function** of X.

In experiments where we need more than one random variables, say $X, Y, Z, \ldots$, we typically use the variable as a subscript to denote the associated distribution function and write $F_X, F_Y, F_Z, \ldots$

Although in the definition we have mentioned that the range of values for F is the set of real numbers $\mathbb{R}$, it is obvious that, since $F(t)$ is a probability for all real t, its values must lie in the interval $[0, 1]$, that is

$$0 \leq F(t) \leq 1 \quad \text{for all } t \in \mathbb{R}.$$

The next proposition presents the main properties of a distribution function. These properties are vital for further discussions and will be frequently used without further reference.

Proposition 4.1 ***(Properties of a distribution function)***

(i) *If $t_1 < t_2$, then $F(t_1) \leq F(t_2)$, so that F is an increasing function;*

(ii) *We have*

$$\lim_{t\to\infty} F(t) = 1, \qquad \lim_{t\to-\infty} F(t) = 0; \tag{4.2}$$

(iii) *Let $\{t_n\}_{n\geq 1}$ be a decreasing sequence of real numbers such that $\lim_{n\to\infty} t_n = t$. Then we have*

$$\lim_{n\to\infty} F(t_n) = F(t).$$

Proof:

(i) Let $t_1 < t_2$ be two real numbers. Consider the following events:

$$A_1 = \{\omega \in \Omega : X(\omega) \leq t_1\},$$
$$A_2 = \{\omega \in \Omega : X(\omega) \leq t_2\}.$$

It is clear that $A_1 \subseteq A_2$, and since

$$F(t_1) = P(A_1), \quad F(t_2) = P(A_2),$$

the result readily follows.

(ii) In order to establish the first limiting relationship in (4.2), it is sufficient to prove that, for every increasing sequence $\{t_n\}_{n\geq 1}$ such that $\lim_{n\to\infty} t_n = \infty$, we have $\lim_{n\to\infty} F(t_n) = 1$. This, in turn, is a consequence of the continuity of probability (Proposition 1.11) if we observe that the sequence of events $\{A_n\}_{n\geq 1}$ defined by $A_n = \{X \leq t_n\}$ is an increasing sequence with

$$\lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n = \bigcup_{n=1}^{\infty} \{X \leq t_n\} = \{X < \infty\}.$$

This gives

$$\lim_{n\to\infty} F(t_n) = \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(X < \infty) = P(\Omega) = 1.$$

The proof of the second relationship in (4.2) is identical.

(iii) If $\{t_n\}_{n\geq 1}$ is a decreasing sequence of real numbers, then the associated sequence of events $\{A_n\}_{n\geq 1}$ with $A_n = P(X \leq t_n)$ is also decreasing. Moreover, we have

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \{X \leq t_n\} = \{X \leq t\},$$

so that, by appealing again to the continuity theorem of probability (Proposition 1.11), we deduce that

$$\lim_{n\to\infty} F(t_n) = \lim_{n\to\infty} P(A_n) = P(\lim_{n\to\infty} A_n) = P(X \leq t) = F(t),$$

which completes the proof. □

We mention that Property (iii) of Proposition 4.1 means that the function $F(t)$ is a **right-continuous function**.

Let us now examine how, with the aid of the last proposition, we can use the distribution function of a random variable X effectively in order to calculate the probabilities of the various forms given in (4.1). First, we note that the probabilities $P(X \leq b)$ and $P(X > a)$ are obtained immediately via the relations

$$P(X \leq b) = F(b),$$

$$P(X > a) = 1 - P(X \leq a) = 1 - F(a).$$

Regarding the probability of the event $\{a < X \leq b\}$, we have the following result.

Proposition 4.2 *Let a and b be two real numbers such that $a < b$. Then we have*

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

Proof: Define the events $A = \{X \leq a\}$ and $B = \{X \leq b\}$. Since we assume that $a < b$, it follows that $A \subset B$ and $AB = A$; we further note that the event $\{a < X \leq b\}$ can be expressed as $\{a < X \leq b\} = B - A = BA'$. Consequently,

$$\begin{aligned} P(a < X \leq b) &= P(BA') = P(B) - P(AB) = P(B) - P(A) \\ &= P(X \leq b) - P(X \leq a) = F(b) - F(a), \end{aligned}$$

which is the desired result. □

Next, in order to calculate a probability of the form $P(X < b)$, we note that the event sequence $\{A_n\}_{n\geq 1}$ defined by

$$A_n = \left\{X \leq b - \frac{1}{n}\right\}$$

is an increasing sequence while its limit is

$$\lim_{n\to\infty} A_n = \lim_{n\to\infty} \left\{X \leq b - \frac{1}{n}\right\} = \{X < b\}.$$

Thus, from the continuity of probability, we derive

$$\begin{aligned} P(X < b) &= P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n) = \lim_{n\to\infty} P\left(X \leq b - \frac{1}{n}\right) \\ &= \lim_{n\to\infty} F\left(b - \frac{1}{n}\right) = F(b-), \end{aligned}$$

using the result in Part (iii) of Proposition 4.1 for the last step and noting that the symbol $F(t-)$ denotes the left-hand limit of the function F at the point $t \in \mathbb{R}$.

For the probability $P(X \geq a)$, we simply note that

$$P(X \geq a) = 1 - P(X < a) = 1 - F(a-)$$

and, finally for the probability $P(X = a)$, we have

$$P(X = a) = P(X \leq a) - P(X < a) = F(a) - F(a-).$$

Table 4.1 presents all the events of the various forms listed in (4.1), associated with a random variable X, and their respective probabilities expressed in terms of the distribution function F of X.

Recall that a distribution function is always a right-continuous function. If, in addition, F is left-continuous at any point a, then $F(a) = F(a-)$, and it follows from above that $P(X = a) = 0$. Further, we observe that if $F(a) = F(a-)$ for all a, then all the events in the third column of Table 4.1 have the same probability, namely,

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

Event	Probability	Event	Probability
$X \leq b$	$F(b)$	$a < X \leq b$	$F(b) - F(a)$
$X < b$	$F(b-)$	$a < X < b$	$F(b-) - F(a)$
$X > a$	$1 - F(a)$	$a \leq X \leq b$	$F(b) - F(a-)$
$X \geq a$	$1 - F(a-)$	$a \leq X < b$	$F(b-) - F(a-)$
$X = a$	$F(a) - F(a-)$		

Table 4.1 Probabilities of events for a variable X given in terms of its distribution function F.

In a similar fashion we see that, if F is continuous at the point a, then $P(X \geq a) = P(X > a)$, and if it is continuous at the point b, then $P(X < b) = P(X \leq b)$.

On the other hand, if F is *n*ot continuous at a, then we have

$$F(a) - F(a-) = P(X = a).$$

In this case, we say that F has a **jump** (or, a discontinuity) of size $P(X = a)$ at the point $a \in \mathbb{R}$.

Example 4.5 Suppose we consider the population of all families with two children in a city. Selecting a family at random, we define the random variable

$$X : \text{number of girls in the family.}$$

Find the distribution function of X.

SOLUTION For the random experiment whose outcome is the genders of the two children in the family, the sample space is

$$\Omega = \{BB, BG, GB, GG\}$$

(here B and G stand for boy and girl, respectively, and we assume that in each pair above, the elder child is given first).

The variable X is then expressed as

$$X = X(\omega) = \begin{cases} 0, & \text{if } \omega = BB, \\ 1, & \text{if } \omega = BG \text{ or } GB, \\ 2, & \text{if } \omega = GG. \end{cases}$$

The range of values for X is the set

$$R_X = \{0, 1, 2\}$$

and, considering the distribution function F, we observe the following:

(i) For $t < 0$, the associated event $\{X \leq t\} = \{\omega \in \Omega : X(\omega) \leq t\}$ is impossible to occur, and therefore

$$F(t) = P(X \leq t) = P(\emptyset) = 0;$$

(ii) For $t = 0$, the event $\{X \leq 0\}$ may only occur if $X = 0$ and this has probability

$$P(X \leq 0) = P(X = 0) = P(\{BB\}) = \frac{1}{4}.$$

Also, if $t \in (0, 1)$, the probability $P(X \leq t)$ is the same as $P(X = 0)$, since X cannot take on any values in the interval $(0, t]$. Thus, we have shown that for $0 \le t < 1$,

$$F(t) = P(X \le t) = P(X = 0) = P(\{BB\}) = \frac{1}{4};$$

(iii) For $t = 1$, we have

$$\begin{aligned}\{X \le 1\} &= \{X = 0 \ \text{ or } \ X = 1\} = \{\omega \in \Omega : X(\omega) = 0 \ \text{ or } \ X(\omega) = 1\}\\ &= \{BB, BG, GB\},\end{aligned}$$

and it is clear that this event has probability $3/4$. If t takes values in the interval $(1, 2)$, we see again that this probability does not change, so that for any $1 \le t < 2$, we have

$$F(t) = P(X \le t) = P(X = 0 \text{ or } X = 1) = P(\{BB, BG, GB\}) = \frac{3}{4};$$

(iv) Finally, arguing as above, we see that for $t \ge 2$ the event $\{X \le t\}$ is the same as the union of the three events $\{X = 0\}, \{X = 1\}$, and $\{X = 2\}$, which is the entire sample space Ω. This gives, for $t \ge 2$,

$$P(X \le t) = P(X = 0 \text{ or } X = 1 \text{ or } X = 2) = P(\Omega) = 1.$$

Putting together all four cases for the values of t that we have examined above, we obtain the distribution function of the variable X as

$$F(t) = P(X \le t) = \begin{cases} 0, & -\infty < t < 0,\\ 1/4, & 0 \le t < 1,\\ 3/4, & 1 \le t < 2,\\ 1, & t \ge 2. \end{cases}$$

The graph of the distribution function is given in Figure 4.1. We observe that F is a *step function*; it consists of horizontal segments, parallel to the x-axis. It is worth noting that, if we were only given either this graph or the formula for the distribution function above, we would be able to calculate any probability associated with the variable X, using the formulas from Table 4.1. For example, the probability $P(1 \le X \le 4)$ is

$$P(1 \le X \le 4) = F(4) - F(1-) = 1 - \frac{1}{4} = \frac{3}{4},$$

the probability $P(X \ge 1.5)$ is

$$P(X \ge 1.5) = 1 - F(1.5-) = 1 - F(1.5) = 1 - \frac{3}{4} = \frac{1}{4},$$

and so on.

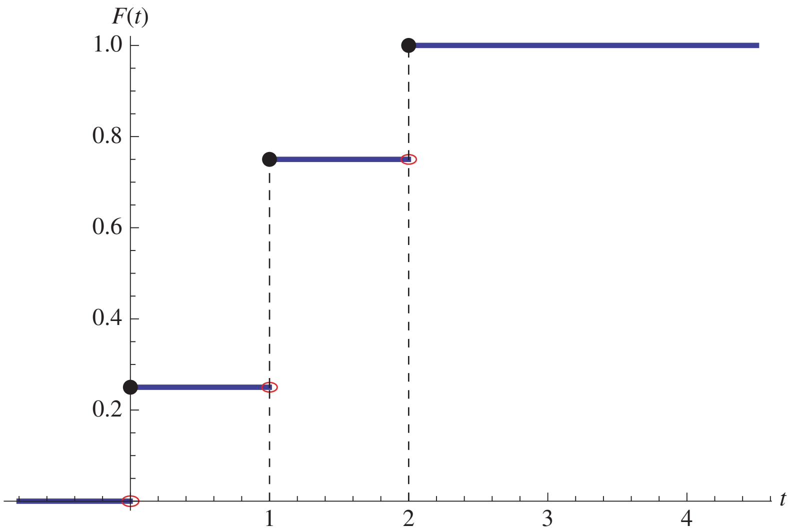


Figure 4.1 Plot of the distribution function for Example 4.5.

Example 4.6 (A continuous distribution function)
Suppose a customer's waiting time at an office (in minutes) is represented by a random variable X, which has the following distribution function:

$$F(t) = P(X \leq t) = \begin{cases} 0, & t < 0, \\ t/4, & 0 \leq t < 1, \\ 1/4, & 1 \leq t < 2, \\ t/8, & 2 \leq t < 8, \\ 1, & t \geq 8. \end{cases}$$

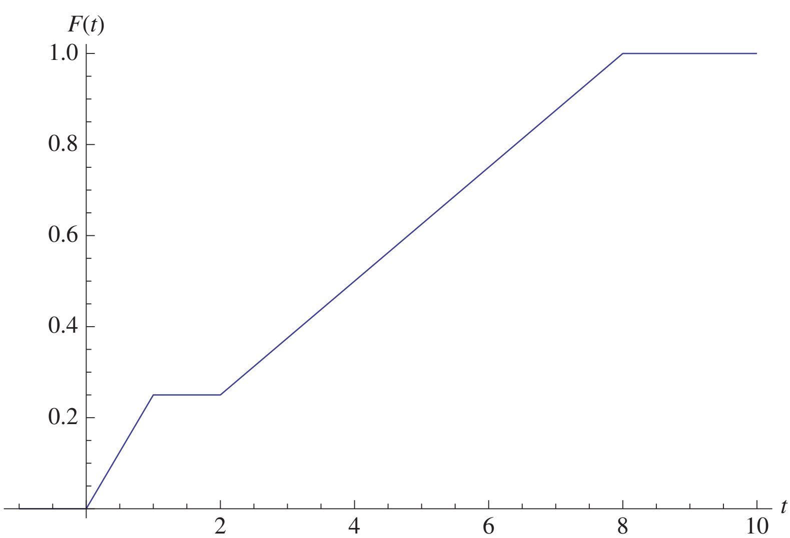


Figure 4.2 Plot of the distribution function for Example 4.6.

Either from this formula or from the graph (see Figure 4.2), we can see that this distribution function is continuous for any $t \in \mathbb{R}$. We also see, for example, that the probability that a customer waits at the bank for three minutes at most is

$$P(X \leq 3) = \frac{3}{8},$$

while the probability that the waiting time is between one and four minutes equals

$$P(1 \leq X \leq 4) = F(4) - F(1-) = F(4) - F(1) = \frac{4}{8} - \frac{1}{4} = \frac{1}{4};$$

in the calculation we used the continuity of F at the point $t = 1$, i.e. that $F(1-) = F(1)$.

Further, suppose that we are interested in the probability that a customer who has waited for 2.5 minutes, will wait for at most 3 additional minutes. Mathematically, this is expressed by the conditional probability

$$P(X \leq 5.5 | X \geq 2.5)$$

and using the definition of conditional probabilities (Definition 3.1), we then obtain

$$\begin{aligned} P(X \leq 5.5 | X \geq 2.5) &= \frac{P(X \leq 5.5, X \geq 2.5)}{P(X \geq 2.5)} = \frac{P(2.5 \leq X \leq 5.5)}{P(X \geq 2.5)} \\ &= \frac{F(5.5) - F(2.5-)}{1 - F(2.5-)} = \frac{F(5.5) - F(2.5)}{1 - F(2.5)} \\ &= \frac{(5.5/8) - (2.5/8)}{1 - (2.5/8)} = \frac{6}{11}, \end{aligned}$$

using again the continuity of F at the point $t = 2.5$.

A rather striking feature for this, but also for any other distribution function which is everywhere continuous, is that

$$P(X = a) = F(a) - F(a-) = F(a) - F(a) = 0$$

for all real a. So, taking $a = 5$ (say) in the example, we see that the probability that a customer waits for exactly five minutes is zero(!), although the event $\{X = 5\}$ is clearly not impossible. We shall not discuss this, rather counter-intuitive, property of continuous distributions here and shall defer the discussion to the last two chapters of the book, which are devoted to a detailed study of such distributions.

We have seen in Proposition 4.1 that the distribution function F of a random variable X is increasing and right-continuous such that

$$\lim_{t \to -\infty} F(t) = 0, \quad \lim_{t \to \infty} F(t) = 1.$$

In fact, it can be shown (although the proof is not given here) that these properties characterize completely the class of functions that may be used as distribution functions for a random variable. More precisely, a real-valued function F is a distribution function (of a suitably defined random variable X) if and only if it is increasing, right-continuous and its limits as $t \to -\infty$ and $t \to \infty$ are 0 and 1, respectively.

Example 4.7 The amount of money, in thousands of dollars, that an insurance company pays to a policyholder when (s)he makes a claim is described by a random variable X whose range is the set $(0, 4)$ and has the following distribution function:

$$F(t) = P(X \le t) = \begin{cases} 0, & t < 0, \\ t^2/8, & 0 \le t < 2, \\ c(8t - t^2), & 2 \le t < 4, \\ 1, & t \ge 4. \end{cases} \tag{4.3}$$

(i) Find the value of the constant c in the above formula.

(ii) Calculate the probability that the amount which the company pays to a customer is

(a) less than \$2000,

(b) more than \$2000,

(c) exactly \$2000,

(d) exactly \$2400.

(iii) Let A be the event that, for a particular claim, the amount to be paid is between \$500 and \$2500, while B is the event that the amount for that claim is at least \$1500. Calculate the probabilities of these two events and examine whether they are independent.

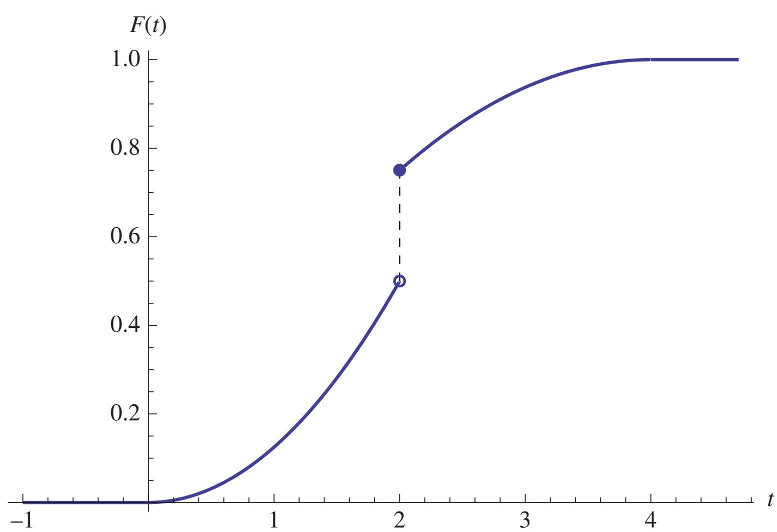


Figure 4.3 Plot of the distribution function for Example 4.7.

SOLUTION

(i) Since we are given that X takes values in the set $(0, 4)$, this means that $P(X < 4) = 1$, and this in turn implies that

$$F(4-) = \lim_{t \to 4-} F(t) = 1.$$

But for $2 \leq t < 4$, the distribution function has the formula $F(t) = c(8t - t^2)$, and so we must have

$$\lim_{t \to 4-} c(8t - t^2) = c(8 \cdot 4 - 4^2) = 1,$$

which gives $16c = 1$, so that $c = 1/16$. For this value of c, the function F is an increasing, right-continuous function with limits 0 and 1 at $-\infty$ and $+\infty$, respectively. Hence, it is well defined as a distribution function of the amount paid for an insurance claim.

(ii) Using the value of c found in (i), we have

(a) $P(X < 2) = F(2-) = \dfrac{1}{8} \cdot 2^2 = \dfrac{1}{2}$,

(b) $P(X > 2) = 1 - F(2) = 1 - \dfrac{1}{16}(8 \cdot 2 - 2^2) = 1 - \dfrac{3}{4} = \dfrac{1}{4}$,

(c) $P(X = 2) = F(2) - F(2-) = \dfrac{3}{4} - \dfrac{1}{2} = \dfrac{1}{4}$,

(d) $P(X = 2.4) = F(2.4) - F(2.4-) = 0$,

the last result being obvious since F is continuous everywhere *except* at the point $t = 2$ (see Figure 4.3). For $t = 2$, F has a discontinuity (an upward jump) of size $P(X = 2) = 1/4$. Let us try to explain how a distribution function of this form may arise in practice and in this particular context. Since we have seen above that $P(X = 2) = 1/4$, this means that a quarter of the payments made for a claim are exactly \$2000. It seems extremely unlikely that this is a result of pure chance, since from the form of the function F (but also from Figure 4.3) we see that payments can take any value in the interval $(0, 4)$, yet only for $t = 2$ there is a positive probability that a payment X equals t. Further, we have seen that $F(2) = 3/4$ and $F(2-) = 1/2$. This means that 75% of the payments for the claims are for amounts up to \$2000, with 25% being exactly equal to 2000. A simple explanation for this might be that, for a certain proportion of customers, the company has a maximum payable amount of 2000; if any of these customers makes a claim exceeding that amount, the company pays exactly \$2000. For the remaining customers, there is a different deal so that they can claim a higher amount; this could be either because the claims described by F belong to two different portfolios, or the company has classified its customers into two different types.

(iii) For the event A, we have

$$P(A) = P(0.5 \le X \le 2.5) = F(2.5) - F(0.5-) = F(2.5) - F(0.5)$$
$$= \frac{1}{16}[8 \cdot (2.5) - (2.5)^2] - \frac{(0.5)^2}{8} = 0.859\ 375 - 0.031\ 25 = 0.828\ 125,$$

while for the event B we obtain

$$P(B) = P(X \ge 1.5) = 1 - F(1.5-) = 1 - F(1.5) = 1 - \frac{(1.5)^2}{8} = 0.718\ 75.$$

For the events A and B to be independent, we must have $P(AB) = P(A)P(B)$. For the probability of the event AB, i.e. for the simultaneous occurrence of these two events, we observe first that $AB = \{1.5 \le X \le 2.5\}$ and this gives, upon using the fact that F is continuous at the point $t = 1.5$,

$$P(AB) = P(1.5 \le X \le 2.5) = F(2.5) - F(1.5-) = F(2.5) - F(1.5)$$
$$= 0.859\ 375 - \frac{(1.5)^2}{8} = 0.578\ 125.$$

Since $P(AB) \ne P(A)P(B)$, we conclude that A and B are *not* independent.

In the three examples of the present section, we have so far seen three different types of distribution functions, corresponding to three different types of random variables that we encounter in probability theory.

First, in Example 4.5, the variable X represented the number of girls in a family with two children, with only three possible values: 0, 1, and 2. Next, in Example 4.6 (waiting time at a bank), X could take any value in a union of intervals of the real line (see Exercise 2 of this section). Looking at the graphs of their distribution functions, we see that in the latter case F is everywhere continuous. Hence, it seems natural to call the associated random variable a *continuous random variable*, as we do in the following definition. In the former case, we observe that F changes values at the points $0, 1$, and 2 and has a discontinuity (from the left) precisely at these points.

Finally, in Example 4.7, we have a distribution that is continuous everywhere except at a single point. The first two types of distributions are the most important ones and they are defined properly next; they will be the primary vehicle for studying probabilities in the remainder of this book. Distributions of the third type occur less frequently and are discussed further in Section 6.5.

Definition 4.3 If a random variable X takes only a finite or at most an infinite but countable number of values, then X is called a **discrete random variable**. We then say that the distribution F of X is a **discrete distribution** that has jumps at every point x in the range of X.

If, on the other hand, a distribution function F has no jumps, then it is called a **continuous distribution**, while a random variable X with a continuous distribution is called a **continuous random variable**.

The simplest examples for a discrete random variable X are of course those where X assumes integer values. We should keep in mind, however, that according to Definition 4.3 this is not restrictive in general; the set of all *rational numbers* for example, or any subset of it, is also countable.

Although in most cases, it is clear from the context whether a random variable is discrete or continuous, in case of ambiguity and, in particular, when only the distribution function F of the variable is given, drawing a picture of F clarifies the situation. If the graph of F has no discontinuities, then the associated random variable is continuous. If F changes value (increases) only when an upward jump occurs, so that F is discontinuous at every such point, then the random variable having distribution F is discrete. The intermediate case is best exemplified by Figure 4.3; the distribution function F there is continuous for any real t except at the point $t = 2$ where there is a jump. This distribution is neither purely continuous nor purely discrete (see Section 6.5 for more details).

EXERCISES

Group A

1. A random variable X has a distribution function F given by

$$F(t) = \begin{cases} 0, & t < 0, \\ 1/16, & 0 \le t < 1, \\ 3/16, & 1 \le t < 2, \\ 1/2, & 2 \le t < 3, \\ 11/16, & 3 \le t < 4, \\ 1, & 4 \le t < \infty. \end{cases}$$

   (i) Explain whether X is a discrete or a continuous random variable and identify the range of its values.

   (ii) Using the distribution function, calculate each of the following probabilities:

$$P(X \le 3),\ P(X < 3),\ P(X = 1),\ P(X > 2),\ P(2 \le X \le 4).$$

2. Identify the range of values for X in Example 4.6. Then calculate the probability $P(1 < X < 2 | X \le 2)$.

3. Jenny is waiting at a bus stop for the bus that takes her to work. It is known that the waiting time (in minutes) for this particular bus, represented by a random variable X, has the following distribution function:

$$F(t) = \begin{cases} 0, & -\infty < t < 0, \\ t/6, & 0 \le t < 2, \\ (t-1)/3, & 2 \le t < 3, \\ (t+6)/12, & 3 \le t < 6, \\ 1, & 6 \le t < \infty. \end{cases}$$

   (i) Draw a graph of this distribution function. Are there any discontinuity points on the graph?

(ii) Calculate the probabilities $P(X = i)$, for $i = 2, 3, 6$.

(iii) Find the conditional probabilities

$$P(X > 3|X > 2), \quad P(1 \leq X < 4|X > 2).$$

Explain what these probabilities represent in words.

4. At the end of each year, a company gives a bonus to all of its employees. The size of this bonus can take one of four possible values (in thousands of dollars): 1, 2, 5, 12. Let X be the size of the bonus given to a randomly chosen employee. The distribution function of X is given by

$$F(t) = \begin{cases} 0, & -\infty < t < 1, \\ 0.35, & 1 \leq t < 2, \\ K, & 2 \leq t < 5, \\ 0.85, & 5 \leq t < 12, \\ 1, & 12 \leq t < \infty. \end{cases}$$

(i) What are the possible values for the constant K above?

(ii) Explain whether the random variable X is discrete or continuous.

(iii) For $K = 0.6$, find the probabilities

$$P(1 < X \leq 5), \quad P(X = 2), \quad P(X \geq 2).$$

5. Examine which of the functions below can be used as distribution functions of a random variable X.

(i) $$F(t) = \begin{cases} 1 - e^{-3t}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

(ii) $$F(t) = \begin{cases} \dfrac{t}{t+1}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

(iii) $$F(t) = e^{-t}, \qquad -\infty < t < \infty.$$

(iv) $$F(t) = \begin{cases} 0, & -\infty < t < 1, \\ 1, & 1 \leq t < 3 \quad \text{or } t \geq 5, \\ 0.5, & 3 \leq t < 5. \end{cases}$$

(v) $$F(t) = \begin{cases} 0, & -\infty < t < 0, \\ t^2/8, & 0 \leq t < 2, \\ (t-1)/3, & 2 \leq t < 4, \\ 1, & t \geq 4. \end{cases}$$

(vi) $$F(t) = \frac{1}{1 + e^{-t}}, \qquad -\infty < t < \infty.$$

6. In a family with four children, let X be the number of girls. Write down the distribution function of X.

7. The orders for CDs (in hundreds of thousands of units) received daily by a factory that produces compact discs are represented by a random variable X with distribution function

$$F(t) = \begin{cases} 0, & -\infty < t < 0, \\ \alpha t^2, & 0 \le t < 0.5, \\ \alpha t - t^2, & 0.5 \le t < 1, \\ 1, & t \ge 1. \end{cases}$$

   (i) Find the value of the constant α above, assuming that the number of CDs ordered daily is less than 100 000.

   (ii) Find the probability that, during a particular day, the number of CDs ordered

      (a) is at least 25 000,

      (b) is at most 70 000,

      (c) is larger than 15 000, but smaller than 45 000,

      (d) exceeds 50 000,

      (e) is exactly 50 000,

      (f) is exactly 80 000.

8. The distribution function of a random variable X is given by

$$F(t) = \begin{cases} 0, & -\infty < t < -1, \\ \dfrac{(1+t)^2}{2}, & -1 \le t < 0, \\ 1 - \dfrac{(1-t)^2}{2}, & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases}$$

   Calculate the following probabilities

$$P\left(|X| > \frac{1}{2}\right) \quad \text{and} \quad P\left(X < 2 \,\middle|\, |X| > \frac{1}{2}\right).$$

9. With reference to Example 4.7, verify analytically (i.e. without the aid of Figure 4.3) that for $c = 1/16$, the distribution function in (4.3) is an increasing function on the real line. At which values of t is F differentiable?

10. From an ordinary deck of 52 cards, we select three cards at random. Let X be the number of aces drawn. Write down the distribution function of X. Is this a discrete or continuous distribution function?

11. A University lecturer ends her lecture in a particular course she gives on Wednesdays between 11:00 a.m. and 11:04 a.m. Suppose that the time X (in minutes) between 11:00 a.m. and the end of the lecture has the following distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ ct^2, & 0 \le t < 4, \\ 1, & t \ge 4, \end{cases}$$

for a positive constant c. We assume that the lecture finishes before 11:04 a.m. with probability one, i.e. $P(X < 4) = 1$.

(i) What is the value of the constant c?

(ii) Find the probability that the lecture finishes

(a) before 11:03 a.m.,

(b) after 11:01 a.m.

(iii) Assuming that on a particular day the lecture has not finished by 11.01, what is the probability that it will finish within the next two minutes?

Group B

12. Let F_1 and F_2 be the distribution functions of two random variables, and λ be a real number such that $0 \le \lambda \le 1$. Verify that the function F defined by

$$F(t) = \lambda F_1(t) + (1 - \lambda)F_2(t), \quad t \in \mathbb{R},$$

is also a distribution function. The distribution $F(t)$ is referred to as a **two-component mixture distribution**.

13. A company manager has two secretaries, Mary and Joanne. In each page typed by Mary, the number of misprints can be 0, 1, or 2 with probabilities 0.50, 0.35, and 0.15, respectively, while the number of misprints in a page typed by Joanne can also be 0, 1, or 2, but with probabilities 0.60, 0.30, and 0.10, respectively. The manager uses Mary more often to type his documents, and in fact she types 50% more pages than Joanne does.

Among the pages typed, we select one at random, without knowing who typed it. Let X be the number of misprints on that page.

(i) Use the law of the total probability (Proposition 3.4) to calculate the probabilities

$$P(X = 0), \quad P(X = 1), \quad P(X \ge 1).$$

(ii) Obtain the distribution function of X.

14. Let X be a random variable with distribution function F. The real number m that satisfies

$$F(m-) \le 0.5 \le F(m)$$

is called the **median** of this distribution function. Assuming that F is known, and that it is continuous for any real t, explain how m can be found from F.

Application: The rate of increase for the price of oil (per gallon) is described by a random variable X with distribution function

$$F(t) = \frac{1}{1 + e^{-t}}, \quad -\infty < t < \infty.$$

Find the median of this distribution and interpret its value. The distribution $F(t)$ is called the **logistic distribution**.

15. In Example 4.7, you were given that the distribution F of the amount paid by an insurance company for a claim that arrives from one of its customers is given by (4.3) and that this amount lies in the range $(0, 4)$. Without this last bit of information, would we able to determine c uniquely? Explain.

16. A random variable is said to be **symmetric around zero** if we have

$$P(X \geq t) = P(X \leq -t)$$

for any $t \in \mathbb{R}$. Prove that, if X is symmetric around zero with a distribution function F, then the following hold:

(i) $P(|X| \leq t) = 2F(t) - 1$;

(ii) $P(X = t) = F(t) + F(-t) - 1$.

If, in addition, the distribution function F is continuous at $t = 0$, show that $F(0) = 0.5$.

(*Hint*: For the last result, use the relations $P(X > 0) + P(X \leq 0) = 1$ and $P(X \geq 0) = 1 - F(0-) = 1 - F(0)$.)

17. The range of values for a random variable X is the interval $R_X = (0, 1)$, while its distribution function is given by

$$F(t) = \begin{cases} 0, & t < 0, \\ t, & 0 \leq t < 1, \\ 1, & t \geq 1. \end{cases}$$

Find the distribution function of the random variable $Y = X/(1 + X)$.

18. Let X be a random variable with distribution function F and Y be another random variable which is defined as

$$Y = \begin{cases} X, & \text{if } X \leq a, \\ b, & \text{if } X > a, \end{cases}$$

where a and b are two real numbers. Write the distribution function F_Y of Y in terms of F, and b.

4.3 DISCRETE RANDOM VARIABLES

We have seen that a random variable X is called discrete if the range of values for X is either finite or countably infinite; for the latter case, recall that any subset of the set $\mathbb{Z}$ of integers is countable. On the other hand, any (finite or infinite) interval (a, b) of the real line, or any union of such intervals is uncountable, and a random variable with values in that set is continuous. For the rest of this chapter as well as in the next chapter, we concentrate on discrete random variables (and their distributions, called discrete distributions), while continuous distributions are considered in the last two chapters of this book.

In the preceding sections, we introduced the concept of a distribution function associated with a variable X that gives us, for any real t, the probabilities $P(X \leq t)$. This function characterizes completely the way in which probabilities are "distributed" into subsets of the real line, i.e. given a set E, knowledge of the distribution function F of X enables us to calculate the probability $P(X \in E)$; see, for example Table 4.1. Yet, it seems more natural and useful (and it is also often simpler) to know for each real x, the probability that X *is equal to* x (rather than less than or equal to x) and use this as the basis for calculating probabilities of various sets. This is particularly simple and insightful when X takes only a small number of values; consider, for example the case where $R_X = \{0, 1\}$ and suppose we are given that $P(X = 0) = P(X = 1) = 1/2$. Then, there is nothing else we need to know in order to calculate the probability associated with an arbitrary event E. Moreover, once we have established this mechanism to calculate probabilities associated with the possible values of X, we no longer need to care about the original sample space Ω of the experiment. Essentially, what matters is how we use this mechanism to calculate probabilities for values of the variable X and not how these values arose from the elements of Ω. In the example above, where X only takes the values 0 and 1, these two values may be associated with the gender of a person selected randomly from a population (where $X(\{\text{"man"}\} = 0, X(\{\text{"woman"}\} = 1)$, or with the outcome of a coin toss (where $X(\{\text{"tails"}\} = 0, X(\{\text{"heads"}\} = 1)$, and so on.

The main tools for the study of discrete random variables are **probability functions**. For any discrete variable X, we associate a real-valued function $f : \mathbb{R} \mapsto \mathbb{R}$ defined by

$$f(x) = P(X = x), \quad x \in \mathbb{R},$$

which is called the probability function of X (sometimes also called the *probability mass function* associated with X).

Discrete random variables arise naturally in a wide range of practical situations, even in cases where the original sample space Ω involving a random experiment that is continuous. We give a few examples:

- When we toss a coin n times, the random variable X that counts the appearances of heads is a discrete random variable whose range is the set

$$R_X = \{0, 1, 2, \dots, n\};$$

- Suppose X represents the lifetime, in hours, of a light bulb. Then X is not a discrete variable, since X may (theoretically, at least) take on any value in the set $(0, \infty)$. However, in the same experiment, the variable $Y = [X]$ that emerges from X by ignoring the decimal part of the value it takes, is a discrete random variable. The range of Y is the set of nonnegative integers

$$\{0, 1, 2, \dots\},$$

which is countably infinite;

- Suppose the items produced in a production line are classified as being defective or nondefective, and let X be the random variable that counts the number of items inspected until the first defective item is found. Then, X is a discrete random variable whose range is the (countably infinite) set

$$R_X = \{1, 2, 3, \dots\}.$$

We may also consider the random variable $Y = X - 1$ that counts the number of nondefective items inspected before the first defective item is found. Y is also discrete with range $R_Y = \{0, 1, 2, \dots\}$. Finally, suppose that n items are selected randomly from the production line and the inspection process stops when the first defective item is found *or* when all n items have been inspected (without any defective item being found). Let X be the number of items inspected until the inspection process ends and $Y = X - 1$. The variables X and Y have now a finite range, namely,

$$R_X = \{1, 2, 3, \dots, n\}, \quad R_Y = \{0, 1, 2, \dots, n-1\}.$$

We now turn our attention about the properties of a probability function f for a discrete random variable X. First, since $f(x) = P(X = x)$ represents a probability, we see that f must be nonnegative. Moreover, the range of values of X is at most countably infinite. This implies that the set of values x for which $f(x)$ is strictly positive is (at most) countable.

More specifically, if

$$R_X = \{x_1, x_2, \dots\},$$

then $f(x) = 0$ for any $x \notin R_X$.

Further, using the fact that the events

$$A_i = \{\omega \in \Omega : X(\omega) = x_i\}, \quad i = 1, 2, \dots$$

form a partition of the sample space Ω (explain why!), we can write

$$1 = P(\Omega) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} P(X = x_i) = \sum_{i=1}^{\infty} f(x_i).$$

Summarizing the above, we arrive at the following proposition.

Proposition 4.3 ***(Properties of a probability function)*** *The probability function f of a random variable X with range the (possibly infinite) set*

$$R_X = \{x_1, x_2, \dots\}$$

satisfies the following properties:

(PF1) $f(x) = 0$ *for any* $x \neq x_1, x_2, \dots$;

(PF2) $f(x_i) \geq 0$ *for any* $i = 1, 2, \dots$;

(PF3) $f(x_1) + f(x_2) + \cdots + f(x_n) + \cdots = \sum_{i=1}^{\infty} f(x_i) = 1.$

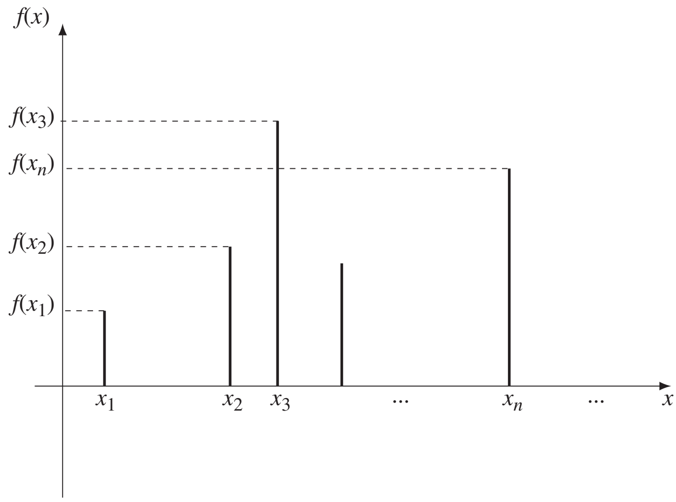


Figure 4.4 Probability function of a discrete random variable.

We note that the converse of the proposition is also true, that is if a function $f : \mathbb{R} \mapsto \mathbb{R}$ satisfies (PF1), (PF2), and (PF3), then f is a probability function of a discrete random variable. For this reason, Proposition 4.3 is often taken as the definition of a probability function. We also mention that, although (PF2) states that $f(x_i) \geq 0$, since $f(x_i) = P(X = x_i)$, it should be clear that in fact the values $f(x_i)$ cannot exceed one,[1] and so the range of a probability function is always in the interval $[0, 1]$.

Figure 4.4 is a typical example of a graph for a probability function. Since the values of the function are zero everywhere except at the points $x_1, x_2, \ldots$, the graph consists of segments parallel to the y-axis. Each segment starts from the x-axis, so that it links the point $(x_i, 0)$ with the point $(x_i, f(x_i))$ on the plane. Note that on the graph we have assumed that the points $x_1, x_2, \ldots$, are in increasing order, so that

$$x_1 < x_2 < x_3 < \cdots < x_n < \cdots$$

This obviously entails no loss of generality and the same assumption will also be made in the sequel without a specific mention of it.

Example 4.8 Let us consider the experiment of throwing two dice and define the random variables

X: the largest of the two outcomes,
Y: the smallest of the two outcomes,
Z: the absolute value of the difference between the two outcomes.

[1] This is guaranteed by the properties of f given in Proposition 4.3, since $f(x_i) \leq \sum_{i=1}^{\infty} f(x_i) = 1$.

Find the probability function for each of these random variables.

SOLUTION We observe initially that the sample space for the experiment consists of the 36 pairs (i, j) with $1 \leq i, j \leq 6$. The following table gives, for each such pair, the value that the variable X takes:

$i \backslash j$	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	2	3	4	5	6
3	3	3	3	4	5	6
4	4	4	4	4	5	6
5	5	5	5	5	5	6
6	6	6	6	6	6	6

Thus, we see that

$$f_X(1) = P(X = 1) = P(\{(1, 1)\}) = \frac{1}{36},$$

$$f_X(2) = P(X = 2) = P(\{(1, 2), (2, 1), (2, 2)\}) = \frac{3}{36},$$

$$f_X(3) = P(X = 3) = P(\{(1, 3), (3, 1), (2, 3), (3, 2), (3, 3)\}) = \frac{5}{36},$$

and so on. Continuing in this way, we obtain the probability function of X, as given in the table below:

x	1	2	3	4	5	6
$f_X(x)$	$\frac{1}{36}$	$\frac{3}{36}$	$\frac{5}{36}$	$\frac{7}{36}$	$\frac{9}{36}$	$\frac{11}{36}$

Note that the values in the table can be expressed concisely through the formula

$$f_X(x) = \frac{2x - 1}{36}, \quad x = 1, 2, 3, 4, 5, 6.$$

We work in the same way for the random variable Y, the smaller between the two dice outcomes. The value of Y in each of the 36 possible outcomes is given in the table below.

$i \backslash j$	1	2	3	4	5	6
1	1	1	1	1	1	1
2	1	2	2	2	2	2
3	1	2	3	3	3	3
4	1	2	3	4	4	4
5	1	2	3	4	5	5
6	1	2	3	4	5	6

The probability function of Y is then as follows:

y	1	2	3	4	5	6
$f_Y(y)$	$\frac{11}{36}$	$\frac{9}{36}$	$\frac{7}{36}$	$\frac{5}{36}$	$\frac{3}{36}$	$\frac{1}{36}$

The probability function of Y can then be expressed concisely as

$$f_Y(y) = \frac{13 - 2y}{36}, \quad y = 1, 2, 3, 4, 5, 6.$$

Finally, for the random variable Z, denoting the absolute difference between the two outcomes, we have the values of Z in each of the 36 possible outcomes as follows:

$i \backslash j$	1	2	3	4	5	6
1	0	1	2	3	4	5
2	1	0	1	2	3	4
3	2	1	0	1	2	3
4	3	2	1	0	1	2
5	4	3	2	1	0	1
6	5	4	3	2	1	0

The probability function of Z is then obtained as follows:

z	0	1	2	3	4	5
$f_Z(z)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$

Can you try to write this function in a concise mathematical form?

Example 4.9 Find the value of the constant c such that the function

$$f(x) = c \cdot \frac{3^{x-1}}{4^x}, \quad x = 1, 2, \ldots,$$

is the probability function of a discrete random variable.

SOLUTION We have to find the value of c such that the conditions (PF1) to (PF3) are satisfied. If we define $R_X = \{1, 2, 3, \ldots\}$, then it is clear that

$$f(x) = 0, \quad \text{for } x \notin R_X.$$

Generally, when we are given the formula for a probability function, we assume that it takes the value 0 at all points not mentioned in the formula. Next, in order to have $f(x) \geq 0$ for all $x = 1, 2, \ldots$, we must have

$$c \cdot \frac{3^{x-1}}{4^x} \geq 0, \quad \text{for all } x = 1, 2, \ldots,$$

which means that c has to be nonnegative. Moreover, from (PF3), we see that the values of f must satisfy the condition

$$\sum_{x=1}^{\infty} c \cdot \frac{3^{x-1}}{4^x} = 1.$$

This gives

$$\frac{c}{4} \sum_{x=1}^{\infty} \left(\frac{3}{4}\right)^{x-1} = 1.$$

But from the formula for the sum of a geometric series, we know that

$$\sum_{x=1}^{\infty} \left(\frac{3}{4}\right)^{x-1} = 1 + \frac{3}{4} + \left(\frac{3}{4}\right)^2 + \cdots = \frac{1}{1 - \frac{3}{4}} = \frac{1}{\frac{1}{4}} = 4,$$

and so we must have $(c/4) \cdot 4 = 1$, that is, $c = 1$.

For a discrete random variable, we have so far defined two functions that can be used to calculate probabilities associated with its values: the (cumulative) distribution function F and the probability function f. An obvious question that arises is whether there is a relationship between them; or, to put it more simply, if we know one of these functions, can we find the other? The answer is affirmative and the precise relation is given in the next proposition.

Proposition 4.4 *Let F be the distribution function and f be the probability function of a discrete random variable X whose range is*

$$R_X = \{x_1, x_2, \ldots, x_n, \ldots\}.$$

Here we assume that $x_1 < x_2 < x_3 < \cdots$, so that the elements of R_X are put in ascending order. Then, we have the following:

(i) *If we know f, then F can be calculated by the formula*

$$F(t) = \begin{cases} 0, & t < x_1, \\ f(x_1), & x_1 \le t < x_2, \\ f(x_1) + f(x_2), & x_2 \le t < x_3, \\ \qquad \cdots & \qquad \cdots \\ \sum_{i=1}^{r-1} f(x_i), & x_{r-1} \le t < x_r, \\ \qquad \cdots & \qquad \cdots \end{cases}$$

(ii) *The probability function f can be calculated from the distribution function F by using the relations*

$$f(x_1) = F(x_1),$$

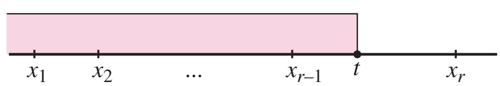
$$f(x_r) = F(x_r) - F(x_{r-1}), \quad r = 2, 3, \ldots$$

Proof:

(i) For $t < x_1$, the event $\{X \le t\}$ cannot happen since x_1 is the smallest of values that X can take on. So clearly

$$F(t) = P(X \le t) = P(\emptyset) = 0.$$

Now, let t be such that $x_{r-1} \le t < x_r$ for $r = 2, 3, \ldots$



Then the event $\{X \le t\}$ contains all elements ω of the sample space Ω for which

$$X(\omega) = x_1, \quad \text{or } X(\omega) = x_2, \ldots, \quad \text{or } X(\omega) = x_{r-1}.$$

In other words, the event $\{X \le t\}$ occurs if and only if X takes one of the values $x_1, x_2, \ldots, x_{r-1}$.

Since for $i = 1, 2, \ldots, r-1$, the events $A_i = \{\omega \in \Omega : X(\omega) = x_i\}$ are disjoint, we obtain

$$F(t) = P(X \le t) = P\left(\bigcup_{i=1}^{r-1} A_i\right) = \sum_{i=1}^{r-1} P(A_i) = \sum_{i=1}^{r-1} P(X = x_i) = \sum_{i=1}^{r-1} f(x_i).$$

(ii) The first relation is immediate by a direct application of

$$F(t) = f(x_1), \quad x_1 \le t < x_2$$

for $t = x_1$. For the second, we notice that for $r = 2, 3, \ldots,$

$$F(x_{r-1}) = \sum_{i=1}^{r-1} f(x_i), \quad F(x_r) = \sum_{i=1}^{r} f(x_i),$$

so that

$$F(x_r) = \sum_{i=1}^{r-1} f(x_i) + f(x_r) = F(x_{r-1}) + f(x_r),$$

and the result now follows immediately. □

Figures 4.5 and 4.6 illustrate graphically the result of Proposition 4.4. Next, we mention that the formula in Part (i) of Proposition 4.4, which gives the distribution function F in

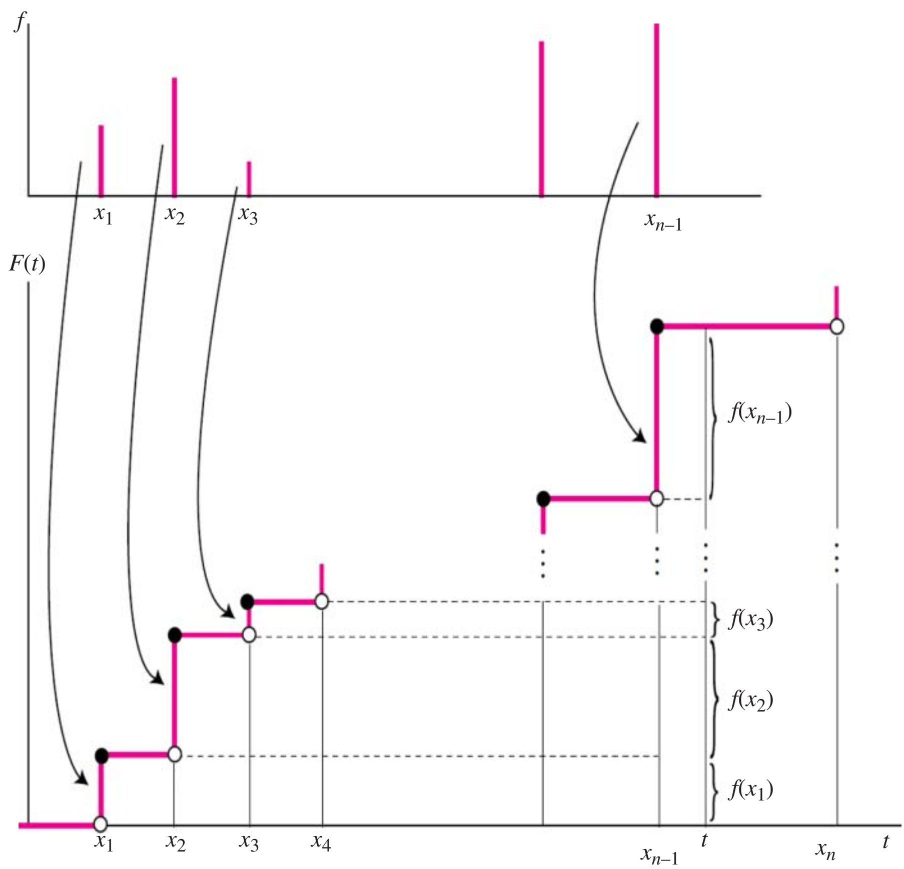


Figure 4.5 From the probability function to the distribution function.

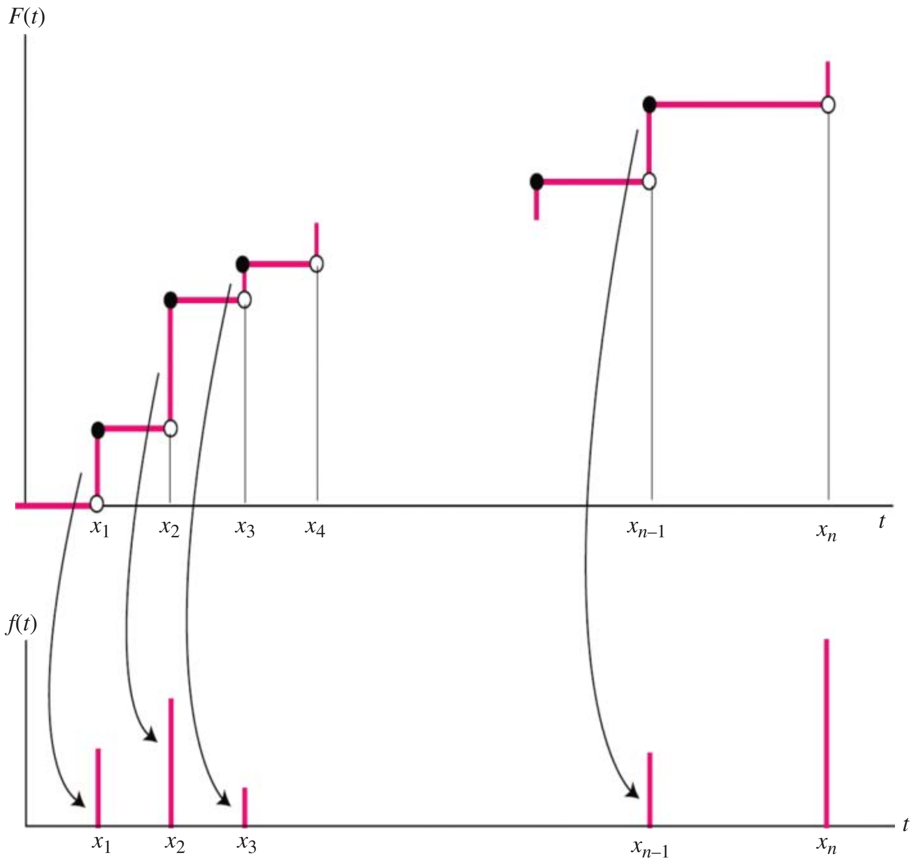


Figure 4.6 From the distribution function to the probability function.

terms of the probability function f, can be written in a more compact form as follows:

$$F(t) = \sum_{i:x_i \le t} f(x_i) = \sum_{x \in R_X : x \le t} f(x),$$

where the index set $\{i : x_i \le t\}$ in the first sum means that the summation extends over all indices i for which the inequality $x_i \le t$ holds. Similarly, in cases where we want to find the probability that the random variable X belongs to a certain subset A of its range R_X, we can write

$$P(X \in A) = \sum_{i:x_i \in A} f(x_i) = \sum_{x \in A} f(x).$$

Example 4.10 A bowl contains n artifacts numbered $1, 2, \ldots, n$, where the numbering is according to the date of their origin (1 is the one which was made first, etc.). We select at random k from these items without replacement. Let X be the largest number on the artifacts drawn.

(i) Find the distribution function of the random variable X.

(ii) Show that the probability function of X can be written in the form

$$f(x) = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}, \quad x = k, k+1, \dots, n.$$

SOLUTION

(i) For integer n, the event $\{X \le x\}$ occurs if and only if the largest number among the k numbers drawn is less than or equal to x, for $x = 1, 2, \dots, n$. Thus, the favorable outcomes for this event are all outcomes such that the k selected numbers belong to the set $\{1, 2, \dots, x\}$ and there are clearly $\binom{x}{k}$ such selections (of course, this is equal to zero if $k > x$). Since the total number of outcomes in this experiment is the number of combinations of k elements from a set of n elements, it readily follows that

$$F(x) = P(X \le x) = \frac{\binom{x}{k}}{\binom{n}{k}}, \quad \text{for } x = k, k+1, \dots, n.$$

For $t < k$, we have $F(t) = 0$, while for any real t such that $t \ge k$ and $t \ne k, k+1, k+2, \dots, n$, the event $\{X \le t\}$ is equivalent to the event $\{X \le x\}$, where x is the largest integer not exceeding t, i.e. $x = [t]$. We thus arrive at the formula

$$F(t) = \begin{cases} 0, & t < k \\ \dfrac{\binom{[t]}{k}}{\binom{n}{k}}, & k \le t \le n. \end{cases}$$

Finally, $F(t) = 1$ if $t > n$.

(ii) Using Proposition 4.4, we find

$$f(k) = F(k) = \frac{\binom{k}{k}}{\binom{n}{k}} = \frac{1}{\binom{n}{k}}$$

and

$$f(x) = F(x) - F(x-1) = \frac{\binom{x}{k} - \binom{x-1}{k}}{\binom{n}{k}}, \quad x = k+1, k+2, \dots, n.$$

From Pascal's triangle (see Proposition 2.7), we now get

$$f(x) = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}, \quad \text{for } x = k+1, \dots, n.$$

Observe that for $x = k$, we have $\binom{x-1}{k-1} = 1$, so that the two cases above can be written in a unified way as

$$f(x) = \frac{\binom{x-1}{k-1}}{\binom{n}{k}}, \quad \text{for } x = k, k+1, \dots, n,$$

as desired.

It is worth noting that the last expression could also be obtained directly as follows: the event $\{X = x\}$, for $x = k, k+1, \dots, n$ means that one of the numbers selected is necessarily the number x while the remaining $k-1$ numbers are selected from the set $\{1, 2, \dots, x-1\}$. Since there are $\binom{x-1}{k-1}$ ways to select these $k-1$ numbers, we arrive immediately at the result.

EXERCISES

Group A

1. Explain which of the following functions can be used as probability functions of a random variable X with the corresponding range R_X given below:

   (i) $f(x) = \dfrac{x+3}{15}, \quad R_X = \{1, 2, 3\};$

   (ii) $f(x) = \dfrac{2x-3}{16}, \quad R_X = \{1, 2, 3, 4\};$

   (iii) $f(x) = \dfrac{3^{x-1}}{4^x}, \quad R_X = \{1, 2, \dots\};$

   (iv) $f(x) = \dfrac{x^2}{n(n+1)(2n+1)}, \quad R_X = \{1, 2, \dots, n\};$

   (v) $f(x) = (-1)^{x-1}\binom{10}{x}, \quad R_X = \{1, 2, \dots, 10\};$

   (vi) $f(x) = 2^{-n}\binom{n}{x}, \quad R_X = \{0, 1, 2, \dots, n\}.$

2. In each of the following cases, find the value of the constant c so that the functions given below can be used as probability functions of a random variable X on the range R_X provided for each case:

   (i) $f(x) = c(x+5), \quad R_X = \{0, 1, 2, 3, 4\};$

   (ii) $f(x) = c(2x+1), \quad R_X = \{1, 2, \dots, 10\};$

(iii) $f(x) = cx, \quad R_X = \{1, 2, \ldots, n\}$;

(iv) $f(x) = \dfrac{c \cdot 4^x}{7^x}, \quad R_X = \{1, 2, \ldots\}$;

(v) $f(x) = c(x+3)^2, \quad R_X = \{-2, -1, 0, 1\}$;

(vi) $f(x) = c \cdot 2^x, \quad R_X = \{-10, -9, \ldots, -2, -1\}$;

(vii) $f(x) = c \cdot \dbinom{n}{x} 2^x, \quad R_X = \{0, 1, 2, \ldots, n\}$;

(viii) $f(x) = \dfrac{c \cdot 4^x}{x!}, \quad R_X = \{0, 1, 2, \ldots\}$.

3. The probability function of a random variable X is given by

$$f(x) = \frac{x}{10}, \quad x = 1, 2, 3, 4.$$

Find the distribution function F of X and plot this function.

4. In the random experiment of throwing two dice, let X be the random variable that represents the sum of the two outcomes. Show that the probability function of X is given by

$$f(x) = \begin{cases} \dfrac{x-1}{36}, & x = 2, 3, \ldots, 7, \\ \dfrac{12-(x-1)}{36}, & x = 8, 9, \ldots, 12. \end{cases}$$

5. Let X be a discrete random variable whose range is the set of positive integers and which has probability function

$$f(x) = \frac{4 \cdot 3^{x-1}}{7^x}, \quad x = 1, 2, 3, \ldots$$

(i) Verify that this function satisfies Properties (PF1), (PF2), and (PF3) of Proposition 4.3, so that it is indeed a probability function.

(ii) Find the distribution function $F(t)$ of the variable X.

(iii) Show that for any positive integers n, k, we have

$$P(X > n + k | X > n) = P(X > k).$$

(iv) Calculate the probability that X takes on an even value.

6. We are going to select at random a family of three children. Let X be the random variable that denotes the number of girls minus the number of boys in that family. Find the range of values for X and determine the probability function and the distribution function associated with it.

7. We toss a coin four times in succession. Let X be the number of times that the ordered pair HT appears in this experiment. Find the probability function and the distribution function of X.

8. From a box that contains 20 balls numbered 1–20, we select 3 balls without replacement. Let X be the largest number among the balls drawn.

(i) Find the probability function of X.

(ii) Nick bets that the largest number in the 3 balls drawn will be at least 17. What is the probability that he wins this bet?

9. Sofia is taking three exams at the end of this semester, one in each of the following courses: Mathematics, Statistics, and Economics. She estimates that the probability of passing the Mathematics exam is 0.85, the Statistics exam 0.75, and the Economics exam 0.90. If the three events "*Success in subject i*" are completely independent, where i takes the values "Mathematics," "Statistics," and "Economics," find the probability function of X, denoting the number of passes in her exams.

10. Let f_1 and f_2 be the probability functions of two discrete random variables with the same range R_X and $0 \le \lambda \le 1$ be a real number. Verify that the function

$$f(x) = \lambda f_1(x) + (1 - \lambda) f_2(x), \quad x \in R_X,$$

defines a probability function on R_X. Observe that this is the probability function of the two-component mixture distribution mentioned in Exercise 12 of Section 4.2.

Group B

11. In each case below, find the value of the constant c so that the corresponding function defines a probability function with the given range of values, R_X:

(i) $f(x) = \frac{c}{x(x+1)}, \quad R_X = \{1, 2, 3, \dots\}$;

(ii) $f(x) = c \cdot 2^{-x}, \quad R_X = \{0, 1, 2, \dots\}$;

(iii) $f(x) = c \cdot 2^{-|x|}, \quad R_X = \{0, \pm 1, \pm 2, \dots\}$.

(*Hint*: for Part (i), observe that $f(x)$ can be expressed as

$$f(x) = \frac{c}{x} - \frac{c}{x+1}.\Bigg)$$

12. In a population of microorganisms, we assume that each individual can produce $0, 1$, or 2 new microorganisms with probabilities $1/5, 3/5$, and $1/5$, respectively. Starting from a single individual (considered as the zeroth generation), let X_i be the number of microorganisms in the ith generation of this population.

(i) Find the probability function and the distribution function of X_1, the number of individuals in the first generation.

(ii) Derive the probability function of X_2, the second-generation individuals.

Here, the assumptions are that each individual is reproduced independently of any other, and that no individual of the ith generation is present in the $(i + 1)$th generation.

13. Empirical studies have suggested that the number of butterflies in a certain area is a random variable, denoted by X, with probability function

$$F(x) = \begin{cases} 0, & -\infty < x < 1, \\ c \sum_{k=1}^{[x]} \frac{\theta^k}{k}, & 1 \le x < \infty, \end{cases}$$

where, as usual, $[x]$ denotes the integer part of x and θ is a parameter such that $0 < \theta < 1$. Obtain the value of the constant c and the probability function

$$f(x) = P(X = x), \quad x = 1, 2, \ldots$$

$\Bigg($*Hint*: You may use the formula

$$\sum_{k=1}^{\infty} \frac{\theta^k}{k} = -\ln(1-\theta), \quad 0 < \theta < 1.\Bigg)$$

14. From a box which contains n lottery tickets numbered $0, 1, \ldots, n-1$, we select tickets *with replacement* until the number 0 appears. Let X be the number of selections made.
    (i) Find the probability function of X.
    (ii) What is the range and the probability function of the random variable $5X - 3$?

15. After a car accident, a hospitalized patient is in need of blood. Among five blood donors, only two have a blood type that matches the one needed for the patient. Find the probability function and the distribution function of the number of persons that will be examined until the first donor with the suitable blood type is found.

16. In a TV quiz show, a contestant is given the names of three countries and three capital cities and she is asked to match every country with its capital. If the contestant makes the correspondence completely at random, find the probability function of the correct matches.

4.4 EXPECTATION OF A DISCRETE RANDOM VARIABLE

We have already seen that early developments in probability theory were motivated primarily by games of chance. The concept of expectation, which is fundamental for random variables, is perhaps best understood in this context. Imagine that you are offered the chance to participate in the following game: you throw a die and, if the result is an even integer, you win \$20; otherwise, you lose \$10. The first dilemma you face is whether you should accept this game at all! The choice is clearly subjective, but an obvious consideration is whether you would be better off "on average" by playing the game. Suppose for instance we play this game 1000 times. Since there is an equal chance to win or lose, we expect intuitively to win 500 times and lose the rest. Our total profit would then be

$$500 \cdot 20 + 500 \cdot (-10)$$

dollars. Therefore, we expect our *profit per game* to be

$$\frac{500 \cdot 20 + 500 \cdot (-10)}{1000} = \frac{1}{2} \cdot 20 + \frac{1}{2} \cdot (-10) = \$5.$$

Now, suppose that the game above is referred to as Game A and you are offered a second game, say Game B. In this, you either win \$15 if the outcome of the die is 5 or 6; otherwise, you lose \$5. If you had to choose between games A and B, which one would you prefer? Neither game seems to be “uniformly” better than the other and the choice might depend on how much you are prepared to gamble. Someone who is afraid of losing \$10 might opt for Game B. In search of a criterion to rank the games from the player’s perspective, we consider again how much the player is expected to win “on average” from the game. Since the probability of winning in Game B is $2/6 = 1/3$ and the probability of losing is $1 - 1/3 = 2/3$, arguing as we did with Game A, we find that the player’s expected profit for Game B is

$$\frac{1}{3} \cdot 15 + \frac{2}{3} \cdot (-5) = \$1.667.$$

Thus, the gain here is smaller than that of Game A and, unless someone cannot afford the loss of \$10, they should choose the first game.

The above idea of an “average” profit is rather heuristic, but seems to agree with our intuition. In the first case, when winning and losing are equiprobable events, we use the arithmetic mean of the potential win and loss to find that average. In the second case, the two events do not have the same probability and so we use the *weighted average* of the winning and losing amounts, with the corresponding probabilities of these events as the respective weights.

The concept of a (possibly, weighted) average is called the *expectation of a random variable*. In the examples above, the variable was the financial win or loss from a chance game. In the following definition, however, this quantity is defined for an arbitrary (discrete) random variable. The analogous definition for continuous variables is given in Chapter 6.

Definition 4.4 Let X be a discrete random variable with probability function $f(x) = P(X = x)$ for any $x \in R_X$, and $f(x) = 0$ otherwise. Then **the expected value (or the expectation)** of X is denoted by $E(X)$ and is defined by

$$E(X) = \sum_{x \in R_X} x f(x),$$

provided that the series converges absolutely. If the series does not converge absolutely, we say that the expectation of X *does not exist*.

Another term that is used for the expectation of a random variable X is **the mean** of X. This reminds us of the arithmetic mean used in statistics to find the average among a number of observations, and is in accordance with the calculation for Game A presented earlier. In fact, another interpretation of the above definition arises if we consider the probability as the limit of a relative frequency. Example 4.11 below illustrates this point.

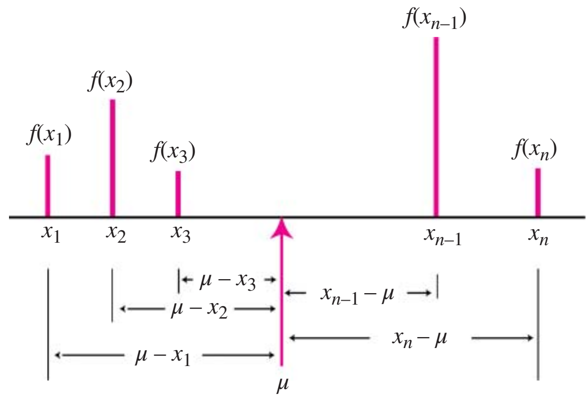


Figure 4.7 Physical interpretation of expectation.

Moreover, to parallel the customary notation in statistics, a symbol that is often used for the expectation of a random variable is the Greek letter μ, provided no ambiguity arises. In the case when we consider more than one random variables in a problem, say X, Y, Z, we use the notation μ_X, μ_Y, μ_Z for the means of X, Y, Z, respectively, so that each variable is represented by a subscript to avoid any confusion.

If the random variable X admits only a finite number of distinct values, say $x_1, x_2, \ldots, x_n$, then the expectation of X has the following physical interpretation, which some readers may find insightful. We assume that at the positions $x_1, x_2, \ldots, x_n$ of an axis, we place masses equal to their respective probabilities, i.e. the values of the associated probability function, $f(x_1), f(x_2), \ldots, f(x_n)$ (see Figure 4.7).

Then, the expectation of X corresponds to the coordinate $\mu = E(X)$, where the axis should be supported so that the entire system is balanced. This is made clear if we denote by

$$m_i = f(x_i), \quad i = 1, 2, \ldots, n,$$

the n masses and put

$$r_i = x_i - \mu, \quad i = 1, 2, \ldots, n,$$

for the distances between the locations of each mass and $\mu = E(X)$. Then we may write the equality

$$\mu = \sum_{i=1}^{n} x_i f(x_i)$$

in the form

$$\sum_{i=1}^{n} (x_i - \mu) f(x_i) = 0$$

or equivalently,

$$\sum_{i=1}^{n} r_i m_i = 0.$$

This is, in fact, the balancing condition for the system of the n masses.

Example 4.11 In the game of bowling, a player rolls a bowling ball trying to bring down as many pins (out of a total of 10 pins) as possible. In each frame, a player has two shots so that if she knocks down six pins with the first ball and three pins with the second, she scores nine points. In order to monitor their performance, a bowling team kept records of their scores in a total of 10 000 frames during a month of practice, and these are as follows:

Score	Frequency of that score
0	399
1	592
2	651
3	707
4	1023
5	1114
6	1085
7	1251
8	1307
9	1123
10	748

To summarize the information contained in the table, and to obtain a measure of the team's overall performance, we calculate the average score per frame, which is simply found as a weighted average of the scores, with the frequency being the weight for each score. In this way, we find that

$$\frac{0 \cdot 399 + 1 \cdot 592 + 2 \cdot 651 + \cdots + 10 \cdot 748}{10\,000} = \frac{56\,987}{10\,000} = 5.70 \tag{4.4}$$

pins were knocked down on average for each frame.

In statistics, this is called the *sample mean* for this experiment. The interpretation of this value *is not* that if a bowler starts a new frame we expect that she will knock down 5.70 pins (which is clearly impossible), but that if a large number, say n, of frames are played by this team, we expect the average number of pins to be knocked down per frame to be close (but not necessarily equal) to 5.7.

Suppose now that, based on these data, we want to estimate the probability that in a future experiment, a bowler knocks down nine pins in a single frame. According to the frequentist (or statistical), definition of probability that we gave earlier in Chapter 1 (see Definition 1.9), this probability is the limit of the *relative frequency* of the event

$$E_9 : \text{a bowler scores nine points in a single frame}$$

as the number of repetitions of a single experiment (here, a frame) tends to infinity. Hence, we can give an estimate for the probability of the event E_9 to be

$$\frac{1123}{10\,000} = 0.1123,$$

and since 10 000 is a rather large number of repetitions, we are fairly confident that this will not be too far from the true value of the probability. The same argument clearly works if we replace the score of 9 points by any other integer from 0 to 10. Specifically, let

$$E_i : \text{a bowler scores } i \text{ points in a single frame}$$

for $i = 0, 1, 2, \ldots, 10$.

Note that the calculation in (4.4) can be written in an equivalent way as

$$0 \cdot \frac{399}{10\ 000} + 1 \cdot \frac{592}{10\ 000} + 2 \cdot \frac{651}{10\ 000} + \cdots + 10 \cdot \frac{748}{10\ 000} = \sum_{i=0}^{10} i \cdot \frac{n_{E_i}}{n},$$

where n_{E_i} is the observed frequency of the event E_i in the total of $n = 10\ 000$ repetitions. Since n is large, we may assume that the relative frequencies

$$f_{E_i} = \frac{n_{E_i}}{n}$$

are quite close to the true probabilities $P(E_i)$ for $i = 0, 1, 2, \ldots, 10$, and if this is indeed the case, the weighted sum

$$\mu_{10\ 000} = \sum_{i=0}^{10} i f_{E_i}$$

should be quite close to the quantity $\sum_{i=0}^{10} iP(E_i)$; here, the subscript 10 000 indicates the number of repetitions on which the relative frequencies were calculated. If now X is a random variable that stands for the number of points scored in a single frame, replacing i by x in the last expression we see that, as the number of frames played tends to infinity, the quantity given in that formula approaches the limit [2]

$$\mu = \lim_{n\to\infty} \mu_n = \sum_{x=0}^{10} xP(E_x) = \sum_{x\in R_X} xP(X = x).$$

This agrees with Definition 4.4 for the expectation of a discrete variable.

Example 4.12 Diane has an arts and crafts internet home business. A few months after she had started her business, she estimated that the number of items she sells per day ranges from one to five, with the respective probabilities as follows.

[2] Historically, this reasoning was essentially used by Christian Huygens (1629–1695) in order to define the expectation of a discrete random variable in his book "*Libellus de ratiociniis in ludo aleae*" ("Reasoning on the games of chance"), published in 1657. This is considered as the first book on probability theory. The French mathematician Blaise Pascal had in fact used the concept of expectation earlier, but in a less rigorous manner.

Number of items sold	Probability
1	0.25
2	0.15
3	0.35
4	0.15
5	0.10

What is the expected number of items she sells during a day?

SOLUTION Let X be the random variable that represents the items sold in a day. Then, X is a discrete random variable whose range is $R_X = \{1, 2, 3, 4, 5\}$. Let $f(x) = P(X = x)$ be the probability function of X. Then, from Definition 4.4, we obtain

$$E(X) = \sum_{x \in R_X} xf(x) = \sum_{x=1}^{5} xf(x)$$
$$= 1 \cdot (0.25) + 2 \cdot (0.15) + 3 \cdot (0.35) + 4 \cdot (0.15) + 5 \cdot (0.1) = 2.7.$$

This reveals an important point to keep in mind about expectations of discrete random variables: even in the case when X is integer-valued, its expectation need not be an integer.

Example 4.13 For a random experiment with sample space Ω and an event $A \subseteq \Omega$, let the random variable X be defined by

$$X = \begin{cases} 1, & \text{if } A \text{ occurs,} \\ 0, & \text{otherwise} \end{cases}$$

Then X is a discrete random variable (its range of values is $R_X = \{0, 1\}$), and the expectation of X equals

$$E(X) = 1 \cdot P(X = 1) + 0 \cdot P(X = 0) = P(A).$$

This variable X is often referred to as the **indicator function** of the set A and is sometimes denoted by I_A. Using this notation, we therefore have

$$E(I_A) = P(A).$$

Example 4.14 Let X be a random variable that takes on the values $a_1, a_2, \dots, a_n$ with equal probabilities, namely,

$$f(x) = P(X = x) = \frac{1}{n}, \quad x \in R_X = \{a_1, a_2, \dots, a_n\}$$

(the distribution of such a random variable is called a **discrete uniform** distribution on the set R_X). In this case, we have

$$\mu = E(X) = \sum_{x \in R_X} xf(x) = \sum_{x \in R_X} x \cdot \frac{1}{n} = \frac{1}{n} \sum_{x \in R_X} x = \frac{a_1 + a_2 + \cdots + a_n}{n},$$

that means that the expectation of X coincides with the usual notion of the arithmetic mean. For this reason, we may regard the expectation of a random variable as a generalization of the arithmetic mean for a finite set of numbers.

As an application, we consider the random variable X that denotes the outcome in a single throw of a die. Then clearly $R_X = \{1, 2, 3, 4, 5, 6\}$ and $f(x) = P(X = x) = 1/6$ for $x \in R_X$. In this case, the expectation is

$$\mu = E(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{7}{2} = 3.5. \tag{4.5}$$

Example 4.15 Let X be a random variable with range $R_X = \{1, 2, \dots\}$ and for which

$$P(X > x) = \frac{c}{x + 1}, \quad x = 0, 1, 2, \dots, \tag{4.6}$$

where c is a real constant.

(i) Find the value of c.

(ii) Prove that the expectation of X does not exist.

SOLUTION

(i) In order to find the value of c, we first obtain the probability function, f, of X and then use the condition that

$$\sum_{x \in R_X} f(x) = 1. \tag{4.7}$$

For the probability function f, we observe initially that for $x = 1, 2, \ldots,$

$$f(x) = P(X = x) = P(X > x - 1) - P(X > x)$$

which readily gives

$$f(x) = \frac{c}{x} - \frac{c}{x+1}, \quad x = 1, 2, \ldots$$

Therefore,

$$\sum_{x=1}^{n} f(x) = \left(\frac{c}{1} - \frac{c}{2}\right) + \left(\frac{c}{2} - \frac{c}{3}\right) + \left(\frac{c}{3} - \frac{c}{4}\right) + \cdots + \left(\frac{c}{n} - \frac{c}{n+1}\right)$$
$$= c - \frac{c}{n+1}.$$

By an appeal to (4.7), we then deduce

$$1 = \sum_{x=1}^{\infty} f(x) = \lim_{n\to\infty} \sum_{x=1}^{n} f(x) = c - \lim_{n\to\infty} \frac{c}{n+1}$$

so that the required value of c is $c = 1$.

(ii) From Part (i) we have that

$$f(x) = P(X > x - 1) - P(X > x) = \frac{1}{x} - \frac{1}{x+1} = \frac{1}{x(x+1)}$$

and, in order to find $E(X)$, we have to calculate the infinite sum

$$\sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} \frac{1}{x+1} = \sum_{r=2}^{\infty} \frac{1}{r}, \tag{4.8}$$

making a change of variable in the last step. However, it is well known from calculus that the series

$$\sum_{r=1}^{\infty} \frac{1}{r^{\rho}}$$

converges only if $\rho > 1$. Thus, the series in (4.8) diverges (the sum is infinite) and, according to Definition 4.4, the expectation of X does not exist.

The reason why this happens ($E(X)$ is infinite) is that the probabilities $f(x)$ do not decrease "quickly enough" for large values of x (as $x \to \infty$). Distributions with this property are in the class of the so-called **heavy-tailed** distributions, although this class also contains some distributions with a finite expectation. Heavy-tailed distributions occur frequently in a number of applied fields, such as actuarial science, economics, engineering, and geology.

Remark A much quicker way to find the value of c in (i) can be provided by the observation that (4.6) holds for $x = 0, 1, 2, \ldots$, while the range of values for X is the set of *positive integers*. This means that the probability $P(X > 0)$ must be equal to one, so that putting $x = 0$ in (4.6) we obtain immediately $c = 1$. In the solution above, we took the much longer path via finding the probability function f, since we needed this function for Part (ii).

If a random variable X can take only one value, say c, with probability one, then one expects that its expected value is also c. As shown formally next, this is an easy deduction from Definition 4.4.

Proposition 4.5 *If X is a random variable such that $P(X = c) = 1$ for some real number c, then $E(X) = c$.*

Proof: Here we have $R_X = \{c\}$, while the probability function of X is

$$f(x) = \begin{cases} 1, & x = c, \\ 0, & x \neq c, \end{cases}$$

so that we obtain immediately

$$E(X) = \sum_{x \in R_X} x f(x) = c \cdot f(c) = c. \qquad \square$$

Often in probability theory, we are interested in the expectation for a *function* of a random variable. For instance, in (4.5), we found the expected value of X, the outcome of the throw of a die. Suppose that each possible outcome, x, is associated with a potential gain or loss, and this is represented by a function $g(x)$. For example, if x is 1 or 2, we lose \$5, if it is 6, we win \$15, and otherwise we do not win or lose anything. In such a case, and in order to find our expected gain from this game, we need to consider a random variable $g(X)$ which represents the gain, and then calculate its expected value.

More generally, let X be a random variable defined on a sample space Ω and assume that $g : \mathbb{R} \mapsto \mathbb{R}$ is a real function. Since we have seen that, in mathematical terms, X is also a function, $X : \Omega \mapsto \mathbb{R}$, we consider the composition of the two functions

$$(g \circ X)(\omega) = g(X(\omega)), \quad \omega \in \Omega,$$

$$\Omega \xrightarrow{X} \mathbb{R} \xrightarrow{g} \mathbb{R}.$$

It follows from the definition of a random variable that $g \circ X$, hereafter denoted by $g(X)$, is also a random variable whose range is the set $\{y : \text{there exists } x \in R_X \text{ with } g(x) = y\}$.

Example 4.16 Nick has invested \$1000 in each of two products. He expects that each successful investment will yield a profit of \$800, while if an investment is unsuccessful he will lose his money. A successful outcome in each investment is independent of the other investment being successful. If he estimates that the probability of success in each of the two investments is 0.6, find the probability function and the expected value of Nick's profit.

SOLUTION Let X be the number of investments that turn out to be successful. Then, clearly, the range of values for X is $R_X = \{0, 1, 2\}$ where, for example, $X = 0$ if both of Nick's investments result in a failure. Since the questions in the example are about his profit, we denote this profit (which is a random variable) by Y and try to find a relationship between X and Y. For each of the X successful investments, Nick receives a total of $1000 + 800 = \$1800$, so that he receives in total $1800X$ dollars at the end of his venture. But he has paid \$2000 for investing in the two products, and so his profit will be

$$Y = 1800X - 2000. \tag{4.9}$$

In order to find the probability function of Y, we first find that of X. For convenience, we label the two investments as 1 and 2, and let S_1 and S_2 be the events that the first and the second investments are successful. Then, the event $\{X = 2\}$ occurs if and only if both S_1 and S_2 occur and since these two events are independent, we have

$$P(X = 2) = P(S_1S_2) = P(S_1)P(S_2) = (0.6)(0.6) = 0.36.$$

Similarly, let F_i be the event that the ith investment is a failure, for $i = 1, 2$; then, F_1 and F_2 are also independent events and their intersection, F_1F_2, is equivalent to the event $\{X = 0\}$. Thus,

$$P(X = 0) = P(F_1F_2) = P(F_1)P(F_2) = (0.4)(0.4) = 0.16.$$

There now remains only one possible value for X, namely, $X = 1$. So, the probability of that event must be

$$P(X = 1) = 1 - P(X = 2) - P(X = 0) = 1 - 0.36 - 0.16 = 0.48.$$

(Alternatively, and arguing directly, observe that $\{X = 1\}$ occurs if and only if exactly one of the events S_1F_2 and F_1S_2 occur, so that $\{X = 1\} = (S_1F_2) \cup (F_1S_2)$. Work out the details to get the same result as above.)

In summary, the probability function of X is as given below:

x	$f(x)$
0	0.16
1	0.48
2	0.36

In order to find the probability function of Y (Nick's profit), we may ask ourselves first about the range of values for Y. Since X and Y are related through (4.9), and X can take on only three values, the number of possible values for Y is also three, and these values are

$$1800 \cdot 0 - 2000 = -2000, \quad 1800 \cdot 1 - 2000 = -200, \quad 1800 \cdot 2 - 2000 = 1600,$$

so that $R_Y = \{-2000, -200, 1600\}$. Next, notice that the event $\{Y = -2000\}$ occurs if and only if the event $\{X = 0\}$ occurs, and so these two events have the same probability. Similarly, $P(Y = -200) = P(X = 1)$ and $P(Y = 1600) = P(X = 2)$. We summarize the probability function of Y, denoted by $h(y)$, in the following table.

y	$h(y)$
−2000	0.16
−200	0.48
1600	0.36

It seems apparent from the two tables of probabilities that the distributions of X and Y are closely related. In fact, the second columns in the two tables are identical. If X takes the value x with a positive probability $f(x)$, then Y takes the value $1800x - 2000$ with the same probability. By a straightforward application of the formula in Definition 4.4, we thus see that Nick's expected profit is

$$E(Y) = \sum_{y \in R_Y} yh(y) = (-2000) \cdot (0.16) + (-200) \cdot (0.48) + 1600 \cdot (0.36) = 160$$

dollars.

From the above example we observe that, if we want to calculate the probability function of the random variable $Y = g(X)$ using the probability function f of the variable X, for each $y \in R_Y$ we have to find the values $x \in R_X$ for which $g(x) = y$ and then add up the associated probabilities $P(X = x) = f(x)$, namely,

$$f_Y(y) = P(Y = y) = P(g(X) = y) = \sum_{x \in R_X : g(x) = y} P(X = x), \quad y \in R_Y. \tag{4.10}$$

We can now state the following result, which gives a formula for calculating the expected value for a function of a discrete random variable.

Proposition 4.6 ***(Expectation for a function of a discrete random variable)*** *Let X be a discrete random variable with probability function f and range R_X. Then, given a real-valued function g, the expectation of the random variable $g(X)$ can be found as*

$$E[g(X)] = \sum_{x \in R_X} g(x)f(x).$$

Proof: Let $Y = g(X)$ and denote by f_Y the probability function of Y. Then, Definition 4.4 gives

$$E[g(X)] = E(Y) = \sum_{y \in R_Y} yf(y).$$

Replacing $f_Y(y)$ from (4.10) in the above formula, we obtain

$$\begin{aligned} E[g(X)] &= \sum_{y \in R_Y} y \sum_{x \in R_X : g(x)=y} f(x) = \sum_{y \in R_Y} \sum_{x \in R_X : g(x)=y} yf(x) \\ &= \sum_{y \in R_Y} \sum_{x \in R_X : g(x)=y} g(x)f(x), \end{aligned}$$

and the desired result follows immediately by observing that the two sums in the last expression can be written as a single sum that extends over all $x \in R_X$; this is because the set $\{x \in R_X : g(x) = y\}$ includes all x-values associated with a certain $y \in R_Y$. Considering now all $y \in R_Y$, we see that any $x \in R_X$ must be associated with one of them. □

The next result is a special case of Proposition 4.6 and it refers to linear combinations of functions of a random variable X.

Proposition 4.7 *Assume that X is a random variable, $\lambda_1, \lambda_2, \ldots, \lambda_k$ are real numbers and $g_1, g_2, \ldots, g_k$ are real-valued functions defined on $\mathbb{R}$. Then, we have*

$$E[\lambda_1 g_1(X) + \cdots + \lambda_k g_k(X)] = \lambda_1 E[g_1(X)] + \cdots + \lambda_k E[g_k(X)].$$

Proof: We define the function

$$g = \lambda_1 g_1 + \lambda_2 g_2 + \cdots + \lambda_k g_k.$$

Proposition 4.6 then yields

$$\begin{aligned} &E[\lambda_1 g_1(X) + \cdots + \lambda_k g_k(X)] = E[g(X)] \\ &= \sum_{x \in R_X} g(x)f(x) = \sum_{x \in R_X} [\lambda_1 g_1(x) + \cdots + \lambda_k g_k(x)]f(x) \\ &= \lambda_1 \sum_{x \in R_X} g_1(x)f(x) + \cdots + \lambda_k \sum_{x \in R_X} g_k(x)f(x) \\ &= \lambda_1 E[g_1(X)] + \cdots + \lambda_k E[g_k(X)], \end{aligned}$$

as required. □

Combining the results of Propositions 4.5 and 4.7, we obtain the following, which is usually referred to as the *linearity of expectation*.

Corollary 4.1 *For any discrete random variable X and real numbers a and b, we have*

$$E(aX + b) = aE(X) + b.$$

Using the results of the present section about the expected value of a discrete variable X, we have, for example, the following equalities for $a, b, c \in \mathbb{R}$:

$$E(aX^2 + bX + c) = aE(X^2) + bE(X) + c,$$
$$E(\ln X + 2e^{aX} - \cos X) = E(\ln X) + 2E(e^{aX}) - E(\cos X),$$
$$E[(X-a)^2] = E(X^2 - 2aX + a^2) = E(X^2) - 2aE(X) + a^2,$$

and so on.

On the other hand, it is important to note that it is generally **not true** that

$$E[g(X)] = g(E(X))$$

for an arbitrary function g, and so in particular the result of Corollary 4.1 does not extend to non-linear functions g. We note for instance that, if X is a random variable which is not concentrated on one point, i.e. there is no c such that $P(X = c) = 1$, then

$$E(X^2) \neq [E(X)]^2, \quad E(\ln X) \neq \ln E(X), \quad E(\cos X) \neq \cos(E(X)).$$

Example 4.17 The number of notebooks sold in a small computer store during a week is described by a random variable X having probability function

$$f(x) = \frac{2x+3}{63}, \quad x \in R_X = \{0, 1, 2, 3, 4, 5, 6\}.$$

(i) Find the expected number of notebooks sold by the store in a given one-week period.

(ii) In order to have enough stock, the store orders every week from its supplier 6 notebooks at a price of \$250 each, under the following agreement: the new notebooks arrive at the store on Monday morning and any notebook not sold during the week can be returned to the supplier at the price of \$210. If the store sells notebooks at a price of \$325, find the store's expected profit during a week.

SOLUTION

(i) Since X is a discrete random variable, we find its expectation by using the formula from Definition 4.4. This gives

$$E(X) = \sum_{0}^{6} x f(x) = \frac{1}{63}\left(2\sum_{x=0}^{6} x^2 + 3\sum_{x=0}^{6} x\right)$$
$$= \frac{1}{63}\left(2\frac{6 \cdot 7 \cdot (2 \cdot 6 + 1)}{6} + 3\frac{6 \cdot 7}{2}\right) = \frac{35}{9},$$

where we have used the formulas

$$\sum_{x=0}^{n} x = \sum_{x=1}^{n} x = \frac{n(n+1)}{2}, \quad \sum_{x=0}^{n} x^2 = \sum_{x=1}^{n} x^2 = \frac{n(n+1)(2n+1)}{6}.$$

(ii) We want to find the expected weekly profit of the store. For this purpose, let Y be the random variable that represents this profit. It is clear that Y is a function of X, the number of notebooks sold. More precisely, if X takes the value $x \in R_X$, then each of the items sold yields a profit of $325 - 250 = \$75$, while there are $6 - x$ unsold items, and each of these corresponds to a loss of $250 - 210 = \$40$ for the store.

In view of the above, we see that Y can be written in terms of X as

$$Y = 75X - 40(6 - X) = 115X - 240.$$

It is now straightforward to obtain the expected value of Y using that of X, already found in Part (i), and the linearity of expectation. In fact, this is

$$E(Y) = E(115X - 240) = 115E(X) - 240 = 115 \cdot \frac{35}{9} - 240 = \frac{1865}{9} \cong 207.22$$

dollars.

Example 4.18 For a random variable X, it is given that $E(X) = 2$ and $E(X^2) = 8$. Calculate the expected value for each of the following random variables:

$$Y = (2X - 3)^2, \quad W = X(X - 1), \quad Z = X^2 + (X + 1)^2.$$

SOLUTION First, we have for the variable Y,

$$\begin{aligned} E(Y) &= E[(2X - 3)^2] = E(4X^2 - 12X + 9) \\ &= 4E(X^2) - 12E(X) + 9 = 4 \cdot 8 - 12 \cdot 2 + 9 = 17. \end{aligned}$$

Similarly for W we get

$$E[X(X - 1)] = E(X^2 - X) = E(X^2) - E(X) = 8 - 2 = 6,$$

and, finally, for the variable Z we find

$$\begin{aligned} E(Z) &= E[X^2 + (X + 1)^2] = E[X^2 + X^2 + 2X + 1] = E[2X^2 + 2X + 1] \\ &= 2E(X^2) + 2E(X) + 1 = 2 \cdot 8 + 2 \cdot 2 + 1 = 21. \end{aligned}$$

The last two examples in this section are devoted to the important idea that a chance game, played between two competing players, is fair to both of them.

Example 4.19 (Fair and unfair games)
The concept of a fair game lies at the very heart not only of the theory of expectations that we consider in this chapter but also to a large extent of probability theory as a whole. In fact, in the first book on probability theory, which was published in 1657, the Dutch mathematician C. Huygens used this concept to find how the stakes should be divided between two competing players if they have unequal probabilities of winning a game.

To explain the idea of a fair game, we use a very simple example, the one we discussed in the beginning of this section (called Game A), which we treated rather informally there. The example we presented was the following: you throw a die and if the result in an even integer, you win \$20; otherwise, you lose \$10. The result we found, which can now be proved formally using the definition of expectation, is that you have an expected profit of \$5 each time you play this game. By the same token, your opponent has an expected loss of \$5 for each game. Obviously, this does not mean that you win every single time that you play against your opponent. If you play just once, you may lose. But this result assures you that *in the long run* you will have more money than what you started with, and your opponent will be losing. This game is clearly *unfair* to him, but it is favorable for you.

More generally, when two players gamble against each other, one's profit will be the other player's loss (negative profit) at each game, so that the total profit of the two players is certain to be zero. This, in turn, implies immediately that the sum of the expected profits will always be zero (in the game above, the expected profits are \$5 for you and -5 for your opponent). The game is in favor of one between the two players unless both expectations are zero, in which case it is called a *fair game*. For simplicity, it is sufficient to consider the expected profit for one of the two players; if his expected profit is zero, his opponent will automatically have the same. If one (hence both) of the expectations is nonzero, we refer to it as an unfair game.

In the example above regarding the throw of the die, if you win \$20 when the outcome is even, you must be prepared to pay *the same amount* when the outcome is odd in order for the game to be fair.

Example 4.20 Danny and Andrew will participate in an important baseball game next Saturday and they are prepared to bet against each other about the weather on that day. Andrew agrees to give Danny \$10 if it rains at the time of the game. They also agree that Andrew will receive c dollars if there is sunshine; if anything else happens (e.g. it is cloudy but no sunshine), there will be no exchange of money. If the probability of rain is 20% and the probability of sunshine is 50%, what should be the value of c so that the game is fair?

SOLUTION As explained in the previous example, it is sufficient to concentrate on one player's profit. Let X be Andrew's profit from this bet. Then, X can take only three values, namely, $-10, 0$, and c. From what is given in the example, we see that the values

c and -10 have respective probabilities 0.5 and 0.2, and so the probability that X takes the value 0 is $1 - 0.5 - 0.2 = 0.3$. Thus, the probability function of X is given by

$$f(x) = \begin{cases} 0.2, & x = -10 \\ 0.3, & x = 0 \\ 0.5, & x = c. \end{cases}$$

The expectation of X then equals

$$E(X) = \sum_{x \in R_X} x f(x) = (-10) \cdot (0.2) + 0 \cdot (0.3) + c \cdot (0.5) = c/2 - 2.$$

For the game to be fair, we must have $E(X) = 0$ that yields $c = 4$. Therefore, Danny should give Andrew \$4 in the event of sunshine so that the expected profit of either player is zero.

EXERCISES

Group A

1. Find the expected value of the random variable X in each of the three cases below, where $f(x)$ denotes the probability function of X:

(i)

x	0	1	2	3
$f(x)$	0.5	0.1	0.3	0.1

(ii)

x	1	2	3	4
$f(x)$	0.2	0.3	0.4	0.1

(iii)

x	2	4	6	8
$f(x)$	0.1	0.2	0.3	0.4

2. The probability that a technician fixes on a given day $0, 1, 2, 3, 4, 5$, and 6 electric appliances is $0.05, 0.10, 0.20, 0.25, 0.20, 0.10$, and 0.10, respectively. What is the expected number of appliances that the technician fixes in a day?

3. The number of boats that arrive at a Greek island during a day is $0, 1, 2, 3$, and 4 with probabilities $0.1, 0.3, 0.25, 0.15$, and 0.2, respectively. What is the expected number of boats that arrive during a day?

4. We toss three coins and suppose X denotes the number of heads that show up. Find the expected value of X.

5. Jimmy participates in a TV quiz show in which he is given two multiple choice questions. In the first one, he has three possible answers and in the second there are four possible answers.

   (i) If he selects his answer to both questions completely at random, what is the expected number of correct answers he will give?

   (ii) If he wins \$1000 for each correct answer he gives, what is the expected amount of money Jimmy will win in the show?

6. A gambler pays €10 to enter the following game: he throws two dice and if the outcomes contain

   - exactly one ace, he receives €15;
   - two aces, he receives €30.

   If no ace appears, he receives nothing. Find the gambler's expected profit from this game.

7. The number of new friends that Jane makes on *Facebook* during a day is a random variable Y with probability function

$$f(y) = \begin{cases} 0.15, & y = 1, \\ 0.25, & y = 2, \\ 0.30, & y = 3, \\ 0.30, & y = 4. \end{cases}$$

   Obtain the probability function and the expected value for the number of friends that she makes during a weekend, i.e. in a two-day period.

8. The roulette wheel in a casino has 38 slots. Two of the slots have a 0 and a 00, and the remaining slots have the integers from 1 to 36. Suppose that a player bets on red or black. Half of the numbers from 1 to 36 are red, the rest are black. The player places a bet of \$10 on red; if the ball stops on a red number, he receives his \$10 back plus \$10 more. If the ball stops on a black number, the 0 or the 00 slot, he loses his bet. Find the player's expected profit or loss from this game.

9. At a Christmas bazaar, 6000 lottery tickets were sold at a price of \$5 each. There are 15 prizes to be won from the lottery: a car with a value of \$15 000, three trips to exotic locations, each of which costs \$2500, a TV set whose value is \$700 and 10 mobile phones each of which costs \$250. Helena bought a ticket for this lottery. What is her expected "net profit" from the lottery draw?

10. A small store buys every week six 2-l bottles of milk at a price of \$1.25 each. If at the end of the week a bottle has not been sold, the store returns it to the provider and receives 30 cents back. The probabilities that the number of 2-l bottles sold during a week is $0, 1, 2, 3, 4, 5,$ and 6 equal $0.05, 0.10, 0.10, 0.20, 0.20, 0.20,$ and 0.15, respectively. What is the expected weekly profit of the store if the bottles are sold at a price of \$1.80?

11. In an army camp, a large number, N, of soldiers are subject to a blood test to examine whether they have a certain disease. There are two ways to carry out this experiment:

    A. each soldier is tested separately, so that N tests take place;

    B. the blood samples of k soldiers ($k < N$) can be pooled and analyzed together. If the test outcome is negative, this means that none of the soldiers in this group have the disease. If it is positive, each of the k persons is then sampled individually, so that $k + 1$ tests are required for a group of k people.[3]

    Assume that each soldier has a probability p of having the disease and that test outcomes for different soldiers are independent.

    (i) What is the probability that the test for a pooled sample of k soldiers will be positive?

    (ii) Find the expected value of the number, X, of blood tests required under Plan B above.

12. A bookstore buys 15 copies of a book at a price of \$20 each and sells them at a price of \$30. Suppose that, after a year, the bookstore may return any unsold copies of the book to the publisher at the original price of \$20. Let X be the number of copies sold during the year and assume that the probability function of X is given by

$$f(x) = \frac{x+2}{150}, \quad x = 1, 2, \ldots, 15.$$

    (i) Check that $f(x)$ is a proper probability function.

    (ii) Find the expected number of copies sold in a year.

    (iii) Find the expected profit of the bookstore from the sales of this particular book.

13. An insurance company has estimated, using previous data that, during a year, an insured person has a probability p of making a claim worth x thousands of dollars (i.e. all claims are of a fixed amount), and probability $1 - p$ of making no claims. What is the annual premium the company should charge per person so that the company's expected annual profit is \$100 for each person insured?

14. A friend of yours is proposing the following game: you put nine balls numbered $1, 2, \ldots, 9$ in a box and you select one at random. If the number on the ball drawn is a multiple of 3, he will give you \$100. How many dollars should you give him if the number of the ball is not divisible by 3 so that the game is fair?

15. Nick throws a die and if the outcome is 1 or 6, he gives Peter \$9, while if the outcome is 3 he gives him \$6. What amount should Peter pay Nick if the outcome is either 2 or 4 so that they have a fair game? (If the outcome of the die is 5, there is no money exchange between the two players.)

[3]Feller (1968) mentions that this technique was developed in World War II by R. Dorfman. In army practice, it achieved savings of up to 80%.

16. Bill plays the following game: he flips a coin four times. If the number of times that heads appear is 0 or 1, he wins \$6, if heads appear 3 or 4 times he loses \$12. What is the amount of money he should receive if heads appear twice so that this is a fair game?

17. A company wants to promote two new products, a and b. For this purpose, the company sends a salesman to visit a number of houses in order to find buyers for these products. If someone does not buy either of the two products, the company gives them a small gift that costs \$10. If a person buys product a, the company's profit is \$150, while for any purchase of product b, the company's profit is \$200. Finally, if the salesman makes a deal for both products, the profit for the company is \$300 (due to a discount made to the buyer in this case). The company estimates that the probability someone buys, after a visit by the salesman, product a alone is $7/30$, product b alone is $1/6$, and for both products it is $2/15$. What is the company's expected profit for each visit made by the salesman?

