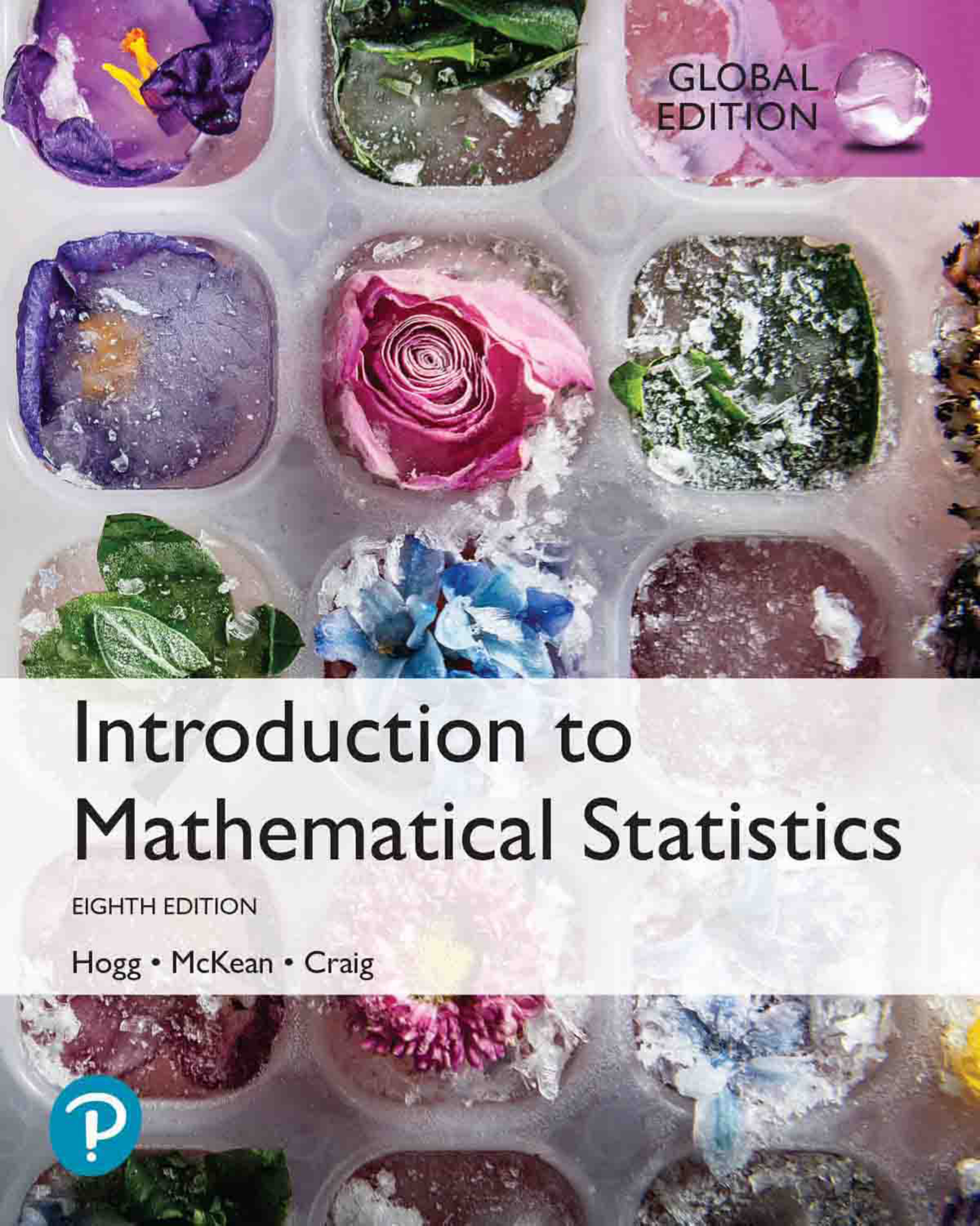

GLOBAL EDITION

Introduction to Mathematical Statistics

EIGHTH EDITION

Hogg • McKean • Craig

Introduction to Mathematical Statistics

Eighth Edition

Global Edition

Robert V. Hogg
University of Iowa

Joseph W. McKean
Western Michigan University

Allen T. Craig
Late Professor of Statistics
University of Iowa

Pearson

Director, Portfolio Management: Deirdre Lynch
Courseware Portfolio Manager: Patrick Barbera
Portfolio Management Assistant: Morgan Danna
Associate Editor, Global Edition: Shaoni Mukherjee
Content Producer: Lauren Morse
Managing Producer: Scott Disanno
Product Marketing Manager: Yvonne Vannatta
Field Marketing Manager: Evan St. Cyr
Marketing Assistant: Jon Bryant
Senior Author Support/Technology Specialist: Joe Vetere
Managing Editor, Digital Media, Global Edition: Gargi Banerjee
Manager, Rights and Permissions: Gina Cheselka
Manufacturing Buyer: Carol Melville, LSC Communications
Manufacturing Buyer, Global Edition: Kay Holman
Art Director: Barbara Atkinson
Production Coordination and Illustrations: Integra
Cover Design: Lumina Datamatics
Cover Image: Anna Mente/Shutterstock

Pearson Education Limited
KAO Two
KAO Park
Hockham Way
Harlow
Essex
CM17 9SR
United Kingdom

and Associated Companies throughout the world

Visit us on the World Wide Web at: www.pearsonglobaleditions.com

Authorized adaptation from the United States edition, entitled Introduction to Mathematical Statistics, *8th Edition, ISBN 978-0-13-468699-8, by Robert V. Hogg, Joseph W. McKean, and Allen T. Craig, published by Pearson Education © 2019.*

ISBN 10: 1-292-26476-4

ISBN 13: 978-1-292-26476-9

eBook ISBN 13: 978-1-292-26478-3

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Typeset by Integra Software Services Pvt. Ltd.