18. The probability function of a discrete random variable X has the form

$$f(x) = c(x^2 + 2|x| + 1), \quad x \in R_X = \{-2, -1, 0, 1, 2\},$$

for a suitable constant c.

(i) Find the value of c.

(ii) Calculate $E(X)$ and then find the expectation of the random variables $Y = |X|$ and $W = 5X - 3|X|$.

19. We consider a random variable X whose range is the set $\{1, 2, 3, \dots\}$ and which has probability function

$$f(x) = P(X = x) = \frac{1}{2^x}, \quad x = 1, 2, \dots$$

Consider also another variable Y defined as

$$Y = \begin{cases} 1, & \text{if the value of } X \text{ is even,} \\ -1, & \text{if the value of } X \text{ is odd.} \end{cases}$$

(i) Find the probability function and the distribution function of the variable Y.

(ii) Calculate $E(Y)$.

20. Suppose that a discrete random variable X has range $\{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$ and the following probability function

$$f_X(x) = \begin{cases} \dfrac{1}{5}, & x = 0, \\ \dfrac{1}{10}, & x = \pm 1, \pm 2, \pm 3, \pm 4. \end{cases}$$

(i) Show that the random variable $Y = |X|$ has probability function

$$f_Y(y) = \frac{1}{5}, \quad x = 0, 1, 2, 3, 4.$$

(ii) Find the expectation for each of the variables X and Y and examine whether $|E(X)| = E(|X|)$.

21. A box contains five cyclical discs, with diameters 1, 2, 3, 4, and 5 cm, respectively. We select a disc at random.

(i) What is the expected *length* of the disc chosen?

(ii) What is the expected *area* of the disc chosen?

Group B

22. A ballot contains N chips numbered $1, 2, \ldots, N$. We select k chips with replacement. Let X be the largest number on the chips drawn.

(i) Find the probability function $f(x) = P(X = x)$.

(ii) Find the expected value of X.

(Note that this is not the same setup as in Example 4.10, because there we had no replacement.)

23. An urn contains 6 red balls numbered 1, 2, 3, 4, 5, 6, and 4 black balls with numbers 7, 8, 9, 10. We select three balls from the urn randomly and without replacement. For each of the following variables, find the probability function and its expectation:

(i) the number of red balls selected,

(ii) the number of black balls selected,

(iii) the largest number on the balls chosen,

(iv) the smallest number on the balls chosen.

24. The following result is known from calculus:

$$\sum_{x=1}^{\infty} \frac{1}{x^2} = \frac{\pi^2}{6}.$$

Suppose X is a random variable whose probability function has the form

$$f(x) = cx^{-2}, \quad x \in R_X = \{1, 2, 3, \ldots\}.$$

(i) Find the value of the constant c.

(ii) Show that the expectation of X does not exist.

25. Give an example of two random variables X and Y such that neither of the expectations $E(X)$ and $E(Y)$ exists, but the expected value of the sum $X + Y$ is finite.

4.5 VARIANCE OF A DISCRETE RANDOM VARIABLE

The expectation, which we discussed in the last section, is only one measure that can be used to summarize properties of the distribution of a random variable. In this section, we discuss another such measure, namely, the *variance* of a distribution. We illustrate this concept with the aid of a familiar example. We recall the example from the beginning of the previous section; in a single throw of a die, you win \$20 if the outcome is an even integer, otherwise you lose \$10. Let X be your profit from this game. We have already seen that $E(X) = 5$. Imagine now you are offered an alternative choice; if the outcome of the throw of a die is even, you win \$60. If it is not, you lose \$50. Let Y be the random variable that denotes the profit from this game. It is very easy to see that your anticipated profit from this game is

$$E(Y) = \frac{1}{2} \cdot 60 + \frac{1}{2} \cdot (-50) = 5$$

dollars again, i.e. X and Y have the same expected value. Yet, it is clear that they are substantially different, at least in a certain aspect. The second game appears to be more *risky*, so that many people (in particular, those who have a conservative attitude toward risk), would rather avoid it. In terms of random variables, and in order to give a general definition which may not necessarily be related to financial win or loss, we say that the distribution of Y has more **variability** (or **dispersion**) around its mean value of \$5.

The most common measure for the variability of a distribution is given in the following definition.

Definition 4.5 Let X be a discrete random variable for which the expectation $\mu = E(X)$ exists. Then, the quantity

$$\text{Var}(X) = E[(X - \mu)^2] = E[(X - E(X))^2]$$

is called the **variance** of X.

From the above definition and Proposition 4.6, it follows that if X has probability function f and its range of values is the set R_X, then its variance is given by the formula

$$\text{Var}(X) = \sum_{x \in R_X} (x - \mu)^2 f(x). \tag{4.11}$$

Readers who have some familiarity with statistics should recognize the similarity between this formula and the formula used for the variance in statistics. Of course, in statistics, we have (relative) frequencies whereas here we have probabilities since we are not interested in any data but rather in the mechanism that produces these data, namely, the distribution function of the random variable in question. However, provided we think of a probability as the limit of the associated relative frequency (cf. Example 4.11 in the last section), the above formula parallels, and motivates, the use of a similar formula for the variance in statistics.

Next, from the above definition of the variance, we observe that if a variable X is measured in a certain unit, then Var(X) is measured in the *squared unit*. Thus, for example, if X is a length variable and is measured in meters, then $E(X)$ is also measured in meters, while Var(X) will be measured in square meters. However, the square of a unit does not always have a simple natural interpretation, and so if X measures time and is expressed in hours, then Var(X) will be measured in hours squared!

In order to express the variability of a random variable X in the same unit as X, which is often desirable, we use an alternative measure of dispersion, called **standard deviation**, which is the *positive* square root of the variance; that is,

$$\sqrt{\operatorname{Var}(X)} = \sqrt{E[(X-\mu)^2]}.$$

The popular symbol used for the standard deviation is the Greek letter σ, and when there are more than one variables involved we use σ_X, σ_Y, and so on, to avoid confusion. In analogy, we use σ^2 or σ_X^2, σ_Y^2, and so on as a symbol for the variance.

Suppose now X has a finite range

$$R_X = \{x_1, x_2, \dots, x_n\}.$$

Then, the variance of X admits the following physical interpretation. We assume that at the positions $x_1, x_2, \dots, x_n$ of an axis, we place masses equal to their respective probabilities, i.e. the values of the associated probability function, $f(x_1), f(x_2), \dots, f(x_n)$, and that the entire system is balanced at the point with coordinate $\mu = E(X)$ (see a similar interpretation after Definition 4.4 for the expected value). Then, denote by

$$m_i = f(x_i), \quad i = 1, 2, \dots, n,$$

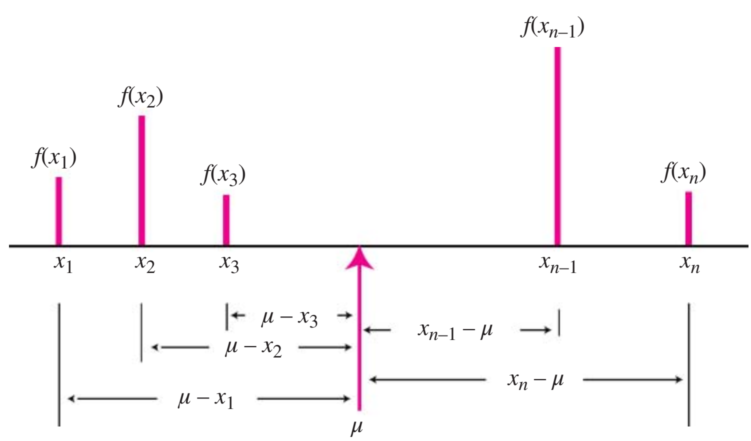
these n masses and put

$$r_i = x_i - \mu, \quad i = 1, 2, \dots, n,$$

for the distances between the locations of each mass and the system's center of mass. In this way, the variance of X takes the form

$$\operatorname{Var}(X) = \sum_{i=1}^{n} m_i r_i^2.$$

In physical terms, this quantity is the *moment of inertia* for the system (with respect to μ).



Example 4.21 We calculate the variance of the two variables X and Y discussed at the beginning of this section. Their probability functions are given by

$$f_X(x) = \begin{cases} 1/2, & x = 20, \\ 1/2, & x = -10, \end{cases}$$

and

$$f_Y(y) = \begin{cases} 1/2, & x = 60, \\ 1/2, & x = -50, \end{cases}$$

and both variables have a mean $\mu = 5$. Using the formula in (4.11), we find the variance of X to be

$$\text{Var}(X) = \sum_{x \in R_X} (x - \mu)^2 f_X(x) = \frac{(20-5)^2}{2} + \frac{(-10-5)^2}{2} = \frac{15^2 + (-15)^2}{2} = 225,$$

while that of Y equals

$$\text{Var}(Y) = \sum_{y \in R_Y} (y - \mu)^2 f_Y(y) = \frac{(60-5)^2}{2} + \frac{(-50-5)^2}{2} = \frac{55^2 + (-55)^2}{2} = 3025.$$

We thus see that the variance of Y is about 13.5 times as much as that of X. Their respective standard deviations are

$$\sigma_X = \sqrt{\text{Var}(X)} = \sqrt{225} = 15, \quad \sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{3025} = 55.$$

From Definition 4.5 it is clear that, unlike the expectation which can take on any real value, the variance of a random variable is always nonnegative. In fact, as the following proposition shows, if a variable assumes more than one value, then its variance is strictly positive.

Proposition 4.8 *Let X be a discrete random variable for which the expectation μ exists. If* $\text{Var}(X) = 0$*, then X takes only one value (with probability one).*

Proof: We shall prove that $P(X = \mu) = 1$. Suppose to the contrary that there exists $x_0 \in R_X$ such that $x_0 \neq \mu$ and $P(X = x_0) = f(x_0) > 0$. Then, we obtain

$$\text{Var}(X) = \sum_{x \in R_X} (x - \mu)^2 f(x) = (x_0 - \mu)^2 f(x_0) + \sum_{x \in R_X, x \neq x_0} (x - \mu)^2 f(x) > 0,$$

since the first summand above is strictly positive. This contradicts the assumption that $\text{Var}(X) = 0$. We thus conclude that X cannot take on any value other than its expected value, μ, with positive probability. □

If a random variable X takes (with probability one) only one value, as in the proposition above, then we say that the distribution of X is *concentrated on a single point*; such distributions are called *degenerate*. Observe that if a variable X assumes only one value, then it follows immediately from the definition of the variance that $\text{Var}(X) = 0$. Proposition 4.8 says that the converse is also true, and so by combining these two results, we see that a random variable has zero variance if and only if its distribution is degenerate.

In practice, when we want to find the variance of a random variable X, instead of using the formula in Definition 4.5, we may use the following result that is much simpler.

Proposition 4.9 *The variance of a random variable X is given by*

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

Proof: Employing Proposition 4.7 and Corollary 4.1, we obtain

$$\begin{aligned}\text{Var}(X) &= E[(X-\mu)^2] = E[X^2 - 2\mu X + \mu^2] = E(X^2) - 2\mu \cdot E(X) + \mu^2 \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 = E(X^2) - \mu^2,\end{aligned}$$

as stated in the proposition. □

Among other things, the last result is useful as a reminder that the quantities $E(X^2)$ and $[E(X)]^2$ are not the same (unless X assumes only one value). In fact, since the variance of a random variable is necessarily nonnegative, we obtain the inequality

$$E(X^2) \geq [E(X)]^2$$

which holds generally. Again, the inequality is strict unless the distribution of X is degenerate.

Example 4.22 The number of faults found during the quality inspection of a manufacturing item is a random variable, X, with probability function

$$f(x) = P(X = x), \quad x \in R_X = \{0, 1, 2, 3, 4\},$$

given in Table 4.2 below:

x	$f(x)$
0	0.29
1	0.23
2	0.21
3	0.17
4	0.10

Table 4.2 Probability function for the number of faults per item.

Find the expected value of faults per item, $E(X)$, and the standard deviation σ_X.

SOLUTION For the expectation, we get

$$\begin{aligned}E(X) &= \sum_{x\in R_X} xf(x) = 0 \cdot (0.29) + 1 \cdot (0.23) + 2 \cdot (0.21) + 3 \cdot (0.17) + 4 \cdot (0.10)\\ &= 1.56\end{aligned}$$

faults per item on average.

For the variance, we use the formula in Proposition 4.9. For this, we need to calculate $E(X^2)$. Detailed calculations are given in the following table (interested readers may verify that calculations using the formula in Definition 4.5 are more cumbersome, although of course the result for the variance is the same).

Thus, we find

$$\begin{aligned}E(X^2) &= \sum_{x\in R_X} x^2 f(x) = 0^2 \cdot (0.29) + 1^2 \cdot (0.23) + 2^2 \cdot (0.21)\\ &\quad + 3^2 \cdot (0.17) + 4^2 \cdot (0.10)\\ &= 4.20,\end{aligned}$$

as shown in Table 4.3.

From the above, we obtain the variance of X as

$$\sigma_X^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 4.2 - (1.56)^2 = 1.7664,$$

from which we get the standard deviation for the number of faults per item to be

$$\sigma_X = \sqrt{\sigma_X^2} = \sqrt{1.7664} = 1.329.$$

x	$f(x)$	$xf(x)$	$x^2f(x)$
0	0.29	0	0
1	0.23	0.23	0.23
2	0.21	0.42	0.84
3	0.17	0.51	1.53
4	0.10	0.40	1.60
Total	1.00	1.56	4.20

Table 4.3 Calculations for $E(X^2)$.

Example 4.23 (Variance of a discrete uniform random variable)
Let X be a random variable that takes the values $a_1, a_2, \ldots, a_n$, each with equal probability, so that

$$f(x) = P(X = x) = \frac{1}{n}, \quad x \in R_x = \{a_1, a_2, \ldots, a_n\}.$$

We have seen in Example 4.14 that the expectation of X equals

$$E(X) = \frac{a_1 + a_2 + \cdots + a_n}{n}.$$

For the expectation $E(X^2)$, we get

$$E(X^2) = \sum_{x \in R_X} x^2 f(x) = \sum_{x \in R_X} x^2 \cdot \frac{1}{n} = \frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n},$$

and so

$$\text{Var}(X) = E(X^2) - [E(X)]^2 = \frac{a_1^2 + a_2^2 + \cdots + a_n^2}{n} - \left(\frac{a_1 + a_2 + \cdots + a_n}{n}\right)^2.$$

In particular, for the example that concerns the throw of a fair die, we have

$$R_X = \{1, 2, 3, 4, 5, 6\}$$

and

$$E(X^2) = \frac{1^2 + 2^2 + \cdots + 6^2}{6} = \frac{91}{6}, \quad \text{Var}(X) = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}.$$

For the general case we note that, when X has a discrete uniform distribution, the inequality $[E(X)]^2 \leq E(X^2)$ leads to the relation

$$(a_1 + a_2 + \cdots + a_n)^2 \leq n(a_1^2 + a_2^2 + \cdots + a_n^2),$$

which is a special case of the (algebraic) Cauchy–Schwarz inequality

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

for $b_1 = b_2 = \cdots = b_n = 1$.

In cases where we are interested in the variance for a linear transformation of a random variable, the following result is useful.

Proposition 4.10 *Let X be a discrete random variable and a and b be two real numbers. Then, the variance and standard deviation of the variable*

$$Y = aX + b$$

are given by

$$\text{Var}(Y) = a^2\text{Var}(X), \quad \sigma_Y = |a|\sigma_X.$$

Proof: From Corollary 4.1, we know the expected value of Y to be

$$\mu_Y = E(aX + b) = a\mu + b,$$

where $\mu = E(X)$. We thus obtain, using Corollary 4.1 again,

$$\begin{aligned}\text{Var}(Y) = \text{Var}(aX + b) &= E[(Y - \mu_Y)^2] = E[((aX + b) - (a\mu + b))^2] \\ &= E[a^2(X - \mu)^2] = a^2E[(X - \mu)^2] \\ &= a^2\text{Var}(X).\end{aligned}$$

The result about the standard deviation follows immediately since

$$\sigma_Y = \sqrt{\text{Var}(Y)} = \sqrt{a^2\text{Var}(X)} = \sqrt{a^2}\sqrt{\text{Var}(X)} = |a|\sigma_X$$

(recall that the standard deviation is the *positive* square root of the variance of a random variable). □

In particular, from the last result we see that, adding a constant to a random variable X does not affect the variance of X at all; to understand why this happens, recall that the variance is a measure of the *variability* for the distribution of that variable. This variability is not affected if all possible values of X are increased by a fixed amount, although the mean of X will clearly change. Imagine, for instance, how the graph of the probability function depicted in Figure 4.4 will be affected if we add the value 3 to the variable having that probability function. The entire graph of the function will be shifted three units to the right, but its shape and variability around the mean will remain unaltered.

Example 4.24 For a random variable X, it is known that $E(X) = 4$ and $E[X(X - 5)] = 5$. Calculate the expected value and the standard deviation for the random variables

$$Y = -3X + 2 \quad \text{and} \quad Z = \frac{X - 4}{3}.$$

SOLUTION We find first the expectations of the two random variables. For Y, we see that

$$E(Y) = E(-3X + 2) = -3E(X) + 2 = -3 \cdot 4 + 2 = -10,$$

while for Z we get, using again the linearity property of expected values,

$$E(Z) = E\left(\frac{X-4}{3}\right) = \frac{E(X)-4}{3} = 0.$$

Next, we find the variance of X. For this, we note that

$$E[X(X-5)] = E(X^2 - 5X) = E(X^2) - 5E(X).$$

Since we are given that this equals 5, we obtain

$$E(X^2) = 5 + 5E(X) = 5 + 5 \cdot 4 = 25.$$

Thus, from Proposition 4.9, we immediately get

$$\mathrm{Var}(X) = 25 - 4^2 = 9,$$

so that $\sigma_X = \sqrt{9} = 3$. For the standard deviation of Y, Proposition 4.10 now yields (taking $a = -3$ and $b = 2$ there)

$$\sigma_Y = |-3| \cdot \sigma_X = 3 \cdot 3 = 9.$$

Similarly, we obtain for Z, putting $a = 1/3$ and $b = -4/3$ in the result of Proposition 4.10,

$$\sigma_Z = \frac{1}{3}\sigma_X = \frac{3}{3} = 1.$$

A random variable that has zero mean and standard deviation 1, just as the variable Z in the last example, is called a **standardized** random variable.

Generally, assume that $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then, for the random variable

$$Z = \frac{X-\mu}{\sigma},$$

it is easy to verify that $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$ (so that the standard deviation of Z is also 1), and this means that Z is a standardized random variable. Such variables play a key role in Chapter 7.

As we have seen, for a random variable X, the expected values

$$\mu = E(X), \quad E[(X-\mu)^2], \quad E(X^2)$$

give us useful information about the distribution of X, that is for the way that X behaves probabilistically. There are other functions $g(X)$ of the variable X, the expected values of which give us further insight in the way that X behaves. For instance, if $c \in \mathbb{R}$ is a constant and $r > 0$ is an arbitrary integer, the expectations of the random variables (functions of X)

$$X^r, \;\; (X)_r = X(X-1)\cdots(X-r+1), \;\; |X|^r, \;\; (X-c)^r, \; (X-\mu)^r, \quad r = 1, 2, \ldots,$$

(provided they exist) are all important quantities that give information about the shape, the location, the presence or lack of symmetry and, more generally, about the distribution of X. We note, in particular, that the quantities $E[(X-c)^r]$ are called the **moments of order** r **around the point** c of the distribution, while in the special case when $c = \mu$ we have the **central moments of order** r. Table 4.4 gives the customary notation and the name given to the expectations for each of the five quantities above.

It is obvious that the following relations are true:

$$\mu'_1 = \mu_{(1)} = E(X), \quad \mu_1 = 0,$$
$$\mu_2 = \mathrm{Var}(X).$$

Example 4.25 In analogy with the expectation, which is simply the first moment of the distribution of a random variable X, we say that the moment of order r *exists* if the series on the right-hand side of the equality

$$E(X^r) = \sum_{x \in R_X} x^r f(x)$$

converges absolutely, i.e. if the sum

$$E(|X|^r) = \sum_{x \in R_X} |x|^r f(x)$$

is finite. With the notation of Table 4.4, this means that X has a finite moment of order r whenever its absolute moment of order r is finite.

Further, in view of the inequality

$$|X|^{r-1} \leq 1 + |X|^r, \quad r = 1, 2, \ldots \tag{4.12}$$

we see that whenever the rth moment of a random variable exists, so does the $(r-1)$th and hence, by induction, all lower-order moments.

$g(X)$	Notation	Name
X^r	$\mu'_r = E(X^r)$	Moment of order r (around zero)
$(X)_r$	$\mu_{(r)} = E[(X)_r]$	Factorial moment of order r
$\|X\|^r$	$E(\|X\|^r)$	Absolute moment of order r
$(X-c)^r$	$E[(X-c)^r]$	Moment of order r around c
$(X-\mu)^r$	$\mu_r = E[(X-\mu)^r]$	Central moment of order r

Table 4.4 Notation and terminology for the various types of moments.

EXERCISES

Group A

1. Find the variance for each of the random variables in Exercise 1 of Section 4.4.

2. With reference to Exercise 5, Section 4.4 wherein Jimmy participates in a quiz show with multiple choice questions:

   (i) Calculate the variance for the number of correct answers that he gives.

   (ii) Calculate the variance and standard deviation of Jimmy's earnings from the show.

3. Find the variance of the random variable X having a probability function f and range R_X in each of the following cases:

   (i) $f(x) = \dfrac{2 - |x-2|}{4}, x \in R_X = \{1, 2, 3\};$

   (ii) $f(x) = \dfrac{4 + |x-12|}{42}, x \in R_X = \{1, 2, 3\};$

   (iii) $f(x) = \dfrac{|x-3| + 1}{28}, x \in R_X = \{-3, -2, -1, 0, 1, 2, 3\};$

   (iv) $f(x) = \dfrac{4 - x^2}{6}, x \in R_X = \{1, \sqrt{2}, \sqrt{3}\}.$

4. If for a variable X it is known that

$$E(X-1) = 3, \quad E[X(X+2)] = 40,$$

   calculate the mean and standard deviation of the random variable

$$Y = \frac{X-4}{3}.$$

5. With reference to Example 4.16, concerning Nick's profit from his investment:

   (i) Find the standard deviation of his profit.

   (ii) What is the probability that Nick's profit is within one standard deviation either way around its expected value?

6. The probability function of a discrete random variable is given by

$$f(x) = c(32 - x), \quad x = 1, 2, \dots, 31,$$

   where c is a constant.

   (i) Show that $c = 1/496$.

   (ii) Find the expectation $E(X)$.

   (iii) Find the expectation of the random variable $Y = 3(X - 11)$ and show that $\text{Var}(Y) = E(Y^2)$.

7. If, for a variable X it is known that $\text{Var}(2018 - X) = 2$, find $\text{Var}(-5X + 1)$.

8. Calculate $\text{Var}(2X + 4)$ if it is known that

$$E(5X - 1) = 9 \quad \text{and} \quad E[(X + 2)^2] = 17.$$

9. Verify the truth of (4.12) in Example 4.25 for any discrete random variable X and any positive integer r.

10. Suppose that a discrete random variable X has probability function

$$f(x) = P(X = x) = \frac{1}{5}, \quad x \in R_X = \{-2a, -a, 0, a, 2a\}.$$

   (i) Show that, regardless of the value of a, we have $E(X) = 0$.

   (ii) Calculate $\text{Var}(X)$ as a function of a. What happens with the variance as a increases? Give an intuitive interpretation of this result.

11. Let for $r = 1, 2, \ldots,$

$$\mu_{(r)} = E[X(X - 1) \cdots (X - r + 1)]$$

be the factorial moments of the random variable X, and

$$\mu_r = [E(X - \mu)^r], \quad \mu'_r = E(X^r), \quad r = 1, 2, \ldots,$$

be the moments of X around its mean, $\mu = E(X)$ (central moments), and around zero, respectively.

Using properties of expectations, show that the following are true:

   (i) $\mu'_1 = \mu_{(1)}, \mu_1 = 0$;

   (ii) $\mu'_2 = \mu_{(2)} + \mu_{(1)}, \mu_2 = \mu'_2 - (\mu'_1)^2$;

   (iii) $\mu'_3 = \mu_{(3)} + 3\mu_{(2)} + \mu_{(1)}, \mu_3 = \mu'_3 - 3\mu'_2\mu'_1 + 2(\mu'_1)^3$;

   (iv) $\mu'_4 = \mu_{(4)} + 6\mu_{(3)} + 7\mu_{(2)} + \mu_{(1)}, \mu_4 = \mu'_4 - 4\mu'_3\mu'_1 + 6\mu'_2(\mu'_1)^2 - 3(\mu'_1)^4$.

Verify also the truth of the identity

$$\text{Var}(X) = \mu_{(2)} + \mu - \mu^2.$$

Group B

12. Walter throws two dice. If the sum of the two outcomes is 10 or more, he receives \$5. If the sum is 8 or 9, nothing happens, while if the sum is less than 8, he loses c dollars. If X denotes Walter's profit from this game, find the variance of X given that the game is fair.

13. Let X be a discrete random variable and assume that it has finite moments of any order up to a positive integer r, i.e. $\mu'_i < \infty$ for any $i = 1, 2, \ldots, r$. Show that the central moment of order r,

$$\mu_r = [E(X - \mu)^r],$$

where $\mu = E(X)$, is given by the formula

$$\mu_r = \sum_{k=0}^{r} (-1)^k \binom{k}{r} \mu^k \mu'_{r-k}.$$

(*Hint*: Begin by expanding the power $(X - \mu)^r$ according to the binomial theorem.)

14. Let X be a random variable and t be a real number such that the moment $E[(X-t)^2]$ exists.

 (i) Verify that
$$E[(X-t)^2] = E(X^2) - 2tE(X) + t^2.$$

 (ii) Show that
$$E[(X-t)^2] = \mathrm{Var}(X) + (\mu - t)^2.$$

 (iii) Use the result in (ii) to establish that the minimum value of the function
$$h(t) = E[(X-t)^2]$$
 is achieved at $t = \mu$, and that
$$\min_{t\in\mathbb{R}} E[(X-t)^2] = \mathrm{Var}(X).$$

15. Consider two random variables X and Y with the same range
$$R_X = R_Y = \{a_1, a_2, a_3\},$$
and suppose that
$$E(X) = E(Y) \quad \text{and} \quad \mathrm{Var}(X) = \mathrm{Var}(Y).$$

 (i) Prove that $E(X^2) = E(Y^2)$.

 (ii) Show that the quantities
$$x_i = P(X = a_i) - P(Y = a_i), \quad i = 1, 2, 3,$$
 satisfy a homogeneous linear system of equations, for which the determinant of the coefficient matrix for the unknowns has the form
$$\begin{vmatrix} 1 & 1 & 1 \\ a_1 & a_2 & a_3 \\ a_1^2 & a_2^2 & a_3^2 \end{vmatrix} \neq 0.$$

 (iii) Explain why the distributions of the random variables X and Y coincide, i.e. we have
$$P(X = a_i) = P(Y = a_i)$$
 for $i = 1, 2, 3$.

16. Consider now two random variables X and Y having the same range

$$R_X = R_Y = \{a_1, a_2, \dots, a_n\},$$

and assume that

$$E(X^r) = E(Y^r) \quad \text{for } r = 1, 2, \dots, n-1.$$

Arguing as in the previous exercise, verify again that the distributions of the variables X and Y are identical.

4.6 SOME RESULTS FOR EXPECTATION AND VARIANCE

When we want to calculate probabilities associated with a random variable X, we need to know its distribution (i.e. either the distribution function of X or the probability function). It is often the case, however, that we can find the mean and the variance of a random variable but we cannot specify completely the distribution of X.

In such cases, although we cannot calculate exact values for the probabilities associated with X, we will still be in a position to give (upper and/or lower) bounds for these probabilities. Two important general results in this direction are two inequalities, both named after Russian mathematicians: the **Markov inequality** and the **Chebyshev inequality**.

Proposition 4.11 *(Markov's inequality) Let X be a discrete random variable taking nonnegative values such that $E(X)$ exists. Then, for any $t > 0$, we have*

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof: Let $t > 0$ and A be the subset of the range, R_X, of values for X, which is defined by

$$A = \{x \in R_X : x \geq t\}.$$

Since $A \subseteq R_X$, we can write

$$E(X) = \sum_{x \in R_X} xf(x) \geq \sum_{x \in A} xf(x) \tag{4.13}$$

and, since for $x \in A$ we have $x \geq t$, we obtain

$$xf(x) \geq tf(x), \quad x \in A.$$

Summing both sides of this inequality over all $x \in A$, we obtain

$$\sum_{x \in A} xf(x) \geq \sum_{x \in A} tf(x) = t \sum_{x \in A} f(x) = tP(X \in A) = tP(X \geq t),$$

the last step following from the definition of the set A. From this and (4.13), we see that $E(X) \geq tP(X \geq t)$, as required. □

An immediate application of Markov's inequality is that for any random variable X with mean μ, the probability that X takes values exceeding the k-fold multiple of μ is at most $1/k$, i.e.

$$P(X \geq k\mu) \leq \frac{1}{k}.$$

For instance, if a telephone center receives on average 7500 calls per day, the probability that on any given day at least 15 000 calls arrive at the center cannot be larger than a half, and we can say this without having any knowledge about the distribution for the daily number of calls arriving at the center.

Proposition 4.12 *(Chebyshev's inequality) Let X be a discrete random variable with mean μ and variance σ^2. Then, for any $t > 0$, we have*

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof: Applying Markov's inequality to the nonnegative random variable

$$Y = (X - \mu)^2$$

(and putting t^2 instead of t), we get

$$P((X - \mu)^2 \geq t^2) \leq \frac{E[(X - \mu)^2]}{t^2}.$$

The required result then follows upon noting that $E[(X - \mu)^2] = \sigma^2$ and that

$$P((X - \mu)^2 \geq t^2) = P(|X - \mu| \geq t). \qquad \square$$

The last result was first formulated by the French mathematician Irénée Bienaymé (1796–1878), and in many books it is referred to as the Bienaymé–Chebyshev inequality. Pafnuty Chebyshev (1821–1894), who was a friend of Bienaymé, later gave a new proof relating this result to the laws of large numbers, a celebrated series of results in probability theory and, mainly because of that, the result of Proposition 4.12 bears only his name in most sources.

Example 4.26 The number of butterflies that live in a certain forested area is a random variable X with mean 200.

(i) Give an estimate for the probability that the number of butterflies is at least 300.

(ii) A zoologist believes that the variance of X is 40. What can we say about the probability that the actual number of butterflies is between 120 and 280? Can we now get a better estimate for the probability that there are at least 300 butterflies in this area?

SOLUTION

(i) For this part, we have no knowledge of the variance (or the standard deviation) of X, and so Chebyshev's inequality cannot be used. Instead, from Markov's inequality, we get

$$P(X \geq 300) \leq \frac{200}{300} = \frac{2}{3}.$$

Note that this is actually an upper bound, not an estimate as required by the question. In the absence of any further information, however, we cannot give a better estimate (although we do not know how close this is to the true probability).

(ii) Here, we have $\mu = 200$ and $\sigma^2 = 40$, so that from Chebyshev's inequality, we get

$$\begin{aligned} P(120 < X < 280) &= P(-80 < X - 200 < 80) \text{ (subtracting 200 on each side)} \\ &= P(|X - 200| < 80) = 1 - P(|X - 200| \geq 80) \geq 1 - \frac{40}{80^2} \\ &= \frac{159}{160} \cong 0.994. \end{aligned}$$

This time, the result we have obtained is lot more useful; the probability that the number of butterflies is between 120 and 280 is higher than 99%, and we can say this without having any knowledge about the distribution of X. Further, it might appear at first sight that the result here is somewhat conflicting with that of part (i); however, this is not so. If the number of butterflies is between 120 and 280 with probability at least 99.4%, then the probability that there are more than 300 butterflies cannot be higher than 0.6%. More precisely, we have

$$\begin{aligned} P(X \geq 300) \leq P(X \geq 280) &= 1 - P(X < 280) \leq 1 - (P(120 < X < 280) \\ &\leq 1 - 0.994 = 0.006. \end{aligned}$$

There is no conflict with the result in (i), however, since the result there is merely an upper bound, and the last result, 0.006, is clearly less than $2/3$. The reason why our answer is much more accurate here, is that in Part (i) we did not have any information about the variance.

Extending a bit further the idea that is implicit in the solution of Example 4.26, we can put forward the role of the variance in describing the probabilistic behavior of a random variable. For this, let us put $t = k\sigma$ in Chebychev's inequality to get

$$P(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

or, equivalently,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

This says that the probability the values of an arbitrary random variable X are spread within k standard deviations around the mean of X is at least $1 - 1/k^2$. Thus, for instance, the probability that

- X takes a value within two standard deviations either way of its mean is at least $1 - \frac{1}{2^2} = 75\%$;
- X takes a value within four standard deviations either way of its mean is at least $1 - \frac{1}{4^2} = 93.75\%$;
- X takes a value within ten standard deviations either way of its mean is at least $1 - \frac{1}{10^2} = 99\%$,

and so on.

The above considerations should not give the (false) impression that either Markov's or Chebyshev's inequality give always very accurate estimates of the true probabilities. For Markov's inequality, we have seen in the last example a case wherein it produces a very poor bound for the actual probability. In fact, both inequalities are valid under minimal assumptions and the price we usually pay for such generality in mathematics is lack of accuracy in individual cases. As one can guess from the statement of both results, the upper bounds in these inequalities are really useful typically only in cases where t is much larger than $E(X)$ (for Markov's inequality) or than σ_X (for Chebyshev's inequality). The next example presents a case where the upper bounds in both inequalities are either useless or give poor approximations for the true probabilities.

Example 4.27 (Applying Markov and Chebyshev inequalities to a throw of a die) Suppose that we throw a single die and denote by X the outcome of the throw. We then have

$$\mu = E(X) = \sum_{x=1}^{6} xf(x) = \sum_{x=1}^{6} x \cdot \frac{1}{6} = \frac{21}{6} = \frac{7}{2}$$

and

$$E(X^2) = \sum_{x=1}^{6} x^2 f(x) = \sum_{x=1}^{6} x^2 \cdot \frac{1}{6} = \frac{91}{6}.$$

Thus,

$$\sigma^2 = \text{Var}(X) = E(X^2) - \mu^2 = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12}$$

(in fact, all three results above have already been obtained in Example 4.23 by observing that the outcome of a fair die has a discrete uniform distribution). Then, a straightforward application of Markov's inequality gives

$$P(X \geq 2) \leq \frac{7/2}{2} = \frac{7}{4} = 1.75,$$

$$P(X \geq 6) \leq \frac{7/2}{6} = \frac{7}{12} \cong 0.583.$$

The first of these results is clearly useless (since we know that $P(X \geq 2)$ cannot be greater than one), while the second is a very poor bound for the true probability that is

$$P(X \geq 6) = P(X = 6) = \frac{1}{6} \cong 0.167.$$

Similarly, applying Chebyshev's inequality to the random variable X, we find

$$P\left(\left|X - \frac{7}{2}\right| \geq \frac{5}{2}\right) \leq \frac{35/12}{25/4} = \frac{7}{15}.$$

Note that the event inside the bracket on the left above occurs if either $P(X \geq 6)$ or $P(X \leq 1)$. Therefore, we obtain that

$$P(X \geq 6 \text{ or } X \leq 1) \leq \frac{7}{15} \cong 0.47.$$

The upper bound is again quite conservative as it is about 50% higher than the true probability on the left, which is

$$P(X \geq 6 \ \text{ or } \ X \leq 1) = P(X = 1) + P(X = 6) = \frac{1}{6} + \frac{1}{6} \cong 0.33.$$

Despite the evidence from the above examples, both Markov's and Chebyshev's inequalities are useful results as they enable us to make probability statements about a random variable in the absence of any information for the underlying distribution. Moreover, Chebyshev's inequality has been used to prove one of the most important results in probability theory, the *weak law of large numbers*, which is discussed in detail in Volume II of this book.

Our final result in this chapter links the distribution function of a random variable to its expected value. Typically, in order to calculate the expectation of a random variable X, we use the probability function of X and the formula from Definition 4.4. The next result, however, shows that when X takes nonnegative integer values, the expectation $E(X)$ can be found from the distribution function F, of X, in case we want to avoid calculation of the probability function.

Proposition 4.13 *Let X be a discrete random variable with range*

$$R_X = \{0, 1, 2, \ldots\}.$$

If the expectation $E(X)$ exists, then it can be expressed in terms of the distribution function F of the variable X using the formula

$$E(X) = \sum_{x=0}^{\infty} [1 - F(x)] = \sum_{x=0}^{\infty} P(X > x).$$

Proof: We have

$$E(X) = \sum_{x \in R_X} xf(x) = \sum_{x=0}^{\infty} xf(x) = \sum_{y=1}^{\infty} yf(y).$$

Upon replacing the coefficient y of $f(y)$ by the sum

$$y = \underbrace{1 + 1 + \cdots + 1}_{y \text{ times}} = \sum_{x=1}^{y} 1,$$

we get

$$E(X) = \sum_{y=1}^{\infty} \left(\sum_{x=1}^{y} 1 \right) f(y) = \sum_{y=1}^{\infty} \sum_{x=1}^{y} f(y).$$

Interchanging the order of summation (which is allowed here since the summands are nonnegative), we may then write

$$E(X) = \sum_{y=1}^{\infty} \sum_{x=1}^{y} f(y) = \sum_{x=1}^{\infty} \left(\sum_{y=x}^{\infty} f(y) \right) = \sum_{x=1}^{\infty} \left[1 - \sum_{y=0}^{x-1} f(y) \right].$$

Taking now into account that

$$\sum_{y=0}^{x-1} f(y) = F(x-1),$$

we arrive at the expression

$$E(X) = \sum_{x=1}^{\infty} [1 - F(x-1)] = \sum_{x=0}^{\infty} [1 - F(x)],$$

which is the desired result. □

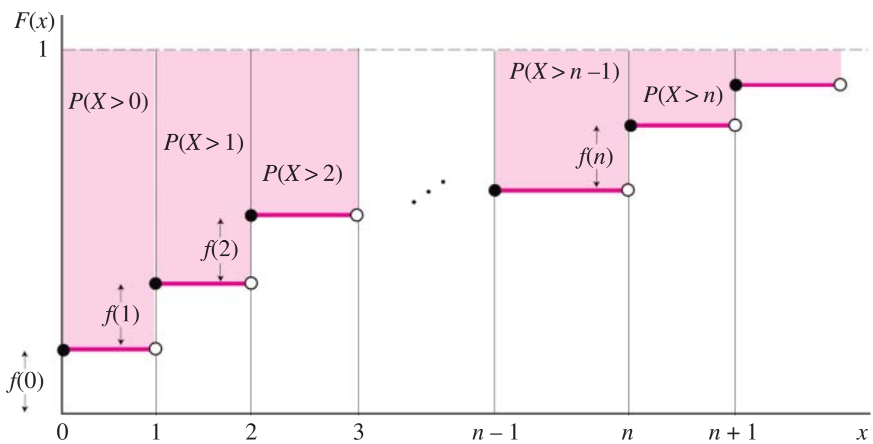


Figure 4.8 Illustration of the result in Proposition 4.13.

The result of Proposition 4.13 is depicted in Figure 4.8. The expectation of X is represented by the shadowed area on the graph.

The next example presents an interesting application of the previous result. We further note that Proposition 4.13 remains valid when the range of values of the variable X is a subset of the set of nonnegative integers, rather than the entire set. For this case, an application of the result is given next.

Example 4.28 Let X and Y be two arbitrary random variables and $c \in \mathbb{R}$. If, for any $t \geq 0$, we have

$$P(|Y - c| \leq t) \leq P(|X - c| \leq t),$$

we understand intuitively that it is more likely for X to be close to c than Y. We may thus say that X **is more concentrated around** c **than** Y.

Suppose now X and Y take on integer values (more specifically, let $R_X = R_Y \subset \{0, 1, 2, \ldots\}$) and they have the same mean, i.e.

$$\mu = E(X) = E(Y).$$

Assume further that X is more concentrated around μ than Y. Then, if we define the variables

$$Y_1 = (Y - \mu)^2, \quad X_1 = (X - \mu)^2,$$

we see that

$$\begin{aligned} P(Y_1 \leq t) &= P[(Y - \mu)^2 \leq t] = P[|Y - \mu| \leq \sqrt{t}] \\ &\leq P[|X - \mu| \leq \sqrt{t}] = P[(X - \mu)^2 \leq t] = P(X_1 \leq t) \end{aligned}$$

for any $t \geq 0$, which in turn gives

$$P(Y_1 > t) \geq P(X_1 > t).$$

Adding these inequalities over all t in their range, we get (note that $R_{X_1} = R_{Y_1}$)

$$\sum_{t \in R_{Y_1}} P(Y_1 > t) \geq \sum_{t \in R_{X_1}} P(X_1 > t)$$

and Proposition 4.13 yields then that $E(Y_1) \geq E(X_1)$. Consequently,

$$\mathrm{Var}(Y) = E[(Y - \mu)^2] \geq E[(X - \mu)^2] = \mathrm{Var}(X).$$

We therefore see that the more concentrated around its mean a variable is, the smaller its variance.

EXERCISES

Group A

1. Markov's inequality (Proposition 4.11) concerns a nonnegative random variable. An extended version of this inequality that applies to *any* random variable X is the following:

$$P(|X| \geq t) \leq \frac{E(|X|)}{t}$$

for any $t > 0$. Explain why this is true (no calculations are needed).

2. A random variable X has the following probability function

$$f(x) = \begin{cases} 3/4, & x = 0, \\ 1/6, & x = 2, \\ 1/12, & x = 4. \end{cases}$$

   (i) Find the expected value of X

      (a) from Definition 4.4,

      (b) using Proposition 4.13,

   and verify that the two results agree.

   (ii) Find the percentage error in the upper bound obtained from Markov's inequality as an estimate for the following probabilities associated with X:

$$P(X \geq 1), \quad P(X \geq 2), \quad P(X \geq 3).$$

3. The number of mail items handled daily by a courier service is a random variable. It has been estimated that this variable has a mean of 3000 items and variance 40 000. Obtain a lower bound for the probability that, in a given day, the company handles more than 2400 but less than 3600 mail items.

4. For a nonnegative discrete random variable X, we know that

$$E(X) = 25, \quad E[X(X-4)] = 900.$$

Find an upper bound for the probability $P(X \geq 50)$ using

   (i) Markov's inequality;

   (ii) Chebyshev's inequality.

5. For the discrete random variable X, suppose that

$$E(X) = \text{Var}(X) = \mu.$$

What can you infer from Chebyshev's inequality for the probability $P(X \geq 2\mu)$?

6. Suppose that X has range $\{-2, 0, 2\}$ and probability function

$$f(x) = \begin{cases} 1/8, & x = -2, \\ 3/4, & x = 0, \\ 1/8, & x = 2. \end{cases}$$

Examine whether Chebychev's inequality is exact, i.e. it holds as an equality in this case.

Group B

7. For a discrete random variable X with mean μ, assume that the central moment of order $2r$,

$$\mu_{2r} = E[(X-\mu)^{2r}]$$

exists for a given integer $r \geq 1$.

Show that for any $a > 0$, we have

$$P(|X-\mu| \geq a) \leq \frac{\mu_{2r}}{a^{2r}}.$$

8. For the random variable X, assume that the expectation $E(\mathrm{e}^{aX})$ exists for a given $a \in \mathbb{R}$. Prove that, for any real t, the following holds:

$$P(X \geq t) \leq \frac{E(\mathrm{e}^{aX})}{\mathrm{e}^{at}}.$$

9. Let X be a discrete random variable, and a and t be two positive real numbers. Assuming that the expectation $E(\mathrm{e}^{aX})$ exists, establish the following inequality

$$P\left(X < \frac{1}{a}\ln t\right) \geq 1 - \frac{1}{t}E(\mathrm{e}^{aX}).$$

10. (One-sided Chebyshev bound) Let X be a random variable such that $E(X) = 0$ and $\mathrm{Var}(X) = \sigma^2$. Show that, for any $a > 0$,

$$P(X \geq a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

This result is known as Cantelli's inequality.

(*Hint*: For $b > 0$, put $W = (X+b)^2$ and apply Markov's inequality to W; then examine which value of b minimizes the quantity on the right.)

4.7 BASIC CONCEPTS AND FORMULAS

Random variable (r.v.) defined on a sample space Ω	A real function $X : \Omega \mapsto \mathbb{R}$ such that the set $\{\omega \in \Omega : X(\omega) \in I\}$ is an event of Ω for any $I \subset \mathbb{R}$
Range of values R_X for the random variable X	$\{x \in \mathbb{R} : X(\omega) = x \text{ for some } \omega \in \Omega\}$
Properties of a distribution function (d.f.) F	F is increasing and right-continuous on $\mathbb{R}$; $\lim_{t\to\infty} F(t) = 1, \quad \lim_{t\to-\infty} F(t) = 0$
Formulas for calculating probabilities via the d.f.	$P(X \le b) = F(b)$; $P(X < b) = F(b-)$; $P(a < X \le b) = F(b) - F(a)$; $P(a < X < b) = F(b-) - F(a)$; $P(a \le X \le b) = F(b) - F(a-)$; $P(a \le X < b) = F(b-) - F(a-)$
Discrete random variable X	A random variable whose range of values R_X is finite or countably infinite
Probability function of a discrete r.v. X with $R_X = \{x_1, x_2, \dots, x_n, \dots\}$	$f(x) = 0$ for any $x \notin R_X$; $f(x) = P(X = x)$ for $x \in R_X$, or $f(x_i) = P(X = x_i)$, $i = 1, 2, \dots$
Properties of the probability function f	(i) $f(x) = 0$ for $x \neq x_1, x_2, \dots$; (ii) $f(x_i) \ge 0$ for $i = 1, 2, \dots$; (iii) $\sum_{x\in R_X} f(x) = \sum_{x=1}^{\infty} f(x_i) = 1$
Relationship between the probability function and the distribution function of a r.v. X with $R_X = \{x_1, x_2, \dots, x_n, \dots\}$, where $x_1 < x_2 < \cdots < x_n < \cdots$	$F(t) = \sum_{i : x_i \le t} f(x_i) = \begin{cases} 0, & t < x_1, \\ \sum_{i=1}^{r-1} f(x_i), & x_{r-1} \le t < x_r, \end{cases}$ for $r = 2, 3, \dots$ $f(x_1) = F(x_1)$; $f(x_r) = F(x_r) - F(x_{r-1}), \quad r = 2, 3, \dots$
$P(X \in A)$ for $A \subset R_X$	$\sum_{i : x_i \in A} f(x_i) = \sum_{x\in A} f(x)$
Probability function of $Y = g(X)$	$f_Y(y) = P(Y = y) = \sum_{x \in R_X : g(x) = y} f(x)$
Expectation (mean value) of X	$\mu = E(X) = \sum_{x\in R_X} xP(X = x) = \sum_{x\in R_X} xf(x)$
Expectation for a function of a discrete random variable	$E[g(X)] = \sum_{x\in R_X} g(x)f(x)$
Properties of expectation	(i) $E\left[\sum_{i=1}^{k} \lambda_i g_i(X)\right] = \sum_{i=1}^{k} \lambda_i E[g_i(X)]$; (ii) $E(aX + b) = aE(X) + b, \quad a, b \in \mathbb{R}$

Fair game	$E(X) = 0$, where X denotes the profit of the player in the game
Variance of X	$\sigma^2 = \text{Var}(X) = E[(X-\mu)^2]$
Standard deviation of X	$\sigma = \sqrt{\text{Var}(X)}$
Alternative formula for the variance	$\text{Var}(X) = E(X^2) - [E(X)]^2$
Properties of variance	(i) If $\text{Var}(X) = 0$, then X is a constant; (ii) $\text{Var}(aX+b) = a^2\text{Var}(X)$, $a, b \in \mathbb{R}$
Standardized random variable	$Z = \dfrac{X-\mu}{\sigma}$, where $\mu = E(X), \sigma^2 = \text{Var}(X)$. Then, we have $E(Z) = 0$ and $\text{Var}(Z) = 1$.
Moment of order r around 0	$\mu'_r = E(X^r), \quad r = 1, 2, \ldots$
Factorial moment of order r	$\mu_{(r)} = E[X_{(r)}] = E[X(X-1)\cdots(X-r+1)], \quad r = 1, 2, \ldots$
Absolute moment of order r	$E(\|X\|^r), \quad r = 1, 2, \ldots$
Moment of order r around c	$E[(X-c)^r], r = 1, 2, \ldots$
Central moment of order r	$\mu_r = E[(X-\mu)^r], r = 1, 2, \ldots$
Markov's inequality	$P(X \geq t) \leq \dfrac{E(X)}{t}$, where X is a nonnegative random variable, and $t > 0$
Chebyshev's inequality	$P(\|X-\mu\| \geq t) \leq \dfrac{\sigma^2}{t^2}$, where $\mu = E(X), \sigma^2 = \text{Var}(X)$ and $t > 0$
Alternative formula for the expectation $E(X)$ when $R_X = \{0, 1, 2, \ldots\}$	$E(X) = \sum_{x=0}^{\infty}[1 - F(x)] = \sum_{x=0}^{\infty} P(X > x)$

4.8 COMPUTATIONAL EXERCISES

1. Let us consider a discrete random variable X with a probability function of the form

$$f(x) = cx^2, \quad x \in R_X = \{1, 2, \ldots, n\},$$

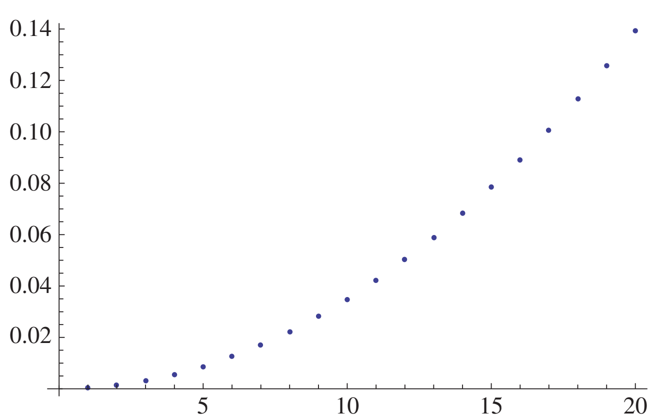
for a suitable constant c. The following sequence of commands enables us to find the value of c, and subsequently the mean and variance of X.

```
In[1]:= f1[x_]:=x^2
        a=Sum[f1[x], {x,1,n}];
        Print["Constant c should be "];c=1/a
        f[x_]:=f1[x]/a
```

```
          Print["The probability function is"];
          Print[f[x]]
          Print["The mean of the distribution is"];
          mu=Sum[x*f[x],{x,1,n}]
          Print["The variance of the distribution is"];
          var=Sum[(x^2)*f[x],{x,1,n}]-mu^2

Constant c should be
Out[3]= 6/(n (1+n) (1+2 n))

The probability function is
(6 x^2)/(n (1+n) (1+2 n))

The mean of the distribution is
Out[8]= (3 n (1+n))/(2 (1+2 n))

The variance of the distribution is
Out[10]= -((9 n^2 (1+n)^2)/(4 (1+2 n)^2))+1/5 (-1+3 n+3 n^2)
```

Note that Mathematica finds the sums above using closed-form expressions in terms of the parameter n. For a specific value of n, we can draw the graph of the probability function f as follows:

```
In[1]:= n=20;
          Print["The probability function for n= ", n, " is as
                  follows"]
          pdftable=Table[{x,f[x]},{x,1,n}]
          Print["Plot of the probability function"]
          ListPlot[pdftable]
The probability function for n= 20 is as follows:
Out[3]= {{1,1/2870},{2,2/1435},{3,9/2870},{4,8/1435},{5,5/574},
          {6,18/1435},{7,7/410},{8,32/1435},{9,81/2870},{10,10/287},
          {11,121/2870},{12,72/1435},{13,169/2870},{14,14/205},
          {15,45/574},{16,128/1435},{17,289/2870},{18,162/1435},
          {19,361/2870},{20,40/287}}
Plot of the probability function
Out[15]=
```



Modifying appropriately the sequence of commands above, do the same for each of the following discrete distributions:

(i) $f(x) = c(x + 10)$, $R_X = \{0, 1, 2, \ldots, 100\}$;

(ii) $f(x) = c(x + 15)^2$, $R_X = \{-30, -29, -28, \ldots, +10\}$;

(iii) $f(x) = c \cdot 3^x/4^x$, $R_X = \{1, 2, 3, \ldots\}$;

(iv) $f(x) = c \cdot 5^x/x!$, $R_X = \{0, 1, 2, \ldots\}$;

(v) $f(x) = c/x^2$, $R_X = \{1, 2, \ldots\}$.

2. Let f be the probability function of a discrete random variable X with range $\{m, m + 1, m + 2, \ldots, n\}$, and let g be a function defined on the same set. Write a program to calculate the expectation $E[g(X)]$. As an application,

   (i) solve Example 4.17;

   (ii) carry out the calculations necessary in Exercise 18 of Section 4.4;

   (iii) calculate the expectations $E(3X + 7)$, $E(5X^3 - 2)$, $E(Y^3)$, $E(\sqrt{Y})$, and $E(e^Y)$ for the random variables X and Y defined in Exercise 20 of Section 4.4.

3. Recall the simulation program of Section 1.8. By suitably modifying this program, find an approximation for the expected profit per game for a player who participates in each of the following games:

   (i) We throw two dice simultaneously. If exactly one die lands on a 6, we win \$10, if both dice land on 6, we win \$30, and if no 6 appears then we lose \$5.

   (ii) We throw two dice and we receive the amount in dollars equal to the sum of the two outcomes. The cost of entering this game is \$7.

   (iii) We throw three dice in succession and the amount we receive, in dollars, is equal to the sum of the first two outcomes minus the third outcome. If the difference is negative, then we have to pay an amount equal to that difference. The cost of entering this game is \$3.5.

   (iv) We throw three dice in succession and the amount we receive, in dollars, is equal to $X_1 + X_2^2 - X_3^2$, where X_1, X_2, and X_3 are the three dice outcomes. If this sum is negative, then we have to pay an amount equal to that sum.

   Decide which of the above games (if any) is a fair game.

4. The next program calculates the probability function for the number of times, X, that the ordered pair TH appears when tossing four coins successively (here, H has been coded as “1” and T has been coded as “0”):

```
In[1]:= n1=1;n2=1;n3=1;n4=1;
        Print["List of all outcomes of the experiment"]
        res=Flatten[Table[{i1,i2,i3,i4},
                {i1,0,n1},{i2,0,n2},{i3,0,n3},{i4,0,n4}],3]
        total=Length[res];
        Print["Total number of outcomes = ", total]
        Do[a[i]=0, {i,0,3}]
        Do[y=res[[i]];
           suc=Sum[(1-y[[j]])*y[[j+1]],{j,1,3}];a[suc]=a[suc]+1,
```

```
           {i,1,total}]
        Print["Probability function"];
        Do[Print[{i,a[i]/total}], {i,0,3}]

List of all outcomes of the experiment
Out[3]= {{0,0,0,0},{0,0,0,1},{0,0,1,0},{0,0,1,1},{0,1,0,0},\\
        {0,1,0,1},{0,1,1,0},{0,1,1,1},{1,0,0,0},{1,0,0,1},\\
        {1,0,1,0},{1,0,1,1},{1,1,0,0},{1,1,0,1},{1,1,1,0},\\
        {1,1,1,1}}

Total number of outcomes = 16

Probability function
{0,5/16}
{1,5/8}
{2,1/16}
{3,0}
```

Working in a similar way, find the probability functions of the random variables given below:

(i) We toss a coin six times. The random variable X counts the occurrences of the ordered pair HT.

(ii) We toss a coin six times. The random variable X counts the occurrences of two *successive tails*. Note that for outcomes of the form

$$HHTTTH, \quad TTHTTH, \quad TTHHTT$$

the random variable takes the value 2, for the outcome

$$TTTTHH$$

X takes the value 3, while for the outcome

$$TTTTTT$$

X takes the value 5.

(iii) When we toss a coin six times, X counts the number of times that three successive heads appear (the enumeration of successive heads is similar to that in Part (ii)).

(iv) Suppose we have two red dice and two black ones. We throw the four dice 30 times. The random variable X counts the number of times in which the sum of the two red dice is equal to the sum of the two black dice.

(v) An urn contains 20 balls numbered $1, 2, \ldots, 20$, while a second urn contains 30 balls numbered $1, 2, \ldots, 30$. We choose at random one ball from each urn. We want to find the probability functions of the variables X and Y defined as follows:

X: the larger of the two numbers in the balls drawn,

Y: the smaller of the two numbers in the balls drawn.

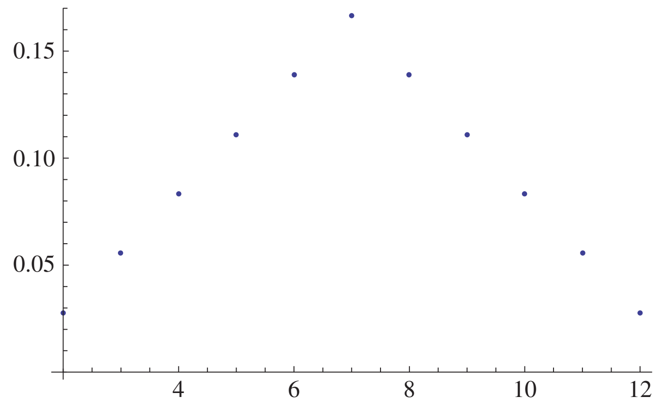
5. With the following program, we study the distribution of the sum of outcomes when two dice are thrown in succession. More specifically, we find the probability function, the expected value and the variance of that distribution, and we draw the graph of the probability function. Note that the values of the random variables that are of interest to us are collected in the variable *val* using the command

   Apply[Plus, res, 1]

   which finds the sums for the 36 possible outcomes of this experiment.

```
In[1]:= m=6;n=6;
        Print["List of all outcomes of the experiment"];
        res=Flatten[Table[{i,j}, {i,1,m}, {j,1,n}],1]
        total=Length[res];
        Print["Total number of outcomes = ", total];
        val=Apply[Plus,res,1];
        Do[s[i]=0,{i,2,12}]
        Do[y=val[[i]];
           s[y]=s[y]+1,
           {i,1,total}]
        Print["Probability function"];
        distr=Table[{i,s[i]/total},{i,2,12}]
        mu=Sum[i*s[i]/total,{i,2,12}];
        var=Sum[(i^2)*s[i]/total,{i,2,12}]-mu^2;
        Print["Mean of the distribution = ",mu]
        Print["Variance of the distribution = ",var]
        Print["Plot of the probability function"]
        ListPlot[distr]

List of all outcomes of the experiment
Out[3]= {{1,1},{1,2},{1,3},{1,4},{1,5},{1,6},{2,1},{2,2},\\
        {2,3},{2,4},{2,5},{2,6},{3,1},{3,2},{3,3},{3,4},\\
        {3,5},{3,6},{4,1},{4,2},{4,3},{4,4},{4,5},{4,6},\\
        {5,1},{5,2},{5,3},{5,4},{5,5},{5,6},{6,1},{6,2},\\
        {6,3},{6,4},{6,5},{6,6}}

Total number of outcomes = 36

Probability function
Out[10]= {{2,1/36},{3,1/18},{4,1/12},{5,1/9},{6,5/36},\\
         {7,1/6},{8,5/36},{9,1/9},{10,1/12},{11,1/18},\\
         {12,1/36}}

Mean of the distribution = 7

Variance of the distribution = 35/6
```

```
Plot of the probability function
Out[16]=
```



By a suitable modification of this program, study in a similar way the distributions for each of the following random variables:

(i) We throw a die twice, and define

X_1 : the absolute difference between the two outcomes,

X_2 : the difference of the first outcome minus the second one,

X_3 : the product of the two dice outcomes,

X_4 : the sum of the squares of the two dice outcomes.

(ii) An urn contains 10 balls numbered $5, 6, 7, \dots, 14$, while a second urn contains 8 balls numbered $3, 4, 5, \dots, 10$. We then define

X_1 : the sum of the two numbers in the balls drawn,

X_2 : the absolute difference between the two numbers in the balls drawn.

(iii) We throw a die three times and let Z_1, Z_2, Z_3 denote the respective outcomes. We then define

$$X_1 = Z_1 + Z_2 - Z_3,$$

$$X_2 = |Z_1 - Z_2| + |Z_3 - Z_2|,$$

$$X_3 = Z_1Z_2 - Z_3,$$

$$X_4 = \begin{cases} Z_1 + Z_2 + Z_3, & \text{if } Z_1 + Z_2 > Z_3, \\ 0, & \text{if } Z_1 + Z_2 = Z_3, \\ Z_1Z_2Z_3, & \text{if } Z_1 + Z_2 < Z_3. \end{cases}$$

(iv) We throw one die and toss three coins and define

X_1 : the sum of the outcome of the die and the number of heads,

X_2 : the difference between the outcome of the die and the numbers of tails,

X_3 : the product between the outcome of the die and the number of tails.

4.9 SELF-ASSESSMENT EXERCISES

4.9.1 True–False Questions

1. When we toss a coin five times, and X denotes the number of heads that appear, then the range of values for X is $\{0, 1, 2, 3, 4, 5\}$.
2. We throw a die repeatedly, and let Y be the number of throws until 5 appears for the first time. Then the range of values for Y is $\{0, 1, 2, 3, 4, 6\}$.
3. The probability function of a random variable is always a nondecreasing function.
4. Let f be a probability function of a random variable X. Then we have that $\lim_{x\to\infty} f(x) = 1$.
5. X is a discrete random variable with distribution function F. Then for any real a, we have
$$P(a < X) = F(a).$$
6. If a continuous random variable X has distribution function F, then for any a and b,
$$P(a < X < b) = F(b) - F(a).$$
7. If X is a discrete random variable with distribution function F, then for any a and b,
$$P(a \le X \le b) = F(b) - F(a).$$
8. Let a function f be such that
$$f(x) = \frac{x-1}{5}.$$
Then f can be the probability function of a random variable X with range $\{0, 1, 2, 3, 4\}$.
9. A random variable X has distribution function
$$F(t) = \begin{cases} 0, & t < 0, \\ \dfrac{1}{4}, & 0 \le t < 1, \\ 1 - \dfrac{3e^{-(t-1)}}{4}, & t \ge 1. \end{cases}$$
Then X is a continuous random variable.
10. Consider the probability function of a discrete random variable X defined by
$$f(x) = \frac{2x-1}{25}, \quad R_X = \{1, 2, 3, 4, 5\}.$$
The expected value of X is $19/5$.

11. For any random variable X, we have $\text{Var}(X) \geq E(X)$.

12. For any random variable X, the standard deviation of X is greater than or equal to its expected value.

13. When we toss a coin three times, the expected number of heads that appear is $3/2$.

14. Let X be the number of heads when we toss a coin twice. Then $E(X^2) = 3/2$.

15. When we toss a coin twice, the variance for the number of tails that appear is $1/2$.

16. For the random variable X, it is known that $E(X) = 1$ and $\text{Var}(X) = 2$. Then for the random variable $Y = 3X^2 - 5X$, we have $E(Y) = 1$.

17. If a random variable X takes two values x_1, x_2 with $x_1 \neq x_2$, then $\text{Var}(X) > [E(X)]^2$.

18. Let X be a random variable with $E(X) = 3$ and $E(X^2) = 9$. Then X can only take one value, i.e. there exists a c such that $P(X = c) = 1$.

19. Jimmy plays a game for which he wins \$4 with probability $2/3$, or else he wins \$8. Let X be his profit for this game. Then, an upper bound for the probability $P(X \geq 6)$ from Markov's inequality is $8/9$.

20. The random variable X takes only the values 10 and 20, each with equal probability. The upper bound we get for the probability $P(12 \leq X \leq 18)$ from Chebyshev's inequality is $25/9$.

4.9.2 Multiple Choice Questions

1. Let f be the probability function of a random variable X, given by

$$f(x) = \frac{cx^3}{9}, \quad R_X = \{1, 2, 3\}.$$

Then the value of c is

(a) 2 (b) $9/14$ (c) 4 (d) $1/4$ (e) $1/2$

2. X is a random variable that takes only the values $1, 2, 3$, and 4 in a way such that the probability of the event $\{X = x\}$ is proportional to x for $x = 1, 2, 3, 4$. Then the probability $P(X = 3)$ equals

(a) $3/10$ (b) $1/10$ (c) $1/4$ (d) $2/5$ (e) $3/5$

3. Andrea throws a die three times and let X be the number of sixes that appear. The range of values for X is

(a) $R_X = \{1, 2, 3\}$ (b) $R_X = \{1, 2, 3, 4, 5, 6\}$ (c) $R_X = \{0, 1, 2, 3\}$

(d) $R_X = \{0, 1, 2, 3, 4, 5, 6\}$ (e) none of the above

4. If the distribution function F of a variable X has a jump at the point a, then

   (a) X is a discrete random variable (b) X is a continuous random variable

   (c) $P(X = a) = F(a)$ (d) $P(X = a) = 0$ (e) $P(X = a) > 0$

5. The probability function f of the variable X is defined by

$$f(x) = \begin{cases} 1/8, & \text{if } x = 1, \\ cx/16, & \text{if } x = 2, 3, \text{ or } 4, \\ 1/64, & \text{if } x = 5. \end{cases}$$

   Then, the value of c is

   (a) $9/55$ (b) $55/36$ (c) $36/55$ (d) $55/9$ (e) $16/55$

6. For the random variable X, we know that

$$P(X = 1) = 0.3, \quad P(X = 2) = 0.2, \quad P(X = 5) = 0.5.$$

   Then, the expected value of X equals

   (a) $8/3$ (b) 2 (c) 1 (d) 3.2 (e) 2.7

7. Let f be a probability function of a random variable X defined by

$$f(x) = cx, \quad R_X = \{1, 2, 3\}.$$

   Then, the expected value of X equals

   (a) $7/6$ (b) $7/3$ (c) $1/6$ (d) $7/2$ (e) $14/9$

8. When Nicholas tosses a coin, he has a probability p of getting heads and a probability $q = 1 - p$ of getting tails. In a single coin toss, let X denote the number of heads. If $\text{Var}(X) = 3/16$, then the value of p is

   (a) $1/2$ (b) $1/4$ (c) $3/4$

   (d) either $1/4$ or $3/4$ (e) either $1/2$ or $3/4$

9. The probability function f of the variable X is defined by

$$f(x) = \begin{cases} 1/3, & \text{if } x = 0, \\ 1/6, & \text{if } x = 1, \\ 1/2, & \text{if } x = 2. \end{cases}$$

   The distribution function of X is then given by

(a) $F(t) = \begin{cases} 0, & t < 0, \\ 1/3, & 0 \leq t \leq 1, \\ 1/2, & 1 \leq t \leq 2, \\ 1, & t > 2 \end{cases}$

(b) $F(t) = \begin{cases} 0, & t \leq 0, \\ 1/3, & 0 < t \leq 1, \\ 1/2, & 1 < t \leq 2, \\ 1, & t > 2 \end{cases}$

(c) $F(t) = \begin{cases} 0, & t < 0, \\ 1/3, & 0 \leq t < 1, \\ 1/2, & 1 \leq t < 2, \\ 1, & t \geq 2 \end{cases}$

(d) $F(t) = \begin{cases} 0, & t \leq 0, \\ 1/3, & 0 < t < 1, \\ 1/2, & 1 \leq t \leq 2, \\ 1, & t > 2 \end{cases}$

(e) $F(t) = \begin{cases} 0, & t < 0, \\ 1/3, & 0 \leq t < 1, \\ 1/2, & 1 \leq t \leq 2, \\ 1, & t > 2. \end{cases}$

10. For the variable X, it is known that

$$E(2X - 4) = 4, \quad E(4X^2 - 3) = 71.$$

The variance of X equals

(a) $7/2$ (b) 1 (c) 55 (d) $21/2$ (e) $5/2$

11. Assume that for a variable X, we have

$$\sigma_X = 1/2, \quad E(X) = 0.$$

The value of $E(6X^2 + 5)$ equals

(a) $3/2$ (b) $13/2$ (c) 13 (d) 3 (e) 8

12. The random variable X takes on only the values $2, 4, 6$, each with equal probability. The standard deviation of X equals

(a) $\dfrac{56}{3}$ (b) $\sqrt{\dfrac{56}{3}}$ (c) $\sqrt{\dfrac{8}{3}}$ (d) $\sqrt{\dfrac{52}{3}}$ (e) $\sqrt{\dfrac{40}{3}}$

13. Jimmy plays a game for which if he wins, he receives $2c$ dollars and if he loses, he pays c dollars. If his probability of winning the game is $3/5$, and his expected earnings from this game are \$2, then the value of c is

(a) $5/2$ (b) $3/2$ (c) $6/5$ (d) 2 (e) 5

14. Jimmy plays another game for which if he wins, he receives λ dollars. If the probability that he wins the game is $2/3$, the amount he should pay if he loses so that the game is fair equals

(a) λ (b) 2λ (c) $3\lambda/2$ (d) $2\lambda/3$ (e) $\lambda/2$

15. For the random variable X, it is known that

$$E(X^4) = 50, \quad E(2X^2 + 7) = 17.$$

Then the variance of the variable $Y = X^2$ equals

(a) 25 (b) 50 (c) 75 (d) 94 (e) 100

16. For the random variable X, we know that

$$P(X = 1) = 1/2, \quad P(X = 2) = 1/4, \quad P(X = 4) = 3/16, \quad P(X = 8) = 1/16.$$

Then an upper bound for the probability $P(X \geq 4)$ from Markov's inequality is

(a) $9/8$ (b) $1/2$ (c) $3/4$ (d) $11/16$ (e) $9/16$

17. A random variable X has range $R_X = \{0, 1, 2, \ldots\}$, and suppose that

$$P(X > k) = \frac{1}{3^k}, \quad k = 0, 1, 2, \ldots$$

Then the expected value of X equals

(a) $1/3$ (b) 3 (c) $3/2$ (d) $2/3$ (e) $1/2$

4.10 REVIEW PROBLEMS

1. Let X be a discrete random variable with probability function

$$f(x) = \begin{cases} \dfrac{1}{2n}, & x = \pm 1, \pm 2, \ldots, \pm(n-1), \\ \dfrac{1}{n}, & x = 0. \end{cases}$$

(i) Show that the variable $Y = |X|$ has probability function

$$f_Y(y) = \frac{1}{n}, \quad y = 0, 1, \ldots, n-1,$$

i.e. Y has the discrete uniform distribution on the set $\{0, 1, 2, 3, \ldots, n-1\}$.

(ii) Calculate the mean and variance for X and Y.

2. A random variable X takes only the values a (with probability p) and $2a$ (with probability $1-p$) for some $0 < p < 1$. If it is known that $p \neq 1/2$ and that the mean and the variance of X are $E(X) = 3$ and $\mathrm{Var}(X) = 1$, find the values of a and p.

3. A discrete random variable X can take on only four values a_1, a_2, a_3, a_4 with respective probabilities

$$\frac{3\lambda - 2}{4}, \frac{2 - \lambda}{4}, \frac{2\lambda - 1}{4}, \frac{-4\lambda + 5}{4}.$$

   (i) What are the admissible values for the parameter λ?

   (ii) For $a_i = i$, find the value of λ such that $E(X) = 13/4$. For this value of λ, calculate the variance of X.

4. Let X be the largest outcome in the throws of two dice. In Example 4.8, we found that the probability function of X is given by

$$f(x) = P(X = x) = \frac{2x - 1}{36}, \quad x = 1, 2, 3, 4, 5, 6.$$

   (i) Write down the distribution function of X.

   (ii) Calculate the expectation of X in two ways: first, using the probability function in the above formula, and then, using the distribution function found in Part (i), and check that the results agree.

   Can you find the probability function and the distribution function of X in the case where we have k dice instead of just two dice?

5. A random variable X has its distribution function as

$$F(t) = \begin{cases} 0, & t < 0, \\ 1/5, & 0 \le t < 4/3, \\ 1/2, & 4/3 \le t < 2, \\ 4/5, & 2 \le t < 5/2, \\ 1, & t \ge 5/2. \end{cases}$$

   (i) What is the range of values for X?

   (ii) Calculate the probabilities

$$P(X = 2), \quad P(X = 2|X < 5/2), \quad P(X \ge 3/2|X < 5/2).$$

   (iii) Obtain the expected value of the variables X, $X^2 - 1$, and $\sqrt{X}$.

6. A small internet company sells ebooks. The company has paid \$200 to obtain a new book electronically and sells each copy of the book at a price of \$30. Let X stand for the number of sales the company will make for this particular book and suppose that X has a probability function

$$f(x) = \frac{7^{x-1}}{8^x}, \quad x = 1, 2, \ldots$$

   (i) Find the probability function of the company's profit (or loss).

   (ii) What is the overall probability that the company will make a profit from the sales of this ebook?

   (iii) What is the probability that the company will make a profit but that will be less than \$80?

7. Simon and Paul place n balls numbered $1, 2, \ldots, n$ in an urn and agree to play the following game. Simon pays the amount of a dollars to Paul, he then selects k balls randomly and receives from Paul an amount of i dollars, where i is the largest number shown in the k balls he selected.

   Show that for the game to be fair, we must have

$$a(k+1) = k(n+1).$$

8. The weekly demand for a magazine in a news shop can be represented by a random variable with the following probability function:

$$f(x) = 1 - \frac{x+5}{18}, \quad x = 7, 8, 9, 10.$$

   The owner of the shop buys each copy of the magazine for \$4 and she sells it at a price of \$6. If there is no possibility to return any unsold copies of the magazine, find the number of copies she should buy every week in order to maximize her expected profit.

9. An electronic system consists of two parts that function independently of one another. The lifetime T (in hours) for each part of the system is a random variable with distribution function

$$F(t) = c - \left(1 + \frac{t}{1000}\right) \mathrm{e}^{-t/1000}, \quad t \geq 0.$$

   (i) Find the value of the constant c.

   (ii) Suppose the system works only when both parts are functioning properly. What is the probability that the system works for at least 2000 hours? If it is known that the system has worked for 2000 hours, what is the probability that it will break down during the next 1000 hours?

10. A coin is tossed until either a head or four tails have occurred, and let X stand for the total number of coin tosses in this experiment.

   (i) What is the range of values for X?

   (ii) Calculate the probability function of X.

   (iii) Find the expected value of X.

11. The probability function of a discrete random variable X is given by

$$f(x) = \frac{c(n-x)}{n(n-1)}, \quad x = 1, 2, \ldots, n-1,$$

   for some integer $n \geq 2$ and a suitable constant c.

   (i) Prove that $c = 2$.

   (ii) Find the expected value $E(X)$ and the variance $\mathrm{Var}(X)$.

(iii) Assuming that $n \geq 5$, calculate the probabilities

$$P(X \geq 2), \quad P(X \leq n-2),$$

$$P(2 \leq X \leq n-2), \quad P(X \leq n-2 | X \geq 2).$$

12. Prove the following, which is a more general form of Markov's inequality than the one given in Proposition 4.11. Let X be a random variable that takes nonnegative values and for which $E(X^k)$ exists for some nonnegative integer k. Then, for any $t > 0$, we have

$$P(X \geq t) \leq \frac{E(X^k)}{t^k}.$$

13. The probability function of a variable X is given by

$$f(x) = \frac{4(x-1)}{3^x}, \quad x \in R_X = \{2, 3, 4, \ldots\}.$$

(i) Show that

$$P(X > n) = \frac{2n+1}{3^n}, \quad n = 2, 3, \ldots$$

(ii) Verify that the variable X satisfies the inequality

$$P(X > n + k | X > k) < P(X > k)$$

for any nonnegative integers n, k.

(*Hint*: Use the formulas

$$\sum_{n=0}^{\infty} t^n = \frac{1}{1-t}$$

and

$$\sum_{n=1}^{\infty} nt^{n-1} = \left(\sum_{n=1}^{\infty} t^n \right)' = \frac{1}{(1-t)^2},$$

valid for $|t| < 1$.)

14. Let X be a discrete random variable and g be a function such that the expectation $E[g(X)]$ exists. Show that for the expectation of the random variable $|g(X)|$, we have

$$|E[g(X)]| \leq E[|g(X)|].$$

15. A machine produces metal discs whose diameter D is a random variable with distribution function

$$F(t) = \begin{cases} 0, & t \leq a, \\ \dfrac{t-a}{b-a}, & a < t \leq b, \\ 1, & t > b, \end{cases}$$

where $a < b$ are two given positive numbers. Find the probability that the area of a metal disc produced by this machine is at least ϵ ($\epsilon > 0$).

4.11 APPLICATIONS

4.11.1 Decision Making Under Uncertainty

As we all know from our experience, a person has to make thousands of decisions every day. Decision making can be broadly defined as the choice of action among a set of alternatives. In most practical situations, decision making is carried out under uncertain conditions; in that case, each possible action is associated with an unknown outcome. The outcome may be financial (e.g. when we have to choose where to invest our money) or otherwise (e.g. whether to take the bus or subway to go to work, or what type of treatment we should have after an injury). In a scientific context, decision making occurs in nearly all areas of human thinking and research, and probability theory lies at the heart of its reasoning. Even from a less formal perspective, probabilistic arguments are a powerful tool in decisions we have to make in our everyday life and work. Let us illustrate this by the following example.

A company has 12 members of staff in managerial positions and wants to send eight of them to a conference so that they become acquainted with new services in their area of interest. The conference lasts for five days, and so the company needs to book accommodation for five nights per person that will be attending. The hotel where the staff members are going to stay offers a discount price of \$100 per person per night if the booking is made at least two months in advance, but charges a cancellation fee of \$200 for each person. The director of the company does not know how many persons will be available to attend the conference, but he estimates from past experience that, if all 12 of them are asked, the probability that x persons will accept is $1/5$ for $x = 8, 9, 10, 11, 12$. If the regular price for a person per night is \$200, should the company make any bookings in advance? If so, how many reservations should it make?

To answer these questions, one has to compute the cost for the company if

(a) the company makes i reservations, for $i = 0$ (no bookings are made in advance), and for $i = 8, 9, 10, 11, 12$;

(b) exactly x persons will be available to attend, for $x = 8, 9, 10, 11, 12$.

Let us denote by X the number of staff members who will be available to attend the meeting. It is clear that X is a random variable that has a discrete distribution with probability function

$$f(x) = \frac{1}{5}, \quad x = 8, 9, 10, 11, 12. \tag{4.14}$$

If the company proceeds to i reservations ($8 \le i \le 12$) at the reduced rate, the total cost (for a five-night accommodation) will be

$$Y_i = \begin{cases} 500X + 200(i - X), & X \le i, \\ 500i + 1000(X - i), & X > i, \end{cases}$$

that is,

$$Y_i = \begin{cases} 300X + 200i, & X \le i, \\ 1000X - 500i, & X > i. \end{cases}$$

For a given i, the variable Y_i is a function of the variable X and, to express this, we write $Y_i = g(X)$. We also write c_i for the expected cost for the company if i reservations are made, so that

$$c_i = E[Y_i] = E[g(X)].$$

Proposition 4.6 now yields

$$\begin{aligned} c_i &= \sum_{x=8}^{12} g(x)f(x) = \sum_{8\le x\le i} (300x + 200i)f(x) + \sum_{i<x\le 12} (1000x - 500i)f(x) \\ &= 300 \sum_{8\le x\le i} xf(x) + 1000 \sum_{i<x\le 12} xf(x) + 200i \sum_{8\le x\le i} f(x) - 500i \sum_{i<x\le 12} f(x). \end{aligned}$$

Taking into account that

$$\sum_{i<x\le 12} f(x) = \sum_{8\le x\le 12} f(x) - \sum_{8\le x\le i} f(x) = 1 - \sum_{8\le x\le i} f(x)$$

and

$$\sum_{i<x\le 12} xf(x) = \sum_{8\le x\le 12} xf(x) - \sum_{8\le x\le i} xf(x) = E[X] - \sum_{8\le x\le i} xf(x),$$

the expression for the expected cost c_i reduces to

$$c_i = 1000E[X] - 500i - 700 \sum_{8\le x\le i} xf(x) + 700i \sum_{8\le x\le i} f(x).$$

Since $f(x)$ is given by (4.14), the expected value of X is

$$E[X] = \sum_{8\le x\le 12} xf(x) = \frac{1}{5} \sum_{8\le x\le 12} x = 10,$$

while it is easy to see that

$$\sum_{8\le x\le i} f(x) = \sum_{x=8}^{i} \frac{1}{5} = \frac{i-7}{5}, \quad \text{for } i = 8, 9, 10, 11, 12,$$

and

$$\sum_{8\le x\le i} xf(x) = \frac{1}{5} \sum_{x=8}^{i} x = \frac{i(i+1) - 56}{10}, \quad \text{for } i = 8, 9, 10, 11, 12.$$

Consequently, if i reservations are made, we derive that the expected cost for the company equals

$$\begin{aligned} c_i &= 1000 \cdot 10 - 500i - 700 \cdot \frac{i(i+1) - 56}{10} + 700 \cdot \frac{i(i-7)}{5} \\ &= 10000 - 500i - 70i(i+1) + 70 \cdot 56 + 140i(i-7) \\ &= 13920 - 1550i + 70i^2. \end{aligned}$$

The numerical values for c_i for all possible choices the company can make are as follows:

i	c_i
8	6000
9	5640
10	5420
11	5340
12	5400

On the other hand, if the company does not make any reservations, the minimum cost will arise when only eight members of staff attend the conference. This cost amounts to $8 \times 1000 = \$8000$. Combining all the above, we see that the best decision for the company, in order to minimize its expected cost, is to proceed to 11 reservations.

Clearly, this choice depends critically on the given values of the *parameters* for the probability model that we have assumed (that is, accommodation costs and cancellation charges) as well as the probability function of the variable X. If (at least) one of them changes, or is misspecified, the optimal choice for the number of reservations might be altered; the process that examines how the result, or a decision, based on a model may change if one or more of the model parameters change is known as **sensitivity analysis**.

Also, in the above discussion, the criterion we have used in order to obtain the optimal choice for the number of reservations i was the **minimization of the expected cost**. In some financial applications, one of the available scenarios may carry a small expected cost, yet it may involve a huge financial loss. Even if this loss has a very small probability to occur, it might be so large that the company (or the individual) that has to make the decision cannot afford it. In such cases, minimizing the expected cost may be inappropriate and some other criteria may be employed for making the "best decision."

KEY TERMS

absolute moments
binary random variable
central moments
Chebyshev inequality
discrete random variable
discrete uniform distribution
distribution function (or cumulative distribution function)
factorial moments
heavy tailed distribution
jump (of a step function)
location parameter
Markov inequality
mean (or mean value, expected value, mathematical expectation)
median
moments
parameter
probability mass function (or probability function)
random variable
right continuous function
standard deviation
standardized random variable
symmetric (around zero) distribution function
variance

CHAPTER 5

SOME IMPORTANT DISCRETE DISTRIBUTIONS

Jacob Bernoulli
(Basel 1654–1705)

He was one of the many outstanding mathematicians in the Bernoulli family; in fact, he initiated the "Bernoulli dynasty" in mathematics. He was born in 1654 which, incidentally, is also the year that many people perceive as the birth year of probability theory. For instance, Tom Apostol (1969) writes "A gambler's dispute in 1654 led to the creation of a mathematical theory of probability by two famous French mathematicians, Blaise Pascal and Pierre de Fermat."

Along with his brother Johann, he was among the first to understand and put forward the work of G. Leibniz on calculus. Although he made profound and extensive advances in calculus, his most important contribution was in probability theory. In his work *Ars Conjectandi* (The Art of Conjecturing) published in 1713, after his death, he derived the first version of one of the most important theorems in probability theory, the law of large numbers. This book is among the first publications that put probability on a sound mathematical basis. He is also the first who distinguished between objective and subjective probability (the latter representing the degree of belief about a statement, supported by evidence which may not be of a statistical nature).

In his book "The History of Statistics: The Measurement of Uncertainty Before 1900," Stephen Stigler states *The Bernoullis are surely the most renowned family in the history of mathematical sciences. Perhaps as many as 12 Bernoullis have contributed to some branch of mathematics and physics and at least five have written on probability. So large is the set of Bernoullis that chance alone may have made it inevitable that a Bernoulli should be designated father of the quantification of uncertainty. The individual in question is Jacob Bernoulli, … contemporary and occasional rival of Isaac Newton.*

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

5.1 BERNOULLI TRIALS AND BINOMIAL DISTRIBUTION

Jacob Bernoulli (1654–1705) is one of the major early contributors in the development of probability theory. As many other prominent mathematicians of that time, he considered the application of probability to games of chance. In his most famous work, *Ars Conjectandi*, he considered what is now known as a *Bernoulli trial*, which forms the basis for some of the most important discrete distributions that we shall discuss in this chapter.

By the term Bernoulli trial, we simply mean an experiment with only two possible outcomes; for convenience, these outcomes are often termed "Success" (abbreviated by S) and "Failure" (denoted by F), although it should be clear from the start that "Success" does not necessarily mean that something positive or pleasant occurs. Usually, the event that we term a success is the one for which we wish to make probability statements about. For instance, when we consider whether or not a certain person (or a group of persons) is affected by a disease within a certain time span, if the interest lies predominantly on the probability that the person(s) gets the disease, then we may call the event of the disease as the "success" outcome for the experiment. Similarly, if we want to see what is the probability that an item produced by a production line is defective, we may consider as success the event that the item is actually defective.

The definition of a Bernoulli trial is so simple and general that it represents numerous practical situations: the gender of a child (girl or boy) to be born, the outcome of a coin toss (heads or tails), the improvement (or not) of the health status for someone receiving a certain medical treatment, the future outcome (profitable or nonprofitable) of a financial investment, the result of a test such as a University exam (pass or failure), and so on.

It is worth noting at this point that there are instances where we have repeated experiments with more than two possible outcomes, yet by grouping these outcomes we may arrive at a sequence of Bernoulli trials. A typical example is tossing a die, wherein we may group the six possible outcomes into two classes: "odd integer" and "even integer." Another grouping, in the same experiment, could be obtained if we call the appearance of 6 a success, and the appearance of $1, 2, 3, 4$, or 5 a failure; thus, we create two disjoint sets of outcomes, $\{6\}$ and $\{1, 2, 3, 4, 5\}$ in which the first corresponds to S and the second to F.

More generally, and using the notation above, we see that the sample space of a single Bernoulli trial is

$$\Omega = \{S, F\}.$$

From here on we shall denote the probability of the event we termed "Success" by p and the probability of the event corresponding to "Failure" by q. Since there are only two outcomes, one of these events must occur and thus

$$p + q = 1, \quad 0 \leq p \leq 1, \quad 0 \leq q \leq 1.$$

Definition 5.1 Let X be the number of successes in a Bernoulli trial in which the probability of success is p and the probability of failure is $q = 1 - p$. The distribution of the binary random variable X is called the **Bernoulli distribution** with parameter p.

It is clear that a variable having a Bernoulli distribution may only take the values 0 and 1 and the probability function of X is given by

$$f(x) = P(X = x) = \begin{cases} p, & \text{if } x = 1, \\ q = 1 - p, & \text{if } x = 0. \end{cases}$$

Alternatively, the probability function may be written in the form

$$f(x) = p^x q^{1-x}, \quad x = 0, 1.$$

The (cumulative) distribution function F of X is then given by

$$F(t) = \begin{cases} 0, & \text{if } t < 0, \\ q, & \text{if } 0 \le t < 1, \\ 1, & \text{if } t \ge 1. \end{cases}$$

Figures 5.1 and 5.2 present, respectively, the probability function and the distribution function of the Bernoulli distribution.

The expectation of X, $\mu = E(X)$, is

$$\mu = \sum_{x=0}^{1} x f(x) = 0 \cdot f(0) + 1 \cdot f(1) = f(1) = p,$$

while the second moment around zero is

$$\mu_2' = E(X^2) = \sum_{x=0}^{1} x^2 f(x) = 0^2 \cdot f(0) + 1^2 \cdot f(1) = f(1) = p.$$

Consequently, the variance of X, $\sigma^2 = \mathrm{Var}(X)$, equals

$$\sigma^2 = \mathrm{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = pq.$$

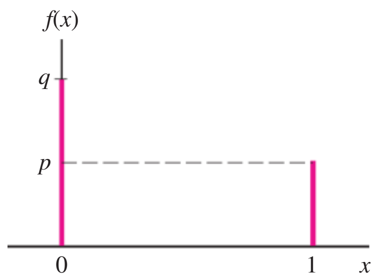


Figure 5.1 The probability function of a Bernoulli random variable.

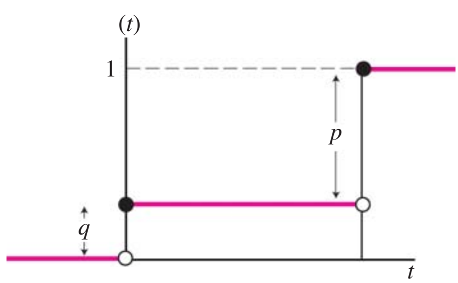


Figure 5.2 The distribution function of a Bernoulli random variable.

Example 5.1 Suppose that one selects a card at random from a pack of 52 cards. Let success denote the event that an ace is drawn, so that $X = 1$ if this occurs, and $X = 0$ otherwise. Find the probability function, distribution function, expectation, and variance of X.

SOLUTION Clearly, X has a Bernoulli distribution. Its parameter is the probability of success (picking out an ace) when selecting a card and, since the pack contains 52 cards of which 4 are aces, this probability equals

$$p = P(X = 1) = \frac{4}{52} = \frac{1}{13}.$$

Thus, the probability function is

$$f(x) = \left(\frac{1}{13}\right)^x \left(1 - \frac{1}{13}\right)^{1-x}, \quad x = 0, 1,$$

while the corresponding distribution function is given by

$$F(t) = \begin{cases} 0, & \text{if } t < 0, \\ \dfrac{12}{13}, & \text{if } 0 \le t < 1, \\ 1, & \text{if } t \ge 1. \end{cases}$$

The mean and variance are, respectively,

$$E(X) = \frac{1}{13}, \quad \mathrm{Var}(X) = \left(\frac{1}{13}\right)\left(\frac{12}{13}\right) = \frac{12}{169}.$$

The Bernoulli distribution appears simple enough to merit further study. When one considers sequences of Bernoulli trials rather than a single trial, the discussion becomes more fruitful and the questions that one might ask are much less trivial. We mentioned above that in a single trial, the sample space consists of two elements, namely, $\Omega = \{S, F\}$. If we have two trials instead (a simple example is provided by two successive tosses of a coin), the sample space has four elements, namely,

$$\Omega = \{SS, SF, FS, FF\}.$$

Here, for example, SF stands for a success in the first trial and a failure in the second trial. For three trials, the sample space has $2^3 = 8$ elements, and so on. It is clear from above that two separate cases may occur, depending on whether we treat the events SF and FS as being distinguishable or not. If one is interested in *the number of successes only*, and not in the order these successes were obtained, then these two events are not distinguished, so that there are only three distinct events which may occur: two successes (SS), one success (SF or FS), zero successes (FF).

More generally, if n coins are tossed successively, we might be interested in the number of heads (successes) that occurred, but not in the order of appearance of these heads in the

n trials. This is a first example of n independent repetitions of a Bernoulli trial, and our main focus here is on the number, X, of successes in these trials.

Definition 5.2 Let X be the number of successes in a sequence of n independent Bernoulli trials, each having the same probability of success, p, so that $q = 1 - p$ is the probability of failure. The distribution of X is then called a **binomial distribution** with parameters n and p, and is denoted by $b(n, p)$.

When a random variable X has a $b(n, p)$ distribution, we denote this by

$$X \sim b(n, p),$$

where, here and in the sequel, the symbol "∼" means "has the distribution." Further, from the definition it is obvious that the range of values for X is

$$R_X = \{0, 1, 2, \ldots, n\}.$$

For the distribution of X, we have the following result.

Proposition 5.1 *The probability function of the $b(n, p)$ distribution is given by*

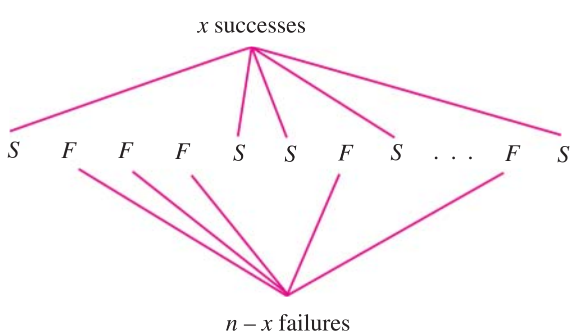
$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n, \tag{5.1}$$

and $P(X = x) = 0$ for all other values of x.

Proof: The elements of the sample space Ω for this experiment (the entire sequence of n independent Bernoulli trials) are n – tuples of successes (S) and failures (F), with a typical element having the form

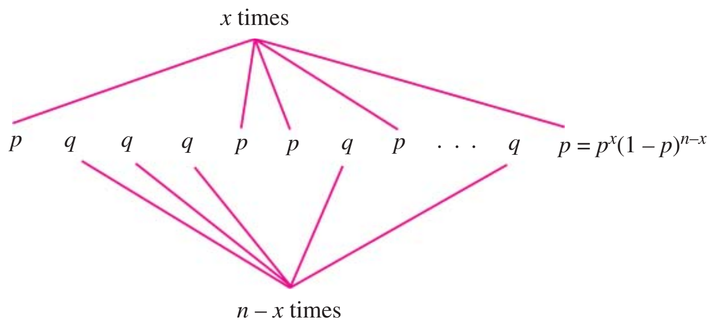
$$SSFSFFS \ldots FS.$$

Now, let x be a fixed number in the range R_X of the random variable X, so that $0 \le x \le n$. The event $\{X = x\}$ consists of all those elements of the sample space Ω that have exactly x successes and $n - x$ failures. Any such element, i.e. a realization of x successes, has the form



$$SFFFSSFS \ldots FS.$$

Because we assume that trials are independent, the probability that the above event is the outcome of the experiment (so that the first trial results in a success, the next three trials result in a failure followed by two successes, and so on), is the product of the corresponding probabilities for each trial, namely



$$pqqqppqp \ldots qp = p^x q^{n-x}.$$

But there are $\binom{n}{x}$ such sequences, consisting of exactly x successes and $n - x$ failures (this is just the same as placing x S's in n available positions, and then filling in the remaining $n - x$ positions with F's.). This yields

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

while $P(X = x) = 0$ for any other value of x. □

It is obvious that when $n = 1$, the binomial distribution reduces to the Bernoulli distribution. Moreover, if we define the variables

$$X_i = \begin{cases} 1, & \text{if the outcome of the } i\text{th trial is } S, \\ 0, & \text{if the outcome of the } i\text{th trial is } F, \end{cases}$$

for $i = 1, 2, \ldots, n$, then each X_i has a Bernoulli distribution with parameter p, and their sum

$$X = X_1 + X_2 + \cdots + X_n$$

gives the number of successes in n trials, so that X has a $b(n, p)$ distribution.

Example 5.2 Jack and Jill meet every Wednesday to play a table tennis match. If the probability that Jack wins any particular game is 0.6, what is the probability that in the first 10 games

(i) Jack wins all 10?

(ii) Jill wins at most two games?

(iii) Jill wins at least six games?

SOLUTION Let X be the number of games that Jill wins in the series of 10 games. The probability that she wins in any game is $1 - 0.6 = 0.4$. Then, under the assumptions of the example, and since results in consecutive games are thought to be independent, X has a $b(10, 0.4)$ distribution.

(i) We want the probability that Jill does not win in any of the 10 games, i.e. $P(X = 0)$. By formula (5.1), we have

$$P(X = 0) = \binom{10}{0}(0.4)^0(0.6)^{10} = (0.6)^{10} = 0.006,$$

i.e. about 0.6%.

(ii) Now we seek the probability that X, the number of Jill's wins, is at most two, i.e. $P(X \leq 2)$. This is

$$\begin{aligned} P(X \leq 2) &= P(X = 0) + P(X = 1) + P(X = 2) \\ &= \binom{10}{0}(0.4)^0(0.6)^{10} + \binom{10}{1}(0.4)^1(0.6)^9 \\ &\quad + \binom{10}{2}(0.4)^2(0.6)^8 \\ &= (0.6)^{10} + 10(0.4)(0.6)^9 + 45(0.4)^2(0.6)^8 \\ &= 0.0060 + 0.0403 + 0.1209 = 0.1672. \end{aligned}$$

(iii) The probability required here is $P(X \geq 6)$, which is given by

$$\begin{aligned} P(X \geq 6) &= P(X = 6) + P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) \\ &= \binom{10}{6}(0.4)^6(0.6)^4 + \binom{10}{7}(0.4)^7(0.6)^3 + \binom{10}{8}(0.4)^8(0.6)^2 \\ &\quad + \binom{10}{9}(0.4)^9(0.6)^1 + \binom{10}{10}(0.4)^{10}(0.6)^0 \\ &= 0.1115 + 0.0425 + 0.0106 + 0.0016 + 0.0001 = 0.1662. \end{aligned}$$

The name of the binomial distribution comes from the fact that the probabilities in (5.1) are in fact the terms of the binomial expansion of $(p + q)^n$. This also shows formally that

$$\sum_{x \in R_X} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} = (p + q)^n = 1,$$

which proves that $f(x)$ is a proper probability mass function.

The cumulative distribution function of the binomial distribution is given by

$$F(t) = \begin{cases} 0, & \text{if } t < 0, \\ \displaystyle\sum_{x=0}^{[t]} \binom{n}{x} p^x q^{n-x}, & \text{if } 0 \leq t < n, \\ 1, & \text{if } t \geq n. \end{cases}$$

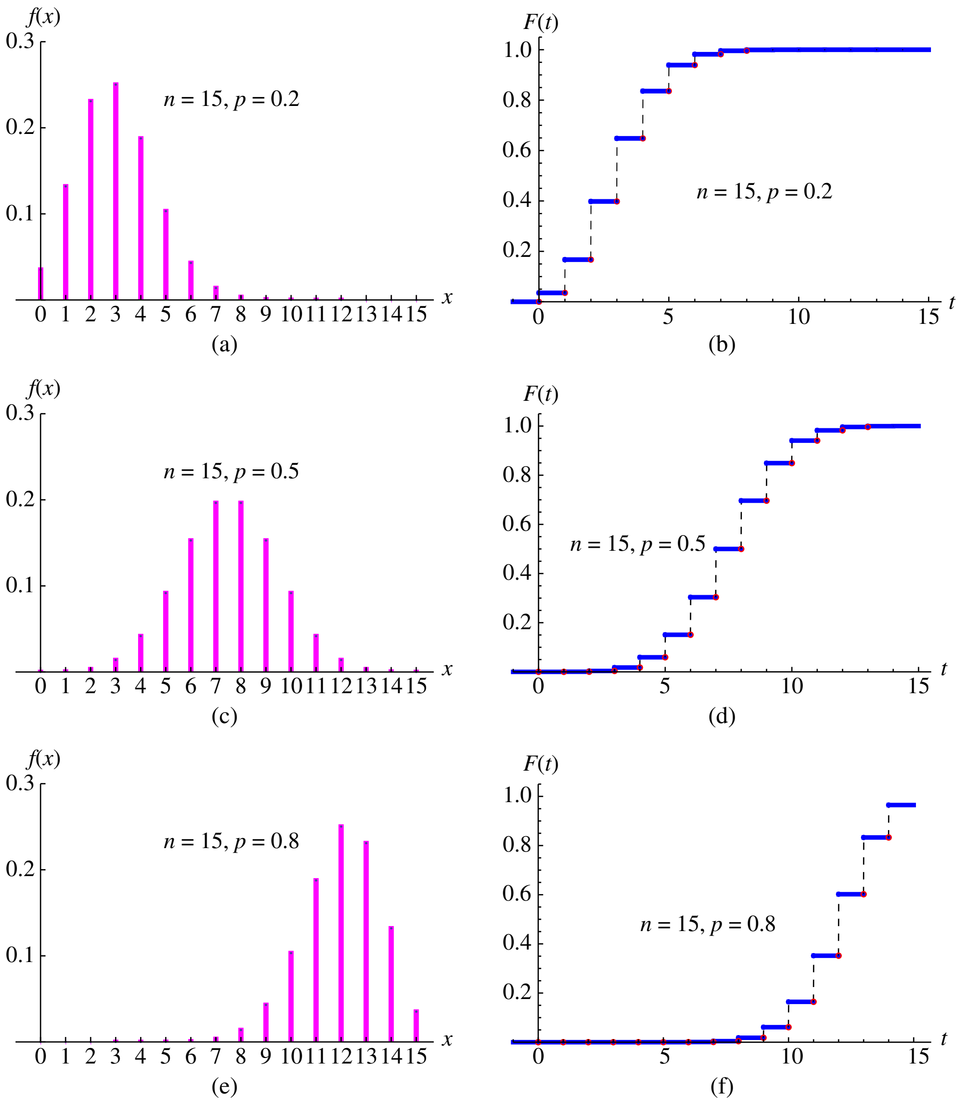


Figure 5.3 The probability function (left) and the cumulative distribution function (right) of the binomial distribution for various choices of n and p.

Figure 5.3 displays the probability function and the cumulative distribution function for a random variable X that follows the binomial distribution $b(n, p)$, for several values of the parameters n and p.

We now calculate the mean and variance of the binomial distribution. First, we recall the following two formulas from Chapter 2:

$$\sum_{x=0}^{n} x \binom{n}{x} p^x q^{n-x} = np, \quad \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} = n(n-1)p^2.$$

The first of these results yields immediately that $E(X) = np$, when $X \sim b(n, p)$, while from the second we obtain the second factorial moment of X as

$$\mu_{(2)} = E[X(X-1)] = n(n-1)p^2.$$

The following proposition, which gives the mean and the variance of a binomial random variable, is now an easy deduction from the above results.

Proposition 5.2 *The mean and variance of a variable X that follows the binomial distribution with parameters n and p are given by*

$$\mu = E(X) = np, \quad \sigma^2 = \mathrm{Var}(X) = npq.$$

Proof: The result about $E(X)$ is already proved above. For the variance, we may write

$$\begin{aligned}\mathrm{Var}(X) = E(X^2) - \mu^2 &= E[X(X-1) + X] - \mu^2 \\ &= E[X(X-1)] + E(X) - \mu^2 = \mu_{(2)} + \mu - \mu^2\end{aligned}$$

and the result follows by using the facts that $\mu = np$ and $\mu_{(2)} = n(n-1)p^2$. □

We now present some examples to illustrate the usefulness of the binomial distribution in practice.

Example 5.3 Thirty people, who are of the same age and the same health status, are insured with the same insurance company. Using life tables, the company estimates that the probability for a randomly chosen person among these 30 to be alive in 15 years from now is 80%.

(i) What is the probability that not all 30 people will be alive in 15 years' time?

(ii) What is the probability that at least one person will be alive?

(iii) Every insured person who will be alive after 15 years will receive \$100000. Find the expected amount of money the company will have to pay at the end of the 15-year period and the standard deviation of that amount.

SOLUTION Let X be the number of individuals who will be alive at the end of the 15-year period. Since lives of different people are assumed to be independent, X has a binomial distribution with parameters $n = 30$ and $p = 0.8$.

(i) We want $P(X < 30)$. This is

$$\begin{aligned}P(X < 30) = 1 - P(X = 30) &= 1 - \binom{30}{30}(0.8)^{30}(0.2)^0 = 1 - (0.8)^{30} \\ &= 0.998\,766 \cong 99.9\%.\end{aligned}$$

(ii) Here we seek $P(X > 0)$, which is equal to

$$P(X > 0) = 1 - P(X = 0) = 1 - \binom{30}{0}(0.8)^0(0.2)^{30} = 1 - (0.2)^{30}$$
$$= 1 - 1.1 \cdot 10^{-21} \cong 100\%.$$

(iii) The total amount the insurance company will have to pay at the end of the 15-year period is represented by the random variable $Y = 100\,000X$. Since

$$E(X) = np = 30 \cdot (0.8) = 24, \quad \mathrm{Var}(X) = npq = 30 \cdot (0.8) \cdot (0.2) = 4.8,$$

we get

$$E(Y) = 10^5 \cdot E(X) = \$2\,400\,000$$

and

$$\mathrm{Var}(Y) = 10^{10} \cdot \mathrm{Var}(X) = 4.8 \cdot 10^{10} \quad \text{(dollars squared)},$$

so that the standard deviation of the total payment equals

$$\sigma_Y = \sqrt{\mathrm{Var}(Y)} = 10^5\sqrt{4.8} = \$219\,089.$$

Example 5.4 At a dinner on a Saturday night at a restaurant, a company of eight friends has the choice of fish or meat as their main course, while for dessert they may choose either créme caramel or black forest cake. Assuming that 70% of people order meat in their main course and 60% order black forest cake as their dessert, find the probability that

(i) a specific person chooses fish as a main course and black forest cake for dessert;

(ii) among the eight persons, no one chooses the combination of fish and black forest cake;

(iii) among the eight persons, at least six choose the combination of meat and créme caramel.

SOLUTION

(i) Assuming that the choice of main course is independent of the choice of dessert, the events

A: a person chooses fish as a main course and
B: a person chooses black forest cake as a dessert

are independent. We know that 70% of people prefer meat; thus, the probability that someone orders fish is 30%, i.e. $P(A) = 0.3$, and since the choice of black forest cake has probability 0.6, the required probability is

$$P(AB) = P(A)P(B) = 0.3 \cdot 0.6 = 0.18.$$

(ii) Treating a person's choice (for the combination of main course and dessert) as a Bernoulli trial, we have eight such trials. Let X be the number of people who choose the combination in the statement of the question, namely, fish and black forest cake. Now, since from (i) this combination has a probability of 0.18 to be chosen, we see that $X \sim b(8, 0.18)$. Thus, we get

$$P(X = 0) = \binom{8}{0}(0.18)^0(1 - 0.18)^8 = 0.2044,$$

or about 20%.

(iii) Next, let Y be the number of persons who order both meat and créme caramel. This order has a probability $0.7 \cdot 0.4 = 0.28$, and so we have $Y \sim b(8, 0.28)$. This gives the required probability to be

$$\begin{aligned} P(Y \geq 6) &= P(Y = 6) + P(Y = 7) + P(Y = 8) \\ &= \binom{8}{6}(0.28)^6(1 - 0.28)^2 + \binom{8}{7}(0.28)^7(1 - 0.28)^1 \\ &\quad + \binom{8}{8}(0.28)^8(1 - 0.28)^0 \\ &= 0.006\,99 + 0.000\,78 + 0.000\,03 = 0.007\,81. \end{aligned}$$

Example 5.5 Suppose Jimmy shoots against a target with probability of hitting the target being p, and successful shootings are independent. He is very keen to win a prize, which is awarded if his successful attempts exceed the ones he missed the target. He is offered a choice of either three shots or five shots. For which values of p should he prefer the five shots?

SOLUTION Let X be the random variable that represents the number of successful attempts if he chooses to shoot five times, and Y be the random variable that represents the number of successful attempts if he chooses to shoot three times. In the former case, he wins if $X \geq 3$, while in the latter if $Y \geq 2$. Therefore, we want to find under what condition on p the statement

$$P(X \geq 3) > P(Y \geq 2)$$

holds true. It is clear that the distributions of X and Y are $b(5, p)$ and $b(3, p)$, respectively. Thus,

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= \binom{5}{3}p^3(1 - p)^2 + \binom{5}{4}p^4(1 - p) + \binom{5}{5}p^5(1 - p)^0 \\ &= 10p^3(1 - p)^2 + 5p^4(1 - p) + p^5, \end{aligned}$$

while

$$\begin{aligned}P(Y \geq 2) &= P(Y = 2) + P(Y = 3)\\ &= \binom{3}{2} p^2(1-p) + \binom{3}{3} p^3(1-p)^0\\ &= 3p^2(1-p) + p^3.\end{aligned}$$

Hence, Jimmy should prefer the five shots if his success probability satisfies

$$10p^3(1-p)^2 + 5p^4(1-p) + p^5 > 3p^2(1-p) + p^3.$$

After some straightforward calculations, we see that this reduces to

$$3p^2(1-p)^2(2p-1) > 0$$

so that the five shots are preferable if

$$p > 1/2.$$

Example 5.6 Assuming that having a boy or a girl is equally likely, how many children should a family have so that the probability of having children of both genders is at least

(i) 90%?

(ii) 99%?

SOLUTION For a family with n children, the distribution of the number X of girls is $b(n, 1/2)$. For the event that the family has children of both genders, we must have $X > 0$ and $n - X > 0$, since $n - X$ is the number of boys in the family. This is equivalent to the simultaneous occurrence of the events $\{X > 0\}$ and $\{X < n\}$, and the probability for this is $P(0 < X < n)$. We may write

$$P(0 < X < n) = 1 - P(X = 0) - P(X = n)$$

and since for $x = 0, 1, 2, \ldots, n$,

$$P(X = x) = \binom{n}{x}\left(\frac{1}{2}\right)^x\left(\frac{1}{2}\right)^{n-x} = \binom{n}{x}\left(\frac{1}{2}\right)^n,$$

from the last two expressions we get

$$\begin{aligned}P(0 < X < n) &= 1 - \binom{n}{0}\left(\frac{1}{2}\right)^n - \binom{n}{n}\left(\frac{1}{2}\right)^n\\ &= 1 - \frac{1}{2^{n-1}}.\end{aligned}$$

If we require that $P(0 < X < n) \geq 1 - \alpha$ for some $\alpha \in (0, 1)$, in view of the last expression we see that this is the same as

$$1 - \frac{1}{2^{n-1}} \geq 1 - \alpha \quad \text{or equivalently} \quad \frac{1}{2^{n-1}} \leq \alpha.$$

Taking logarithms on both sides, we deduce that the desired condition is

$$-(n-1)\ln 2 \leq \ln \alpha$$

and we finally obtain that

$$n \geq 1 - \frac{\ln \alpha}{\ln 2}.$$

(i) For $\alpha = 0.1$, the last expression yields

$$n \geq 1 - \frac{\ln 0.1}{\ln 2} \cong 4.32.$$

Thus, the family must have at least five children, so that that the probability of having children of both genders is at least 90%.

(ii) Here we have $\alpha = 0.01$, so that we obtain

$$n \geq 1 - \frac{\ln 0.01}{\ln 2} \cong 7.64.$$

Therefore, the minimum number of children a family must have in order to have children of both genders with a probability at least 99% is 8.

EXERCISES

Group A

1. If four dice are thrown simultaneously, find the probability that
   (i) the number six appears in exactly one die;
   (ii) the number six appears in at least one die.

2. An urn contains 100 balls, numbered 1–100. We select a ball at random. We consider as success the event that the number on the ball selected is divisible by 6. Write down the probability function, the cumulative distribution function, the mean, and variance of the Bernoulli distribution for this experiment.

3. If we throw a well-balanced die seven times, what is the probability that
   (i) no sixes or aces will appear?
   (ii) at least twice the outcome will be an odd integer?
   (iii) no more than three times the outcome will be greater than two?

4. The probability that a car is stolen when it is parked overnight in an unsafe area of a city is 10%. If there are 12 cars parked on a particular street of that area, what is the probability that during a night

   (i) no car is stolen?

   (ii) at most two cars are stolen?

   (iii) at least nine cars are stolen?

5. Suppose a mouse moves on a line at random so that, at each step, it either makes a move to the left or to the right with equal probability. Find the probability that the mouse will be at the position where it started after (i) 10 moves; (ii) 20 moves.

6. Peter plays a game in which he throws a die four times and he wins a dollars if at least one six appears. He is offered an alternative option: to throw two dice 24 times and win a dollars if a double six appears at least once. Should he switch to this new game?

7. (i) In a family with four children, what is the probability of having two boys and two girls, assuming that both genders are equally likely?

   (ii) If we select 10 families, each having four children, find

      (a) the probability that at least seven of them will have two boys and two girls;

      (b) the expected number of families with two boys and two girls.

8. Plates are produced at a certain factory department in dozens. Assume that the probability of a plate being nondefective is p, while the probability of a defective plate is $q = 1 - p$. Let X be the number of nondefective plates in a dozen, so that $X \sim b(12, p)$. From past experience, it has been estimated that it is twice more likely for a dozen to have no defective plates compared with the probability of having exactly one defective plate.

   (i) Find the probability that an item is defective.

   (ii) Hence estimate the probability that, in a given dozen, there are at most two defective plates.

9. A computer selects digits at random. How many digits are needed to be chosen so that the digit seven appears with a probability of at least $1/2$?

10. At a hospital, 15 babies are born on a given day. If we know that seven of the newborn babies are boys, what is the probability that all seven boys were born in the last seven births at the hospital on that day?

11. A mathematics professor, when leaving for work, forgets his car keys with a probability p_1 and his office keys with a probability p_2 (we assume that these two events are independent).

    (i) What is the probability that during a week (i.e. in five working days) there will be at least one day when he forgets both sets of keys?

    (ii) What is the probability that in exactly three days he will forget at least one set of keys?

12. Let X be a random variable having the binomial distribution with parameters n and p for some $p \in (0, 1)$ and Y be another random variable having the binomial distribution with parameters n and q, where $q = 1 - p$. Show that for $x = 0, 1, 2, \ldots, n$, we have
$$P(X = x) = P(Y = n - x).$$

Group B

13. Among the claims that are received by an insurance company, it is estimated that 3% exceed a specified amount c and those are registered as "large claims."
    (i) What is the probability that in the first 30 claims at most three will be large?
    (ii) What is the probability that the third large claim received by the company will arrive after at least 30 claims have been received in total?

14. Let X be a random variable having the binomial distribution with parameters n and p.
    (i) Show that its probability function $f(x), x = 0, 1, \ldots, n$, can be calculated via the recursive relation
    $$f(x) = \left(a + \frac{b}{x}\right) f(x-1), \quad x = 1, 2, 3, \ldots, n,$$
    where
    $$a = -\frac{p}{1-p}, \quad b = \frac{(n+1)p}{1-p},$$
    with initial condition $f(0) = (1-p)^n$.
    (ii) Show that $f(x) > f(x-1)$ if and only if $x < (n+1)p$.
    (iii) Verify that, if the product $(n+1)p$ *is not an integer*, f attains its maximum at the point $x_0 = [(n+1)p]$. What happens if $(n+1)p$ is an integer?

    Application: A footballer, who likes to take penalty kicks, has an 80% probability of scoring in each penalty. If he shoots 10 penalty kicks for his team during a football season, for what value of k is the probability that he scores k goals in these 10 attempts maximized? For this value of k, what is the probability that he scores k goals in 10 penalty kicks?

15. An investor invests the same amount of money in three funds. In each of these, he estimates that there is a probability $p = 3/5$ of making a profit of \$15 000, or else he loses \$10 000. Find the investor's expected profit.

16. Lisa, Tony, and Tom take part in a TV quiz and at some point they have to answer the same question. They do not have the same probability of answering the question correctly, but their respective probabilities are p, q, and r.
    (i) Find the probability that k persons answer the question correctly, for $k = 0, 1, 2, 3$ (do not forget to check that this is a proper probability function by adding these probabilities).

(ii) If the probability of a correct answer for each person remains the same in subsequent questions, derive the probability that when they are asked five questions,

(a) Lisa will be the only person to answer all five correctly;

(b) there will be three questions answered correctly by all three contestants.

17. At the end of the academic year, Sophie will sit in six exams, of which three she considers "easy" as she feels that she has a probability p of passing the exam. Sophie thinks the other three exams are more difficult. In particular, for each of these the probability of passing the exam is d, with $d < p$. The outcomes of all the exams are considered to be independent.

(i) What is the probability that she passes all six exams?

(ii) What is the probability that she passes five among the six exams?

(iii) What is the probability that she passes as many "easy" exams as "difficult" ones?

18. A jury has three members. Each of these members makes the correct decision with probability p, independent of the other jury members. The jury's decision is based on the majority rule.

(i) What is the probability that the jury makes the correct decision?

(ii) What is the probability that all three members make the same decision (whether correct or not)?

(iii) Show that for $p \geq 3/4$, the probability that the jury makes the correct decision is at least 84%.

19. A basketball player makes free throws successfully with a probability 0.85. In a particular game, he attempted 16 free throws and scored in 11 of them.

(i) What is the probability that the five throws he missed were the first five he attempted?

(ii) What is the probability that the first three shots he attempted were successful?

20. Let X be a random variable with a $b(n, p)$ distribution. Show that for any positive integer r, the rth factorial moment of X

$$\mu_{(r)} = E[(X)_r] = E[X(X-1)(X-2)\cdots(X-r+1)]$$

is given by the formula

$$\mu_{(r)} = (n)_r p^r = n(n-1)(n-2)\cdots(n-r+1)p^r, \quad 1 \leq r \leq n.$$

In particular, for $r = 1, 2$, verify that we get the same results as those derived in Proposition 5.2.

$\Bigg($*Hint*: You may find it useful to use the identity

$$\sum_{k=r}^{n} \binom{n}{k}\binom{k}{r} x^k = \binom{n}{r} x^r (1+x)^{n-r}.\Bigg)$$

21. Let X be a random variable that follows the binomial distribution with parameters n and p, and denote by X/n the proportion of successes in the n trials. Show that for any positive real number ϵ such that $\epsilon < p$, the probability that Y belongs to the interval $[p - \epsilon, p + \epsilon]$ is at least

$$1 - \frac{p(1-p)}{n\epsilon^2}.$$

(*Hint*: Use Chebychev's inequality.)

5.2 GEOMETRIC AND NEGATIVE BINOMIAL DISTRIBUTIONS

In an experiment modeled by the binomial distribution, we fix the number of trials, n, and consider probability statements about the number X of "successes," which is random. Interest in many occasions may lie in the opposite direction; that is, it may be natural to ask *how long* one has to wait until the first success occurs in a Bernoulli sequence. For example, if we throw a die repeatedly, how long do we have to wait until the first appearance of a six? Or, if a shooter shoots against a target with a constant success probability p, what is the distribution of the number of trials until he hits the target for the first time?

Definition 5.3 Let X denote the number of trials until the first success occurs in a sequence of independent Bernoulli trials, where the probability of success in each trial equals p. We then say that the distribution of X is a **geometric distribution** with parameter p, which is denoted by $G(p)$.

If X has a $G(p)$ distribution, we write $X \sim G(p)$. It is typically very easy to answer questions regarding probability statements about X. In order to find the probability function of X, we observe that the event $\{X = x\}$ means that the first $x - 1$ trials resulted in a failure while the nth trial is a success:

$$\underbrace{F \quad F \quad \cdots \quad F}_{x-1 \text{ times}} \quad S$$

Since trials are independent, the event $\{X = x\}$ has probability $(1 - p)^{x-1}p$, or writing q for $1 - p$:

$$\underbrace{q \quad q \quad \cdots \quad q}_{x-1 \text{ times}} \quad p = q^{x-1}p.$$

We have therefore shown the following.

Proposition 5.3 *If a random variable X follows the G(p) distribution, then the probability function of X is given by*

$$f(x) = P(X = x) = (1-p)^{x-1}p, \quad \textit{for} \quad x = 1, 2, \ldots \tag{5.2}$$

Notice that, in contrast to the binomial (or the Bernoulli) distribution, if a random variable X follows a geometric distribution, the range of values for X is infinite. Specifically, the set of possible values for X is the set of positive integers $\{1, 2, 3, \ldots\}$.

The fact that the probabilities of the geometric distribution add up to one is an immediate consequence of the geometric series formula (thus giving the name to the distribution), since we have

$$\sum_{x=1}^{\infty} (1-p)^{x-1} p = p \sum_{r=0}^{\infty} (1-p)^r = \frac{p}{1-(1-p)} = 1.$$

It should be noted that there is another definition of the geometric distribution that differs from the one given above, but also has a very simple physical interpretation. To be specific, let us consider the *number of failures* rather than the number of trials until the first success appears in a Bernoulli sequence. Set X_0 for the random variable that counts this number of failures (prior to the success); it is then clear that

$$X = X_0 + 1,$$

and the range of values for X_0 is now the set $\{0, 1, 2, \dots\}$, so that in particular the additional value zero is now included. The probability function of X_0 is

$$P(X_0 = x) = (1-p)^x p = f(x+1), \quad x = 0, 1, 2, \dots$$

Example 5.7 If we throw a die until six appears for the first time, what is the probability that this will happen

(i) in the 6th throw?

(ii) no sooner than the 10th throw?

SOLUTION Let X be the number of throws needed until the first appearance of a six. Then, the distribution of X is geometric with parameter $p = 1/6$, i.e.

$$f(x) = P(X = x) = (1-p)^{x-1} p = \left(\frac{5}{6}\right)^{x-1} \cdot \left(\frac{1}{6}\right) = \frac{5^{x-1}}{6^x}, \quad x = 1, 2, \dots$$

(i) We want to find $P(X = 6)$, which is

$$P(X = 6) = f(6) = \frac{5^5}{6^6} = 0.067,$$

i.e. less than 7%. This seems to contradict our intuition, as with $p = 1/6$, we might think that $X = 6$ is the most likely value for X (in fact, the highest probability associated with the geometric distribution is when $X = 1$, as one checks easily, no matter what the value of p is).

(ii) The probability we seek is now $P(X \geq 10)$ and, using the formula for geometric probabilities, we deduce that

$$P(X \geq 10) = \sum_{k=10}^{\infty} (1-p)^{k-1} p = p(1-p)^9 \sum_{x=0}^{\infty} (1-p)^x = (1-p)^9, \qquad (5.3)$$

where we used again the well-known formula for a geometric series,

$$\sum_{x=0}^{\infty} t^x = 1 + t + t^2 + t^3 + \cdots = \frac{1}{1-t}, \quad |t| < 1. \tag{5.4}$$

Upon making the substitution $p = 1/6$ in (5.3), we get

$$P(X \geq 10) = \left(\frac{5}{6}\right)^9 = 0.1938.$$

In (5.3), we have shown that $P(X \geq 10) = q^9$, but this is in fact very easy to see without any calculations: the event $\{X \geq 10\}$ means that at least 10 trials are needed for the first success, and this happens *if and only if* the first 9 trials result in a failure. By the same token, we get for any positive integer k, $P(X \geq k) = q^{k-1}$. Since k takes only integer values, we may recast this as

$$P(X > k) = q^k. \tag{5.5}$$

This implies immediately that $P(X \leq k) = 1 - q^k$ for any $k = 1, 2, \ldots$ and, since the distribution function changes value only at positive integers, we see that for a geometrically distributed random variable X, the associated distribution function is given by

$$F(t) = P(X \leq t) = 1 - q^{[t]}, \quad t \geq 1,$$

while $F(t) = 0$ for $t < 1$ (see also Figure 5.4).

Although being obvious probabilistically, the identity (5.5) has an important consequence, which is a key property of the geometric distribution. In Example 5.7, we found that the probability we have to wait until at least the 10th trial for a success (the appearance of a six) is 0.067. Suppose now we have thrown the die five times and no six has turned up. What is the probability that at least 10 *more* trials are needed until a six occurs? It may seem (and, to many people, it is) strange, but the answer is the same, i.e. 0.067. Although one is tempted to think that if six does not appear in more and more trials, the probability of getting a six in the next trial gets higher, mathematically this cannot be true. Since we assume that trials are independent, a sequence of failures (no matter how long) does not affect the probability of getting a six in the next trial. To formalize this argument, and establish the aforementioned property of the geometric distribution, known as the *memoryless property* (or *lack of memory property*), we argue that for any positive integers n and k, we have

$$P(X > n + k | X > n) = P(X > k).$$

To check its validity, notice first that the conditional probability in the left-hand side above is

$$P(X > n + k | X > n) = \frac{P(X > n + k, X > n)}{P(X > n)}.$$

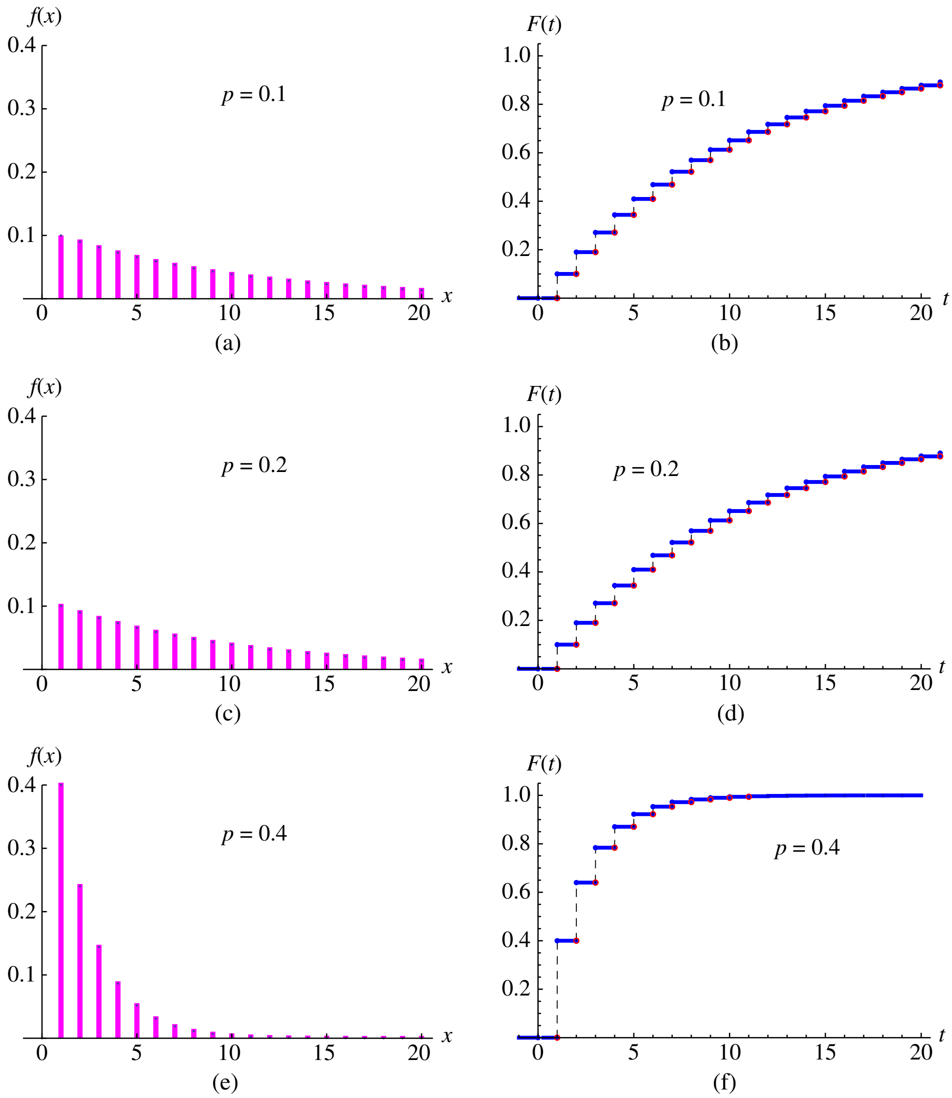


Figure 5.4 The probability function (left) and the cumulative distribution function (right) of the geometric distribution for various choices of p.

But the intersection of the events $\{X > n + k\}$ and $\{X > n\}$ is simply $\{X > n + k\}$, and by an appeal to (5.5) we get that

$$P(X > n + k | X > n) = \frac{P(X > n + k)}{P(X > n)} = \frac{q^{n+k}}{q^n} = q^k = P(X > k).$$

We now turn our attention to the mean and variance for a geometric random variable.

Proposition 5.4 *If the random variable X has the geometric distribution with parameter p, then*

$$\mu = E(X) = \frac{1}{p}, \quad \sigma^2 = \mathrm{Var}(X) = \frac{q}{p^2}.$$

Proof: The mean of $X \sim G(p)$ is given by

$$\mu = E(X) = \sum_{x \in R_X} xf(x) = \sum_{x=1}^{\infty} xpq^{x-1}. \tag{5.6}$$

Note that p can be taken outside the last summation and that the function xq^{x-1} is the derivative (with respect to q) of the function q^x. This leads us to use the geometric series in (5.4). Differentiating both sides of this expression, we derive that

$$\sum_{x=1}^{\infty} xt^{x-1} = 1 + 2t + 3t^2 + 4t^3 + \cdots + nt^{n-1} + \cdots = \left(\frac{1}{1-t}\right)' = \left(\frac{1}{1-t}\right)^2, \tag{5.7}$$

for $|t| < 1$. Replacing t by q, and inserting the result into (5.6), we obtain

$$\mu = E(X) = p\frac{1}{(1-q)^2} = \frac{p}{p^2} = \frac{1}{p},$$

as required.

For the variance, we find first the factorial moment of second order,

$$\mu_{(2)} = E[(X)_2] = \sum_{x \in R_X} x(x-1)f(x) = \sum_{x=1}^{\infty} x(x-1)pq^{x-1}. \tag{5.8}$$

Now we notice that the function $x(x-1)q^{x-1}$, which appears on the right, is the *second derivative* of the function q^x, multiplied by the term q. This suggests differentiating (5.4) once more. Since the first derivative of the terms appearing there was found in (5.7), we differentiate both sides of the latter equation to get

$$\begin{aligned}\sum_{x=2}^{\infty} x(x-1)t^{x-2} &= 2 + 6t + 12t^2 + \cdots + n(n-1)t^{n-2} + \cdots = \left(\frac{1}{1-t}\right)'' \\ &= \left(\frac{1}{(1-t)^2}\right)' = \left(\frac{2}{1-t}\right)^3.\end{aligned}$$

Replacing again t by q and inserting into (5.8), we obtain

$$\mu_{(2)} = pq\frac{2}{(1-q)^3} = \frac{2pq}{p^3} = \frac{2q}{p^2}.$$

It is now easy to obtain the variance of X by using the expression (see the proof of Proposition 5.2)

$$\text{Var}(X) = \mu_{(2)} + \mu - \mu^2.$$

This yields

$$\text{Var}(X) = \frac{2q}{p^2} + \frac{1}{p} - \left(\frac{1}{p}\right)^2 = \frac{2(1-p)+p-1}{p^2} = \frac{q}{p^2},$$

which is the desired result. □

According to the result of Proposition 5.4,

- if we throw a die successively, the expected number of trials until the first six is

$$\frac{1}{p} = \frac{1}{1/6} = 6$$

(the same is true for any other outcome of the die, e.g. if we wait until a five appears). Similarly, the expected number of trials until the outcome is greater than four is

$$\frac{1}{p} = \frac{1}{2/6} = 3.$$

- when we toss a coin repeatedly, the expected number of tosses until "heads" appears for the first time is

$$\frac{1}{p} = \frac{1}{1/2} = 2.$$

Example 5.8 Let X have a geometric distribution with parameter p.

(i) Find the probability that X takes on an even value.

(ii) Show that the probability derived in Part (i) is less than or equal to $1/2$.

(iii) Is there a value of p that maximizes the probability found in Part (i)?

SOLUTION

(i) The probability that X is even is the sum over k of the probabilities that X takes the value $2k$, for $k = 1, 2, 3, \ldots$ This is

$$P(X \text{ is even}) = \sum_{k=1}^{\infty} P(X = 2k) = \sum_{k=1}^{\infty} (1-p)^{2k-1} p = p(1-p) \sum_{k=1}^{\infty} [(1-p)^2]^{k-1}.$$

Employing (5.4), we then get

$$P(X \text{ is even}) = \frac{p(1-p)}{1-(1-p)^2}.$$

(ii) Using the result in (i), we want to demonstrate that for all values of p, the following inequality holds:

$$\frac{p(1-p)}{1-(1-p)^2} \leq \frac{1}{2}.$$

This leads to the quadratic inequality $2p(1-p) \leq 1-(1-p)^2$, and this in turn reduces to $p^2 \geq 0$, which is obviously true for any value of p. So, for all values of $p \in (0,1)$, the probability that the number X of trials is even is less than a half. This seems to be an obvious result if one notices that for the geometric distribution we have $P(X = n) > P(X = n - 1)$ for any $n = 2, 3, \ldots$, and so adding up the probabilities $P(X = n)$ for odd values of n produces a result which is always (strictly) greater than the sum of the probabilities for even values of n.

(iii) In view of the result in (i), we need to maximize the function

$$A(p) = \frac{p(1-p)}{1-(1-p)^2} = \frac{p(1-p)}{p(2-p)} = \frac{1-p}{2-p} = 1 - \frac{1}{2-p},$$

which is a strictly decreasing function of p in $(0, 1)$. If we were to allow the value $p = 0$, the function $A(p)$ would attain a maximum at zero. However, it is clear that a zero value cannot be allowed as a parameter for the geometric distribution (if $p = 0$, "success" is impossible, so we would have to wait indefinitely for the first success to occur).

In view of the above, a maximum is not attained for the function $A(p)$. Notice however that, by virtue of (ii), we may write that

$$A(p) \le \frac{1}{2} = \lim_{p \to 0} A(p).$$

Example 5.9 (The St. Petersburg paradox) Paul plays the following game: he tosses a fair coin until it lands on heads for the first time. If this happens at the first throw, he receives $2^1 = \$2$, if it happens in the second throw he receives $2^2 = \$4$, and so on so that, if it happens on the rth throw, he receives 2^r dollars.

(i) Find Paul's expected gain from this game.

(ii) Consider now a variation of this game: Paul receives nothing if heads have not turned up after N tosses. Find Paul's expected profit for this new game.

(iii) What price should Paul pay to enter each of the two games in (i) and (ii) above, assuming that the game is fair on both occasions?

SOLUTION

(i) Let X be the number of trials needed until a toss results in heads for the first time, and let V be Paul's gain from this game. Then, it is clear that X has a geometric distribution with $p = 1/2$, so that

$$P(X = r) = \left(\frac{1}{2}\right)^r = 2^{-r}, \quad \text{for} \quad r = 1, 2, 3, \ldots$$

It is obvious that the amount of money Paul receives from the game depends on the number of trials needed, i.e. on the value of X. More explicitly, we have, under the assumptions of the example, that V takes the value 2^r if and only if X takes the value r, that is,

$$P(V = 2^r) = P(X = r) = 2^{-r}.$$

As a result, we see that the expected value of the variable V equals

$$E(V) = \sum_{v \in R_V} vP(V = v) = \sum_{r=1}^{\infty} 2^r \cdot 2^{-r} = 1 + 1 + 1 + \cdots$$

Since we have a series, that is an infinite sum, with each term equal to one, it is clear that the series diverges; i.e. $E(V) = \infty$, and so Paul's expected profit is infinite!

(ii) Denote by Y the random variable that takes the value k if the first k trials resulted in tails (Y takes the value zero if heads occur with the first trial). Notice that Y cannot be greater than N since, if the first N trials are all tails, the game is terminated and Paul receives no money. For $r = 0, 1, \ldots, N - 1$, we have

$$P(Y = r) = \left(\frac{1}{2}\right)^r \cdot \frac{1}{2} = 2^{-r-1},$$

while $P(Y = N)$ is the probability that no heads occur in the first N trials, and this equals (see (5.5))

$$P(Y = N) = P(X > N) = \left(\frac{1}{2}\right)^N.$$

As a result, arguing as in (i) above, we obtain Paul's expected gain (we denote this again by V) in this case to be

$$E(V) = \sum_{v \in R_V} vP(V = v) = \sum_{r=0}^{N-1} 2^r \cdot 2^{-r} + 0 \cdot P(Y = N) = N,$$

so that the expected gain is now finite.

(iii) For the game in Part (i), since the expected profit is infinite, mathematical formalism suggests that a person may enter *at any price*. So, Paul can pay any amount of money to enter this and yet expect to make a profit.

However, it is natural to believe that most people would not be willing to pay a very large amount to enter this game. In practical terms, the fallacy here lies on the fact that people do not have unlimited resources.

As a matter of fact, the problem in Part (i) is known as the *St. Petersburg paradox*; it owes its name to the statement and solution of the problem that Daniel Bernoulli, another member of the Bernoulli family, gave in 1738 and appeared in the *Commentaries of the Imperial Academy of Science of Saint Petersburg*.[1]

For the modified game in Part (ii), which has a finite expectation, N, it is clear that if Peter pays less than N dollars, the game is favorable to him. If he pays N dollars, the game is fair. However, if N is very large, so that paying N dollars (or even, say, $N/2$ dollars) is beyond Peter's available resources, we may come close to the situation described for the game in Part (i).

Until now, in this section we have been concerned with the situation wherein one waits until the first success occurs in a sequence of Bernoulli trials. Naturally, and as will be illustrated with various examples below, there are many situations where the interest lies in the occurrence of the second, third or, more generally, the rth success in such a sequence. We define next the distribution that emerges in this more general set-up.

Definition 5.4 Let X denote the number of trials until the rth success in a sequence of independent Bernoulli trials, where the probability of success in each trial is equal to p. Then, the distribution of X is called a **negative binomial distribution** with parameters r and p, and is denoted by $\text{Nb}(r, p)$.

Before proceeding to the study of the main properties of the negative binomial distribution, let us present an example from medical statistics in which the distribution arises in a natural way.

Example 5.10 The negative binomial distribution has direct applications to sampling theory in statistics. As an illustration assume that, for a certain clinical trial, a sample size of $k = 20$ patients who have developed a certain disease, is needed. Doctors have available a large (for our purposes, we may regard it as infinite) population of patients who have had symptoms of that disease in the last 12 months. Suppose that each patient in that population has a probability p of developing the disease, something that may be detected through a special medical examination. Under this scenario, what is the distribution of the number, X, of patients that have to undergo the (potentially, very expensive) medical examination, until 20 of them have tested positive for the disease?

SOLUTION Considering the medical examination of each patient as a trial with two possible outcomes (S, if a patient tests positive for the disease, and F if the patient tests negative) and making the plausible assumption that successive trials are independent, it is then clear from the statement of the example that X has a negative binomial distribution with parameters 20 and p.

[1]It is thought that the problem was invented originally in 1713 by yet another mathematician with the same surname, Nicholas Bernoulli, who was Daniel's cousin. It should be noted that Daniel Bernoulli's solution to the problem brought a new light to many ideas in economic theory.

If $X \sim \mathrm{Nb}(r, p)$, then the set of possible values for X is $R_X = \{r, r+1, \dots\}$. In order to find the probability function of X, it suffices to observe that the event $\{X = x\}$ is equivalent to the intersection of the independent events

A: trial x results in a success,

B: in the first $x - 1$ trials there have been exactly $r - 1$ successes.

Therefore,

$$P(X = x) = P(AB) = P(A)P(B)$$

and since for the event A we have obviously $P(A) = p$, while the probability of the event B is

$$P(B) = \binom{x-1}{r-1} p^{r-1} q^{(x-1)-(r-1)} = \binom{x-1}{r-1} p^{r-1} q^{x-r},$$

the following result ensues.

Proposition 5.5 *The probability function of the negative binomial distribution* $\mathrm{Nb}(r, p)$ *is given by*

$$f(x) = P(X = x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, \dots \tag{5.9}$$

Comparing either Definitions 5.3 and 5.4 or Equation (5.2) with the result of Proposition 5.5, it should be obvious that the geometric distribution is a special case of the negative binomial distribution when $r = 1$. The term "negative binomial distribution" is due to the fact that the probability function $f(x)$ appears in the *binomial expansion* of the *negative* power $(1 - q)^{-r}$. More precisely, we have

$$(1 - q)^{-r} = \sum_{k=0}^{\infty} \binom{r+k-1}{k} q^k = \sum_{k=0}^{\infty} \binom{r+k-1}{r-1} q^k.$$

Changing the variable of summation in the last sum to $x = r + k$, we obtain

$$(1 - q)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} q^{x-r}, \tag{5.10}$$

and, multiplying both sides by p^r, we deduce that

$$p^r (1 - q)^{-r} = \sum_{x=r}^{\infty} \binom{x-1}{r-1} p^r q^{x-r} = \sum_{x=r}^{\infty} f(x).$$

The last expression verifies that f is a probability function of a random variable that takes values in the set $R_X = \{r, r+1, \dots\}$, because we have $f(x) \geq 0$ for all $x \in R_X$ and, in addition,

$$\sum_{x=r}^{\infty} f(x) = p^r (1 - q)^{-r} = p^r p^{-r} = 1.$$

The cumulative distribution function of the negative binomial distribution with parameters r and p is given by

$$F(t) = \begin{cases} 0, & \text{if } t < r \\ \displaystyle\sum_{x=r}^{[t]} f(x) = \sum_{x=r}^{[t]} \binom{x-1}{r-1} p^r q^{x-r}, & \text{if } t \geq r. \end{cases}$$

Plots of the probability function $f(x)$ and the cumulative distribution function $F(t)$ are presented in Figure 5.5 for some choices of r and p.

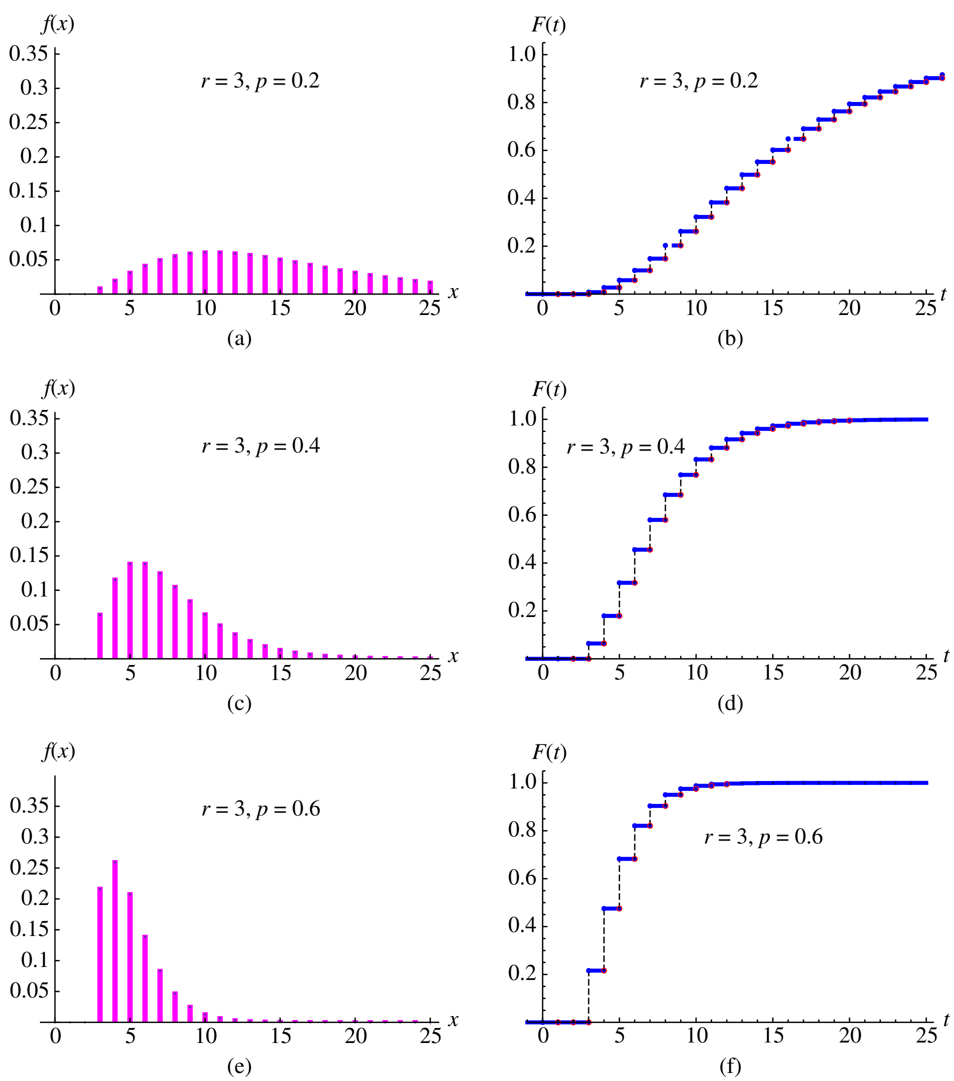


Figure 5.5 The probability function (left) and the cumulative distribution function (right) of the negative binomial distribution for some choices of r and p.

Example 5.11 In Example 5.10, if the probability p of developing the disease is $p = 1/5$, find the probability that the number of patients that needs to be tested until 20 of them have tested positive for the disease is larger than 50 but no larger than 80.

SOLUTION As we mentioned in Example 5.10, the distribution of the number of patients, X, that need to be tested is negative binomial $\text{Nb}(20, p)$, and so for $p = 1/5$ it is $\text{Nb}(20, 1/5)$. The probability function of X thus is

$$f_X(x) = P(X = x) = \binom{x-1}{20-1}\left(\frac{1}{5}\right)^{20}\left(\frac{4}{5}\right)^{x-20}.$$

We want the probability that X takes values larger than 50 but no larger than 80, which is

$$P(50 < X \leq 80) = \sum_{x=51}^{80} f(x) = \sum_{x=51}^{80}\binom{x-1}{19}\left(\frac{1}{5}\right)^{20}\left(\frac{4}{5}\right)^{x-20}.$$

It is obvious that this is a very long and tedious calculation if one uses a pocket calculator; it is much less so if one resorts to the aid of computers, e.g. using a spreadsheet. Moreover, we mention that there are a number of more sophisticated (algebraic and statistical) software packages, using which finding the above result becomes a rather straightforward task. Using such a program, we find that the required probability is 0.16248 (students are encouraged to verify this!).

We note that in Definition 5.4, the negative binomial distribution has been defined as the "waiting time" of an event in a sequence of Bernoulli trials. In this context, it is clear that the parameter r of that distribution must be an integer; we should point out, however, that if we extend the definition in (5.9) for any (fixed) *real* number $r > 0$, we also get a probability distribution since the values $f(x)$ for this more general case are still nonnegative and add up to one. In some advanced textbooks, this more general notion with positive real r is defined as the negative binomial distribution, while the case we have considered here is called a **Pascal distribution**.

The next proposition provides formulas for the mean and the variance of a random variable having a negative binomial distribution.

Proposition 5.6 *When X has a* $\text{Nb}(r, p)$ *distribution, its mean and variance are given by*

$$\mu = E(X) = \frac{r}{p}, \quad \sigma^2 = \text{Var}(X) = \frac{rq}{p^2}.$$

Proof: The expected value of X is given by the expression

$$\mu = E(X) = \sum_{x \in R_X} xf(x) = \sum_{x=r}^{\infty} x\binom{x-1}{r-1}p^r q^{x-r}.$$

Upon using the combinatorial identity

$$x\binom{x-1}{r-1} = x\frac{(x-1)!}{(r-1)!(x-r)!} = r\binom{x}{x-r} = r\binom{x}{r},$$

we get

$$\mu = rp^r \sum_{x=r}^{\infty} \binom{x}{r} q^{x-r}.$$

Setting $y = x - r$, we obtain

$$\mu = rp^r \sum_{y=0}^{\infty} \binom{y+r}{r} q^y = rp^r \sum_{y=0}^{\infty} \binom{y+(r+1)-1}{(r+1)-1} q^y$$

and substituting the last sum of this expression with the aid of (5.10), we get

$$\mu = rp^r(1-q)^{-(r+1)} = \frac{rp^r}{p^{r+1}} = \frac{r}{p},$$

as required. For the variance Var(X), it is useful to compute first an expression for the quantity

$$E[X(X+1)] = \sum_{x \in R_X} x(x+1)f(x) = \sum_{x=r}^{\infty} x(x+1) \binom{x-1}{r-1} p^r q^{x-r}.$$

In order to calculate this sum, we observe that

$$\begin{aligned} x(x+1)\binom{x-1}{r-1} &= x(x+1)\frac{(x-1)!}{(r-1)!(x-r)!} \\ &= r(r+1)\frac{(x+1)!}{(r+1)!(x-r)!} = r(r+1)\binom{x+1}{x-r} \end{aligned}$$

and, by an appeal to (5.10) again, we find successively

$$\begin{aligned} E[X(X+1)] &= r(r+1)p^r \sum_{x=r}^{\infty} \binom{x+1}{x-r} q^{x-r} \\ &= r(r+1)p^r \sum_{y=r+2}^{\infty} \binom{y-1}{(r+2)-1} q^{y-(r+2)} \\ &= r(r+1)p^r(1-q)^{-(r+2)} \\ &= \frac{r(r+1)p^r}{p^{r+2}} = \frac{r(r+1)}{p^2}. \end{aligned}$$

It is now easy to find the variance of X to be

$$\mathrm{Var}(X) = E[X(X+1)] - E(X) - [E(X)]^2 = \frac{r(r+1)}{p^2} - \frac{r}{p} - \frac{r^2}{p^2} = \frac{rq}{p^2},$$

which is the desired result. □

Example 5.12 A millionaire donates every Christmas a large amount of money to 3 charity organizations, which she selects randomly from a list of 10 charities. Donations in successive years are independent, so that the same organization may be selected again and again. Suppose that one of the charities is UNICEF, and their directors have decided that all donations received from the millionaire in the future will be kept in a special account with a view to offer free education for children in a developing country. They estimate they will have sufficient funds to accomplish that when they receive the donation at least twice.

(i) Using the negative binomial distribution, find the probability that UNICEF will have to wait at least six years until they can implement their project.

(ii) Derive the result using the binomial distribution instead.

(iii) Find the expected value and the variance for the number of years until UNICEF receives the donation for the second time.

SOLUTION

(i) Let X be the number of years until UNICEF receive the donation for the second time. Since the probability that they are selected by the millionaire at any particular year is $3/10 = 0.3$, it is clear that $X \sim \text{Nb}(2, 0.3)$. Moreover, X takes values on the set $\{2, 3, \ldots\}$, with respective probabilities

$$f(x) = \binom{x-1}{2-1}\left(\frac{3}{10}\right)^2\left(\frac{7}{10}\right)^{x-2} = (x-1)\frac{3^2 \cdot 7^{x-2}}{10^x}, \quad x = 2, 3, \ldots$$

The probability we are seeking is then

$$\begin{aligned}
P(X \geq 6) =& 1 - P(X < 6) = 1 - \sum_{x=2}^{5} f(x) \\
=& 1 - \left[(2-1)\frac{3^2 \cdot 7^{2-2}}{10^2} + (3-1)\frac{3^2 \cdot 7^{3-2}}{10^3} + (4-1)\frac{3^2 \cdot 7^{4-2}}{10^4}\right. \\
&\left. + (5-1)\frac{3^2 \cdot 7^{5-2}}{10^5}\right] \\
=& 1 - (0.3)^2[1 + 1.4 + 1.47 + 1.372] = 0.528\,22.
\end{aligned}$$

(ii) Observe that the event in (i) occurs, i.e. UNICEF has to wait at least six years for the second donation, if and only of they receive *at most* one donation in the first five years. So, let us now use Y to denote the number of donations received in the first five years. Then, the required probability is

$$P(Y \leq 1) = P(Y = 0) + P(Y = 1)$$

and since the distribution of Y is binomial with parameters $n = 5$ and $p = 0.3$, we deduce

$$\begin{aligned}P(Y \leq 1) &= \binom{5}{0}(0.3)^0(0.7)^5 + \binom{5}{1}(0.3)^1(0.7)^{5-1}\\ &= (0.7)^5 + 5 \cdot (0.3)(0.7)^4 = 0.528\,22,\end{aligned}$$

which is obviously the same as the result in (i).

(iii) Let X be as in Part (i). We want $E(X)$ and $\mathrm{Var}(X)$ and, for a negative binomial random variable, these are given by Proposition 5.6. In particular, we have

$$E(X) = \frac{r}{p} = \frac{2}{0.3} = 6.67 \text{ years},$$

and

$$\mathrm{Var}(X) = \frac{rq}{p^2} = \frac{2 \cdot 0.7}{(0.3)^2} = 15.56 \text{ years}^2.$$

The first two parts of the preceding example illustrate the fact that there are instances where probabilities associated with an experiment involving sequences of Bernoulli trials can be calculated using *either* the binomial or the negative binomial distribution. More explicitly, we have the following result. Let X have a negative binomial distribution with parameters r and p, and Y have a binomial distribution with parameters n and p (assuming that $n > r$). Then, we have

$$P(X > n) = P(Y < r). \tag{5.11}$$

No calculations are needed to verify this. Simply note that both events inside the brackets on the two sides of (5.11) correspond to the fact that the first n trials in a Bernoulli sequence result in *at most* $r - 1$ successes. It goes without saying that, when r is small as in Example 5.12, we may prefer to use the binomial distribution, i.e. the RHS of (5.11) instead of the LHS there, which calls for the evaluation of an infinite series involving negative binomial probabilities.

Note that (5.11) can obviously be restated in terms of the probabilities of the complementary events, i.e.

$$P(X \leq n) = P(Y \geq r).$$

To illustrate the relationship between binomial and negative binomial distributions further, we now give another example, which is historically associated with the early development of probability theory. The first part of the following example is, essentially, one of the problems posed to the seventeenth century French mathematician Blaise Pascal by the gambler Chevalier de Méré. Pascal used a recursion argument to solve the problem but, more importantly, initiated a correspondence[2] with another French mathematician of that time, Pierre de Fermat (1607–1665).

The problem discussed in the following example is known as the *problem of points*.

[2]This famous correspondence, which soon became well-known and attracted interest outside France, laid down the foundations for the calculus of probability. It is also worth noting that, unlike Pascal, Fermat (who was trained as a lawyer and was only an amateur mathematician) used combinatorial methods to tackle most of the problems presented to him by Pascal.

Example 5.13

(i) In a series of Bernoulli trials, with success probability p, calculate the probability that the kth success occurs prior to the nth failure.

(ii) A game in table tennis is won by the first player who reaches 11 points. Jack and Jill play each other in their regular Wednesday meetings, and at some point the score is 8–4 in favor of Jack. Assuming that, for each point, Jack has a 60% chance to win it, what is the probability that Jill will first accumulate 11 points?

SOLUTION

(i) Let $\mathcal{E}$ be the event of interest, namely,

$$\mathcal{E}\ :\ \text{the } k\text{th success occurs prior to the } n\text{th failure.}$$

Then, for $\mathcal{E}$ to occur, we must have the kth success to occur on or before the $(n+k-1)$th trial; this is so because, by the end of the first $n+k-1$ trials, **exactly one** of the events "k successes" and "n failures" has occurred. Thus, if X is the number of trials until the kth success occurs so that $X \sim \text{Nb}(k, p)$, we seek the probability

$$P(X \leq n+k-1) = \sum_{x=k}^{n+k-1} \binom{x-1}{k-1} p^k (1-p)^{x-k}.$$

As explained above, the same result can be obtained by using the binomial distribution instead of the negative binomial distribution. A key observation here is that $\mathcal{E}$ occurs if and only if there are at least k successes in the first $n+k-1$ trials. In fact, the required probability can be expressed as $P(Y \geq k)$, where Y has a binomial distribution with parameters $n+k-1$ and p. From the formula for binomial probabilities, the above expression equals

$$P(Y \geq k) = \sum_{x=k}^{n+k-1} \binom{n+k-1}{x} p^x (1-p)^{n+k-1-x}.$$

(ii) This is an application of the result found in (i). The required probability is in fact the probability that with the score at 8–4, in the following series of points played (regarded as Bernoulli trials), Jill will win seven points before Jack wins three points. The success probability in each trial is 0.4, so from the result in Part (i) (with $k = 7$ and $n = 3$), we deduce the desired probability to be (using the binomial distribution first)

$$\begin{aligned}\sum_{x=7}^{9} \binom{9}{x} (0.4)^x (1-0.4)^{9-x} &= \binom{9}{7} (0.4)^7 (1-0.4)^2 + \binom{9}{8} (0.4)^8 (1-0.4)^1 \\ &\quad + \binom{9}{9} (0.4)^9 (1-0.4)^0 \\ &= 0.021\,233\,7 + 0.003\,538\,9 + 0.000\,262\,1 \\ &= 0.025\,034\,7,\end{aligned}$$

or about 2.5%. Upon using the negative binomial distribution, we get the required probability to be

$$
\begin{aligned}
P(X \leq n+k-1) &= \sum_{x=k}^{n+k-1} \binom{x-1}{k-1} p^k q^{x-k} = \sum_{x=7}^{9} \binom{x-1}{6} (0.4)^6 (0.6)^{x-6} \\
&= (0.4)^6 \left[\binom{6}{6} (0.6)^1 + \binom{7}{6} (0.6)^2 + \binom{8}{6} (0.6)^3 \right] \\
&= 0.025\,034\,7,
\end{aligned}
$$

and it is no surprise that this agrees with the result found above.

The next example is a notorious problem that is historically associated with the name of a famous Polish mathematician, Stefan Banach (1892–1945).[3]

Example 5.14 (Banach matchbox problem) A mathematician always carries two matchboxes, one in his right pocket and one in his left. When he wants a match, he selects a pocket at random so that successive selections may be regarded as Bernoulli trials with $p = 1/2$. Assume that each matchbox contains initially N matches. If he carries on doing this, there will be a moment when he realizes that one of the boxes is empty. The number Y of matches in his *other pocket* is obviously a random variable. What is the distribution of Y?

SOLUTION We note first that the range of values for Y is the set $\{0, 1, \ldots, N\}$. Observe in particular that Y can take the value 0 if the pocket which is emptied first is not the one that the mathematician *discovers first* that it is empty. Suppose that we call a success the choice of the matchbox which is in his right pocket and let X be the number of trials until the $(N+1)$th success occurs (this is the point where the mathematician looks for a match in his right pocket and finds out there isn't one).

Since we are given that successive selections of a pocket constitute Bernoulli trials with success probability $p = 1/2$, it is clear that X has a $\mathrm{Nb}(N+1, 1/2)$ distribution with probability function

$$
\begin{aligned}
f(x) = P(X = x) &= \binom{x-1}{(N+1)-1} \left(\frac{1}{2}\right)^{N+1} \left(\frac{1}{2}\right)^{x-(N+1)} \\
&= \binom{x-1}{N} \left(\frac{1}{2}\right)^x, \quad x = N+1, N+2, \ldots
\end{aligned}
$$

Now, consider the situation wherein there are y matches in the left pocket at the time the mathematician finds the right pocket to be empty. This means that he has selected his right pocket $N+1$ times and his left pocket $N-y$ times, in a total of $2N-y+1$

[3]The problem is attributed to Hugo Steinhaus (1887–1972) in an address to honor Banach, who was a heavy smoker.

trials. The probability of this event is

$$P(X = 2N - y + 1) = f(2N - y + 1) = \binom{2N - y}{N} \left(\frac{1}{2}\right)^{2N-y+1}.$$

So far, we have assumed that the first pocket found to be empty is the right one; by symmetry, and since there are the same number of matches in each pocket initially, we obtain that this is also the distribution of the number of matches in his right pocket if the left one is first found empty. Consequently, the probability function of the random variable Y is given by the formula

$$P(Y = y) = 2f(2N - y + 1) = \binom{2N - y}{N} \left(\frac{1}{2}\right)^{2N-y}, \quad y = 0, 1, 2, \ldots, N.$$

In closing this section, let us introduce a second version of the Nb distribution. Consider a sequence of independent Bernoulli trials with a success probability p and define the random variables

X: number of **trials** until r successes occur, and

Y: number of **failures** until r successes occur.

This parallels, and generalizes, the similar distinction we made for geometrically distributed random variables. The variables X and Y above are related via the relationship $Y = X - r$, and so their respective probability functions satisfy the relationship

$$f_Y(y) = P(Y = y) = P(X - r = y) = P(X = y + r) = f_X(y + r).$$

Since X has a negative binomial distribution, its probability function, as given by Proposition 5.5, is

$$f_Y(y) = \binom{y + r - 1}{r - 1} p^r q^y, \quad y = 0, 1, 2, \ldots$$

In some textbooks, the distribution of the random variable X is referred to as a Pascal distribution, to honor Blaise Pascal (1605–1662) who first studied it, while the term negative binomial distribution is used for the distribution of the random variable Y; see also the discussion before Proposition 5.6.

EXERCISES

Group A

1. If we throw a fair die successively, find the probability that a four appears for the third time

   (i) at the 10th throw;

   (ii) after the first 20 throws of the die.

2. The probability that a digital signal is transmitted incorrectly is p. Suppose a particular signal is transmitted repeatedly and that successive transmissions occur every two minutes. What is the distribution of the time T until it is transmitted correctly

   (i) for the first time?

   (ii) for the fourth time?

   For each case above, find the expectation and variance of T.

3. Danai and Bill each have a die and throw it simultaneously. If both dice land on an even number, the game ends. If not, they continue with another round of each throwing their die.

   (i) What is the probability that the game lasts

      (a) five rounds?

      (b) at least eight rounds?

      (c) between four and six rounds?

   (ii) What is the expected number of rounds to be played?

4. The probability that someone tests positive in a medical examination for a disease is 6%. What is the probability that

   (i) the 8th person who is examined is the first who tests positive?

   (ii) the 12th person who is examined is the third who tests positive?

   (iii) more than 20 persons are needed until 5 people have tested positive?

5. Peter plays the following game: he rolls a die successively and he receives k^2 dollars if a five appears for the first time in the kth row. Find Peter's expected turnover from this game.

6. In a match of women's tennis, the winner is the player who first wins two sets. Suppose two players, Martina and Steffi, play against each other and that the probability that Martina wins any particular set is p for some $0 < p < 1$.

   (i) Find, as a function of p, the probability that Martina wins the match.

   (ii) Let Y be the number of sets to be played. Find $E(Y)$.

   (iii) Show that the answer in (ii) is maximized when $p = 1/2$.

7. A computer selects decimal digits of a random number in the interval $(0, 1)$ with equal probability among the digits $0, 1, 2, \dots, 9$. In a number generated in this way, what is the expected number of decimal digits

   (i) before the digit 5 appears for the first time?

   (ii) before one of the digits $2, 3, 4$ appears for the first time?

   (iii) before the sixth appearance of the digit 3?

   (iv) before the fourth appearance of an odd digit?

8. Rodney, who sells newspapers on the street, buys each newspaper for 35 cents and he sells it for 50 cents. However, he cannot return any unsold newspapers. He has estimated that the number of newspapers he sells daily has a negative binomial distribution with parameters $r = 25$ and $p = 2/5$. Find how many newspapers he should buy so that he maximizes his expected profit.

9. Let X be a random variable that has the geometric distribution with parameter p $(0 < p < 1)$, and let f be the probability function of X.

   (i) Verify that f satisfies the recursive relationship

   $$f(x) = (1-p)f(x-1), \quad x = 2, 3, \dots,$$

   with the initial condition $f(1) = p$.

   (ii) Verify that $f(x) < f(x-1)$ for any $x = 2, 3, \dots$

   (iii) Show that the rth factorial moment of X

   $$\mu_{(r)} = E[X(X-1)(X-2)\cdots(X-r+1)]$$

   is given by the formula

   $$\mu_{(r)} = r!\,\frac{(1-p)^{r-1}}{p^r}, \quad r \geq 1.$$

   (*Hint*: For Part (iii), differentiate r times the geometric series

   $$1 + t + t^2 + \cdots + t^n + \cdots = \frac{1}{1-t}, \quad \text{for} \quad |t| < 1,$$

   and check that the following identity ensues:

   $$\sum_{x=r}^{\infty} (x)_r t^{x-r} = \frac{r!}{(1-t)^{r+1}}.$$

   Then use this identity to complete your derivation.)

10. Emily is in the final semester of her studies at the University. She has completed successfully all compulsory courses needed for her degree and she must pass at least four exams to finish her studies among seven possible optional courses offered. The probability of passing an exam in any course she takes is $p = 4/5$, but exams vary in difficulty. More specifically, Emily feels that in order to pass the first two exams she needs to study a further 15 hours for each, for the next three she has to study 20 hours for each course, while for each of the last two exams she has to study 25 hours. Assuming that she knows the outcome of each exam immediately after the exam finishes, she is obviously not willing to sit any further exams as long as she has passed four.

    (i) What is the probability that she takes all seven exams?

    (ii) Find the expected number of hours she needs to study from now.

Group B

11. At the NBA playoff finals, the championship is awarded in a best-out-of-seven series of games, i.e. the first team to reach four wins is the champion. Assume that in each game, the two teams, say A and B, have the same probability of winning.

    (i) What is the probability that team A wins the trophy? (You should be able to guess the answer by symmetry. Then you may verify formally that this is correct.)

    (ii) What is the probability that team A wins the trophy if we know that team B has won the first game?

    (iii) What is the probability that the series lasts for k games (for $k = 4, 5, 6, 7$)?

    (iv) If we know that the series was decided after k games, what is the probability that team B won the championship? Again, try to give an intuitive answer to this question without doing any calculations, and then check that this is correct (establishing in particular that the probability is the same for any value of k).

    (v) What is the expected number of games in the series?

12. We throw a die repeatedly and we stop when the outcome is either five or six.

    (i) What is the probability that k throws will be needed *and* the outcome of the kth throw is six?

    (ii) What is the probability that throws are terminated when a five occurs?

    (iii) If we know that at least three throws of the die will be needed, what is the probability that the series ends with a five?

    (iv) If we know that the series ends with a five, what is the probability that at least three throws of the die will be needed?

13. In a series of Bernoulli trials with success probability p, find the probability that k successes *in a row* occur prior to n failures *in a row*.

    (*Hint*: Let $\mathcal{E}$ be the event of interest, namely, $\mathcal{E}$: k successes in a row prior to n failures in a row. Consider now the (conditional) probabilities

    $$p_1 = P(\mathcal{E}|\text{the first trial is a success}), \quad p_2 = P(\mathcal{E}|\text{the first trial is a failure}).$$

    If we know that the first trial resulted in a success, for $\mathcal{E}$ to occur we must have that

    - the trials numbered $2, 3, \ldots, k$ all result in a success, or
    - there is a first failure at the mth trial for some $2 \leq m \leq k$, and then $\mathcal{E}$ occurs starting from a failure.

    This gives us one equation relating p_1 with p_2. We can get another one in a similar way. Thus, we obtain both p_1 and p_2 and hence the required probability.

14. The maximum number of independent Bernoulli trials in an experiment is $3r$, and the success probability in each trial is p. Trials terminate when either

- r successes have occurred, or
- it is not possible with the remaining trials to accumulate r successes.

For $r \leq x \leq 3r$, what is the probability that at the end of the experiment, x trials have been performed?

15. Solve the Banach matchbox problem (Example 5.14) if the two matchboxes do not contain initially the same number of matches. In other words, assume that the box in the right-hand pocket has N_1 matches and the one in the left-hand pocket has N_2 matches.

16. Anna and Steve play the following game. A die is thrown until either a five or a six appears for the first time. If the number of throws X is even, then Anna gives Steve the amount of a dollars; if X is odd, Anna receives from Steve b dollars.

    (i) What is the probability function of X?

    (ii) What is the probability that Anna wins this game?

    (iii) Show that for the game to be fair, we must have $2a = 3b$.

17. Let X be a random variable that follows the $\mathrm{Nb}(r, p)$ distribution. Show that for any positive integer k, the quantity

$$\mu_{[k]} = E[X(X+1)\cdots(X+k-1)]$$

is given by

$$\mu_{[k]} = \frac{r(r+1)\cdots(r+k-1)}{p^k}.$$

Incidentally, the quantity $\mu_{[k]}$ is called an *ascending factorial moment* of order k for the variable X. In particular, for $k = 1, 2$, check that the above result agrees with those given in Proposition 5.6.

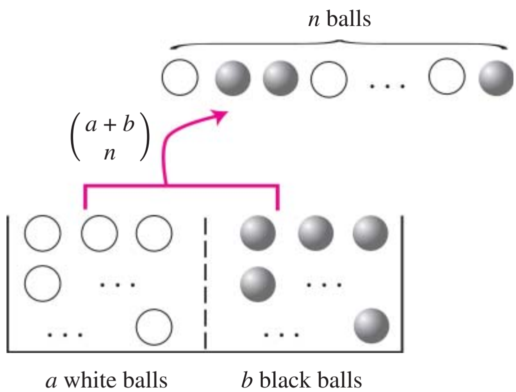
5.3 THE HYPERGEOMETRIC DISTRIBUTION

Suppose that members of a finite population are classified into two groups, say A and B, according to a certain characteristic. This could be, for example, gender, status of health (classified as healthy or not), the presence of defect in items produced by a production unit (so that these items are either defective or not), digit (with values 0 and 1) in the transmission of binary signals, and so on.

Then, we select a sample of size n from the above population, *without replacement*. When we are interested in how many units in the sample are in each of these groups (e.g. how many males and females we have in our sample), the resulting probability distribution is the **hypergeometric distribution**. Traditionally, the definition of this distribution is given in terms of an *urn model*.

Definition 5.5 From an urn that contains a white and b black balls, we select n balls without replacement. Let X be the number of white balls in the sample. Then, the distribution of X is called hypergeometric, and is denoted by $h(n; a, b)$.

The sample space for this experiment contains all possible selections of n balls out of a total of $a + b$ balls that are originally in the urn.



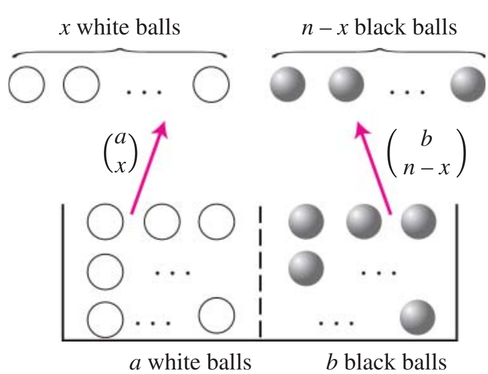
Therefore, the total number of these possible selections is $\binom{a+b}{n}$. Since balls are chosen at random, by symmetry, we have all these selections to have the same probability, and thus the probability for each of these is $\binom{a+b}{n}^{-1}$.

Employing the classical definition of the probability, we arrive at the following result for the probability function of the hypergeometric distribution.

Proposition 5.7 *The probability function of the hypergeometric distribution is given by*

$$f(x) = P(X = x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \quad x = \max\{0, n - b\}, \ldots, \min\{n, a\}.$$

Proof: The favorable outcomes for the event $\{X = x\}$, as shown in the following picture,



arise by selecting x white balls out of the a white balls that are available and $n - x$ black balls out of the b black balls that are originally in the urn. There are $\binom{a}{x}$ ways to select the white balls and $\binom{b}{n-x}$ ways to select the black balls. As a consequence, by the multiplication rule, the number of favorable outcomes equals

$$\binom{a}{x} \cdot \binom{b}{n-x}.$$

Since the number of possible selections is $\binom{a+b}{n}$, the probability of the event $\{X = x\}$ is simply

$$\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}.$$

As implied by the statement of the proposition, some care is needed when considering the range of values for x. In fact, x can in principle be any nonnegative integer such that x does not exceed the total sample size, n, so that

$$x = 0, 1, 2, \ldots, n.$$

But, there are only a white balls in the urn, and so we must have

$$x \leq a$$

and also there are only b black balls in the urn, and so in addition we must have

$$0 \leq n - x \leq b.$$

Combining these restrictions, we see that the following must simultaneously hold:

$$x \leq a, \quad x \leq n \quad \text{and} \quad x \geq 0, \quad x \geq n - b.$$

Thus, the smallest value that x can take is $\max\{0, n - b\}$ and the largest is $\min\{n, a\}$. □

For the probability function of the hypergeometric distribution, $h(n; a, b)$, we can also write

$$f(x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \quad x = 0, 1, 2, \ldots, n,$$

under the convention that

$$\binom{m}{r} = 0 \quad \text{if} \;\; m < r \;\; \text{or} \;\; r < 0.$$

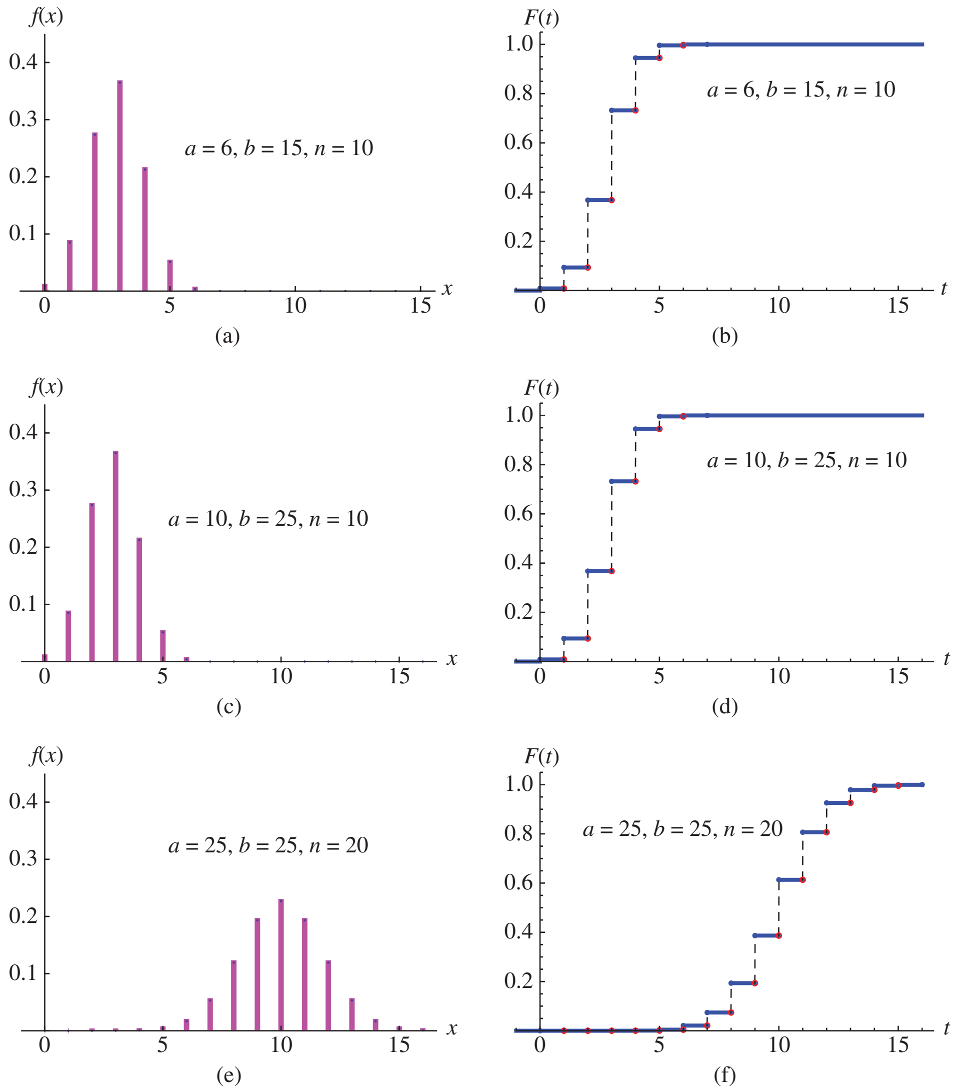


Figure 5.6 The probability function (left) and the cumulative distribution function (right) of the hypergeometric distribution for various choices of n, a, and b.

For the distribution function of the $h(n; a, b)$ distribution, we have

$$F(t) = \begin{cases} 0, & \text{if } t < \max\{0, n-b\}, \\ \displaystyle\sum_{x=\max\{0,n-b\}}^{[t]} \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, & \text{if } \max\{0, n-b\} \leq t < \min\{n, a\}, \\ 1, & \text{if } t \geq \min\{n, a\}. \end{cases}$$

In Figure 5.6, we have plotted the probability function and the cumulative distribution function of the hypergeometric distribution, $h(n; a, b)$, for various choices of n, a, b.

Example 5.15 In a factory, out of 50 machines produced during a day, 8 are defective having an operational error. An engineer selects six machines at random to examine whether they have this error or not. What is the probability that at least two of the machines selected are defective?

SOLUTION Let X denote the number of defective machines in the sample of the six machines selected by the engineer. Then, and since these 6 machines were selected out of the 50 without replacement, X has a hypergeometric distribution with parameters $n = 6$, $a = 8$, and $b = 42$ (the number of nondefective machines among the 50 ones produced that day). Hence, the probability function of X is

$$f(x) = P(X = x) = \frac{\binom{8}{x}\binom{42}{6-x}}{\binom{50}{6}}, \quad x = 0, 1, 2, \ldots, 6.$$

We want to find the probability $P(X \geq 2)$ and, as usual, it is easier to work with the probability of the complementary event. This gives

$$\begin{aligned}
P(X \geq 2) &= 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) \\
&= 1 - \frac{\binom{8}{0}\binom{42}{6}}{\binom{50}{6}} - \frac{\binom{8}{1}\binom{42}{5}}{\binom{50}{6}} \\
&= 1 - \frac{5\,245\,786}{15\,890\,700} - \frac{6\,805\,344}{15\,890\,700} = 0.241\,62.
\end{aligned}$$

Example 5.16 A music teacher wants to select eight high-school students from a class for the school choir. If in this class there are 17 girls and 13 boys, and assuming that all students are equally likely to be selected, what is the probability that among the 8 students selected there will be

(i) five girls and three boys?

(ii) no boys?

SOLUTION Let X be the number of girls selected for the choir. Then, under the assumptions of the example, $X \sim h(8; 17, 13)$ with probability function

$$f(x) = P(X = x) = \frac{\binom{17}{x}\binom{13}{8-x}}{\binom{17+13}{8}}, \quad x = 0, 1, 2, \ldots, 8.$$

(i) We require $P(X = 5)$ which, from the probability function of the hypergeometric distribution, is

$$\frac{\binom{17}{5}\binom{13}{3}}{\binom{17+13}{8}} = \frac{\frac{17!}{12!5!} \cdot \frac{13!}{10!3!}}{\frac{30!}{22!8!}} = \frac{6188 \cdot 286}{5\,852\,925} = 0.302,$$

that is, just over 30%.

(ii) Here, we want the probability that all eight students selected are girls, i.e. $P(X = 8)$, which is

$$\frac{\binom{17}{8}\binom{13}{0}}{\binom{17+13}{8}} = \frac{\frac{17!}{8!9!} \cdot 1}{\frac{30!}{22!8!}} = \frac{24\,310}{5\,852\,925} = 0.004,$$

or, about 0.4%.

Recall now the Cauchy formula in (2.16). This gives that, for positive n, a, b, we have

$$\sum_{x=0}^{n} \binom{a}{x} \cdot \binom{b}{n-x} = \binom{a+b}{n}$$

which, in turn, yields immediately that the sum of hypergeometric probabilities is equal to one, that is,

$$\sum_{x=0}^{n} \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} = \sum_{x=0}^{n} f(x) = 1.$$

Since we have clearly $f(x) \geq 0$ for all x, this shows that $f(x)$ is indeed a valid probability function of a discrete random variable. Essentially exploiting the same formula, one can find the rth factorial moment of the distribution as

$$\mu_{(r)} = E[(X)_r] = \sum_{x=0}^{n} (x)_r f(x) = \sum_{x=r}^{n} (x)_r \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}, \quad r = 1, 2, \ldots$$

In fact, using the identity (you may prove this as an exercise!)

$$(x)_r \binom{a}{x} = (a)_r \binom{a-r}{x-r},$$

we obtain

$$\mu_{(r)} = \frac{(a)_r}{\binom{a+b}{n}} \sum_{x=r}^{n} \binom{a-r}{x-r}\binom{b}{n-x}.$$

Next, making the substitution $y = x - r$, we notice that the sum on the right-hand side above equals

$$\sum_{y=0}^{n-r} \binom{a-r}{y} \binom{b}{n-(r+y)} = \sum_{y=0}^{n-r} \binom{a-r}{y} \binom{b}{(n-r)-y} = \binom{a-r+b}{n-r}$$

and this gives finally for $r = 1, 2, \ldots$,

$$\mu_{(r)} = \frac{(a)_r}{\binom{a+b}{n}} \cdot \binom{a+b-r}{n-r} = \frac{(a)_r (n)_r}{(a+b)_r}. \tag{5.12}$$

In view of the above, we can present the following proposition, which gives the mean and variance of the hypergeometric distribution.

Proposition 5.8 *If the variable X follows the hypergeometric distribution with parameters a, b, and n, then*

$$\mu = E(X) = n \cdot \frac{a}{a+b}, \quad \sigma^2 = \mathrm{Var}(X) = n \cdot \frac{a}{a+b} \cdot \frac{b}{a+b} \left(1 - \frac{n-1}{a+b-1}\right).$$

Proof: The result about the mean is an easy consequence of (5.12) setting $r = 1$. For the variance, it suffices to observe that

$$E[X(X-1)] = \mu_{(2)} = \frac{(a)_2 (n)_2}{(a+b)_2} = \frac{a(a-1)n(n-1)}{(a+b)(a+b-1)}$$

and use, as we did in the proof of Propositions 5.2 and 5.4, the fact that

$$\mathrm{Var}(X) = \mu_{(2)} + \mu - \mu^2.$$

The result is now a matter of straightforward calculations. □

An important point to note is the relation between the hypergeometric distribution and the binomial distribution, discussed in Section 5.1. First, we mention that in Definition 5.5, if the balls are chosen from the urn *with replacement* (i.e. we select a ball, note its color and return it to the urn), then the distribution of X that represents the number of white balls drawn is no longer hypergeometric, but it is the binomial distribution. This is so because before each selection, the number of white and black balls remains the same, and so in fact the successive selections of a ball from the urn constitute Bernoulli trials, where in each trial we regard the selection of a white ball as success and the selection of a black ball as a failure.

Thus, if there were a white balls and b black balls inside the urn before drawing the first ball, the same number from each color will be in the urn in every subsequent trial. Therefore, the probability of success in the binomial distribution will be

$$p = \frac{a}{a+b}$$

and clearly the probability of a failure is $q = 1 - p = b/(a+b)$.

Intuitively one expects that, when the total number of balls, $a + b$, becomes sufficiently large, there will effectively be little difference in the probabilities obtained by the hypergeometric and binomial distributions, since putting back one ball into the urn after each selection does not affect much the proportion of balls from each color there. So, for instance, if we take a sample of 50 persons from a population of half a million, and we are interested in the number of females in the sample, it matters little whether sampling is done with or without replacement.

The above heuristic considerations are made rigorous in the following result.

Proposition 5.9 *Suppose X has a hypergeometric distribution with parameters a, b, and n and probability function f. Then, as $a \to \infty$ and $b \to \infty$ in such a way that*

$$\lim_{a,b\to\infty} \frac{a}{a+b} = p,$$

we have that for $x = 0, 1, 2, \ldots, n$,

$$\lim_{a,b\to\infty} f(x) = \lim_{a,b\to\infty} P(X = x) = \binom{n}{x} p^x q^{n-x},$$

where $q = 1 - p$.

Proof: From the probability function of the hypergeometric distribution, we obtain

$$f(x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} = \frac{\dfrac{(a)_x}{x!} \cdot \dfrac{(b)_{n-x}}{(n-x)!}}{\dfrac{(a+b)_n}{n!}}$$

$$= \frac{n!}{x!(n-x)!} \cdot \frac{(a)_x (b)_{n-x}}{(a+b)_n} = \binom{n}{x} \frac{(a)_x}{(a+b)^x} \cdot \frac{(b)_{n-x}}{(a+b)^{n-x}} \cdot \frac{(a+b)^n}{(a+b)_n}. \tag{5.13}$$

But since $a/(a + b) \to p$ as $a, b \to \infty$, we have that for both a, b tending to infinity, the limit of the ratio $b/(a + b)$ is equal to $q = 1 - p$. Further,

$$\lim_{a,b\to\infty} \frac{1}{a+b} = \lim_{a,b\to\infty} \left(\frac{1}{a} \cdot \frac{a}{a+b}\right) = 0 \cdot p = 0,$$

so that writing the ratio $(a)_x/(a + b)^x$ in the form

$$\frac{(a)_x}{(a+b)^x} = \frac{a(a-1)\cdots(a-x+1)}{(a+b)(a+b)\cdots(a+b)} = \frac{a}{a+b} \cdot \frac{a-1}{a+b} \cdots \frac{a-(x-1)}{a+b}$$

$$= \frac{a}{a+b} \cdot \left(\frac{a}{a+b} - \frac{1}{a+b}\right) \cdots \left(\frac{a}{a+b} - \frac{x-1}{a+b}\right),$$

we obtain that

$$\lim_{a,b\to\infty} \frac{(a)_x}{(a+b)^x} = p^x.$$

In a similar way, we find

$$\lim_{a,b\to\infty} \frac{(b)_{n-x}}{(a+b)^{n-x}} = (1-p)^{n-x}, \qquad \lim_{a,b\to\infty} \frac{(a+b)_n}{(a+b)^n} = 1.$$

The result of the proposition now follows by inserting the last three displayed expressions into (5.13). □

The approximation of the hypergeometric distribution by the binomial is illustrated graphically in Figure 5.7. We note at this point that the approximation is remarkably good if both parameters a and b are much larger than n. For example, if $a = 200, b = 200$, and $n = 6$, the exact values from the hypergeometric distribution for $X = 3$, $X = 4$, and $X = 5$ are

$$P(X = 3) = 0.314\,867\,5, \quad P(X = 4) = 0.234\,957\,9, \quad P(X = 5) = 0.092\,566\,3,$$

while those from a binomial approximation using $p = a/(a + b) = 0.5$ are

$$P(X = 3) \cong 0.3125, \quad P(X = 4) \cong 0.234\,375, \quad P(X = 5) \cong 0.093\,75,$$

respectively. The relative percentage error for these probabilities is 0.75%, 0.25%, and 1.28%, respectively.

The next example, however, illustrates that the binomial approximation can be poor if a and/or b is of the same magnitude as n.

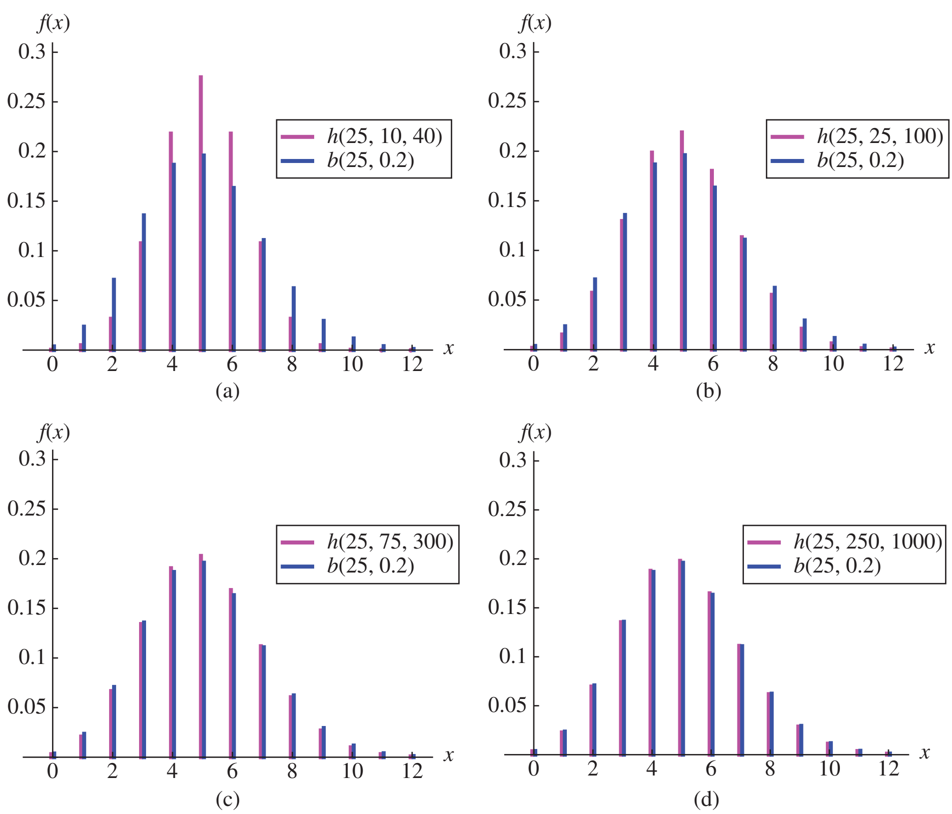


Figure 5.7 Approximation of the hypergeometric distribution by the binomial distribution for some choices of n, a, and b.

Example 5.17 In the top 100 tennis players, there are 13 left-handed players. We select randomly 16 players from the top 100.

(i) Find the mean and variance of the left-handed players in the sample.

(ii) Find the probability that, among the 16 selected players, 2 are left-handed. Compare your result with that obtained by the binomial approximation to the hypergeometric.

SOLUTION Let X be the number of left-handed tennis players in the sample of 16 players. Then, under the assumptions of the example and since the selection of players has been made without replacement, the distribution X is hypergeometric with parameters $a = 13, b = 100 - 13 = 87$, and $n = 16$.

(i) We know that the mean and variance of a hypergeometric random variable are

$$E(X) = n \cdot \frac{a}{a+b}, \quad \mathrm{Var}(X) = n \cdot \frac{a}{a+b} \cdot \frac{b}{a+b}\left(1 - \frac{n-1}{a+b-1}\right).$$

Thus, in our case and for the values of a, b, and n above, we get

$$E(X) = 16 \cdot \frac{13}{100} = 2.08$$

and

$$\mathrm{Var}(X) = 16 \cdot \frac{13}{100} \cdot \frac{87}{100}\left(1 - \frac{15}{99}\right) = 1.535.$$

(ii) We want to find $P(X = 2)$. Using the formula for probabilities from Proposition 5.7, we readily find

$$P(X = 2) = \frac{\binom{13}{2}\binom{87}{14}}{\binom{100}{16}} = 0.313\,464.$$

Next, if we want to use the binomial approximation to the hypergeometric, according to Proposition 5.9 the probability of success will be

$$p = \frac{a}{a+b} = \frac{13}{100} = 0.13,$$

and for $n = 16$ we obtain

$$P(X = 2) = \binom{16}{2}(0.13)^2(0.87)^{14} = 0.288\,63.$$

The binomial approximation is poor in this case, because the parameter $a = 13$ in the hypergeometric is smaller than the sample size $n = 16$.

EXERCISES

Group A

1. On a supermarket shelf, there are 45 packs of cereals. Among these, there are five packs whose sell-by date is less than a week from now. Lena selects four packs of cereals at random and intends to consume them after a week, since she has another pack of cereals at home. What is the probability that at least one of the four packs she bought will have past its sell-by date a week from now?

2. 16 players enter a table tennis tournament. Three of the players who enter the tournament are left handers while the rest are right handers. Assuming that initially all players are equally likely to reach the final,

   (i) what is the probability that two left handers meet in the final?

   (ii) what is the probability that the final will be between a left hander and a right hander?

3. Assume that a lake contains a total of N fish. Suppose we catch r among these fish, mark them in some way, and then put them back in the lake. After a certain amount of time, so that the marked fish have mixed well with the other ones in the lake, we select randomly n fish from the lake. Let X be the number of marked fish in that sample.

   (i) Find the probability function of X.

   (ii) What is the expected number of marked fish that will be caught in the sample of n fish?

   (The answer to Part (i) has been found already in Section 2.10, without any reference to the hypergeometric distribution. See that section for applications of this particular problem.)

4. An urn contains 15 balls numbered 1–15. We select three balls without replacement. Find the probability that

   (i) at least one number drawn is a prime;

   (ii) exactly one number drawn is a prime.

5. During an excavation, an archaeologist has discovered bones of 12 animals; 5 of these bones belong to rhinoceros, 4 belong to a mammoth, and 3 to a certain type of hippopotamus. She wants to select the bones of three animals in order to date them. Let X be the number of mammoths selected.

   (i) What is the sample space for this experiment?

   (ii) Write down the probability function of X.

   (iii) Find the mean and variance of X.

6. There are 10 blue and 16 red chips in a bowl. If we select five chips at random, what is the probability that more red chips are drawn than blue ones?

7. From a usual pack of 52 cards, we select 3 cards at random without replacement. Find the probability that

   (i) no ace is selected;

   (ii) a queen is selected;

   (iii) at least one queen is selected;

   (iv) no ace but at least one queen are selected.

8. A matchbox contains normally 40 matches. We select three matchboxes at random and we find seven matches in total to be defective. What is the probability that there are at least two defective matches in the first among these three boxes? Explain whether the binomial approximation to the hypergeometric would be appropriate here.

9. An urn contains a red balls and b black balls. We select n balls from this urn *with replacement*. If X denotes the number of red balls in the sample, derive the probability function and the expected value of the random variables X and $n - X$.

10. We select cards from a pack of 52 cards and let X be the number of cards picked until the first queen is drawn. Find the probability function and the expected value of X.

11. In a school class with 20 male students, 7 of them are white. The sports teacher wants to select five students for the basketball team (regular team) and seven more students as substitutes in that team.

    (i) What is the probability that there are exactly three white students in the regular team?

    (ii) What is the probability that there are exactly three white students in the regular team while two further white students are selected as substitutes?

12. A box contains six white stones and six black stones. Peter selects three stones at random without replacement. Let Y be the random variable that represents the number of white stones minus the number of black stones drawn.

    (i) What is the range of values of Y?

    (ii) Write down the probability function of Y.

    (iii) Suppose for each white ball drawn, Peter wins \$2 while for each black ball he loses \$1.50. What is his expected profit from this experiment?

Group B

13. In a lottery played in several countries, players have to choose 6 numbers from 1 to 49. Alice wants to buy a new iPad, and she would like to try her chances with the lottery for the first time. After a little research, she finds out that a single ticket of the lottery (i.e. if you select just 6 numbers out of the 49) costs \$1, while in order to win enough money to buy the iPad, she must predict at least 4 numbers, out of the 6 drawn, correctly. She feels that, if she buys a single ticket, her chances are very slim. She therefore decides to select eight numbers and, because she is quite

good at maths (and also, she has just finished a course in Probability), she works out that she has to pay \$28 (she is right, but explain why!). Then, she wants to find out what are her chances of predicting exactly four out of the six numbers in the next draw, and she comes up with an answer

$$\frac{\binom{8}{4}\binom{41}{2}}{\binom{49}{6}}$$

(she is right again, as you can verify). After that, she meets her friend Bill and tells him that she bought a ticket with eight numbers for the next lottery draw, but doesn't tell him what these numbers are.

On the night of the draw, Bill watches his favorite program on television and suddenly the six numbers drawn in the lottery appear on the screen. She remembers what Alice told him and wants to find out what are the chances that she won, so he argues as follows.

Alice may have chosen any 8 out of the 49 numbers, and there are $\binom{49}{8}$ ways to do this. The favorable outcomes (remember that he knows only the numbers drawn, not the ones that Alice chose) are $\binom{6}{4}$ (since Alice must have picked 4 numbers of the 6 drawn) multiplied by $\binom{43}{4}$, since the other 4 numbers on Alice's ticket are not winning numbers. Therefore, Bill thinks that the probability that Alice has picked four correct numbers is

$$\frac{\binom{6}{4}\binom{43}{4}}{\binom{49}{8}}$$

Show that this result agrees with Alice's result above.

The fact that the two results agree is an immediate consequence of the combinatorial identity

$$\frac{\binom{b}{x}\binom{N-b}{n-x}}{\binom{N}{n}} = \frac{\binom{n}{x}\binom{N-n}{b-x}}{\binom{N}{b}}$$

for any positive integers N, n, b, x such that $x \leq \min\{n, b\}$ and $N \geq \max\{n, b\}$. Verify this identity!

14. Under the assumptions for the lottery in the previous exercise, suppose that Wendy has just bought a lottery ticket.

    (i) What is the probability that among the six numbers she selected, at least four will appear in the next draw?

    (ii) Wendy decides to buy a lottery ticket for each of the following five draws. Find the probability that, in at least one draw, she has four winning numbers or more.

15. A referendum is going to take place in an European country in order to decide whether the country will adopt the euro as its currency unit or not. Suppose that the number of eligible voters in the referendum is 15 million people, and 53% of them are in favor of adopting the euro. If 750 000 (randomly selected) people do not turn up to vote,

    (i) give a formula for the probability function of the number X of people who do not turn up to vote and they are in favor of the euro;

    (ii) calculate the probability that the result of the referendum is against the adoption of the euro.

16. Jimmy likes to play a poker game with his friends once in a while. In the first hand, he receives 5 cards out of a pack of 52 cards.

    (i) What is the probability that he has a full house (i.e. three of a kind and two of another kind, e.g. three aces and two kings)?

    (ii) In the first 20 hands, he has never had a full house. Is Jimmy right in feeling particularly unlucky?

    (iii) Estimate the expected number of hands until Jimmy gets a full house.

17. Let X be a random variable such that $X \sim h(n; a, b)$. Show that the probability function f of X can be calculated via the following recursive scheme:

$$f(x+1) = \frac{(n-x)(a-x)}{(x+1)(b-n+x+1)} f(x), \quad \max\{0, n-b\} \leq x \leq \min\{a, n\},$$

with the initial conditions

$$f(0) = \frac{\binom{b}{n}}{\binom{a+b}{n}}, \quad \text{if } n < b,$$

and

$$f(n-b) = \frac{\binom{a}{n-b}}{\binom{a+b}{n}}, \quad \text{if } n \geq b.$$

5.4 THE POISSON DISTRIBUTION

The discrete distributions we have encountered so far arise frequently in situations involving sampling, with or without replacement, from a population which consists of units that fall into one of two types. More generally, we have considered situations where a sequence of experiments, or repetitions of the same experiment, is performed. Our experience suggests that the distributions we have met cannot serve as a model if, for example, one counts the occurrence of an event in continuous time. For instance, what is the distribution of the number of cars passing through a motorway during an hour? Or, how many telephone calls one receives during a day? One might argue that it is not totally unrealistic to use the binomial distribution for such phenomena – for example, to split the time axis in very small intervals, say one second, and define on each such interval a Bernoulli trial, defining as success the occurrence of the event we are interested in (in the above examples, a car passing through the motorway or an incoming telephone call), assuming that in such a small time length no more than one "success" may occur. Although this seems possible, it is rather inconvenient, and the formulation of the problem depends on the choice of the interval length to be used.

A common feature of the above examples (as in many other applications in practice) is that n is very large and, at the same time p is very small, whereas their product

$$\lambda = np$$

is of moderate size. Computationally, if one uses the binomial distribution in such a case, the calculation of probabilities becomes complicated. Imagine, for instance, the case where $X \sim b(n, p)$ with $n = 10^6$, $p = 3 \cdot 10^{-5}$ and we want to calculate the probability $P(X \geq 80\,000)$. Of course, in the current computer era, such calculations can be done efficiently and very quickly; note in particular the recursive relation (see Exercise 14 of Section 5.1)

$$f(x) = \left(a + \frac{b}{x}\right) f(x-1), \quad x = 1, 2, 3, \dots, n,$$

where

$$a = -\frac{p}{1-p}, \quad b = \frac{(n+1)p}{1-p}.$$

This relation, along with the initial condition $f(0) = (1-p)^n$, offers a rather easy recursive scheme to calculate the probabilities $P(X = x)$ for all $x = 0, 1, 2, \dots, n$. One disadvantage of this approach is that if we want the probabilities $P(X = x)$ for some large values of x, we need to find first the corresponding probabilities for all values $1, 2, \dots, x-1$.

In the nineteenth century, at a time when computers did not exist, the French mathematician Siméon Denis Poisson realized (and tackled) the problem of calculating binomial probabilities when n is very large and p is small. The solution he gave, in the form of a limiting theorem, in a book, which appeared in 1837, has given rise to modeling a very large number of phenomena in practice. In particular, under the above assumptions, Poisson found the limit, as $n \to \infty$, of the binomial probabilities. These limiting values constitute a new probability distribution, which has been aptly named after Poisson and is, along with the binomial, the most useful and commonly used discrete probability distribution.

Specifically, the assumptions we need are as follows:

- The number of trials n is very large ($n \to \infty$);
- The success probability p is very small ($p \to 0$);
- The product np, which is the expected number of successes after n trials, is of moderate magnitude. More precisely, we assume that np converges to a fixed $\lambda > 0$.

The next proposition describes explicitly the result established by Poisson.

Proposition 5.10 (*Limit of the binomial distribution, $b(n, p)$, for large n*) *Assume that the random variable X has a binomial distribution with parameters n and p so that*

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2 \dots, n.$$

If, as $n \to \infty$, the success probability[4] $p \to 0$ *in such a way that* $E(X) = np$ *converges to a positive constant* λ*, then*

$$\lim_{n\to\infty} f(x) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Proof: Introducing the notation $\lambda_n = np$, the binomial probability function f can be written in the form

$$f(x) = \frac{(n)_x}{x!} \cdot \frac{p^x q^n}{q^x} = \frac{(n)_x}{x!} \cdot \frac{\left(\dfrac{\lambda_n}{n}\right)^x \left(1 - \dfrac{\lambda_n}{n}\right)^n}{\left(1 - \dfrac{\lambda_n}{n}\right)^x}$$

$$= \frac{\lambda_n^x}{x!} \cdot \frac{(n)_x}{n^x} \cdot \frac{\left(1 - \dfrac{\lambda_n}{n}\right)^n}{\left(1 - \dfrac{\lambda_n}{n}\right)^x}. \tag{5.14}$$

We now calculate the limit, as $n \to \infty$, for each of the three terms in the last expression. First, as $n \to \infty$, we have by assumption $\lambda_n^x \to \lambda^x$, and since we further assume that

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = \lim_{n\to\infty} p = 0,$$

we get

$$\lim_{n\to\infty} \left(1 - \frac{\lambda_n}{n}\right)^x = 1.$$

Moreover, it follows from a standard calculus result that

$$\lim_{n\to\infty} \left(1 - \frac{\lambda_n}{n}\right)^n = \mathrm{e}^{-\lim_{n\to\infty} \lambda_n} = \mathrm{e}^{-\lambda}.$$

Considering the middle term in (5.14), we see that

$$\frac{(n)_x}{n^x} = \frac{n(n-1)\cdots(n-x+1)}{n^x} = \frac{n}{n} \cdot \frac{n-1}{n} \cdot \frac{n-2}{n} \cdots \frac{n-x+1}{n}$$
$$= 1\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{x-1}{n}\right) \to 1 \cdot 1 \cdots 1 = 1.$$

Using all the above facts, it follows from (5.14) that

$$\lim_{n\to\infty} f(x) = \frac{\lambda^x}{x!} \cdot 1 \cdot \frac{\mathrm{e}^{-\lambda}}{1} = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!},$$

which completes the proof of the proposition. □

[4]More formally, we should have written here $p_n \to 0$ as the success probability varies with n. However, for simplicity, we have throughout suppressed the dependence on n.

It is clear that the limiting probabilities $e^{-\lambda}\lambda^x/x!$ in the proposition define a proper probability distribution since

$$\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} = e^{-\lambda}e^{\lambda} = 1.$$

Definition 5.6 Let X be a random variable that takes values on the set $R_X = \{0, 1, 2, \dots\}$ and has probability function

$$f(x) = e^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots,$$

for some positive λ. Then, we say that X follows a Poisson distribution with parameter λ, and we denote this distribution by $\mathcal{P}(\lambda)$.

The cumulative distribution function of the $\mathcal{P}(\lambda)$ distribution is given by

$$F(t) = \begin{cases} 0, & \text{if } t < 0 \\ \displaystyle\sum_{x=0}^{[t]} f(x) = e^{-\lambda}\sum_{x=0}^{[t]}\frac{\lambda^x}{x!}, & \text{if } t \geq 0. \end{cases}$$

Figure 5.8 shows the Poisson probability function and cumulative distribution function for various values of λ, while Figure 5.9 illustrates the convergence of the binomial distribution to the Poisson distribution (Proposition 5.10).

Using Definition 5.6, we may restate the result of Proposition 5.10 as follows:

Proposition 5.11 *Assume that $X \sim b(n, p)$ and that, as $n \to \infty$, the success probability p tends to zero ($p \to 0$) in such a way that the mean $E(X) = np$ converges to some $\lambda > 0$. Then, the distribution of X can be approximated by the Poisson distribution with parameter λ.*

Mathematically, the result of Proposition 5.11 can be expressed (under the assumptions of the proposition) as follows:

$$f(x) = \binom{n}{x} p^x q^{n-x} \cong e^{-np}\frac{(np)^x}{x!}, \quad x = 0, 1, 2 \dots, n.$$

We note at this point that, for practical purposes, the quality of the approximation is generally satisfactory when the number of Bernoulli trials is at least 20 ($n \geq 20$) and $np \leq 10$, that is when the success probability p satisfies the condition $p \leq 10/n$.

Since the Poisson distribution emerges as the limiting distribution in situations involving the occurrence of an event with very small probability in each trial ($p \to 0$), it is often referred to as the **distribution of rare events**. Another term that is in use for the same distribution is the **law of small numbers**.

Example 5.18 (Car accidents) The probability that a car, passing through a dangerous spot in a national highway, has an accident is 0.0001. If 4500 cars pass through this spot during a weekend, what is the probability that there are at least three accidents?

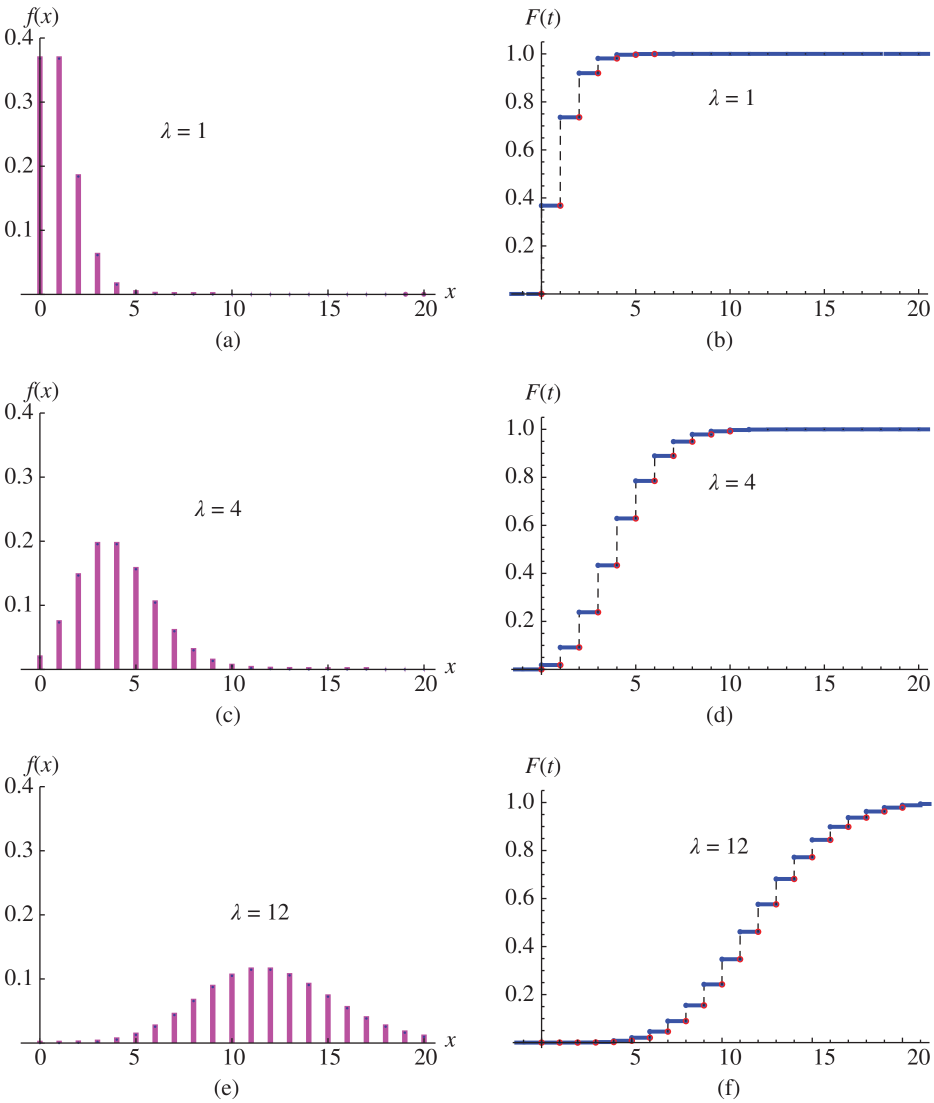


Figure 5.8 The probability function (left) and the cumulative distribution function (right) of the Poisson distribution for various choices of λ.

SOLUTION We assume that the involvement of any two different cars in an accident constitutes a pair of independent events. Then the cars passing through the spot may be thought of as a sequence of Bernoulli trials, each with success probability $p = 0.0001$ (note that in this way we regard a car accident as a "success"!). Then, the total number of accidents over the weekend, say X, has a binomial distribution with $n = 4500$ and $p = 0.0001$. Thus, for $x = 0, 1, 2, \ldots, 4500$,

$$P(X = x) = \binom{4500}{x}(0.0001)^x(0.9999)^{n-x}.$$

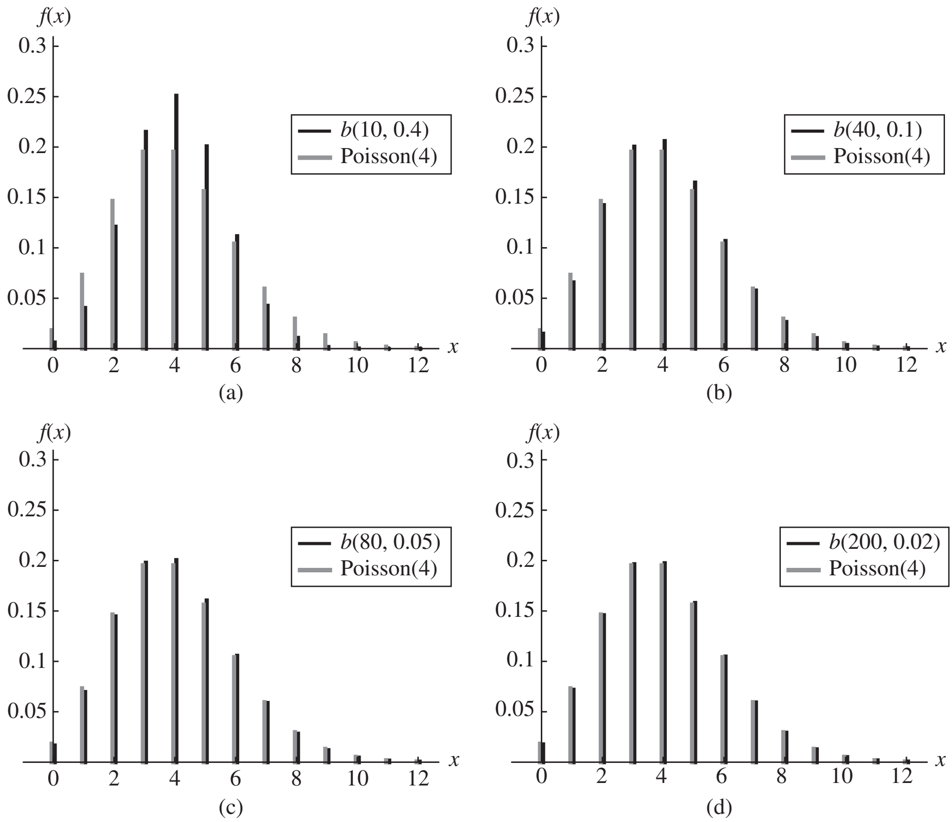


Figure 5.9 The probability function of the Poisson distribution with $\lambda = 4$ and the binomial approximation with the same mean, for various choices of n.

We seek $P(X \geq 3)$ and it is clearly much easier to find the probability of the complementary event. Thus, we find

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) = 1 - P(X = 0) - P(X = 1) - P(X = 2) \\ &= 1 - \binom{4500}{0}(0.0001)^0(0.9999)^{4500} - \binom{4500}{1}(0.0001)^1(0.9999)^{4500-1} \\ &\quad - \binom{4500}{2}(0.0001)^2(0.9999)^{4500-2} \\ &= 1 - 0.6376 - 0.2870 - 0.0646 = 0.010\,874. \end{aligned}$$

However, the required probability may also be found by applying the Poisson approximation to the binomial. The Poisson parameter in this case is

$$\lambda = np = 4500 \cdot (0.0001) = 0.45.$$

Thus,

$$P(X = x) \cong e^{-0.45} \frac{(0.45)^x}{x!}, \quad x = 0, 1, 2, \dots,$$

so that we have

$$\begin{aligned} P(X \geq 3) &= 1 - P(X < 3) \cong 1 - e^{-0.45} \frac{(0.45)^0}{0!} - e^{-0.45} \frac{(0.45)^1}{1!} - e^{-0.45} \frac{(0.45)^2}{2!} \\ &= 1 - e^{-0.45} \left(1 + 0.45 + \frac{(0.45)^2}{2} \right) = 1 - 1.551\,25e^{-0.45} \\ &= 0.010\,879. \end{aligned}$$

Clearly, the Poisson approximation is extremely good in this case, as the exact result and the approximation coincide to five decimal places and in fact the relative error of the approximation is less than 0.05%.

Example 5.19 Suppose that the number of goals scored in a football match has a Poisson distribution with parameter $\lambda = 2$. What is the percentage of football matches with

(i) two goals?

(ii) at least two goals?

(iii) more than two but less than six goals?

SOLUTION Under the assumptions of the example, if X denotes the number of goals scored in a football match, then $X \sim \mathcal{P}(2)$. Therefore,

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

(i) For $x = 2$, we simply get

$$P(X = 2) = e^{-2} \frac{2^2}{2!} = 2e^{-2} = 0.271,$$

i.e. the percentages of matches with exactly two goals is 27.1%.

(ii) Here we have

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-2} - 2e^{-2} = 0.594,$$

that is in 59.4% of football matches we expect two or more goals.

(iii) The required probability is $P(2 < X < 6)$ and, since the integer values between 2 and 6 are 3, 4, and 5, we get

$$\begin{aligned} P(2 < X < 6) &= P(X = 3) + P(X = 4) + P(X = 5) \\ &= e^{-2}\frac{2^3}{3!} + e^{-2}\frac{2^4}{4!} + e^{-2}\frac{2^5}{5!} \\ &= e^{-2}\left(\frac{8}{6} + \frac{16}{24} + \frac{32}{120}\right) = 0.3068. \end{aligned}$$

Thus, in 30.7% of the matches, we expect to have more than two but less than six goals scored.

The next proposition gives the mean and variance of the Poisson distribution.

Proposition 5.12 *If X has a Poisson distribution with parameter λ, then both the mean and the variance of X are equal to λ, that is*

$$\mu = E(X) = \lambda, \quad \sigma^2 = \mathrm{Var}(X) = \lambda.$$

Proof: Inserting the formula for Poisson probabilities into the general expression for the mean of a discrete random variable, we obtain

$$\begin{aligned} E(X) = \sum_{x \in R_X} f(x) &= \sum_{x=0}^{\infty} x e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!} \\ &= \lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!} = \lambda e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!} \\ &= \lambda e^{-\lambda}e^{\lambda} = \lambda. \end{aligned}$$

For the variance, it is again more convenient to start with the second factorial moment. We have

$$E[X(X-1)] = \sum_{x=0}^{\infty} x(x-1)e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=2}^{\infty}\frac{\lambda^x}{(x-2)!}.$$

Notice here that the last summation starts from $x = 2$, since for $x = 0$ and $x = 1$ the corresponding terms in the middle sum are zero. Making the substitution $x - 2 = m$ we obtain, in a similar fashion as above, that

$$E[X(X-1)] = \lambda^2 e^{-\lambda}\sum_{m=0}^{\infty}\frac{\lambda^m}{m!} = \lambda^2.$$

This, in turn, yields

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = \mu_{(2)} + \mu - \mu^2 = \lambda^2 + \lambda - \lambda^2 = \lambda,$$

as required. □

Example 5.20 In the French restaurant "Amélie," the average number of red wine bottles sold during a day is 12. We may assume that the number of bottles sold has a Poisson distribution.

(i) What is the probability that on a particular day the number of bottles sold will be at least three standard deviations less than its mean?

(ii) If there have been two orders for a bottle of red wine already, what is the probability that no more than five bottles of red wine will be sold in total?

SOLUTION If X denotes the number of bottles sold during a day, we know that $E(X) = 12$, and since X is assumed to follow a Poisson distribution, the parameter of this distribution will be $\lambda = 12$. This means that the variance of X will also be 12.

(i) We seek the probability

$$P\left(X \leq E(X) - 3\sqrt{\mathrm{Var}(X)}\right).$$

Since $E(X) = \mathrm{Var}(X) = \lambda = 12$, we subsequently obtain this to be

$$P(X \leq 12 - 3\sqrt{12}) = P(X \leq 1.608).$$

But X is integer-valued, and so the only possible values for it are 0 and 1. Therefore, the required probability equals

$$P(X \leq 1) = P(X = 0) + P(X = 1) = \mathrm{e}^{-\lambda} + \lambda \mathrm{e}^{-\lambda} = 13 \cdot \mathrm{e}^{-12} \cong 8 \cdot 10^{-5}. \tag{5.15}$$

(ii) Here we know that X, the number of bottles, is at least two and we seek the probability that it will be at most five, i.e. we want to calculate the conditional probability

$$P(X \leq 5 | X \geq 2) = \frac{P(X \leq 5, X \geq 2)}{P(X \geq 2)} = \frac{P(2 \leq X \leq 5)}{P(X \geq 2)}.$$

We now find each of the two terms in the last fraction. For the denominator, we have

$$P(X \geq 2) = 1 - P(X < 2) = 1 - P(X \leq 1) \cong 1 - 8 \cdot 10^{-5},$$

using the result in (i) above. For the numerator, we get

$$\begin{aligned} P(2 \leq X \leq 5) &= P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) \\ &= \mathrm{e}^{-12}\frac{12^2}{2!} + \mathrm{e}^{-12}\frac{12^3}{3!} + \mathrm{e}^{-12}\frac{12^4}{4!} + \mathrm{e}^{-12}\frac{12^5}{5!} \\ &\cong 3297.6 \cdot \mathrm{e}^{-12} = 0.020\,261. \end{aligned} \tag{5.16}$$

Thus, the required probability is

$$P(X \leq 5|X \geq 2) = \frac{0.020\,261}{1 - 8 \cdot 10^{-5}} = 0.020\,263.$$

We observe that the conditional probability $P(X \leq 5|X \geq 2)$ is very close to the unconditional probability $P(X \leq 5)$, which can be found by adding the results in (5.15) and (5.16). This can be attributed to the fact that the probability of the event $\{X \geq 2\}$ is very close to one, and so knowledge that this event has occurred does not affect much the probabilities $P(X = x)$ for $x \geq 2$.

In the examples of this section, we have given three practical situations in which the Poisson distribution might offer a reasonable probability model for the random variable of interest. There are many other circumstances in which the Poisson distribution is used frequently, as empirical studies suggest that it conforms with the data (or with the structure of the problem). We list some of them below:

- the number of claims arriving at an insurance company within a fixed time period;
- the number of telephone calls that arrive at a call center during a particular period;
- the number of customers visiting a shop during a day;
- the number of teenagers in a population who will be at least 6 ft 5 in. tall on their 18th birthday;
- the number of persons in a population who will live at least 90 years;
- the number of emissions from a radioactive source during a given time interval;
- the number of mutations in a given stretch of DNA after a certain amount of radiation;
- the number of particles emitted by a radioactive source in a second;
- the number of photons emitted by a celestial source of constant luminosity.

Suppose, for instance, that a motor insurance company has n customers insured in a certain portfolio. We denote by p the probability that a customer makes a claim for a car accident within a certain time period, such as a day or a week. Considering as "success" the claim made by a customer, the total number X of claims that the insurance company will have to pay within that period has a binomial distribution with parameters n and p. It is natural to assume that, in reality, n (the total number of insured customers) will be very large while p, the probability of a claim by a customer within such a short time period, will be very small. Moreover, the product $np = E(X)$, the expected number of claims arrived at the insurance company will typically be of modest magnitude (compared with the total number of insured persons, n). Thus, it seems appropriate to use the approximation

$$b(n, p) \approx \mathcal{P}(\lambda)$$

with $\lambda = E(X) = np$. In practice, the value of the parameter λ will not be known but will have to be estimated, usually not using the values of n and p (the latter may also be unknown) but from statistical data available to the company. The most common (and the simplest) estimate for $\lambda = E(X)$ an insurance company can use is the average number of claims (for the specified time period, daily or weekly) from a large number of such periods in the past.

Another area where the Poisson distribution is used quite frequently is ecology, especially in population ecology which studies the population growth, or spatial distribution, of animals or other organisms as they evolve over time and/or space. Further examples where the Poisson distribution can be an useful model for practical applications will be given in the next section of this chapter when we consider a concept generalizing the Poisson distribution, that of a Poisson process.

It should be kept in mind that, although a random variable having the Poisson distribution takes only integer values, the parameter λ of that distribution may not be an integer. We have already seen this in Example 5.18 and the next example illustrates this further.

Example 5.21 The number of fish caught by a fisherman during a day follows a Poisson distribution with parameter $\lambda = 3.5$.

(i) Find the probability that the fisherman catches at least two fish during a day.

(ii) If the fisherman goes fishing six days a week, find the probability

  (a) that in a particular week, on five days, he catches at least two fish, while on one day, he catches less than two fish;

  (b) that during a week, he catches at least one fish every day.

SOLUTION Let X denote the number of fish caught during a day. Then, $X \sim \mathcal{P}(\lambda)$ with $\lambda = 3.5$.

(i) The probability that at least two fish are caught in a day is $P(X \geq 2)$, and this equals

$$P(X \geq 2) = 1 - P(X = 0) - P(X = 1) = 1 - e^{-\lambda} - e^{-\lambda} \cdot \lambda = 1 - 4.5 \cdot e^{-3.5}$$
$$\simeq 0.8641.$$

(ii) Since we can assume that fishing in successive days constitutes a sequence of Bernoulli trials, for Part (a), we define the success event to be "at least two fish are caught during a day." Now, let Y be the number of successes during a week, i.e. in a sequence of six Bernoulli trials.

Then, Y has a Binomial distribution, $b(n, p)$, with parameters $n = 6$ and $p = 0.8641$ (the probability of success found in (i) above). We seek $P(Y = 5)$ and, from the formula for binomial probabilities, this is

$$P(Y = 5) = \binom{n}{5} p^5(1-p)^{6-5} = \binom{6}{5}(0.8641)^5(1 - 0.8641)^1 = 0.3928,$$

which is the required answer to Part (a).

For Part (b), we have another sequence of six Bernoulli trials, but now the success event is "at least one fish caught during a day." The success probability, p_1, in analogy with Part (i), equals

$$p_1 = P(X \geq 1) = 1 - P(X = 0) = 1 - \mathrm{e}^{-\lambda} = 1 - \mathrm{e}^{-3.5} = 0.9698.$$

If now W is a random variable that represents the number of days during a week that the fisherman catches at least one fish, then the distribution of W is again binomial with parameters $n = 6$ and $p_1 = 0.9698$. Consequently, the desired result is

$$P(W = 6) = \binom{n}{6}(p_1)^6(1 - p_1)^{6-6} = \binom{6}{6}(0.9698)^6(1 - 0.9698)^0 = 0.8319.$$

EXERCISES

Group A

1. The number of defects in a 100 ft long wire produced in a factory is Poisson distributed with a mean of 3.4. What is the probability that in a wire of that length,

   (i) there are at least three defects?

   (ii) there are at most three defects?

   (iii) there are at least five defects if, at a preliminary check, two defects have already been found?

2. At a zoo, the probability that a visitor requires medical attention during a day is 0.0004. Use the Poisson distribution to find the probability that, on a particular day with 3500 visitors, at least 2 will need medical attention.

3. In a company with 300 employees, find the probability that exactly 4 employees have their birthday on February 14th (assuming a nonleap year of 365 days)

   (i) using the binomial distribution;

   (ii) using the Poisson approximation to the binomial.

4. The number of animals caught in an animal trap during an hour has the Poisson distribution with parameter λ. If it is known that the probability that no animal is trapped during an hour equals $\mathrm{e}^{-2.5}$, find

   (i) the expected number of trapped animals during an hour;

   (ii) the probability that at least three animals are trapped;

   (iii) the probability that at least three animals are trapped in an hour, if we know that at least one has been trapped.

5. Razors produced by a machine are either defective, with probability 0.01, or nondefective.

   (i) In a batch of 200 razors, find the probability that

      (a) none is defective;

      (b) exactly two are defective

      using the binomial distribution.

   (ii) For each case in (i), find the percentage error if the Poisson approximation to the binomial is used instead.

6. The number of deaths at a large hospital during a month has the Poisson distribution. If the probability that at most one death occurs during a month is equal to the probability of having exactly two deaths, find the probabilities

   (i) to have at least one death during a month;

   (ii) to have at most three deaths during a month.

Group B

7. The number of burglaries in a city is thought to have a Poisson distribution, and it is known that the mean of this distribution is three burglaries per week.

   (i) What is the probability that in a given week, there are at most two burglaries?

   (ii) Find the probability that in a 40-week period, there are 5 weeks with at most two burglaries in each of them.

8. Let f_1, f_2 be the probability functions of the random variables X_1, X_2, which follow the Poisson distribution with parameters λ_1, λ_2, respectively, that is

$$f_i(x) = e^{-\lambda_i} \frac{\lambda_i^x}{x!}, \quad x = 0, 1, 2, \dots ,$$

   for $i = 1, 2$. Let μ_1, μ_2 be two constants with $0 \leq \mu_1, \mu_2 \leq 1$ and define the probability function $f(x)$ by the formula

$$f(x) = \mu_1 f_1(x) + \mu_2 f_2(x), \quad x = 0, 1, 2, \dots$$

   Find the condition that must be satisfied by the μ_i's so that f is a valid probability function of a random variable, which takes values on the nonnegative integers. Then, find the mean and variance of the resulting distribution.

   Application: In an automobile insurance portfolio, the insurance company estimates that the number of claims made per year by a male driver has a Poisson distribution with parameter λ_1, while the number of claims made per year by a female driver has a Poisson distribution with parameter λ_2. If it is known that 60% of the insured drivers are males, find the probability function and the expected value of the number of accidents per driver (either male or female) in this portfolio.

9. In the section that sells swimming suits at a large department store, the number of purchases in an hour follows the Poisson distribution with $\lambda = 4$. Purchases are made for both men's and women's suits, so that for a male suit one piece is sold, while for a female suit two pieces (top and bottom) are sold. The store has estimated that two-thirds of the swimming suits sold are for women. Let X denote the number of *pieces of suits* sold within an hour. Find the probabilities

   (i) $P(X = 0)$, (ii) $P(X = 2)$, (iii) $P(X = 6)$.

10. The number of particles emitted by a radioactive source in one minute is a random variable following the Poisson distribution with parameter $\lambda = 2.5$. In order to count the number of emissions in one minute, we put a Geiger counter device close to the source. But since this device has only a 1-digit counter, it registers all particles emitted only if their number is a 1-digit number, while if it is 10 or more it gives a result 9. If Y is the variable that represents the number shown in the counter, write down the probability function of Y, and hence find its mean and variance.

    (The distribution of Y in this exercise is an example of what is called a *clumped Poisson distribution*.)

11. Let X be a random variable with the Poisson distribution with parameter λ. Show that the rth factorial moment of X,

$$\mu_{(r)} = E[X(X-1)\cdots(X-r+1)], \quad r \geq 1,$$

    is given by

$$\mu_{(r)} = \lambda^r.$$

12. Consider the probability function of the $\mathcal{P}(\lambda)$ distribution

$$f(x) = P(X = x) = \mathrm{e}^{-\lambda}\frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

    (i) Show that $f(x)$ can be calculated via the recursion

$$f(x) = \left(a + \frac{b}{x}\right)f(x-1), \quad x = 1, 2, \dots,$$

    where $a = 0$ and $b = \lambda$, with initial condition $f(0) = \mathrm{e}^{-\lambda}$.

    (ii) Verify that $f(x) > f(x-1)$ if and only if $x < \lambda$.

    (iii) Prove that

    (a) if λ is not an integer, then f attains its maximum value at the point $x_0 = [\lambda]$;

    (b) if λ is a positive integer, then f has two maxima, attained at the points $x_0 = \lambda$ and $x_0' = \lambda - 1$.

13. Let $X \sim b(n, p)$ and $Y \sim \mathcal{P}(\lambda)$. If $\lambda = np$, so that X and Y have the same mean, consider the sequence

$$\alpha_n = \frac{P(X = n)}{P(Y = n)}, \quad n = 0, 1, 2, \dots$$

    Show that the sequence α_n is first increasing and then decreasing, attaining a maximum at the largest integer not exceeding λ.

5.5 THE POISSON PROCESS

We have seen in the previous section that the Poisson distribution is typically used for a random variable that counts occurrences of an event *in a fixed interval of time*. But what if we want to study the number of occurrences of an event *through time*? For instance, we mentioned in the last section that a natural assumption for the (random) number of emissions from a radioactive source or telephone calls arriving at a call center in a small time interval, such as one minute, is that it has a Poisson distribution. Is it true then that the number of emissions or telephone calls in a larger interval, such as an 10 minutes or an hour, still has the Poisson distribution? We may also be interested in drawing a graph of this number of emissions (or its expected value) as a function of time.

Moreover, as a motivation for the usefulness of the Poisson distribution (which is historically linked with the emergence of this distribution), we mentioned that it serves as an approximation to the binomial distribution. For this, as described formally in Proposition 5.10, we need a very large number of trials (n) and a very small probability of "success" (p). So, if we want to study occurrences of an event through time, we may split the time axis into very small intervals, such that within each interval the event of interest either happens (with a small probability) or it doesn't.

This partitioning of the time axis into "very small intervals" is made precise, and mathematically formal, in Definition 5.7 below by a passage to the limit as the length of each such interval tends to zero. Before that, we need to mention that the phenomenon we are currently studying cannot be described, in terms of probability theory, by a single probability distribution. For instance, if we are studying the number of telephone calls arriving at a call center between 10:59 a.m. and 11:00 a.m. on a particular day, this is a random variable, say N_1, which we may assume to follow the Poisson distribution. If we study instead the number of calls between 10:59 a.m. and 11:20 a.m., this is another random variable, say N_2, which will typically be much larger[5] than N_1, and it may (or may not) follow a Poisson distribution. In general, if we study the number of events that occur from some point in time, which we take as time zero, onward, we shall use the following notation. Let us denote by $N(t)$ the number of events that occur in the interval $[0, t]$. For a *fixed* time point t, $N(t)$ is a random variable in the usual sense. However, when we allow t to vary, we shall be interested in the *collection of random variables* $\{N(t), t \geq 0\}$. Any such collection, sometimes also called a family of random variables, is termed a *stochastic process*. The process we define below in connection with the Poisson distribution is both one of the simplest and one of the most important stochastic processes in probability theory.

Definition 5.7 A collection of random variables $\{N(t), t \geq 0\}$ is called a Poisson process if it satisfies the following conditions:

PP1. $N(0) = 0$ with probability one;

PP2. For any two time intervals of equal length, $(t, t + u]$ and $(s, s + u]$, the distribution of the number of occurrences in these intervals is the same. So this distribution depends only on the length of the interval and not its starting point in time;

[5] More formally, since N_1 and N_2 are both random variables, N_2 takes large values with higher probability than N_1.

PP3. For any nonnegative integer k and any interval $(t, t+u]$, the event that there are k occurrences in that interval is independent of the number of occurrences at any interval prior to time t. More specifically, if $0 \leq t_1 < t_2 < \cdots < t_n$ and we denote by A_i the event that there are r_i occurrences in the interval $(t_i, t_{i+1}]$ for $i = 1, 2, \ldots, n-1$, then the events $A_1, A_2, \ldots, A_{n-1}$ are (completely) independent;

PP4. The probability that there are two or more occurrences in an interval of length h tends to zero as h tends to zero from above, denoted by $h \to 0^+$, that is,

$$\lim_{h \to 0^+} \frac{P(N(h) \geq 2)}{h} = 0.$$

Property PP2 in the definition is usually called the property of *stationary increments*. Recall that $N(t)$ represents the (random) number of occurrences in $[0, t]$ for any t. Thus, $N(t+u) - N(t)$, which is called the increment of the stochastic process $\{N(t), t \geq 0\}$, gives the number of occurrences in the half-open interval $(t, t+u]$. Mathematically, this property tells us that for any nonnegative k,

$$P(N(t+u) - N(t) = k) = P(N(u) - N(0) = k)$$

and since $N(0) = 0$ by property PP1 (which simply says there is no occurrence at time zero), we get

$$P(N(t+u) - N(t) = k) = P(N(u) = k)$$

Further, property PP3 is called the property of *independent increments*.

From the discussion in the beginning of this section, one might feel that the Poisson process emerges from the Poisson distribution when we allow the time interval of observation to vary. Yet, in the above definition, there does not seem to be any link between the process $\{N(t), t \geq 0\}$ and the Poisson distribution. In fact, this link is given by the following result, proved by the Russian mathematician Alexander Khinchin (1894–1959).

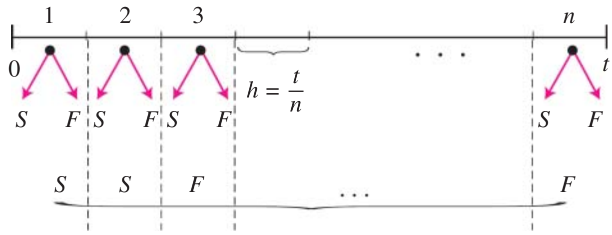
Proposition 5.13 *If an event $\mathcal{E}$ occurs through time in such a way that PP1–PP4 are satisfied, then there exists a $\lambda > 0$ such that the distribution of $N(t)$, i.e. the number of occurrences of $\mathcal{E}$ in an interval of length t, is given by*

$$P(N(t) = x) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^x}{x!}, \quad x = 0, 1, \ldots$$

The parameter λ is called the ***rate*** *associated with the Poisson process.*

Proof: Although there is obviously a rigorous proof of the proposition, using the law of the total probability and differential equations, this is beyond the scope of the present textbook and can be found in more advanced texts. Here, we proceed by giving a more or less intuitive argument which shows why the result of the proposition holds true.

We divide the interval $[0, t]$ into intervals of length $h = t/n$.



$N(t)$: number of successes

Assuming that n is large enough (theoretically, $n \to \infty$ so that $h \to 0$), PP4 shows that the event $\mathcal{E}$ may occur at most once in each interval of length h. If the occurrence of $\mathcal{E}$ in an interval of length h is a "success" and failure means that $\mathcal{E}$ does not occur, then the random variable $N(t)$ represents the number of successes in n trials. Consequently, for each fixed t the distribution of $N(t)$ is binomial with parameters n, p. Now, let λ denote the expected number of occurrences of $\mathcal{E}$ in a time unit (this unit is much larger than the small intervals of length h we considered earlier). Since there are t time units in the interval $[0, t]$, by the stationary increments property (PP2 in the definition), we deduce that the expected number of times $\mathcal{E}$ occurs in $[0, t]$ is λt. In other words, we must have

$$\lambda t = E[N(t)] = np,$$

which implies that $p = \lambda t/n$. But assuming that $n \to \infty$, we have that $p \to 0$ in such a way that $np \to \lambda t$ and the result follows now from Proposition 5.10. □

Note that, since the parameter $\lambda = E[N(t)]/t$ in a Poisson process is actually a *rate*, it is always important to specify the time unit that this refers to. Once this has been done, we may estimate the (often unknown in practice) value of λ by an appeal to the formula $\lambda = E[N(1)]$ (i.e. λ is simply the number of occurrences of $\mathcal{E}$ in one time unit).

Example 5.22 At an insurance company, claims for an insurance portfolio (e.g. for car insurance) arrive according to a Poisson process with a rate $\lambda = 3$ claims per day.

(a) Find the probability that there will be

  (i) five claims on a given day;

  (ii) no claims for a three-day period;

  (iii) at least six claims in a five-day period.

(b) What is the probability that in a five-day period, there will be exactly two days with no claims?

SOLUTION Let $N(t)$ be the number of claims that arrive at the company during the time interval $[0, t]$, when time is measured in days. Then

$$P(N(t) = x) = e^{-\lambda t} \cdot \frac{(\lambda t)^x}{x!} = e^{-3t} \cdot \frac{(3t)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(a) (i) Setting $t = 1$, $N(1)$ denotes the number of claims that arrive during a day. Thus, for $t = 1$, so that $\lambda t = 3$, the distribution of $N(1)$ will be $\mathcal{P}(3)$. As a consequence, we get

$$P(N(1) = 5) = e^{-3} \cdot \frac{3^5}{120} = 0.1001,$$

that is about a 10% chance.

(ii) For a three-day period, and since time is measured in days, we have $t = 3$, and this implies that $N(3)$, the number of claims in that period will have a Poisson distribution with parameter $\lambda t = 3 \cdot 3 = 9$. Hence, the required probability is

$$P(N(3) = 0) = e^{-9} \cdot \frac{9^0}{1} = e^{-9} = 0.000\,12,$$

which is about 0.01%, i.e. there is an extremely small chance that there will be no claims for three days in a row.

(iii) Here we have $t = 5$, so that $\lambda t = 15$ and, in analogy with above, we see that the number of claims in a five-day period, denoted by $N(5)$, is Poisson distributed with parameter 15. We want $P(N(5) \geq 6)$ and since the range of values for the Poisson is infinite, we work with the complementary event. More specifically, we have

$$\begin{aligned} P(N(5) \geq 6) &= 1 - [P(N(5) = 0) + P(N(5) = 1) + P(N(5) = 2) \\ &\qquad + P(N(5) = 3) + P(N(5) = 4) + P(N(5) = 5)] \\ &= 1 - e^{-15} \left[\frac{15^0}{0!} + \frac{15^1}{1!} + \frac{15^2}{2!} + \frac{15^3}{3!} + \frac{15^4}{4!} + \frac{15^5}{5!} \right] \\ &= 1 - 9128.5 \cdot e^{-15} = 0.9972, \end{aligned}$$

which means that it is almost certain (a 99.7% chance) that there will be at least six claims in a five-day period.

(b) For a one-day period, we have $t = 1$, so that $\lambda t = 3$. Therefore the probability that in any given day there will be no claims arriving at the company is

$$P(N(1) = 0) = e^{-3}.$$

Now, let X denote the number of days in the five-day period in which no claims will occur. Then, assuming independence between the numbers in successive days, X has a binomial distribution with parameters $n = 5$ and $p = e^{-3}$. Thus, the required probability is

$$P(X = 2) = \binom{5}{2} (e^{-3})^2 (1 - e^{-3})^{5-2} = \frac{5!}{3!2!} e^{-6} (1 - e^{-3})^3 = 0.0213,$$

i.e. just above 2%.

Example 5.23 The number of customers who enter a shop follows a Poisson process with a rate of three customers for a five-minute period.

(i) What is the probability that between 11:00 a.m. and 11:10 a.m. at least three customers will enter the shop?

(ii) Find the probability that, between 11:00 a.m. and 12:00 noon, 20 customers will arrive at the shop if we know that 12 customers have arrived between 11:00 a.m. and 11:30 a.m.

SOLUTION First, we note that the rate $\lambda = 3$ of the Poisson process is given for a five-minute period. Therefore, in order to find the parameter of the Poisson distribution for the number of customers entering the shop at any given interval, we have to consider how many five-minute intervals are included in that time interval.

(i) In view of the above, for a 10-minute interval, we have $t = 2$ and so the number of customers, $N(2)$, who arrive between 11:00 a.m. and 11:10 a.m. has a Poisson distribution with parameter $\lambda t = 3 \cdot 2 = 6$. This gives

$$\begin{aligned} P(N(2) \geq 3) &= 1 - [P(N(2) = 0) + P(N(2) = 1) + P(N(2) = 2)] \\ &= 1 - \mathrm{e}^{-6} \left[\frac{6^0}{0!} + \frac{6^1}{1!} + \frac{6^2}{2!} \right] = 1 - 25 \cdot \mathrm{e}^{-6} = 0.938. \end{aligned}$$

(ii) Noting again that the time unit is five minutes and taking 11:00 a.m. as the start time of the process, we see that the required probability is

$$P(N(12) = 20 | N(6) = 12).$$

But this is the same as the probability that there will be eight customers entering the shop between 11:30 a.m. and noon given that 12 people had entered by 11:30 a.m., or,

$$P(N(12) = 20 | N(6) = 12) = P(N(12) - N(6) = 8 | N(6) = 12).$$

We now use the fact that the Poisson process has independent increments, which means that the events $\{N(12) - N(6) = 8\}$ and $\{\mathrm{N(6)=12}\}$are independent; thus, the conditional probability above is equal to the unconditional probability of the event $\{N(12) - N(6) = 8\}$. Therefore we have

$$P(N(12) = 20 | N(6) = 12) = P(N(12) - N(6) = 8).$$

Next, we recall a second property of a Poisson process, namely that of stationary increments. This implies that the distribution of the number of customers in intervals of equal length is the same, or that the variables $N(12) - N(6)$ and $N(6)$ have the same distribution. As a result, we obtain that

$$P(N(12) - N(6) = 8|N(6) = 12) = P(N(6) = 8).$$

It is then straightforward to find the probability on the right-hand side by noticing simply that the variable $N(6)$ has a Poisson distribution with parameter $\lambda t = 3 \cdot 6 = 18$. This readily yields

$$\begin{aligned} P(N(12) = 20|N(6) = 12) = P(N(6) = 8) &= e^{-\lambda t} \cdot \frac{(\lambda t)^8}{8!} \\ &= e^{-18} \cdot \frac{18^8}{8!} = 0.004\,16, \end{aligned}$$

which is the required probability.

Example 5.24 The number of car accidents in a city follows a Poisson process with a rate of λ accidents per week. From police records, it has been found that a proportion p of these accidents are fatal (meaning there is at least one casualty).

(i) What is the probability that during a week there are at least x fatal accidents?

(ii) If $\lambda = 6$ and $p = 0.18$, find the probability that there are at least four fatal accidents in a four-week period.

(iii) For the same values of λ and p as in (ii), what is the expected number of nonfatal accidents during a four-week period?

SOLUTION Let $N(t)$ be the number of car accidents that occur in the city during the time interval $[0, t]$ and $X(t)$ the number of fatal accidents in the same time period. The events

$$\{N(t) = n\}, \quad n = 0, 1, 2, \ldots$$

form a *partition* of the sample space. Therefore, using the law of the total probability (see Section 3.3), we may write

$$P(X(t) = x) = \sum_{n=0}^{\infty} P(X(t) = x|N(t) = n) \cdot P(N(t) = n). \tag{5.17}$$

Notice now that, for $x > n$, the probability $P(X(t) = x|N(t) = n)$ equals zero, as the number of fatal accidents cannot exceed the total number of accidents. Moreover, for values of x such that $x \leq n$, if we know that n accidents have occurred, the number of fatal accidents has the binomial distribution with parameters n and p, since each accident may be regarded as a Bernoulli trial with success probability p (in this context, "success" means there is at least one fatality!). In other words, the conditional distribution of $X(t)$, given $N(t) = n$, is $b(n, p)$ and so

$$P(X(t) = x|N(t) = n) = \binom{n}{x} p^x q^{n-x}, \quad n \geq x,$$

where, as usual, $q = 1 - p$. Next, we note that $N(t) \sim \mathcal{P}(\lambda t)$ and, consequently, the term $P(N(t) = n)$ on the right-hand side of (5.17) is given by the Poisson probability $P(N(t) = n) = e^{-\lambda t}(\lambda t)^n/n!$.

In view of (5.17), the above considerations now yield that

$$\begin{aligned} P(X(t) = x) &= \sum_{n=x}^{\infty} \frac{n!}{x!(n-x)!} p^x q^{n-x} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \\ &= \frac{e^{-\lambda t}(\lambda pt)^x}{x!} \sum_{n=x}^{\infty} \frac{(\lambda qt)^{n-x}}{(n-x)!} = \frac{e^{-\lambda t}(\lambda pt)^x}{x!} e^{\lambda qt}. \end{aligned}$$

A straightforward calculation shows that the last expression is the same as

$$P(X(t) = x) = e^{-\lambda pt} \frac{(\lambda pt)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

(i) The last result shows that $\{X(t),\ t \geq 0\}$ is *also* a Poisson process with rate λp. Therefore, the probability that there will be at least x fatal accidents in a given week (so $t = 1$) is

$$P(X(t) \geq x) = \sum_{r=x}^{\infty} e^{-\lambda p} \frac{(\lambda p)^r}{r!}.$$

(ii) For a four-week period, we have $t = 4$ and thus, from Part (i) above, the number of fatal accidents has a Poisson distribution with parameter $\lambda pt = 6 \cdot (0.18) \cdot 4 = 4.32$ accidents. The probability that there are at least four fatal accidents in a four-week period is then given by

$$\begin{aligned} P(X(4) \geq 4) &= 1 - [P(X(4) = 0) + P(X(4) = 1) + P(X(4) = 2) + P(X(4) = 3)] \\ &= 1 - \left[e^{-4.32} + e^{-4.32} \frac{(4.32)^1}{1!} + e^{-4.32} \frac{(4.32)^2}{2!} + e^{-4.32} \frac{(4.32)^3}{3!} \right] \\ &= 1 - e^{-4.32}[1 + 4.32 + 9.3312 + 13.4369] = 1 - 28.0881 \cdot e^{-4.32} \\ &= 0.6264. \end{aligned}$$

(iii) We have seen in Part (ii) above that $X(4) \sim \mathcal{P}(4.32)$. Consequently, the expected number of fatal accidents during a four-week period is simply 4.32. The expected number of total accidents (both fatal and nonfatal) during that period is $6 \cdot 4 = 24$, and consequently the expected number of nonfatal accidents is $24 - 4.32 = 19.68$.

This section has only served as a beginner's introduction to a very important stochastic process, namely, the Poisson process. There are various other important properties that have not been mentioned here, as they require concepts and material which we have not covered so far. Here we merely mention that, although a Poisson process typically measures occurrences of an event through time, we may also have *spatial Poisson processes*, wherein we count the occurrence of an event as it evolves in space (in one, two, or three dimensions). Finally, before closing this section, we mention that there is a large number of natural phenomena, in time or space, which can be adequately modeled by a Poisson process in practice. Below, we list some of these applications:

- The arrival of customers in a queue, for example, at a cashier in a bank or in a large store;
- the number of goals scored, as time evolves, in a football match;
- the number of raindrops falling over an area;
- the distribution of stars in a galaxy;
- the locations of weeds in the lawn;
- the number of claims arriving at an insurance company[6] ;
- the number of particles emitted via radioactive decay by an unstable substance (generally, in radioactive decay, the nucleus of an atom which is in an unstable condition, loses energy by emitting ionizing particles or radiation).

EXERCISES

Group A

1. Bees sit on a particular flower according to a Poisson process with a rate of $\lambda = 6$ bees per minute. What is the probability that
   (i) exactly four bees visit the flower in two minutes?
   (ii) at least one bee visits the flower in a half-minute period?

2. The arrival of airplanes at an airport can be modeled by a Poisson process with a rate $\lambda = 5$ arrivals per hour.
   (i) What is the probability that there will be at least one arrival between 3:30 p.m. and 5:00 p.m. on a particular day?
   (ii) Find the expected value and the variance of the number of airplanes that will arrive between 3:30 p.m. and 5:00 p.m. of that day.

3. The number of hits to a website follows a Poisson process with a rate of 8 hits per minute. What is the probability that there will be 20 hits in a five-minute period, if we know that there have been 7 hits during the first minute of that period?

[6]This may be plausible for certain types of insurance such as for vehicle insurance, while for other types, such as for insurance related to natural disasters (floods, earthquakes, etc.) the conditions for a Poisson process may not be satisfied.

4. Irene has a Facebook page and she is very keen to have a large number of friends there. The number of friends added to her page follows a Poisson process with a rate of $\lambda = 3$ persons per week.

   (i) What is the probability that on a particular week she makes less than three new friends?

   (ii) Find how long she has to wait from now on so that she makes at least one new friend with a probability of 99%.

   (iii) Suppose that during a particular week she made no new friends between Sunday and Friday, and so on Friday night she felt very disappointed and wanted to know the probability that there would be at least one new friend on Saturday of that week. Can you give an answer to her question?

5. Customers arrive at a bank according to a Poisson process with a rate of $\lambda = 2$ customers for a five-minute period. Find the probability that

   (i) three customers will enter the bank between 2:00 p.m. and 2:15 p.m.;

   (ii) three customers will enter the bank between 2:00 p.m. and 2:15 p.m. *and* two customers will enter the bank between 2.15 p.m. and 2:30 p.m.;

   (iii) three customers will enter the bank between 2:00 p.m. and 2:15 p.m. *and at most* two customers will enter the bank between 2:15 p.m. and 2:30 p.m.

6. In the wire production unit of a factory, there is an employee who inspects the quality of the wire as it comes out of the machine that produces it. It has been estimated that the number of defects on the wire follows a Poisson process with a rate of one defect per 100 m of wire produced. One day, during work, the employee is called to answer an urgent phone call and he is absent from his post for 20 minutes. If the machine produces 30 m of wire per minute, find the probability that the employee has missed during his absence more than one defects on the wire.

7. The number of goals that Real Madrid scores in the football Champions League competition follows a Poisson process with a rate of $\lambda = 0.12$ goals in a five-minute period. Find the probability that in their next three matches, Real will score exactly one goal in two of them and three goals in the other.

Group B

8. In seismology, an earthquake is said to be "strong" if it has a magnitude of at least six measured on the Richter scale. Imagine that in an area which is frequently hit by earthquakes, the number of strong earthquakes follows a Poisson process with a rate of 2.5 per year.

   (i) What is the probability that there will be at least three earthquakes of that magnitude (a) in a three-month period; (b) in a year?

   (ii) Find the probability that in the next 10 years there will be exactly 3 years in each of which there will be at least 3 strong earthquakes.

9. The number of trees in a forested area of S square feet follows a Poisson process with rate λ.

(i) Find the probability that the distance from a given tree to the one closest to it will be at least α feet.

(ii) Calculate the probability that the distance between a given tree to the tree which is *the third closest to it* will be at least α feet.

10. (i) Using the identity $\sum_{r=0}^{\infty} t^r/r! = e^t$, prove that for any real t, the following identity holds:

$$\sum_{r=0}^{\infty} \frac{t^{2r}}{(2r)!} = \frac{1}{2}(e^t + e^{-t}).$$

(ii) Telephone calls arrive at a large company according to a Poisson process with a rate λ per minute. If time is measured in minutes, and $t < s$ are two real numbers, what is the probability that in the time interval $(t, s]$ there will be an even number of phone calls (that includes zero phone calls) arriving at the company?
(*Hint*: You may find the result in Part (i) to be useful.)

11. Nick, who is a car mechanic, receives damaged cars for repair at a rate of three cars every two hours. He has estimated that, among the cars arriving 75% are minor repairs, which he can fix by himself, while the rest should be sent to the central repair store. Find the probability that there will be at least two cars needing a major repair between 8:00 a.m. and 11:30 a.m. on a particular day.

12. The number of fraudulent credit card uses in a large store follows a Poisson process with a rate of $\lambda = 3$ per day. The store is open 12 hours each day (except Sundays).

(i) What is the expected number of frauds during a week, i.e. in six working days?

(ii) If the average cost of a fraud is \$240, what is the expected cost that will be incurred as a result of fraudulent credit card use between 12:00 noon and 4:30 p.m. on a particular day?

5.6 BASIC CONCEPTS AND FORMULAS

Bernoulli trial	An experiment with two possible outcomes: success (S) and failure (F)
Bernoulli distribution with parameter p	$f(x) = p^x q^{1-x}, \quad x = 0, 1 \ (q = 1 - p)$; $E(X) = p, \quad \mathrm{Var}(X) = pq$
Binomial distribution $b(n, p)$	The distribution of the number of successes in n independent Bernoulli trials with the same success probability p. $f(x) = \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, \ldots, n$; $E(X) = np, \quad \mathrm{Var}(X) = npq$
Geometric distribution $G(p)$	The distribution of the number of Bernoulli trials (independent, all with success probability p) until the first success. $f(x) = q^{x-1}p, \quad x = 1, 2, \ldots$; $E(X) = \frac{1}{p}, \quad \mathrm{Var}(X) = \frac{q}{p^2}$

Negative binomial distribution $\mathrm{Nb}(r, p)$	The distribution of the number of Bernoulli trials (independent, all with success probability p) until the rth success. $f(x) = \binom{x-1}{r-1} p^r q^{x-r}, \quad x = r, r+1, r+2, \ldots;$ $E(X) = \frac{r}{p}, \quad \mathrm{Var}(X) = \frac{rq}{p^2}$
Hypergeometric distribution $h(n; a, b)$	The distribution of white balls selected in a random sample of size n, taken without replacement from an urn that contains a white and b black balls. $f(x) = P(X = x) = \frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}};$ $E(X) = n \cdot \frac{a}{a+b},$ $\mathrm{Var}(X) = n \cdot \frac{a}{a+b} \cdot \frac{b}{a+b} \cdot \left(1 - \frac{n-1}{a+b-1}\right)$
Approximation to $h(n; a, b)$ by $b(n, p)$	If $a \to \infty$ and $b \to \infty$ such that $\frac{a}{a+b} \to p$, then $\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}} \longrightarrow \binom{n}{x} p^x q^{n-x}, \quad x = 0, 1, 2, \ldots, n$
Poisson distribution $\mathcal{P}(\lambda)$	$f(x) = \mathrm{e}^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots;$ $E(X) = \lambda, \quad \mathrm{Var}(X) = \lambda$
Approximation to $b(n, p)$ by $\mathcal{P}(\lambda)$	If $n \to \infty$ and $p \longrightarrow 0$ such that $np \to \lambda > 0$, then $\binom{n}{x} p^x q^{n-x} \longrightarrow \mathrm{e}^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$
Poisson process with rate λ	If $N(t)$ is the number of occurrences of an event $\mathcal{E}$ in the interval $[0, t]$, then under certain assumptions (see Section 5.5), $P(N(t) = x) = \mathrm{e}^{-\lambda t} \frac{(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots$

5.7 COMPUTATIONAL EXERCISES

Mathematica has a very powerful machinery to handle discrete (and, as we shall see in the following chapters, continuous) distributions. If *dist* stands for any particular discrete distribution, Table 5.1 lists some of the main Mathematica functions that can be used to obtain quantities associated with *dist*.

Mathematica has implemented[7] all the discrete distributions we have met in this chapter, along with several others. The ones we use more frequently are listed in Table 5.2.

[7]For earlier versions of Mathematica, fewer distributions are available and, in order to make use of them, one needs to upload first the corresponding package; this is accomplished with the command Statistics "Discrete Distributions."

Mathematical function	Result
PDF[dist,x]	Gives the value of the probability function for the distribution *dist* at the point *x*
CDF[dist,x]	Gives the cumulative distribution function of *dist* calculated at the point *x*
Mean[dist]	Gives the mean of *dist*
Variance[dist]	Gives the variance of *dist*
StandardDeviation[dist]	Gives the standard deviation of *dist*
Random[dist]	Produces a set of (pseudo)random numbers from *dist*

Table 5.1 Mathematica functions used to obtain various quantities associated with a distribution.

Mathematica function	Distribution
BernoulliDistribution[p]	$b(1, p)$
BinomialDistribution[n, p]	$b(n, p)$
GeometricDistribution[p]	$G(p)$
NegativeBinomialDistribution[r, p]	$\mathrm{Nb}(r, p)$
HypergeometricDistribution[$n, a, a + b$]	$h(n; a, b)$
PoissonDistribution[s]	$\mathcal{P}(s)$
DiscreteUniformDistribution[$\{1, n\}$]	$\mathcal{DU}(n)$

Table 5.2 Mathematica functions for discrete distributions.

Notice that, apart from the distributions we have discussed in this chapter, Mathematica has a special function for a distribution called the *Discrete Uniform Distribution* (see Table 5.2). This is used rather less frequently than other distributions listed in Table 5.2, but we have already seen examples that can be modeled by this distribution. More explicitly, we say that a random variable X has a Discrete Uniform distribution with parameter n, denoted by $X \sim$ DU (n), if X can take any of the values $1, 2, \ldots, n$, each with the same probability. Since these probabilities must add up to one, it follows that the probability function for the $\mathcal{DU}(n)$ distribution is

$$P(X = i) = \frac{1}{n}, \quad \text{for } i = 1, 2, \ldots, n.$$

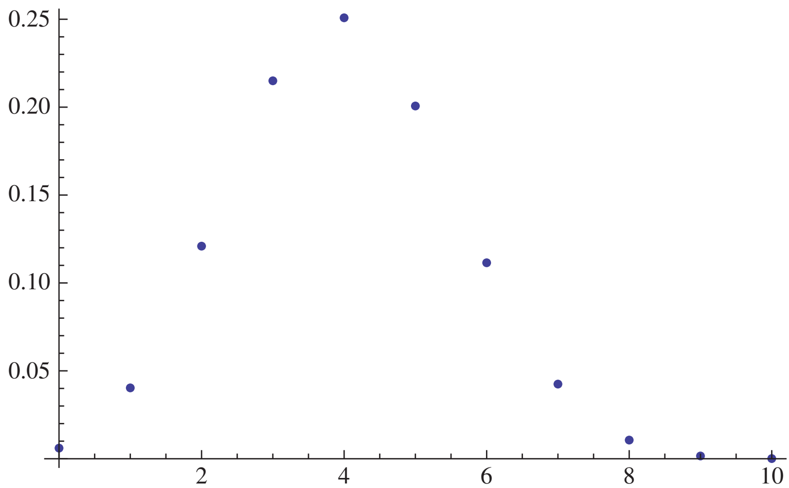
This distribution is sometimes referred to as the "equally likely outcomes" distribution. One of the simplest examples of the discrete uniform distribution is when throwing a fair die, in which case $n = 6$. The following short program gives an example of the use of the various functions mentioned above, showing how Mathematica could be used to answer the questions in Example 5.2.

```
In[1]:= n=10;p=0.40;
        f[x_]:=PDF[BinomialDistribution[n,p],x]
        Print["Probability function"];
        probfun=Table[{x,f[x]},{x,0,n}]
        Print["Plot of the Probability function"];
        ListPlot[probfun]
```

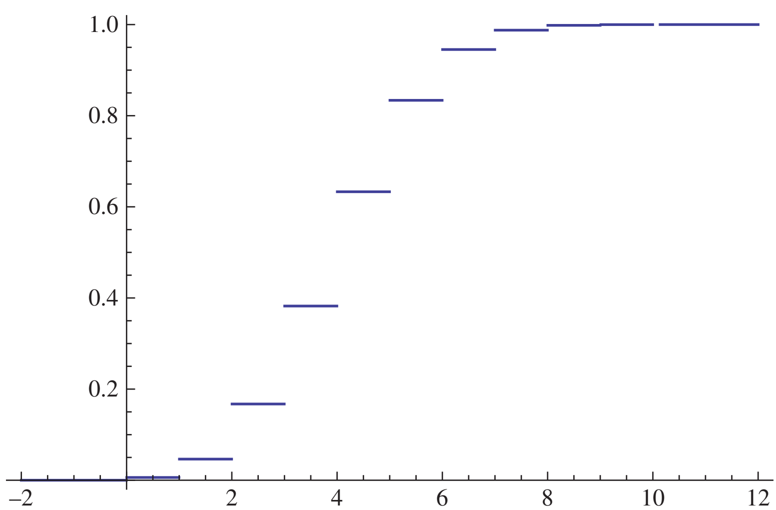
```
        F[t_]=CDF[BinomialDistribution[n,p],t];
        Print["Plot of the cumulative distribution function"];
        Plot[F[t],{t,-2,12}]
        Print["P[X=0]  = ",f[0]]
        Print["P[X<=2]  = ",F[2]]
        Print["P[X>=6]  = ",Sum[f[x],{x,6,10}]]
```

```
Probability function
Out[4]= {0, 0.00604662}, {1, 0.0403108}, {2, 0.120932}, {3, 0.214991},
{4, 0.250823}, {5, 0.200658}, {6, 0.111477}, {7, 0.0424673},
{8, 0.0106168}, {9, 0.00157286}, {10, 0.000104858}
```

Plot of the Probability function



Plot of the cumulative distribution function




```
P[X=0]  = 0.00604662
P[X<=2]  = 0.16729
P[X>=6]  = 0.166239
```

EXERCISES

1. Use Mathematica to draw a graph of the probability functions and the cumulative distribution functions of the seven discrete distributions given in Table 5.2, for various choices of their parameters.

2. As in Example 5.5, suppose Jimmy shoots against a target with the probability of hitting the target being p, and successive shootings are assumed independent. This time he is told that a prize is awarded if he chooses to shoot $2n$ times and hits the target at least n times. He has been offered the choice of n. Find for which values of p he wins the prize if $n = 1, 2, 3, 4, 5$.

   Note: For $n \geq 2$, it may not be easy to solve the corresponding inequality for p, and so you can find the solution either graphically or, preferably, using the command NSolve[f[x]==0,x].

3. An airline estimates that 3% of the persons who book a seat to travel do not show up. On that basis, suppose that the company makes 350 reservations for an aircraft capacity of 340 seats. What is the probability that every passenger who shows up for travel finds a seat?

4. How many times do we have to throw a die so that the outcome in at least one throw is less than three and in at least one throw it is greater than two with probability more than or equal to

   (i) 50%; (ii) 90%; (iii) 95%; (iv) 99%.

   Suggestion: Work as in Example 5.6 and use Mathematica to solve the resulting inequality.

5. Create a large number (e.g. 10 000) of observations from the geometric distribution with parameter $p = 0.10, 0.50, 0.90, 0.99$. In each case, calculate the arithmetic mean of these observations. How close are these values to the theoretical mean of the distribution?

6. In a certain lottery, there is a total of 300 tickets and 40 of them win some amount. If we purchase 15 tickets, what is the probability that we have

   (i) exactly two winning tickets?

   (ii) at least six winning tickets?

   (iii) at most 10 winning tickets?

7. Make a graph of the hypergeometric distribution, $h(n; a, b)$, for $n = 30$, $a = 3k$, $b = 2k$, where $k = 50,100, 150, \ldots$, and compare this with the graph of the binomial distribution with $n = 30, p = 0.6$. What do you observe? Can you justify this result?

8. Draw the graph of the binomial distribution with $n = 10, 50, 100, 200, 300, 500$, and $p = 5/n$. Then compare, in each case, the graph with that of the Poisson distribution with $\lambda = 5$. Again, try to justify your findings.

9. Fill in the table below with the values of the probability functions for the distributions mentioned in the table. What do you observe? Make a similar table when the values of the probability function are replaced by those of the cumulative distribution function.

	$h(n; a, b)$					$b(20, p)$
x	$a = 30$ $b = 20$	$a = 90$ $b = 60$	$a = 150$ $b = 100$	$a = 900$ $b = 600$	$a = 1500$ $b = 1000$	$p = \dfrac{a}{a+b}$
0						
1						
2						
.						
.						
.						
15						

10. Make a table similar to the one above, and fill in the corresponding values of the probability functions, for the distributions in Exercise 7.

5.8 SELF-ASSESSMENT EXERCISES

5.8.1 True–False Questions

1. The owner of a restaurant allocates the shifts to the waiters of the restaurant in a way such that each waiter has a probability of $3/10$ of having a day off on a Saturday (where tips are usually higher). If the random variable X denotes the number of weeks a waiter has to wait until he works on a Saturday, then X follows the Poisson distribution.

2. Let X be a random variable denoting the number of roulette spins until a black-colored number appears. Then, the distribution of X is geometric. (To answer the question, it is immaterial how many numbers, such as 36, 37, 38, etc., there are on the roulette wheel).

3. Twenty persons of the same age and the same health status are insured with a certain company. From the company's records, it is estimated that each person of that age has a chance of 0.65 of being alive in 20 years' time. The expected number of persons that will be alive in 20 years' time is 15.

4. If X has a binomial distribution with parameters n and p, then the random variable $Y = n - X$ also has a binomial distribution with parameters n and p.

5. If X has a geometric distribution, then

$$P(X > k + r | X > k) = P(X > r)$$

for any nonnegative integers k and r.

6. If X has a geometric distribution, then

$$P(X \leq k + r | X \leq k) = P(X \leq r)$$

for any nonnegative integers k and r.

7. Consider a sequence of Bernoulli trials with the same success probability, p. The expected number of trials until the first failure is equal to $1/p$.

8. An urn contains a white and b black balls. We select a ball randomly, note its color and return it to the urn. We repeat this experiment and let X be the number of trials until a black ball is selected for the first time. Then, X has a geometric distribution with parameter $p = a/(a + b)$.

9. For $n \geq 20$ and $p \leq 10/n$, the following approximation can be used:

$$(1 - p)^n \approx \mathrm{e}^{-np}.$$

10. If it is known that the number of motor accidents in a city during a week has the Poisson distribution with parameter 1, then the probability of having at least one accident during a week is $1 - 2\mathrm{e}^{-1}$.

11. An urn contains a red and b black balls. We select a ball randomly, note its color, but do not return it to the urn. We repeat this experiment n times (where $n \leq \min\{a, b\}$) and let X be the number of red balls selected. Then the expected value of X is an/b.

12. Assuming that both genders are equally likely, the probability that a family with n children has at least one girl is $1 - 2^{-n}$.

13. The minimum number of children a family must have so that it has at least one boy with a probability not less than 90% is four (assume again that it is equally likely for a boy or a girl to be born).

14. In a lottery with 49 balls numbered 1–49, 6 numbers are selected at any draw. If someone selects six numbers in advance, the probability that all six will be drawn is $\binom{49}{6}^{-1}$.

15. The number of characters in a page that are misprinted has the Poisson distribution with parameter 2. Then, the probability of having at least three misprinted characters on any particular page equals $5\mathrm{e}^{-2}$.

16. The number of house burglaries in a city during a day follows the Poisson distribution. If the probability that there will be no burglaries on a particular day is e^{-5}, the expected value of burglaries for a day is 5.

17. When we throw five fair coins, the probability that three or more heads will appear is 0.5.

18. Claims arrive at an insurance company according to a Poisson process $\{N(t) : t \geq 0\}$ with rate $\lambda = 3$ claims per day. The probability that there will be exactly one claim in a two-day period is e^{-6}.

5.8.2 Multiple Choice Questions

1. A box contains a red and b black balls. We select a ball randomly, note its color but do not return it to the urn. We repeat this experiment n times (where $n \leq \min\{a, b\}$) and let X be the number of red balls selected. The distribution of X is

   (a) binomial (b) negative binomial (c) hypergeometric

   (d) Poisson (e) none of the above

2. The number of customers served by a bank employee during an hour has the Poisson distribution with mean 10. Then, the probability that the employee serves exactly four customers in a particular one-hour interval is

   (a) $e^{-10}\dfrac{10^4}{4}$ (b) $e^{-10}\left(1 + \dfrac{10}{1!} + \dfrac{10^2}{2!} + \dfrac{10^3}{3!}\right)$ (c) $e^{-10}\dfrac{1}{4!}$

   (d) $e^{-10}\dfrac{10^4}{4!}$ (e) $e^{-10}\dfrac{4!}{10^4}$

3. We throw a fair die repeatedly until an outcome of 4 or greater appears for the fifth time. If X denotes the number of trials until this happens, then X follows the

   (a) Poisson distribution with parameter $\lambda = 1/2$

   (b) hypergeometric distribution

   (c) geometric distribution with parameter $p = 1/4$

   (d) geometric distribution with parameter $p = 4/6$

   (e) negative binomial distribution with parameters $r = 5$ and $p = 1/2$.

4. In a lottery with 49 numbered balls, 6 are selected on any particular draw. If we select 10 numbers before a certain draw, the probability that we have exactly 5 winning numbers is

   (a) $\dfrac{\binom{10}{5} \cdot 39}{\binom{49}{6}}$ (b) $\dfrac{5 \cdot 10}{\binom{49}{5}}$ (c) $\dfrac{\binom{10}{6} \cdot \binom{39}{1}}{\binom{49}{6}}$

   (d) $\dfrac{\binom{10}{5} \cdot 39 + \binom{10}{6} \cdot 39}{\binom{49}{6}}$ (e) $\dfrac{\binom{10}{5} \cdot \binom{39}{1}}{\binom{49}{5}}$

5. The probability that a students fail a certain exam has been estimated to be 0.05. The probability that at least three scripts will be needed until the professor who marks the scripts finds the *second script* to be a fail is

   (a) $1 - (0.05)^2(0.95)$ (b) $1 - (0.05)^2$ (c) $(0.05)^2 + (0.05)^2(0.95)$

   (d) $(0.95)^2$ (e) $(0.95)^2(0.05)$

6. At an industrial unit which produces TFT 22″ computer screens, the probability that a new screen is defective is 0.05. In a sample of 60 screens, the probability that *at most one* defective screen is included equals

   (a) 0.191 55 (b) 0.046 07 (c) 0.145 48 (d) 0.225 88 (e) 0.002 42

7. We throw a fair die repeatedly until an outcome of either 3 or 4 appears for the *fourth time*. If X denotes the number of trials until this happens, then which of the following statements is correct?

   (a) $E(X) = 12$ (b) $E(X) = 3$ (c) $P(X = 4) = (2/3)^4$
   (d) $P(X = 3) = (1/3)^4$ (e) $P(X = 5) = 2^6/3^5$

8. Apples are packaged at a fruit packaging unit in boxes with 20 fruits each. It is known that 5% of the apples that arrive at the packaging unit cannot be sold, and for this purpose each apple is checked before packaging and put in a box only if found suitable for sale. The distribution of the number of apples that are needed to be checked in order to fill a box is

   (a) Poisson (b) binomial (c) hypergeometric
   (d) negative binomial (e) none of the above

9. During the last 10 years, 18 fatal accidents have been recorded on a certain motorway. The distribution of the number of accidents in this motorway per year can be reasonably assumed to be

   (a) Poisson (b) binomial (c) hypergeometric
   (d) negative binomial (e) none of the above

10. The percentage of companies that had their stock price increased at the NY Stock Exchange on a particular day is 45%. If we take a sample of 15 companies, the number of those who *did not* have an increase on that day has a distribution which is

    (a) $\mathrm{Nb}(15, 0.45)$ (b) $\mathcal{P}(6.75)$ (c) $G(0.45)$
    (d) $b(15, 0.55)$ (e) $b(15, 0.45)$

11. Faye asks George to select a nonnegative integer less than 1000. Assuming that George selects this number completely at random, the probability that this number has at least one digit equal to 2 is

    (a) $1 - (9/10)^3$ (b) $1 - (8/10)^3$ (c) $8^3/10^3$
    (d) $9^3/10^3$ (e) $1 - (8^3 + 9^3)/10.$

12. An electric wire of length 30 m has on average 0.5 faults. Assuming that the number of faults follows a Poisson process, the probability that there are exactly two faults in a wire which is 60 m long is

    (a) $2\mathrm{e}^{-2}$ (b) e^{-1} (c) $\mathrm{e}^{-1}/2$ (d) e^{-2} (e) $\mathrm{e}^{-2}/2$

13. The number, X, of births at a hospital during an hour has the Poisson distribution with parameter λ. We know that the probability a single birth occurs in an hour is four times the probability of having two births in an hour. The value of the parameter λ is

    (a) $1/4$ (b) $1/2$ (c) 2 (d) e^{-2} (e) 4

14. For a family with n children ($n \geq 2$), the probability that there exists at least one boy and at least one girl is

    (a) $1 - \dfrac{1}{2^{n-1}}$ (b) $1 - \dfrac{1}{2^n}$ (c) $\dfrac{1}{2^{n-1}}$ (d) $\dfrac{1}{2^n}$ (e) $\dfrac{1}{2^{n-1}} + \dfrac{1}{2^n}$

15. From an urn that contains 60 balls numbered 1–60, we select 3 balls with replacement. Let X be the random variable which denotes the number of balls selected, having a number which is a multiple of 3. The distribution of X is

(a) hypergeometric (b) binomial (c) Poisson
(d) negative binomial (e) none of the above

16. If X has a $\mathcal{P}(\lambda)$ distribution and we know that the standard deviation of X is three times higher than its mean, then the value of λ is

(a) $1/9$ (b) $1/3$ (c) 1 (d) 3 (e) 9

17. The number of cars arriving at a parking station follows a Poisson process $\{N(t) : t \geq 0\}$ with rate λ cars for a five-minute period. Suppose that the variance of the number of cars to arrive in the next 10 minutes is 9. Then the value of λ is

(a) 18 (b) $9/2$ (c) $9/10$ (d) $9/4$ (e) $3/2$

18. The number of cars arriving at a gas station follows a Poisson process $\{N(t) : t \geq 0\}$ with rate $\lambda = 7$ cars for a five-minute period. The expected number of cars to arrive between 9:00 a.m. and 10:00 a.m. on a particular day, if we know that six cars arrived between 9:00 a.m. and 9.15 a.m., equals

(a) 63 (b) 78 (c) 90 (d) 84 (e) 69

19. Assume that X has a $b(n, p)$ distribution. If it is known that the mean of X is 6 and its standard deviation is 2, then the values of n and p are

(a) $n = 18, p = 2/3$ (b) $n = 9, p = 2/3$ (c) $n = 18, p = 1/3$
(d) $n = 12, p = 2/3$ (e) $n = 9, p = 1/3$

5.9 REVIEW PROBLEMS

1. Susan throws three dice and Andrew also throws three dice. What is the probability that they get the same number of sixes?

2. The probability of error in the transmission of a binary digit over a communication channel is $2 \cdot 10^{-4}$. If 5000 digits are transmitted,

(i) find the probability that at least two errors occur;
(ii) calculate the mean and the variance of the number of errors.

3. The price of a certain stock either increases during a trading day by \$1 and this happens with probability 0.7 or decreases by \$1. What is the probability that, after 20 trading days, the price of the stock will be

(i) the same as it was before the first day of trading;
(ii) \$4 higher than its value before the first day of trading?

4. We ask five persons which day of the week they were born. Find the probability that

(i) no one was born on a Tuesday;
(ii) at least three persons were born during a weekend;
(iii) not all five were born on the same day of the week.

5. A bus traveling between Paris and Lyon in France is about to depart in five minutes. The bus has 52 seats and they have all been booked. Anne and Sophie have just arrived at the bus station without a reservation, hoping that there will be two empty seats. If the probability that a person who has booked will not turn up for traveling is 0.04, what is the probability that both girls will find a seat in the bus?

6. Screws produced by a machine in a large company are packaged in packs of 50 screws each. The company believes that a pack cannot be sold if it contains at least three defective screws. If it is known that the probability of a screw being defective is 0.03,

   (i) find the probability that a pack cannot be sold;

   (ii) to decide whether a pack should be sold, an employee checks five screws at random and decides that a pack should not be sold if it contains at least one defective screw. Find the probability that a *rejected pack* contains less than three defective screws, so that it could have been sold.

7. A box contains five red and six yellow balls. We select four balls at random.

   (i) What is the probability that at least three are red? Compare the answers you get if the selection is made

   - without replacement;
   - with replacement.

   (ii) From each of the two sampling strategies in (i), find the expected number of red balls in the sample of size 4. Again, compare the results if sampling is made with or without replacement.

8. Susan rolls two dice and Adam rolls three dice. What is the probability

   (i) that they get the same number of sixes?

   (ii) that Susan gets more sixes than Adam?

   (iii) that Adam gets exactly one more six than Susan?

9. Daniel goes with his father to an amusement park and heads straight for the shooting game. He pays \$3 to enter the game and he is offered five shots; if he finds the target in at least four of them, he wins a prize which his dad estimates to be worth \$6. Assuming that the probability of hitting the target in each shot is p, find Daniel's (or his dad's!) expected profit if

   (i) $p = 0.7$;

   (ii) $p = 0.9$.

10. John rolls a die, and if the outcome is k, he rolls the dice k times successively.

    (i) What is the sample space for this experiment?

    (ii) Write down the probability function for the number of trials (rolls of the die).

    (iii) Let Y be the number of 3's which turn up. Find the probability function of Y.

11. Consider the discrete uniform distribution, $\mathcal{DU}(n)$, defined in Section 5.7. For this distribution,

 (i) calculate the first three moments around zero, that is, μ_i' for $i = 1, 2, 3$;

 (ii) show that the third moment around the mean is always zero;

 (iii) verify that the variance of the random variable having this distribution is

$$\frac{(n-1)(n+1)}{12}.$$

12. Six school girls discuss about their color preferences. They decide to write in a piece of paper their two favorite colors among a possible set of seven colors: pink, purple, yellow, maroon, red, green, and brown.

 If we assume that all colors are equally likely to be selected, what is the probability that

 (i) all six girls select maroon as one of their choices?

 (ii) exactly five girls select pink and maroon as their choices?

 (iii) the total number of colors the girls wrote down is five, so that two colors are not selected by anybody?

13. In Section 5.2, we showed that the geometric distribution has the memoryless property, i.e. for any positive integers n, k, we have

$$P(Y > n + k | Y > n) = P(Y > k).$$

 Verify that the converse is true; that is, if a random variable Y on the nonnegative integers satisfies the above for all n and k, then the distribution of Y is geometric. This type of a result, meaning a two-way implication, is called a *characterization of the distribution*.

14. Reservations for a theater performance are made according to a Poisson process with rate $\lambda = 5$ reservations per hour. If there are currently 20 seats available, what is the probability that all the seats will have been booked within the next three hours?

15. The number of molecules that a gas contains can be described by a Poisson process with rate λ, so that in an area of w units the expected number of molecules is λw. Find the probability that a specific molecule

 (i) has a distance from its closest molecule less than a fixed positive number α;

 (ii) has a distance from its *second closest molecule* greater than α.

16. Nicky, who is a wine taster, tries five California wines and six Bordeaux wines and she is asked to assign a number to each wine, according to her preference, so that the best wine receives number one, the one she likes after that receives number two, and so on.

(i) Let Y be the number assigned to the *Bordeaux wine* that Nicky likes best. Find the probability function of Y, assuming that initially all wines have the same probability of being selected in any order.

(ii) Let now Y' be the number assigned to Nicky's *second most favorite* Bordeaux wine. Find the probability function of Y'.

17. The number of phone calls that Nicky receives on her mobile phone follows a Poisson process with a rate $\lambda = 3$ calls per hour.

(i) Find the probability that she will receive three calls between 6:00 p.m. and 6:20 p.m.

(ii) Calculate the probability that she will receive three calls between 6:40 p.m. and 7:00 p.m. *and* at least two calls between 7:10 p.m. and 7:30 p.m. on the same day.

(iii) Find the probability that she will receive three calls between 6:40 p.m. and 7:00 p.m. *and* at least six calls between 6:00 p.m. and 7:40 p.m. on the same day.

18. An employee in a telephone sales company estimates that she has a 15% chance of closing a sale after each phone call.

(i) What is the expected number of sales she will make after 40 calls?

(ii) What is the number of calls needed so that the probability of at least one sale is 95% or more?

19. George and Faye enter a quiz in which they have to answer independently 10 "True or False" questions.

(i) Assuming that they are both purely guessing, so that both have a probability of $1/2$ of answering each question correctly, show that the probability they give the same number of correct answers is

$$\binom{20}{10} 2^{-20}.$$

(*Hint*: you may find a combinatorial identity useful here.)

(ii) What is the probability that Faye gives two more correct answers than George?

20. Let X be a random variable with the negative binomial distribution with parameters r and p.

(i) Let $Y = X - r$. Find the probability function, the expectation, and variance of Y.

(ii) Assume that, as $r \to \infty$, the probability of a failure $q = 1 - p$ converges to zero (so that $p \to 1$) in such a way that the product rq converges to a positive constant λ. Show that

$$\lim_{y \to \infty} P(Y = y) = e^{-\lambda} \frac{\lambda^y}{y!}, \quad y = 0, 1, 2, \ldots$$

21. When Lucy goes shopping, the number of items she buys follows a Poisson distribution with rate $\lambda_1 = 3$. However, if her friend Irene, who is a shopaholic, goes with her, then she tends to buy more items, so that the number of purchased items still has a Poisson distribution but with a new rate $\lambda_2 = 4.5$. Lucy is going to the shops next Saturday and she has invited Irene to come with her, but she is not sure if she will be coming. If the probability that Irene accompanies Lucy to the shops is 0.7, find

(i) the probability that Lucy buys at least five items;

(ii) the expected number of items she will buy.

22. In Exercise 14 of Section 5.1 and in Exercise 12 of Section 5.4, you were asked to show that the probability function of both the binomial and the Poisson distribution satisfies the recursive formula

$$f(x) = \left(a + \frac{b}{x}\right) f(x-1). \tag{5.18}$$

In the first case, the formula holds for $x = 1, 2, \ldots, n$, in the latter for $x = 1, 2, \ldots,$ i.e. for all positive integers.

Show that this recursive relationship also holds for $x = 1, 2, \ldots,$ in the case of the negative binomial distribution and identify the choices of a and b for this case.

23. Let X have a distribution on the nonnegative integers $\{0, 1, 2, \ldots\}$ such that (5.18) holds.

(i) Assuming that $E(X)$ exists, show that

$$E(X) = \frac{a+b}{1-a}.$$

(ii) If we assume that $E[X(X-1)]$ exists, then it is given by

$$E[X(X-1)] = \frac{(2a+b)(a+b)}{(1-a)^2}.$$

From this, deduce an expression for the variance of X in terms of a and b.

(iii) Combine the results of the first two parts with those of the previous exercise to reestablish the mean and the variance of the Poisson and the negative binomial distributions.
(*Hint*: Use the identities

$$xf(x) = a(x-1)f(x-1) + (a+b)f(x-1)$$

and

$$x(x-1)f(x) = a(x-1)(x-2)f(x-1) + (2a+b)(x-1)f(x-1)$$

and sum the two parts of the first identity for $x = 1, 2, \ldots,$ and of the second identity for $x = 2, 3, \ldots$)

24. Customers arrive at a bookstore with a rate of $\lambda = 20$ customers per hour. It is estimated that 30% of the customers buy a particular book which is this month's best-seller. What is the probability that the store sells at least four copies of the book within an hour?

25. From an urn, which contains a white and b black balls, we select n balls with replacement. However, each time a ball is put back into the urn, we also put s more balls of the same color into the urn. Let X be the number of white balls in the sample of n balls selected.

    (i) Show that the probability function of X is given by the formula

$$f(x) = P(X = x) = \binom{n}{x} \frac{\prod_{i=1}^{x}(a + (i-1)s) \prod_{i=1}^{n-x}(b + (i-1)s)}{\prod_{i=1}^{n}(a + b + (i-1)s)}.$$

    The distribution of the random variable X is often referred to as the **Pólya distribution**, and is in fact one of several distributions named after the Hungarian mathematician George Pólya (1887–1985).

    (ii) Which distribution arises in the special case $s = 0$?

    (iii) Suppose now that $s = -1$ so that, rather than putting more balls, *we remove one ball* from the urn, which is of the same color as the last ball chosen. What is the distribution of X in this case?

26. A bowl contains a blue and b yellow chips. From the bowl, we select successively and without replacement, until the nth blue chip is selected. Let X be the number of chips selected until this happens. The distribution of X is called a **negative hypergeometric distribution**.

    (i) Prove that the probability function of X is given by

$$f(x) = P(X = x) = \frac{\binom{a}{n-1}\binom{b}{x-n}}{\binom{a+b}{x-1}} \cdot \frac{a-n+1}{a+b-n+1},$$

    where $x = n, n+1, \ldots, n+b$.

    (ii) Arguing as in the proof of Proposition 5.9, show that, as $a \to \infty, b \to \infty$ in a way such that

$$\lim_{a\to\infty, b\to\infty} \frac{a}{a+b} = p,$$

    the following holds true:

$$\lim_{a\to\infty, b\to\infty} f(x) = \binom{x-1}{n-1} p^n (1-p)^{x-n}, \quad x = n, n+1, \ldots$$

    (convergence of the negative hypergeometric distribution to the negative binomial). How do you interpret the last result intuitively?

Application: We have 25 male and 35 female students and we want to select a committee that has exactly 3 female members. For this reason, we select students at random (among the 60 students) and we stop when the third female student is selected. Let X be the number of committee members. Write down the probability function of X. Hence, calculate the expected number of committee members.

27. An electronic game machine makes a profit of c dollars per hour of service. The machine, however, breaks down at random times. In particular, it is assumed that the number of breakdowns follows a Poisson process with rate $\lambda > 0$. Moreover, if the number of breakdowns equals x, the financial damage incurred due to the repair costs and the fact that the machine is not working equals $ax^2 + bx$, where a and b are known positive constants.

   (i) Write down a formula for the profit K from the operation of the machine during the time interval $[0, t]$, if time is measured in hours.

   (ii) Show that the expected profit, $E(K)$, is given by

$$E(K) = ct - (a + b)\lambda t - a\lambda^2 t^2.$$

   (iii) Find the time t for which the expected profit is maximized from the operation of the machine during the interval $[0, t]$.

28. A scientific institute has n_1 male and n_2 female members. In the last elections of the institute, n members did not vote ($n < n_1 + n_2$). If all institute members have the same probability of abstaining from voting, find

   (i) the distribution of the number of female members who did vote;

   (ii) the probability that the number of members who voted is less than double the number of female members who voted in the elections.

29. A new law proposal is submitted to the parliament of a country. There are n members of the government party, who will all certainly vote in favor of the proposal, m members of the leading opposition party who will all vote against the proposal, while there are also k parliament members who belong to other parties; we assume that $m < n, k < n$, but $m + k > n$. In the last group of k parliament members, each person will vote in favor of the proposal with a probability p, and against it with a probability $q = 1 - p$, and we assume that they reach a decision independently of one another.

   (i) Find the probability function for the number X of votes that will be cast in favor of the law proposal.

   (ii) Find the probability that the proposal is approved by the parliament.

30. While information is transmitted to a computer system, a so-called parity check takes place as follows: suppose that the system uses words with n binary digits (bits), i.e. each digit is either 0 or 1. Then, next to each word which is transmitted, another parity digit is placed such that the total number of 1's representing a word is an even integer. When the word with the $n + 1$ digits is transmitted to the receiver, this receiver checks the number of 1's received and if this is not an even number,

it reports an error in the data transportation. Suppose that each digit is transmitted correctly with probability p $(0 < p < 1)$.

(i) Find the distribution of the number X of digits that are correctly transmitted.

(ii) What is the probability that a word which is not transmitted correctly passes unnoticed?

(iii) Give numerical answers to the previous two parts for $p = 0.98$ when $n = 3$ and $n = 6$.

31. Let X be a variable that follows the Poisson distribution with parameter $\lambda > 0$. If r is an integer with $r \geq 2$, define the random variable

$$Y = \prod_{i=1}^{r-1}(X + i) = (X + 1)(X + 2) \cdots (X + r - 1).$$

(i) Prove that

$$E\left(\frac{1}{Y}\right) = \frac{1}{\lambda^{r-1}}\left(1 - \sum_{i=0}^{r-2} \mathrm{e}^{-\lambda}\frac{\lambda^i}{i!}\right).$$

(ii) Calculate the expectation

$$E\left(\frac{1}{X+1}\right)$$

and compare this with the quantity

$$\frac{1}{E(X) + 1}.$$

32. Let X_n be the number of successes in n independent Bernoulli trials, each with a success probability p, and

$$a_n = P(X_n \text{ is an even integer}).$$

(i) Using the law of total probability and the partition $\{B_1, B_2\}$ with $B_1 = \{X_1 = 0\}$, and $B_2 = \{X_1 = 1\}$, prove that the following recursion holds:

$$a_n = (1 - p)a_{n-1} + p(1 - a_{n-1}), \quad n = 2, 3, \ldots$$

(ii) Show by induction that the numbers a_n, $n = 1, 2, \ldots$, are given by

$$a_n = \frac{1}{2}[1 + (1 - 2p)^n], \quad n = 1, 2, \ldots$$

(iii) Using the result from (ii), deduce the combinatorial identity

$$\sum_{x=0}^{[n/2]} \binom{n}{2x} p^{2x}(1 - p)^{n-2x} = \frac{1}{2}[1 + (1 - 2p)^n].$$

33. In a sequence of $n + k$ independent Bernoulli trials with success probability p in each trial, let X be the number of successes in the first n trials and Y be the total number of successes out of the $n + k$ trials. Establish that for $0 \le x \le y$, $0 \le x \le n$, and $0 \le y \le n + k$, the following expressions are true:

(i) $P(X = x,\ Y = y) = \binom{n}{x}\binom{k}{y-x} p^y q^{n+k-y}.$

(ii) $P(Y = y \mid X = x) = \binom{k}{y-x} p^{y-x} q^{k-y+x}.$

(iii) $P(X = x \mid Y = y) = \dfrac{\binom{n}{x}\binom{k}{y-x}}{\binom{n+k}{y}}.$

How do you interpret the last result intuitively?

5.10 APPLICATIONS

5.10.1 Overbooking

Overbooking is a common practice in the travel industry, especially in airlines, trains, and ships. This practice is applied when sellers (air carriers, railway, and maritime companies) expect that some passengers will not show up. This allows them to have a fuller or nearly full plane/train/ship even if some passengers miss the trip, cancel their ticket at the last minute (which is quite common in the case of flexible tickets, which are rebookable), or do not show up at all.

Excessive overbooking may result in increased cost for the carrier because, in the event of not being able to accommodate all the passengers who show up, a search for volunteers to give up their seats will be necessary, at the expense of offering them a travel credit or free ticket.

In view of the above, the carriers would wish to calculate the amount of overbooking that can secure (at a certain probability level, say 90% or 95% or 99%) that all passengers who show up will be able to travel.

Let us now examine a simple probability model that can facilitate the above requirement. Assume that an air carrier has estimated (from past experience) that a percentage $p = 5\%$ of the passengers for a specific flight does not show up. This flight is serviced by a small plane with seat capacity $c = 50$. Let us assume that the air carrier sells $n = 52$ tickets. If we denote by X the number of passengers who do not show up to travel (out of the 52 ticket holders), it is clear that X follows a binomial distribution with parameters $n = 52$ and $p = 0.05$. Therefore,

$$P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x} = \binom{52}{x} 0.05^x\ 0.95^{52-x}, \quad x = 0, 1, \ldots, 52,$$

and the probability that all ticket holders who showed up will be able to travel is given by

$$P(X \ge n - c) = 1 - P(X < n - c) = 1 - P(X < 2) = 1 - P(X = 0) - P(X = 1).$$

Replacing $P(X = 0)$ and $P(X = 1)$ by

$$P(X = 0) = \binom{52}{0} 0.05^0 \;\; 0.95^{52} = 0.95^{52},$$
$$P(X = 1) = \binom{52}{1} 0.05^1 \;\; 0.95^{51} = 52 \cdot 0.05 \cdot 0.95^{51},$$

we obtain $P(X \geq 2) = 0.74$, that is, with probability 74%, the overbooking will not cause any problem to the air carrier.

A question that is of special interest for the air carrier is the following: for a given airplane capacity c and passenger non-show-up probability p, how many tickets can be sold so that the probability of being able to accommodate on the plane all the passengers who will turn up is not less than a desired level, say p_0. Evidently, the condition set in the previous question leads to the inequality

$$P(X \geq n - c) \geq p_0 \iff \sum_{x=n-c}^{n} \binom{n}{x} p^x (1-p)^{n-x} \geq p_0$$

and since the unknown quantity is the positive integer n, we can only solve the problem by tabulating the LHS of the last inequality for $n = c, c + 1, \ldots$ and look for the largest value

	$c = 50$		$c = 150$		$c = 345$	
p	n	Probability	n	Probability	n	Probability
0.1	50	1.000 000	152	0.999 998	355	1.000 000
	51	0.995 362	153	0.999 984	360	0.999 988
	52	0.971 706	157	0.996 536	369	0.992 853
	53	0.910 201	158	0.991 517	370	0.988 367
	54	0.801 542	159	0.981 606	375	0.919 387
	55	0.654 852	160	0.964 095	380	0.720 272
0.05	50	1.000 000	150	1.000 000	347	1.000 000
	51	0.926 902	151	0.999 567	348	0.999 997
	52	0.740 503	152	0.996 30	349	0.999 98
	53	0.498 184	153	0.983 883	350	0.999 905
	54	0.284 136	154	0.952 222	355	0.984 424
	55	0.139 653	155	0.891 273	360	0.798 631
0.01	50	1.000 000	150	1.000 000	345	1.000 000
	51	0.401 044	151	0.780 763	346	0.969 113
	52	0.095 576	152	0.449 714	347	0.862 243
	53	0.016 155	153	0.198 118	348	0.676 825
	54	0.002 124	154	0.069 803	349	0.461 740
	55	0.000 230	155	0.020 402	350	0.274 078

Table 5.3 Probability that all passengers can be accommodated in an airplane of capacity c when an overbooking of n tickets has been made.

of n that corresponds to a probability greater than or equal to p_0. In Table 5.3 we provide the values of the quantities

$$P(n) = P(n; c, p) = \sum_{x=n-c}^{n} \binom{n}{x} p^x (1-p)^{n-x}, \quad n \geq c,$$

for $p = 0.1, 0.05, 0.01$ and capacities $c = 50, 150, 345$ (corresponding to typical capacities of small, medium, and large airplanes, respectively). Note that for $n = c$, we always have $P(c) = 1$ while for $c = 345$ and $p = 0.1$, we see that the company can sell 10 more tickets than the available seats and be safe that all passengers who will show up for the flight will have a seat (more precisely, this event has probability one, with accuracy of 6 decimal places).

Further, according to Table 5.3, if the non-show-up probability is 10% and we wish to guarantee a problem-free overbooking at level 99%, we can sell up to 51 tickets for a flight with seat capacity $c = 50$, up to 158 for a flight with $c = 150$, and up to 369 for a flight with $c = 345$.

It is worth noting that the overbooking potential decreases dramatically when the non-show up probability decreases. For example, when $p = 0.01$ and $p_0 = 0.95$, no overbooking should be made in airplanes with a small or medium capacity while a safe overbooking, at this level, for a large airplane ($c = 345$) can be maintained by selling just one extra ticket.

KEY TERMS

Banach matchbox problem
Bernoulli distribution
Bernoulli trial
binomial distribution
geometric distribution
hypergeometric distribution
memoryless (or lack of memory) property
negative binomial distribution
Pascal distribution
Poisson distribution
Poisson process
success/failure
waiting time distribution

CHAPTER 6

CONTINUOUS RANDOM VARIABLES

Pierre Simon Marquis de Laplace
(Beaumont-en-Auge, Normandy 1749–Paris 1827)

French mathematician and astronomer. Apart from making key contributions to the early development of probability theory, he carried out pioneering work on calculus, but perhaps is known most of all for his ingenuous ideas in celestial mechanics. His five-volume work *Traité de mécanique céleste* (published between 1799 and 1825) not only included everything known at his time about the movement of planets, but presented his own new methods for calculating the motions of the planets, determining the average distance of the planets from the earth, resolving tidal problems, and so on.

One of Laplace's major contributions to probability theory was the proof, in 1810, of the Central Limit Theorem, a celebrated result that will be discussed in Volume II of the present book. Among other things, this consolidated the importance of the normal distribution as a theoretical tool in probability and statistics.

In 1812, he introduced a multitude of new ideas and mathematical techniques in his book *Théorie Analytique des Probabilités*. Before Laplace, probability theory was solely concerned with developing a mathematical analysis for problems related to games of chance. Laplace applied probabilistic ideas to many scientific and practical problems. Astronomy, statistical mechanics, and actuarial mathematics are examples for some of the important applications of probability theory developed in the nineteenth century.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

6.1 DENSITY FUNCTIONS

This chapter and the next one deal with continuous random variables. Recall that a random variable X (and the associated distribution function) is said to be continuous if the cumulative distribution function

$$F(t) = P(X \leq t), \quad -\infty < t < \infty,$$

is continuous everywhere. In this case, the range of values of X, denoted by R_X, is uncountable. Typically, and nearly in every case we shall consider here, R_X will be a finite or infinite interval on the real line. An important observation we made in Chapter 4 is that, both for discrete and continuous variables, the distribution function $F(t)$ of X maps a real number t to the probability of the event $\{X \leq t\}$. Two important properties of F are that it is increasing and right-continuous everywhere. When X is discrete and has probability function f and range $\{x_1, x_2, \ldots\}$, then F is related to f by

$$F(t) = \sum_{i:x_i \leq t} f(x_i). \tag{6.1}$$

It is clear that even a small change in the value of t can have a dramatic impact on $F(t)$. This is due to the upward jumps at the points x_i, because $\{X = x_i\}$ occurs with positive probability. As we have discussed in Chapter 4, however, when X is continuous, then $P(X = x) = 0$ for all real x. The distribution function $F(t)$ increases in a smooth way (formally, it is a continuous function) and is free of jumps.

It should be obvious from the above that the probability function, which is a major tool for discrete variables, cannot be used for continuous ones. What should one use instead? Let us try to find an answer to that question with the aid of an example. In the discrete case, we have seen examples where we ask a computer to generate digits in a "random" manner sequentially so that, at each place, any of the digits $0, 1, \ldots, 9$ appears with equal probability $1/10$. We assume for convenience that these digits occur as decimal places (in the order that they were obtained) of a number whose integer part is zero. As a result, for the first two digits, there are 10^2 possible selections, namely, the numbers $0.00, 0.01, \ldots, 0.99$, and each of them occurs with probability $1/100$. Similarly, if we ask the computer to select 10 digits, we have 10^{10} possible numbers between 0 and 1, each with 10 decimal places and any of these numbers occurs with probability 10^{-10}.

To illustrate the situation better, consider now the case where at each point, the digits $0, 1, \ldots, 9$ do not have the same probability of being selected, but obviously the sum of their probabilities is one. Then imagine the experiment of selecting, say, two digits. The possible values for this experiment are still $0.00, 0.01, \ldots, 0.99$, with their probabilities adding to one, but these probabilities are not equal any more (see Figure 6.1). As there are 10^2 outcomes, the distance between any adjacent outcomes on the x-axis is $10^{-2} = 0.01$. If we decide to continue the experiment, so that 10 digits are selected, there are 10^{10} outcomes, and the graph of the probability function will still look like Figure 6.1, but now there are 10^{10} possible outcomes and the distance between two adjacent points is 10^{-10}. In the case where we continue the experiment further, so that 40 digits are selected, we have 10^{40} possible outcomes with the distance between any two consecutive points on the

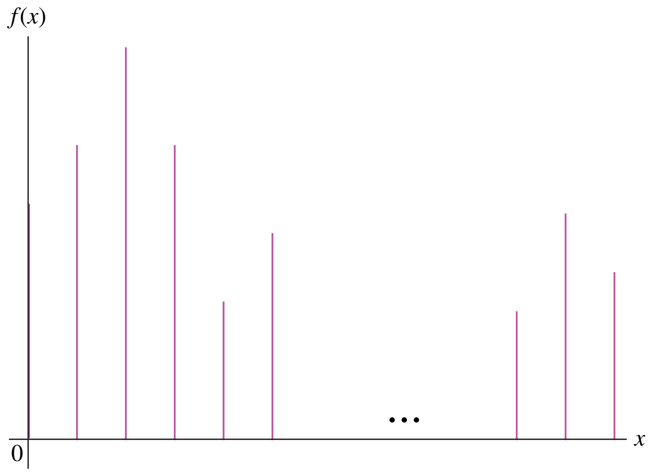


Figure 6.1 The probability function of a discrete random variable.

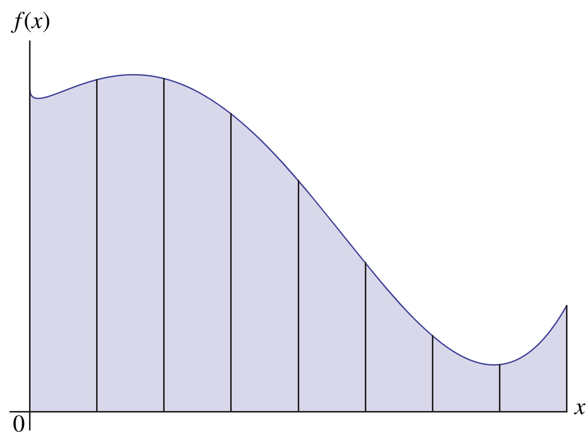


Figure 6.2 The density function of a continuous random variable.

x-axis decreasing to 10^{-40}, and so on. We thus see that, as the number of digits increases, the numbers which can be produced as outcomes of this experiment "fill in" the interval $[0, 1]$ on the x-axis, as they get closer and closer to one another.

The passage from the discrete to the continuous case is achieved by letting this experiment to carry on *indefinitely*. In such a (theoretical) experiment, where we anticipate an infinite number of decimal digits, any real number between 0 and 1 is possible and so, if the random variable X denotes the outcome of the experiment, the range of values for X is $R_X = [0, 1]$, which is an uncountable set. The situation is now illustrated graphically in Figure 6.2. As the number of possible outcomes tends to infinity, the distance between any two adjacent points tends to zero so that, on the graph, instead of a series of segments vertical to the x-axis, we now have a "smooth" function $f(x)$ that takes nonzero values for any $x \in [0, 1]$. Such a function is called a **density function** (associated with the continuous random variable X).

The formal definition of this concept is as follows:

Definition 6.1 Let X be a random variable. Assume that there exists a nonnegative real function $f : \mathbb{R} \mapsto [0, \infty)$ such that for any subset A of $\mathbb{R}$ which can be written as the union of a finite or infinite number of intervals, we have

$$P(X \in A) = \int_A f(x)\mathrm{d}x. \tag{6.2}$$

Then, we say that the random variable X is **absolutely continuous** and it has (probability) **density function** (or, simply **density**) f.

The term *absolutely continuous* is new, but we shall not use it any further. In more advanced textbooks, there is a distinction between continuous and absolutely continuous random variables, because random variables having a continuous distribution function can be studied under more general conditions than (6.2) given in the above definition. However, this is beyond the scope of the present book and, from now on, a variable X for which there exists a density f satisfying (6.2) will simply be referred to as a **continuous random variable**.

Let us now look at some consequences of (6.2). First, taking $A = (-\infty, t]$ there, we obtain

$$F(t) = P(X \leq t) = P(X \in A) = \int_A f(x)\mathrm{d}x = \int_{-\infty}^{t} f(x)\mathrm{d}x. \tag{6.3}$$

This is the continuous analog of (6.1) (see also the discussion following Proposition 4.4), which is valid for a discrete random variable.

A glance at (6.1) and the last equation gives us a general rule-of-thumb for calculating probabilities associated with a discrete or a continuous random variable:

- Whenever we use sums for discrete variables, we use integrals for continuous variables.

Thus, most of the relations that we have seen in Chapter 4 for a discrete variable are also valid in the continuous case provided we replace the sums involved by integrals.

To give a first example, property (PF3) in Proposition 4.3 for the probability function f which tells us that

$$\sum_{x \in R_X} f(x) = 1$$

is now replaced by

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$$

where, in the last expression, f denotes the density function. This follows immediately if we take $A = (-\infty, \infty)$ either in (6.2) or in (6.3); in the latter case, we also need that $\lim_{t\to\infty} F(t) = 1$, which holds for any distribution function (see Proposition 4.1), so that

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \lim_{t\to\infty} \int_{-\infty}^{t} f(x)\mathrm{d}x = \lim_{t\to\infty} F(t) = 1.$$

In fact, the properties

DF1. $f(x) \geq 0$ for all $x \in \mathbb{R}$,

DF2. $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$,

are characteristic properties of a density function for a continuous random variable, in the sense that if and only if these properties hold simultaneously, there exists a random variable X having density f.

Note that a density function need not be continuous. We merely require that it is an integrable function over $(-\infty, \infty)$. If, however, f is continuous (as will be in most applications) then, since f is related with the distribution function F via (6.3), we know from calculus that this ensures that F is differentiable everywhere with derivative at $x \in \mathbb{R}$ given by

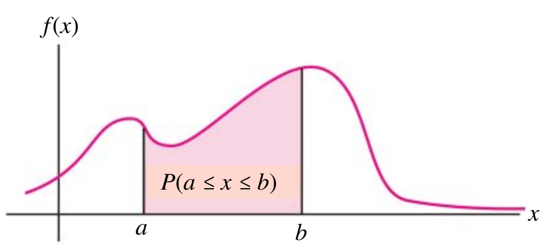
$$F'(x) = f(x).$$

Even if f is not everywhere continuous, (6.3) still implies that the last relationship holds at each continuity point of f.

Next, taking $A = [a, b]$ in (6.2), we get

$$\begin{aligned} P(a \leq X \leq b) = P(X \in A) &= \int_A f(x)\mathrm{d}x = \int_a^b f(x)\mathrm{d}x \\ &= \int_{-\infty}^b f(x)\mathrm{d}x - \int_{-\infty}^a f(x)\mathrm{d}x = F(b) - F(a). \end{aligned}$$

In view of the geometric interpretation for a definite integral, we thus see that the probability $P(a \leq X \leq b)$ is given by the area enclosed by the graph of the density function, the x-axis, and the lines $x = a$ and $x = b$ in a set of Cartesian coordinates (see the following figure).



A somewhat surprising result for a continuous random variable X, discussed already in Chapter 4, is that the probability for X to take any particular value equals zero. From the last relation, we see that this is consistent with the above interpretation of probabilities as areas on the plane; for $a = b$ in the last relation, we get

$$P(a \leq X \leq a) = P(X = a) = F(a) - F(a) = 0,$$

which is reflected geometrically in the fact that when a and b coincide, the relevant part of the graph for the density function has area zero.

An immediate consequence of the last result is that

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

This is in contrast with the discrete case, where only the relation $F(b) - F(a) = P(a < X \leq b)$ is valid in general (the other relations are only valid when there is no jump at either a or b); compare to Table 4.1.

The above results are summarized in the following proposition.

Proposition 6.1 *Let X be a continuous random variable having distribution function F and density function f. Then, the following hold:*

(a) $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$;

(b) $F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$;

(c) $f(x) = F'(x)$ *for each x at which f is continuous;*

(d) $P(a \leq X \leq b) = F(b) - F(a) = \int_{a}^{b} f(x)\mathrm{d}x$.

The last result is also valid if one, or both, of the $\leq$ signs inside the brackets on the left are replaced by $<$.

From the above proposition, it is apparent that, for specifying the distribution of a continuous random variable and for calculating probabilities associated with it, we need to know either the density function or the distribution function of that variable. In fact, if we know one of these two functions, calculating the other is now a rather more straightforward task than in the discrete case (compare with Proposition 4.4).

Example 6.1 The time required, in hours, to repair a car at a garage is a random variable X with density function

$$f(x) = c(4x - x^2), \quad 0 < x \leq 4.$$

(i) Find the value of the constant c.

(ii) Find the probability that for a car which arrives now at the garage, the amount of time needed to get repaired will be

(a) at least one but less than three hours;

(b) more than two hours.

SOLUTION

(i) For f given in the example to be a probability density function, conditions DF1 and DF2 should be satisfied, that is, $f(x) \geq 0$ for all x and $\displaystyle\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$.

The first condition is met when $c \geq 0$, because for $x \in (0, 4]$ the quantity $4x - x^2$ is nonnegative. The second condition gives

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^4 f(x)\mathrm{d}x = 1$$

(the first equality holds since the density is zero outside the interval $(0, 4]$). Replacing f by the specific form given in the statement of the example, we get that

$$\int_0^4 f(x)\mathrm{d}x = \int_0^4 c(4x - x^2)\mathrm{d}x = c\left(2x^2 - \frac{x^3}{3}\right)\Big|_0^4$$
$$= c\left(32 - \frac{64}{3}\right) = \frac{32c}{3}.$$

Since this must be equal to 1, we obtain that $c = 3/32$. Therefore, the density for the time to repair a car at the garage is

$$f(x) = \frac{3}{32}(4x - x^2) = \frac{12x - 3x^2}{32}, \quad 0 < x \leq 4.$$

(ii) (a) The probability that the car will need at least one but less than three hours to be repaired is given by $P(1 \leq X < 3)$. From Proposition 6.1, we have it as

$$P(1 \leq X < 3) = P(1 \leq X \leq 3) = \int_1^3 f(x)\mathrm{d}x.$$

Upon using the density function f, which was found in Part (i), we get

$$P(1 \leq X < 3) = \int_1^3 \frac{12x - 3x^2}{32}\mathrm{d}x$$

$$= \frac{1}{32}(6x^2 - x^3)\Big|_1^3 = \frac{1}{32}[(6 \cdot 3^2 - 3^3) - (6 \cdot 1^2 - 1^3)]$$

$$= \frac{22}{32} = \frac{11}{16}.$$

(b) The probability that the car will need more than two hours to be repaired is given by $P(X > 2)$. To emphasize the link with the distribution function, we deal with the probability of the complementary event, $P(X \leq 2)$ which, by an appeal to (6.3), equals

$$P(X \leq 2) = F(2) = \int_{-\infty}^2 f(x)\mathrm{d}x = \int_0^2 f(x)\mathrm{d}x,$$

because f is zero for negative argument. Substituting for f from Part (i), we get

$$P(X \leq 2) = \int_0^2 \frac{12x - 3x^2}{32} \mathrm{d}x = \frac{1}{32}(6x^2 - x^3)|_0^2$$
$$= \frac{6 \cdot 2^2 - 2^3}{32} = \frac{1}{2}.$$

Recall that this is the probability that X is at most 2, while we want the probability of the complementary event, which is

$$P(X > 2) = 1 - P(X \leq 2) = 1 - \frac{1}{2} = \frac{1}{2}.$$

Rather than dealing each time with the complementary event, in order to calculate probabilities of the form $P(X > t)$, as in Part (ii)b of the last example, a more convenient way is to make use of the following observation. Since for any continuous variable X and $t \in \mathbb{R}$, we have

$$P(X \leq t) = F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x,$$

changing the signs and adding one to either side, we obtain

$$1 - P(X \leq t) = 1 - F(t) = 1 - \int_{-\infty}^{t} f(x)\mathrm{d}x.$$

But $1 - P(X \leq t)$ equals $P(X > t)$, and so by using Part (b) of Proposition 6.1, we deduce that

$$P(X > t) = \int_{-\infty}^{\infty} f(x)\mathrm{d}x - \int_{-\infty}^{t} f(x)\mathrm{d}x = \int_{t}^{\infty} f(x)\mathrm{d}x.$$

This is a general result which holds for any continuous variable X having density function f, and any real t.

Example 6.2 The time, in minutes, between two consecutive arrivals of a train at a metro station is a continuous random variable X with distribution function

$$F(t) = c - \mathrm{e}^{-t/3} - \frac{t}{3} \cdot \mathrm{e}^{-t/3}, \quad t > 0,$$

for some constant c.

(i) Find the value of c and then obtain the density function f of X.

(ii) Suppose that Mary arrives at the platform of the metro station exactly at the time when a train departs. What is the probability that she has to wait until the arrival of the next train

(a) less than three minutes?

(b) more than one but less than four minutes?

SOLUTION

(i) In order to find the value of c, we use that the distribution function F must satisfy

$$\lim_{t\to\infty} F(t) = 1.$$

For the function F given in the statement, this gives

$$\lim_{t\to\infty}\left(c - e^{-t/3} - \frac{t}{3}\cdot e^{-t/3}\right) = 1. \tag{6.4}$$

It is clear that $\lim_{t\to\infty} e^{-t/3} = 0$, while for the last term inside the bracket we have

$$\lim_{t\to\infty}\frac{t}{3}\cdot e^{-t/3} = \lim_{t\to\infty}\frac{t}{3e^{t/3}} = 0,$$

by applying l'Hôpital's rule. Substituting these results into (6.4), we get immediately that $c = 1$. This means that the distribution function of X is

$$F(t) = 1 - e^{-t/3} - \frac{t}{3}\cdot e^{-t/3}, \quad t > 0 \tag{6.5}$$

and the density of X can now be obtained by differentiating this expression. This yields

$$f(x) = F'(x) = \frac{1}{3}e^{-x/3} - \left(\frac{1}{3}e^{-x/3} - \frac{x}{3}\cdot\frac{1}{3}e^{-x/3}\right) = \frac{x}{9}e^{-x/3}.$$

Figure 6.3 shows the distribution function and the density function obtained above.

(ii) The probabilities we seek for this part are

(a) $P(X < 3)$,

(b) $P(1 < X < 4)$.

For the former, we use Proposition 6.1(b) which gives, since X is continuous, that $P(X < 3) = F(3)$ and now putting $t = 3$ in (6.5) we obtain immediately

$$P(X < 3) = 1 - 2e^{-1} \cong 0.264.$$

Next, in order to find the probability $P(1 < X < 4)$, we first note this is the same as $P(1 \le X \le 4)$ and then use the formula in Part (d) of Proposition 6.1. In this way, we obtain by an appeal to (6.5) again,

$$\begin{aligned} P(1 < X < 4) &= F(4) - F(1) \\ &= \left(1 - e^{-4/3} - \frac{4}{3}\cdot e^{-4/3}\right) - \left(1 - e^{-1/3} - \frac{1}{3}\cdot e^{-1/3}\right) \\ &= \frac{4}{3}e^{-1/3} - \frac{7}{3}e^{-4/3} \cong 0.34. \end{aligned}$$

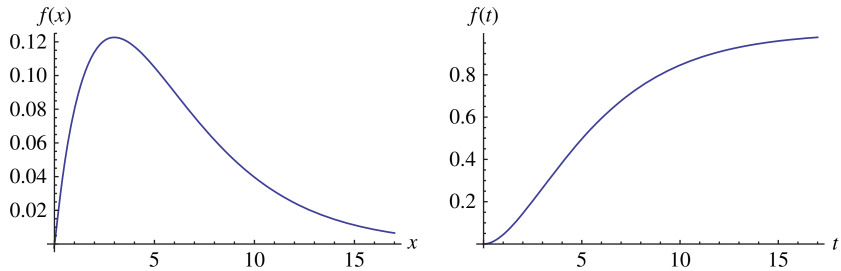


Figure 6.3 The density function and the distribution function of the variable X for Example 6.2.

Remark It should be clear that the probabilities for Part (ii) of the last example could also be obtained by using the density function f, found in Part (i). For instance, for Part (ii)b we find that

$$P(1 < X < 4) = F(4) - F(1) = \int_1^4 \frac{x}{9} e^{-x/3} dx.$$

Using integration by parts, we get

$$\begin{aligned} P(1 < X < 4) &= -3 \cdot \frac{x}{9} \cdot e^{-x/3} \Big|_1^4 + \int_1^4 \frac{3}{9} e^{-x/3} dx \\ &= \left(-\frac{4}{3} e^{-4/3} + \frac{1}{3} e^{-1/3} \right) + (-e^{-4/3} + e^{-1/3}) \\ &= \frac{4}{3} e^{-1/3} - \frac{7}{3} e^{-4/3} \cong 0.34, \end{aligned}$$

which obviously agrees with the result found above. However, we mention that once we have available the distribution function, it is usually easier and quicker to work with this rather than the density of a random variable, since in particular we avoid the integration.

For a discrete random variable X, we have seen in Chapter 4 that the probability function $f(x)$ expresses the probability that $X = x$. However, when X is continuous, $P(X = x) = 0$ for all real x. This clearly suggests that a density function does not have a direct interpretation as a function that assigns probabilities to an event. In fact, if f denotes a density, then $f(x)$ does not represent the probability of any event and, as we will see in several examples, it can take values greater than one! So, it is natural to wonder what is the interpretation of $f(x)$ in this case. Intuitively, and since $f(x)$ *is used* for the calculation of probabilities in (6.2) and (6.3), one expects that the higher the value of f at a point $x = a$, the more probable it should be that the random variable takes values "close to a."

Trying to get a better understanding of this, choose $\epsilon > 0$ to be small and, for a given a in the range of X, put $A = (a - \epsilon/2, a + \epsilon/2)$ in (6.2). This gives

$$P\left(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}\right) = \int_{a-\frac{\epsilon}{2}}^{a+\frac{\epsilon}{2}} f(x) dx.$$

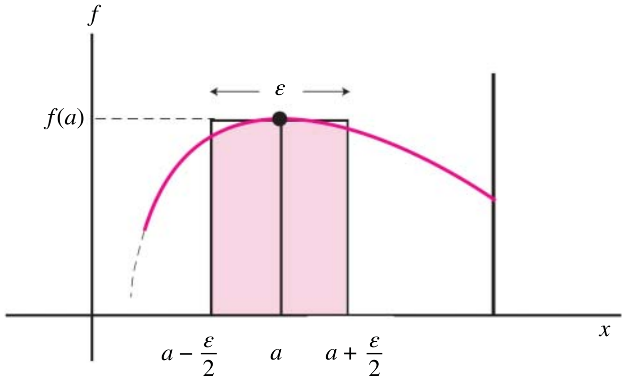


Figure 6.4 Approximation of $P\left(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}\right)$.

This is the shaded area on the graph in Figure 6.4. Since ϵ is assumed to be small, this area should be close to the area of a rectangle that has width ϵ and length $f(a)$, assuming that f is continuous at the point a. Thus,

$$P\left(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}\right) \cong \epsilon f(a). \tag{6.6}$$

Now, let b be another point in the range of X at which f is continuous. Then, we obtain in a similar manner that

$$P\left(b - \frac{\epsilon}{2} < X < b + \frac{\epsilon}{2}\right) \cong \epsilon f(b).$$

The last two expressions suggest that $f(b) \leq f(a)$ occurs if and only if

$$P\left(b - \frac{\epsilon}{2} < X < b + \frac{\epsilon}{2}\right) \leq P\left(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}\right).$$

Therefore, we see that the higher the value of the density f at a point a, the more probable it is for the variable X to be inside an interval of very small width centered at a.

An alternative interpretation is provided if we write (6.6) as follows:

$$f(a) \cong \frac{P\left(a - \frac{\epsilon}{2} < X < a + \frac{\epsilon}{2}\right)}{\epsilon}.$$

In words, this says that the value of a density at the point a is the ratio of the probability that X is contained in a "small" interval around a divided by the width of that interval. We stress again, however, that $f(a)$ itself **is not** a probability (and can take values larger than one).

Finally, if we use a different approximation,

$$f(a) \cong \frac{P(a < X < a + \epsilon)}{\epsilon} = \frac{F(a + \epsilon) - F(a)}{\epsilon},$$

which is justified since ϵ is small, we get an understanding of the fact that f is the derivative of F (see Part (c) of Proposition 6.1); taking the limit as $\epsilon \to 0$ above, we arrive at the derivative of F at the point a.

EXERCISES

Group A

1. In each of the following cases, identify the value of c so that the function f is a probability density (f vanishes outside the range of values indicated below).

   (i) $f(x) = 3c - 1, \quad 2 \le x \le 5;$

   (ii) $f(x) = c(1 - x^2), \quad 0 \le x \le 1;$

   (iii) $f(x) = \dfrac{cx^2}{3}, \quad 1 < x \le 3;$

   (iv) $f(x) = c\mathrm{e}^{-3x}, \quad 0 < x < \infty;$

   (v) $f(x) = \dfrac{c}{(1+x)^4}, \quad 0 < x < \infty.$

2. A continuous random variable X has density function

$$f(x) = \frac{c(3x+1)}{4}, \quad 1 \le x \le 4.$$

   (i) Find the value of c.

   (ii) Obtain the distribution function of X.

   (iii) Calculate the probabilities $P(X \le 2)$, $P(X > 2.5)$, $P(1.5 < X < 3)$.

3. The daily amount of time, in hours, that Nicky spends surfing on the internet is a random variable X with density function

$$f(x) = \begin{cases} \dfrac{3(x-2)^2}{16}, & 0 \le x \le 4, \\ 0, & \text{elsewhere.} \end{cases}$$

   Find the proportion of days on which she spends

   (i) less than two hours;

   (ii) between one and three hours.

4. The measurement error of a certain instrument is a continuous random variable X with density function

$$f(x) = \begin{cases} c(9 - x^2), & -3 \le x \le 3, \\ 0, & \text{elsewhere.} \end{cases}$$

   (i) Find the value of the constant c.

   (ii) Obtain the distribution function of X.

   (iii) Find the probability that the measurement error is

   (a) positive;

   (b) less than 1 in absolute value;

   (c) smaller than $-3/2$ or greater than $3/2$.

5. The total distance, in miles, that a taxi driver drives during a day is a continuous random variable with density function

$$f(x) = \begin{cases} \dfrac{x}{4800}, & 0 \le x \le 80, \\ -\dfrac{x}{2400} + \dfrac{1}{20}, & 80 < x \le 120, \\ 0, & \text{elsewhere.} \end{cases}$$

   (i) Verify that this function satisfies properties DF1 and DF2 so that it is indeed a valid probability density and sketch this function.

   (ii) Find the probability that the number of miles that he drives during a particular day is

      (a) more than 60;

      (b) between 60 and 100.

6. The borrowing period, in days, for a particular book at a University library can be regarded as a continuous random variable X with density function

$$f(x) = \begin{cases} 0.08x, & 0 < x \le 5, \\ 0, & \text{elsewhere.} \end{cases}$$

   (i) What is the maximum period allowed for borrowing this book?

   (ii) Calculate the probability that a new borrower keeps the book

      (a) for at least one day;

      (b) between two and three days;

      (c) exactly one or two or three days.

   (iii) If the borrower has not returned the book within two days since he took it from the library, what is the probability that the book will be returned during the third day?

7. Dr Smith finishes a particular lecture at the University between 2:58 p.m. and 3:04 p.m. The time X, in minutes after 2:58 p.m., that she finishes her class is a random variable with density function

$$f(x) = \begin{cases} \dfrac{x^2}{72}, & 0 < x \le 6, \\ 0, & \text{elsewhere.} \end{cases}$$

   (i) Find the probability that the class finishes

      (a) after 3:00 p.m.;

      (b) between 2:59 p.m. and 3:01 p.m.

   (ii) If by 3:00 p.m. on a particular day, Dr. Smith has not finished her class, what is the probability that she will do so within the next minute?

8. The lifetime of a light bulb, in thousands of hours, is a continuous variable with density function
$$f(x) = \begin{cases} cx, & 0 < x \le 2, \\ c(4-x), & 2 < x \le 4. \end{cases}$$
   (i) What proportion of this type of bulbs will work for at least 2500 hours?
   (ii) Find the proportion of light bulbs with a life length between 2000 and 3200 hours.

9. The response time (in minutes) of a patient to a new medical treatment for a disease is a continuous random variable X with distribution function
$$F(x) = 1 - (x+1)\mathrm{e}^{-x}, \quad x > 0.$$
   (i) Obtain the density of this distribution.
   (ii) Find the probability that, if three patients are subject to this treatment, at least two of them will have a response time of less than a minute.

10. For what value of c the function
$$f(x) = c(1-x), \quad -1 < x < 1,$$
is the probability density of a continuous random variable? Obtain the distribution function of this random variable.

11. The quantity of gasoline (in tens of thousands of gallons) sold at a gas station during a day has the density function
$$f(x) = \begin{cases} \dfrac{cx^2}{3}, & 0 < x \le 1, \\ c, & 1 < x \le 2, \\ c(4-x), & 2 < x \le 4 \end{cases}$$
for a suitable constant c (which you should be able to identify). Calculate the probability that the daily sales of gasoline at this gas station are
   (i) at most 4000 gallons;
   (ii) between 7000 and 12 500 gallons.

12. Let f be a nonnegative real function for which we have
$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = a,$$
for some positive number a. Show that the function
$$f^*(x) = \frac{f(x)}{a}$$
can be a probability density for a continuous random variable.

13. The distribution function of a continuous random variable X is given by

$$F(t) = \begin{cases} 0, & t \le 0, \\ \dfrac{t}{4} + \dfrac{t}{4} \ln\left(\dfrac{4}{t}\right), & 0 < t \le 4, \\ 1, & t > 4. \end{cases}$$

(i) Calculate the probabilities $P(X \le 3)$, $P(X \ge 1)$, $P(1 \le X \le 3)$.

(ii) Calculate the conditional probabilities $P(X \le 3 | X \ge 1)$, $P(X \le 4 | X \ge 1)$.

(iii) Find the density function of X.

14. The pressure X, measured in psi (pound-force per square inch), at the wings of a turbine that is tested in a tunnel, follows the so-called Rayleigh distribution with density function

$$f(x) = cxe^{-ax^2}, \quad x > 0,$$

where $a = 1/20\,000$ and c is a positive constant.

(i) What is the value of c?

(ii) Find the distribution function of X and use it to calculate the probabilities that the value of the pressure

(a) is at most 200 psi;

(b) is between 100 and 200 psi.

(iii) If it is known that, at a certain instant, the value of the pressure exceeds 100 psi, what is the probability that it has not exceeded twice that value?

15. If f_1, f_2, and f_3 are three density functions, define a new function f by

$$f(x) = \frac{2}{5} f_1(x) + \frac{1}{3} f_2(x) + \lambda f_3(x).$$

For which value(s) of λ is f a probability density? What is the range of values associated with f in terms of the ranges corresponding to f_1, f_2, and f_3?

16. Assume that X is a continuous random variable with density function f and distribution function F. Suppose that a is a real number for which $P(X \le a) < 1$. Show that the function h defined by

$$h(x) = \begin{cases} \dfrac{f(x)}{1 - F(a)}, & x \ge a, \\ 0, & x < a, \end{cases}$$

is also a density function.

17. The sizes of claims (in thousands of dollars) arriving at an insurance company can be modeled by a random variable X having density function

$$f(x) = \frac{4x}{(x+2)^3}, \quad x > 0.$$

What is the proportion of claims arriving at the company that are

(i) at least \$10 000?

(ii) between \$5000 and \$8000?

18. Let the potential losses from an investment, in thousands of dollars, be represented by a random variable X having density function

$$f(x) = \begin{cases} -4x - cx^2, & -1 < x < 0, \\ 0, & \text{elsewhere,} \end{cases}$$

for some real constant c.

(i) Find the value of c.

(ii) Calculate the probability that the losses from this investment are at most \$400.

(iii) If we make five such investments, what is the probability that the loss is at most \$400 in exactly three of them?

Group B

19. The time X, in hours, from the production time of a dairy product until it is safe to consume is a random variable with distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 - e^{-(t/\beta)^\alpha}, & t \geq 0, \end{cases}$$

with $\alpha = 20, \beta = 100$.

(i) Find the density function of X.

(ii) What is the percentage of products of that type that become unsafe to use between 98 and 102 hours after production?

(iii) If we select five such products at random, what is the probability that all five of them are safe to consume 105 hours after their production?

(iv) The company that produces this product wants to estimate the value q, in hours, such that the probability a new product is safe to consume after q hours is 0.99. Find the value of q. (Such a value is called the 99th **percentile** of the distribution F.)

20. The lifetime, X (in days), of a sea microorganism has density function

$$f(x) = cx^2 e^{-4x}, \quad x \geq 0,$$

for a suitable constant c.

(i) Obtain the value of c.

(ii) Find the distribution function of X and hence the probability that a newly-born microorganism survives for at least three days.

21. Let X be a continuous random variable with density function f. We say that X has a **symmetric distribution** around the point a if we have

$$P(X \geq a + x) = P(X \leq a - x)$$

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

(i) Show that the distribution of X is symmetric around a if and only if

$$f(a - x) = f(a + x).$$

(ii) Establish that each of the following distributions, with densities given below, is symmetric around a point a which should be identified:

(a) $f(x) = \dfrac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad x \in \mathbb{R};$

(b) $f(x) = \dfrac{1}{2} - \dfrac{1}{4}|x - 3|, \qquad 1 \leq x \leq 5;$

(c) $f(x) = 5e^{-10|x-7|}, \qquad x \in \mathbb{R};$

(d) $f(x) = \dfrac{5}{2} e^{-5|x|}, \qquad x \in \mathbb{R};$

(e) $f(x) = \dfrac{1}{\pi[1 + (x - 2)^2]}, \qquad x \in \mathbb{R}.$

22. A continuous random variable X has density function

$$f(x) = \begin{cases} 0, & x < 0, \\ \dfrac{e^{-x/4}}{4}, & 0 \leq x < 2, \\ \dfrac{2c}{(x+3)^5}, & x \geq 2. \end{cases}$$

(i) Obtain the value of c.

(ii) Calculate the probabilities $P(0 < X < 2)$ and $P(1 < X < 3)$.

(iii) Find the distribution function, $F(x)$, of X. Is it true that $F'(x) = f(x)$ for all real x? Explain.

6.2 DISTRIBUTION FOR A FUNCTION OF A RANDOM VARIABLE

We have seen in Section 4.4 (see in particular Example 4.16 and the discussion before and after that example) how we can calculate the expectation of a function $Y = g(X)$ of a random variable X, in the case when X is discrete. More generally, it is often the case that we know the distribution of a variable X, either discrete or continuous, but we are interested in the *distribution function* (not just the expectation) of a function, $g(X)$, of that variable. For continuous variables, interest lies of course in the density function of $g(X)$, but it may often be easier to work with the associated distribution function. The idea is that we obtain from the distribution function, say F, of X the distribution of $Y = g(X)$, and then, if needed, we differentiate this distribution function so as to obtain the density function of Y.

To be more precise, a general approach to implement this is as follows:

(a) First, we express the distribution function of Y,

$$F_Y(t) = P(Y \leq t) = P(g(X) \leq t)$$

in terms of the distribution function of X. To accomplish this, we have to specify a set A (which will typically be an interval, or an union of intervals) such that the relation $g(X) \leq t$ is equivalent to $X \in A$. Then, we get

$$F_Y(t) = P(X \in A) = \int_A f(x)\mathrm{d}x;$$

in most cases, this readily yields an expression in terms of the distribution F of X.

(b) Next, having obtained F_Y, we find the density function of Y by differentiating it, that is,

$$f_Y(t) = F'_Y(t),$$

provided, of course, that F_Y is differentiable at t.

This procedure is illustrated in the following examples.

Example 6.3 (Distribution for a linear transformation of a random variable)
Let X be a continuous random variable with density function f and let

$$Y = aX + b,$$

be another random variable that arises as a linear transformation of X; here, $a > 0$ and b are given real numbers.

(i) Find an expression for the density f_Y of Y in terms of f.

(ii) Apply the result in (i) to the case when the density of X is

$$f(x) = \mathrm{e}^{-x}, \quad x \geq 0.$$

SOLUTION Let F and F_Y be the distribution functions of X and Y, respectively.

(i) We have

$$F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P\left(X \leq \frac{y-b}{a}\right)$$

using, for the last step, the fact that $a > 0$. Thus, we have

$$F_Y(y) = F\left(\frac{y-b}{a}\right)$$

so that, upon differentiating both sides with respect to y, we derive

$$f_Y(y) = F_Y'(y) = \left(F\left(\frac{y-b}{a}\right)\right)' = F'\left(\frac{y-b}{a}\right)\left(\frac{y-b}{a}\right)' = f\left(\frac{y-b}{a}\right)\cdot\frac{1}{a}.$$

Thus, we have shown that the two densities f and f_Y are related by

$$f_Y(y) = \frac{1}{a}\cdot f\left(\frac{y-b}{a}\right).$$

(ii) For f as specified, we have

$$f(x) = \begin{cases} 0, & x < 0, \\ e^{-x}, & x \geq 0. \end{cases}$$

Therefore, we see that the density f_Y will take the value zero if we have

$$\frac{y-b}{a} < 0,$$

i.e. if $y < b$, while f_Y will take the value

$$\frac{1}{a}\cdot f\left(\frac{y-b}{a}\right) = \frac{1}{a}\cdot e^{-(y-b)/a}$$

for values of y greater than or equal to b. We thus see that f_Y is given by

$$f_Y(y) = \begin{cases} 0, & y < b, \\ \frac{1}{a}\cdot e^{-(y-b)/a}, & y \geq b. \end{cases}$$

In Figure 6.5, the probability density functions $f(x)$ and $f_Y(y)$ are displayed graphically.

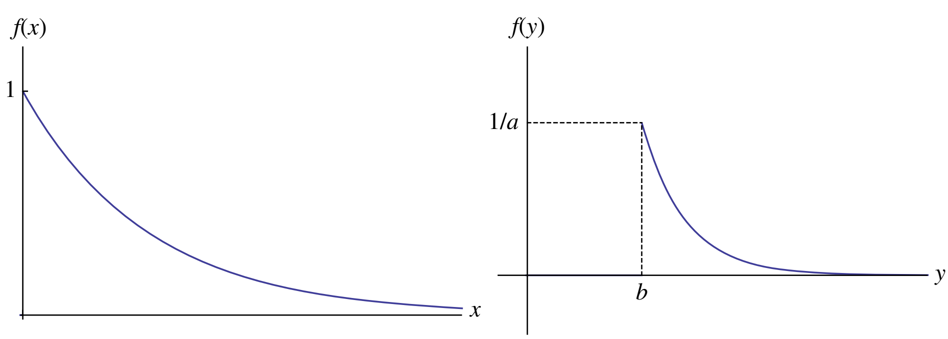


Figure 6.5 The density functions of X and $Y = aX + b$.

Example 6.4 Suppose that some amount of oil is spilled onto the surface of an ocean (for instance, due to an oil-tanker accident). The oil spot then spreads away on the sea surface. The radius, R, of the oil spot in kilometers, 24 hours after the accident, has density function

$$f(r) = \frac{3}{4}\{1 - (20 - r)^2\}, \quad 19 \le r \le 21.$$

Assuming that the area covered by the oil spot is circular, find the density function for the size of this area.

SOLUTION Let Y be the random variable denoting the size of the area. Then, R and Y are related by

$$Y = \pi R^2.$$

Let also F_R and F_Y denote the distribution functions of R and Y, respectively. Then, we have, for any $y \ge 0$,

$$\begin{aligned} P(Y \le y) = P(\pi R^2 \le y) = P\left(R^2 \le \frac{y}{\pi}\right) &= P\left(-\sqrt{\frac{y}{\pi}} \le R \le \sqrt{\frac{y}{\pi}}\right) \\ &= F_R\left(\sqrt{\frac{y}{\pi}}\right) - F_R\left(-\sqrt{\frac{y}{\pi}}\right). \end{aligned}$$

But the range of the variable R is the interval $[19, 21]$, which means that $F_R(r) = 0$ for $r < 19$ and $F_R(r) = 1$ for $r > 21$. This implies, in particular, that the last term on the right-hand side above is zero. Taking this into account, we see that we can express the function F_Y as follows:

$$F_Y(y) = F_R\left(\sqrt{\frac{y}{\pi}}\right) = \begin{cases} 0, & \sqrt{\frac{y}{\pi}} < 19, \\ F_R\left(\sqrt{\frac{y}{\pi}}\right), & 19 \le \sqrt{\frac{y}{\pi}} \le 21, \\ 1, & \sqrt{\frac{y}{\pi}} > 21, \end{cases}$$

or, upon simplifying the conditions on the right, we can write this in the form

$$F_Y(y) = F_R\left(\sqrt{\frac{y}{\pi}}\right) = \begin{cases} 0, & y < 361\pi, \\ F_R\left(\sqrt{\frac{y}{\pi}}\right), & 361\pi \le y \le 441\pi, \\ 1, & y > 441\pi. \end{cases}$$

Differentiating the last expression with respect to y, we find that

$$f_Y(y) = F'_Y(y) = 0, \quad \text{for } y < 361\pi \text{ or } y > 441\pi,$$

while for y in the range $[361\pi, 441\pi]$ we get

$$f_Y(y) = F'_Y(y) = \left[F_R\left(\sqrt{\frac{y}{\pi}}\right)\right]' = F'_R\left(\sqrt{\frac{y}{\pi}}\right) \cdot \left(\sqrt{\frac{y}{\pi}}\right)' = f_R\left(\sqrt{\frac{y}{\pi}}\right) \cdot \frac{1}{2\sqrt{\pi y}}.$$

Thus, we obtain finally the required density to be

$$f_Y(y) = \frac{3}{8\sqrt{\pi}}\, y^{-1/2}\left[1 - \left(20 - \sqrt{\frac{y}{\pi}}\right)^2\right], \quad 361\pi \le y \le 441\pi.$$

Example 6.5 (Measurement error)
It has been found that an electronic device exhibits a measurement error, denoted by X, which is regarded as a random variable with density function

$$f(x) = \begin{cases} \dfrac{1}{7}, & -3 \le x \le c, \\ 0, & \text{elsewhere}, \end{cases}$$

for some suitable constant $c > 0$, which represents the maximum value of that error.

(i) Find the value of c.

(ii) Obtain the density function of the random variable $Y = |X|$, which denotes the magnitude of the error (or the absolute error) made by the device.

SOLUTION

(i) Since we are given that f is a probability density, it must satisfy conditions DF1 and DF2. The first of them is trivially true, while from the second we obtain

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_{-3}^{c} \frac{1}{7}\,\mathrm{d}x = 1,$$

which yields $(c+3)/7 = 1$, or the value of c to be $c = 4$.

(ii) From Part (i) we have, for $x \in [-3, 4]$, the density as $f(x) = 1/7$.

For the distribution function of Y, we find

$$F_Y(y) = P(Y \le y) = P(|X| \le y) = P(-y \le X \le y) = F(y) - F(-y),$$

where F is the distribution function of X. By differentiation, we then derive

$$f_Y(y) = F'(y) = (F(y) - F(-y))' = F'(y) - F'(-y) \cdot (-1),$$

which yields

$$f_Y(y) = f(y) + f(-y). \tag{6.7}$$

We observe that, since X takes values in the interval $[-3, 4]$, the range of values for $Y = |X|$ will be $[0, 4]$.

Taking into account the specific form of the density f, we consider the following two cases for values of y in (6.7):

- For $0 \le y \le 3$, we have $-3 \le -y \le 0$, and so $f(y) = f(-y) = 1/7$ and for any such y, (6.7) yields
$$f_Y(y) = \frac{1}{7} + \frac{1}{7} = \frac{2}{7}.$$
- For y in $(3, 4]$, we have $-y$ to be outside the range of values for the variable X, and so $f(-y) = 0$, hence for values of y in that interval, (6.7) yields that
$$f_Y(y) = \frac{1}{7} + 0 = \frac{1}{7}.$$

Upon combining these expressions, the density function of Y is obtained as

$$f_Y(y) = \begin{cases} \frac{2}{7}, & 0 \le y \le 3, \\ \frac{1}{7}, & 3 < y \le 4. \end{cases}$$

Intuitively, this makes sense since for a small $\epsilon > 0$, it is twice more likely that Y takes a value in an interval $(y - \epsilon, y + \epsilon)$ when $y \in [0, 3]$ compared to the values of y in $(3, 4]$. This is so because the former occurs both for positive and negative values of the original error X, while the latter occurs only when X takes values in $(3, 4]$.

The procedure followed in Examples 6.3 and 6.4 can be used more generally to obtain a result for the distribution of a function $g(X)$ of a random variable X. We shall only discuss the case where g is a differentiable function and g' does not change sign on the interval where it is defined. More explicitly, assume that X takes values in (a, b) with $-\infty \le a < b \le \infty$ and that g is differentiable throughout (a, b) such that either

$$g'(x) > 0, \quad \text{for any } x \in (a, b),$$

or

$$g'(x) < 0, \quad \text{for any } x \in (a, b),$$

holds true. In the first case, g is strictly increasing; in the latter, it is a strictly decreasing function. In either case, for an arbitrary y in the range of g there exists a unique $x \in (a, b)$ such that $g(x) = y$. Further, for a strictly monotone function, its inverse function, g^{-1}, exists and we can safely write $x = g^{-1}(y)$. It is also known from calculus that, for a strictly increasing g, its inverse is also a strictly increasing function so that its derivative satisfies

$$(g^{-1})'(y) = \frac{\mathrm{d}}{\mathrm{d}y} g^{-1}(y) > 0$$

for any y in the range of values for g; if g is strictly decreasing, the same is true with the sign of the inequality in the last expression reversed, i.e. g^{-1} has a negative derivative throughout its domain.

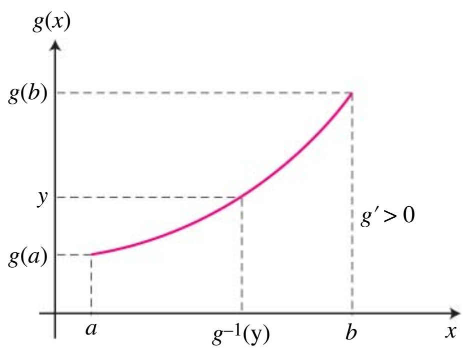
In view of the above, we can now state the following:

Proposition 6.2 *Let X be a continuous random variable with density function f and suppose $f(x) > 0$ for $a < x < b$. Assume that the real function g is such that g' exists and has the same sign throughout the interval (a, b). Then, the density function of the random variable $Y = g(X)$ is given by the formula*

$$f_Y(y) = |(g^{-1})'(y)|f(g^{-1}(y)), \quad y \in \Gamma,$$

where Γ is the interval on the real line with endpoints $g(a)$ and $g(b)$.

Proof: We first assume that g' is positive in the interval (a, b), i.e. g is strictly increasing in that interval.



Then the inequality $g(x) \le y$ is equivalent to $x \le g^{-1}(y)$, so that for the distribution function F_Y of Y, we can write

$$F_Y(y) = P(Y \le y) = P(g(X) \le y) = P(X \le g^{-1}(y)) = F(g^{-1}(y))$$

for $g(a) < y < g(b)$. Differentiating this expression and employing the chain rule for derivatives, we get

$$f_Y(y) = F'_Y(y) = F'(g^{-1}(y)) \cdot [(g^{-1})'(y)] = f(g^{-1}(y))(g^{-1})'(y)$$

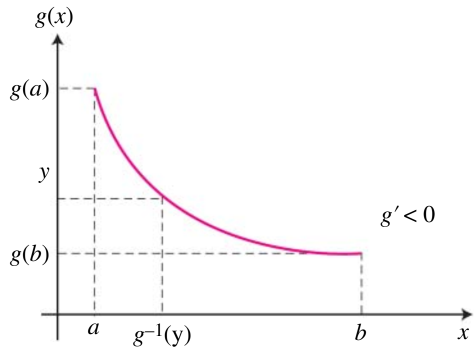
for $y \in (g(a), g(b))$, which proves the result for this case.

For the case where g' is negative on (a, b), g is strictly decreasing in that interval.

We therefore get

$$\begin{aligned} F_Y(y) &= P(Y \le y) = P(g(X) \le y) = P(X \ge g^{-1}(y)) \\ &= 1 - P(X < g^{-1}(y)) = 1 - F(g^{-1}(y)) \end{aligned}$$

for $g(b) < y < g(a)$.



A differentiation yields in this case

$$\begin{aligned} f_Y(y) = F'_Y(y) &= -F'(g^{-1}(y)) \cdot [(g^{-1})'(y)] = -f(g^{-1}(y))(g^{-1})'(y) \\ &= f(g^{-1}(y))|(g^{-1})'(y)|, \quad g(b) < y < g(a), \end{aligned}$$

and this completes the proof. □

We mention that the result of the last proposition is in fact valid under more general conditions, i.e. without assuming necessarily that g is strictly monotone. However, for most of the examples that we will consider, g will be a monotone function. In cases where it is not, we may find the density of Y by a direct argument as we did earlier in Example 6.5.

Finally, we note that there are cases in which, while X is a continuous random variable, $Y = g(X)$, which is a function of X, is discrete. To present a simple example, consider the case when X is continuous with density f, and Y is defined by

$$Y = g(X) = \begin{cases} 1, & \text{if } X \geq 0, \\ 0, & \text{if } X < 0. \end{cases}$$

In this case, the distribution of Y is Bernoulli with success probability

$$p = P(Y = 1) = P(X \geq 0) = \int_0^\infty f(x)\mathrm{d}x.$$

Another example is the integer part of a random variable, $Y = [X]$. If $f(x) > 0$ for all $x \in \mathbb{R}$, then the range of Y will be the set $\{0, \pm 1, \pm 2, \dots\}$ (see Exercise 13 at the end of this section).

EXERCISES

Group A

1. The temperature C (in degrees Celsius) where a certain chemical reaction takes place can be considered as a continuous random variable with density function

$$f(t) = \begin{cases} \dfrac{3}{4}(t - 29)(31 - t), & 29 \leq t \leq 31, \\ 0, & \text{elsewhere.} \end{cases}$$

Obtain the density function of the temperature that the reaction takes place, in degrees Farenheit (recall that if F and C denote temperature in Farenheit and Celcius degrees, respectively, then $F = 1.8C + 32$).

2. Let X be a continuous random variable with probability density function f and assume that $a \neq 0$ and b are two real numbers.

   (i) Show that the probability density f_Y of the variable $Y = aX + b$ is given by the formula
$$f_Y(y) = \frac{1}{|a|} f\left(\frac{y-b}{a}\right).$$

   (ii) If the probability density of X is given by
$$f(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \quad x \in \mathbb{R},$$
find the density function of $Y = \mu + \sigma X$, where $\sigma > 0$ and $\mu \in \mathbb{R}$ are two real constants.

3. Let X be a continuous random variable with density function $f(x), x \in R_X$, and distribution function $F(t), t \in \mathbb{R}$. Express the density function and the distribution function of the random variable $Y = X^2$ in terms of f and F.

   Application: Find the density function of $Y = X^2$ when the density of X is
$$f_X(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}, \quad x \in \mathbb{R}.$$

4. The density of a random variable X is given by
$$f(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$
Derive the density function of the random variables $Y = X^a$ and $Z = \mathrm{e}^X$, where a is a given positive real number.

5. The diameter of a bubble, at the time when it breaks, is a random variable X with density function
$$f(d) = \begin{cases} \dfrac{3}{4} d(2-d), & 0 < d < 2, \\ 0, & \text{elsewhere.} \end{cases}$$
Find the distribution function of the area E and the volume V of the bubble at the time when it breaks.

6. The value I of the electric current passing through a resistance, R, of a circuit is a continuous random variable with probability density
$$f(x) = \begin{cases} \dfrac{1}{4}, & 6 < x < 10, \\ 0, & \text{elsewhere.} \end{cases}$$

Derive the density function of the electric power, given by $P = RI^2$, assuming the value of the resistance R to be kept constant.

7. A continuous variable X has density function

$$f(x) = \begin{cases} \dfrac{2x}{9}, & 0 < x < 3, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the density function of the variables

(i) $Y = 3X - 2$;

(ii) $W = (3 - X)(3 + X)$.

8. Let X be a continuous random variable with density f. Suppose Y is another random variable which is related to X by

$$Y = e^{-X}.$$

(i) Prove that the density function, $f_Y(y)$, of Y is related to f by the formula

$$f_Y(y) = \frac{1}{y} f(-\ln y), \quad y > 0.$$

(ii) Apply the result from Part (i) to find f_Y in the case when the density f is given by

$$f(x) = \begin{cases} \dfrac{x}{2}, & 0 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

9. For a continuous variable X with density f, show that the density function, f_Y, of the variable $Y = X^2$ is given by

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f(\sqrt{y}) + f(-\sqrt{y})].$$

Group B

10. The velocity of the molecules in a homogeneous gas that is in the state of equilibrium can be described by a continuous variable V with density function

$$f(v) = \frac{4}{\sqrt{\pi}} v^2 e^{-av^2}, \quad v > 0,$$

for some positive constant a.

(i) Find the value of a using the following result (known from calculus):

$$\int_0^\infty e^{-x^2} dx = \frac{\sqrt{\pi}}{2}.$$

(ii) Obtain the density function of the variable

$$E = \frac{1}{2} m V^2$$

which represents the kinetic energy of the molecules (here, m is the molecule mass, which is assumed to be known).

11. Assume that X is a continuous random variable with distribution function

$$F(x) = \begin{cases} 0, & x < \theta, \\ 1 - \exp\left[-\left(\frac{x-\theta}{\lambda}\right)^a\right], & x \geq \theta, \end{cases}$$

where $a > 0$, $\lambda > 0$, and $\theta \in \mathbb{R}$ are known parameters of this distribution. Find the density function of the random variable

$$Y = \left(\frac{X - \theta}{\lambda}\right)^a.$$

12. Two continuous random variables X and Y with $R_X = \mathbb{R}$ and $R_Y = (-\pi/2, \pi/2)$ are related by $X = \tan Y$. Find the density function of the random variable Y if it is known that the density of X is given by

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad x \in \mathbb{R}.$$

The distribution of the variable X above is known as **Cauchy distribution**.

13. Let X be a continuous random variable whose distribution function F is strictly increasing throughout $\mathbb{R}$. Show that the random variable $Y = [X]$, the integer part of X, is a discrete variable with probability function

$$f_Y(y) = F(y + 1) - F(y), \quad y \in R_Y = \{0, \pm 1, \pm 2, \ldots\}.$$

14. Let X be a random variable with density function

$$f(x) = e^{-x}, \quad x > 0.$$

Obtain the density function for each of the following variables:

$$Y = -2X, \quad Z = (1 + X)^{-1}, \text{ and}$$

$$W = \begin{cases} X, & \text{if } X \leq 1, \\ 1/X, & \text{if } X > 1. \end{cases}$$

6.3 EXPECTATION AND VARIANCE

The expectation of a continuous random variable is defined in a similar way as we have seen for discrete variables in Chapter 4. More specifically, replacing the sum by an integral, we have the following definition.

Definition 6.2 Let X be a continuous random variable with density function f. Then, the **expected value or the expectation** of X is given by the expression

$$E(X) = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x \tag{6.8}$$

provided that the integral exists (converges) absolutely, i.e.

$$\int_{-\infty}^{\infty} |x|\, f(x)\mathrm{d}x < \infty. \tag{6.9}$$

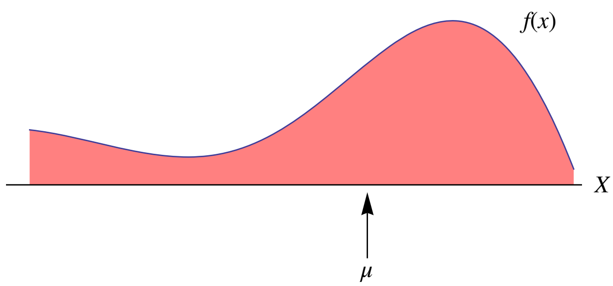
If the integral in (6.8) fails to converge absolutely, we typically distinguish between two cases: if X takes only nonnegative values, then both the integrals appearing in (6.8) and (6.9) are equal to $+\infty$, and we often write this as $E(X) = \infty$. In that case, the expression "X has no (finite) expectation" and "the expectation of X is infinite" are often used interchangeably. If, in contrast, X takes both positive and negative values, and the integral in (6.9) is not finite, the expectation of X cannot be assigned any value and we simply say that this expectation is *undefined* or that *it does not exist*.

As with discrete variables, other expressions we use for the expectation of a random variable X are the mean value, or simply the mean of X. Moreover, and in analogy with the discrete case, the symbol we use for expectations is the Greek letter μ, while if more than one variables $X, Y, Z, \ldots$ are involved, we use μ_X, μ_Y, μ_Z, and so on for their expected values.

From (6.8), it is clear that

$$E(X - \mu) = \int_{-\infty}^{\infty} (x - \mu)f(x)\mathrm{d}x = \int_{-\infty}^{\infty} xf(x)\mathrm{d}x - \int_{-\infty}^{\infty} \mu f(x)\mathrm{d}x = \mu - \mu = 0.$$

This admits the following natural geometric interpretation. Suppose that the graph of a density f is drawn on a homogeneous rod and we cut the part of the graph between the values of the function f and the horizontal axis (for simplicity, assume that X takes values in a finite interval $[a, b]$ so that the part of the graph we cut is over a finite range of values). Then the mean value, μ, of X is, in physical terms, the point on the x-axis where the homogeneous body that arises from the above procedure should be supported so that it is balanced.



Example 6.6 We consider a family of distributions that is widely used as a model for the income distribution of a population, the claim sizes in an insurance portfolio, etc. The density function of the **Pareto** distribution is given by

$$f(x) = \begin{cases} \dfrac{k\theta^k}{x^{k+1}}, & x \geq \theta, \\ 0, & \text{elsewhere}, \end{cases} \tag{6.10}$$

where $k > 0$ and $\theta > 0$ are given **parameters** of this distribution. For the expectation of this distribution, we have that

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x)\mathrm{d}x = \int_{\theta}^{\infty} xf(x)\mathrm{d}x = \int_{\theta}^{\infty} x\frac{k\theta^k}{x^{k+1}}\mathrm{d}x \\ &= k\theta^k \int_{\theta}^{\infty} x^{-k}\mathrm{d}x. \end{aligned} \tag{6.11}$$

If the parameter k is such that $k > 1$, then it is clear that the integral above converges (it is finite). In particular, we obtain

$$\begin{aligned} E(X) &= k\theta^k \left[\frac{x^{-k+1}}{-k+1}\right]_{\theta}^{\infty} = \frac{k}{k-1}\ \theta^k\ \left[-\frac{1}{x^{k-1}}\right]_{\theta}^{\infty} \\ &= \frac{k}{k-1}\theta^k\left(-0+\frac{1}{\theta^{k-1}}\right) = \frac{k\theta}{k-1}. \end{aligned}$$

It should be clear from (6.11) that for all other values of k, the expectation of the Pareto distribution does not exist (finitely). For instance, when $k = 1$, we obtain for the mean of a Pareto random variable that

$$E(X) = k\theta^k \int_{\theta}^{\infty} x^{-1}\mathrm{d}x = k\theta^k[\ln x]_{\theta}^{\infty},$$

which is not finite. Similarly, we can see that the expectation does not exist for values of k strictly less than one.

Example 6.7 Suppose the time X (in hours) that an ATM outside a bank is in use during a day is a random variable with density function

$$f(x) = \begin{cases} \dfrac{x}{9}, & 0 \leq x \leq 3, \\ -\dfrac{1}{9}(x-6), & 3 < x \leq 6. \end{cases}$$

Find the mean time that the machine operates during a day.

SOLUTION In this case, we have to split the range of integration in formula (6.8), because the expression for $f(x)$ given is different in the intervals $[0, 3]$ and $(3, 6]$. Thus, we obtain

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x) = \int_0^3 \frac{x}{9}\, x\mathrm{d}x + \int_3^6 x\left\{-\frac{1}{9}(x-6)\right\}\mathrm{d}x$$

$$= \frac{1}{9}\left[\frac{x^3}{3}\right]_0^3 + \left(-\frac{1}{9}\right)\left\{\left[\frac{x^3}{3}\right]_3^6 - 6\left[\frac{x^2}{2}\right]_3^6\right\}$$

$$= \frac{1}{9}\cdot\frac{3^3-0}{3} - \frac{1}{9}\left(\frac{6^3-3^3}{3} - 6\frac{6^2-3^2}{2}\right) = 3,$$

and so the machine operates on average 3 hours per day.

Next, we prove a proposition about the expectation for a function of a random variable.

Proposition 6.3 *Let X be a continuous random variable with density function f and $g : \mathbb{R} \mapsto \mathbb{R}$ be a real function. Then the expectation of the random variable $Y = g(X)$ is given by*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x.$$

Proof: Here, we shall only give the proof of the proposition when the conditions of Proposition 6.2 are satisfied. An alternative proof, which covers the general case, i.e. without assuming that g is monotone, is given at the end of Section 6.4.

If the function g is strictly increasing in the set $\mathbb{R}$, the density function of Y takes the form

$$f_Y(y) = (g^{-1})'(y)\cdot f(g^{-1}(y)), \quad y \in \mathbb{R},$$

which implies that

$$E[g(X)] = E(Y) = \int_{-\infty}^{\infty} yf_Y(y)\mathrm{d}y$$

$$= \int_{\mathbb{R}} y(g^{-1})'(y)\cdot f(g^{-1}(y))\mathrm{d}y.$$

Upon making the substitution (change of variable) $g^{-1}(y) = x$, we obtain $\mathrm{d}x = (g^{-1})'(y)\mathrm{d}y$, $y = g(x)$, whence

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x,$$

as required. The proof for the case where the function g is strictly decreasing is almost identical; the details are left for the reader to verify. □

An interesting remark concerning the applicability of the result in Proposition 6.3 is that, in order to find the expectation of the random variable $g(X)$, we do not have to know *the density* (or the distribution function) of that random variable. Both of these might be difficult to obtain, especially when the function g is complicated and the assumptions of Proposition 6.2 are not met.

The following two results arise by a straightforward application of Proposition 6.3, combined with Proposition 4.5, which refers to the expectation of a constant random variable (a variable which takes only one value).

Proposition 6.4 *Let X be a continuous random variable, $g_1, g_2, \ldots, g_k$ be real functions, and $\lambda_1, \lambda_2, \ldots, \lambda_k$ be real numbers. Then, the following holds:*

$$E[\lambda_1 g_1(X) + \cdots + \lambda_k g_k(X)] = \lambda_1 E[g_1(X)] + \cdots + \lambda_k E[g_k(X)].$$

Proof: We introduce the function

$$g = \lambda_1 g_1 + \cdots + \lambda_k g_k = \sum_{i=1}^{k} \lambda_i g_i.$$

Then, by using Proposition 6.3 and the properties of the integral, we obtain

$$\begin{aligned}
E[\lambda_1 g_1(X) + \cdots + \lambda_k g_k(X)] = E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \left(\sum_{i=1}^{k} \lambda_i g_i(x) \right) f(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \sum_{i=1}^{k} \lambda_i [g_i(x) f(x)]\mathrm{d}x = \sum_{i=1}^{k} \lambda_i \int_{-\infty}^{\infty} g_i(x) f(x)\mathrm{d}x \\
&= \sum_{i=1}^{k} \lambda_i E[g_i(X)],
\end{aligned}$$

which is the same as the required result. □

An important special case of the above proposition is given next, with regard to the expectation of a linear function of a random variable.

Corollary 6.1 *Let X be a continuous random variable and a and b be two real numbers. Then,*

$$E(aX + b) = aE(X) + b.$$

It is worth noting at this point that, since the above properties of expectation are identical for the discrete and continuous cases, the calculations given after Corollary 4.1 or those in Example 4.18 are also valid for continuous random variables.

Example 6.8 At one end of a stick with length $OA = R$, we place a sphere. We push the sphere using a force of random magnitude so that it is rotated around point O (see Figure 6.6) creating an angle of Θ degrees, and the density function of this angle is given by

$$f_\Theta(\theta) = \begin{cases} \dfrac{2}{\pi}, & 0 < \theta < \dfrac{\pi}{2}, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the expected value of the position, $(E(X), E(Y))$, of the sphere after its rotation.

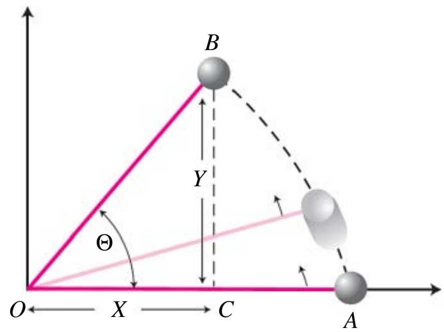


Figure 6.6 Rotation of a sphere around point O.

SOLUTION As can be seen from Figure 6.6, we have

$$X = R\cos\Theta, \quad Y = R\sin\Theta,$$

so that we obtain, first for the expectation $E(X)$,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} R\cos\theta f_\Theta(\theta)\mathrm{d}\theta = \int_0^{\pi/2} R\cos\theta \cdot \frac{2}{\pi}\,\mathrm{d}\theta = \frac{2R}{\pi}\int_0^{\pi/2}\cos\theta\mathrm{d}\theta \\ &= \frac{2R}{\pi}[\sin\theta]_0^{\pi/2} = \frac{2R}{\pi}\left(\sin\frac{\pi}{2} - \sin 0\right) = \frac{2R}{\pi}. \end{aligned}$$

In a similar fashion, we find for $E(Y)$ that

$$\begin{aligned} E(Y) &= \int_{-\infty}^{\infty} R\sin\theta f_\Theta(\theta)\mathrm{d}\theta = \int_0^{\pi/2} R\sin\theta \cdot \frac{2}{\pi}\,\mathrm{d}\theta = \frac{2R}{\pi}\int_0^{\pi/2}\sin\theta\mathrm{d}\theta \\ &= \frac{2R}{\pi}[-\cos\theta]_0^{\pi/2} = -\frac{2R}{\pi}\left(\cos\frac{\pi}{2} - \cos 0\right) = \frac{2R}{\pi}. \end{aligned}$$

Consequently, the expected position of the sphere lies on the angle bisector of the first quadrant of the Cartesian axes.

Next, we consider the variance of a continuous random variable. Note that, although the definitions for the expectation are different for discrete and continuous variables, the definition below concerning the variance of a continuous variable is identical to the one in Definition 4.5 for discrete variables.

Definition 6.3 Let X be a continuous random variable for which the expectation $\mu = E(X)$ exists. The quantity

$$\sigma^2 = \mathrm{Var}(X) = E[(X-\mu)^2] = E[(X - E(X))^2]$$

is called the **variance** of X. The quantity $\sigma = \sqrt{\mathrm{Var}(X)}$ is called the **standard deviation** of the variable X and is expressed in the same units as X. Here, the symbol $\sqrt{\mathrm{Var}(X)}$ means the nonnegative square root of the variance.

From the definition we see, by an appeal to Proposition 6.3 that, for a continuous variable X, the variance of X is given by the formula

$$\mathrm{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x)\, \mathrm{d}x.$$

As in the discrete case, we would rarely use this expression for the calculation of the variance. Instead, we use again the formula

$$\mathrm{Var}(X) = E(X^2) - \mu^2 = E(X^2) - [E(X)]^2, \tag{6.12}$$

while we note that for $a, b \in \mathbb{R}$ we also have

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X), \quad \sigma_{aX+b} = |a|\sigma_X.$$

The last two expressions are completely analogous to the results in the discrete case, and thus we may, from now on, use them irrespective of whether the random variable is discrete or continuous. The proofs are not given since they are identical to the proofs of Propositions 4.9 and 4.10.

Example 6.9 We revisit the Pareto distribution, introduced in Example 6.6. This has density as in (6.10) with parameters k, θ. Suppose now we are interested in the second moment (around zero) of that distribution,

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\mathrm{d}x = k\theta^k \int_{\theta}^{\infty} x^{-k+1}\mathrm{d}x.$$

It is clear that, when $k > 2$, we have

$$E(X^2) = k\theta^k \left[\frac{x^{-k+2}}{-k+2}\right]_{\theta}^{\infty} = k\theta^k \cdot \frac{\theta^{-k+2}}{k-2} = \frac{k\theta^2}{k-2},$$

while for $k \leq 2$ the expectation does not exist (it is infinite). So, for the Pareto distribution, we conclude that while it is defined for any $k > 0$,

- for $0 < k \leq 1$, the expectation does not exist (and hence, the second and higher-order moments also do not exist);
- for $1 < k \leq 2$, the expectation exists, but not the variance (which is infinite);
- for $k > 2$, both the expectation and the variance exist. The latter is given by

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{k\theta^2}{k-2} - \left(\frac{k\theta}{k-1}\right)^2 = \frac{k\theta^2}{(k-2)(k-1)^2},$$

after a little algebra.

The remarks we made at the end of Example 4.15 for discrete variables, about the nonexistence of the expectation when the distribution of the random variable takes values far away from the mean with nonnegligible probability, are also applicable in the case of continuous variables. In fact, the Pareto distribution in the last example is one of the most widely used *heavy-tailed* continuous distributions.

In closing this section, we note that a number of results that we have proved for discrete variables also hold for continuous ones, since their proof relies on properties shared by both types of variables; we mention, for example, Exercises 7, 8, 11, and 14 of Section 4.5 and Exercise 15 of Section 4.10.

EXERCISES

Group A

1. The time, in hours, it takes to repair the fault in a machine has a continuous distribution with density function

$$f(x) = \begin{cases} 1, & 0 \le x \le 1, \\ 0, & \text{elsewhere.} \end{cases}$$

If the cost associated with the machine not working for x hours is $3x + 2$, calculate the expected cost for each fault of the machine.

2. The ash concentration (as a percentage) in a certain type of coal is a continuous random variable with probability density function

$$f(x) = \begin{cases} \dfrac{1}{1125}(x-5)^2, & 5 \le x \le 20, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the mean percentage concentration of ash for this type of coal.

3. The repair time, in hours, for a certain type of laptop is a continuous variable with density function

$$f(x) = \begin{cases} 1/2, & 0 \le x \le 2, \\ 0, & \text{elsewhere.} \end{cases}$$

   (i) What is the expected time to repair a laptop of this type when it breaks down?

   (ii) If the repair cost depends on the time that the repair takes and, when this time is x hours, the associated cost in dollars is estimated to be $30 + 5\sqrt{x}$, find the expected cost for each repair.

4. The weekly circulation, in tens of thousands, of a magazine is a random variable X whose density function is

$$f(x) = \begin{cases} 2\left(1 - \dfrac{1}{x^2}\right), & 1 \le x \le 2, \\ 0, & \text{elsewhere.} \end{cases}$$

What is the expected value and the standard deviation for the number of magazines sold weekly?

5. The monthly income (in thousands of dollars) of a family in a city, represented by a random variable X, has the Pareto distribution (see Examples 6.6 and 6.9) with parameters $k = 4$ and $\theta = 2$.
   (i) Obtain the distribution function of X.
   (ii) Find the probability that the monthly income of a randomly chosen family is between 4000 and 5000.

6. The density function of a random variable X is

$$f(x) = a + bx^2, \quad 0 \le x \le 1.$$

If we know that $E(X) = 3/5$, what are the values of a and b?

7. For what values of a and b is the function

$$f(x) = a(b - x)^2, \quad 0 \le x \le b,$$

the probability density function of a continuous random variable X with $E(X) = 1$?

8. The time, in hours, that a student needs to complete a Mathematics exam is a random variable X with density function

$$f(x) = \begin{cases} 6(x-1)(2-x), & 1 < x < 2, \\ 0, & \text{elsewhere.} \end{cases}$$

What is the expected value and the variance of X?

9. Suppose the density function of a random variable X is

$$f(x) = \begin{cases} \dfrac{c}{x^3}, & 1 < x < 3, \\ 0, & \text{elsewhere,} \end{cases}$$

where c is a real constant.
   (i) Find the value of c.
   (ii) Calculate the expected value of the random variable $Y = \ln X$.

10. Let X be a random variable with density function

$$f(x) = \frac{1}{\pi(1 + x^2)}, \quad -\infty < x < \infty,$$

which is the density of the Cauchy distribution – see Exercise 12 of Section 6.2. Prove that the expectation of X does not exist.

Group B

11. Find the expectation and the variance of a random variable X whose density function is

$$f(x) = \frac{3}{x^4}, \quad x \geq 1.$$

Then, find the density function, the expectation and the variance of the random variable

$$Y = \frac{1}{X},$$

and verify that the following hold true in this case:

$$E\left(\frac{1}{X}\right) \neq \frac{1}{E(X)}, \quad E(XY) \neq E(X)E(Y).$$

12. For a random variable X with mean μ, standard deviation σ, and third central moment $\mu_3 = E[(X - \mu)^3]$, the **coefficient of skewness** for the distribution of X is given by

$$\gamma_1 = \frac{\mu_3}{\sigma^3}$$

(this is a measure of the departure from symmetry for the distribution of X; note that γ_1 can take both positive and negative values, while it is zero for symmetric distributions).

Obtain the coefficient of skewness for the distribution of X when X has the density of the Pareto distribution in (6.10). What conditions are needed for the parameters of that distribution so that γ_1 exists (i.e. it is finite)?

13. Let X be a continuous variable with density

$$f(x) = \begin{cases} \dfrac{3x^2 + 1}{2}, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere,} \end{cases}$$

and let $Y = [3X + 1]$, where [ ] denotes the integer part.

(i) Calculate the distribution of the *discrete* random variable Y and then use Definition 4.4 to derive its expectation.

(ii) Is it true that the expected value of Y equals three times the integer part of the expectation of X plus one?

14. The lifetime, in hours, of an electrical appliance is described by a random variable X, having distribution function

$$F(t) = 1 - e^{-\theta t^3}, \quad t \geq 0,$$

where $\theta > 0$ is a known parameter of this distribution. A company sells this appliance making a profit of k dollars and gives its customers a guarantee that the appliance will work properly for at least α hours. In case the machine breaks down before that period, the company will repair it without a charge and the cost for this repair is k_1 dollars. If Y denotes the profit that the company makes for each item sold,

(i) identify the range of values and the probability function of Y;

(ii) show that the expectation of Y is given by

$$E(Y) = (k - k_1) + k_1 e^{-\theta\alpha^3}.$$

Application: Suppose $\theta = 10^{-6}$, $k = \$140$, and $k_1 = \$45$. If the company wishes the mean profit for each item sold to be \$120, what is the guarantee working time α it should offer its customers?

15. The production time, in minutes, of a manufacturing item is a continuous random variable X with density function

$$f(x) = \begin{cases} c(3x^2 + 2x), & 0 \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

for some suitable constants b and c.

(i) If it is known that $E(X) = 13/9$, what are the values of b and c?

(ii) Calculate the standard deviation of X.

(iii) Find the probability that, among the next three items which will be produced, none will have a production time of less than one minute.

16. The number of light bulbs of a certain type (measured in hundreds), sold by a large store during a year, is a random variable X having the distribution function

$$F(t) = \begin{cases} 0, & t \le 0, \\ \dfrac{t}{60}, & 0 < t \le 60, \\ 1, & t > 60. \end{cases}$$

The profit that the store makes for each lightbulb sold is \$1. If by the end of a year, a lightbulb has not been sold, it results in a loss of \$0.2 to the store.

(i) Calculate the expected profit of the store if, at the beginning of a year, it makes an order for y hundreds of light bulbs.

(ii) What is the quantity y that the store must order in order to maximize its expected profit?

6.4 ADDITIONAL USEFUL RESULTS FOR THE EXPECTATION

In this section, we shall present some further results for the expectation, similar to those given in Section 4.6 for discrete distributions. More specifically, we shall see that both Markov's and Chebyshev's inequalities are valid in the case of continuous variables.

Proposition 6.5 *Let X be a nonnegative continuous random variable for which the expectation $E(X)$ exists. Then, for any $t > 0$, we have*

$$P(X \geq t) \leq \frac{E(X)}{t}.$$

Proof: Since X takes nonnegative values and t is positive, we can write

$$E(X) = \int_0^\infty xf(x)\mathrm{d}x \geq \int_t^\infty xf(x)\mathrm{d}x,$$

where f is the density of X. Now, for $x \geq t$, we have $xf(x) \geq tf(x)$ so that, integrating both sides over the integral $[t, \infty)$, we deduce that

$$\int_t^\infty xf(x)\mathrm{d}x \geq \int_t^\infty tf(x)\mathrm{d}x = t\int_t^\infty f(x)\mathrm{d}x = tP(X \geq t).$$

The result of the proposition is now immediate from the last two displayed expressions. □

Recall that Chebyshev's inequality for discrete random variables was a straightforward deduction from Markov's inequality. We thus anticipate the former inequality to be valid in the continuous case as well.

Proposition 6.6 *Let X be a continuous random variable with mean μ and variance σ^2. Then, for any $t > 0$, we have*

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}.$$

Proof: The proof is the same as in the discrete case (see Proposition 4.12). □

Example 6.10 Consider the random variable X having density function

$$f(x) = \begin{cases} 1, & 0 \leq x \leq 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Then we get

$$E(X) = \int_{-\infty}^\infty xf(x)\mathrm{d}x = \int_0^1 x\mathrm{d}x = \left[\frac{x^2}{2}\right]_0^1 = \frac{1}{2}$$

and

$$E(X^2) = \int_{-\infty}^\infty x^2f(x)\mathrm{d}x = \int_0^1 x^2\mathrm{d}x = \left[\frac{x^3}{3}\right]_0^1 = \frac{1}{3},$$

so that the variance of X is given by

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.$$

By Chebyshev's inequality, we now obtain that

$$P(|X - 0.5| \geq t) \leq \frac{1}{12t^2},$$

or

$$P(0.5 - t \leq X \leq 0.5 + t) \geq 1 - \frac{1}{12t^2}. \tag{6.13}$$

We note that, as in the discrete case, the bound we obtain from Chebyshev's inequality can be very conservative. For instance, in this case, we see that the exact value for the probability on the left-hand side above equals, for $t \leq 1 = 2$,

$$P(0.5 - t \leq X \leq 0.5 + t) = \int_{0.5-t}^{0.5+t} f(x)\mathrm{d}x = \int_{0.5-t}^{0.5+t} \mathrm{d}x = 2t,$$

so that the error of Chebyshev's bound is equal to

$$2t - \left(1 - \frac{1}{12t^2}\right) = \frac{24t^3 - 12t^2 + 1}{12t^2}.$$

For $t = 0.4$, the error is 0.42, while for $t = 0.3$, it becomes 0.53. Moreover, it is clear that for $t < 1/(2\sqrt{3})$, the bound in (6.13) is useless since it takes negative values.

Despite its apparent disadvantages, Chebyshev's inequality is a powerful tool for establishing some useful identities, in particular of an asymptotic nature. Also, since both Markov's and Chebyshev's inequalities are valid (and have the same form) both for discrete and continuous random variables, they can be used to obtain general results, which are valid for *any* random variable.

Example 6.11 If a random variable X is constant, i.e. there exists a real c such that $P(X = c) = 1$, then $E(X) = c$ and $\mathrm{Var}(X) = 0$. We employ Chebyshev's inequality and the continuity of probability to show that the converse is also true.

For this purpose, assume that X is such that $E(X) = \mu$ and $\mathrm{Var}(X) = 0$. Consider the sequence of events

$$A_n = \left\{|X - \mu| < \frac{1}{n}\right\}, \quad n = 1, 2, \ldots$$

It is clear that the sequence $(A_n)_{n\geq 1}$ is decreasing and that

$$\lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \bigcap_{n=1}^{\infty} \left\{|X - \mu| < \frac{1}{n}\right\} = \{X = \mu\}.$$

We therefore deduce, by an appeal to the continuity property of probability (Proposition 1.11), that

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

But Chebyshev's inequality now yields

$$P\left(|X-\mu| \geq \frac{1}{n}\right) \leq \frac{\mathrm{Var}(X)}{(1/n)^2} = n^2 \mathrm{Var}(X) = 0$$

for any $n = 1, 2, \ldots$ Consequently,

$$P(A_n) = P\left(|X-\mu| < \frac{1}{n}\right) = 1 - P\left(|X-\mu| \geq \frac{1}{n}\right) = 1,$$

and we finally obtain

$$P(X = \mu) = \lim_{n\to\infty} P(A_n) = 1.$$

Thus, X assumes only the value μ with probability one, as asserted.

The proof of the result in Example 6.11 covers both the discrete and continuous cases. Note that for the case when X is discrete, a different proof has been given in Proposition 4.8.

Example 6.12 Prove that there does not exist a random variable X with $E(X) = \mu$ and $\mathrm{Var}(X) = \sigma^2$ such that

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 0.70.$$

SOLUTION From Chebyshev's inequality we have, for any variable X with mean μ and variance σ^2,

$$P(|X-\mu| \geq 2\sigma) \leq \frac{\sigma^2}{(2\sigma)^2} = \frac{1}{4},$$

which in turn implies that

$$\begin{aligned} P(\mu - 2\sigma < X < \mu + 2\sigma) &= P(|X-\mu| < 2\sigma) = 1 - P(|X-\mu| \geq 2\sigma) \geq 1 - \frac{1}{4} \\ &= \frac{3}{4} > 0.70. \end{aligned}$$

This shows that the equation in the statement of the example is impossible for any random variable X.

In Proposition 4.13, we have seen that, when X is discrete and nonnegative, the expectation $E(X)$ can be found from the associated distribution function F, in cases where the probability function of X is not easy to find. The same is true in the continuous case, even without the nonnegativity condition for X, as the next proposition shows.

Proposition 6.7 *Let X be a continuous random variable having distribution function F. Provided the expectation of X exists, it can be expressed in terms of F as follows:*

$$E(X) = \int_0^\infty (1 - F(x))\mathrm{d}x - \int_{-\infty}^0 F(x)\mathrm{d}x. \tag{6.14}$$

In particular, if X takes only nonnegative values, the last result takes the form

$$E(X) = \int_0^\infty (1 - F(x))\mathrm{d}x.$$

Proof: In order to establish (6.14), we calculate each of the two integrals on the right-hand side of that equation. For the first integral, since

$$1 - F(x) = P(X > x) = \int_x^\infty f(y)\mathrm{d}y,$$

we have

$$\int_0^\infty (1 - F(x))\mathrm{d}x = \int_0^\infty \left(\int_x^\infty f(y)\mathrm{d}y \right) \mathrm{d}x = \int_0^\infty \int_x^\infty f(y)\mathrm{d}y\mathrm{d}x.$$

Interchanging the order of variables in the integration (which is allowed, since the integrand is nonnegative), we see that

$$\int_0^\infty (1 - F(x))\mathrm{d}x = \int_0^\infty \int_0^y f(y)\mathrm{d}x\mathrm{d}y = \int_0^\infty \left(\int_0^y \mathrm{d}x \right) f(y)\mathrm{d}y = \int_0^\infty yf(y)\mathrm{d}y.$$

Similarly, substituting in the second integral on the right of (6.14), for $x < 0$, the function

$$F(x) = \int_{-\infty}^x f(y)\mathrm{d}y$$

and interchanging the order of integration, we derive

$$\begin{aligned} \int_{-\infty}^0 F(x)\mathrm{d}x &= \int_{-\infty}^0 \left(\int_{-\infty}^x f(y)\mathrm{d}y \right) \mathrm{d}x = \int_{-\infty}^0 \int_{-\infty}^x f(y)\mathrm{d}y\mathrm{d}x = \int_{-\infty}^0 \int_y^0 f(y)\mathrm{d}x\mathrm{d}y \\ &= \int_{-\infty}^0 \left(\int_y^0 \mathrm{d}x \right) f(y)\mathrm{d}y = \int_{-\infty}^0 (-y)f(y)\mathrm{d}y = -\int_{-\infty}^0 yf(y)\mathrm{d}y. \end{aligned}$$

We thus see that the right-hand side in (6.14) equals

$$\int_0^\infty yf(y)\mathrm{d}y + \int_{-\infty}^0 yf(y)\mathrm{d}y = \int_{-\infty}^\infty yf(y)\mathrm{d}y = E(X).$$

The result for nonnegative X is immediate from this, since in that case, $F(x) = 0$ for all $x < 0$, and so $\int_{-\infty}^0 F(x)\mathrm{d}x = 0$. □

We note at this point that the result of Proposition 6.7 admits the following two equivalent representations:

$$E(X) = \int_0^\infty (1 - F(x))\mathrm{d}x - \int_0^\infty F(-x)\mathrm{d}x = \int_0^\infty \{(1 - F(x) - F(-x))\}\mathrm{d}x$$

and

$$E(X) = \int_0^\infty P(X > x)\mathrm{d}x - \int_0^\infty P(X \le -x)\mathrm{d}x. \tag{6.15}$$

Example 6.13 A nonnegative random variable X has distribution function

$$F(x) = 1 - \frac{3}{5}e^{-5x} - \frac{2}{5}e^{-2x}, \quad x \geq 0.$$

(i) Verify that F is a proper distribution function.

(ii) Find $E(X)$.

SOLUTION

(i) We need to check that all conditions for a distribution function are satisfied. First, it is obvious that $F(x) \geq 0$ for all x (in particular, $F(x) = 0$ for $x < 0$, since X is nonnegative). Next, we have

$$\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$$

and finally, we note that F is right continuous (in fact, it is both left and right continuous for any real x). It is also straightforward to check that F is a nondecreasing function. Thus, it is a proper distribution function of a random variable.

(ii) We use the second formula in Proposition 6.7, because X takes only nonnegative values. This gives

$$\begin{aligned} E(X) &= \int_0^\infty (1 - F(x))\mathrm{d}x = \int_0^\infty \left(\frac{3}{5}e^{-5x} + \frac{2}{5}e^{-2x}\right)\mathrm{d}x \\ &= \frac{3}{5}\int_0^\infty e^{-5x}\mathrm{d}x + \frac{2}{5}\int_0^\infty e^{-2x}\mathrm{d}x \\ &= \frac{3}{5}\left[\frac{e^{-5x}}{-5}\right]_0^\infty + \frac{2}{5}\left[\frac{e^{-2x}}{-2}\right]_0^\infty = \frac{3}{5}\cdot\frac{1}{5} + \frac{2}{5}\cdot\frac{1}{2} \\ &= \frac{8}{25}. \end{aligned}$$

The result of Proposition 6.7 enables us to provide an alternative proof of Proposition 6.3, about the expectation for a function of a random variable X. In this proof, we do not need any assumptions about the monotonicity of g. In fact, setting $Y = g(X)$, we deduce from (6.15) that

$$\begin{aligned} E[g(X)] = E(Y) &= \int_0^\infty P(Y > y)\mathrm{d}y - \int_0^\infty P(Y \leq -y)\mathrm{d}y \\ &= \int_0^\infty P(g(X) > y)\mathrm{d}y - \int_0^\infty P(g(X) \leq -y)\mathrm{d}y. \end{aligned} \tag{6.16}$$

We now notice that

$$P(g(X) > y) = P(X \in \{x : g(x) > y\}) = \int_{x: g(x) > y} f(x)\mathrm{d}x$$

and

$$P(g(X) \leq -y) = P(X \in \{x : g(x) \leq -y\}) = \int_{x:g(x)\leq -y} f(x)\mathrm{d}x.$$

Substituting these two expressions into (6.16), we get

$$E[g(X)] = \int_0^{\infty} \int_{x:g(x)>y} f(x)\mathrm{d}x\mathrm{d}y - \int_0^{\infty} \int_{x:g(x)\leq -y} f(x)\mathrm{d}x\mathrm{d}y.$$

We interchange the order of integration to obtain

$$\begin{aligned} E[g(X)] &= \int_{x:g(x)>0} \int_0^{g(x)} f(x)\mathrm{d}y\mathrm{d}x - \int_{x:g(x)\leq 0} \int_0^{-g(x)} f(x)\mathrm{d}y\mathrm{d}x \\ &= \int_{x:g(x)>0} \left(\int_0^{g(x)} \mathrm{d}y \right) f(x)\mathrm{d}x - \int_{x:g(x)\leq 0} \left(\int_0^{-g(x)} \mathrm{d}y \right) f(x)\mathrm{d}x \\ &= \int_{x:g(x)>0} g(x)f(x)\mathrm{d}x - \int_{x:g(x)\leq 0} (-g(x))f(x)\mathrm{d}x. \end{aligned}$$

This yields the desired result that

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x.$$

EXERCISES

Group A

1. Show that there does not exist a random variable X having mean μ and variance σ^2 such that
$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 0.80.$$

2. Prove that there does not exist a positive random variable X with $E(X) = \mu$ for which we have
$$P(X \geq 3\mu) = 0.5.$$

3. The running time in minutes, X, of a computer program on a PC has a distribution function
$$F(t) = 1 - (3t + 1)\mathrm{e}^{-3t}, \quad t \geq 0.$$

   (i) John has started running this program half a minute ago and it is still running. What is the probability that it will stop within the next 15 seconds?

   (ii) Obtain the expected value of the running time, $E(X)$, in two ways:

   (a) by first finding the density function of X;

   (b) by using the result of Proposition 6.7.

4. The maximum daily rainfall, measured in inches, during a year for a particular city is represented by a random variable X having density function
$$f(x) = \begin{cases} 81x^{-4}, & x \geq 3, \\ 0 & x < 3 \end{cases}$$
(this is an example of the Pareto distribution discussed in Examples 6.6 and 6.9).

(i) Let the function $S(t)$ be defined by the relationship

$$S(t) = P(X \geq t), \quad t \geq 0.$$

Obtain the value $S(t)$ for $t = 1,\ 3,\ 10,\ 15$.

(ii) For the same values of t as in Part (i), calculate the percentage error using Markov's inequality, i.e. if $S_M(t)$ is an upper bound for $S(t)$, find the values of the quantity

$$\frac{S_M(t) - S(t)}{S(t)} \cdot 100.$$

5. Assume that X is a nonnegative random variable having distribution function F. Arguing as in the proof of Proposition 6.7, show that the moment of order r of X (around zero) can be expressed in the form

$$E(X^r) = r \int_0^\infty x^{r-1}(1 - F(x))\mathrm{d}x.$$

Group B

6. The time, in seconds, that a butterfly sits on a leaf of a tree is a random variable X having distribution function

$$F(t) = 1 - \frac{2}{3}\mathrm{e}^{-2t} - \frac{1}{3}\mathrm{e}^{-4t}, \quad t \geq 0.$$

(i) Verify that F is a proper distribution function of a random variable.

(ii) Calculate the probabilities

$$P(X \geq 2), \quad P(1 \leq X \leq 2), \quad P(X \leq 2 | X \geq 1).$$

(iii) Calculate the expected value for a butterfly's visit on a leaf.

(iv) Find the variance of X

(a) by finding the density function of X;

(b) directly from the distribution function $F(t)$ given, by using the result of the previous exercise

and check that the two results coincide.

7. Let X be a continuous variable for which the expectation $E(X)$ exists. Show that the following holds:

$$\sum_{n=1}^{\infty} P(|X| \geq n) \leq E(|X|) \leq 1 + \sum_{n=1}^{\infty} P(|X| \geq n).$$

(*Hint*: Making use of Proposition 6.7 for the nonnegative variable $Y = |X|$, express the expected value of Y as

$$E(Y) = \int_0^\infty P(Y > t)\mathrm{d}t = \sum_{n=0}^{\infty} \int_n^{n+1} P(Y > t)\mathrm{d}t.$$

Then observe that for $t \in [n, n+1)$, we have

$$P(Y \geq n+1) \leq P(Y > t) \leq P(Y \geq n).)$$

8. For the random variable X, suppose that we have $E(X) = 5$ and $E(X^2) = 30$. Find
   (i) an upper bound for the probability
   $$P(X \leq -1 \text{ or } X \geq 11);$$
   (ii) an upper bound for the probability $P(X \leq -1)$;
   (iii) a lower bound for the probability $P(1 \leq X \leq 9)$.

6.5 MIXED DISTRIBUTIONS

Discrete random variables have probability functions while continuous variables have density functions. A random variable X, which has a mixed distribution, has both a discrete part and a continuous part and, as a result, it has both a probability function, say $f_1(x)$, and a density function, $f_2(x)$. If X takes any of the values in a set R_{d} with positive probability, i.e. if

$$P(X = x) > 0, \quad \text{for any } x \in R_{\mathrm{d}},$$

then we say that the discrete part of X has range R_{d}, while if X has a density on the set R_{c} (which could, for example, be an interval or a union of intervals) then we say that the continuous part of X has range R_{c}, and the overall range of X is $R_X = R_{\mathrm{d}} \cup R_{\mathrm{c}}$.

It is clear that, in analogy to the purely discrete and purely continuous cases, we must have

$$\sum_{x \in R_{\mathrm{d}}} f_1(x) + \int_{R_{\mathrm{c}}} f_2(x)\mathrm{d}x = 1,$$

or, in a rather more easy-to-remember form,

$$\sum_{\text{discrete part}} f_1(x) + \int_{\text{continuous part}} f_2(x)\mathrm{d}x = 1. \tag{6.17}$$

Example 6.14 An insurance company A, in order to avoid paying excessive amounts to its customers, makes the following deal[1] with a larger insurance company, B. In case that a claim arrives in excess of \$25 000, company A pays 25 000 to the customer and company B pays the remaining amount. The amount X that company A pays per customer is therefore a random variable such that $P(0 \leq X \leq 25\,000) = 1$.

Suppose that the distribution F of X is given by

$$F(t) = \begin{cases} 0, & t < 0, \\ 0.6 + 0.3\left[1 - \left(1 - \dfrac{t}{25\,000}\right)^2\right], & 0 \leq t < 25\,000, \\ 1, & t \geq 25\,000. \end{cases}$$

[1] In the insurance industry, this type of deal is known as excess of loss reinsurance.

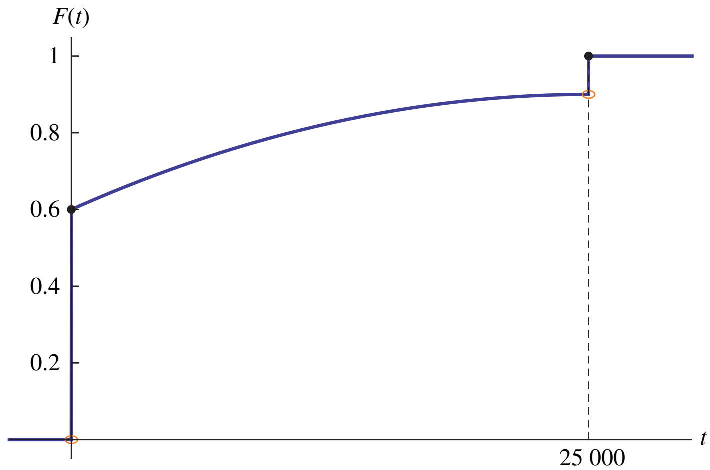


Figure 6.7 The distribution function of X in Example 6.14.

As can also be seen from Figure 6.7, F is suitable as a distribution function for a random variable, since it is increasing, right continuous and it satisfies the conditions

$$\lim_{t\to-\infty} F(t) = 0, \qquad \lim_{t\to\infty} F(t) = 1.$$

Using the formulas from Table 4.1, we verify that each of the events $\{X = 0\}$ and $\{X = 25\,000\}$ has positive probability. Specifically, we have

$$P(X = 0) = F(0) - F(0-) = 0.6 - 0 = 0.6$$

and

$$P(X = 25\,000) = F(25\,000) - F(25\,000) = 1 - 0.9 = 0.1,$$

respectively. This means that 60% of the customers make no claims while 10% of the customers make a claim larger than \$25 000 (we assume implicitly that no customer can make more than one claim for the period under consideration).

For $x \in (0, 25\,000)$, it is apparent from above that

$$P(X = x) = F(x) - F(x-) = 0.$$

It follows that F is a continuous distribution over the interval $(0, 25\,000)$ (continuous part), and it has jumps at the points $x = 0$ and $x = 25\,000$ (discrete part). In real terms, the jump at $x = 25\,000$ is explained by the deal made between the two companies, A and B, while the jump at $x = 0$ simply means that there is a positive probability for a customer to make no claims.

Next, in order to handle the continuous part, we calculate the density function of X in $(0, 25\,000)$. This is

$$f(x) = F'(x) = \frac{3}{125\,000}\left(1 - \frac{x}{25\,000}\right), \qquad 0 < x < 25\,000.$$

In cases where the distribution of a variable X possesses both a continuous and a discrete part, as in the present example, we can calculate probabilities associated with X

(i) directly using the distribution function F, if this is known,

(ii) using the probability density of X at intervals (or unions of intervals) where X is continuous, and the probability function of X at the points where X has jumps.

Let us illustrate the above by using the distribution function F of the present example.

(a) Suppose we want to calculate the probability that the company pays between \$10 000 and \$15 000 to a customer. Then this equals

$$F(15\,000) - F(10\,000) = \left\{0.6 + 0.3\left[1 - \left(1 - \frac{15\,000}{25\,000}\right)^2\right]\right\}$$
$$- \left\{0.6 + 0.3\left[1 - \left(1 - \frac{10\,000}{25\,000}\right)^2\right]\right\} = 0.06,$$

i.e. the probability is 6%. Alternatively, and since F is continuous in the interval $[10\,000, 15\,000]$, we may use the probability density function, as follows:

$$\int_{10\,000}^{15\,000} f(x)\mathrm{d}x = \int_{10\,000}^{15\,000} \frac{3}{125\,000}\left(1 - \frac{x}{25\,000}\right)\mathrm{d}x$$
$$= \frac{3}{125\,000}\left\{\left[x - \frac{x^2}{50\,000}\right]_{10\,000}^{15\,000}\right\} = 0.06.$$

(b) Assume now we are interested in the probability that the company pays at most \$15 000 to a customer. This can be found as follows:

$$P(X \le 15\,000) = F(15\,000) = 0.6 + 0.3\left[1 - \left(1 - \frac{15\,000}{25\,000}\right)^2\right]$$
$$= 0.852 = 85.2\%.$$

Alternatively, we separate the discrete and continuous parts of X in the interval $[0, 15\,000]$ and we have

$$P(X \le 15\,000) = P(X = 0) + P(0 < X \le 15\,000)$$
$$= 0.6 + \int_0^{15\,000} f(x)\mathrm{d}x = 0.852,$$

after some straightforward calculations which you may verify.

Trying to put some of the ideas we have used in the last example into a general framework, suppose we have a random variable X that can take each of the (finitely or countably many) values on a set $R_{\mathrm{d}} = \{x_1, x_2, \ldots, x_n, \ldots\}$ with positive probability; more specifically, let

$$P(X = x_i) = f_1(x_i) > 0, \quad \text{for } x_i \in R_{\mathrm{d}},$$

so that the total "probability mass" for the discrete part is

$$\sum_{i:x_i \in R_{\mathrm{d}}} f_1(x_i) = 1 - p < 1.$$

Assume further that X has a continuous part, with range R_{c} (which may be an interval or a union of intervals) and with a density function f_2 for that part. Then, for X to be a proper random variable, we must have

$$\int_{-\infty}^{\infty} f_2(x)\mathrm{d}x = p,$$

so that the probability that X takes any value is 1. Moreover, the probability $P(a \le X \le b)$ can be found by

$$P(a \le X \le b) = \sum_{i:x_i \in R_{\mathrm{d}}, a \le x_i \le b} f_1(x_i) + \int_a^b f_2(x)\mathrm{d}x.$$

Similar formulas hold in the case when we have strict inequalities inside the brackets or when $a = -\infty$ or $b = \infty$.

For the expectation, an easy-to-remember formula is

$$E(X) = \sum_{\text{discrete part}} x f_1(x) + \int_{\text{continuous part}} x f_2(x)\mathrm{d}x$$

or, in a slightly more formal fashion,

$$E(X) = \sum_{i:x_i \in R_{\mathrm{d}}} x_i f_1(x_i) + \int_{R_{\mathrm{c}}} x f_2(x)\mathrm{d}x.$$

Note that the range of integration on the right may safely be taken to be the entire range of values for X, or even the whole real line, since R_{d} contains at most a countable number of points, so that the value of the integral is not affected.

More generally, for a real-valued function g, we have

$$E[g(X)] = \sum_{i:x_i \in R_{\mathrm{d}}} g(x_i) f_1(x_i) + \int_{-\infty}^{\infty} g(x) f_2(x)\mathrm{d}x.$$

Example 6.15 A random variable X has a mixed distribution whose discrete and continuous parts, $f_1(x)$ and $f_2(x)$, are described, respectively, by

$$f_1(x) = \begin{cases} \dfrac{1}{10}, & x = 4, \\ \dfrac{1}{10}, & x = 8, \end{cases}$$

and

$$f_2(x) = \begin{cases} \dfrac{1}{12}, & 0 < x < 4, \\ \dfrac{1}{15}, & 8 < x \le 15. \end{cases}$$

(i) What is the range of values for X?

(ii) Calculate the probability $P(2 < X < 9)$.

(iii) Find the expectation and variance of X.

SOLUTION

(i) The range for the discrete part of X is the set $R_{\mathrm{d}} = \{4, 8\}$, while that for the continuous part, as can be seen from the formula for f_2 above, is the union of the intervals $(0, 4)$ and $(8, 15]$. The range of X is thus the union $R_{\mathrm{d}} \cup R_{\mathrm{c}}$, which is the set $R_X = (0, 4] \cup [8, 15]$.

(ii) We use formula (6.17), separating the discrete from the continuous part of X, which gives

$$\begin{aligned} P(2 < X < 9) &= P(2 < X < 4) + P(X = 4) + P(4 < X < 8) \\ &\quad + P(X = 8) + P(8 < X < 9) \\ &= \int_2^4 f_2(x)\mathrm{d}x + f_1(4) + \int_4^8 f_2(x)\mathrm{d}x + f_1(8) + \int_8^9 f_2(x)\mathrm{d}x \\ &= \int_2^4 \frac{1}{12}\mathrm{d}x + \frac{1}{10} + 0 + \frac{1}{10} + \int_8^9 \frac{1}{15}\mathrm{d}x \\ &= \frac{1}{12} \cdot (4-2) + \frac{1}{10} + 0 + \frac{1}{10} + \frac{1}{15} \cdot (9-8) = \frac{13}{30}. \end{aligned}$$

(iii) In order to find the expectation of X, we have to consider again separately the discrete and the continuous parts of the distribution. More specifically, we have

$$\begin{aligned} E(X) &= \sum_{\text{discrete part}} x f_1(x) + \int_{\text{continuous part}} x f_2(x)\mathrm{d}x \\ &= 4 \cdot \frac{1}{10} + 8 \cdot \frac{1}{10} + \int_0^4 \frac{x}{12}\,\mathrm{d}x + \int_8^{15} \frac{x}{15}\,\mathrm{d}x \\ &= \frac{2}{5} + \frac{4}{5} + \frac{1}{12} \cdot \left[\frac{x^2}{2}\right]_0^4 + \frac{1}{15} \cdot \left[\frac{x^2}{2}\right]_8^{15} \\ &= \frac{2}{5} + \frac{4}{5} + \frac{8}{12} + \frac{15^2 - 8^2}{2 \cdot 15} = \frac{217}{30} \cong 7.23. \end{aligned}$$

For the variance of X, in a similar way, we first find $E(X^2)$ as

$$E(X^2) = \sum_{\text{discrete part}} x^2 f_1(x) + \int_{\text{continuous part}} x^2 f_2(x)\mathrm{d}x$$

$$= 4^2 \cdot \frac{1}{10} + 8^2 \cdot \frac{1}{10} + \int_0^4 \frac{x^2}{12}\mathrm{d}x + \int_8^{15} \frac{x^2}{15}\mathrm{d}x$$

$$= \frac{8}{5} + \frac{32}{5} + \frac{1}{12} \cdot \left[\frac{x^3}{3}\right]_0^4 + \frac{1}{15} \cdot \left[\frac{x^3}{3}\right]_8^{15} = \frac{367}{5} = 73.4.$$

We therefore find the variance of X to be

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = 73.4 - (7.23)^2 = 21.08.$$

The formula $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$, which we used for Part (iii) of the last example, as we have seen, applies to random variables X which are either discrete or continuous. However, although not explicitly mentioned so far, it is equally valid when the distribution of X is a mixed distribution (so that it holds for *any* random variable X). In fact, the proof of the identity we gave in Proposition 4.9 does not make any use of the nature of X.

EXERCISES

Group A

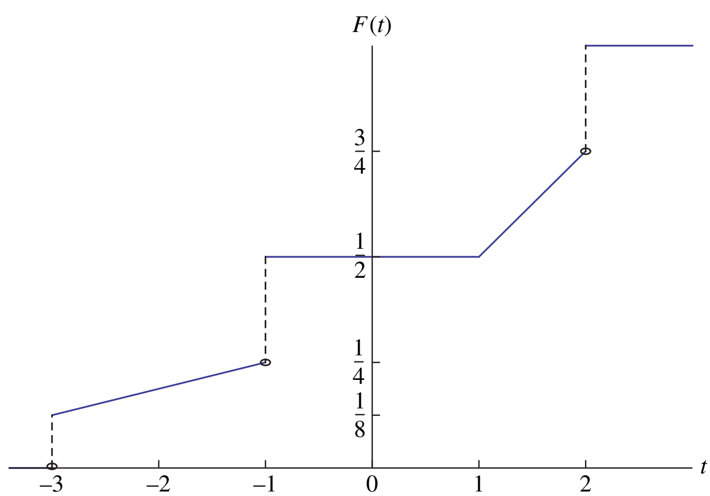
1. Using the following diagram, which gives the distribution function of a random variable X, calculate the probabilities

   (i) $P(X < -2)$;

   (ii) $P(X = i)$, for $i = -3, -1, 1, 2$;

   (iii) $P(X \leq -1)$;

   (iv) $P(-2 < X < 2)$.



2. When a car passes through a traffic junction, the delay (in minutes) caused by the red light in the traffic lights is a random variable X with distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 - 0.75e^{-t}, & t \geq 0. \end{cases}$$

   (i) Explain why this is a mixed distribution and find the probability that a car has no delay in passing through the traffic lights of the junction.

   (ii) For a car passing through the traffic lights, what is the probability that it is delayed by

      (a) less than 60 seconds?

      (b) at least 45 but no more than 90 seconds?

      (c) more than 120 seconds?

   (iii) What is the expected delay caused by the red light in the junction?

3. A random variable X has a jump at the point $x = 2$, while its continuous part is described by the density

$$f(x) = \frac{x^2 - 2x}{10}, \quad 2 < x \leq 4.$$

   (i) What is the value of the probability $P(X = 2)$?

   (ii) Find the expected value of X.

4. The distribution function of a random variable X is given by

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 - \left(\dfrac{a}{a+t}\right)^k, & 0 \leq t < a, \\ 1 & t \geq a, \end{cases}$$

   for some positive constants a and k.

   (i) Find the value of the probabilities $P(X = 0)$ and $P(X = a)$.

   (ii) For what value of k, we have $P(X = a) = 1/4$?

   (iii) For $a = 4, k = 2$, obtain the value of t such that

$$P(X \leq t) = \frac{5}{9}.$$

   (Distributions of this form occur frequently in insurance modeling [the scientific area where such models are studied and used is known as actuarial science], in particular in excess of loss reinsurance; see the footnote earlier in this section.)

5. The distribution function of a random variable X is

$$F(t) = \begin{cases} 0, & t < 0, \\ \dfrac{1}{16}(t^2 + 2t + 5), & 0 \leq t < 1, \\ \dfrac{1}{8}(t + 5), & 1 \leq t < 3, \\ 1, & t \geq 3. \end{cases}$$

(i) Which are the points of discontinuity for F?

(ii) Calculate the probabilities $P(0 < X < 1/2)$ and $P(1 \le X < 2)$.

(iii) Find the expected value and the variance of X.

6. A random variable X has discrete range $R_d = \{1, 3\}$, while its continuous range is the open interval $(1, 3)$, with density in that interval given by

$$f_2(x) = \frac{x-1}{3}, \quad 1 < x < 3.$$

If it is known that $E(X) = 7/3$, calculate the probabilities

$$P(X = 1), \quad P(X = 3).$$

7. A random variable X has a mixed distribution whose discrete and continuous parts, $f_1(x)$ and $f_2(x)$, are given, respectively, by

$$f_1(x) = \begin{cases} \frac{1}{3}, & x = 0, \\ \frac{1}{6}, & x = 4, \end{cases}$$

and

$$f_2(x) = \begin{cases} \frac{c(x+1)}{6}, & 0 < x < 4, \\ 0, & \text{elsewhere,} \end{cases}$$

for some suitable constant c.

(i) What is the range of values for X?

(ii) Find the value of the constant c.

(iii) Calculate the probability $P(2 \le X \le 4)$.

(iv) Obtain the expectation and variance of X.

Group B

8. An insurance company classifies the claims it receives as being either small (if they are up to \$5000) or large. During a calendar year, an insured customer may either make one claim or none. It has been estimated that 85% of the company's customers make no claims during a year, 10% make a small claim and the rest make a large claim. If Y denotes the size (in thousands of dollars) of a small claim that arrives at the company, Y has distribution function

$$F_Y(t) = \begin{cases} \frac{t}{5}, & 0 \le t < 5, \\ 0, & \text{otherwise,} \end{cases}$$

while if Z denotes the size of a large claim, in thousands of dollars, the distribution function of Z is given by

$$F_Z(t) = \begin{cases} 1 - \frac{125}{t^3}, & t \ge 5, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the distribution function of the amount X that the company pays during a year to a randomly selected customer. Identify the discrete and the continuous parts of that distribution.

(ii) What is the expected amount the company pays per customer each year?

9. Let F_X be the distribution function and f_X be the density function of a variable X for which we know that $P(X \geq a) = 1$, where $a \in \mathbb{R}$ is a given constant. We define the function

$$F(t) = \begin{cases} 0, & t < a, \\ 1 - p(1 - F_X(t)), & t \geq a, \end{cases}$$

where $0 < p < 1$.

(i) Verify that F is the distribution function of a random variable Y, which is neither (purely) discrete nor continuous, and which takes the value a with probability $1 - p$. Express the density function of Y in terms of f_X.

(ii) Derive an expression for the expected value of Y in terms of $E(X)$.

(iii) Write F in the form

$$F = (1 - p)F_1 + pF_2,$$

where F_1 and F_2 are suitable distribution functions of two random variables.

10. A random variable X has distribution function

$$F(t) = \begin{cases} 1 - e^{-\lambda t}, & 0 \leq t < m, \\ 1, & t \geq m, \end{cases}$$

where λ and m are positive constants.

(i) Identify the point(s) at which F has jumps.

(ii) Find the density function associated with the continuous part of this distribution.

(iii) Show that the expectation of X is

$$E(X) = \frac{1}{\lambda}(1 - e^{-\lambda m}).$$

11. An insurance contract with deductible amount a and retention level b (with $0 < a < b$) entails the following agreement: Let Y be the size of the loss incurred to the insured. If $Y \leq a$, the customer pays the full amount of this loss, while if $a < Y \leq b$, the customer is responsible for the amount a and the company pays the difference $Y - a$. Finally, for $Y > b$, the company pays the amount $b - a$. The size of the customer's loss Y is a continuous random variable with distribution function F_Y and density f_Y.

(i) Verify that the distribution function F of the amount X that the company pays for a claim is given by the expression

$$F(t) = \begin{cases} 0, & t < 0, \\ F_Y(a + t), & 0 \leq t < b - a, \\ 1, & t \geq b - a. \end{cases}$$

(ii) Show that the density, f_2, for the continuous part of F is

$$f_2(x) = f(a+x), \quad 0 < x < b-a,$$

while the probability function associated with the discrete part of F is given by

$$f_1(0) = F_Y(a), \quad f_1(b-a) = 1 - F_Y(b).$$

(iii) Prove that

$$E(X) = (b-a)(1-F_Y(b)) + \int_a^b y f_Y(y)\mathrm{d}y.$$

(iv) Give numerical answers to Parts (ii) and (iii) when Y has the density

$$f_Y(y) = \frac{3}{1000}\left(\frac{1000}{1000+y}\right)^4, \quad y > 0,$$

while a and b can be determined by knowing that

$$a = \frac{2}{5}E(Y), \quad b = 2E(Y).$$

6.6 BASIC CONCEPTS AND FORMULAS

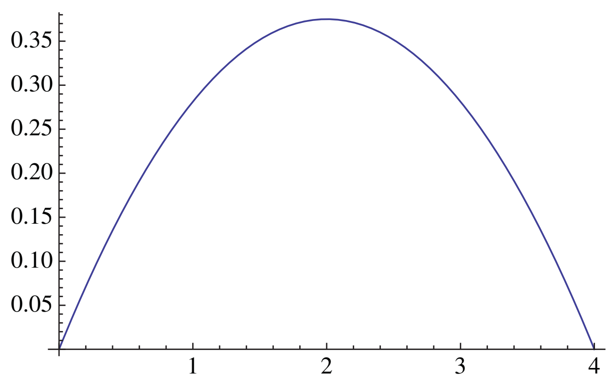
Probability density function of a continuous random variable X	A nonnegative real function $f : \mathbb{R} \mapsto [0,\infty)$ such that $P(X \in A) = \int_A f(x)\mathrm{d}x$
Properties of a probability density function	• $f(x) \geq 0$ for any $x \in \mathbb{R}$; • $\int_{-\infty}^{\infty} f(x)\mathrm{d}x = 1$
Relation between a density function and the corresponding distribution function	$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x$; $f(x) = F'(x)$ at all continuity points x of f
Formulas for calculating probabilities for continuous random variables	$P(a < X < b) = P(a < X \leq b) = P(a \leq X < b) = P(a \leq X \leq b) = \int_a^b f(x)\mathrm{d}x$
Density of a function $Y = g(X)$ of the continuous random variable X	$f_Y(y) = \lvert (g^{-1})'(y) \rvert f(g^{-1}(y))$, provided that g is a strictly monotone function
Expectation of X	$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\mathrm{d}x$
Expectation of a function $Y = g(X)$ of the continuous random variable X	$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)\mathrm{d}x$
Properties of expectation	• $E\left[\sum_{i=1}^{k} \lambda_i g_i(X)\right] = \sum_{i=1}^{k} \lambda_i E[g_i(X)]$ • $E(aX+b) = aE(X)+b, \quad a,b \in \mathbb{R}$

Variance of X	$\sigma^2 = \text{Var}(X) = E[(X-\mu)^2]$
Standard deviation of X	$\sigma = \sqrt{\text{Var}(X)}$
Alternative formula for calculating the variance	$\text{Var}(X) = E(X^2) - [E(X)]^2$
Properties of variance	• If $\text{Var}(X) = 0$, then X is a constant; • $\text{Var}(aX+b) = a^2\text{Var}(X), \quad a, b \in \mathbb{R}$
Moment of order r around zero	$\mu_r' = E(X^r), \quad r = 1, 2, \ldots$
Factorial moment of order r	$\mu_{(r)} = E[(X)_r] = E[X(X-1)\cdots(X-r+1)]$, $r = 1, 2, \ldots$
Absolute moment of order r	$E(\lvert X\rvert^r), \quad r = 1, 2, \ldots$
Moment of order r around c	$E[(X-c)^r)], \quad r = 1, 2, \ldots$
Central moment of order r	$\mu_r = E[(X-\mu)^r)], \quad r = 1, 2, \ldots$
Markov's inequality	$P(X \geq t) \leq \dfrac{E(X)}{t}$, where X is a nonnegative variable and $t > 0$
Chebyshev's inequality	$P(\lvert X-\mu\rvert \geq t) \leq \dfrac{\sigma^2}{t^2}$, where $\mu = E(X), \sigma^2 = \text{Var}(X)$, and $t > 0$
Alternative formulas for calculating the expectation	$E(X) = \int_0^\infty (1-F(x))\mathrm{d}x - \int_{-\infty}^0 F(x)\mathrm{d}x$ $= \int_0^\infty (1-F(x)-F(-x))\mathrm{d}x$ $= \int_0^\infty P(X > x)\mathrm{d}x - \int_0^\infty P(X \leq -x)\mathrm{d}x$
Random variable X with a mixed distribution: this has a probability function f_1 on $A = \{x_1, x_2, \ldots\}$ and a density function f_2	$\sum_{i:x_i \in A} f_1(x_i) = \sum_{i:x_i \in A} P(X = x_i) = 1-p$ and $\int_{-\infty}^\infty f_2(x)\mathrm{d}x = p$, where $p \in (0,1)$
Calculation of probabilities for random variables with a mixed distribution	$P(a \leq X \leq b) = \sum\limits_{i:x_i \in [a,b]} f_1(x_i) + \int_a^b f_2(x)\mathrm{d}x$
Expected value for a random variable X with a discrete part f_1 on $A = \{x_1, x_2, \ldots\}$ and a continuous part f_2	$E(X) = \sum\limits_{i:x_i \in A} x_i f_1(x_i) + \int_{-\infty}^\infty x f_2(x)\mathrm{d}x$

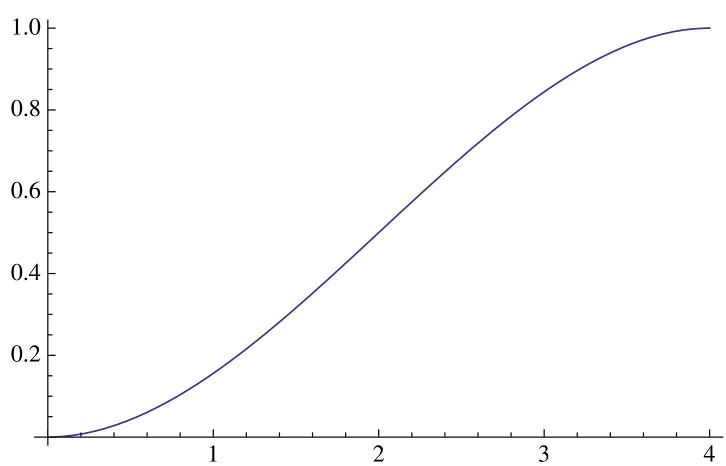
6.7 COMPUTATIONAL EXERCISES

In the following program, we study the distribution of the random variable X described in Example 6.1. More specifically, we calculate the value of the constant c and then the probabilities $P(1 \leq X < 3)$ and $P(X > 2)$. In addition, we plot the density function and the distribution function of X and, finally, we compute the expected value and the variance of X.

```
In[1]:= f1[x_]:=If[(x>0 && x<=4), 4x-x^2,0] ;
        a=Integrate[f1[x],{x,-Infinity,Infinity}];
        Print["The constant c should be equal to"]
        c=1/a
        Print["Plot of probability density function"]
        f[x_]:= c*f1[x];Plot[f[x],{x,0,4}]
        Print["Plot of cumulative distribution function"]
        F[t_]:=Integrate[c*(4x-x^2),{x,0,t}];
        Plot[F[t],{t,0,4}]
        Print["P[1<=X<3] = ",F[3]-F[1]];
        Print["P[X>2] = ",1-F[2]];
        mu=Integrate[x*f[x],{x,0,4}];
        Print["Mean = ", mu]
        Print["Variance = ",Integrate[x^2*f[x],{x,0,4}] -mu^2]

        The constant c should be equal to
Out[4]= 3/32
        Plot of probability density function
Out[6]=
```



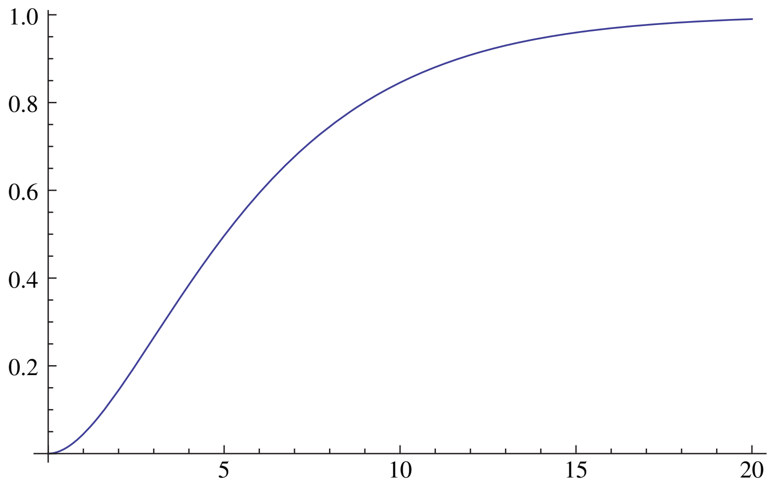

```
Out[9]=Plot of cumulative distribution function
```



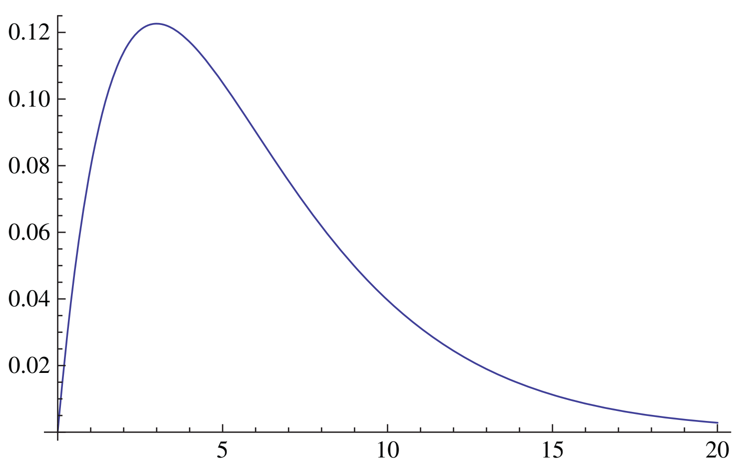

```
P[1<=X<3] = 11/16
P[X>2] = 1/2
Mean = 2
Variance = 4/5
```

For the solution of Example 6.2 and the construction of the relevant plots, we use the following sequence of commands:

```
In[1]:= F1[t_]:=Exp[-t/3]+(t/3)*Exp[-t/3]
        c=1-Limit[F1[t],t->Infinity];
        F[t_]:=c-F1[t]
        f[x_]:=F'[x]
        Print["Plot of cumulative distribution function F(t) = ",F[t]]
        Plot[F[t],{t,0,20}]
        Print["Plot of probability density function f(x) = ", f[x]]
        Plot[f[x],{x,0,20}]
        Print["P[X<3] = ",F[3]," = ", N[F[3]]];
        Print["P[1<X<4] = ",F[4]-F[1]," = ", N[F[4]-F[1]]]

        Plot of cumulative distribution function
              F(t) = 1-E^(-t/3)-1/3 E^(-t/3) t
Out[6]=
```




```
Plot of probability density function f(x) = 1/9 E^(-x/3) x

Out[8]=
```




```
P[X<3] = 1-2/E = 0.264\,241
P[1<X<4] = -(7/(3 E^(4/3)))+4/(3 E^(1/3)) = 0.340\,315
```

Using a similar set of commands, solve each of the following problems:

1. The lifetime of an electrical appliance (in thousands of hours) is a random variable X with density function

$$f(x) = \begin{cases} 0, & x < 0, \\ 0.5 \cdot (x/20)^9 \cdot \mathrm{e}^{-(x/20)^{10}}, & x \geq 0. \end{cases}$$

   (i) Find the distribution function of X.

   (ii) Find the proportions of appliances of this type with a lifetime

      (a) more than 20 000 hours;

      (b) at most 20 500 hours.

   (iii) If it is known that an appliance has operated for 20 000 hours, what is the probability that it will break down within the next 500 hours? Compare this probability with the probability that a *new appliance* will break down in the first 500 hours of its operation.

   (iv) Find the expectation and the variance of X.

2. Draw a graph of the density functions for the distributions given in Exercise 21 of Section 6.1. Hence find, in each case, the point a around which the density is symmetric. Moreover, verify numerically that for each of these distribution functions, we have

$$F(a - x) + F(a + x) = 1$$

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

3. Suppose that X is a continuous random variable with distribution function

$$F(t) = \begin{cases} 0, & t < 2, \\ 0.5t - 1, & 2 \leq t \leq 4, \\ 1, & t > 4. \end{cases}$$

   Draw a graph of the *difference* between the exact value of the probability

$$P(|X - 3| \geq t), \quad t > 0,$$

   and the upper bound for this that we get from Chebyshev's inequality, for different values of t.

4. Find the value of the constant $c \in \mathbb{R}$ for which the function f defined by

$$f(x) = \begin{cases} c \cdot \dfrac{2 + \cos(\sqrt{x})}{\mathrm{e}^{x/3}}, & 0 \leq x \leq \pi^2, \\ 0, & \text{elsewhere}, \end{cases}$$

   is a density function for a continuous random variable X. Using this value of c, find the mean and the variance of X.

5. The value of an investment at the end of a certain period is described by a random variable X with density function

$$f(x) = \begin{cases} \dfrac{1}{x\sqrt{2\pi}} e^{-(\ln x - 4)^2/2}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

(i) Find the probability that, at the end of the period, the value of the investment is

(a) at least 35 monetary units;

(b) at most 50 monetary units;

(c) between 60 and 75 monetary units.

(ii) Calculate the "median" value, that is, the value m for which we have

$$P(X \le m) = \frac{1}{2}.$$

(iii) Find the expected value for this investment and compare your answer to that found in Part (ii).

6. The density function of a random variable X is given by

$$f(x) = c \cdot e^{-|x|}, \quad -\infty < x < \infty,$$

for a real constant c.

(i) Find the value of c and draw the graphs of the density function and the distribution function of the variable X.

(ii) Obtain the expectation $\mu = E(X)$ and the variance $\mathrm{Var}(X)$.

(iii) Calculate the following moments of order three for the variable X:

$$E[X^3], \quad E[|X|^3], \quad E[(X-\mu)^3].$$

7. After reading the applications section of this chapter (Section 6.10), draw a graph of the mean monthly profit, as a function of the ordered quantity z, and find the value of z that maximizes this profit, in each of the following cases:

(i) $\alpha = 10, \beta = 5, \quad f(x) = \dfrac{3x^2}{8} e^{-x^3/8}, \quad x > 0;$

(ii) $\alpha = 100, \beta = 5, \quad f(x) = \dfrac{3x^2}{8} e^{-x^3/8}, \quad x > 0;$

(iii) $\alpha = 10, \beta = 5, \quad f(x) = \dfrac{1}{3 \cdot 2^8} x^4 e^{-0.5x}, \quad x > 0;$

(iv) $\alpha = 100, \beta = 5, \quad f(x) = \dfrac{1}{3 \cdot 2^8} x^4 e^{-0.5x}, \quad x > 0.$

6.8 SELF-ASSESSMENT EXERCISES

6.8.1 True–False Questions

1. A probability density function can take positive, zero, and negative values.
2. A continuous variable X has density function

$$f(x) = 2c + 3, \quad 2 \le x \le 6.$$

Then the value of c is $1/2$.

3. The density function of a random variable is always a decreasing function.
4. If a continuous random variable X with density function f and distribution function F takes only positive values then, for any $x > 0$, we have

$$F(x) = \int_0^x f(y)\mathrm{d}y.$$

5. For a continuous random variable X with distribution function F, we have $F'(x) = P(X = x)$.
6. The function

$$f(x) = 3(4x - x^2)/8, \quad 0 \le x \le 4,$$

can be the density function of a continuous random variable.

7. Let f be a density function of a random variable X. Then, we have $\lim_{x\to\infty} f(x) = 1$.
8. Let X be a continuous variable with density

$$f_X(x) = (2x + 1)/6, \quad 0 < x < 2.$$

Then the density function of $Y = X^2$ is

$$f_Y(y) = (2y + 1)^2/36, \quad 0 < y < 4.$$

9. A random variable X has density

$$f(x) = \frac{2(3x - x^2)}{9}, \quad 0 < x < 3.$$

Then the probability $P(1 < X < 2)$ equals $13/27$.

10. If a continuous random variable X has distribution function F, then for any a, b,

$$P(a < X < b) = F(b) - F(a).$$

11. The distribution function of X is given by

$$F(x) = \begin{cases} 0, & x < 0 \\ (x^2+1)/5, & 0 \le x < 2 \\ 1, & x \ge 2. \end{cases}$$

Then X has a mixed distribution.

12. If X has a mixed distribution, then its range contains infinitely many points.

13. If the distribution of X has a jump at the point a, then $P(X = a) = 0$.

14. The density function of X is

$$f(x) = 32x^{-3}, \quad x \ge 4.$$

The expected value of X is 2.

15. The manufacturing time, X, for an item produced by a machine has distribution function

$$F(t) = 1 - \mathrm{e}^{-2t^2}, \quad t \ge 0.$$

Then, the probability $P(X > 2)$ equals e^{-8}.

16. A random variable has probability density

$$f(x) = x^3/4, \quad 0 \le x \le 2.$$

The variance of X is $8/75$.

17. Let X be a continuous variable whose range is the entire real line and which has distribution function F_X, and let $Y = |X|$. Then for the distribution function, F_Y, of Y, we have

$$F_Y(y) = F_X(y) - F_X(-y), \quad y \ge 0.$$

18. Let X be a random variable with density function

$$f(x) = \frac{1}{3}, \quad 0 \le x \le 3.$$

Then we have $P(X^2 \le 1) = 1/3$.

19. If X has distribution function

$$F(t) = \begin{cases} 0, & t < 2, \\ (t-2)/4, & 2 \le t < 6, \\ 1, & t \ge 6, \end{cases}$$

then $E(X) = 5$.

6.8.2 Multiple Choice Questions

1. The density function of a random variable X is

$$f(x) = x\mathrm{e}^{-x}, \quad x > 0.$$

The distribution function of X (for $x > 0$) is

(a) $1 - \mathrm{e}^{-x}$ (b) $1 - x\mathrm{e}^{-x}$ (c) $1 - x - \mathrm{e}^{-x}$
(d) $1 - (1 + x)\mathrm{e}^{-x}$ (e) $1 - x^2\mathrm{e}^{-x}$

2. The density function of X is

$$f(x) = \frac{c}{(x+3)^2}, \quad 0 \le x \le 3.$$

Then the value of c is

(a) 2 (b) 6 (c) 3 (d) 1/6 (e) 1/2

3. A random variable X has distribution function

$$F(t) = \begin{cases} 0, & t < -2, \\ (t+2)/8, & -2 \le t < 1, \\ 5/8, & 1 \le t < 3, \\ 1, & t \ge 3. \end{cases}$$

Then the probability $P(-1 \le X < 3)$ equals

(a) $\frac{5}{8}$ (b) $\frac{3}{8}$ (c) $\frac{1}{2}$ (d) $\frac{7}{8}$ (e) $\frac{3}{4}$

4. A continuous variable X has density function

$$f(x) = 3x^2, \quad 0 \le x \le b.$$

The value of b is

(a) 1 (b) 1/3 (c) 3 (d) $\sqrt[3]{3}$ (e) $\sqrt[3]{2}$

5. A random variable has probability density function

$$f(x) = c|x|, \quad -2 \le x \le 2.$$

The value of c is

(a) 1 (b) 2 (c) 4 (d) 1/2 (e) 1/4

6. The density function of a random variable is given by

$$f(x) = \frac{2x(c+1)}{5}, \quad 0 < x < 2.$$

The distribution function of X is

(a) $$F(t) = \begin{cases} 0, & t < 0, \\ t/2, & 0 \leq t < 2, \\ 1, & t \geq 2 \end{cases}$$

(b) $$F(t) = \begin{cases} 0, & t < 0, \\ t^2/2, & 0 \leq t < 2, \\ 1, & t \geq 2 \end{cases}$$

(c) $$F(t) = \begin{cases} 0, & t < 0, \\ t^2/4, & 0 \leq t < 2, \\ 1, & t \geq 2 \end{cases}$$

(d) $$F(t) = \begin{cases} 0, & t < 0, \\ t/4, & 0 \leq t < 2, \\ 1 + t/4, & t \geq 2 \end{cases}$$

(e) $$F(t) = \begin{cases} 0, & t < 0, \\ t/5, & 0 \leq t < 2, \\ 1, & t \geq 2 \end{cases}$$

7. For the random variable X we have $P(X = 0) = 1/4$, while X has a density

$$f(x) = \frac{9}{4}e^{-3x}, \quad x > 0.$$

The expected value of X is

(a) $1/4$ (b) $3/4$ (c) $9/4$ (d) 3 (e) 4

8. The distribution function of the random variable X is

$$F(t) = 1 - e^{-5t^3}, \quad t \geq 0.$$

The density function of X (for $x \geq 0$) is

(a) $f(x) = e^{-5x^2}$ (b) $f(x) = 5e^{-5x^2}$ (c) $f(x) = 15e^{-5x^3}$

(d) $f(x) = 5x^2e^{-5x^3}$ (e) $f(x) = 15x^2e^{-5x^3}$

9. If the random variable X has density

$$f(x) = \begin{cases} x, & 0 \le x < 1, \\ 2 - x, & 1 \le x \le 2, \end{cases}$$

then the conditional probability $P(X \le 3/2 | X \ge 1/2)$ equals

(a) $3/7$ (b) $7/8$ (c) $3/4$ (d) $6/7$ (e) $2/7$

10. A random variable X has distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ t/6, & 0 \le t < 3, \\ 3/5, & 3 \le t < 5, \\ 1, & t \ge 5. \end{cases}$$

Then

(a) F has no jumps (it is a continuous distribution)

(b) F has one jump at the point $t = 5$

(c) F has two jumps at the points $t = 0$ and $t = 3$

(d) $P(X = 3) = 1/10$

(e) $P(X \ge 5) = 1$

11. A random variable X has a discrete range $R_\text{d} = \{1, 2\}$, while its continuous range is the open interval $(1, 2)$ with density

$$f_2(x) = x/6, \quad 1 < x < 2.$$

If it is known that $P(X = 1) = 2P(X = 2)$, then the probability $P(X = 2)$ equals

(a) $\dfrac{1}{8}$ (b) $\dfrac{3}{8}$ (c) $\dfrac{1}{4}$ (d) $\dfrac{1}{6}$ (e) $\dfrac{1}{12}$

12. If X has distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ t/2, & 0 \le t < 2, \\ 1, & t \ge 2, \end{cases}$$

then the standard deviation of X is

(a) $\sqrt{\dfrac{1}{3}}$ (b) $\sqrt{\dfrac{2}{3}}$ (c) 1 (d) $\sqrt{\dfrac{3}{2}}$ (e) $\sqrt{3}$

13. Let X be a continuous variable with distribution function

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 - e^{-2t}(1 + 2t + 2t^2), & t \geq 0. \end{cases}$$

The probability $P(X \geq 2 | X \geq 1)$ equals

(a) $13e^{-2}$ (b) $\dfrac{5}{13}e^{-2}$ (c) $\dfrac{13}{5}e^{-2}$ (d) $\dfrac{5}{13}$ (e) $\dfrac{1}{5}e^{-2}$

14. The monthly income (in thousands of dollars) of employees in a large firm is represented by a continuous variable X having density

$$f(x) = \begin{cases} \dfrac{160}{x^6}, & x \geq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Then the expected value of X is

(a) 5 (b) 10 (c) $15/2$ (d) $5/2$ (e) $5/4$

15. The time X (in minutes) for a customer to be served at a post office has density function

$$f(x) = 2e^{-2x}, \quad x \geq 0.$$

The probability that each of the next three customers will be served within one minute equals

(a) e^{-6} (b) $3e^{-2}$ (c) $2e^{-6}$ (d) $2e^{-3}$ (e) $(1 - e^{-2})^3$

6.9 REVIEW PROBLEMS

1. The density function of a random variable X is given by

$$f(x) = \begin{cases} -\beta x, & -1 < x \leq 1, \\ \alpha e^{-3x}, & x > 1, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the values of α and β if it is known that $E(X) = 1$.

2. The density function of a variable X is given by

$$f(x) = \begin{cases} 24/x^4, & x \geq 2 \\ 0, & x < 2. \end{cases}$$

(i) Find the mean and variance of X.

(ii) Obtain the distribution function, F, of X and use it to calculate the conditional probability $P(X \leq 7|X \geq 4)$.

(iii) Find the probability density, the mean and the variance of the random variable $Y = 1/X$.

(iv) Which, if any, of the following five equations are true for the variables X and Y defined above?

$$\begin{aligned} E(1/X) &= 1/E(X), \\ \mathrm{Var}(1/X) &= 1/\mathrm{Var}(X), \\ E(X+Y) &= E(X)+E(Y), \\ E(X+Y^{-1}) &= 2E(X), \\ E(XY) &= E(X)E(Y). \end{aligned}$$

3. The quantity of petrol, X (in thousands of liters), sold by a gas station daily is a continuous random variable with density function

$$f(x) = \begin{cases} cx, & 0 \leq x < 2, \\ 2c, & 2 \leq x < 4, \\ c(6-x), & 4 \leq x \leq 6. \end{cases}$$

(i) Find the value of the constant c.

(ii) Calculate the probabilities

$$P(X > 2), \quad P(1 < X < 3), \quad P(X \leq 5|X > 1).$$

(iii) Find the expected value and the standard deviation for the quantity of petrol sold daily.

4. A super market sells potato sacks whose weight, X (in kilograms), is a continuous random variable with density function f as

$$f(x) = \begin{cases} |x-2|, & 1 \leq x \leq 3, \\ 0, & \text{elsewhere}. \end{cases}$$

(i) What proportion of sacks will have a weight between 1.6 and 2.4 kg?

(ii) Find the standard deviation for the weight of a potato sack.

(iii) If the price of potatoes per kilogram is \$0.65, calculate the mean and standard deviation of the amount the super market receives for a sack.

5. Let X be a continuous random variable with density function f. Recall that the *median* of the distribution of X is the real number m such that

$$P(X \leq m) = P(X \geq m).$$

Show that, if the distribution of X is symmetric around a point a (see Exercise 21 of Section 6.1), and the expectation $E(X)$ exists, then

$$m = E(X) = a.$$

6. Let X be a continuous random variable with density f, and Y be another variable related to X by $Y = X^3$. Show that the density function of Y is

$$f_Y(y) = \frac{f\left(\sqrt[3]{y}\right)}{3\sqrt[3]{y^2}}.$$

7. Assume that X is a random variable with density function

$$f(x) = \begin{cases} \dfrac{x \sin x}{\pi}, & 0 \leq x < \pi, \\ 0, & \text{otherwise.} \end{cases}$$

Let

$$\mu_r' = E(X^r), \quad r = 1, 2, \ldots,$$

be the moments of X (around zero). Then, verify that these moments satisfy the recursive relationship

$$\mu_{r+1}' = \pi^{r+1} - (r+1)(r+2)\mu_{r-1}', \quad r = 2, 3, \ldots$$

8. The daily orders for a particular product at a factory, in hundreds of kilograms, are represented by a random variable X having density function

$$f(x) = \begin{cases} a(x-1)^2, & 0 \leq x \leq 6, \\ 0, & \text{elsewhere.} \end{cases}$$

(i) Obtain the value of a.

(ii) Find the proportion of days in which the total amount of the product ordered is

(a) at least 450 but less than 550 kg;

(b) at most 500 kg.

(iii) Find the expected amount of the product ordered daily.

9. The pressure P, in lb ft^{-2}, developed at the wings of an aircraft is given by

$$P = 3 \cdot 10^{-3} V^2,$$

where V is the velocity (in miles per hour) of the wind surrounding the wings. The pressure can be considered as a random variable with density function

$$f(p) = \frac{c}{\sqrt{p}}, \quad 0 \le p \le a,$$

for $c > 0$ and $a > 0$.

(i) Find the value of c if it is known that

$$E(P) = 0.1 \text{ lb ft}^{-2}.$$

(ii) Find the distribution function and the expected value of the velocity V.

10. The measurement error of an instrument can be described by a random variable X with distribution function

$$F(t) = \begin{cases} 0, & t < -2, \\ \dfrac{1}{2} + \dfrac{3}{32}\left(4t - \dfrac{t^3}{3}\right), & -2 \le t < 2, \\ 1, & t \ge 2. \end{cases}$$

(i) Calculate the following probabilities

$$P(X < 0), \quad P(X < 1 | X \ge 0), \quad P(|X| < 1).$$

(ii) Find the density function and the expectation of the random variable X.

(iii) Obtain the distribution for the magnitude of the error (i.e. that of the random variable $|X|$) along with the expected value of this magnitude.

11. The percentage concentration, in alcohol, of a medical substance is a continuous variable X with density function

$$f(x) = cx(1-x)^n, \quad 0 < x < 1,$$

where c is a real constant and n is a positive integer.

(i) Find the value of c, as a function of n.

(ii) For $n = 7$, show that the average content in alcohol for this substance is 20%.

(iii) For $n = 7$, obtain the distribution function of X and then find each of the probabilities

$$P(0 < X \le 0.25), \quad P(0.25 < X \le 0.5), \quad P(0.5 < X \le 0.75).$$

(iv) When the percentage concentration is between $0.25(j-1)$ and $0.25j$, the substance is sold at a price a_j dollars per liter, while the cost of production is b_j per liter, for $j = 1, 2, 3, 4$. Calculate, in terms of n, a_j, and b_j, the average profit made by the sale of 1 l from this substance.

12. The time, in minutes, that a medicine for pain relief takes until it starts to have effect is a random variable X with density function

$$f(x) = \frac{2(\theta - x)}{\theta^2}, \quad 0 \le x < \theta,$$

where $\theta > 0$ is the parameter of the distribution of X. Find

(i) the distribution function $F(t)$ and the probability $P(a < X \le b)$ for $0 < a < b \le \theta$;

(ii) the mean and variance of X.

13. A continuous random variable X has distribution function

$$F(t) = \begin{cases} \frac{1}{2}e^t, & t < 0, \\ 1 - \frac{1}{2}e^{-t}, & t \ge 0. \end{cases}$$

(i) Obtain an expression for the probability $P(a < X \le b)$ for $a < 0$ and $b > 0$.

(ii) Verify that the density function of X is given by

$$f(x) = \frac{1}{2}e^{-|x|}, \quad x \in \mathbb{R}.$$

(iii) Calculate the expected value and the variance of X.

(iv) Find the distribution function, the mean, and the variance of the random variable $Y = \alpha X$, where $\alpha > 0$ is a given constant.

14. Let X be a random variable having the Pareto distribution (Example 6.6). The density of X is given by

$$f(x) = \begin{cases} \frac{k\theta^k}{x^{k+1}}, & x \ge \theta, \\ 0 & x < \theta. \end{cases}$$

(i) What is the distribution function $F(t)$ of X?

(ii) For a given positive integer n, show that

$$P[n\theta \le X \le (n+1)\theta] = \frac{1}{n^k} - \frac{1}{(n+1)^k}.$$

(iii) Derive the value of the limit

$$\lim_{n \to \infty} \{n^k P[n\theta \le X \le (n+1)\theta]\}.$$

15. The lifetime, in hundreds of hours, of a light bulb is a continuous random variable with density function

$$f(x) = \lambda^2 x e^{-\lambda x}, \quad x > 0,$$

where $\lambda > 0$.

(i) Calculate the distribution function of X.

(ii) What is the expected value and the variance for the lifetime of a bulb?

(iii) From past data, it has been estimated that the average length of lifetime for a bulb of this type is 2500 hours. Which value of λ should be used to describe the distribution of bulb lifetimes? Using this value of λ, calculate the probability that a randomly selected light bulb will last for more than 2500 hours.

16. Let X_1 and X_2 be two continuous random variables with density functions f_1 and f_2, expected values μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Consider now a function f defined by

$$f(x) = \lambda f_1(x) + (1 - \lambda)f_2(x), \quad x \in \mathbb{R},$$

where $0 \leq \lambda \leq 1$.

(i) Verify that f is the density function of some random variable X.

(ii) Show that the expected value and the second moment around zero of X are given by the formulas

$$\mu = E(X) = \lambda\mu_1 + (1 - \lambda)\mu_2,$$
$$E(X^2) = \lambda E(X_1^2) + (1 - \lambda)E(X_2^2).$$

(iii) Deduce from above an expression for the variance of X in terms of $\lambda, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$.

17. Let X and Y be two continuous random variables with densities f_X and f_Y, respectively, and with the same range of values, R. Is it possible that

$$f_X(x) < f_Y(x) \quad \text{for all } x \in R?$$

If it is, give an example, otherwise explain why this cannot happen.

18. A manufacturing unit produces metal tubes with a nominal diameter of 12 in. The actual diameter of a tube can be regarded as a random variable having density function (in inches)

$$f(x) = \begin{cases} \dfrac{12}{11}[1 - (12 - x)^2], & 11.5 \leq x \leq 12.5, \\ 0, & \text{elsewhere}. \end{cases}$$

Assuming that the length of each tube is 20 ft, answer the following:

(i) Find the density function and the distribution function of the surface area for a randomly selected tube. (The formula for the surface area is $2\pi r(r + h)$, where r is the radius and h is the length [or height] of a tube.)

(ii) If the cost of producing the tubes is k dollars per square inch, calculate the expected cost for each tube produced.

19. In Example 6.3, we obtained the distribution function of $Y = aX + b$, when the distribution of X is known and $a > 0$. Obtain a similar result for the case when $a < 0$, i.e. by assuming that the distribution function, F, of the variable X is known, find the distribution function and the density of

$$Y = aX + b.$$

Application: Find the density function of

$$Y = -4X + 1$$

for the case where X has density

$$f(x) = 4\mathrm{e}^{-4x}, \quad x \geq 0.$$

20. Let X be a random variable with density function

$$f(x) = \mathrm{e}^{-x}, \quad x > 0.$$

Let Y be another variable defined by

$$Y = \begin{cases} X, & \text{if } X \leq 1, \\ 1/X, & \text{if } X > 1. \end{cases}$$

Find the distribution function, the density function and the expectation of Y.

21. Suppose that a continuous variable X has density function

$$f(x) = \begin{cases} cx^s, & 0 \leq x \leq a, \\ 0 & \text{elsewhere,} \end{cases}$$

where a and s are two given positive real numbers, while c is a suitable constant.

(i) Find the value of c in terms of a and s.

(ii) Obtain the distribution function for each of the following random variables:

$$X_1 = \mathrm{e}^X, \quad X_2 = X^2, \quad X_3 = (X-1)^2, \quad X_4 = \sqrt{X}, \quad X_5 = 1/X.$$

(iii) Calculate the mean and variance for each of the variables X_2, X_3, X_4, X_5 in Part (ii).

22. Assuming that for the random variables X and Y, we have

$$E[(X - Y)^2] = 0,$$

establish that $P(X = Y) = 1$.

(*Hint*: Work with the random variable $Z = X - Y$.)

23. The density function of a random variable X is given by

$$f(x) = \frac{1}{n!}x^n e^{-x}, \quad x > 0,$$

where n is a positive integer. Show that

$$P(X < 2n + 2) > \frac{n}{n+1}.$$

(*Hint*: Use the formula

$$\int_0^\infty x^k e^{-x} dx = k!, \quad k = 0, 1, 2 \ldots,$$

to find the mean and variance of X. Then, use Chebyshev's inequality.)

6.10 APPLICATIONS

6.10.1 Profit Maximization

Profit maximization is the primary long-term objective of any business/enterprise. In theoretical economics, the term profit maximization is used to describe the process by which a company determines the parameters of its commercial activity such as price of a product, frequency of orders, order level, etc., so that the resulting profit attains its largest value.

In this section, we shall present a simple model by which the order level of a grocery product (for a specific time cycle, say one month) is determined, so as to maximize the mean profit obtained from the product.

Let us assume that the monthly sales of a product (in kilograms) at a particular supermarket can be described by a continuous random variable X with density function f and distribution function F. The supermarket makes a (net) profit of α dollars for each kilogram of the product which is sold, while any kilogram unsold by the end of a month incurs a loss of β dollars. Suppose at the beginning of each month, the supermarket orders z kilograms of this product.

In order to determine the optimal value z for the monthly order level, we shall first give a formula for the monthly profit of the supermarket from the sale of this product. If the monthly demand, say x, of the product is greater than z, then the entire quantity ordered will be sold and the total profit will be

$$P = \alpha z.$$

However, if the demand x is less than z, only $x < z$ kilograms will be sold (offering a profit of αx dollars) while the rest of the quantity, $z - x$, will remain unsold by the end of the month and will incur a loss of β dollars per unit. Consequently, in this case, the final profit will be

$$P = \alpha x - \beta(z - x).$$

According to the above discussion, and taking into account that the demand is in fact a continuous random variable X, we may state that the monthly profit is also a random variable, $P = P(X)$, which is given by

$$P = P(X) = \begin{cases} \alpha z, & \text{if } z \le X, \\ \alpha X - \beta(z - X), & \text{if } z > X. \end{cases}$$

The average (mean) monthly profit can then be expressed as

$$\begin{aligned} m(z) = E[P(X)] &= \int_{-\infty}^{\infty} P(x)f(x)\mathrm{d}x \\ &= \int_0^z (\alpha x - \beta(z - x))f(x)\mathrm{d}x + \int_z^{\infty} \alpha z f(x)\mathrm{d}x \\ &= \int_0^z ((\alpha + \beta)x - \beta z)f(x)\mathrm{d}x + \alpha z \left(1 - \int_0^z f(x)\mathrm{d}x\right) \\ &= \alpha z + (\alpha + \beta)\int_0^z (x - z)f(x)\mathrm{d}x. \end{aligned}$$

Writing the integral on the RHS as a difference of two integrals, we may easily derive the following more convenient form:

$$m(z) = \alpha z + (\alpha + \beta)\int_0^z xf(x)\mathrm{d}x - (\alpha + \beta)zF(z), \qquad z \ge 0. \tag{6.18}$$

Figure 6.8 displays the mean monthly profit as a function of z for $\alpha = 2$, $\beta = 1$ and several choices of the density function $f(x)$. The function $m(z)$ attains a maximum value in all three cases presented in Figure 6.8.

In order to determine the value of z that maximizes $m(z)$, we may observe that

$$\begin{aligned} m'(z) &= \alpha + (\alpha + \beta)\left(\int_0^z xf(x)\mathrm{d}x\right)' - (\alpha + \beta)zF'(z) - (\alpha + \beta)F(z) \\ &= \alpha + (\alpha + \beta)zf(z) - (\alpha + \beta)zf(z) - (\alpha + \beta)F(z) \\ &= \alpha - (\alpha + \beta)F(z), \\ m''(z) &= -(\alpha + \beta)f(z) < 0. \end{aligned}$$

Consequently, the value of z that maximizes $m(z)$ is the solution of the equation

$$m'(z) = 0 \iff \alpha - (\alpha + \beta)F(z) = 0 \iff F(z) = \frac{\alpha}{\alpha + \beta}.$$

As an example we mention that, for the three distributions considered in Figure 6.8, the corresponding cumulative distribution functions are

$$\text{(a) } F(x) = 1 - \mathrm{e}^{-x^3},\ x > 0; \quad \text{(b) } F(x) = x,\ 0 \le x \le 1; \quad \text{(c) } F(x) = \frac{x^2}{4},\ 0 \le x \le 2,$$

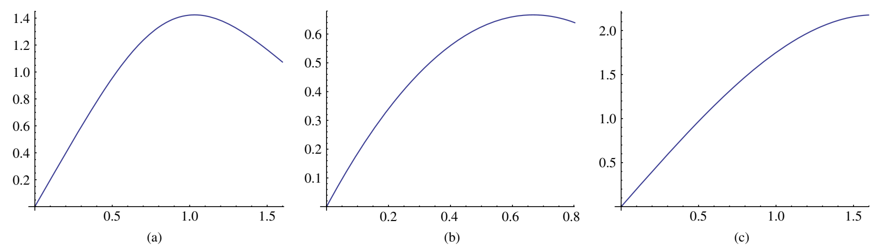


Figure 6.8 Plot of the mean monthly profit $m(z)$ for (a) $f(x) = 3x^2e^{-x^3}$, $x > 0$; (b) $f(x) = 1$, $0 \leq x \leq 1$; and (c) $f(x) = x/2$, $0 \leq x \leq 2$.

while the order levels that maximize the mean profit are easily determined as follows:

(a) $1 - \mathrm{e}^{-z^3} = \dfrac{\alpha}{\alpha + \beta} \Longleftrightarrow z = \left(\ln \dfrac{\alpha + \beta}{\beta} \right)^{1/3}$,

(b) $z = \dfrac{\alpha}{\alpha + \beta}$,

(c) $\dfrac{z^2}{4} = \dfrac{\alpha}{\alpha + \beta} \Longleftrightarrow z = \sqrt{\dfrac{4\alpha}{\alpha + \beta}}$.

KEY TERMS

(absolutely) continuous random variable
Cauchy distribution
expectation of a continuous random variable
expectation of a function of a random variable
functions of a random variable
median
mixed distribution
Pareto distribution
probability density function
symmetric distribution
variance of a continuous random variable

CHAPTER 7

SOME IMPORTANT CONTINUOUS DISTRIBUTIONS

Carl Friedrich Gauss (Braunschweig 1777–Göttingen 1855)

Named the *Prince of Mathematicians* and considered to be one of the most influential mathematicians in history, Gauss advanced almost every branch of mathematics and physics known at his time. Neither of his parents were educated; his mother was illiterate while his father was a stone mason. Perhaps the prime example of a child prodigy, he showed exceptional mathematical skills even before going to school; according to one story, at the age of three, while his father was performing some arithmetic calculations on paper, he spotted an error and corrected him by performing all calculations mentally and faultlessly.

Being self-educated in his early years, his genius was recognized by the Duke of Brunswick, whose funding allowed Gauss to pursue a formal education. After attending Caroline College from 1792 to 1795, he studied at Göttingen University. Gauss developed the concept of complex numbers and the University of Helmstedt granted him a PhD in 1799.

In his early investigations, he was fascinated by number theory and in 1801 he published his famous book *Disquisitiones Arithmeticae* (Arithmetical Investigations). The book developed the modern approach to modular arithmetic, and it is considered to be one of the most brilliant achievements in the history of mathematics.

From the age of 23, he became interested in astronomy and predicted the position of the asteroid Ceres in December 1801, within a half-degree accuracy. In order to achieve this, Gauss used what is currently known as the *least squares method*, which he had developed a few years earlier. The method is used even today in various scientific areas as a means to minimize measurement error. In an attempt to rationalize the method of least squares, he discovered the normal distribution, suggesting this as a generic model for experimental error. This important distribution is in fact described in this chapter.

During the late 1820s, Gauss collaborated with the physicist Wilhelm Weber and made significant advances in the areas of optics, acoustics, mechanics, and magnetism. Together, they built and first used in 1833, for regular communication, the electromagnetic telegraph. This invention revolutionized communications in the nineteenth century.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

7.1 THE UNIFORM DISTRIBUTION

Suppose that a continuous random variable X takes values in a finite interval $[a, b]$. In many cases, it will be reasonable to assume that, any subinterval of $[a, b]$ with a fixed length, has the same probability that X falls into it. Informally, we might be tempted to say that "X takes any value in $[a, b]$ with the same probability" and this is how people often understand this, although in a formal mathematical way this is inappropriate, as we have seen that if X is continuous, then $P(X = x) = 0$ for *any* real x. We might then express this by requiring that "the probability that X takes a value close to $x \in (a, b)$ is the same for all x in that interval."

Moreover, we have already seen in Section 6.1 that the value of the density $f(x)$ of a variable X at the point x is proportional to the probability that the value of X "is close to x." Thus, the value of f must be constant throughout the interval $[a, b]$, and so it should have the form

$$f(x) = \begin{cases} c, & a \le x \le b, \\ 0, & \text{elsewhere.} \end{cases} \tag{7.1}$$

With f defined as above, the original requirement that X takes values with the same probability in an interval of fixed width, say h, is satisfied, since for $y \in [a, b - h]$, we have that

$$P(y \le X \le y + h) = \int_y^{y+h} f(x)\,\mathrm{d}x = \int_y^{y+h} c\,\mathrm{d}x = ch,$$

which is independent of y and depends only on the width h of the interval.

Further, in order to determine the constant c in (7.1), we note that

$$1 = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_a^b c\,\mathrm{d}x = c(b - a),$$

which immediately shows that c must be equal to $1/(b - a)$.

We thus arrive at the following definition.

Definition 7.1 Let X be a continuous random variable with density function

$$f(x) = \begin{cases} \dfrac{1}{b - a}, & a \le x \le b, \\ 0, & \text{elsewhere.} \end{cases}$$

We then say that X has the uniform distribution over the interval $[a, b]$. We denote this by $X \sim \mathcal{U}[a, b]$.

In fact we have already seen, in the previous chapter, examples of random variables having a uniform distribution. Also, in cases where we "generate randomly" a number within a specified range, e.g. [0, 1], the random variable that represents the outcome of this experiment is typically assumed to follow a uniform distribution.

When a variable X is uniformly distributed in a given interval $[a, b]$, the distribution function of X given by

$$F(t) = P(X \le t) = \int_{-\infty}^{t} f(x)\,\mathrm{d}x$$

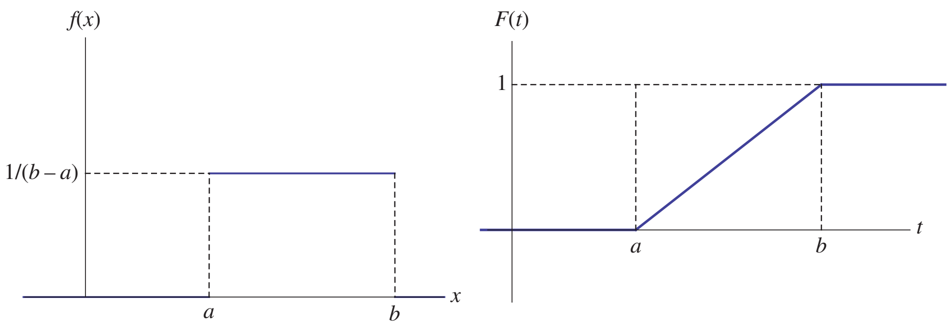


Figure 7.1 Plot of probability density and cumulative distribution function of $\mathcal{U}[a, b]$.

is easily found to be

$$F(t) = \begin{cases} 0, & t < a, \\ \dfrac{t-a}{b-a}, & a \le t \le b, \\ 1, & t > b, \end{cases} \tag{7.2}$$

and it can be seen in Figure 7.1, along with the density function of the $\mathcal{U}[a, b]$ distribution. Because of the shape of its density, the uniform distribution is sometimes referred to as the **rectangular distribution**.

In Figure 7.2, we have plotted the density function and the distribution function of the $\mathcal{U}[a, b]$ distribution for different choices of the parameters a and b.

For the remainder of this section, it is useful to keep in mind that, since all events of the form $\{X = x\}$ for a continuous X have zero probability, it does not matter whether we use inequalities of the form $X \le x$ (and $X \ge x$) or $X < x$ (resp., $X > x$). Thus, in particular, in the definition of the uniform distribution (Definition 7.1), it does not matter if we use inequalities of the form $a \le x$ and $x \le b$ or $a < x$ and $x < b$, respectively.

In practice, when we use a distribution as a probability model for an unknown quantity, we typically have to ask ourselves which probability distribution is suitable for the quantity under consideration. In view of the fact that the uniform distribution has a constant density (over a finite interval), it is used in probability modeling for quantities whose values we consider to be "equally likely" or, more formally, for situations where all intervals of the same width have the same probability.

Proposition 7.1 *The expectation and variance of a random variable X having the $\mathcal{U}[a, b]$ distribution are given by*

$$E(X) = \frac{a+b}{2}, \qquad \mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Proof: For the expectation, we have

$$E(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_a^b x \cdot \frac{1}{b-a}\,\mathrm{d}x = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{a+b}{2}.$$

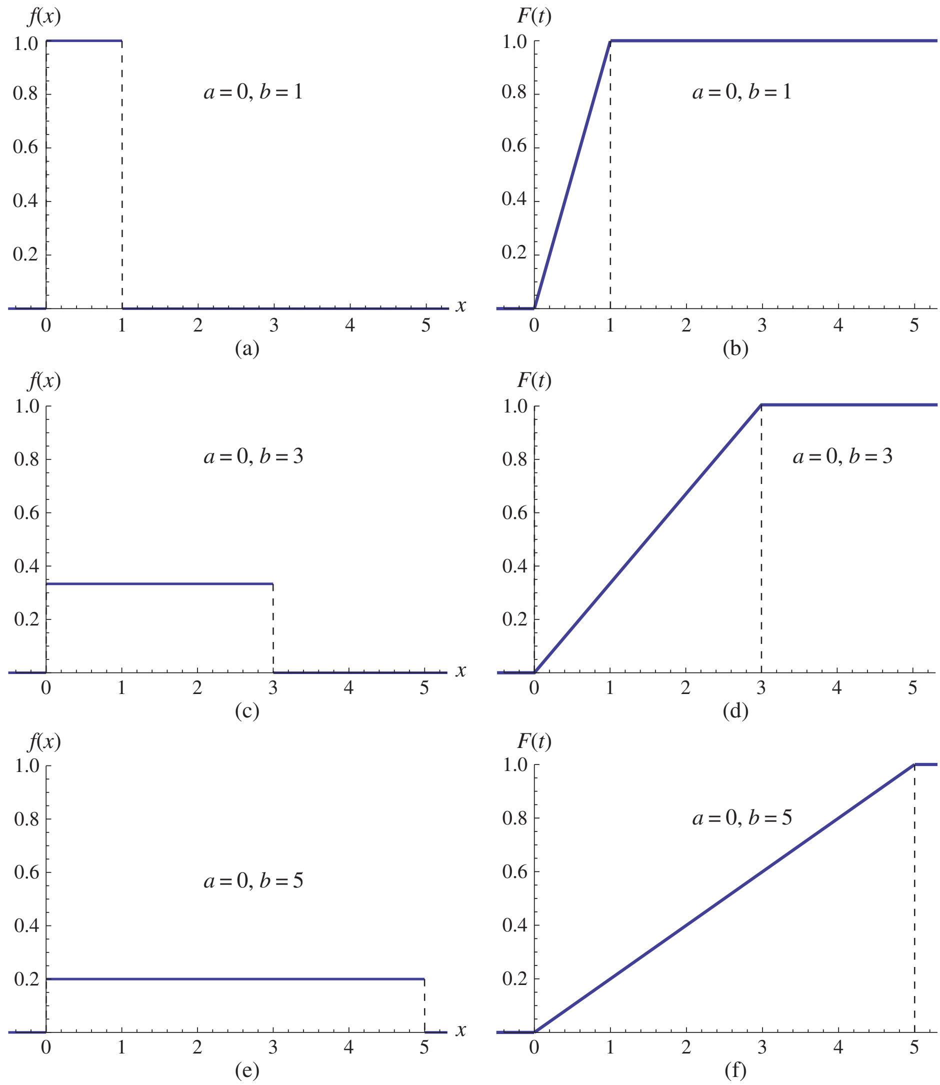


Figure 7.2 The density function (left) and the distribution function (right) of the uniform distribution for different choices of a and b.

For the variance, we calculate first the moment of second order around zero. This is

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, \mathrm{d}x = \int_a^b x^2 \cdot \frac{1}{b-a}\, \mathrm{d}x = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b$$
$$= \frac{b^3 - a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \frac{b^2+ab+a^2}{3}.$$

Making use of this in the formula $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$, we derive

$$\mathrm{Var}(X) = \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12},$$

as required. □

Example 7.1 Simon goes to his University class every weekday by using a train that leaves at 8:00 a.m. We assume that the duration of the journey, in minutes, is a random variable following the uniform distribution in the interval $[58, 63]$. Further, suppose that from the train platform he needs a 15-minute walk to enter the classroom and that his class starts precisely at 9:15a.m.

(i) What is the probability that Simon arrives in time for his class?

(ii) What is the probability that he arrives at the class at least two minutes after it has begun?

(iii) Find Simon's expected arrival time in the class.

SOLUTION Since the only source of uncertainty is the duration of the train journey, we define a random variable X, which represents the number of minutes that this journey lasts. Then $X \sim \mathcal{U}[58, 63]$, so that X has density function

$$f(x) = \begin{cases} \dfrac{1}{63-58}, & 58 \le x \le 63, \\ 0, & \text{elsewhere}, \end{cases} = \begin{cases} \dfrac{1}{5}, & 58 \le x \le 63, \\ 0, & \text{elsewhere}, \end{cases}$$

while the corresponding distribution function of X is given by

$$F(t) = \begin{cases} 0, & t < 58, \\ \dfrac{t-58}{63-58}, & 58 \le x \le 63, \\ 1, & t > 63. \end{cases}$$

(i) For Simon to be in time (i.e. before 9:15 a.m.) for his class, he must arrive at the platform by 9:00 a.m., which means that the train journey lasts for 60 minutes at most. The probability for this is

$$P(X \le 60) = \frac{60-58}{63-58} = \frac{2}{5}.$$

(ii) Now, we seek the probability that Simon arrives at the classroom after 9:17 a.m., which in turn implies that he arrives at the platform after 9:02 a.m. This happens if the train journey takes 62 minutes or more, with associated probability

$$P(X > 62) = 1 - F(62) = 1 - \frac{62-58}{63-58} = \frac{1}{5},$$

i.e. there is a 20% chance that he arrives at least two minutes late for the class.

(iii) The expected value of X, the journey time, by Proposition 7.1 is

$$E(X) = \frac{58+63}{2} = 60.5 \text{ minutes}.$$

This implies that Simon is expected to arrive at the train platform of the University station half a minute after 9:00 a.m. and, as a result, his expected arrival time at the classroom is half a minute after the beginning of the class (9:15 a.m.).

The following two examples present two important properties of the uniform distribution. The first of these shows that a linear transformation of an uniform random variable is again uniform; the second relates the distribution function of the uniform distribution over the unit interval to the distribution function of an arbitrary continuous random variable.

Example 7.2 Let X have a uniform distribution in the interval $[\alpha, \beta]$. What is the distribution of the linear transformation

$$Y = \gamma X + \delta$$

for $\gamma > 0$ and $\delta \in \mathbb{R}$?

SOLUTION It is easier to work with distribution functions rather than densities. The distribution function of X is

$$F_X(t) = P(X \leq t) = \begin{cases} 0, & t < \alpha, \\ \dfrac{t-\alpha}{\beta-\alpha}, & \alpha \leq t \leq \beta, \\ 1, & t > \beta. \end{cases}$$

Before we consider the distribution function of Y, we first find its range of values. Since $X \sim \mathcal{U}[\alpha, \beta]$, we have $\alpha \leq X \leq \beta$ and so, since γ is positive,

$$\gamma\alpha + \delta \leq Y = \gamma X + \delta \leq \gamma\beta + \delta$$

(formally, we should say that this statement holds with probability one). Next, for y in the interval $R_Y = [\gamma\alpha + \delta, \gamma\beta + \delta]$, we have

$$P(Y \leq y) = P(\gamma X + \delta \leq y) = P\left(X \leq \frac{y-\delta}{\gamma}\right),$$

which, by the formula for F_X given above (replacing t by $(y - \delta)/\gamma$ there), equals

$$\frac{(y-\delta)/\gamma - \alpha}{\beta - \alpha} = \frac{y - \delta - \gamma\alpha}{\gamma(\beta-\alpha)} = \frac{y - (\delta + \gamma\alpha)}{\gamma(\beta-\alpha)}.$$

This gives the distribution function of Y as

$$F_Y(y) = P(Y \le y) = \begin{cases} 0, & y < \gamma\alpha + \delta, \\ \dfrac{y - (\delta + \gamma\alpha)}{\gamma(\beta - \alpha)}, & \gamma\alpha + \delta \le y \le \gamma\beta + \delta, \\ 1, & y > \gamma\beta + \delta. \end{cases}$$

It is now easy to see (just put $a = \alpha\gamma + \delta$ and $b = \beta\gamma + \delta$ in (7.2)) that this is the distribution function of the uniform distribution over the interval $[a, b]$. In words, we have shown that a linear transformation of an uniformly distributed random variable also has an uniform distribution. More precisely, we have

$$X \sim \mathcal{U}[\alpha, \beta] \;\Rightarrow\; \gamma X + \delta \sim \mathcal{U}[\gamma\alpha + \delta, \gamma\beta + \delta].$$

As a special case, assume that $X \sim \mathcal{U}[\alpha, \beta]$ and consider the case when

$$\gamma = \frac{1}{\beta - \alpha}, \qquad \delta = -\frac{\alpha}{\beta - \alpha}.$$

Then, it follows that $Y = \gamma X + \delta$ is uniformly distributed in the unit interval, $[0, 1]$. This distribution is commonly referred to as the *standard uniform distribution*.

Example 7.3 (Transformation $Y = F(X)$) Let X be a continuous random variable with distribution function F. Then, for the random variable $Y = F(X)$, we can write

$$F_Y(y) = P(Y \le y) = P(F(X) \le y), \quad 0 \le y \le 1.$$

Even though F may not be a strictly increasing function, an inverse function can be defined by

$$F^{-1}(y) = \min\{x : F(x) \ge y\}.$$

Using this definition of the inverse function F^{-1}, we get

$$F_Y(y) = P(X \le F^{-1}(y)) = F(F^{-1}(y)) = y, \quad 0 \le y \le 1.$$

We have thus arrived at the fact that, no matter what the original distribution F of the variable X is, the random variable $Y = F(X)$ is uniformly distributed in the unit interval. This is a powerful result with diverse applications (e.g. in the theory of simulation), as it gives us a convenient way to reach the uniform distribution function from an arbitrary continuous distribution and vice versa. The transformation $Y = F(X)$ considered here is called *Probability Integral Transformation*.

EXERCISES

Group A

1. A train leaves exactly every 45 minutes from London to Manchester. Emily, who wants to take the train to Manchester, arrives at the train station in London but is completely unaware of the train times. What is the expected time she has to wait until the next train leaving to Manchester?

2. The repair time, in hours, of a faulty machine has the uniform distribution in the interval $[0, 1]$. If the cost of service for the machine for X hours is $X\sqrt{X}$, find the expected cost associated with each fault.

3. The time X, in hours, required to service a PC is a continuous random variable X, having the $\mathcal{U}[0.5, 2.5]$ distribution. Then, calculate

   (i) the expected time for the service of a PC;

   (ii) the expected cost of a service, if the repair company charges

   $$12\sqrt{X} + \frac{5}{X}$$

   dollars for a service lasting X hours.

4. If X has the $\mathcal{U}[-a, a]$ distribution, find the value of a in each of the following cases (in each case, you are given a statement about X):

   (a) $P\left(\frac{1}{8} \le X \le \frac{1}{4}\right) = \frac{1}{32}$;

   (b) $P\left(X \ge \frac{3}{2}\right) = 3P\left(X \le -\frac{5}{2}\right)$;

   (c) $P(X \le 1) = P(|X| \ge 1)$;

   (d) $P(|X| \le 2) = P(|X| > 1)$;

   (e) $\mathrm{Var}(X) = \frac{16}{3}$.

5. A number X is randomly selected from the interval $[4, 12]$. Calculate

   (i) the density function of the area E of a square whose side has size X (in meters);

   (ii) the probability that this area is at most 36 m^2;

   (iii) the expected area of a square whose side has size $X/3$.

6. A computer selects at random a number X from the interval $[0, 3]$. Find the probability that

   (i) the first decimal place of X is a multiple of 3;

   (ii) the sum of the two first decimal places of X is equal to 5;

   (iii) the first decimal place of the cubic root of X is 4.

7. A university lecturer finishes a particular class on Thursdays between 10:00 a.m. and 10:05 a.m. Let X be the time, in minutes, between 10:00 a.m. and the end of the class, and let us assume that X is uniformly distributed in the interval $[0, 5]$.

   (i) Find the probability that on a particular day, the class finishes

      (a) before 10:03 a.m.;

      (b) between 10:01 a.m. and 10:03 a.m.

   (ii) If, on a particular day, the class has not finished by 10:01 a.m., what is the probability that it will finish after 10:03 a.m.?

   (iii) Suppose that a friend of yours, who attends this class with you, offers you the following bet: if the lecture finishes before 10:01 a.m., he will give you α dollars, while if it finishes after that time you have to pay him β dollars. Show that, for the game to be fair, we must have $\alpha = 4\beta$.

8. A number β is selected randomly from the interval $[-7, 8]$. What is the probability that the quadratic equation

$$3t^2 + \beta t + 3 = 0$$

   has at least one real root?

9. Let X be a random variable that has the uniform distribution over the interval $[a, b]$. Show that the central moments of order r,

$$\mu_r = E[(X - \mu)^r], \quad r = 1, 2, 3, \ldots,$$

   where $\mu = (a + b)/2$, are given by the formula

$$\mu_r = \begin{cases} 0, & r = 1, 3, 5, \ldots, \\ \dfrac{(b-a)^r}{(r+1)2^r}, & r = 2, 4, 6, \ldots. \end{cases}$$

   For the special case $r = 2$, deduce the formula given in Proposition 7.1, namely,

$$\mathrm{Var}(X) = \frac{(b-a)^2}{12}.$$

Group B

10. A stick of length a is broken at an arbitrary point due to the effect of some weight forced on it. What is the probability that

   (i) the larger of the two parts of the stick that are formed after the break is at least three times bigger than the smaller part?

   (ii) neither of the two parts has a length less than $a/4$?

11. Find the distribution of the random variable $X = \tan\Theta$, where Θ is a random variable with the uniform distribution in the interval $(-\pi/2, \pi/2)$.

12. Assume that X is a random variable with the $\mathcal{U}[0, 1]$ distribution. Obtain the distribution

    (i) of the random variable $Y = [nX]$, where n is a positive integer and [.] denotes the integer part;

    (ii) of the variable $Z = -\ln(1 - X)$;

    (iii) of the variable $W = X^n$, $n \in \mathbb{N}$.

13. Let X be a number randomly chosen from the interval [0, 1], Y_1 be the first decimal place of X, and Y_2 be the second decimal place of $\sqrt{X}$. Show that while the probabilities $P(Y_1 = n)$, $n = 0, 1, \ldots, 9$, are the same for all positive integers n, the probabilities $P(Y_2 = n)$, $n = 0, 1, \ldots, 9$, increase as n gets larger.

14. Let X be an uniform random variable on $[0, \theta]$, with $\theta > 0$. Find a function of the form $g(x) = ax^2 + bx + c$, so that we have

$$E[g(X)] = \theta^2, \quad \text{for all } \theta > 0.$$

15. (i) A stick of length $2l$ is broken at a random point, and let X be the length of the smallest between the two parts that are formed. Find the density function and the expected value of X.

    (ii) We have two sticks, each of which has length $2l$ and is broken at a random point into two pieces. Let Y be the length of the smallest among the 4 pieces.

    (a) Show that the density function of Y is given by

$$f_Y(y) = 2(l - y)/l^2, \quad 0 \le y \le l;$$

    (b) Verify that the expectation and variance of Y are given by

$$E(Y) = \frac{l}{3}, \qquad \mathrm{Var}(Y) = \frac{l^2}{18}.$$

16. A sugar manufacturing factory consists of three sections. Each of these sections processes daily a quantity X of sugarcane (in tons), which has an uniform distribution in the interval $[0, a]$ and has an expectation of 5 tons. The three sections work independently of one another.

    (i) What is the probability that, on a given day, the first two sections process less than 8 tons each and the third section processes more than 6 tons of sugarcane?

    (ii) If the quantity that can be processed by the first section is increased by 50%, find the density function and the mean value of the quantity that this section processes per day.

    (iii) Answer Part (i) using the new capacity for the first section given in Part (ii) above.

7.2 THE NORMAL DISTRIBUTION

The fact that this is the longest section in this book should come as no surprise. Among all probability distributions, the normal distribution has a paramount role and is the most commonly used distribution in a wide array of applications. Its significance will be fully understood in Volume II of this book, when we consider sums of random variables, but here we set up the theoretical framework and present its main properties, some of which are unique among distributions on the real line.

The prominent position of the normal distribution in probability theory, statistics and their applications can be ascribed primarily to the following reasons:

- For many random variables that are frequently studied, such as height, weight, shoe size, student performance in an exam, IQ scores, tree diameters, molecule velocities, etc., the normal distribution has been seen to offer a satisfactory fit;
- Random errors that appear in various measurements are usually (approximately) normal. For this reason, the normal distribution is sometimes called the *error distribution*;
- The sum and the average of *a large number of observations* from a random variable X can be approximated by the normal distribution, and this is irrespective of the original distribution of X;
- Many other distributions, both discrete and continuous, can be approximated, under certain conditions, suitably by the normal distribution.

Historically, the normal distribution was first studied by the French mathematician Abraham De Moivre (1667–1754), but today it is more commonly associated with the German Karl Friedrich Gauss (1777–1855). Gauss put the normal distribution into practice and used it in astronomy (with his famous method of least squares) in order to describe the orbits of planets and other celestial bodies. As a result, another name for the normal distribution that is often used is the *Gaussian distribution*. De Moivre's original work, motivated like that of many others in his time by gambling problems, involved the search for an approximation to the binomial distribution, $b(n, p)$, for the case when n is large and $p = 1/2$, such as the distribution of heads in a large number of coin tosses. The result was the following theorem, which rightfully bears his name.

Proposition 7.2 *(De Moivre's Theorem) Let X be a random variable having the binomial distribution, $b(n, p)$, with $p = 1/2$. Then, for any real numbers $a < b$, we have*

$$\lim_{n\to\infty} P\left(a < \frac{X - \frac{n}{2}}{\frac{\sqrt{n}}{2}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\,\mathrm{d}x. \tag{7.3}$$

The proof of this, and the following result which is a generalization of it, is not given here. Both results are in fact special cases of the *central limit theorem*, one of the most prominent results in probability theory and statistics, which is presented and proved formally in Volume II of this book. In the statement of De Moivre's theorem, note in particular the appearance of the random variable

$$Z = \frac{X - \frac{n}{2}}{\frac{\sqrt{n}}{2}} = \frac{X - \mu}{\sigma},$$

where

$$\mu = E(X) = np = \frac{n}{2}, \quad \sigma^2 = \mathrm{Var}(X) = npq = \frac{n}{4},$$

since X is binomially distributed. The random variable Z, which arises from X by subtracting its mean and dividing by its standard deviation, is the *standardized random variable* introduced in Section 4.5 (see Example 4.24). It plays a crucial role while dealing with the normal distribution.

In 1812, several decades after the publication of De Moivre's limit theorem, the French mathematician Pierre-Simon Laplace (1749–1827) published a book in which he laid down many fundamental results in probability theory and statistics. One of them is the following, known today as the De Moivre–Laplace theorem. This is an extension of Proposition 7.2, treating now the case for arbitrary p.

Proposition 7.3 *(De Moivre–Laplace theorem) Let X be a random variable having the binomial distribution with parameters n and p. Then, for any real numbers $a < b$, we have*

$$\lim_{n\to\infty} P\left(a < \frac{X - np}{\sqrt{npq}} < b\right) = \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\, \mathrm{d}x. \tag{7.4}$$

Since for the variable X in the above proposition we have $E(X) = np$ and $\mathrm{Var}(X) = npq$ (see Proposition 5.2), the standardized variable associated with X now takes on the form

$$Z = \frac{X - np}{\sqrt{npq}}.$$

We have seen in Chapter 5 that the Poisson approximation also emerged as an approximation to the binomial distribution. In fact, Poisson's result was published in 1837, a quarter of a century after Laplace's result above and the Poisson approximation to the binomial (Proposition 5.10) provides satisfactory results only when $n \to \infty$ and $p \to 0$ in such a way that $np \to \lambda > 0$. On the other hand, the De Moivre–Laplace theorem, which suggests the approximation

$$P\left(a < \frac{X - np}{\sqrt{npq}} < b\right) \cong \frac{1}{\sqrt{2\pi}} \int_a^b \mathrm{e}^{-x^2/2}\, \mathrm{d}x, \tag{7.5}$$

yields good (and, in some cases, excellent) results when n is large and $np(1 - p) \geq 10$.

According to Proposition 7.3 by letting $a \to -\infty$ and $b = t$ there we see that, for each real t, the sequence of probabilities

$$P\left(\frac{X - np}{\sqrt{npq}} \leq t\right), \quad n = 1, 2, \ldots,$$

converges, as $n \to \infty$, to the function

$$\Phi(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} \mathrm{e}^{-x^2/2}\, \mathrm{d}x, \tag{7.6}$$

and it is reasonable to expect that this function is itself a distribution function. If this happens to be the case, from the fundamental theorem of calculus, the density of this distribution has to be the function $\mathrm{e}^{-x^2/2}/(\sqrt{2\pi})$. The proof of the fact that Φ above is a proper distribution function is given after the following definition, which introduces the standard normal distribution.

Definition 7.2 A continuous random variable Z is said to follow the **standard normal distribution** when the density function of Z is given by

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-z^2/2}, \quad -\infty < z < \infty. \tag{7.7}$$

The corresponding distribution function of Z is denoted by Φ and is as given in (7.6).

We use Z rather than X for a standard normal random variable, in accordance with the standard notation in most textbooks, while X is used below when we consider an arbitrary normal random variable (i.e. not a standard one).

We now prove that Φ satisfies the properties of Proposition 4.1, so that it is a valid probability distribution function. Since the function $\mathrm{e}^{-x^2/2}$ takes positive values, it is immediate that the function Φ is a (strictly) increasing function, while the property in Part (iii) of Proposition 4.1 follows from the continuity of the integral. Further, we have

$$\lim_{t\to-\infty} \Phi(t) = \lim_{t\to-\infty} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x = 0.$$

One may wonder about the appearance of the term $(\sqrt{2\pi})^{-1}$ in (7.6). This is in fact to ensure that Φ satisfies the remaining property for a distribution function, namely, that $\lim_{t\to\infty}\Phi(t) = 1$. The proof of this result is not straightforward. In order to find the value of the integral

$$I = \int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x, \tag{7.8}$$

we argue as follows. Consider the square of that integral and observe that

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x\right)\left(\int_{-\infty}^{\infty} \mathrm{e}^{-y^2/2}\, \mathrm{d}y\right) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{e}^{-(x^2+y^2)/2}\, \mathrm{d}x\, \mathrm{d}y. \end{aligned}$$

We now make the following change of variables to polar coordinates. Let $x = r\cos\theta$ and $y = r\sin\theta$ (see the following figure). Then, we have[1] $\mathrm{d}x\, \mathrm{d}y = r\, \mathrm{d}\theta\, \mathrm{d}r$, while

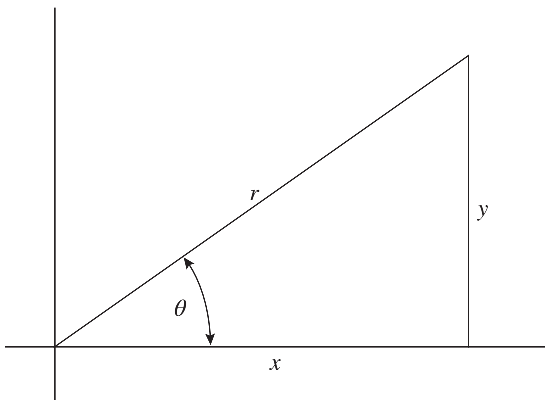
$$x^2 + y^2 = r^2(\cos^2\theta + \sin^2\theta) = r^2.$$

[1]The proof of this is not trivial, as one has to consider the Jacobian of the transformation when making the change of variables in the double integrals. We leave the details to the interested reader.

We subsequently obtain

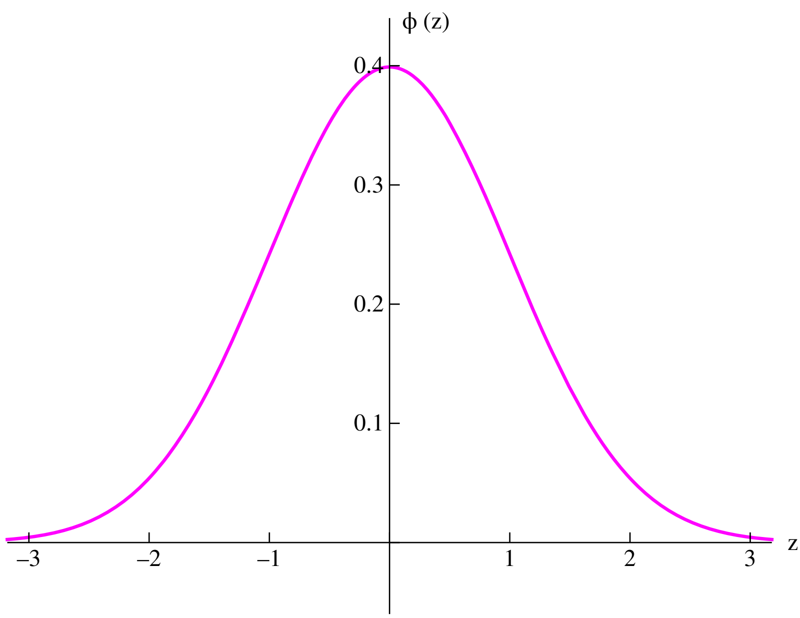
$$I^2 = \int_0^\infty \int_0^{2\pi} e^{-r^2/2} r \, d\theta \, dr = \int_0^\infty r e^{-r^2/2} \left(\int_0^{2\pi} d\theta \right) dr$$
$$= 2\pi \int_0^\infty r e^{-r^2/2} \, dr = 2\pi[-e^{-r^2/2}]_0^\infty = 2\pi.$$

It then follows that $I = \sqrt{2\pi}$, which in turn implies that the standard normal density function, ϕ, in Definition 7.2 integrates to one (or that $\lim_{t\to\infty} \Phi(t) = 1$, which is the same).



The following figure shows the graph of the function $\phi(z)$. The graph reveals that ϕ is *bell-shaped* and symmetric around the y-axis; the latter property can be verified formally from the immediate relationship

$$\phi(-z) = \frac{1}{\sqrt{2\pi}} e^{-(-z)^2/2} = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} = \phi(z).$$



z	$\Phi(z)$	z	$\Phi(z)$
0.0	0.5000	0.0	0.5000
−0.5	0.3085	0.5	0.6915
−1.0	0.1587	1.0	0.8413
−1.5	0.0668	1.5	0.9332
−2.0	0.0228	2.0	0.9772
−2.5	0.0062	2.5	0.9938
−3.0	0.0013	3.0	0.9987

Table 7.1 Extract from the table of values of the standard normal distribution function.

Calculation of probabilities associated with the standard normal distribution requires evaluation of the distribution function

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \phi(y)\mathrm{d}y = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \mathrm{e}^{-y^2/2}\, \mathrm{d}y$$

for various values of $z \in \mathbb{R}$.

Unfortunately, none of the known integration techniques provides an analytic expression for the above integral. In view of this, from the early days of the normal distribution, the values of $\Phi(z)$ were tabulated for different values of z. Today, tables of the standard normal distribution form an essential complement to most probability and statistics textbooks (as also in this book, see Appendix B), while the values of Φ are also available in nearly every statistics and algebraic software, such as R, SAS, Mathematica, etc. A small set of values of $\Phi(z)$ is given in Table 7.1.

The table in Appendix B gives $\Phi(z)$ for $z = 0.00, 0.01, 0.02, \ldots, 3.59$ (for higher values of z, $\Phi(z)$ is very close to 1). The way to read the table of values $\Phi(z)$ in Appendix B is as follows: the values in the first column of the table (in bold) give the integer part and the first decimal place of z while the values, also in bold, in the first row of the table give the second decimal place of z. Then we look for the associated value $\Phi(z)$ inside the table, in the corresponding row and column. Suppose, for instance, we want to find $\Phi(z)$ when $z = 0.23$. Then, we select the row having 0.2 in its left margin and the column having 0.03 in its top margin ($z = 0.23 = 0.2 + 0.03$). The associated value in the table is 0.5910, and so $\Phi(0.23) = 0.5910$. In a similar fashion, we find, for example,

$$\Phi(0.07) = 0.5279, \quad \Phi(1.94) = 0.9738, \quad \Phi(2.45) = 0.9929, \quad \Phi(3.02) = 0.9987,$$

and so on (you may check and practice the use of this table).

Note that the table only gives the values of the distribution function for nonnegative values of z. What happens if we want $\Phi(z)$ when $z < 0$? A glance at the values in Table 7.1, which has negative values of z also reveals that, for the values of z in that table, we have

$$\Phi(z) + \Phi(-z) = 1.$$

This is in fact valid for any $z \in \mathbb{R}$, as the next proposition shows.

Proposition 7.4 *The distribution function* Φ *of the standard normal distribution satisfies*

$$\Phi(-z) = 1 - \Phi(z), \quad -\infty < z < \infty.$$

Proof: This is a simple consequence of the symmetry of the standard normal distribution mentioned earlier. More specifically, we obtain for any real z,

$$\Phi(-z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-z} \mathrm{e}^{-y^2/2}\, \mathrm{d}y = \frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x,$$

by making the substitution $y = -x$ in the second step. But the standard normal density function ϕ integrates to one over the interval $(-\infty, \infty)$, which means that

$$\frac{1}{\sqrt{2\pi}} \int_{z}^{\infty} \mathrm{e}^{-x^2/2}\, \mathrm{d}x = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \mathrm{e}^{-x^2/2}\, \mathrm{d}x = 1 - \Phi(z),$$

and the required result follows from the last two expressions. □

Applying this result for $z = 0$, we obtain $\Phi(0) = 0.5$. Also, we have

$$P(-1 \le Z \le 1) = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\Phi(1) - 1,$$

and similarly we obtain

$$P(-2 \le Z \le 2) = 2\Phi(2) - 1, \quad P(-3 \le Z \le 3) = 2\Phi(3) - 1.$$

By inserting the values of $\Phi(z)$, for $z = 1, 2, 3$, into the last expressions, we readily obtain

$$P(-1 \le Z \le 1) = 2(0.8413) - 1 = 0.6826 \cong 68\%, \tag{7.9}$$

$$P(-2 \le Z \le 2) = 2(0.9772) - 1 = 0.9544 \cong 95\%$$

and

$$P(-3 \le Z \le 3) = 2(0.9987) - 1 = 0.9974 \cong 99.7\%.$$

Thus, we see that 68% of the area under the standard normal curve lies between the values -1 and 1, 95% lies between -2 and 2, while nearly all the area (99.7%) is between the values $z = -3$ and $z = 3$ (see Figure 7.3).

Example 7.4 (Quantiles of the standard normal distribution) We have just seen how we can use the table of values for the standard normal distribution to find, for a given real z, the corresponding probability $\Phi(z) = P(Z \le z)$. In many cases, interest lies in the opposite direction, so that for a given value α in the interval $(0, 1)$, we want to find the value of z, denoted by z_α, such that

$$P(Z > z_\alpha) = \alpha.$$

In such a case, the value z_α is called the (**upper**) **α-quantile** of the distribution Φ (Figure 7.4).

Find the α-quantiles of Φ for the cases when

(i) $\alpha = 0.01$, (ii) $\alpha = 0.05$, (iii) $\alpha = 0.10$.

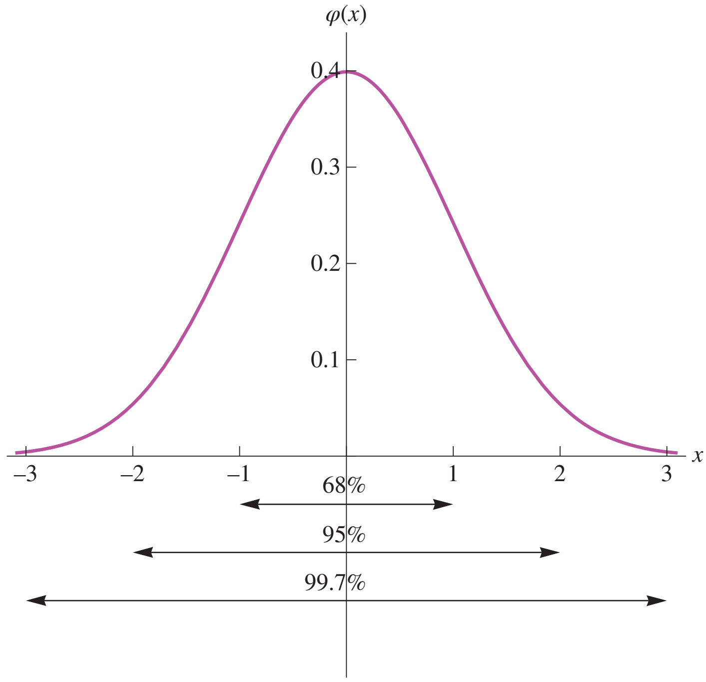


Figure 7.3 Area of the standard normal density function inside each of the intervals $[-1, 1]$, $[-2, 2]$, and $[-3, 3]$.

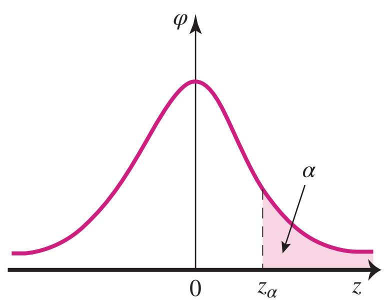


Figure 7.4 Graphical representation of the α-quantile of the standard normal distribution.

SOLUTION Since for any real z

$$P(Z > z) = 1 - P(Z \leq z) = 1 - \Phi(z),$$

we see that, if z_α is the upper α-quantile of Φ, we have $1 - \Phi(z_\alpha) = \alpha$, or

$$\Phi(z_\alpha) = 1 - \alpha.$$

We thus look at the tables of the standard normal distribution and we seek, for different values of α, the closest z-value (up to two decimal places) for which the value of the distribution function is $1 - \alpha$.

(i) For $\alpha = 0.01$, the last equation becomes $\Phi(z_{0.01}) = 0.99$, and we see from the table in Appendix B that the corresponding quantile is

$$z_{0.01} \cong 2.33.$$

(ii) Similarly for $\alpha = 0.05$, we get that $\Phi(z_{0.05}) = 0.95$, and from the tables we find the upper 0.05-quantile to be

$$z_{0.05} \cong 1.645$$

(since the value appears to be halfway between 1.64 and 1.65).

(iii) Now we have $\Phi(z_{0.10}) = 0.10$ and this gives, by using the table in Appendix B again, that

$$z_{0.10} \cong 1.28.$$

Next, we calculate the mean and variance of the standard normal distribution. Let Z be a random variable having this distribution. For the expectation of Z, we have

$$E(Z) = \int_{-\infty}^{\infty} z\phi(z)\,\mathrm{d}z = \int_{-\infty}^{0} z\phi(z)\,\mathrm{d}z + \int_{0}^{\infty} z\phi(z)\,\mathrm{d}z. \tag{7.10}$$

Setting $g(z) = z\phi(z)$, since the function ϕ satisfies $\phi(-z) = \phi(z)$, we obtain that, for any real z,

$$g(-z) = -g(z).$$

Putting $x = -z$ in the first integral on the right of (7.10), we deduce that

$$E(Z) = -\int_{0}^{\infty} g(x)\,\mathrm{d}x + \int_{0}^{\infty} g(z)\,\mathrm{d}z.$$

Since both integrals on the right are finite,[2] we obtain immediately that

$$E(Z) = 0.$$

For the variance

$$\mathrm{Var}(Z) = E(Z^2) - [E(Z)]^2 = E(Z^2),$$

we have that

$$\mathrm{Var}(Z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z^2 \mathrm{e}^{-z^2/2}\,\mathrm{d}z = -\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} z(\mathrm{e}^{-z^2/2})'\,\mathrm{d}z$$

and by integration by parts, we get

$$\mathrm{Var}(Z) = -\frac{1}{\sqrt{2\pi}} \left[z\mathrm{e}^{-z^2/2}\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \phi(z)\mathrm{d}z = 0 + 1 = 1.$$

[2] It is easy to see that this is true: observe first that $g(z) = -\phi'(z)$, so that $\int_0^\infty g(z)\mathrm{d}z = -[\phi(z)]_0^\infty = \phi(0) < \infty$.

Although certain real quantities can be modeled by the standard normal distribution in a satisfactory way, the lack of parameters of this distribution (unlike the binomial distribution, which has two parameters, n and p, or the $\mathcal{U}[a, b]$ distribution, whose parameters are the endpoints of its range, a and b) renders it to be inflexible for modeling in practice. For example, if a quantity is modeled by the standard normal distribution, its mean must necessarily be equal to zero, which may be violated in many situations. For this reason, we now define a general form of the normal distribution, i.e. essentially the family of normal distributions.

Definition 7.3 A random variable X is said to have the normal distribution with parameters $\mu \in \mathbb{R}$ and $\sigma^2 > 0$ if its density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \qquad -\infty < x < \infty.$$

In this case, we use the notation $N(\mu, \sigma^2)$, and we write in short $X \sim N(\mu, \sigma^2)$.

The parameter σ in the definition above is assumed to be positive ($\sigma > 0$).

It is obvious that the standard normal distribution is a particular case of a normal distribution when $\mu = 0$ and $\sigma = 1$. For the function f in the above definition, we see that it takes nonnegative values and, in addition, we have

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, \mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \sigma \mathrm{e}^{-z^2/2}\, \mathrm{d}z = \frac{1}{\sqrt{2\pi}} I = 1,$$

making the substitution $z = (x - \mu)/\sigma$ and using the value of the integral I defined in (7.8), that was found earlier. Hence, f satisfies the properties of a density function. Some further mathematical properties of this function are presented in the following result.

Proposition 7.5 *(Properties of the density function of the $N(\mu, \sigma^2)$ distribution)*

(i) *The function f has a unique local maximum (which is, in fact, a global maximum) point at $x = \mu$, and the maximum value of f at this point is $(\sigma\sqrt{2\pi})^{-1}$;*

(ii) *The function f is symmetric around the point $x = \mu$, which means that for any real x, we have $f(\mu + x) = f(\mu - x)$;*

(iii) *The points $\mu \pm \sigma$ are inflection points of f.*

Proof:

(i) Differentiating $f(x)$ with respect to x, we get

$$f'(x) = \frac{1}{\sigma^3\sqrt{2\pi}} \mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}\, (\mu - x),$$

from which we readily see that $f'(x) > 0$ for $x < \mu$ and $f'(x) < 0$ for $x > \mu$. This shows that f is strictly increasing over the interval $(-\infty, \mu)$ and strictly decreasing over the interval (μ, ∞), so that μ is indeed a maximum point for f. The value of the maximum at $x = \mu$ is found immediately by substitution.

(ii) For any real x, we see that

$$f(\mu+x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-[(\mu+x)-\mu]^2/2\sigma^2}=\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2}$$

and

$$f(\mu-x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-[(\mu-x)-\mu]^2/2\sigma^2}=\frac{1}{\sigma\sqrt{2\pi}}e^{-x^2/2}.$$

Thus, $f(\mu+x)=f(\mu-x)$ and so f is symmetric around the point μ.

(iii) Now we find the second derivative of $f(x)$ with respect to x, which is

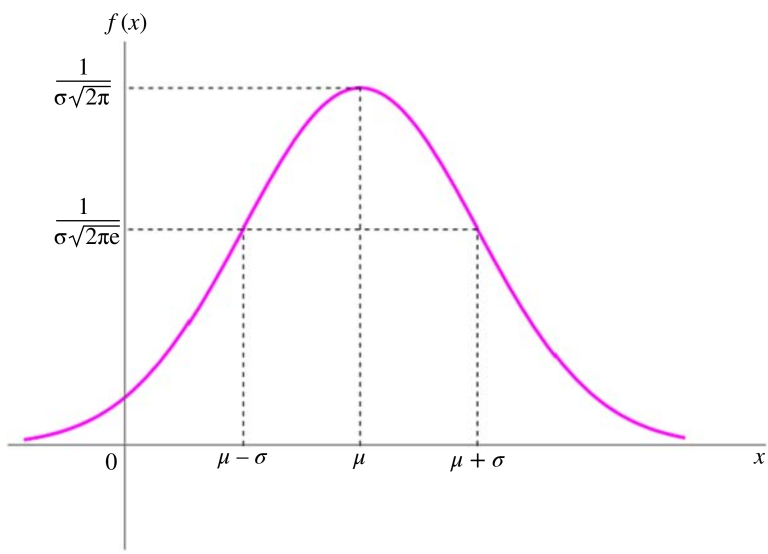
$$f''(x)=[x-(\mu+\sigma)][x-(\mu-\sigma)]\frac{1}{\sigma^5\sqrt{2\pi}}e^{-(x-\mu)^2/(2\sigma^2)}.$$

We observe that this function changes sign at the points $x=\mu+\sigma$ and $x=\mu-\sigma$. We note that the value of the function f at each of these points equals

$$\frac{1}{\sigma\sqrt{2\pi}}\ e^{-0.5}.$$

□

The following figure shows the maximum and the inflection points for the $N(\mu,\sigma^2)$ distribution, according to Proposition 7.5.



We note in particular that for the case $\mu=0,\sigma^2=1$, we see that the density $\phi(z)$ of the standard normal distribution is symmetric around the y-axis, has a maximum at $z=0$ (with maximum value $\phi(0)=(\sqrt{2\pi})^{-1}\cong 0.4$), while the points $z=\pm 1$ are inflection points for ϕ.

Having studied the mathematical properties of the normal density function, we now present some more graphs to get a better understanding of how this function looks like for various values of the parameters μ and σ^2. Figure 7.5 shows three density functions of the normal distribution with the same mean, $\mu=2$ and different values for σ, namely, $\sigma=0.5$, $\sigma=1$, and $\sigma=2$.

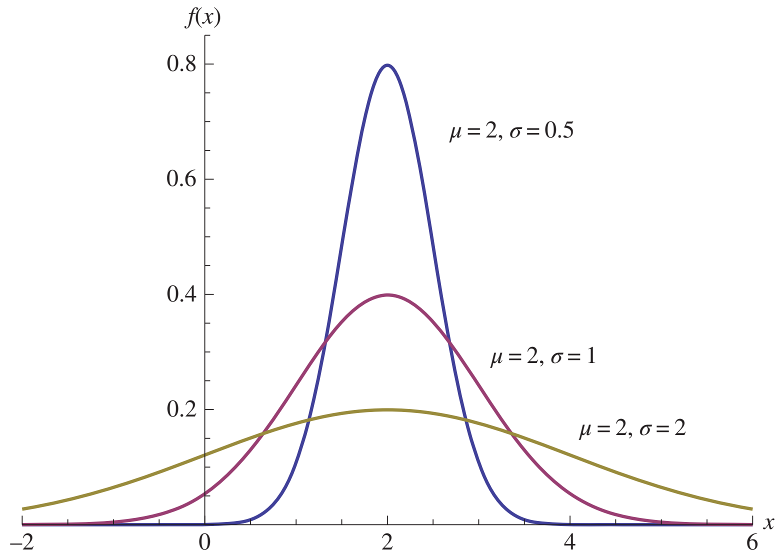


Figure 7.5 A comparison of the density functions for the $N(2, \sigma^2)$ distributions, for $\sigma = 0.5,\ 1,$ and 2.

Figure 7.6 depicts the density function and the distribution function of the $N(\mu, \sigma^2)$ distribution for different values of its parameters μ and σ^2.

Next, we have seen in the previous section that a linear combination of a uniform random variable also has a uniform distribution. This is not a common property among probability distributions (e.g. it is *not satisfied* by the discrete distributions we met in Chapter 5), and it is natural to wonder whether this holds for the normal distribution. A result in this direction is Part (i) of the following proposition; for the (affirmative) answer to the general problem, see Exercise 22 at the end of this section. The result of the proposition that follows enables us also to calculate probabilities associated with a random variable $X \sim N(\mu, \sigma^2)$ for arbitrary values of μ and σ.

Proposition 7.6 *(Calculation of probabilities for $N(\mu, \sigma^2)$) If the random variable X has the $N(\mu, \sigma^2)$ distribution, then we have the following:*

(i) *The variable*

$$Z = \frac{X - \mu}{\sigma}$$

follows the standard normal distribution, that is, $Z \sim N(0, 1)$;

(ii) *The probability $P(a \le X \le b)$ can be found from the formula*

$$P(a \le X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right), \quad a < b;$$

(iii) *For any real a and b, the probabilities $P(X \le b)$ and $P(X > a)$ can be found by using the relations*

$$P(X \le b) = \Phi\left(\frac{b-\mu}{\sigma}\right), \quad P(X > a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{\mu-a}{\sigma}\right).$$

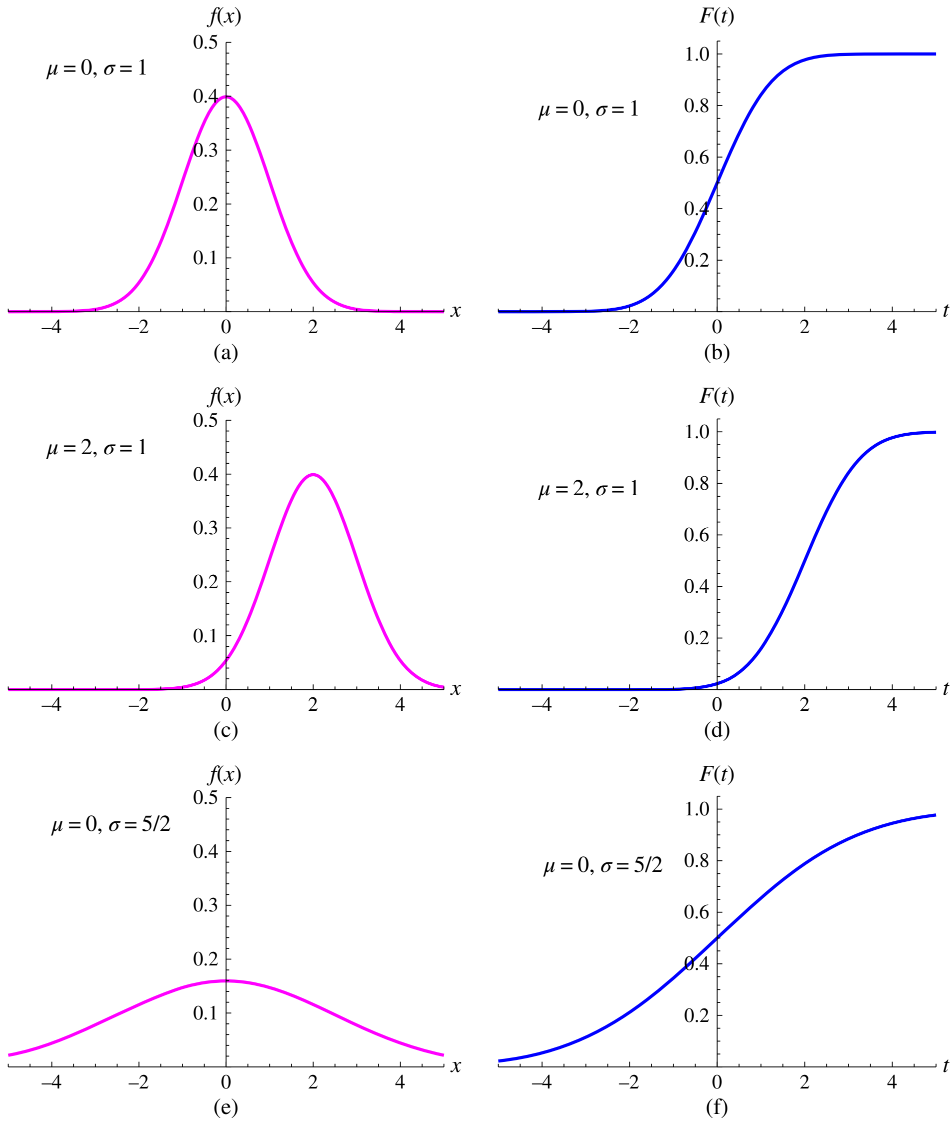


Figure 7.6 The density function (left) and the distribution function (right) of the normal distribution, for various values of μ and σ^2.

Proof:

(i) Let us denote by f and F the density function and the distribution function of the variable X, respectively. Then the distribution function of $Z = (X - \mu)/\sigma$ is given by (see also Example 6.3)

$$F_Z(z) = \left(\frac{X-\mu}{\sigma} \le z\right) = P(X \le \mu + \sigma z) = F(\mu + \sigma z),$$

and by differentiation we obtain the density of Z to be

$$f_Z(z) = F_Z'(z) = \sigma f(\mu + \sigma z) = \sigma \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{[(\mu+\sigma z)-\mu]^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} = \phi(z),$$

where $\phi(.)$ is the density function of the standard normal distribution (Definition 7.2). Thus, we see that $Z \sim N(0, 1)$.

(ii) We have

$$P(a \leq X \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right), \tag{7.11}$$

where the variable $Z = (X-\mu)/\sigma$ has the standard normal distribution, with distribution function $\Phi(z)$. The result then follows.

(iii) The result for this part can be obtained from Part (ii) by letting $a \to -\infty$ and $b \to \infty$, respectively. Alternatively, we may argue directly as follows:

$$P(X \leq b) = P\left(\frac{X-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(Z \leq \frac{b-\mu}{\sigma}\right) = \Phi\left(\frac{b-\mu}{\sigma}\right)$$

and

$$P(X > a) = 1 - P(X \leq a) = 1 - \Phi\left(\frac{a-\mu}{\sigma}\right) = \Phi\left(-\frac{a-\mu}{\sigma}\right) = \Phi\left(\frac{\mu-a}{\sigma}\right),$$

using Proposition 7.4 for the penultimate step above. □

The most important message from the above proposition is that, when we are dealing with a problem related to an arbitrary normal distribution, we can convert it to a problem of the standard normal distribution and then use the table of the latter to calculate the required probabilities. This method, which involves subtracting the mean from an arbitrary normal random variable, and then dividing by its standard deviation, is illustrated in (7.11), and is called **standardization** of that variable.

The following proposition shows that the parameters μ and σ^2 of the normal distribution coincide with the mean and the variance of that distribution, respectively (this explains the choice of these symbols for the parameters of a normal distribution, in fact).

Proposition 7.7 *The expected value, the variance and the standard deviation of a random variable following the $N(\mu, \sigma^2)$ distribution are given, respectively, by*

$$E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2, \quad \sqrt{\mathrm{Var}(X)} = \sigma.$$

Proof: This is an immediate consequence of Proposition 7.6 and the properties of expectation and the variance we have seen in earlier chapters. In particular, if $Z = (X-\mu)/\sigma \sim N(0, 1)$, then X can be written as

$$X = \mu + \sigma Z,$$

so that we obtain

$$E(X) = E(\mu + \sigma Z) = \mu + \sigma E(Z), \quad \mathrm{Var}(X) = \mathrm{Var}(\mu + \sigma Z) = \sigma^2 \mathrm{Var}(Z),$$

and the result is now obvious since $E(Z) = 0$ and $\mathrm{Var}(Z) = 1$. □

Example 7.5 It is common among social scientists to believe that the distribution of IQ scores in the general population follows a normal distribution with a mean value $\mu = 100$ and a standard deviation $\sigma = 15$. Let X denote the IQ score of a randomly chosen person from the population.

(i) The probability that the person has an IQ score of at least 130 is

$$\begin{aligned} P(X \geq 130) &= P\left(\frac{X-\mu}{\sigma} \geq \frac{130-\mu}{\sigma}\right) = P\left(Z \geq \frac{130-100}{15}\right) \\ &= 1 - P(Z < 2) = 1 - \Phi(2), \end{aligned}$$

and from the standard normal table we see that $\Phi(2) = 0.9772$, so that the required probability is 0.0228 or 2.28%. Since the person was selected randomly from the population, this can also be interpreted as the proportion of persons in the population with an IQ score of 130 or more.

(ii) The proportion of persons who achieve an IQ score between 100 and 120 is

$$\begin{aligned} P(100 \leq X \leq 120) &= P\left(\frac{100-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{120-\mu}{\sigma}\right) \\ &= P\left(\frac{100-100}{15} \leq Z \leq \frac{120-100}{15}\right) \\ &= \Phi\left(\frac{4}{3}\right) - \Phi(0). \end{aligned}$$

Since $4/3 \cong 1.33$, we see from the standard normal table that $\Phi(4/3)$ is approximately equal to 0.9082 (a more accurate result could be achieved by interpolating between the values of $\Phi(x)$ for $x = 1.33$ and 1.34 in the table). We already know that $\Phi(0) = 0.5$, and we therefore get

$$P(100 \leq X \leq 120) = 0.9082 - 0.5 = 0.4082,$$

that is, about 40.8% of people in the population have an IQ score between 100 and 120.

Example 7.6 The time that John needs to get from his house to the University every morning is a continuous random variable which is assumed to follow the normal distribution with mean $\mu = 35$ minutes and a standard deviation $\sigma = 5$ minutes.

(i) Find the probability that on a particular day his journey takes

(a) less than 30 minutes;

(b) between 30 and 40 minutes.

(ii) Tomorrow, John's first lecture starts at 10:15 a.m. and he does not want to be late. Estimate what time he should leave his house so that he arrives at the classroom before the lecture starts with probability 99%.

SOLUTION Let X be the time that John's journey takes on that day. Then, it is given that $X \sim N(35, 5^2)$.

(i) (a) The required probability is $P(X < 30)$ which is

$$P(X < 30) = P\left(\frac{X-\mu}{\sigma} < \frac{30-\mu}{\sigma}\right) = P\left(\frac{X-35}{5} < \frac{30-35}{5}\right)$$
$$= P\left(Z < \frac{-5}{5}\right) = P(Z < -1) = \Phi(-1) = 1 - \Phi(1),$$

wherein we have used Proposition 7.4 in the last step (recall that the table in Appendix B gives $\Phi(z)$ only for nonnegative z). From this table, we see that $\Phi(1) = 0.8413$, and so the required result is $1 - 0.8413 = 0.1587$, i.e. a probability of about 16%.

(b) The probability we seek for this part is $P(30 \le X \le 40)$, which is

$$P(30 \le X \le 40) = P\left(\frac{30-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{40-\mu}{\sigma}\right)$$
$$= P\left(\frac{30-35}{5} \le \frac{X-35}{5} \le \frac{40-35}{5}\right)$$
$$= P(-1 \le Z \le 1),$$

and we have already seen in (7.9) that this equals 0.6826. Therefore, the probability that John's journey to the University will take between 30 and 40 minutes is about 68%.

(ii) Now the problem is of a different type. Specifically, we want to find the value of x such that the journey time, X, is at most x with probability 99%, i.e.

$$P(X \le x) = 0.99. \tag{7.12}$$

For this, we express the probability $P(X \le x)$ in terms of the standard normal distribution, working in a familiar way, as follows:

$$P(X \le x) = P\left(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\right) = P\left(Z \le \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right).$$

If we put for simplicity $z = (x-\mu)/\sigma$, we then want to find the value of z for which

$$\Phi(z) = 0.99.$$

So, in this case, we are given the (cumulative) probability $P(Z \le z)$ and we seek the value of z. From the standard normal table in Appendix B, we see that this value is $z = 2.33$. Since the value of x which satisfies (7.12) is given by $x = \mu + \sigma z$, for $z = 2.33$ we finally obtain

$$x = \mu + \sigma z = 35 + 5 \cdot 2.33 = 46.65 \cong 47 \text{ minutes}.$$

This is the maximum journey time that John should allow himself in order to arrive in time for the lecture with probability 99%. Since the lecture starts at 10:15 a.m., this means that he should leave his house by 9:28 a.m.

Suppose now X represents a random quantity of interest, e.g. the height of a student in a University, and that $X \sim N(\mu, \sigma^2)$. When we take a set of measurements (observations) from X, such as a set of students' heights, then the percentage of these measurements that differ from μ by k standard deviations at most will be

$$P(|X-\mu| \le k\sigma) = P\left(\left|\frac{X-\mu}{\sigma}\right| \le k\right) = P(|Z| \le k) = P(-k \le Z \le k).$$

From this, we get

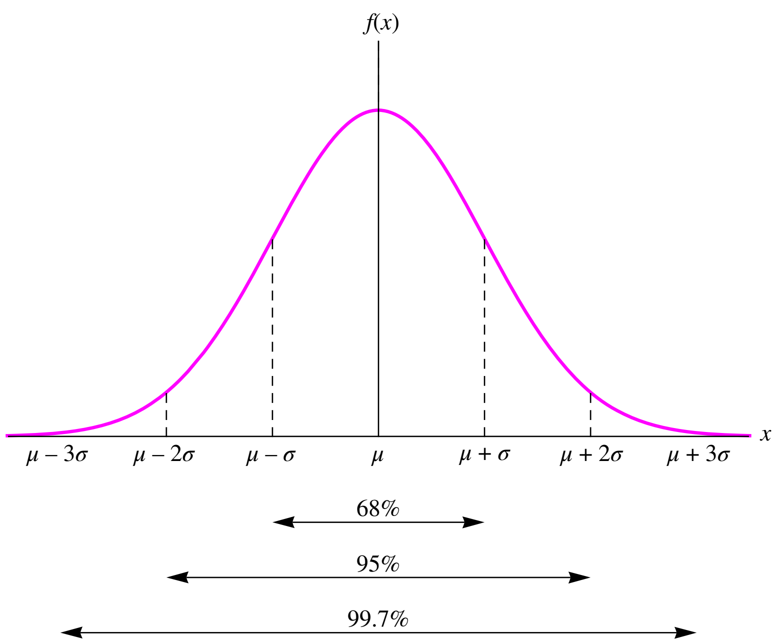
$$P(|X-\mu| \le \sigma) = P(-1 \le Z \le 1) = 2(0.8413) - 1 = 0.6826 \cong 68\%,$$
$$P(|X-\mu| \le 2\sigma) = P(-2 \le Z \le 2) = 2(0.9772) - 1 = 0.9544 \cong 95\%$$

and

$$P(|X-\mu| \le 3\sigma) = P(-1 \le Z \le 1) = 2(0.9987) - 1 = 0.9974 \cong 99.7\%$$

(see the calculations after the proof of Proposition 7.4).

Thus, *about 68% of the values from a normal population are within one standard deviation distance from the mean μ, about 95% of the values are within two standard deviations distance from the mean, while about 99.7% of the values are within three standard deviations distance from the mean of that population*.



The above conclusions may be used to justify the use of normal distribution as a model for random quantities which assume only nonnegative values. This has been done earlier in Examples 7.5 and 7.6. For instance, in Example 7.6 X is John's traveling time so that, in reality, X cannot take negative values. Can we assume that such a variable follows a normal distribution? At first sight, this might seem inappropriate, since any random variable which has a $N(\mu, \sigma^2)$ distribution takes, by Definition 7.3, any real value from $-\infty$ to $+\infty$. However, the calculations above illustrate that if μ is larger than 3σ, the probability that X takes negative values will be very small, and so in most cases it can be ignored. Consider for instance the situation in Example 7.6. We have $\mu = 35$ and $\sigma = 5$, so that the probability that John's traveling time X takes a negative value is

$$P(X < 0) = P\left(\frac{X - 35}{5} < \frac{0 - 35}{5}\right) = P(Z < -7) = \Phi(-7),$$

which is of order 10^{-12}.

In view of the above, we shall freely use the normal distribution in the sequel as a model for quantities that cannot be negative by their nature, provided that the mean of that distribution is at least three times larger than the standard deviation. In fact, due to the several advantages that the normal distribution possesses as a model, this distribution has been used widely in practice to describe quantities such as length, height, lifetimes, time duration of a certain event, etc. (bear in mind that a probability distribution we use to model, i.e. to represent the uncertainty about a real phenomenon, is always only an approximation).

Example 7.7 An automatic machine is used to dispense a particular soft drink in bottles with a nominal quantity (volume) of 1.5l. If the machine puts more than 1.6l of the soft drink in a bottle, the excess quantity is wasted.

The actual volume that the machine dispenses in the bottles is a random variable which has the normal distribution with mean μ (in liters) and a standard deviation 0.02l. What is the mean quantity μ that should be set for the machine so that the probability that it dispenses more than 1.6l in a bottle is 0.2%?

SOLUTION Let X denote the random variable corresponding to the amount dispensed by the machine in a bottle. We are given that $X \sim N(\mu, (0.02)^2)$, and we then seek the value of μ such that

$$P(X > 1.6) = 0.002.$$

Since, by Proposition 7.6,

$$P(X > 1.6) = 1 - \Phi\left(\frac{1.6 - \mu}{0.02}\right),$$

we want to find μ such that

$$1 - \Phi\left(\frac{1.6 - \mu}{0.02}\right) = 0.002$$

which gives

$$\Phi\left(\frac{1.6 - \mu}{0.02}\right) = 1 - 0.002 = 0.998.$$

From the tables of the standard normal distribution, we see that the value z such that $\Phi(z) = 0.998$ is $z = 2.05$. Since Φ is a one-to-one function (as it is strictly increasing), we deduce that we must have

$$\frac{1.6 - \mu}{0.02} = 2.05.$$

Upon solving this, we obtain $\mu = 1.6 - (0.02)(2.05) = 1.559$, which is the desired result.

We have mentioned earlier in this section that the normal distribution offers a satisfactory approximation to the binomial distribution. Figure 7.7 depicts the probability function of the binomial distribution when $p = 0.6$ and $n = 1, 2, 5, 10, 25, 100$.

It is apparent that as n increases, the shape of the probability function resembles that of a normal distribution. Figure 7.8 shows, on the same graph, the binomial probability function

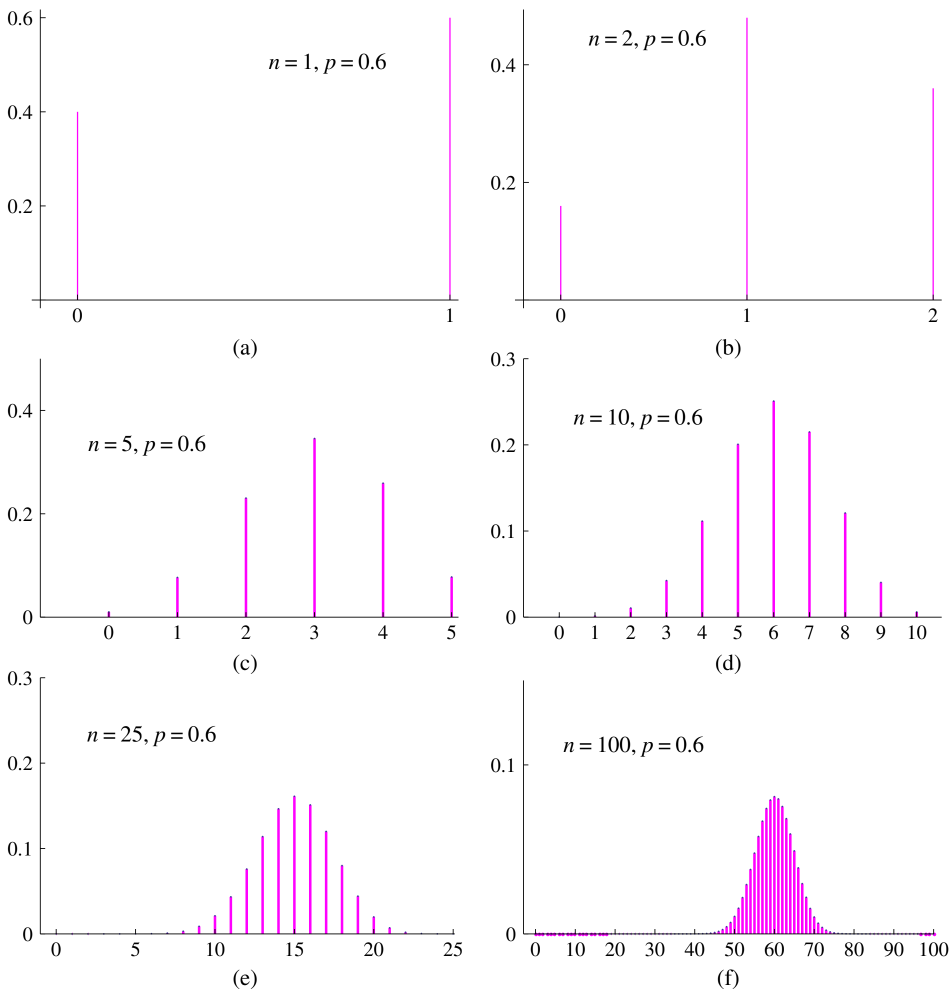


Figure 7.7 The probability function of the binomial distribution function for $p = 0.6$ and various values of n.

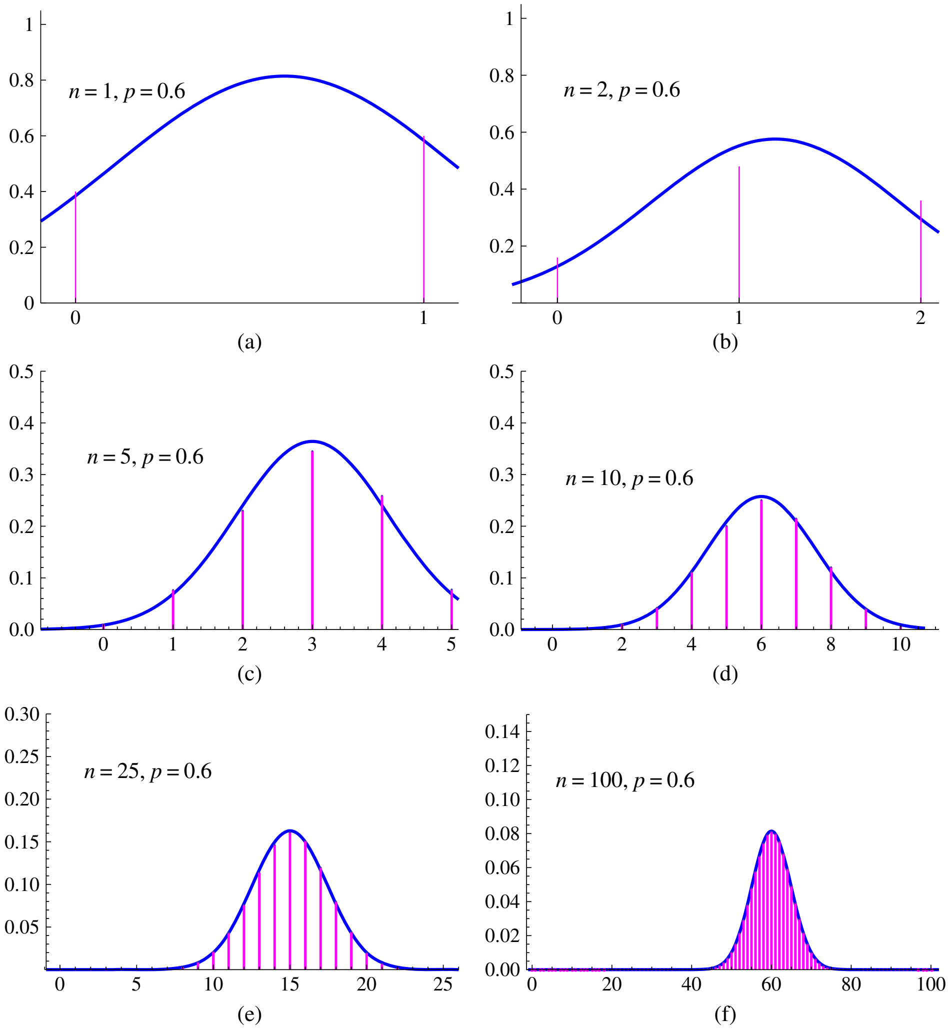


Figure 7.8 Binomial probability function and approximating normal form.

with parameters n and p and the density function of the $N(\mu, \sigma^2)$ distribution so that the two distributions have the same mean and variance, i.e. the values of the parameters are chosen in each case such that

$$\mu = np, \quad \sigma^2 = npq = np(1-p).$$

Observe that for large n ($n = 100$) the graphs of the two distributions are very close to each other.

The De Moivre–Laplace theorem (Proposition 7.3) offers indeed the theoretical validation to the above comments. Specifically, if X is a random variable following the binomial distribution with parameters n and p, then we can write

$$P(a < X < b) = P\left(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}\right)$$

and the De Moivre–Laplace theorem suggests the approximation, for large values of n,

$$P(a < X < b) = \Phi\left(\frac{b - np}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np}{\sqrt{npq}}\right).$$

In general, the normal distribution approximates very well the probabilities of the binomial distribution when npq is large. However, we must bear in mind that in the De Moivre–Laplace theorem, we approximate a discrete distribution by a continuous one. For this reason, and in particular when the quantity npq is of moderate magnitude, a simple correction to that approximation which is an improvement of that theorem, is often used.

Assume that X is a discrete random variable and suppose that we are interested in calculating the probability $P(i \le X \le j)$, where i and j are integers. This probability can be expressed as

$$P(i \le X \le j) = \sum_{k=i}^{j} P(X = k)$$

and equals the shaded area in Figure 7.9. If we attempt now to approximate this area by a smooth curve f (associated with a continuous distribution), it seems natural to use the area surrounded by the x-axis, the graph of f and the perpendicular axes

$$x = i - \frac{1}{2} \quad \text{and} \quad x = j + \frac{1}{2}$$

as an approximation for the probability $P(i \le X \le j)$ (see Figure 7.10).

The use of the terms $i - 0.5$ (for the lower limit) and $j + 0.5$ (for the upper limit) above is known as **continuity correction**, and is considered appropriate, whenever we approximate a discrete distribution by a continuous one. Using similar arguments, we can justify the approximate formulas

$$P(X = i) \cong \int_{i-0.5}^{i+0.5} f(x)\,\mathrm{d}x,$$

$$P(X \ge i) \cong \int_{i-0.5}^{\infty} f(x)\,\mathrm{d}x,$$

$$P(X \le j) \cong \int_{-\infty}^{j+0.5} f(x)\,\mathrm{d}x.$$

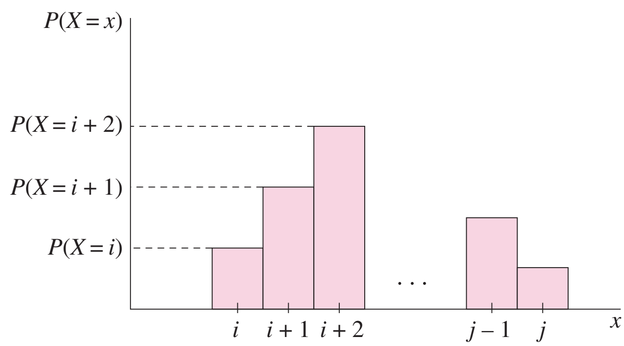


Figure 7.9 Shaded region is $P(i \le X \le j)$ for a discrete variable.

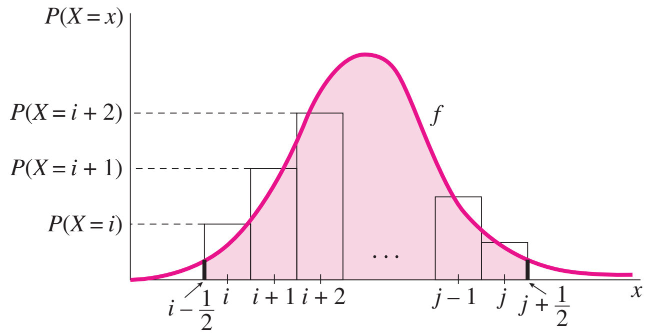


Figure 7.10 A corrected normal approximation to discrete probabilities.

Especially for the case when X has the binomial distribution with parameters n and p, we know that the corresponding normal approximation is made through the $N(\mu, \sigma^2)$ distribution, with $\mu = E(X) = np$ and $\sigma^2 = \text{Var}(X) = npq$. Consequently, we have in this case

$$
\begin{aligned}
P(i \le X \le j) &\cong \int_{i-0.5}^{j+0.5} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}\, dx \\
&= P\left(\frac{(i-0.5)-\mu}{\sigma} \le Z \le \frac{(j+0.5)-\mu}{\sigma}\right),
\end{aligned}
\tag{7.13}
$$

where $Z \sim N(0, 1)$, so that we finally arrive at the corrected approximation

$$
P(i \le X \le j) \cong \Phi\left(\frac{(j+0.5)-np}{\sqrt{npq}}\right) - \Phi\left(\frac{(i-0.5)-np}{\sqrt{npq}}\right).
$$

Similar relations hold for probabilities of the form $P(X = i), P(X \ge i)$ or $P(X \le j)$.

Example 7.8 Assume that 20% of people who take up a particular diet drop out before they complete it. If we select a random sample of 100 persons who have started this diet, what is the probability that more than a quarter of them will not complete the diet program?

SOLUTION Suppose X represents the number of persons in the sample who do not complete their diet schedule. It then follows from the statement that $X \sim b(n, p)$ with $n = 100$ and $p = 0.2$. Therefore, we have from the formula for binomial probabilities that

$$
P(X = k) = \binom{100}{k}(0.2)^k(0.8)^{n-k}, \quad k = 0, 1, \ldots, 100.
$$

The probability we require is $P(X > 25) = P(X \ge 26)$, since X takes only integer values. The exact calculation for this probability requires evaluation of the sum

$$
P(X \ge 26) = \sum_{k=26}^{100} \binom{100}{k}(0.2)^k(0.8)^{100-k}. \tag{7.14}
$$

Even with the aid of modern computers, some effort is required to find this, such as writing a little program in Mathematica (see Exercise 5 in Section 7.6). However, using the normal approximation to the binomial distribution (formula (7.13) with $i = 26, j = \infty$ there) instead, we obtain

$$P(X \geq 26) \cong P\left(Z \geq \frac{(26-0.5) - 100 \cdot (0.2)}{\sqrt{100 \cdot (0.2) \cdot (0.8)}}\right) = P(Z \geq 1.375)$$
$$= 1 - \Phi(1.375) = 1 - 0.9154 = 0.0846.$$

It is worth mentioning that the exact value for the required probability, calculated with Mathematica, is 0.0875. If we use a normal approximation without a continuity correction, we get

$$P(X \geq 26) \cong P\left(Z \geq \frac{26 - 100 \cdot (0.2)}{\sqrt{100 \cdot (0.2) \cdot (0.8)}}\right) = P(Z \geq 1.5) = 1 - \Phi(1.5) = 0.0668,$$

which is a much poorer approximation to the true value.

Example 7.9 Suppose we want to estimate the proportion p (or equivalently, the percentage $100p\%$) of the persons who intend to vote for a certain political party in the forthcoming general elections. For this reason, we plan to take a sample of size n of voters and ask them about their intention (assuming they are all going to vote), from the entire population of eligible voters. How large should the sample size n be so that the error we make in this estimate (sometimes in everyday language called the "statistical error") does not exceed 1%?

SOLUTION This is a sampling problem. We have seen problems of this type in relation to the hypergeometric distribution earlier in Chapter 5. The problem here, as is very often in case of opinion polls, is to determine the sample size. If we select n persons from the population of eligible voters, the number, X, who will vote in favor of the party is a random variable which follows the binomial distribution with parameters n and p (neither of which is known at present). Then, $\hat{p} = X/n$ will be our estimate for p and we want the following condition to be satisfied:

$$|p - \hat{p}| \leq 0.01. \tag{7.15}$$

We thus see that the problem here is stated in an ill-posed manner! No matter how large n is we cannot be sure that the above condition is met, unless we select the whole population. Otherwise it is even possible, for example, that no one in our sample votes in favor of the party in question. The best we can do to overcome this is to impose the condition that (7.15) holds with a high preassigned probability, such as 95%. Consequently, the problem now becomes that of finding n such that (7.15) holds with a probability of (at least) 0.95, i.e.

$$P(|p - \hat{p}| \le 0.01) = P(|np - n\hat{p}| \le 0.01n) \ge 0.95.$$

But, $X = n\hat{p}$ has a $b(n, p)$ distribution and since we expect n to be large (typically several hundreds of persons will be needed for opinion polls of this type), we approximate the distribution of X by the $N(\mu, \sigma^2)$ distribution, with

$$\mu = np, \quad \sigma^2 = np(1 - p).$$

We thus obtain

$$\begin{aligned} P(|np - n\hat{p}| \le 0.01n) &= P(|\mu - X| \le 0.01n) = P(-0.01n \le X - \mu \le 0.01n) \\ &= P\left(\frac{-0.01n}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{0.01n}{\sigma}\right). \end{aligned}$$

Setting for simplicity $\alpha = 0.01n/\sigma$, we see that the last expression equals $\Phi(\alpha) - \Phi(-\alpha)$. But this, in turn, is

$$\Phi(\alpha) - \Phi(-\alpha) = \Phi(\alpha) - [1 - \Phi(\alpha)] = 2\Phi(\alpha) - 1,$$

in view of Proposition 7.4. So, the condition we need in order to determine the sample size n, becomes now $2\Phi(\alpha) - 1 \ge 0.95$; that is, we need to find the value of α such that

$$\Phi(\alpha) \ge \frac{1 + 0.95}{2} = 0.975.$$

From the tables of the standard normal distribution in Appendix B, we see that α must be greater than or equal to 1.96. Recalling now the definition of α, we get

$$\alpha = \frac{0.01n}{\sigma} = \frac{0.01n}{\sqrt{np(1-p)}} \ge 1.96$$

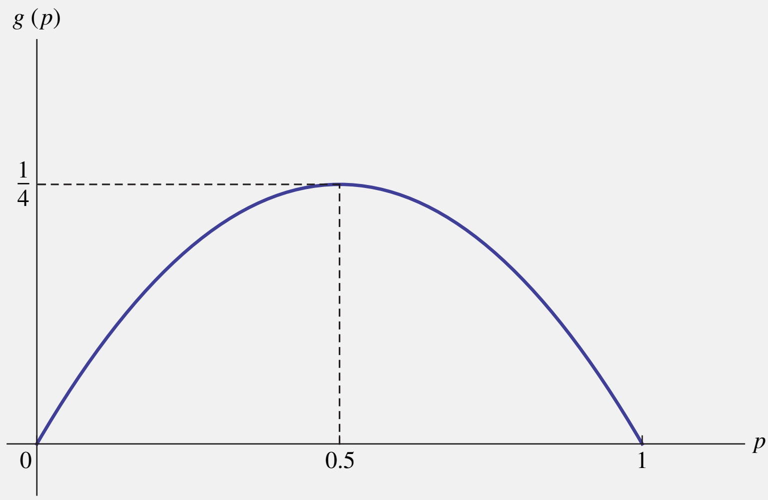
and, solving this for n, we derive that

$$0.01n \ge 1.96\sqrt{np(1-p)},$$

or equivalently

$$\sqrt{n} \ge 1.96\,\frac{\sqrt{p(1-p)}}{0.01} = 196\sqrt{p(1-p)}.$$

Note that the quantity $\sqrt{p(1-p)}$ is unknown since the population proportion of persons voting for that party, p, is not known. However, and since we only require a lower bound for n, we may use the fact that $g(p) = p(1-p) \leq 1/4$ for all values of $p \in [0, 1]$ (see the next figure).



Hence, for any $p \in [0, 1]$, $\sqrt{p(1-p)}$ is at most $1/2$. This finally shows that a minimum sample size of

$$\left(196 \cdot \frac{1}{2}\right)^2 = 9604$$

is sufficient for our purposes.

We close this section by noting that the normal distribution may be used as an approximation to the Poisson distribution (this comes as no surprise since we have seen that both distributions, under suitable conditions, serve as approximations to the binomial distribution). To be more precise, assume that X is a random variable having the Poisson distribution with parameter $\lambda > 0$. Then, **provided $\boldsymbol{\lambda}$ is large enough**, the distribution of X can be approximated by the $N(\mu, \sigma^2)$ distribution with $\mu = E(X) = \lambda$ and $\sigma^2 = \text{Var}(X) = \lambda$. Thus we may write, for example,

$$P(a < X < b) \cong \Phi\left(\frac{b-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{a-\lambda}{\sqrt{\lambda}}\right), \quad a < b,$$

or upon using a continuity correction,

$$P(i \leq X \leq j) \cong \Phi\left(\frac{(j+0.5)-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{(i-0.5)-\lambda}{\sqrt{\lambda}}\right)$$

for any nonnegative integers i and j such that $i < j$. □

EXERCISES

Group A

1. Assume that X is a random variable having the normal distribution with parameters $\mu = 7$ and $\sigma^2 = 4$. Find the probabilities

   (a) $P(X < 7)$

   (b) $P(X \geq 5)$

   (c) $P(X \leq 9 | X > 7)$

   (d) $P[(X - 6)^2 < 1]$

   (e) $P(X = 7)$

   (f) $P(|X - 7| < 2)$

   (g) $P(\ln |X| < 1.95)$.

2. The gestation period for live births in humans can be represented by a random variable X following the normal distribution with mean $\mu = 270$ days and a standard deviation $\sigma = 10$ days. Find the proportion of children who are born after a gestation period

   (i) of less than 240 days;

   (ii) between 255 and 285 days.

3. Suppose that the concentration of sodium (Na) in the human blood follows a normal distribution with mean 140 (measured in mM) and standard deviation 5 mM. Calculate

   (i) the probability that the level of sodium in the blood of a person is

      (a) less than 130;

      (b) between 135 and 145;

      (c) at least 145.

   (ii) the percentage of persons in the population for which the level of sodium is

      (a) between 140 and 150;

      (b) below 130 or above 160.

4. A random variable X has its density function as

$$f(x) = \left(2\sqrt{2\pi}\right)^{-1} e^{-(x+4)^2/8}, \quad x \in \mathbb{R}.$$

   Find the following probabilities:

   (i) $P(X \leq -2)$

   (ii) $P(-5 \leq X \leq -2)$

   (iii) $P(|X + 3| \leq 1)$.

5. In a certain population, the level of cholesterol in the human blood (measured in mg dl^{-1}) follows a normal distribution with mean $\mu = 220$ and standard deviation $\sigma = 40$.

   (i) Find the percentage of persons in that population with a cholesterol level between 200 and 260.

   (ii) For what value c we have 10% of persons in the population to have a cholesterol level of at least c?

6. If X is a random variable having the normal distribution with mean μ and standard deviation σ, and c is a real number such that

$$P(X > c) = 2P(X \leq c),$$

   show that

$$c + 0.43\sigma = \mu.$$

   Application: The level of Fe in the blood for a male population is a random variable having the normal distribution with a mean value of 110 mg dl^{-1} and a standard deviation of 25 mg dl^{-1}. Find the value of c such that the percentage of males whose level of Fe in their blood is above c (mg dl^{-1}) is twice as much as the percentage of those with a value of Fe less than c (mg dl^{-1}).

7. The height of females in a population has the normal distribution with mean 167 cm and a standard deviation 5 cm.

   (i) What percentage of females in that population have a height

      (a) more than 167 cm?

      (b) between 167 and 175 cm?

   (ii) If we select randomly 6 females from that population, what is the probability that

      (a) all six have a height between 167 and 175 cm?

      (b) two persons have a height above the mean and four persons have a height below the mean?

8. Female shirts are classified as S, M, L, and XL according to their size. Shirts of size S are suitable for women with a chest size between 29 and 32 (in inches); size M is suitable for women with chest size between 32 and 34, size L is suitable for women with chest size between 34 and 38, while size XL is suitable for women with a chest size over 38.

   Suppose we select a woman at random from the population, and her chest size, X, has the normal distribution with mean 34.25 and standard deviation 1.75 in.

   (i) Find the proportion of women with a chest size of less than 29 in., so that for them size S is too large.

   (ii) A manufacturing unit that produces female shirts makes 5000 shirts every week. Find how many shirts of each size it should produce, so that its products are in line with demand.

(iii) If we are to redetermine the chest size limits for female shirts, so that each of the four sizes corresponds to 25% of the women population, find the limits which should be used to distinguish the four sizes.

9. Find the probability that in 600 throws of a die, an ace turns up at least 120 times.

   (*Hint*: use a normal approximation to the exact distribution.)

10. The internal diameter, in inches, of copper tubes made at by a factory has a $N(15, \sigma^2)$ distribution. Tubes that fall outside the tolerance limits 15 ± 0.1 in. are recycled.

    (i) If $\sigma = 0.1$, find the probability that exactly 4 tubes are recycled in a sample of 6 tubes which have been inspected.

    (ii) What should be the value of the standard deviation σ so that the percentage of tubes that are recycled is 6%?

11. The daily demand for a product in a shop (in kilograms) has the normal distribution with parameters $\mu = 80$ and $\sigma = 6$.

    (i) If the current stock of the shop is 95 kg, estimate the probability that all this stock will be sold within a day.

    (ii) What is the probability that after the end of the next working day, the shop will have at least 30 kg of the product still in stock?

12. For a certain flight, there are 250 seats available. However, the company which operates the flight estimates that 6% of the passengers who buy a ticket do not show up. Therefore, the company has sold tickets to 265 persons for the flight. Find the probability that

    (i) at least one passenger who turns up for the flight does not find a seat;

    (ii) at least five passengers who turn up for the flight do not find a seat.

13. The price of a certain stock, during a period of 200 trading days, has increased in 115 of these days, while in the remaining days it has decreased. Explain whether it is reasonable to assume that there is an even chance that on a particular day, the stock price increases or decreases.

    (*Hint*: assuming there is an even chance of increase or decrease in each day, find the probability that in *at least 115 days* the price has gone up.)

14. A newsagent shop orders 200 copies of *The Sunday Times* every week. It has been estimated that the number of copies of the newspaper sold weekly has a normal distribution with parameters $\mu = 180$ and $\sigma = 8$.

    (i) Find the probability that, in a given week, the shop sells all 200 copies of the newspaper.

    (ii) What is the probability that in a period of 12 weeks, there are at least two Sundays on which all 200 copies are sold?

    (iii) What is the distribution of the number of weeks until all available copies are sold? What is the expected number of weeks until this happens for the first time?

15. A large number of students take an exam and their scripts are marked on the scale 1–100. The distribution of marks allocated is supposed to be normal with parameters $\mu = 63$ and $\sigma = 9$.

    (i) If the pass mark for this exam is 45, estimate the percentage of students who fail the exam.

    (ii) What proportion of students get a mark between 55 and 70?

    (iii) Estimate the mark which Laura, who is taking this exam, should get so that her performance is in the top 10% of students.

16. A random variable X has probability density function

$$f(x) = c\mathrm{e}^{-8(x-3)^2}, -\infty < x < \infty,$$

where c is a suitable constant.

    (i) Find the value of c.

    (ii) Calculate the mean and variance of X.

17. It is expected that 60% of patients who undertake a certain surgery develop a high level of cholesterol. If 110 patients had this surgery within a month at a hospital, what is the probability that the number of those who developed high cholesterol levels is at least 30, but no more than 50?

18. Under the assumptions of Example 7.5, calculate the probability that when a sample of 100 persons is drawn randomly, at least 3 of them will have an IQ score of 140 or more.

19. The number of customers at a super market during a week follows the Poisson distribution with a mean value of 4650 customers.

    (i) Using a suitable approximation estimate the probability that, in a given week, the number of customers who visit the super market is

        (a) at least 4500;

        (b) between 4500 and 4800.

    (ii) Estimate the probability that in a period of four weeks, there is at least one week in which the super market has less than 4500 customers.

20. Find the value of the constant $a > 0$ for which the function

$$f(x) = \sqrt{a}\, \mathrm{e}^{-a^2x^2-2ax-1}, \quad x \in \mathbb{R},$$

is a valid density function.

21. Assume that a random variable X follows the $N(\mu, \sigma^2)$ distribution. Find the density function for each of the variables:

    (i) $Y_1 = |X| - \mu$;

    (ii) $Y_2 = |X - \mu|$.

22. Suppose $X \sim N(\mu, \sigma^2)$ and $a, b \in \mathbb{R}$ are given real numbers. Show that the distribution of the random variable $Y = aX + b$ is $N(a\mu + b, a^2\sigma^2)$.

23. Let X be a random variable that has the $N(0, \sigma^2)$ distribution. Find the distributions of

    (i) $Y = X^2$;

    (ii) $W = (X/\sigma)^2$ (the distribution of W is called a *chi-squared* distribution);

    (iii) $V = \sqrt{|X|}$.

24. In Exercise 5 of Section 6.9, it was mentioned that the median of the distribution associated with a continuous random variable X is the real number m such that

$$P(X \leq m) = P(X \geq m).$$

Find the median of the $N(\mu, \sigma^2)$ distribution.

Group B

25. It is known that 13.57% of the persons who take a certain psychology test achieve a score of less than 50 (out of 100), while 8.08% achieve a score of 80 or higher. Assuming that X, a person's score in that test, follows the normal distribution, find the mean and the variance of X.

26. Andy and Phil work in a garage. When a car goes to the garage for a regular service, one of them undertakes the job. The amount of time, X, to complete the service has a normal distribution $N(75, 10^2)$ if it is done by Andy, while it has a $N(90, 15^2)$ distribution if it is done by Phil (times are in minutes). This morning a car came for service and the time it took to finish the job was less than an hour. What is the probability that the car was serviced by Andy?

27. The insurance company *Riska* has insured a large number of homeowners against potential destruction of their homes by tornadoes. The company director feels that it is too risky for his company to cover losses from this portfolio, so he makes an agreement with another company, called *FRisk*. The agreement is that the total losses from the portfolio are divided as follows: if *Riska* pays, in a given month, an amount X to the policyholders, *FRisk* pays an amount $Y = 3X - 10$.

    Suppose $X \sim N(20, 16)$ (amounts are in millions of dollars). Then:

    (i) Calculate the probability that in a given month, *FRisk* pays to the policyholders at least twice as many dollars as *Riska* does.

    (ii) What is the probability that in a period of six months, there is only one month in which *Riska* pays more than \$35 million?

    (iii) Find the probability that the event "*Riska* pays more than \$35 million in a month" occurs for the third time within a six-month period.

    (iv) Write down the density function of the variable Y.

    (v) Let $\mathcal{E}$ be the event that "the monthly amount which FRisk pays to the policyholders is between 40 and 50 million dollars." What is the probability that the event $\mathcal{E}$ occurs for the first time during the 5th month of the agreement between the two companies?

28. A female athlete in long jump, looking at data from her training, has estimated that 90% of her jumps are over 6.30 m, while 20% of her jumps are over 6.80 m. If X represents her performance in a jump, we assume that X has a normal distribution.

    (i) Find the mean and standard deviation of X.

    (ii) In the next world championships where she is participating, the athlete will have six attempts and she thinks that she needs a 6.85 m jump to win a medal. What is the probability that her best attempt is at least 6.85 m?

    (iii) If on a regular training she makes 150 jumps, find the largest value of X she is expected to achieve at least once in these attempts.

    (To solve this question, assume that the athlete never commits a foul jump!)

29. A factory produces metal cylinders. The requirements for each cylinder are that it must have a length, L, between 8.45 and 8.65 cm and a diameter, D, between 1.55 and 1.60 cm. The quality control section of the factory knows that the distribution of the random variable L is $N(8.54, (0.05)^2)$, while the distribution of D is $N(1.57, (0.01)^2)$. It is also assumed that the length of a cylinder that is produced is independent of its diameter (this means that any event associated with L is independent of any event associated with D). Calculate

    (i) the percentage of cylinders produced that do not meet the length requirements;

    (ii) the percentage of cylinders produced that do not meet the requirements for the diameter;

    (iii) the percentage of cylinders produced that cannot be sold (that is, they do not meet *either* the length *or* the diameter requirements);

    (iv) the probability that in a sample of five cylinders, exactly one of them cannot be sold.

30. A University lecturer marks scripts on a scale from 1 to 100. For a particular course, the lecturer has estimated from previous years that the marks of male students have a $N(67, 9^2)$ distribution, while the marks of female students have a $N(72, 8^2)$ distribution. Find the expected value and the standard deviation of the marks for this course, if it is known that 60% of the students are males.

31. The voltage supplied to a factory is a random variable X that follows the normal distribution with mean 110 V and a standard deviation 5 V. For the proper operation of the factory, there is an automatic regulator, which increases the voltage to 105 V when $X < 105$ and reduces the voltage to 120V when $X > 120$. Let Y denote the random variable that denotes the actual voltage at which the factory works.

    (i) Obtain the distribution function, $F_Y(y)$, of Y.

    (ii) Calculate the probabilities $P(Y = 105)$, $P(Y = 120)$, $P(Y \leq 120)$, and $P(Y > 117)$.

32. In order to estimate the proportion p of smokers in a population, we need to take a sample of n persons. What should be the value of n such that the proportion of smokers in the sample differs from p at most by 0.02 with a probability 99%? If it

is known a priori that the percentage of nonsmokers is at least 80%, what should be the value of n in this case? (Use a normal approximation.)

33. In a city with 50 000 inhabitants, every person has to be vaccinated for an infectious disease. If each person selects completely at random any of the $k \geq 2$ vaccination centers, use a normal approximation to find how many vaccines (at least) should each center possess so that, with probability 95%, it can meet the demand. Obtain a numerical answer to your result for $k = 5$.

7.3 THE EXPONENTIAL DISTRIBUTION

Although the uniform distribution, discussed in Section 7.1, seems to be the simplest continuous distribution, it is not a distribution that is widely used to model real-life phenomena as mentioned there. It is rather rare in practice to have an unknown quantity that takes values on a finite interval and to assume that its density is constant over that interval. In fact, as pointed out in the preceding section, the normal distribution is by far the most commonly used probability distribution in many applied areas. There are two special features of the normal distribution, however, that we have to take into consideration before using it to model the uncertainty underlying a quantity of interest: (i) any normal distribution is symmetric and (ii) the range of values for a quantity that follows the normal distribution is the entire real line. Although we have explained in the last section that the second of these attributes is not a forbidding condition for the use of the normal distribution, the first of these features is clearly inappropriate when we know that the distribution of a variable (as dictated by the nature of the problem or from past experience) is heavily asymmetric.

Among all probability distributions, those with a range on the set of positive (or nonnegative) real numbers play a key role in many applications. For example, it is often the case that the quantity we consider represents time, some simple examples of which are as follows:

- the lifetime of an individual, a machine, or a manufacturing item;
- the time we have to wait until a certain event occurs, e.g. the next bus to arrive at the bus stop, the waiting time of a customer at a bank, or the time until the next thunderstorm in a certain area;
- the duration of a telephone conversation;
- the time it takes to repair a faulty item;
- processing time of a computer, which executes a line of commands or a prescribed schedule;
- the time a patient needs to recover from a surgery, or for a medical treatment to take effect.

In all these cases, either common knowledge or experience based on data available indicates that the normal distribution will fail to provide an adequate model for the quantity of interest.

For variables assuming only nonnegative values, the distribution which is most frequently used is the exponential distribution, although several other distributions will be discussed in the following section as well as in the exercises of this chapter. We have already seen the exponential distribution in several places in the previous chapter (see, e.g. Example 6.3 and Exercise 14 in Section 6.2). The prominence of this distribution in probability theory is largely due to its close connection with a special type of probability models, known as stochastic processes. Although here we shall not explore this concept any further, we mention that we have already met one of the most important stochastic processes in Section 5.5, namely, the Poisson process.

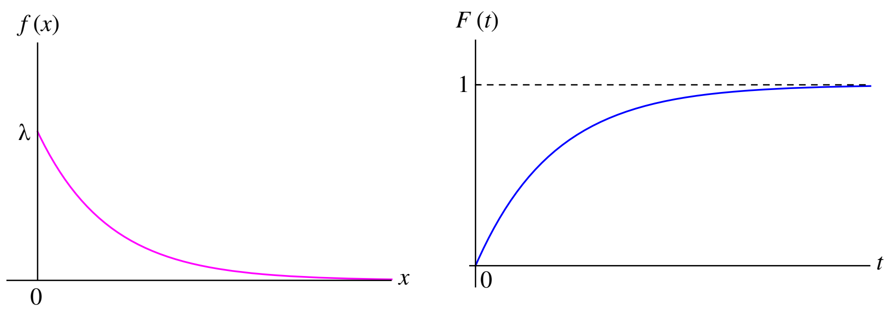
Definition 7.4 Let X be a continuous random variable with density function

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$$

for some $\lambda > 0$. Then we say that X has the exponential distribution with parameter λ, and denote this by $X \sim \mathcal{E}(\lambda)$.

It can be readily checked that the density of the exponential distribution, given above, satisfies the conditions required for a function to be a valid probability density. The distribution function of the exponential distribution is given by

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x = \begin{cases} 0, & t < 0, \\ \int_0^t f(x)\mathrm{d}x, & t \geq 0, \end{cases}$$



which gives immediately, by a simple integration,

$$F(t) = \begin{cases} 0, & t < 0, \\ 1 - e^{-\lambda t}, & t \geq 0. \end{cases} \tag{7.16}$$

Figure 7.11 provides plots of the density function and the distribution function given above for various values of λ.

The exponential distribution is most commonly used to model times until occurrence of an event or the time between successive occurrences of events, etc.

Proposition 7.8 *The expectation and variance of a random variable X following the exponential distribution with parameter λ are*

$$E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}.$$

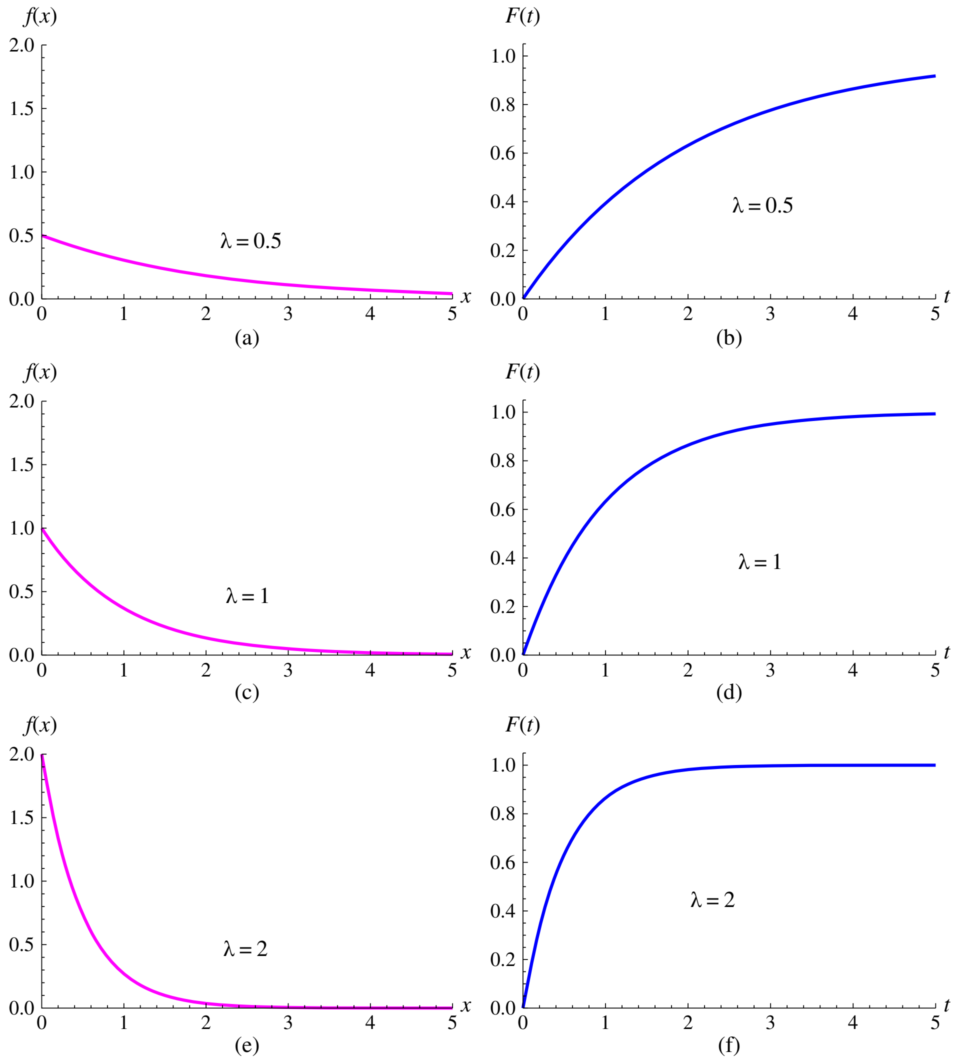


Figure 7.11 The density function (left) and the distribution function (right) of the $\mathcal{E}(\lambda)$ distribution for various choices of λ.

Proof: First for the expectation, we have

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = \int_{0}^{\infty} x\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x.$$

Integrating by parts, and making use of the fact the $\lambda \mathrm{e}^{-\lambda x}$ is negative of the derivative of the function $\mathrm{e}^{-\lambda x}$, we obtain

$$E(X) = \left[-x\mathrm{e}^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} \mathrm{e}^{-\lambda x}\,\mathrm{d}x.$$

The integral on the right equals λ^{-1} and the proof for the expectation gets completed upon noting that

$$\left[-x\mathrm{e}^{-\lambda x}\right]_0^{\infty} = -\lim_{x\to\infty} \frac{x}{\mathrm{e}^{\lambda x}} + 0 \cdot \mathrm{e}^0 = 0,$$

which follows by an application of l' Hôpital's rule. For the variance, as usual we find $E(X^2)$ first. This is given by

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_0^{\infty} x^2 \lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x \\ &= \left[-x^2 \mathrm{e}^{-\lambda x}\right]_0^{\infty} + \int_0^{\infty} 2x\mathrm{e}^{-\lambda x}\,\mathrm{d}x, \end{aligned} \tag{7.17}$$

where we have used integration by parts again. The values of the function $-x^2\mathrm{e}^{-\lambda x}$ are zero both at $x = 0$ and as $x \to \infty$ (the latter is seen by l' Hôpital's rule). Moreover, we observe that the integral on the right-hand side of (7.17) can be expressed as

$$\int_0^{\infty} 2x\mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda}\int_0^{\infty} x\lambda \mathrm{e}^{-\lambda x}\,\mathrm{d}x = \frac{2}{\lambda}E(X).$$

Since we have already determined $E(X) = 1/\lambda$, we immediately have

$$E(X^2) = \frac{2}{\lambda^2},$$

from which it follows that

$$\mathrm{Var}(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2},$$

as required. □

Since the variance of a random variable X with the $\mathcal{E}(\lambda)$ distribution is $1/\lambda^2$, the standard deviation of X is $1/\lambda$. We thus see that for an exponentially distributed variable, the mean is equal to its standard deviation. The same is also true for the geometric distribution, one of the discrete distributions discussed in Chapter 5. In fact, as will be seen later in this section, the exponential distribution can be regarded as the continuous analogue of the geometric distribution. In contrast, for the Poisson distribution, we have seen in Chapter 5 that the mean is equal to the *variance*, rather than the standard deviation.

Example 7.10 A convenience store has only one cashier. The time in minutes he requires to serve a customer, denoted by X, is a random variable that has the exponential distribution with parameter $\lambda = 2$. Sarah is first in the queue at the cashier and the person who was in front of her has just started being served. Find the probability that Sarah will have to wait until her service starts

(i) at least three minutes;

(ii) between two and four minutes.

SOLUTION We observe first from (7.16) that, for $t \geq 0$, the probability $P(X \leq t)$ is $F(t) = 1 - \mathrm{e}^{-2t}$.

(i) The event that Sarah will have to wait at least three minutes before she is served is $\{X \geq 3\}$. Therefore, the required probability is

$$P(X \geq 3) = P(X > 3) = 1 - F(3) = 1 - (1 - e^{-3\cdot 2}) = e^{-6},$$

noting that the first equality above follows since X is a continuous random variable, and so the events $\{X \geq 3\}$ and $\{X > 3\}$ have the same probability.

(ii) Next, we seek the probability $P(2 \leq X \leq 4)$. This is the same as $P(2 < X \leq 4)$ and by a standard property of distribution functions, it is equal to

$$F(4) - F(2) = (1 - e^{-4\cdot 2}) - (1 - e^{-2\cdot 2}) = e^{-4} - e^{-8}.$$

Example 7.11 The lifetime X of a device is assumed to follow an exponential distribution with parameter $\lambda > 0$.

(i) Calculate the probability that the lifelength of the device

(a) exceeds its mean value;

(b) is at least twice as much as its mean value.

(ii) Find the length of time t_0, such that the probability the device works at time t_0 is 95%.

SOLUTION Let F denote the distribution function of the lifetime X. Then by (7.16), $F(t) = 1 - e^{-\lambda t}$ for any $t \geq 0$, while $F(t) = 0$ for $t < 0$. Further, the expected lifetime of the device is λ^{-1}.

(i) (a) The required probability is $P(X > \lambda^{-1})$, which is given by

$$P(X > \lambda^{-1}) = 1 - F(\lambda^{-1}) = e^{-\lambda\cdot\lambda^{-1}} = e^{-1} \cong 0.3679 \cong 37\%.$$

(b) We now seek $P(X > 2\lambda^{-1})$, which by an appeal to (7.16) again, equals

$$P(X > 2\lambda^{-1}) = 1 - F(2\lambda^{-1}) = e^{-\lambda\cdot 2\lambda^{-1}} = e^{-2} \cong 0.1353,$$

or about 13.5%. Observe that both probabilities for this part are independent of the parameter λ.

(ii) Compared to the previous part, we see here that the problem is in the opposite direction, since we seek the value t_0 such that

$$P(X \geq t_0) = 0.95.$$

Using once more the particular form of the distribution function for X, this becomes $1 - e^{-\lambda t_0} = 0.05$, which in turn yields

$$e^{-\lambda t_0} = 0.95.$$

Taking logarithms on both sides, we get $-\lambda t_0 = \ln(0.95)$, so that we obtain finally

$$t_0 = -\frac{1}{\lambda}\ln(0.95).$$

In the terminology used in Example 7.4, this is the 95%-quantile of the $\mathcal{E}(\lambda)$ distribution.

An important property of the exponential distribution is the **lack of memory property** (or the **memoryless property**) of the distribution. We say that a nonnegative random variable X has this property if for any $t, s \geq 0$, we have

$$P(X > t + s | X > s) = P(X > t). \tag{7.18}$$

As mentioned earlier, a random variable having the exponential distribution often represents time. Assume for instance that X denotes the lifetime of a certain device. Then, the property in (7.18) states that the probability $P(X > t)$ that a *new device* will last for at least t units of time is the same as the probability that a device which has already lasted for s units will operate for an additional t units of time, the latter being $P(X > t + s | X > s)$.

To check that (7.18) holds, first note that for any $t \geq 0$,

$$P(X > t) = 1 - P(X \leq t) = 1 - F(t) = 1 - (1 - e^{-\lambda t}) = e^{-\lambda t},$$

while

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

In view of this property, we see that an exponential distribution cannot, in many cases, be a realistic model for the lifelength of a device or a system and also for the lifetime of living organisms. This is because in both these cases, we expect that the further a device or an organism lives, the more prone it will become to failure or death. Note, however, that the exponential distribution can be used to model *part* of the duration of the life for the subject under consideration. For example, in engineering, we speak about the **useful period** of a device; this spans the period after the **early period** or **failure period** (during which any defects or other production problems in the product are identified and removed) and before the **wear-out period** (during which the first signs of fatigue are seen in the device), as displayed in Figure 7.12.

It is worth noting that the lack of memory property characterizes the exponential among all continuous distributions (that is, no other continuous distribution has that property; see

Figure 7.12 Early period, useful period and wear-out period of a device.

Exercise 18 at the end of this section). Among discrete distributions, however, this is also a property of the geometric distribution, as seen earlier in Section 5.2.

Example 7.12 The time between two successive emails arriving at Julia's email account follows the exponential distribution with a mean value of 90 minutes.

(i) What is the probability that the time between two successive emails exceeds one hour?

(ii) Given that 90 minutes have passed since Julia received her last email, what is the probability that she has to wait for at least 30 minutes until the next one?

SOLUTION Let us choose one hour as the time unit.[3] Then, if X denotes the time between successive emails arriving, we are given that $E(X) = 3/2$. Since from Proposition 7.8 the mean value of an exponential random variable is the reciprocal of the parameter of its distribution, we see that $X \sim \mathcal{E}(\lambda)$ with $\lambda = 2/3$.

(i) We need the probability $P(X > 1)$, which is given by

$$P(X > 1) = 1 - F(1) = \mathrm{e}^{-2/3}.$$

(ii) If no email arrives within the next 30minutes, this means that the time between the last email and the next one that arrives will be at least $90 + 30 = 120$ minutes, i.e. two hours. Moreover, due to the lack of memory property, we have

$$P\left(X > 2|X > \frac{3}{2}\right) = P\left(X > \frac{3}{2} + \frac{1}{2}|X > \frac{3}{2}\right) = P\left(X > \frac{1}{2}\right).$$

Hence,

$$P(X > 2|X > 3/2) = 1 - F(1/2) = \mathrm{e}^{-1/3},$$

which is the desired probability.

In Example 6.3 of the previous chapter, we have obtained the distribution for a linear transformation of an exponential random variable. In the following example, we use the result of Proposition 6.2 to obtain the distribution of a nonlinear transformation for such a variable.

[3]A different choice would obviously have no effect on the result. Bear in mind, however, that we have to be consistent while dealing with a particular problem, in that we should use the same unit throughout.

Example 7.13 (Weibull distribution) Let X be a continuous random variable with density

$$f(x) = \begin{cases} e^{-x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

Obtain the density of the random variable

$$Y = \alpha X^{\beta},$$

where $\alpha > 0$ and $\beta > 0$ are arbitrary constants.

SOLUTION We apply the formula

$$f_Y(y) = |(g^{-1})'(y)| f(g^{-1}(y)) \tag{7.19}$$

from Proposition 6.2, where the function g is defined by $g(x) = \alpha x^{\beta}$.

The derivative of this function is

$$g'(x) = \alpha\beta x^{\beta-1},$$

which is positive for $x > 0$, and so g is a strictly increasing function. In order to find the inverse function, g^{-1}, we solve the equation $g(x) = y$. This gives

$$\alpha x^{\beta} = y \iff x^{\beta} = \frac{y}{\alpha} \iff x = \left(\frac{y}{\alpha}\right)^{1/\beta}.$$

Therefore,

$$(g^{-1})(y) = \left(\frac{y}{\alpha}\right)^{1/\beta}.$$

The derivative of this function, with respect to y, is

$$(g^{-1})'(y) = \frac{1}{\beta} \cdot \frac{y^{1/\beta-1}}{\alpha^{1/\beta}}.$$

We thus see from (7.19) that the density of Y is

$$\begin{aligned} f_Y(y) &= \left|\frac{1}{\beta} \cdot \frac{y^{1/\beta-1}}{\alpha^{1/\beta}}\right| \exp\left\{-\left(\frac{y}{\alpha}\right)^{1/\beta}\right\} \\ &= \frac{1}{\beta\alpha^{1/\beta}} \cdot y^{1/\beta-1} \cdot \exp\left\{-\left(\frac{y}{\alpha}\right)^{1/\beta}\right\}, \quad y > 0. \end{aligned}$$

This distribution is known as the Weibull distribution, after the famous Swedish mathematician Waloddi Weibull.

EXERCISES

Group A

1. Assume that X is an exponential random variable, i.e. $X \sim \mathcal{E}(\lambda)$. Find the value of λ in the following cases, using in each case the information given below:

   (i) $E(X) = 5/3$

   (ii) $\text{Var}(X) = E(X)$

   (iii) $P(X \leq 1) = P(X > 1)$

   (iv) $P(X \leq 3) = 3P(X > 3)$.

2. The lifelength, X, for a certain type of bacteria has the exponential distribution with a mean value of three weeks. What is the probability that a bacterium lives for

   (i) at least five weeks?

   (ii) at least five weeks, if we know that it has lived for three weeks so far?

3. The time until a fault appears in the engine of a certain type of airplane has the exponential distribution with expected value of 1200 hours. The airplane has four such engines and suppose that it can operate safely during a flight if at least two of the engines function properly. Find the probability that the airplane operates without a problem during a

   (i) 3-hour flight;

   (ii) 10-hour flight.

4. The number of accidents that occur at a dangerous spot in a traffic junction of a main road has a mean value of three accidents per month. If the time between two successive accidents has an exponential distribution, find the probability that there will be at least two months from the last accident until the next one occurs.

5. Suppose that the duration of phone calls in Paula's mobile phone is an exponentially distributed random variable with a mean time of one minute. Calculate the probability that the duration of the next call on her mobile will be

   (i) less than 15 seconds;

   (ii) between 30 and 90 seconds;

   (iii) at most three minutes given that it will be at least two minutes.

6. Suppose the duration X of a phone call has the exponential distribution and the expected time length of a phone call is 15 minutes.

   (i) Marc arrives at a phone booth in a street immediately after Sandra, who has just entered the booth and starts making a call. If no one arrived between Sandra and Marc, so that Marc is the first person in the queue, what is the probability that he has to wait

(a) at least 15 minutes,

(b) between 20 and 25 minutes,

until Sandra finishes her call?

(ii) Assuming now that, when Marc arrives, Sandra has already been talking for 15 minutes. Find the probability that Marc has to wait

(a) at least 15 minutes,

(b) between 20 and 25 minutes,

until Sandra finishes her call.

7. The duration, X (in seconds), that a PC takes to run a program has the exponential distribution with mean $1/\lambda = 20$seconds. If we give a command for 10 such programs to be executed by the computer, find the probability that at least two programs will still be running 20 seconds after the start of their execution.

8. A type of a light bulb used in radiophones has a life span that is exponentially distributed with expected value 25 (in thousands of hours). The factory that produces the bulbs wishes to offer a guarantee to its customers. If the product fails before the guarantee period, it is returned to the factory and replaced immediately. What should be set as the guarantee period, in thousands of hours, if at most 1% of the bulbs are to be returned to the factory?

9. If a random variable X has a distribution function that is strictly increasing, the point x_a which is (uniquely) defined by the relationship

$$F(x_a) = 1 - a, \quad 0 < a < 1,$$

is the upper-a quantile of the random variable X (see also Examples 7.4 and 7.11). Assuming that $X \sim \mathcal{E}(\lambda)$, show that the upper-$a$ quantile of X is

$$x_a = -\frac{1}{\lambda} \ln a.$$

Hence find numerically the points $x_{0.25}$, $x_{0.5}$, and $x_{0.75}$ for the case when $\lambda = 1$.

10. Let X be a random variable with an exponential distribution with $\lambda = 1$. Find the distribution function for each of the following variables:

(i) $Y = \sqrt{X}$;

(ii) $W = \alpha X^{1/\beta} + \gamma$, where $\alpha, \beta > 0$ and $\gamma \geq 0$ are constants.

11. If X has the exponential distribution with parameter $\lambda = 1$, find the distribution function of the random variable $Y = -\ln X$. (The distribution of Y is a special case of the **generalized extreme value distribution**.)

12. Let X be a random variable following the exponential distribution with parameter λ. Find the distribution function, and hence the density function for the following random variables:

    (i) $Y = 5X$;

    (ii) $W = (2X + 3)/4$;

    (iii) $V = |X - 2|$.

Group B

13. The length of a phone call, in minutes, is assumed to be exponentially distributed with parameter $\lambda > 0$. If the cost, Y, in dollars, of a phone call is given by $Y = \alpha[X] + \beta$ for given constants α and β, calculate the expected value and the variance of Y. Then, give a numerical answer to your result for $\lambda = 1/10, \alpha = 0.04$, and $\beta = 0.02$.

    (*Hint*: Calculate first the mean and variance of the random variable $[X]$.)

14. The size X, in thousands of dollars, of a claim on an insurance portfolio is a continuous random variable. The insurance company that operates the portfolio has estimated from past data that the probability the size of a claim exceeds a value $x \geq 0$ is given by

    $$P(X > x) = \frac{1}{3}\mathrm{e}^{-2x} + \frac{2}{3}\mathrm{e}^{-4x}, x \geq 0.$$

    Calculate

    (i) the probability that when a claim arrives, its size will be between \$2000 and \$2500;

    (ii) the density function for the claim size X;

    (iii) the expectation and variance of X.

15. Peggy has sent an email to her friend Susan and she is waiting for her reply. The waiting time T, in minutes, until Peggy receives the reply message has density function $f_T(t)$. If by time t_0, she has not received a reply, find the distribution function of the **residual waiting time** $X = T - t_0$, that is

    $$F(t) = P(X \leq t | T > t_0).$$

    Apply the general result that you find for F to the special case when T has an exponential distribution with parameter λ.

16. A random variable X is said to follow the **translated Weibull distribution** if its density function has the form

    $$f(x) = \frac{b}{a}\left(\frac{x-c}{a}\right)^{b-1} \exp\left[-\left(\frac{x-c}{a}\right)^{b}\right], \quad x \geq c,$$

    where $a, b > 0$ and $c \geq 0$ are given constants. Verify that the random variable

    $$Y = \left(\frac{X-c}{a}\right)^{b}$$

has the exponential distribution with parameter $\lambda = 1$ and then find the distribution function of the translated Weibull distribution.

Application: The delivery time (in minutes), of a pizza delivery service, has the translated Weibull distribution with $a = 6/5$, $b = 3/2$, and $c = 10$. What is the probability that if someone orders pizza, it will arrive within 15 minutes from the time the order was made?

17. Suppose X has the exponential distribution with parameter $\lambda > 0$. Show that for the moments $\mu_r' = E(X^r)$, the following recursive relation holds:

$$\mu_r' = \frac{r}{\lambda} \cdot \mu_{r-1}', \quad r = 2, 3, \ldots .$$

Hence verify that, for $r = 1, 2, \ldots,$

$$\mu_r' = \frac{r!}{\lambda^r}.$$

18. Suppose X is a random variable having the lack-of-memory property in (7.18). We define the function

$$R(t) = P(X > t) = 1 - F(t), \quad t \geq 0,$$

where F is the distribution function of X.

(i) Check that, for any $t > 0$ and $h > 0$, we have

$$\frac{R(t+h) - R(t)}{h} = -R(t)\frac{1 - R(h)}{h}.$$

(ii) Letting $h \to 0$, show that the derivative with respect to t, of the function $\ln R(t)$ is equal to $-\lambda$, where

$$\lambda = \lim_{h \to 0} \frac{1 - R(h)}{h} > 0.$$

(iii) Prove that the distribution function of X is given by

$$F(t) = 1 - \mathrm{e}^{-\lambda t}, \quad t \geq 0,$$

so that X has the exponential distribution with parameter λ.

7.4 OTHER CONTINUOUS DISTRIBUTIONS

Apart from the three continuous distributions discussed so far, there is a large number of other distributions that are used in practice. Their relative importance among all possible continuous distributions that may be constructed lies primarily on their ability to model adequately some quantities of interest arising in real-life phenomena. Since probability distributions are used these days in a wide array of disciplines, different distributions have been considered in different areas. We have seen, for instance, the Pareto distribution in Example 6.6 of the previous chapter. This is a distribution that is prominently used in economics and insurance modeling, areas where the normal or the uniform distributions are not so useful.

In this section, we discuss in detail two other continuous distributions. Many other distributions that are used in practice are mentioned in the Exercises at the end of this and the previous section, as well as in Section 7.8.

7.4.1 The Gamma Distribution

In the early twentieth century, the Danish mathematician and engineer A. K. Erlang (1878–1929) studied various problems connected with *telephone networks* in his country. Experimenting with real data he showed, among other things, that the Poisson distribution provides a good fit to the number of phone calls arriving at a call center. Considering the duration of phone calls and the waiting times until an operator is free, he used the following probability distribution, which now bears his name.

Let X be a random variable with density function

$$f(x) = \begin{cases} \dfrac{\lambda^n}{(n-1)!} x^{n-1} \mathrm{e}^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Then we say that X has an **Erlang distribution** with parameters n and λ, where n is a positive integer and $\lambda > 0$. It is obvious that for $n = 1$, the Erlang distribution coincides with the exponential distribution with parameter λ.

The Gamma distribution is a generalization of the Erlang distribution when the parameter n is not necessarily an integer. A problem that arises in the above formula for this case is how we interpret $(n - 1)!$ for noninteger n. The problem is solved with the use of an important function in mathematics, called the **Gamma function**, $\Gamma : (0, \infty) \mapsto \mathbb{R}$, which is defined by

$$\Gamma(a) = \int_0^\infty t^{a-1} \mathrm{e}^{-t} \, \mathrm{d}t.$$

For $a = 1$, it is easy to see that

$$\Gamma(1) = \int_0^\infty t^0 \mathrm{e}^{-t} \, \mathrm{d}t = \int_0^\infty \mathrm{e}^{-t} \, \mathrm{d}t = [-\mathrm{e}^{-t}]_0^\infty = 1.$$

Moreover, for an arbitrary $a > 0$, integration by parts yields

$$\begin{aligned} \Gamma(a+1) &= \int_0^\infty t^a \mathrm{e}^{-t} \, \mathrm{d}t = \int_0^\infty t^a (-\mathrm{e}^{-t})' \, \mathrm{d}t = [-t^a \mathrm{e}^{-t}]_0^\infty - \int_0^\infty a t^{a-1} (-\mathrm{e}^{-t}) \, \mathrm{d}t \\ &= 0 + a \int_0^\infty t^{a-1} \mathrm{e}^{-t} \, \mathrm{d}t, \end{aligned}$$

which reduces to

$$\Gamma(a+1) = a \cdot \Gamma(a), \quad a > 0. \tag{7.20}$$

Employing this expression successively, we see that

$$\Gamma(2) = 1 \cdot \Gamma(1) = 1,$$
$$\Gamma(3) = 2 \cdot \Gamma(2) = 2 \cdot 1 = 2!,$$
$$\Gamma(4) = 3 \cdot \Gamma(3) = 3 \cdot (2 \cdot 1) = 3!$$

and so on, so that we have for any nonnegative integer n,

$$\Gamma(n+1) = n\Gamma(n) = n(n-1)\Gamma(n-1) = \cdots = n(n-1)(n-2)\cdots 2 \cdot 1 = n!.$$

This shows that the Gamma function provides indeed a generalization for the concept of a factorial.

Values of the Gamma function, $\Gamma(a)$, for noninteger a are in general difficult to obtain by hand. We can use Mathematica instead; writing *Gamma*[a], Mathematica returns the value of $\Gamma(a)$. In this way, we find, for example, that $\Gamma(1/2) = \sqrt{\pi}$ (for an analytic proof of this, see Exercise 4 at the end of this section). Then, formula (7.20) readily gives

$$\Gamma\left(\frac{3}{2}\right) = \frac{1}{2} \cdot \Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi},$$
$$\Gamma\left(\frac{5}{2}\right) = \frac{3}{2} \cdot \Gamma\left(\frac{3}{2}\right) = \frac{3}{4}\sqrt{\pi},$$

and so on.

We can now present the following definition for the Gamma distribution.

Definition 7.5 A random variable X has the Gamma distribution with parameters $a > 0$ and $\lambda > 0$ if its density function is given by

$$f(x) = \begin{cases} \dfrac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

Note that we use different letters for the parameters of the Erlang and the Gamma distribution to emphasize the fact that, in the former, n is an integer while the parameter a of the Gamma distribution can be any positive real number.

To check that the function f in the above definition is indeed a valid density function for any $a, \lambda > 0$, we first note that $f(x) \geq 0$ for any real x and then we determine the integral of this function as

$$\int_{-\infty}^{\infty} f(x)\mathrm{d}x = \int_0^{\infty} \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}\, \mathrm{d}x = \frac{1}{\Gamma(a)} \int_0^{\infty} (\lambda x)^{a-1} e^{-\lambda x} d(\lambda x)$$
$$= \frac{1}{\Gamma(a)} \int_0^{\infty} y^{a-1} e^{-y}\, \mathrm{d}y = \frac{\Gamma(a)}{\Gamma(a)} = 1.$$

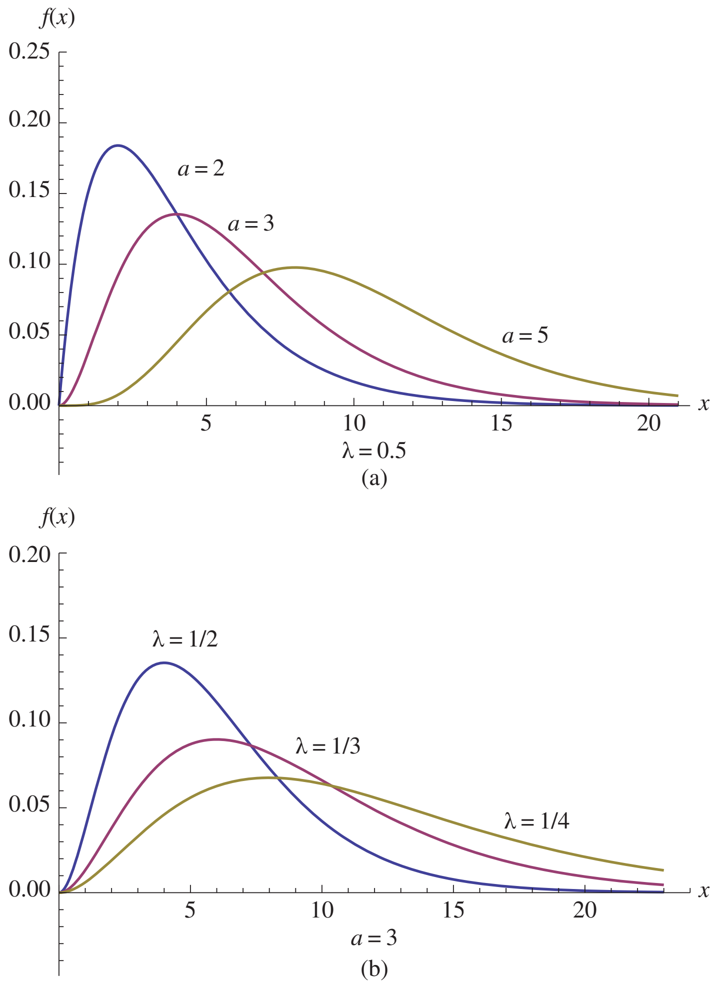


Figure 7.13 The density function of the Gamma distribution for different values of a (top) and λ (bottom).

Figures 7.13 and 7.14 present the plot of the density function and the latter also presents the cumulative distribution of the Gamma distribution for various values of the parameters a and λ.

Example 7.14 A radio show offers its listeners two pairs of tickets for a concert. The tickets will be given to the first two persons who call the radio station. Let X be the time (in minutes) until both pairs of tickets have been given, and suppose the density function of X is

$$f(x) = 4xe^{-2x}, \quad x \geq 0.$$

This is the density of the Gamma distribution with parameters $a = 2$ and $\lambda = 2$. (In fact, since a is an integer we could also say this is an Erlang density with $n = 2$ and $\lambda = 2$.)

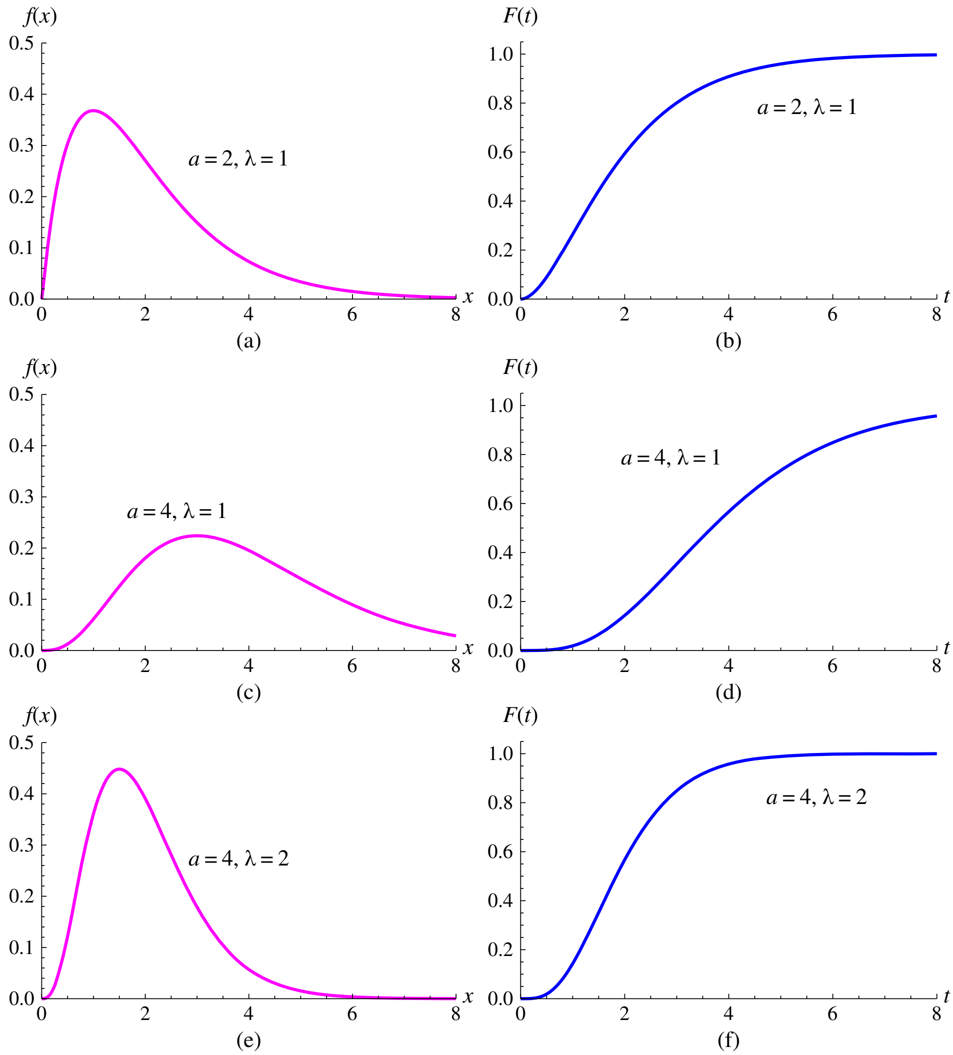


Figure 7.14 The density function (left) and cumulative distribution function (right) of the Gamma distribution for various values of the parameters a and λ.

The distribution function of X is given by

$$
\begin{aligned}
F(t) = P(X \le t) &= \int_0^t 4x\mathrm{e}^{-2x}\,\mathrm{d}x = \int_0^t 2x(-\mathrm{e}^{-2x})'\,\mathrm{d}x \\
&= [-2x\mathrm{e}^{-2x}]_0^t + \int_0^t 2\mathrm{e}^{-2x}\,\mathrm{d}x \\
&= -2t\mathrm{e}^{-2t} + [-\mathrm{e}^{-2x}]_0^t = -2t\mathrm{e}^{-2t} + (1-\mathrm{e}^{-2t}) = 1-(2t+1)\mathrm{e}^{-2t}.
\end{aligned}
$$

Using this we have, for example, the following:

(i) The probability that at least one of the tickets is available after one minute is

$$P(X > 1) = 1 - F(1) = 1 - [1 - (2 \cdot 1 + 1)e^{-2\cdot 1}] = 3e^{-2} \cong 0.406.$$

(ii) The probability that the waiting time until both pairs of tickets have been given lies between one and three minutes equals

$$\begin{aligned} P(1 \le X \le 3) &= F(3) - F(1) = [1 - (2 \cdot 3 + 1)e^{-2\cdot 3}] - [1 - (2 \cdot 1 + 1)e^{-2\cdot 1}] \\ &= 3e^{-2} - 7e^{-6} \cong 0.389, \end{aligned}$$

that is, about 39%.

(iii) Suppose that there is at least one pair of tickets available after one minute. The probability that no tickets will be available within the next two minutes is

$$P(X \le 3|X > 1) = \frac{P(1 < X \le 3)}{P(X > 1)} = \frac{F(3) - F(1)}{1 - F(1)} \cong \frac{0.389}{0.406} \cong 0.957,$$

using the results found earlier.

Note that, in general, there is no closed-form expression for the distribution function of the Gamma distribution. For the Erlang distribution, the integral of the density function is tractable and can be written as a sum of finitely many terms.

Proposition 7.9 *Assume that a random variable X follows the Gamma distribution with parameters a and λ. Then the expected value and the variance of X are given by*

$$E(X) = \frac{a}{\lambda}, \quad \mathrm{Var}(X) = \frac{a}{\lambda^2}.$$

Proof: For the expectation, we have

$$E(X) = \int_0^\infty x \cdot \frac{(\lambda x)^{a-1}}{\Gamma(a)} \lambda e^{-\lambda x}\, dx.$$

Putting $\lambda x = y$, we obtain

$$E(X) = \frac{1}{\lambda\Gamma(a)} \int_0^\infty y \cdot y^{a-1} \cdot e^{-y}\, dy = \frac{\Gamma(a+1)}{\lambda\Gamma(a)} = \frac{a}{\lambda},$$

using (7.20). Next, we find $E(X^2)$. Making again the change of variable $\lambda x = y$ in the integral, we get

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2 \cdot \frac{(\lambda x)^{a-1}}{\Gamma(a)} \lambda e^{-\lambda x}\, dx = \frac{1}{\lambda^2\Gamma(a)} \int_0^\infty y^{a+1} e^{-y}\, dy \\ &= \frac{\Gamma(a+2)}{\lambda^2\Gamma(a)} = \frac{a(a+1)}{\lambda^2}, \end{aligned}$$

by an appeal to (7.20) again. This yields

$$\operatorname{Var}(X) = E(X^2) - [E(X)]^2 = \frac{a(a+1)}{\lambda^2} - \left(\frac{a}{\lambda}\right)^2 = \frac{a}{\lambda^2}.$$

A more general result, concerning moments of arbitrary positive order for the Gamma distribution, is given in Exercise 3 at the end of this section. □

7.4.2 The Beta Distribution

This is another continuous distribution, which offers a satisfactory model for random variables that take values between two known endpoints. As the range of values for the Beta distribution is the interval $(0, 1)$, it is often used when X denotes a percentage (scaled to take values in that interval). For example, the percentage of persons in a population with a certain characteristic, student performance (out of 100) in a test, etc. More generally, when a continuous random variable X can only take values in a finite interval (a, b) (or $[a, b]$), a simple linear transformation which maps that interval onto $(0, 1)$ results in a variable with values in the unit interval, and the Beta distribution defined below often offers a satisfactory model in such cases.

Definition 7.6 A random variable X has the Beta distribution with parameters $\alpha > 0$ and $\beta > 0$ if its density function is given by

$$f(x) = \begin{cases} \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where $B(\alpha, \beta)$ denotes the *Beta* function defined by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx.$$

The Beta function defined above is related to the Gamma function via the formula (the proof of this is omitted)

$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}. \tag{7.21}$$

It is immediate from above that the function f in Definition 7.6 is a valid probability density function. We now find the mean and the variance of a variable X having the Beta distribution.

Proposition 7.10 *Let X be a random variable having the Beta distribution with parameters α and β. Then the mean and variance of X are given by*

$$E(X) = \frac{\alpha}{\alpha+\beta}, \qquad \operatorname{Var}(X) = \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}.$$

Proof: We first observe that, for $r = 1, 2, \dots$, the rth moment of X is given by

$$E(X^r) = \int_0^1 x^r f(x)\mathrm{d}x = \int_0^1 \frac{1}{B(\alpha, \beta)} x^{\alpha+r-1}(1-x)^{\beta-1}\,\mathrm{d}x$$
$$= \frac{1}{B(\alpha, \beta)} \int_0^1 x^{\alpha+r-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{B(\alpha+r, \beta)}{B(\alpha, \beta)}$$

which, in view of Eq. (7.21), gives

$$E(X^r) = \frac{\dfrac{\Gamma(\alpha+r)\Gamma(\beta)}{\Gamma(\alpha+\beta+r)}}{\dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}} = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+r)}{\Gamma(\alpha+\beta+r)\Gamma(\alpha)}, \qquad r = 1, 2, \dots .$$

For $r = 1$, this gives, by an appeal to (7.20), that

$$\mu = E(X) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)} = \frac{\Gamma(\alpha+\beta)\alpha\Gamma(\alpha)}{(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)} = \frac{\alpha}{\alpha+\beta},$$

while for $r = 2$, upon employing (7.20) once again, we obtain

$$E(X^2) = \frac{\Gamma(\alpha+\beta)\Gamma(\alpha+2)}{\Gamma(\alpha+\beta+2)\Gamma(\alpha)} = \frac{\Gamma(\alpha+\beta)(\alpha+1)\alpha\Gamma(\alpha)}{(\alpha+\beta+1)(\alpha+\beta)\Gamma(\alpha+\beta)\Gamma(\alpha)}$$
$$= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}.$$

The result for the variance is now an easy consequence of the above and the formula $\mathrm{Var}(X) = E(X^2) - [E(X)]^2$. □

In Mathematica, the command *Beta*[a, b] returns the value of the Beta function evaluated at the pair of points (a, b), with $a, b > 0$, but not necessarily integers. In this way, we determine, for example,

$$B(2,3) = \frac{1}{12}, \quad B\left(4, \frac{1}{2}\right) = \frac{32}{35}, \quad B\left(\frac{2}{3}, 3\right) = \frac{27}{40},$$

and so on.

Figure 7.15(a) presents the plot of the density of the Beta distribution when $\alpha = \beta$. As can be seen easily, for $\alpha = \beta$ the Beta distribution is symmetric around the point $1/2$. Note that for $\alpha = \beta = 1$, the Beta distribution reduces to the Uniform distribution discussed in Section 7.1.

Figure 7.15(b) shows the graph of three Beta densities for which the mean $\alpha/(\alpha+\beta)$ is kept fixed. In all cases, we have $\alpha < \beta$ and we see that the distribution is *skewed to the right*; that is, the density function has a peak at a point to the left of the midpoint $x = 1/2$ and large values are less likely. When $\alpha > \beta$, the situation is reversed and the Beta distribution is *skewed to the left*. Figure 7.15(c) presents (in the same system of coordinates) the density function of the Beta distribution for several choices of the parameters α and β.

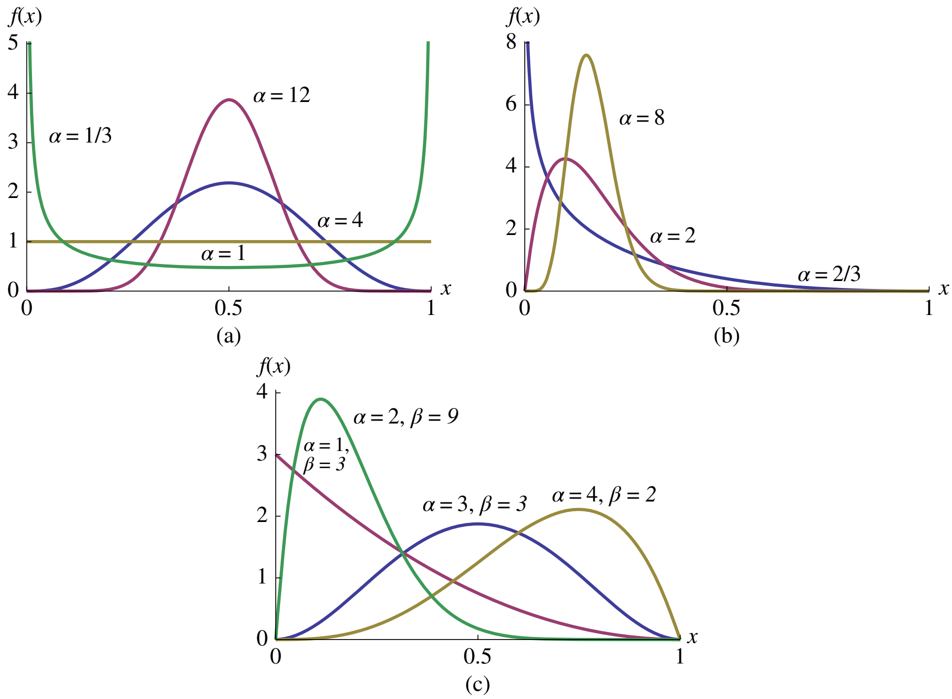


Figure 7.15 The density function of the Beta distribution for different choices of the parameters α and β.

Figure 7.16 gives the plot of the density function and the distribution function of the Beta distribution for different choices of the parameters α, β.

Example 7.15 The maximum time allowed for an exam is three hours. The time required, X, as a proportion of this maximum duration by a student to complete the exam (that is, if a student completes the exam in t hours, X takes the value $t/3$) has a Beta distribution with parameters $\alpha = 5$ and $\beta = 2$.

(i) Find the distribution function of X.

(ii) Obtain the mean and the standard deviation of X.

(iii) What proportion of students complete the exam within two hours?

SOLUTION The density function of X is given by

$$f(x) = \begin{cases} \dfrac{1}{B(5,2)} x^{5-1}(1-x)^{2-1}, & 0 < x < 1, \\ 0, & \text{elsewhere,} \end{cases}$$

where, from (7.21), we see that the value of $B(5,2)$ is

$$B(5,2) = \frac{\Gamma(5)\Gamma(2)}{\Gamma(5+2)} = \frac{4!1!}{6!} = \frac{1}{30}.$$

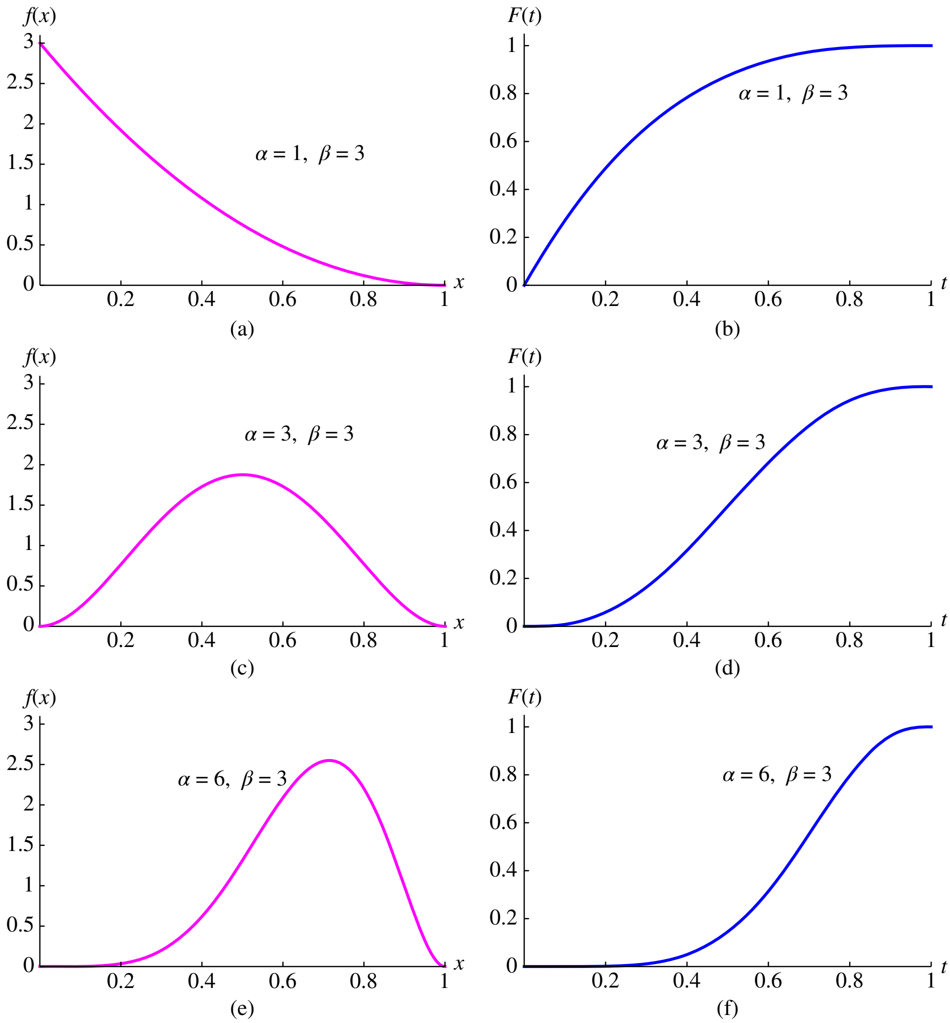


Figure 7.16 The density function (left) and the distribution function (right) of the Beta distribution for various choices of the parameters α and β.

Thus, the density takes the form

$$f(x) = \begin{cases} 30x^4(1-x), & 0 < x < 1, \\ 0, & \text{elsewhere.} \end{cases}$$

(i) For the distribution function, we have

$$F(t) = \int_{-\infty}^{t} f(x)\mathrm{d}x = \begin{cases} 0, & t < 0, \\ 6t^5 - 5t^6, & 0 \le t < 1, \\ 1, & t \ge 1. \end{cases}$$

(ii) Using the formulas from Proposition 7.10 for the mean and variance, we find

$$E(X) = \frac{\alpha}{\alpha + \beta} = \frac{5}{5+2} = \frac{5}{7}$$

and

$$\begin{aligned}\operatorname{Var}(X) &= \frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2} = \frac{5\cdot 2}{(5+2+1)(5+2)^2}\\ &= \frac{10}{8\cdot 49} = \frac{10}{392} \cong 0.0255,\end{aligned}$$

so that the standard deviation is

$$\sigma = \sqrt{\operatorname{Var}(X)} \cong \sqrt{0.0255} = 0.1597.$$

(iii) As explained in the statement, if a student completes the exam in t hours, the associated value of X is $t/3$. Thus, in order to find the probability that a student completes the exam within two hours (or, the proportion of students who complete the exam within two hours, which is the same), we need to find $P(X \le 2/3)$. This can be obtained easily upon substituting $t = 2/3$ in the distribution function F found in Part (i). Thus, we see that

$$P\left(X \le \frac{2}{3}\right) = F\left(\frac{2}{3}\right) = 6\left(\frac{2}{3}\right)^5 - 5\left(\frac{2}{3}\right)^6 = \frac{256}{729} \cong 0.3512,$$

and so about 35% of the students complete the exam within two hours.

EXERCISES

Group A

1. The running time, X (in seconds), for a computer program on a PC is a continuous random variable which follows the Erlang distribution with parameters $n = 2$ and $\lambda = 3$. Find the probabilities

   (i) $P(X < 1)$;

   (ii) $P(1 < X < 2)$;

   (iii) $P(X < 2 | X > 1)$.

2. The time X, in hours, to repair a car at a garage has an Erlang distribution with $n = \lambda = 2$.

   (i) Obtain the distribution function of X.

   (ii) Find the probability that, if two cars arrive at the garage for repair at the same time, they will both be repaired within an hour. (Assume that each car will be repaired by a different mechanic so that no car has to wait.)

3. Let X be a random variable having the Gamma distribution with parameters a and λ. Show that for any positive integer r, the r-th moment of X around zero is given by

$$E(X^r) = \frac{\Gamma(a+r)}{\lambda^r \Gamma(a)}$$

and hence obtain the coefficient of skewness (see the definition given in Exercise 12 of Section 6.3) for the Gamma distribution.

4. Let a be a positive real number.

   (i) Making a change of variable for the calculation of the integral, show that

$$\int_0^\infty t^{2a-1} \mathrm{e}^{-t^2/2}\, \mathrm{d}t = 2^{a-1}\Gamma(a).$$

   (ii) Applying the above relation for $a = 1/2$ and using the integral I in (7.8), the value of which was found in Section 7.2, prove that

$$\Gamma(1/2) = \sqrt{\pi}.$$

   (iii) Verify that for any positive integer $k = 1, 2, \ldots$, we have

$$\Gamma\left(k + \frac{1}{2}\right) = \frac{1 \cdot 3 \cdot 5 \cdots (2k-1)}{2^k}\sqrt{\pi}.$$

   (iv) If X is a random variable with the $N(\mu, \sigma^2)$ distribution, establish that the central moments of X are given by

$$\mu_r = \begin{cases} 0, & r = 1, 3, 5, \ldots \\ 1 \cdot 3 \cdot 5 \cdots (2k-1) \cdot \sigma^{2k}, & r = 2k \text{ with } k = 1, 2, \ldots . \end{cases}$$

5. The lifetime X (in thousands of hours) of an electronic device can be described by a random variable X with probability density function

$$f(x) = cx\sqrt{x}\mathrm{e}^{-4x}, \quad x \geq 0,$$

for some real constant c. Check that the distribution of X is Gamma, and then find the value of the constant c and the expected value of X.

6. (i) Let X be a standard normal random variable. Show that the variable $Y = X^2$ has a Gamma distribution and identify the parameters a and λ of that distribution.

   (ii) If $X \sim N(\mu, \sigma^2)$, find the distributions of the random variables

$$W = \left(\frac{X - \mu}{\sigma}\right)^2, \qquad \sigma^2 W = (X - \mu)^2.$$

   Application: Find the distribution of the kinetic energy $E = mV^2/2$ of a body with mass m which has speed V, assuming that m is known while $V \sim N(\gamma, \sigma^2)$.

7. For a random variable X having the Gamma distribution with parameters a and λ, what is the distribution of $Y = cX$, where c is a positive constant?

8. In Exercise 21 of Section 6.1, we gave the definition of a symmetric distribution around a point $a \in \mathbb{R}$. Prove that, when the two parameters α and β of the Beta distribution are equal, then the distribution is symmetric around the point $a = 1/2$.

9. Assuming that the variable X has a Beta distribution with parameters α and β, show that the distribution of $Y = 1 - X$ is also Beta and identify the parameters of that distribution.

10. Let X be a continuous random variable with density function

$$f(x) = \begin{cases} cx^{\alpha-1}(\gamma - x)^{\beta-1}, & 0 < x < \gamma, \\ 0, & \text{elsewhere,} \end{cases}$$

where α, β, γ, c are positive constants.

(i) Express the value of c in terms of α, β.

(ii) What is the distribution of the variable X/γ?

Group B

11. We have seen that the uniform distribution over the interval $[0, 1]$ is a special case of the Beta distribution. Here is another connection between the two distributions. Let $X \sim \mathcal{U}[0, 1]$. Then, for $a > 0$ the variable $Y = 1 - X^{1/a}$ has a Beta distribution. Prove this result and obtain the expectation of Y, by identifying first the parameters of the Beta distribution.

12. A continuous random variable X is said to have a **chi-squared distribution with n degrees of freedom** if it has a density of the form

$$f(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

(i) Check that this is a special case of a Gamma distribution.

(ii) Calculate the mean and the variance of the variable X.

(iii) Which random variable(s) among those stated in Exercise 6 above follow a chi-squared distribution?

13. Let α and β be two positive real numbers and X be a continuous random variable which has the Beta distribution with parameters α and β. In addition, let Y be

a discrete random variable following the binomial distribution with parameters $n = \alpha + \beta - 1$ and p $(0 < p < 1)$.

(i) Putting

$$I_p(\alpha, \beta) = \int_0^p x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x,$$

use integration by parts to prove the following recursive relationship:

$$I_p(\alpha, \beta) = \frac{1}{\alpha}[p^{\alpha}(1-p)^{\beta-1} + (\beta-1)I_p(\alpha+1, \beta-1)].$$

(ii) Show that

$$P(X \leq p) = P(Y \geq \alpha).$$

7.5 BASIC CONCEPTS AND FORMULAS

Uniform distribution $\mathcal{U}[a, b]$, $a < b$	$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{elsewhere;} \end{cases}$ $F(t) = \begin{cases} 0, & t < 0, \\ \dfrac{t-a}{b-a}, & a \leq t \leq b, \\ 1, & t > b; \end{cases}$ $E(X) = \dfrac{a+b}{2}, \quad \mathrm{Var}(X) = \dfrac{(b-a)^2}{12}$
Normal distribution $N(\mu, \sigma^2)$, $\mu \in \mathbb{R}, \sigma > 0$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}}\,\mathrm{e}^{-(x-\mu)^2/(2\sigma^2)}, \quad -\infty < x < \infty;$ $E(X) = \mu, \quad \mathrm{Var}(X) = \sigma^2$
Standard normal distribution $N(0, 1)$	$\phi(z) = \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}, -\infty < z < \infty;$ $\Phi(t) = \displaystyle\int_{-\infty}^{t} \dfrac{1}{\sqrt{2\pi}}\,\mathrm{e}^{-z^2/2}\,\mathrm{d}z;$ $E(X) = 0, \quad \mathrm{Var}(X) = 1$

Properties of the normal distribution	If $X \sim N(\mu, \sigma^2)$, then • $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$; • $P(a < X < b) = \Phi\left(\dfrac{b-\mu}{\sigma}\right) - \Phi\left(\dfrac{a-\mu}{\sigma}\right)$; • $\Phi(-z) = 1 - \Phi(z)$; • $\Phi(0) = 0.5$
Normal approximation to the binomial distribution (De Moivre–Laplace theorem)	If $X \sim b(n,p)$, then for the standardized variable $Z = (X-np)/\sqrt{npq}$, we have $$\lim_{n\to\infty} P(a < Z < b) = \Phi(b) - \Phi(a)$$
Normal approximation to the binomial distribution with continuity correction	If $X \sim b(n,p)$ for large n ($n \to \infty$), and i, j are nonnegative integers with $i \leq j$, then $$P(i \leq X \leq j) \cong \Phi\left(\frac{j+0.5-np}{\sqrt{npq}}\right) - \Phi\left(\frac{i-0.5-np}{\sqrt{npq}}\right)$$
Normal approximation to the Poisson distribution with continuity correction	If $X \sim \mathcal{P}(\lambda)$, then for large values of λ, the following approximation holds: $$P(i \leq X \leq j) \cong \Phi\left(\frac{j+0.5-\lambda}{\sqrt{\lambda}}\right) - \Phi\left(\frac{i-0.5-\lambda}{\sqrt{\lambda}}\right)$$ (i and j are nonnegative integers with $i \leq j$).
Exponential distribution $\mathcal{E}(\lambda)$	$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0; \end{cases}$$ $$F(t) = \begin{cases} 0, & t < 0, \\ 1 - e^{-\lambda t}, & t \geq 0; \end{cases}$$ $$E(X) = \frac{1}{\lambda}, \quad \mathrm{Var}(X) = \frac{1}{\lambda^2}$$

Memoryless (lack-of-memory) property of the exponential distribution	$P(X > s + t \mid X > s) = P(X > t), \quad \text{for } s, t \geq 0$
Gamma distribution with parameters a and λ ($a > 0, \lambda > 0$)	$f(x) = \begin{cases} \dfrac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, & x \geq 0, \\ 0, & x < 0, \end{cases}$ where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\, dt$ is the Gamma function. $E(X) = \dfrac{a}{\lambda}, \quad \mathrm{Var}(X) = \dfrac{a}{\lambda^2}$
Properties of the Gamma function	$\Gamma(a+1) = a\Gamma(a),\ a > 0$; $\Gamma(1) = 1$; $\Gamma(n+1) = n!$, for integers $n \geq 0$; $\Gamma(1/2) = \sqrt{\pi}$
Beta distribution with parameters α and β ($\alpha > 0, \beta > 0$)	$f(x) = \begin{cases} \dfrac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, & 0 < x < 1, \\ 0, & \text{otherwise}, \end{cases}$ where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\, dx = \dfrac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ is the Beta function; $E(X) = \dfrac{\alpha}{\alpha+\beta}, \quad \mathrm{Var}(X) = \dfrac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

7.6 COMPUTATIONAL EXERCISES

We have seen in Chapter 5 how we can deal with the most common discrete distributions in Mathematica. The first point to note here is that all functions listed in Table 5.1, and many others, are identical for discrete and continuous distributions; note, in particular, the dual use of the command *PDF[dist,x]* mentioned in Table 5.1.

Table 7.2 presents the notation used in Mathematica for the continuous distributions we have met in this chapter.

Notice especially the fact that, in the command NormalDistribution[m, s], the second parameter of the normal distribution is the **standard deviation**, and not the variance of the distribution.

Mathematica function	Distribution
UniformDistribution[a, b]	A uniform distribution over the interval $[a, b]$ with $a \leq b$
NormalDistribution[m, s]	A normal distribution with mean m and standard deviation s
ExponentialDistribution[s]	An exponential distribution with parameter s (mean $1/s$)
GammaDistribution[a, s]	A Gamma distribution with parameters a and s
BetaDistribution[a, b]	A Beta distribution with parameters a and b

Table 7.2 Mathematica functions for continuous distributions.

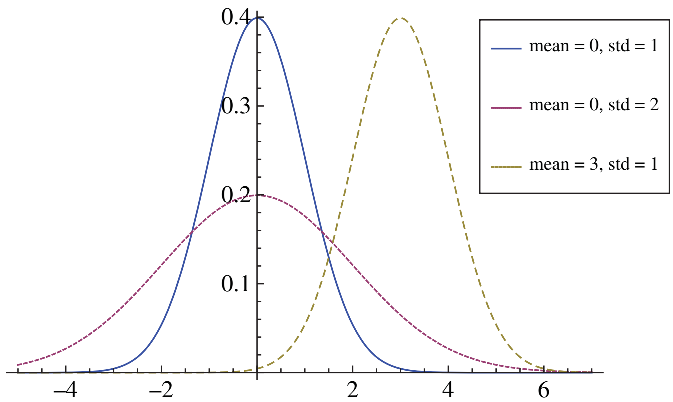
In the command sequence that follows, we give an example regarding the use of the normal distribution. Specifically, the commands below are used to answer the questions in Example 7.6:

```
In[1]:= mu=35;s=5;
        F[t_]:=CDF[NormalDistribution[mu,s],t]
        Print["Probability of the event X<30"]
        N[F[30]]
        Print["Probability of the event 30<X<40"]
        N[F[40]-F[30]]
        Print["Value of x so that P(X<x)=0.99"]
        a=Quantile[NormalDistribution[mu,s],0.99]

        Probability of the event X<30
Out[4]= 0.158655
        Probability of the event 30<X<40
Out[6]= 0.682689
        Value of x so that P(X<x)=0.99
        Out[8]= 46.6317
```

1. The following sequence of commands creates a plot of the density of the normal distribution, for various values of its parameters, on the same system of coordinates:

```
In[1]:= m1=0;s1=1;
        m2=0;s2=2;
        m3=3;s3=1;
        f1[x_]:=PDF[NormalDistribution[m1,s1],x]
        f2[x_]:=PDF[NormalDistribution[m2,s2],x]
        f3[x_]:=PDF[NormalDistribution[m3,s3],x]
        Needs["PlotLegends"]
        Plot[{f1[x],f2[x],f3[x]},{x,-5,7},PlotLegend->{"mean=0,
        std=1", "mean=0, std=2 ", "mean=3, std=1 "},PlotStyle->{Blue,
        Dashing[0.005],Dashing[0.01]}, LegendShadow->None,
        LegendPosition->{0.5,0.0},LegendSpacing->-0,LegendSize->0.6]
```

Out[8]=



Draw similar graphs for each of the following distributions: (i) exponential; (ii) Gamma; (iii) Beta.

2. Verify graphically, as well as using a table the limiting result in Proposition 7.3 (De Moivre–Laplace theorem). Start with $n = 5$ and then use successively $n = 10, 50, 100, 500$. Compare the results for five different choices of the success probability: $p = 0.5$ (symmetric case), $p = 0.1, p = 0.01$ (success is "rare'), and $p = 0.9$, $p = 0.99$ (success is almost certain). For which values of p is the approximation better?

3. The commands

   dist = NormalDistribution[m,s]

   rtable = Table[Random[dist],n]

   generate n random numbers from the normal distribution with mean m and standard deviation s (after we have assigned numerical values to each of n, m, s).

   Using these commands, generate a set of ν random numbers from the standard normal distribution. Repeat this for $\nu = 10, 50, 100, 500, 3000$. Find the usual arithmetic mean of the numbers produced for each value of ν, and compare that mean in each case with the mean $\mu = 0$ of the distribution where these numbers are supposed to come from. Try to explain what you observe as ν grows.

   Now repeat the above procedure for different values of $\sigma = 4, 10, 40, 200$. What do you observe in this case?

4. Use simulation to solve the following problem. Consider the quadratic equation

$$At^2 + Bt + C = 0,$$

   where the coefficients A, B, and C are random quantities with the uniform distribution on the interval $[-1, 3]$, the normal distribution with parameters $\mu = 1$ and $\sigma^2 = 1$

and the exponential distribution with mean $1/2$, respectively. Hence, derive an approximation for the probability that the quadratic equation has two real roots.

5. Use Mathematica to obtain the exact value of the probability associated with the binomial distribution in (7.14) of Example 7.8. For this example, calculate the percentage error of the normal approximation to the above value, with and without the continuity correction.

6. Let X be a random variable following the binomial distribution with parameters $n = 200$ and $p = 0.05$. Find the probabilities

$$P(X = x) = \binom{n}{x} p^x q^{n-x}, \quad x = 6, 7, \dots, 15,$$

and compare these values with the corresponding probabilities arising from the normal approximation (with continuity correction).

7. Let X be a random variable following the binomial distribution with parameters $n = 500$ and $p = 0.05$. For $x = 20, 21, \dots, 30$, calculate the probabilities $P(X \le x)$ using

   (i) the exact formula, from the binomial distribution;

   (ii) a normal approximation of the form

$$P(X \le x) \cong \Phi\left(\frac{x - np}{\sqrt{npq}}\right);$$

   (iii) a normal approximation with continuity correction,

$$P(X \le x) \cong \Phi\left(\frac{x - np + 0.5}{\sqrt{npq}}\right).$$

   For which values of x are the two approximations more accurate?

8. In Exercise 12 of Section 7.4, the formula of the so-called chi-squared distribution has been given. Draw a graph, on the same plot, of the density function of the chi-squared distribution with n degrees of freedom along with the normal density with mean n and variance $2n$. Perform this for $n = 5, 15, 50$. What do you observe?

9. There are various approximations available for the cumulative distribution function, Φ, of the standard normal distribution. One of them which, despite its simplicity, is quite accurate, employs the following formula:

$$\Phi_1(z) = 0.5 + \frac{z(4.4 - z)}{10},$$

   suggested by Shah (1985, *The American Statistician*).

Use Mathematica to draw a graph of Φ and Φ_1 over the interval $[-3, 3]$ and check (numerically) that, for $0 \le z \le 2.2$, we have

$$|\Phi(z) - \Phi_1(z)| \le 0.005.$$

Using this fact, explain how the approximation could be used so that the magnitude of the error is at most 0.005 for $z \in [-2.2, 2.2]$.

7.7 SELF-ASSESSMENT EXERCISES

7.7.1 True–False Questions

1. If $X \sim \mathcal{U}[2, 4]$, then $E(X^2) = 9$.
2. The duration of a train journey (in minutes) is a random variable $Y = 60 + X$, where X has the uniform distribution over the interval $[0, 6]$. Then, $\text{Var}(X) = 63$.
3. The repair time, X, of a machine (in hours) has the uniform distribution on the interval $[0, 4]$. Then,
$$P(X \le 3|X > 1) = 2/3.$$
4. If $X \sim N(\mu, \sigma^2)$, then the density of X attains its maximum value at $x = \mu$.
5. Let X be a random variable that has the standard normal distribution. Then the distribution of the variable $Y = X/3$ is also standard normal.
6. For a standard normal random variable Z, we have (up to three decimal places)
$$P(-1 \le Z \le 1) = 0.997.$$
7. Let X be a random variable having the standard normal distribution. Then, the variable X^2 also has a standard normal distribution.
8. Let X and Y be two random variables such that
$$X \sim \mathcal{E}(1/3), \quad Y \sim N(10, 9^2).$$
Then X and Y have the same variance.
9. If $X \sim N(\mu, \sigma^2)$, then the random variable $Y = 3X - 8$ also has a normal distribution.
10. If a random variable X is such that $X \sim N(1, 5^2)$, then $E[X(X - 1)] = 25$.
11. Suppose that the random variable X has a $N(10, 3^2)$ distribution. Then the probability $P(X < 9)$ equals $\Phi(2/3)$.
12. If a random variable X is such that $X \sim \mathcal{E}(\lambda)$, then $Y = 3X \sim \mathcal{E}(3\lambda)$.
13. If X has a Gamma distribution with parameters a and λ, then $Y = X - 1$ has a Gamma distribution with parameters $a - 1$ and λ.

14. If x is such that $\Phi(x) = 0.05$, then x is the upper 5% quantile of the distribution Φ.

15. The value of the Beta function, $B(\alpha, \beta)$, for $\alpha = 3$ and $\beta = 4$, is equal to $1/35$.

16. With Γ denoting the Gamma function, if it is known that $\Gamma(1/2) = \sqrt{\pi}$, then
$$\Gamma\left(\frac{5}{2}\right) = \frac{5}{2}\sqrt{\pi}.$$

17. If a continuous variable X has density function
$$f(x) = cx^2 e^{-3x}, \quad x \geq 0,$$
the value of c is $27/2$.

18. Let X be a continuous random variable having the uniform distribution over the interval $[0, n]$, for some positive integer n. Then the distribution of the integer part of X, i.e. of the variable $Y = [X]$ is discrete uniform (see Example 4.14 for its definition).

19. The uniform distribution over the interval $[0, 1]$ is a special case of the Beta distribution.

20. Assume that the distribution of a variable X is uniform over the interval $[1, 4]$. The upper-10%-quantile of this distribution is the point $x = 3.7$.

21. Let $X \sim \mathcal{U}[0, 1]$. Then, the distribution of the variable $Y = -\ln X$ is exponential.

7.7.2 Multiple Choice Questions

1. Let X be a random variable which has the uniform distribution over the interval $[0, a]$ for some $a > 0$. If it is known that $\mathrm{Var}(X) = 2E(X)$, then the value of a is

   (a) 3 (b) 12 (c) $\sqrt{12}$ (d) $\sqrt{6}$ (e) 6

2. Assume that the random variable X has the $\mathcal{U}[0, 2]$ distribution. Then the expectation of the variable $Y = X^3 - 2X + 4$ equals

   (a) 2 (b) 4 (c) 6 (d) 8 (e) 12

3. The time X to repair a fault follows a uniform distribution over the interval $[0, 6]$. Then, the expected value of $\sqrt{X}$ is

   (a) 3 (b) $\frac{1}{2}$ (c) $\frac{\sqrt{6}}{4}$ (d) $4\sqrt{6}$ (e) $\frac{72\sqrt{6}}{5}$

4. The profit X from a financial endeavor is assumed to follow a $\mathcal{U}[-a, a]$ distribution. If it is known that $\mathrm{Var}(X) = 12$, the value of a is

   (a) 3 (b) 6 (c) 9 (d) 12 (e) 72

5. If Φ denotes the distribution function of the standard normal distribution, and Z is a variable following this distribution, then for any $a \geq 0$ the probability $P(-a \leq Z \leq a)$ equals

   (a) $2\Phi(a) - 1$ (b) $1 - 2\Phi(a)$ (c) $\dfrac{1 - 2\Phi(a)}{2}$

   (d) $\dfrac{2\Phi(a) - 1}{2}$ (e) $2\Phi(a) + 1$

6. Assume that $Z \sim N(0, 1)$. Then, the probability $P(0 \leq Z \leq 1)$ equals

   (a) 0.5398 (b) 0.8413 (c) 0.8643 (d) 0.3413 (e) 0.1587

7. If $X \sim N(1, 25)$, then the expected value of the random variable $Y = (3 - X)/4$ is

   (a) $\dfrac{1}{4}$ (b) $\dfrac{1}{2}$ (c) $\dfrac{1}{25}$ (d) $\dfrac{1}{50}$ (e) none of the above

8. The density of a random variable X is given by

$$f(x) = \frac{1}{4\sqrt{2\pi}} \exp\left\{-\frac{(x-10)^2}{32}\right\}, \quad -\infty < x < \infty.$$

   Then, the probability $P(X > 20)$ equals

   (a) 0.5 (b) 0.0013 (c) 0.0062 (d) 0.9938 (e) 0.9876

9. The time that John needs to get from his house to the University every morning is a continuous random variable following the normal distribution with mean $\mu = 35$ minutes and a standard deviation $\sigma = 5$ minutes. The probability that on a particular day his journey takes at least 40 minutes is

   (a) 0.5398 (b) 0.8413 (c) 0.8643 (d) 0.3413 (e) 0.1587

10. Suppose the random variable X has a $N(27, 4)$ distribution. Then, the probability $P(25 < X < 31)$ equals

    (a) 0.1359 (b) 0.1498 (c) 0.5328 (d) 0.6628 (e) 0.8185

11. Student marks in a University examination, on a scale 0–100, are assumed to follow a $N(\mu, 12^2)$ distribution. If it is known that 9.68% of the students get a mark higher than 75, then the value of μ is

    (a) 59.4 (b) 62 (c) 90.6 (d) 88 (e) 73.7

12. Let X be a random variable having the exponential distribution with mean $1/2$. Then, the upper α-quantile of the distribution of X is

    (a) $\ln \alpha$ (b) $-2 \ln \alpha$ (c) $-\dfrac{\ln(1-\alpha)}{2}$

    (d) $-\dfrac{\ln \alpha}{2}$ (e) $-\ln(2\alpha)$

13. Let X be a random variable having density function

$$f(x) = 4e^{-4x}, \quad x \geq 0.$$

The probability $P(X > 4|X > 3)$ equals

(a) $e^{-3} - e^{-4}$ (b) $e^{-3/4}$ (c) e^{-4} (d) $e^{-1/4}$ (e) $4e^{-1}$

14. Let X be a random variable having a $N(\mu, 3^2)$ distribution. Then, approximately 95% of the values of X lie in the interval

(a) $[\mu - 3, \mu + 3]$ (b) $[\mu - 6, \mu + 6]$ (c) $[\mu - 18, \mu + 18]$

(d) $[\mu - 9, \mu + 9]$ (e) $[\mu - 36, \mu + 36]$

15. An important property of the exponential distribution is the **memoryless property**. Which of the following statements is correct?

(a) The exponential is the only probability distribution with that property.

(b) The exponential is the only continuous probability distribution with that property.

(c) The exponential and the geometric are the only continuous probability distributions with that property.

(d) The exponential and the normal are the only continuous probability distributions with that property.

(e) The exponential and the Gamma are the only continuous probability distributions with that property.

16. A continuous random variable X has density function

$$f(x) = cx^3 e^{-5x}, \quad x \geq 0.$$

Then the value of c is

(a) $\frac{1}{4}$ (b) $\frac{1}{2}$ (c) $\frac{625}{6}$ (d) $\frac{2}{625}$ (e) 6

17. A continuous variable X has density function

$$f(x) = cx^6 e^{-x}, \quad x \geq 0,$$

for a suitable constant c. Then the following holds for the value of c and the expectation $E(X)$:

(a) $c = 720,\ E(X) = 7$ (b) $c = \frac{1}{5040},\ E(X) = 1$

(c) $c = \frac{1}{720},\ E(X) = 7$ (d) $c = \frac{1}{5040},\ E(X) = 7$

(e) none of the above

18. A continuous variable X has density function

$$f(x) = cx^7(1 - x)^3, \quad 0 \leq x \leq 1.$$

The value of c is

(a) 252 (b) 840 (c) 360 (d) 1320 (e) 42

19. The lifetime X of an electronic equipment (in thousands of hours) has an Erlang distribution with parameters $n = 2$ and $\lambda = 2$. The probability that a particular equipment of this type has a lifetime exceeding its mean value equals

(a) $3e^{-2}$ (b) $2e^{-2}$ (c) $5e^{-4}$ (d) $3e^{-4}$ (e) $7e^{-6}$

20. The time X (in minutes) for a customer to be served at a post office has density function

$$f(x) = 2e^{-2x}, \quad x \geq 0.$$

The probability that each of the next three customers will be served within one minute equals

(a) $3e^{-2}$ (b) $1 - e^{-3}$ (c) $2e^{-6}$ (d) e^{-6} (e) $(1 - e^{-2})^3$

21. The daily change in the closing price of a certain stock, denoted by X, is assumed to follow a normal distribution with $\mu = 1$ and $\sigma = 1$. We select randomly three days and we want to find in how many of those days the price of the stock has increased. The probability that the stock price has gone up in exactly two of the three days is given by

(a) $(1 - \Phi(-1))^2$ (b) $\dfrac{(\Phi(1))^2}{1 - \Phi(1)}$ (c) $(\Phi(-1))^2(1 - \Phi(-1))$

(d) $3(\Phi(1))^2(1 - \Phi(1))$ (e) $(\Phi(1))^2(1 - \Phi(1))$

7.8 REVIEW PROBLEMS

1. Suppose X has the uniform distribution on the interval $[-1, \beta]$ with $\beta > -1$. If we have

$$\frac{\mathrm{Var}(X)}{E(X)} = \frac{4}{3},$$

(i) find the value of β;

(ii) obtain the density function, the mean and variance of the random variable $Y = |X|$.

2. Let X be the rounding error that arises when, after taking a certain measurement, we round it to the nearest real number with k decimal places.

(i) Derive the distribution of X.

(ii) Calculate the probability

$$P(10^k|X| < 0.3).$$

3. If $X \sim \mathcal{U}[0, 1]$, obtain the distribution function and the mean for the variables below:

(a) $Y = (\ln X)^2$;

(b) $W = e^{tX}$, where t is an arbitrary positive number.

4. Assume that X has the uniform distribution over the interval $[0,1]$. Find the distribution function and the expected value for the following random variables:
   (i) $Y = \ln(1/X) = -\ln X$;
   (ii) $Z = -\ln(1-X)$;
   (iii) $W = -\frac{1}{\lambda}\ln(1-X)$, where $\lambda > 0$ is a constant.

5. If $X \sim \mathcal{U}[0,1]$, obtain the distribution function and the expected value of the random variable $Y = \sqrt{X^2+1}$.

6. Assume that X is a continuous variable having the uniform distribution over the interval $[1-\theta, 1+\theta]$ for $0 < \theta < 1$. Find a function of the form

$$g(x) = ax^2 + bx + c, \qquad a, b, c, \in \mathbb{R},$$

such that $E[g(X)] = \mathrm{Var}(X)$ for any $\theta \in [0,1]$.

7. Suppose X has a Poisson distribution with parameter λ, and let a be a real number in the interval $(-1,1)$. Assuming that λ is sufficiently large so that the normal distribution can be used and c is a real number such that

$$F(c) = 1 - \frac{a}{2},$$

show that

$$P(\lambda - c\sqrt{\lambda} \le X \le \lambda + c\sqrt{\lambda}) \cong 1 - a.$$

8. Let X be a random variable having the standard normal distribution. Find the expected value for the following variables:
   (a) $X^3 \cos X$;
   (b) $\sin X + X\cos X$;
   (c) $X/(1+X^4)$;
   (d) $X\cos X/(1+X^2)$.

9. A factory produces 10 000 car tires per day, of which 5000 are of type A and 5000 are of type B. From the daily production, 90 tires are selected randomly for quality control. Using the normal approximation, calculate the probability that at least 40 tires from each type are selected for quality control on a given day.

10. Assume that X is a random variable such that $X \sim N(\mu, \sigma^2)$ and $a > 0$ is a real number. Find the value of x that maximizes the quantity

$$P(x \le X \le x + a).$$

11. An algorithm generates random numbers. If 3000 random digits $0, 1, \ldots, 9$ are generated,
    (i) what is the probability that the digit 4 appears at least 300 times?
    (ii) what is the probability that at least 800 digits which are multiples of 3 appear?
    (iii) find the probability that at least 1400 even digits (including zero) and 1400 odd digits are generated.

12. If Φ denotes the distribution function of the standard normal distribution, show that the function defined by

$$F(t) = \begin{cases} 0, & t < 0, \\ 2\Phi(t) - 1, & t \geq 0, \end{cases}$$

is another distribution function of a random variable X. Verify also that if $Z \sim N(0, 1)$, then the function F above is the distribution function of the variable $X = |Z|$.

13. The distribution function of a variable X is given by

$$F(t) = \int_{-\infty}^{t} ka^{-x^2}\, \mathrm{d}x,$$

where $k > 0$ and $a > 0$ are two given real numbers. Prove that the distribution of X is $N(\mu, \sigma^2)$ and express the constant k and the parameters μ and σ^2 of the normal distribution in terms of a.

14. Let Z be a random variable that has the standard normal distribution, $N(0, 1)$, with distribution function Φ and density function ϕ.

(i) Prove that for each $z > 0$, we have

$$\frac{1}{z}\left(1 - \frac{1}{z^2}\right) \leq 1 - \Phi(z) \leq \frac{1}{z}\phi(z).$$

(ii) Use the result of Part (i) to show the relationship

$$\lim_{z\to\infty} \frac{z(1 - \Phi(z))}{\phi(z)} = 1.$$

(Incidentally, this result shows that for large values of z, the distribution function $\Phi(z)$ of the $N(0, 1)$ distribution can be approximated by using the formula

$$\Phi(z) \cong 1 - \frac{1}{z}\phi(z).\ )$$

(iii) Show that for any $z > 0$, we have

$$\lim_{t\to\infty} P\left(Z > t + \frac{z}{t} | Z \geq t\right) = \mathrm{e}^{-z}.$$

(*Hint*: For Part (i), integrate, over a suitable range, each side of the inequality

$$(1 - 3t^{-4})\mathrm{e}^{-t^2/2} < \mathrm{e}^{-t^2/2} < (1 + t^{-2})\mathrm{e}^{-t^2/2}.\ )$$

15. In order to check the validity of an algorithm used to generate random numbers from the set $\{1, 2, \ldots, 50\}$, we select 200 000 numbers produced by the algorithm. We observe that 4800 of these numbers are equal to 1. Given that the expected number of 1's in this sample is $200\,000/50 = 4000$, can we conclude that the algorithm is not working well?

16. The recovery time of a patient from a medical operation is a random variable following the exponential distribution with a mean value of 15 days. Find the probability that

    (i) a patient takes between 13 and 16 days to recover;

    (ii) a patient who has not recovered by the end of the 13th day will recover within the next three days;

    (iii) if 15 patients are subject to this operation, at least 5 of them will have a recovery time between 13 and 16 days (the remaining patients have a recovery time of less than 13 days or more than 16 days).

17. The radius R of spherical bubbles produced by an automatic dispensing machine is a random variable with density function

$$f_R(r) = ce^{-50r^2+100r}, \quad -\infty < r < \infty.$$

    (i) Verify that R follows a normal distribution and obtain its mean and variance.

    (ii) Derive the density function for the *area* $S = 4\pi R^2$ of these bubbles.

18. For each of the following cases, calculate the probability

$$P(\mu - 2\sigma \le X \le \mu + 2\sigma),$$

    where $\mu = E(X)$ and $\sigma^2 = \text{Var}(X)$ and compare this probability with the corresponding lower bound from Chebyshev's inequality (Proposition 6.6):

    (i) X has a $N(\mu, \sigma^2)$ distribution;

    (ii) X has a uniform distribution over the interval $[0, a]$, for $a > 0$;

    (iii) X has an exponential distribution with mean λ^{-1}.

19. A continuous random variable X has density function

$$f(x) = ce^{-|x|}, \quad -\infty < x < \infty.$$

    (i) Find the value of c.

    (ii) Derive the density function of the random variable $Y = aX + b$, where $a \neq 0$ and b are given real numbers.

    (iii) Verify that the moments of order r,

$$\mu_r = E(X^r), \qquad r = 1, 2, \dots,$$

    of the distribution of X are given by

$$E(X^r) = \begin{cases} (2k)!, & r = 2k \quad \text{with } k = 1, 2, \dots \\ 0, & r = 2k + 1 \quad \text{with } k = 0, 1, 2, \dots \end{cases}$$

    [The distribution considered in this exercise is known as the **Laplace** (or **double exponential**) **distribution**.]

20. Suppose that the lifetime, T, of an electronic device (in hours) until it needs service has an exponential distribution with mean $1/\lambda$ ($\lambda > 0$). The operation of the machine that uses this device costs a_1 dollars per hour. A machine operator is

hired for a *specified* number of hours, h, and is paid a_2 dollars per hour (that is, the operator will receive $a_2 h$ dollars no matter how many hours the machine works). We assume further that for each hour of operation, the profit of the company that uses this machine is a_3 dollars.

(i) Show that the total profit, K, for the company until servicing the electronic device is given by

$$K = \begin{cases} (a_3 - a_2 - a_1)h, & \text{if } T > h, \\ (a_3 - a_1)T - a_2 h, & \text{if } T \leq h. \end{cases}$$

(ii) Obtain an analytic expression for the expected cost $E(K)$ and verify that this is maximized when the specified number of hours, h, that the operator is hired for, equals

$$h = -\frac{1}{\lambda} \ln\left(\frac{a_2}{a_3 - a_1}\right).$$

Then, examine the restrictions that the quantities a_1, a_2 and a_3 must satisfy.

(iii) Calculate the optimal value of h for the special case when $E(T) = 1000$ (hours), $a_1 = 35, a_2 = 40$, and $a_3 = 90$ (in dollars).

21. The lifetime of a bulb has the exponential distribution with parameter λ. If n new light bulbs are installed in a room, what is the probability that at time t at least half of them are still working?

22. An area that has a small lake is visited frequently by a rare bird. A scientist arrives at the lake and the time X, in minutes, until she sees the bird is a continuous random variable with density function

$$f(x) = \frac{1}{4}\mathrm{e}^{-x/2} + \frac{1}{2}x\mathrm{e}^{-x}, \quad x > 0.$$

Calculate

(i) the distribution function of the scientist's waiting time X;

(ii) the probability that the time period until she sees the bird is more than two minutes, but less than four minutes;

(iii) the expectation and variance of X.

[This exercise illustrates the fact that we can form new probability distributions by considering linear combinations of known ones (provided that the respective weights are nonnegative and their sum equals one). See also Exercise 14 of Section 7.3.]

23. The total lifetime of an electronic equipment has the exponential distribution. For the production of this equipment, the manufacturing company has two alternative procedures, denoted by I and II, respectively. When procedure I is followed, the mean lifetime of the equipment is α hours, while when procedure II is followed the mean lifetime of the equipment is $3\alpha/2$ hours.

However, the total cost for the company when procedure II is followed is twice as much compared with procedure I, which has a cost of β dollars. Suppose finally that the company offers a guarantee of 2α hours for this equipment, so that each item that fails before that length of time is returned to the manufacturer, incurring a loss of γ dollars.

(i) Calculate the expected cost of the manufacturer per item, for each of the two alternative procedures used in the production.

(ii) Obtain a condition that must hold for β and γ so that procedure I is preferable for the manufacturer compared to procedure II.

24. A general result in probability theory says that the distance between the mean and the median of an arbitrary distribution cannot exceed the standard deviation for this distribution. More explicitly, let μ, m, and σ denote the expected value, the median and the standard deviation of a random variable X. Then, we have

$$|\mu - m| \leq \sigma.$$

(i) Use Cantelli's inequality (see Exercise 10 in Section 4.6) to prove this result.

(ii) Verify that the inequality holds for

(a) a uniform distribution over the interval $[a, b]$,

(b) an exponential distribution with parameter λ,

by finding first the median of the distribution in each case (for the normal distribution, the inequality holds trivially, since $|\mu - m| = 0$).

25. Let X be a continuous random variable having density function

$$f(x) = \begin{cases} \lambda x e^{-\lambda x^2/2}, & x \geq 0, \\ 0, & x < 0. \end{cases}$$

(i) Find the distribution function of X and calculate the probability $P(1 < X < 3)$.

(ii) Show that the random variable $Y = X^2/2$ has an exponential distribution with parameter λ. Hence, find the mean and variance of Y.

26. A random variable X follows the continuous **power function distribution** with parameters $\beta > 0$ and $\gamma > 0$ if its probability density is given by

$$f(x) = \begin{cases} \dfrac{\gamma x^{\gamma-1}}{\beta^\gamma}, & 0 < x < \beta, \\ 0, & \text{otherwise.} \end{cases}$$

(i) Find the distribution of the random variable $1/X$.

(ii) Verify that the distribution of the random variable $W = X/\beta$ is a special case of the Beta distribution. Thus, by finding the mean and variance of W, obtain the mean and variance of the variable X.

(iii) Show that the random variable $-\ln(X/\beta) = -\ln W$ follows an exponential distribution and calculate the quantities $E(\ln X)$ and $\mathrm{Var}(\ln X)$.

(iv) Prove that X has the same distribution as the random variable $\beta U^{1/\gamma}$, where U is uniformly distributed over the interval $[0, 1]$.

27. Let X be a variable having the $\mathcal{U}[0, 4]$ distribution. Determine the density function of the random variable Y, which is related to X through the transformation

$$Y = \begin{cases} X, & \text{if } 0 < X < 2, \\ 4 - X, & \text{if } 2 \le X \le 4, \\ 0, & \text{otherwise.} \end{cases}$$

28. Assume that a variable X has the Beta distribution with parameters α and β where α and β are both positive integers. Then, show that the distribution function of X is given by the formula

$$F(t) = P(X \le t) = t^{\alpha} \sum_{k=0}^{\beta-1} \binom{\alpha+\beta-1}{\alpha+k} t^{k}(1-t)^{\beta-1-k}, \qquad 0 \le t \le 1.$$

29. An automatic machine cuts a loaf of bread into toast bread slices. The weight, X, of each slice has a normal distribution with mean 32 g and a standard deviation 1.5 g. These slices are then put into packs of 20 to be sold. A pack is considered unacceptable if it has at least two slices whose weight is more than two standard deviations away from the mean value of 32 g. An inspection team selects 100 packs randomly to check whether they meet the production standards. Find the probability that at least five packs in this sample are unacceptable for selling purposes.

30. Let X be a random variable with density function

$$f(x) = c\,\frac{e^{-x}}{(1+e^{-x})^2}, \qquad -\infty < x < \infty.$$

(i) Find the constant c.

(ii) Obtain an expression for the associated distribution function.

(iii) Plot the density function $f(x)$ and comment on it.

(iv) Find $E(X)$ and the median of the distribution, compare these two quantities and comment.

(This distribution is called the **logistic distribution** and it has found numerous applications in diverse areas including economics, finance, demography, and dose–response studies.)

31. Let X have a Gamma distribution with parameters $a > 0$, and $\lambda > 0$, as given in Definition 7.5. Further, let $Y = \ln X$.

(i) Derive the density function of Y.

(ii) Give an expression for the distribution function of Y.

(iii) How would you find the probability that Y is negative?

This distribution is called the **log-gamma distribution** and it has found important applications in reliability and survival analysis.

32. Let X be a positive continuous random variable, describing the lifetime in hours of a unit, e.g. the time until the first breakdown of a pump or an electric oven. Denote by $f(x)$ the density function of X and by $F(t)$ its cumulative distribution function.

    (i) If the unit has survived for $x > 0$ hours, what is the probability that it does not survive for an additional time of t hours? Express the conditional probability in terms of F.

    (ii) The function
    $$r(x) = \frac{f(x)}{1 - F(x)}, \qquad x > 0$$
    is called the **failure rate** (or **hazard rate**) function of the unit. Verify that
    $$r(x) = \lim_{t \to 0} \frac{P(X \le x + t | X > x)}{t},$$
    where the term in the numerator on the right is the conditional probability calculated in (i). The last expression permits writing the approximate formula
    $$P(X \le x + t | X > x) \cong r(x)t$$
    for small values of t ($t \to 0$).

    (iii) Let $0 < x_1 < x_2$ and assume that $r(x_1) < r(x_2)$. Explain which of the following two events is more likely to occur:
    A: an x_1-hour old unit will fail in the next few t hours ($t \to 0$);
    B: an x_2-hour old unit will fail in the next few t hours ($t \to 0$).

    (iv) Show that the cumulative distribution function $F(t)$ and the density function $f(x)$ can be expressed in terms of $r(x)$ through the formulas
    $$F(t) = 1 - \exp\left[-\int_0^t r(x)dx\right] \quad \text{and} \quad f(x) = r(x)\exp\left[-\int_0^x r(s)ds\right].$$

33. Calculate the failure rate function (see the previous exercise) of a random variable X that follows the exponential distribution with parameter $\lambda > 0$. Next, make use of the approximate formula
    $$P(X \le x + t | X > x) \cong r(x)t$$
    to justify the memoryless property of the exponential distribution, by noting that for fixed (small) t, the conditional probability $P(X \le x + t | X > x)$ is independent of x and therefore
    $$P(X \le x + t | X > x) = P(X \le t | X > 0) = P(X \le t).$$

34. If X is a random variable with density function
    $$f(x) = \frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}, \qquad x \ge 0$$
    (Gamma distribution with parameters $a > 0$ and $\lambda > 0$), show that for the corresponding failure rate we may write
    $$\frac{1}{r(x)} = \int_x^\infty \left(\frac{u}{x}\right)^{a-1} e^{-\lambda(u-x)}\, du = \int_0^\infty \left(1 + \frac{s}{x}\right)^{a-1} e^{-\lambda s} ds.$$

Then prove that

(i) if $a > 1$, then $r(x)$ is an increasing function of x. In such a case we say that X has an Increasing Failure Rate (IFR);

(ii) if $a < 1$, then $r(x)$ is an decreasing function of x. In such a case we say that X has a Decreasing Failure Rate (DFR);

(iii) if $a = 1$, then $r(x)$ is constant (to be expected, since this corresponds to the exponential case).

35. Calculate the failure rate function $r(x)$ of the Weibull distribution (see Example 7.13) with density function

$$f(x) = \frac{1}{\beta}\, x^{1/\beta - 1} \cdot \exp(-x^{1/\beta}), \qquad x > 0.$$

Investigate the monotonicity of $r(x)$ for the cases $\beta > 1$, $\beta < 1$ and $\beta = 1$.

7.9 APPLICATIONS

7.9.1 Transforming Data: The Lognormal Distribution

We have seen in this chapter that the normal distribution is the most important probability distribution for applications. But, there are several practical situations in which the (*natural) logarithm* of a random variable X, which is of interest, has a normal distribution. For instance,

- the distribution of blood pressure for female adults does not follow a normal distribution, but its logarithm provides a rather good fit to that distribution;
- if X represents the financial loss from a certain natural disaster (fire, typhoon, etc.), then in many cases $Y = \ln X$ is closer to a normal distribution than X itself;
- for many medical substances, the logarithm of the amount of the substance that stays in the human organism after a certain period of time has approximately a normal distribution;
- in hydrology, logarithms of the annual maximum for daily rainfall amounts are approximately normal (the same is often true for river discharge volumes);
- a more recent application was suggested in an article by Sobkowicz et al. (2013). Specifically, the logarithm of the length of comments posted in Internet discussion pages can, in many cases, be well approximated by a normal distribution.

A random variable X whose natural logarithm, $\ln X$, has a $N(\mu, \sigma^2)$ distribution is said to follow a **lognormal distribution** with parameters μ and σ^2. This is abbreviated by writing $X \sim LN(\mu, \sigma^2)$. The lognormal distribution is one of the continuous distributions available in Mathematica. For example, the command

```
F[x_]:=CDF[LogNormalDistribution[0, 1], x]
```

defines a function $F(x)$ as the distribution function of the lognormal distribution with parameters $\mu = 0$ and $\sigma = 1$. Note that, in analogy with the normal distribution, Mathematica accepts as input parameters μ and σ, rather than μ and σ^2; so, with our notation,

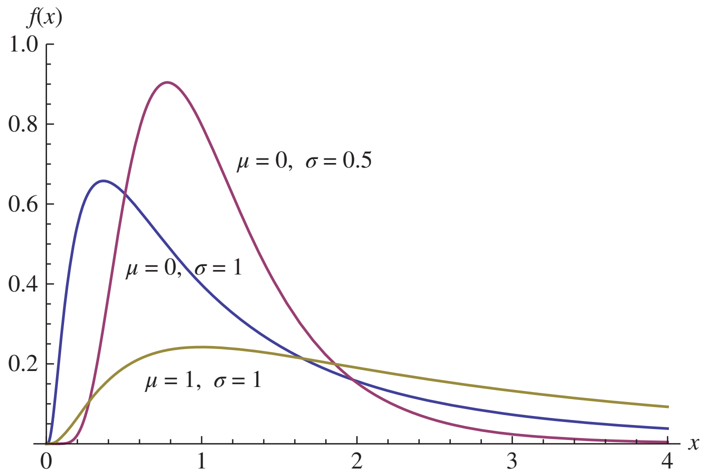


Figure 7.17 The density function of the lognormal distribution for different values of the parameters μ and σ.

if we want, e.g. to define F as the distribution function of the lognormal with $\mu = 2$ and $\sigma^2 = 9$, we must write

```
F[x_]:=CDF[LogNormalDistribution[2, 3], x]
```

Observe that, even without using Mathematica, it is very easy to find an expression for the distribution function F associated with $X \sim LN(\mu, \sigma^2)$. Specifically, for such a random variable X, set $Y = \ln X$ so that $Y \sim N(\mu, \sigma^2)$. Then, for $t > 0$, we have

$$\begin{aligned} F(t) &= P(X \le t) = P(\ln X \le \ln t) = P(Y \le \ln t) \\ &= P\left(\frac{Y-\mu}{\sigma} \le \frac{\ln t - \mu}{\sigma}\right) = \Phi\left(\frac{\ln t - \mu}{\sigma}\right). \end{aligned} \tag{7.22}$$

Differentiating this distribution function, we obtain the density function of X as

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp\left[-\frac{(\ln x - \mu)^2}{2\sigma^2}\right], \quad x > 0.$$

Moreover, for $0 < a < b$, we see from (7.22) that

$$P(a < X \le b) = P(X \le b) - P(X \le a) = \Phi\left(\frac{\ln b - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a - \mu}{\sigma}\right). \tag{7.23}$$

Figure 7.17 shows the density function of the lognormal distribution for three choices of its parameters: (i) $\mu = 0, \sigma = 1$; (ii) $\mu = 0, \sigma = 1/2$; and (iii) $\mu = 1, \sigma = 1$.

A common feature of the three curves on the graph is that they are all *skewed to the right* (intuitively, the density function has a longer "tail" at the right end of the graph); this is more pronounced in the curves with $\sigma = 1$, and less so when the parameter σ decreases. This characteristic underlies the use of the lognormal distribution in a variety of practical studies. A number of diverse applications involve data sets that exhibit some form of "skewness to the right," thus producing a histogram of the data with a shape that would resemble one of the curves in Figure 7.17 (or at least exhibit a similar lack of symmetry).

On the other hand, from a statistical viewpoint it is often desirable to have data coming from a distribution which is, at least approximately, normal; many statistical methods rely on the crucial assumption of normality. Can we ever use any of these methods if our data are skewed? The lognormal distribution provides a comforting, affirmative answer to this. Suppose we have a sample of blood pressure measurements from a set of female patients. As mentioned at the beginning of this section (and has been found in a number of empirical studies), taking the logarithms of these values may provide us with a data set which is consistent with the assumption of normality. We can then analyze these transformed data and draw conclusions from them. All we have to do next is to translate these conclusions back to our original data set.

Of course, taking the logarithms for a data set does not always produce data that are normally distributed (if this happens, it means from the discussion above that the original data arose from a lognormal distribution!). However, this trick often works, so that a logarithmic transformation is something like a "rule of thumb' for statisticians facing skewed data. Some other transformations are also used in practice, such as the square root transformation, the cubic root transformation and so on. All these are special cases of the more general family of **power transformations** (also known as Box–Cox transformations, named after the two statisticians who proposed them in 1964).

Let us now look at a particular example (theoretical, i.e. no data involved!) in which the lognormal distribution may serve as a model. The fuel produced at an oil refinery contains a liquid product which improves the pre- and post-combustion stages. The percentage concentration of this product is represented by a random variable X, which is assumed to follow the lognormal distribution with parameters μ and σ^2. The ideal requirements for the oil production are that X must lie in an interval (a, b), where a and b are given positive real numbers. To be precise, if $X \in (a, b)$, the company producing the fuel estimates a profit of c_1 cents per liter of fuel produced.

If it happens that $X \leq a$, the company needs to repeat the combustion stage, and this results in a loss of c_2 cents per liter produced. If, on the other hand, $X \geq b$, there is a loss of c_3 cents (per liter) due to unnecessary waste of the chemical improver.

Suppose now that the company wants to calculate the expected profit per liter of oil produced. Write W for the profit per liter; it is then clear that W is a discrete random variable that assumes only three values, namely, $c_1, -c_2, -c_3$ (in the last two cases we have a loss rather than a profit) with respective probabilities $P(a < X < b)$, $P(X \leq a)$ and $P(X \geq b)$. Using (7.22) and (7.23), we thus see that the expected profit per liter is given by

$$
\begin{aligned}
E(W) &= c_1 P(a < X < b) - c_2 P(X \leq a) - c_3 P(X \geq b) \\
&= c_1 \left[\Phi\left(\frac{\ln b - \mu}{\sigma}\right) - \Phi\left(\frac{\ln a - \mu}{\sigma}\right) \right] - c_2 \Phi\left(\frac{\ln a - \mu}{\sigma}\right) \\
&- c_3 \left[1 - \Phi\left(\frac{\ln b - \mu}{\sigma}\right) \right] \\
&= (c_1 + c_3)\Phi\left(\frac{\ln b - \mu}{\sigma}\right) - (c_1 + c_2)\Phi\left(\frac{\ln a - \mu}{\sigma}\right) - c_3.
\end{aligned}
$$

It is clear that this is a linear function of c_1 (a and b are assumed known), and so it is very easy for the company to estimate its expected profit for a given value of c_1 (or, vice versa, to set the value of c_1 for a desired level of profit).

As a final illustration, assume that all parameters are explicitly known except μ; the company wants to calculate the value of μ that maximizes its expected profit. To find this, let $g(\mu)$ be the expected profit, viewed as a function of the parameter μ. In order to find the maximum, we differentiate $g(\mu)$ with respect to μ; a straightforward calculation, invoking the chain rule for derivatives, gives

$$g'(\mu) = \frac{c_1 + c_2}{\sigma} \cdot \phi\left(\frac{\ln a - \mu}{\sigma}\right) - \frac{c_1 + c_3}{\sigma} \cdot \phi\left(\frac{\ln b - \mu}{\sigma}\right),$$

where ϕ is the standard normal density function, given by $\phi(x) = (\sqrt{2\pi})^{-1} \mathrm{e}^{-x^2/2}$ for $x \in \mathbb{R}$. Setting this derivative equal to zero yields, after a little rearrangement,

$$\exp\left[\frac{(\ln a - \mu)^2 - (\ln b - \mu)^2}{2\sigma^2}\right] = \frac{c_1 + c_2}{c_1 + c_3}.$$

Taking logarithms on both sides, we get

$$\frac{(\ln a - \mu)^2 - (\ln b - \mu)^2}{2\sigma^2} = \ln\left(\frac{c_1 + c_2}{c_1 + c_3}\right).$$

The numerator on the left-hand side is equal to $(\ln a - \ln b)(\ln a + \ln b - 2\mu)$, so that the last equation takes the form

$$\ln a + \ln b - 2\mu = \frac{2\sigma^2}{\ln a - \ln b} \cdot \ln\left(\frac{c_1 + c_2}{c_1 + c_3}\right).$$

Solving for μ from the above equation, we finally obtain that

$$\mu = \frac{\ln a + \ln b}{2} + \frac{\sigma^2}{\ln b - \ln a} \cdot \ln\left(\frac{c_1 + c_2}{c_1 + c_3}\right).$$

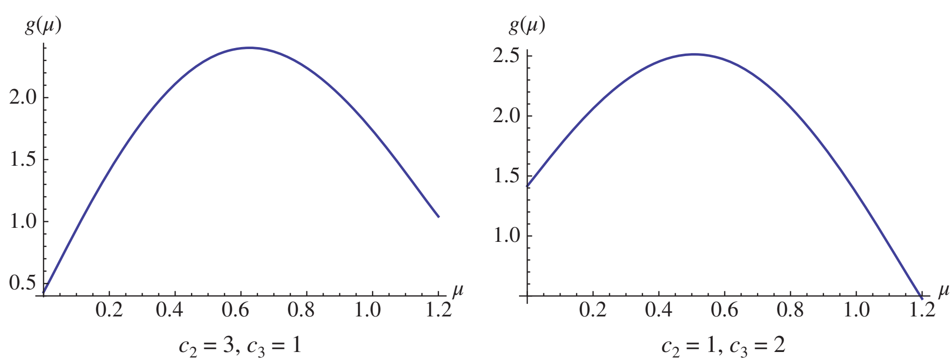


Figure 7.18 The expected profit of the company as a function of μ.

Though it is a rather tedious exercise to do, it can be verified that the second derivative of g at this point is negative, and so μ given above is indeed the value that maximizes the expected profit of the company.

Further, it is clear from the last expression that, when $c_2 = c_3$, the second term on the right above vanishes so that the value of μ, where the maximum occurs, is simply the arithmetic average of $\ln a$ and $\ln b$. For $c_1 \neq c_2$, the function $g(\mu)$ first increases and then decreases so that it attains a maximum at the value of μ found above. Figure 7.18 shows the function $g(\mu)$ for $a = 1, b = 3, \sigma = 1/2, c_1 = 4$ and two sets of values for c_2, c_3.

KEY TERMS

Beta distribution
Beta function
continuity correction
De Moivre–Laplace theorem
exponential distribution
failure period
Gamma distribution
Gamma function
lack of memory (or memoryless) property
lognormal distribution
normal distribution (or Gaussian distribution)
quantile of a distribution
standard normal distribution
standardization of a variable
uniform distribution
useful period
wear-out period
Weibull distribution

APPENDIX A

SUMS AND PRODUCTS

Suppose we have a given set of n real numbers $a_1, a_2, \ldots, a_n$. The symbol

$$\sum_{i=1}^{n} a_i$$

is an abbreviation for the sum $a_1 + a_2 + \cdots + a_n$. Often, and when no confusion arises, we simply write $\sum a_i$. It is apparent that in the above notation, i is merely used as a symbol to indicate the starting ($i = 1$) and the final ($i = n$) index in the summation and that in no way affects the value of the sum. Therefore, for instance,

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j = \sum_{x=1}^{n} a_x = a_1 + a_2 + \cdots + a_n.$$

It is also clear that the range of indices in the summation could be any subset of $\{1, 2, \ldots, n\}$ such that the lower limit is smaller than or equal to the upper limit. For example, if n and k are two given integers such that $k \le n$, we can write

$$\sum_{i=k}^{n} a_i = \sum_{x=k}^{n} a_x = a_k + a_{k+1} + \cdots + a_n,$$

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

or

$$\sum_{j=0}^{n-k} a_{k+j} = a_k + a_{k+1} + \cdots + a_n.$$

Note that the right-hand side in the last two equations is the same. In fact, we say that the last equation arises from the previous one by a "change of variable". More explicitly, if we consider the sum

$$S = \sum_{i=k}^{n} a_i$$

and make the substitution $j = i - k$, we have that $i = k + j$ so that the term a_i in the summation has to be replaced by a_{k+j}, that is

$$S = \sum a_{k+j}.$$

In order to determine the limits in the summation above, we observe that for $i = k$ we get $j = k - k = 0$ while for $i = n$ we have $j = n - k$. Therefore,

$$S = \sum_{j=0}^{n-k} a_{k+j}.$$

Using the $\sum$ notation we can write in a succinct way a large variety of simple, or more complicated, expressions. For instance, the sum of the first n positive integers is written as

$$1 + 2 + 3 + \cdots + n = \sum_{i=1}^{n} i,$$

while the sum of squares of the real numbers $a_1, a_2, \ldots, a_n$ is

$$a_1^2 + a_2^2 + \cdots + a_n^2 = \sum_{i=1}^{n} a_i^2.$$

Suppose now that we have another set of real numbers, $b_1, b_2, \ldots, b_n$. Then we can write for instance

$$\sum_{i=1}^{n} a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n,$$

$$\sum_{i=1}^{n} (i a_i + b_i) = (1 a_1 + b_1) + (2 a_2 + b_2) + \cdots + (n a_n + b_n),$$

$$\sum_{i=1}^{n} a_i^3 b_i = a_1^3 b_1 + a_2^3 b_2 + \cdots + a_n^3 b_n,$$

$$\sum_{i=k}^{n} \left(a_i + \frac{b_i}{i}\right) = \left(a_k + \frac{b_k}{k}\right) + \left(a_{k+1} + \frac{b_{k+1}}{k+1}\right) + \cdots + \left(a_n + \frac{b_n}{n}\right).$$

The next proposition collects some elementary properties for operations involving the summation symbol.

Proposition A.1 *Let $a_1, a_2, \dots, a_n, b_1, b_2, \dots b_n$ and λ be any real numbers. Then the following hold:*

$$\sum_{i=1}^{n}(a_i + b_i) = \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i, \tag{A.1}$$

$$\sum_{i=1}^{n}(\lambda a_i) = \lambda \sum_{i=1}^{n} a_i, \tag{A.2}$$

$$\sum_{i=1}^{n}(a_i - b_i) = \sum_{i=1}^{n} a_i - \sum_{i=1}^{n} b_i. \tag{A.3}$$

Proof: We have

$$\begin{aligned}\sum_{i=1}^{n}(a_i + b_i) &= (a_1 + b_1) + (a_2 + b_2) + \cdots + (a_n + b_n)\\ &= (a_1 + a_2 + \cdots + a_n) + (b_1 + b_2 + \cdots + b_n)\\ &= \sum_{i=1}^{n} a_i + \sum_{i=1}^{n} b_i\end{aligned}$$

and

$$\sum_{i=1}^{n}(\lambda a_i) = \lambda a_1 + \lambda a_2 + \cdots + \lambda a_n = \lambda \sum_{i=1}^{n} a_i.$$

The last result is an immediate deduction of the first two (by putting $\lambda = -1$). □

In the case where all a_i's are equal, so that $a_i = a$ for $i = 1, 2, \dots, n$, it is obvious that

$$\sum_{i=1}^{n} a_i = na. \tag{A.4}$$

Example A.1 Let $\overline{a}$ be the arithmetic average of $a_1, a_2, \dots, a_n$. Then

(i) $\displaystyle\sum_{i=1}^{n}(a_i - \overline{a}) = 0,$

(ii) $\displaystyle\sum_{i=1}^{n}(a_i - \overline{a})^2 = \sum_{i=1}^{n} a_i^2 - n\overline{a}^2.$

SOLUTION We put

$$S = \sum_{i=1}^{n} a_i,$$

so that

$$\overline{a} = \frac{a_1 + a_2 + \cdots + a_n}{n} = \frac{S}{n}.$$

Therefore,

$$\begin{aligned}\sum_{i=1}^{n}(a_i-\overline{a}) &= \sum_{i=1}^{n} a_i - \sum_{i=1}^{n}\overline{a} \quad \text{(from Proposition 1.1)}\\ &= S - n\overline{a} \quad \text{(from (1.4))}\\ &= S - n\cdot\frac{S}{n} = 0.\end{aligned}$$

For Part (ii), we observe first that (put $b_i = -\overline{a}$ for $i = 1, 2, \ldots, n$ in the result of Exercise 8)

$$\sum_{i=1}^{n}(a_i-\overline{a})^2 = \sum_{i=1}^{n} a_i^2 - 2\sum_{i=1}^{n} a_i\overline{a} + \sum_{i=1}^{n}\overline{a}^2.$$

But from the second part of Proposition 1.1 and (1.4) we obtain respectively that

$$\sum_{i=1}^{n}(a_i\overline{a}) = \overline{a}\sum_{i=1}^{n} a_i = \frac{S}{n}\cdot S = \frac{S^2}{n}$$

and

$$\sum_{i=1}^{n}\overline{a}^2 = n\overline{a}^2 = n\left(\frac{S}{n}\right)^2 = \frac{S^2}{n}.$$

Thus, we finally obtain

$$\begin{aligned}\sum_{i=1}^{n}(a_i-\overline{a})^2 &= \sum_{i=1}^{n} a_i^2 - 2\frac{S^2}{n} + \frac{S^2}{n} = \sum_{i=1}^{n} a_i^2 - \frac{S^2}{n}\\ &= \sum_{i=1}^{n} a_i^2 - n\left(\frac{S}{n}\right)^2 = \sum_{i=1}^{n} a_i^2 - n\overline{a}^2,\end{aligned}$$

as required.

Example A.2 Consider the sum

$$S = a + aw + aw^2 + \cdots + aw^{n-1} = \sum_{i=1}^{n} aw^{i-1}$$

of the first n terms in a geometric series. Show that the sum S in the above formula equals

$$\sum_{i=1}^{n} aw^{i-1} = \frac{a(1-w^n)}{1-w},$$

provided that $w \neq 1$. (For $w = 1$, we have trivially $S = na$.)

SOLUTION Multiplying each term of S by w we obtain that

$$wS = aw + aw^2 + aw^3 + \cdots + aw^n.$$

Therefore,

$$\begin{aligned} wS - S &= (aw + aw^2 + aw^3 + \cdots + aw^n) - (a + aw + aw^2 + \cdots + aw^{n-1}) \\ &= aw^n - a, \end{aligned}$$

and for $w \neq 1$ this gives immediately

$$S = \frac{a(w^n - 1)}{w - 1} = \frac{a(1 - w^n)}{1 - w}.$$

The following result is particularly useful.

Proposition A.2 *The infinite sum of the geometric series whose first term is a and the ratio between two successive terms is w is given by*

$$\sum_{i=1}^{\infty} aw^{i-1} = \frac{a}{1 - w}, \quad \textit{for } |w| < 1.$$

Proof: We notice that for $|w| < 1$, the limit of the sequence $\{w^n : n = 1, 2, \ldots\}$, as $n \to \infty$, is zero. The result is now immediate by taking the limit $(n \to \infty)$ in the result of Example A.2. □

The shorthand notation we use writing $\sum$ for a sum can be extended in the case where we consider a "double sum." This arises when we have symbols with two indices, say i, j, as subscripts. For instance, consider the real numbers

$$a_{11}, a_{12}, \ldots, a_{1k}, a_{21}, a_{22}, \ldots, a_{2k}, \ldots, a_{n1}, a_{n2}, \ldots, a_{nk},$$

or, more concisely,

$$a_{ij}, \text{ for } i = 1, 2, \ldots, n \quad \text{and} \quad j = 1, 2, \ldots, k.$$

In this case, for the sum of these numbers we use the abbreviation

$$\sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij}.$$

When both n, k are finite, the order in which the two sums above appear is irrelevant; that is

$$\sum_{i=1}^{n} \sum_{j=1}^{k} a_{ij} = \sum_{j=1}^{k} \sum_{i=1}^{n} a_{ij}.$$

For infinite sums, however, this is not always true, but certain conditions are needed. For the purposes of the present book, we may safely assume that these conditions are met so that we can interchange the order of summation without further reference.

To make the "double sum" notation more transparent and accessible to the reader, we give a couple of examples illustrating its use. First,

$$\sum_{i=1}^{n}\sum_{j=1}^{k} w^{i+j} = (w^{1+1} + w^{1+2} + \cdots + w^{1+k}) + (w^{2+1} + w^{2+2} + \ldots + w^{2+k}) + \cdots$$
$$+ (w^{n+1} + w^{n+2} + \cdots + w^{n+k}),$$

while

$$\begin{aligned}\sum_{i=1}^{n}\sum_{j=1}^{k}(i+j)a_{ij} =& [(1+1)a_{11} + (1+2)a_{12} + \cdots + (1+k)a_{1k}]\\ &+ [(2+1)a_{21} + (2+2)a_{22} + \cdots + (2+k)a_{2k}]\\ &+ \cdots\\ &+ [(n+1)a_{n1} + (n+2)a_{n2} + \cdots + (n+k)a_{nk}].\end{aligned}$$

Particular care is needed in cases where the range in the inner summation depends on the variable used in the outer summation. For example, consider the double sum

$$S = \sum_{i=1}^{n}\sum_{j=1}^{i} a_{ij}, \tag{A.5}$$

which can be written explicitly as

$$S = a_{11} + (a_{21} + a_{22}) + (a_{31} + a_{32} + a_{33}) + \cdots + (a_{n1} + a_{n2} + \cdots + a_{nn}).$$

Rearranging the order of the terms in this summation, we have the equivalent expression

$$S = (a_{11} + a_{21} + a_{31} + \cdots + a_{n1}) + (a_{22} + a_{32} + \cdots + a_{n2}) + \cdots + a_{nn},$$

or, with the aid of the summation symbol,

$$S = \sum_{j=1}^{n}\sum_{i=j}^{n} a_{ij},$$

so that this must be equal to the expression on the right of (A.5).

An easy way to make the change of variable in such cases is the following: we write carefully the range of the variables in the initial summation. Then

(i) we find the range of values for the variable that occurs in the outer summation in the new sum;

(ii) once the limits for the variable in the outer summation have been fixed, we calculate the range of values for the variable in the inner summation.

As an example, we return to the calculation of the sum

$$S = \sum_{i=1}^{n} \sum_{j=1}^{i} a_{ij}.$$

It is evident that the restrictions for the two variables are

$$1 \leq i \leq n \quad \text{and} \quad 1 \leq j \leq i,$$

which can be written in a combined way as

$$1 \leq j \leq i \leq n. \tag{A.6}$$

Since we want to reverse the order of the variables in the summation, our new sum must have the form

$$S = \sum_{j} \sum_{i} a_{ij}.$$

From (A.6) we see that j, which is the variable in the outside sum above, can take the values $1, 2, \ldots, n$. For any (fixed) value of j in that range, because of (A.6) we see that the admissible values for i are $j, j+1, \ldots, n$. We therefore get that

$$S = \sum_{j=1}^{n} \sum_{i=j}^{n} a_{ij}.$$

In view of the above, we have established the truth of the following identity

$$\sum_{i=1}^{n} \sum_{j=1}^{i} a_{ij} = \sum_{j=1}^{n} \sum_{i=j}^{n} a_{ij}.$$

Before we close this section, we note that a similar notation with sums can be used for products, replacing the symbol $\sum$ by $\prod$. For example, let $a_1, a_2, \ldots, a_n$ be real numbers. Then their product $a_1 a_2 \cdots a_n$ can be written as

$$\prod_{i=1}^{n} a_i$$

and we say that this is "the product of a_i for $i = 1$ to n." When no confusion arises about the range of values for the index i, we can write simply $\prod a_i$.

For a double-indexed array a_{ij}, we write

$$\prod_{i=1}^{n} \prod_{j=1}^{n} a_{ij}$$

for the product of all a_{ij} for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, n$.

The following proposition lists some properties for products, analogous to those we have seen earlier for sums. The proof of these properties is left as an exercise for the reader.

Proposition A.3 *Let $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_n, \lambda$ be real numbers, while n, k are positive integers. Then the following are true:*

$$(i)\ \prod_{i=1}^{n} a_i b_i = \left(\prod_{i=1}^{n} a_i\right)\left(\prod_{b=1}^{n} b_j\right), \qquad (ii)\ \prod_{i=1}^{n} (\lambda a_i) = \lambda^n \prod_{i=1}^{n} a_i,$$

$$(iii)\ \prod_{i=1}^{n} \left(\frac{a_i}{b_i}\right) = \frac{\prod_{i=1}^{n} a_i}{\prod_{i=1}^{n} b_i}\ (\textit{for } b_i \neq 0,\ i = 1, 2, \dots, n), \qquad (iv)\ \prod_{i=1}^{n} a = a^n,$$

$$(v)\ \prod_{i=1}^{n} \prod_{j=1}^{k} a_{ij} = \prod_{j=1}^{k} \prod_{i=1}^{n} a_{ij}, \qquad (vi)\ \prod_{i=1}^{n} \prod_{j=1}^{i} a_{ij} = \prod_{j=1}^{n} \prod_{i=j}^{n} a_{ij}.$$

EXERCISES

1. Calculate the value for each of the following sums:

(i) $\displaystyle\sum_{i=0}^{8} (3i+2) - \sum_{i=2}^{8} (2i+4)$

(ii) $\displaystyle\sum_{i=3}^{8} (2i+1) - \sum_{i=2}^{5} (i^2-2)$

(iii) $\displaystyle\sum_{i=1}^{10} \frac{2}{3^i} + \sum_{i=2}^{10} \frac{4}{6^{i+1}}$

(iv) $\displaystyle\sum_{i=1}^{50} (-1)^{i-1} i$

(v) $\displaystyle\sum_{i=1}^{10} i(10-i) - 10\sum_{i=1}^{10} i - 10\sum_{i=1}^{10} i^2.$

2. Given two sequences $(a_i), (b_i)$ of real numbers, simplify the following expressions by writing down explicitly each sum which appears:

(i) $\displaystyle\sum_{i=0}^{6} (i+1)a_i - \sum_{i=0}^{5} (i+2)a_i$

(ii) $\displaystyle\sum_{i=0}^{8} a_i b_{8-i} - \sum_{i=0}^{8} a_{8-i} b_i$

(iii) $\displaystyle\sum_{i=0}^{10} a_i(b_i+1) - \sum_{i=0}^{9} (a_i+1)b_i + \sum_{i=1}^{9} (b_i+1) - \sum_{i=1}^{10} (a_i+2)$

(iv) $\displaystyle\left(\sum_{i=1}^{3} a_i\right)\left(\sum_{i=1}^{3} b_i\right) - \sum_{i=1}^{3} a_i b_i$

(v) $\displaystyle\sum_{i=3}^{10} i^2 a_i - \sum_{i=1}^{7} (i+2)^2 a_i.$

3. Write each of the following expressions in a concise form with the aid of the $\sum$ symbol:

   (i) $(a_1 - b_1) + (a_2 - b_2) + \cdots + (a_n - b_n)$
   (ii) $1^3 + 2^3 + \cdots + n^3$
   (iii) $a_1 + 2a_2 + 3a_3 + \cdots + na_n$
   (iv) $1 \cdot 2 + 2 \cdot 3 + \cdots + n \cdot (n+1)$
   (v) $1 \cdot (n-1) + 2 \cdot (n-2) + \cdots + k \cdot (n-k)$ (assuming $1 \leq k \leq n$)
   (vi) $a_0 b_n + a_1 b_{n-1} + \cdots + a_n b_0$.

4. Let

$$S = \sum_{i=1}^{n} a_i, \quad T = \sum_{i=1}^{n} b_i.$$

   Express each of the following sums in terms of S and T:

   (i) $\sum_{i=1}^{n}(3a_i + 7)$
   (ii) $\sum_{i=1}^{n}(2 - 4b_i)$
   (iii) $\sum_{i=1}^{n}(-3a_i - 4b_i)$
   (iv) $\sum_{i=1}^{n}\left(4a_i - \frac{b_i}{9}\right)$.

5. Write each of the following expressions in an equivalent form such that the $\sum$ symbol appears only once in each case (assume that $1 \leq k < n$):

   (i) $\sum_{i=1}^{n} a_i + 2\sum_{i=1}^{n} b_i$
   (ii) $\sum_{i=1}^{n}(2a_i) - \frac{1}{5}\sum_{i=1}^{n} b_i$
   (iii) $\sum_{i=1}^{k}(6a_i) + 6\sum_{i=k+1}^{n} a_i$
   (iv) $\sum_{i=1}^{n} a_i - \sum_{i=1}^{k-1} a_i$

6. Assume that

$$S = 1 + 2 + \cdots + n = \sum_{i=1}^{n} i$$

   is the sum of the first n positive integers.

(i) Verify that

$$\sum_{i=0}^{n}(n-i) = S.$$

(ii) Adding side-by-side each of these two expressions for S, prove that

$$S = \frac{n(n+1)}{2}.$$

(iii) Show, more generally, that the sum of the first n terms of an arithmetic progression whose first term is a and the common difference between two consecutive terms is d is given by

$$\sum_{i=1}^{n}[a+(i-1)d] = \frac{n[2a+(n-1)d]}{2}.$$

7. Find the value of x if it is known that

$$\sum_{i=1}^{n}(a_i - x) = 0.$$

8. Show that for any real numbers a_i, b_i, for $i = 1, 2, \dots, n$, the following identity holds:

$$\sum_{i=1}^{n}(a_i+b_i)^2 = \sum_{i=1}^{n}a_i^2 + 2\sum_{i=1}^{n}a_ib_i + \sum_{i=1}^{n}b_i^2.$$

9. Let $a_1, a_2, \dots, a_n$ be real numbers and

$$\overline{a} = \frac{a_1 + a_2 + \cdots + a_n}{n}$$

be their arithmetic mean. Establish that for any $x \in \mathbb{R}$,

$$\sum_{i=1}^{n}(a_i - x)^2 = \sum_{i=1}^{n}(a_i - \overline{a})^2 + n(\overline{a} - x)^2.$$

Hence, find the value of x that minimizes the sum

$$\sum_{i=1}^{n}(a_i - x)^2.$$

10. Let $a_1, a_2, \dots, a_n$ be real numbers and denote by $\overline{a}$ their arithmetic mean. Put

$$b_i = ca_i + d, \quad i = 1, 2, \dots, n,$$

where $c, d \in \mathbb{R}$. Show that

$$\overline{b} = c\overline{a} + d,$$

where $\overline{b}$ is the arithmetic mean of b_i, and

$$\sum_{i=1}^{n} (b_i - \overline{b})^2 = c^2 \sum_{i=1}^{n} (a_i - \overline{a})^2.$$

11. Given $a_i, b_i \in \mathbb{R}$ for $i = 1, 2, \ldots, n$, define the following sums:

$$S_1 = \sum_{i=1}^{n} a_i, \quad S_2 = \sum_{i=1}^{n} a_i^2, \quad T_1 = \sum_{i=1}^{n} b_i, \quad T_2 = \sum_{i=1}^{n} b_i^2$$

and

$$R = \sum_{i=1}^{n} a_i b_i.$$

Express each of the sums below in terms of S_1, S_2, T_1, T_2, and R:

(i) $\sum_{i=1}^{n} (2a_i - 3)^2$

(ii) $\sum_{i=1}^{n} (4a_i - 2)^2$

(iii) $\sum_{i=1}^{n} (6a_i^2 - 3)$

(iv) $\sum_{i=1}^{n} (a_i^2 + 3b_i^2)$

(v) $\sum_{i=1}^{n} (3a_i - 2b_i)^2$

12. Investigate whether each of the following relations is generally true or not (here a_i, b_i are real numbers with $b_i \neq 0$ for $i = 1, 2, \ldots, n$ and n is a positive integer):

(i) $\left(\sum_{i=1}^{n} a_i\right)^2 = \sum_{i=1}^{n} a_i^2$

(ii) $\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{i=1}^{n} b_i\right) = \sum_{i=1}^{n} a_i b_i$

(iii) $\dfrac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} = \sum_{i=1}^{n} \dfrac{a_i}{b_i}.$

(*Hint*: You may check first whether these relations are true for $n = 2$.)

13. Show that for any real numbers a_i $(i = 1, 2, \dots, n)$ and b_j $(j = 1, 2, \dots, k)$, the following holds

$$\left(\sum_{i=1}^{n} a_i\right)\left(\sum_{j=1}^{k} b_j\right) = \sum_{i=1}^{n}\sum_{j=1}^{k} a_i b_j.$$

Application: Calculate the double sum

$$\sum_{i=1}^{n}\sum_{j=1}^{k} x^i y^j.$$

What is the value of the infinite (double) sum

$$\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} x^i y^j$$

for $|x| < 1$ and $|y| < 1$?

14. Calculate each of the following sums; here, a and b are two given real numbers:

(i) $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n}(3i + 5j)$

(ii) $\displaystyle\sum_{i=1}^{n}\sum_{j=1}^{n} a$

(iii) $\displaystyle\sum_{i=1}^{n}\frac{1}{i}\sum_{j=1}^{i}(a + 2bj)$

(iv) $\displaystyle\sum_{j=1}^{n}\sum_{i=j}^{n}\frac{a + 2bj}{i}.$

(*Hint*: Use the results from Exercise 6.)

15. Write each of the expressions below using the $\prod$ symbol:

(i) $c_k c_{k+1} \cdots c_n$

(ii) $\dfrac{a_2}{b_2} \cdot \dfrac{a_3}{b_3} \cdots \dfrac{a_n}{b_n}$

(iii) $1 \cdot 2 \cdots n$

(iv) $a(a + b)(a + 2b) \cdots (a + (n - 1)b)$

(v) $a(a + b)\left(a + \dfrac{b}{2}\right) \cdots \left(a + \dfrac{b}{n - 1}\right)$

(vi) $s_{11}s_{22} \cdots s_{nn}$

16. Let $a > 0$, $b > 0$ and n be a positive integer. Calculate each of the products below:

(i) $\displaystyle\prod_{i=1}^{n}(ab^i)$

(ii) $\displaystyle\prod_{i=0}^{n} b^{3i+5}$

(iii) $\dfrac{\prod_{i=1}^{n} a^{2i}}{\prod_{i=1}^{n} a^{i+1}}$

(iv) $\prod_{i=1}^{n}(a^{2i}b^{3i})$

(v) $\prod_{i=1}^{n}\dfrac{a^{2i}}{b^{4i}}$

(vi) $\dfrac{\prod_{i=1}^{n} a_i}{\prod_{i=1}^{n} b_i}$, given that $a_i = \lambda b_i,\ i = 1, 2, \ldots, n$, for some $\lambda \neq 0$.

(*Hint*: You may find the results of Exercise 6 useful.)

USEFUL FORMULAS

1. $\lim_{x\to\infty}\left(1 + \dfrac{a}{x}\right)^x = \mathrm{e}^a, \quad -\infty < a < \infty$

2. $\sum_{i=0}^{\infty}\dfrac{x^i}{i!} = \mathrm{e}^x, \quad -\infty < x < \infty$

3. $\sum_{i=0}^{\infty}\dfrac{x^i}{(i+1)!} = \dfrac{\mathrm{e}^x - 1}{x}, \quad -\infty < x < \infty$

4. $\sum_{i=1}^{\infty}\dfrac{x^{2i-1}}{(2i-1)!} = \dfrac{\mathrm{e}^x - \mathrm{e}^{-x}}{2}, \quad -\infty < x < \infty$

5. $\sum_{i=1}^{\infty}\dfrac{x^i}{i} = \ln\left(\dfrac{1}{1-x}\right) = -\ln(1-x), \quad |x| < 1$

6. $\sum_{i=0}^{n}\binom{n}{i}a^i b^{n-i} = (a+b)^n, \quad -\infty < a, b < \infty$

7. $\sum_{i=0}^{\infty}\binom{n+i-1}{i}x^i = \dfrac{1}{(1-x)^n}, \quad |x| < 1$

8. $\sum_{i=0}^{k}x^i = \dfrac{1-x^{k+1}}{1-x}, \quad x \neq 1$

9. $\sum_{i=0}^{\infty}x^i = \dfrac{1}{1-x}, \quad |x| < 1$

10. $\sum_{i=1}^{k}ix^i = x\,\dfrac{(1-x^k) - kx^k(1-x)}{(1-x)^2}, \quad x \neq 1$

11. $\sum_{i=0}^{\infty}\dfrac{(i+n)!}{i!}x^i = \dfrac{n!}{(1-x)^{n+1}}, \quad |x| < 1 \text{ and } n \geq 0$

12. $\sum_{i=1}^{\infty} i x^i = \frac{x}{(1-x)^2}, \quad \sum_{i=1}^{\infty} i^2 x^i = \frac{x(1+x)}{(1-x)^3}, \quad \sum_{i=1}^{\infty} i(i+1) x^i = \frac{2x}{(1-x)^3}, \quad |x| < 1$

(these arise from 11. for $n = 1, 2$)

13. $\sum_{i=0}^{\infty} \binom{n+i-1}{i} x^{-i} = \left(\frac{x}{x-1}\right)^n, \quad |x| < 1$ and $n \geq 1$

14. $\sum_{i=1}^{k} i = \frac{k(k+1)}{2}$

15. $\sum_{i=1}^{k} i^2 = \frac{k(k+1)(2k+1)}{6}$

16. $\int_0^t x\mathrm{e}^{-\lambda x}\mathrm{d}x = -\frac{1+\lambda t}{\lambda^2}\mathrm{e}^{-\lambda t} + \frac{1}{\lambda^2}, \quad \lambda > 0$

17. $\int_0^t x^2\mathrm{e}^{-\lambda x}\mathrm{d}x = -\left[\frac{t^2}{\lambda} + \frac{2t}{\lambda^2} + \frac{2}{\lambda^3}\right]\mathrm{e}^{-\lambda t} + \frac{2}{\lambda^3}, \quad \lambda > 0$

18. $\int_0^\infty t^{\alpha-1}\mathrm{e}^{-\lambda t}\mathrm{d}t = \frac{\Gamma(\alpha)}{\lambda^\alpha}, \quad \alpha > 0, \lambda > 0$

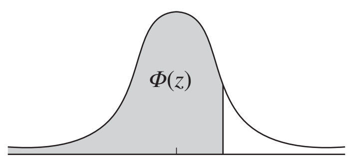
19. $\int_{-\infty}^\infty \mathrm{e}^{-x^2/2}\mathrm{d}x = 2\int_0^\infty \mathrm{e}^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}$

20. $\int_0^\infty x^{\alpha-1}\exp(-\lambda x^\beta)\mathrm{d}x = \frac{\Gamma(\alpha/\beta)}{\beta\lambda^{\alpha/\beta}}, \quad \alpha > 0, \beta > 0$

21. $\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\mathrm{d}x = B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}, \quad \alpha > 0, \beta > 0.$

APPENDIX B

DISTRIBUTION FUNCTION OF THE STANDARD NORMAL DISTRIBUTION



$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx$$

$$\Phi(-z) = 1 - \Phi(z)$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.70.995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9889	0.9889	0.9890	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998

APPENDIX C

SIMULATION

The whole essence of probability is randomness. Of course uncertainty, and therefore randomness, is omnipresent in nature including all sorts of human activities. However, in the study of natural phenomena and human activities, it is often very useful to be able to *produce randomness* at a lab or, nowadays, at work or even at home using a personal computer. This is achieved by simulation. Byron Morgan writes in the preface of his book *Elements of Simulation* (1984) “The use of simulation in statistics dates from the start of the 20th century, coinciding with the beginnings of radio broadcasting and the advent of television. Just as radio and television are now commonplace in our everyday lives, simulation methods are now widely used throughout the many branches of statistics… ” Moreover, the use and applications of simulation (not only in probability and statistics) have exploded in recent decades primarily due to the ever growing capability of modern computers. In brief, simulation offers a source of randomness by imitating the realization of a random experiment, such as the throw of a die or the future growth of a company’s sales. However, one has to bear in mind that computers are unable to produce truly random realizations, so that computer results are essentially deterministic but they possess the properties of the model we have developed for a random experiment.

A simple such model can be a probability distribution that is used to mimic a real mechanism that produces values of some quantity of interest. For instance, we may assume that the measurement errors of an instrument that weighs the amount of a chemical

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

substance are normally distributed. But what is a "*typical value*" of such an error? The most common use of simulation in probability and statistics is when one wants to generate values from a given probability distribution. There are various ways to accomplish this; however, in the present context, we shall avoid going into the technicalities of this issue. Below we merely give a simple example for the outcome of a throw of a die to illustrate the basic ideas of a simulation experiment, while for more details the interested reader is referred to one of the many excellent books on the subject, including the book by B. Morgan mentioned above.

Suppose we want to calculate the probability that, in the throw of three dice, the sum of outcomes equals 10. A simple simulation algorithm can be based on the following three steps:

1. Produce n triples of outcomes (X, Y, Z) such that $X \in \{1, 2, 3, 4, 5, 6\}$, $Y \in \{1, 2, 3, 4, 5, 6\}$, $Z \in \{1, 2, 3, 4, 5, 6\}$ and
$$P(X = k) = P(Y = k) = P(Z = k) = \frac{1}{6}, \quad k = 1, 2, 3, 4, 5, 6;$$
2. Count the number n_A out of the n triples from Step 1 that satisfy
$$X + Y + Z = 10;$$
3. Find the ratio
$$f_A = \frac{n_A}{n}$$
which, for sufficiently large n, approximates the true probability $P(A)$.

To implement the first of these three steps, we need a mechanism that **produces random integers** chosen from the set $\{1, 2, 3, 4, 5, 6\}$. A variety of mathematical and statistical software has ready-to-use commands for such a purpose. For instance, the Mathematica command
$$X = \text{Random[Integer}, \{m, n\}]$$
returns a randomly chosen integer X from the set $A = \{m, m + 1, m + 2, \ldots, n\}$ so that X takes any value from the set A with the same probability $1/(n - m + 1)$. Consequently, the command
$$X = \text{Random[Integer}, \{1, 6\}]$$
can be used to "simulate a die outcome in a single throw". In Maple, which is another computer algebra package, the corresponding command to select an integer from the set A is
$$X := \text{rand}(m, n),$$
while the command
$$X := \text{Die}(k, n)$$
produces the outcomes of n throws from a "die with k faces (so that each side has a probability $1/k$ to turn up).

If the software we are using does not have a specific command for selecting *integer random numbers* from a given set, we can implement this task easily using **random numbers selected from the interval** $[0, 1]$. A command to obtain such numbers exists in nearly every programming language (Fortran, Basic, C, Pascal, etc.) and has usually one of the following forms

$$\text{RND( ),}\quad \text{RAN( ),}\quad \text{RANDOM( ).}$$

Any such command produces random numbers from the interval $[0, 1]$, under the assumption that the probability to generate a number Y which belongs to an interval $[a, b]$ (or $(a, b]$, or $[a, b)$, or (a, b)), for $0 \le a \le b \le 1$, is $b - a$; that is

$$P(a < Y < b) = P(a \le Y < b) = P(a < Y \le b) = P(a \le Y \le b) = b - a.$$

With the terminology of Section 7.1, the distribution of Y generated in this way is the uniform distribution over the unit interval.

Once we have obtained a collection of (real) random numbers in $[0, 1]$, it is very easy to transform these into a collection of integers from the set $\{m, m + 1, \dots, n\}$. For instance:

- Suppose we want to simulate random outcomes from a series of coin tosses. Let 0 denote a "tails" outcome and 1 represent "heads" in a throw. The following simple algorithm is then used:

  Step 1 Generate a random number Y from the interval $[0, 1]$.

  Step 2 If $Y < 0.5$, put $X = 0$ (tails). Otherwise $X = 1$ (heads).

- Consider now the experiment of throwing a die, so that we want to generate integers from the set $A = \{1, 2, 3, 4, 5, 6\}$. The two-step procedure for this case is

  Step 1 Generate a random number Y from the interval $[0, 1]$.

  Step 2 If $Y < 1/6$, put $X = 1$.
  If $1/6 \le Y < 2/6$, put $X = 2$.
  If $2/6 \le Y < 3/6$, put $X = 3$.
  If $3/6 \le Y < 4/6$, put $X = 4$.
  If $4/6 \le Y < 5/6$, put $X = 5$.
  If $5/6 \le Y$, put $X = 6$.

It has to be noted that, although using simulation we can take satisfactory approximations to a wide range of difficult mathematical problems, this cannot serve as an alternative to the formal mathematical proof. Simulation is particularly useful in cases where a precise modeling, in mathematical terms, of a real-life situation is either unfeasible or it is possible but presents formidable theoretical difficulties.

APPENDIX D

DISCRETE AND CONTINUOUS DISTRIBUTIONS

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

Discrete distributions

Distribution	Parameters	Probability or density function	Range of values	$E(X)$	$V(X)$
Binomial $b(n,p)$	n is a positive integer $0<p<1$	$\binom{n}{x} p^x(1-p)^{n-x}$	$x=0,1,2,\dots,n$	np	$np(1-p)$
Poisson $\mathcal{P}(\lambda)$	$\lambda>0$	$e^{-\lambda}\frac{\lambda^x}{x!}$	$x=0,1,2,\dots$	λ	λ
Geometric $G(p)$	$0<p<1$	$p(1-p)^{x-1}$	$x=1,2,\dots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative binomial $Nb(r,p)$	r is a positive integer $0<p<1$	$\binom{x-1}{r-1} p^r(1-p)^{x-r}$	$x=r,r+1,\dots$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$
Hypergeometric $h(n,a,b)$	a,b,n are positive integers $n\le a+b$	$\frac{\binom{a}{x}\binom{b}{n-x}}{\binom{a+b}{n}}$	$x=0,1,2,\dots,n$ $(\max\{0,n-b\})\le$ $\le x\le\min\{n,a\})$	$\frac{na}{a+b}$	$\frac{nab}{(a+b)^2}\times$ $\left(1-\frac{n-1}{a+b-1}\right)$

Continuous distributions					
Uniform $\mathcal{U}[a,b]$	$a, b \in \mathbb{R}, a < b$	$\frac{1}{b-a}$	$a \le x \le b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Normal $N(\mu, \sigma^2)$	$\mu \in \mathbb{R},\ \ \sigma^2 > 0$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$-\infty < x < \infty$	μ	σ^2
Exponential $\mathcal{E}(\lambda)$	$\lambda > 0$	$\lambda e^{-\lambda x}$	$0 \le x < \infty$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Gamma (a, λ)	$a > 0, \lambda > 0$	$\frac{\lambda^a}{\Gamma(a)} x^{a-1} e^{-\lambda x}$	$0 \le x < \infty$	$\frac{a}{\lambda}$	$\frac{a}{\lambda^2}$
Beta (α, β)	$\alpha > 0, \beta > 0$	$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$	$0 \le x \le 1$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta+1)(\alpha+\beta)^2}$

BIBLIOGRAPHY

Bassett, E.E., Bremner, J.M., Jolliffe, I.T. et al. (2000). *Statistics: Problems and Solutions*. Singapore: World Scientific.

Bean, M.A. (2001). *Probability: The Science of Uncertainty, Brooks/Cole Series in Advanced Mathematics*. Pacific Grove.

Blom, G., Holst, S., and Sandell, D. (1994). *Problems and Snapshots from the World of Probability*. New York: Springer.

Devore, J.L. (2012). *Probability and Statistics for Engineering and the Sciences*, 8e. Brooks Cole/Cengage Learning.

Dorogovtsev, A.Y., Silvestrov, D.S., Skorohod, A.V., and Yadrenko, M.I. (1997). *Probability Theory: Collection of Problems, Translations of Mathematical Monographs*. Providence, RI: American Mathematical Society.

Durret, R. (2009). *Elementary Probability for Applications*. Cambridge, NY: Cambridge University Press.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, 3e. New York: Wiley.

Ghahramani, S. (2005). *Fundamentals of Probability with Stochastic Processes*, 3e. London: Pearson Education Ltd.

Grimmett, G. and Stirzaker, D.R. (2001a). *One Thousand Exercises in Probability*. Oxford: Oxford University Press.

Grimmett, G. and Stirzaker, D.R. (2001b). *Probability and Random Processes*, 3e. Oxford: Oxford University Press.

Grimmett, G. and Welsh, D. (1986). *Probability: An introduction*. Oxford: Clarendon Press.

Hald, A. (1998). *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.

Pitman, J. (1993). *Probability*. New York: Springer.

Ross, S.M. (2006). *A First Course in Probability*, 7e. London, New Jersey: Pearson Prentice Hall (International Edition).

Ross, S.M. (2007). *An Introduction to Probability Models*, 9e. Academic Press.

Shah, A. (1985). A Simpler Approximation for Areas under the Standard Normal Curve. *The American Statistician*, Vol. 39, p. 80.

Sobkowicz, P., Thelwall, M., Buckley, K. et al. (2013). Lognormal distributions of user post lengths in Internet discussions - a consequence of the Weber-Fechner law? *EPJ Data Science* 2: 2.

Weiss, N.A. (2006). *A Course in Probability*. Pearson Education.

Other non-technical books

Gardner, M. (2008). *Hexaflexagons, Probability Paradoxes, and the Tower of Hanoi: Martin Gardner's First Book of Mathematical Puzzles and Games*. Cambridge University Press.

Gardner, M. (2009). *Aha!: Aha! Insight and Aha! Gotcha, Spectrum Series*. The Mathematical Association of America.

Weaver, W. (1977). *Lady Luck: The Theory of Probability*. New York: Penguin Books.

Web sources

Earliest Known Uses of Some of the Words of Mathematics http://jeff560.tripod.com/ (Jeff Miller web pages)

www.wikipedia.org.

INDEX

Introduction to Probability: Models and Applications, First Edition.
N. Balakrishnan, Markos V. Koutras, and Konstadinos G. Politis.