Dedicated to my wife Marge
and to the memory of Bob Hogg

This page intentionally left blank

Contents

This page intentionally left blank

Preface

We have made substantial changes in this edition of *Introduction to Mathematical Statistics*. Some of these changes help students appreciate the connection between statistical theory and statistical practice while other changes enhance the development and discussion of the statistical theory presented in this book.

Many of the changes in this edition reflect comments made by our readers. One of these comments concerned the small number of real data sets in the previous editions. In this edition, we have included more real data sets, using them to illustrate statistical methods or to compare methods. Further, we have made these data sets accessible to students by including them in the free R package `hmcpkg`. They can also be individually downloaded in an R session at the url listed on page 12. In general, the R code for the analyses on these data sets is given in the text.

We have also expanded the use of the statistical software R. We selected R because it is a powerful statistical language that is free and runs on all three main platforms (Windows, Mac, and Linux). Instructors, though, can select another statistical package. We have also expanded our use of R functions to compute analyses and simulation studies, including several games. We have kept the level of coding for these functions straightforward. Our goal is to show students that with a few simple lines of code they can perform significant computations. Appendix B contains a brief R primer, which suffices for the understanding of the R used in the text. As with the data sets, these R functions can be sourced individually at the cited url; however, they are also included in the package `hmcpkg`.

We have supplemented the mathematical review material in Appendix A, placing it in the document *Mathematical Primer for Introduction to Mathematical Statistics*. It is freely available for students to download at the listed url. Besides sequences, this supplement reviews the topics of infinite series, differentiation, and integration (univariate and bivariate). We have also expanded the discussion of iterated integrals in the text. We have added figures to clarify discussion.

We have retained the order of elementary statistical inferences (Chapter 4) and asymptotic theory (Chapter 5). In Chapters 5 and 6, we have written brief reviews of the material in Chapter 4, so that Chapters 4 and 5 are essentially independent of one another and, hence, can be interchanged. In Chapter 3, we now begin the section on the multivariate normal distribution with a subsection on the bivariate normal distribution. Several important topics have been added. This includes Tukey's multiple comparison procedure in Chapter 9 and confidence intervals for the correlation coefficients found in Chapters 9 and 10. Chapter 7 now contains a discussion on standard errors for estimates obtained by bootstrapping the sample. Several topics that were discussed in the Exercises are now discussed in the text. Examples include quantiles, Section 1.7.1, and hazard functions, Section 3.3. In general, we have made more use of subsections to break up some of the discussion. Also, several more sections are now indicated by * as being optional.

Content and Course Planning

Chapters 1 and 2 develop probability models for univariate and multivariate variables while Chapter 3 discusses many of the most widely used probability models. Chapter 4 discusses statistical theory for much of the inference found in a standard statistical methods course. Chapter 5 presents asymptotic theory, concluding with the Central Limit Theorem. Chapter 6 provides a complete inference (estimation and testing) based on maximum likelihood theory. The EM algorithm is also discussed. Chapters 7–8 contain optimal estimation procedures and tests of statistical hypotheses. The final three chapters provide theory for three important topics in statistics. Chapter 9 contains inference for normal theory methods for basic analysis of variance, univariate regression, and correlation models. Chapter 10 presents nonparametric methods (estimation and testing) for location and univariate regression models. It also includes discussion on the robust concepts of efficiency, influence, and breakdown. Chapter 11 offers an introduction to Bayesian methods. This includes traditional Bayesian procedures as well as Markov Chain Monte Carlo techniques.

Several courses can be designed using our book. The basic two-semester course in mathematical statistics covers most of the material in Chapters 1–8 with topics selected from the remaining chapters. For such a course, the instructor would have the option of interchanging the order of Chapters 4 and 5, thus beginning the second semester with an introduction to statistical theory (Chapter 4). A one-semester course could consist of Chapters 1–4 with a selection of topics from Chapter 5. Under this option, the student sees much of the statistical theory for the methods discussed in a non-theoretical course in methods. On the other hand, as with the two-semester sequence, after covering Chapters 1–3, the instructor can elect to cover Chapter 5 and finish the course with a selection of topics from Chapter 4.

The data sets and R functions used in this book and the R package `hmcpkg` can be downloaded from this title's page at the site:
`www.pearsonglobaleditions.com`

Acknowledgments

Bob Hogg passed away in 2014, so he did not work on this edition of the book. Often, though, when I was trying to decide whether or not to make a change in the manuscript, I found myself thinking of what Bob would do. In his memory, I have retained the order of the authors for this edition.

As with earlier editions, comments from readers are always welcomed and appreciated. We would like to thank these reviewers of the previous edition: James Baldone, Virginia College; Steven Culpepper, University of Illinois at Urbana-Champaign; Yuichiro Kakihara, California State University; Jaechoul Lee, Boise State University; Michael Levine, Purdue University; Tingni Sun, University of Maryland, College Park; and Daniel Weiner, Boston University. We appreciated and took into consideration their comments for this revision. We appreciate the helpful comments of Thomas Hettmansperger of Penn State University, Ash Abebe of Auburn University, and Professor Ioannis Kalogridis of the University of Leuven. A special thanks to Patrick Barbera (Portfolio Manager, Statistics), Lauren Morse (Content Producer, Math/Stats), Yvonne Vannatta (Product Marketing Manager), and the rest of the staff at Pearson for their help in putting this edition together. Thanks also to Richard Ponticelli, North Shore Community College, who accuracy checked the page proofs. Also, a special thanks to my wife Marge for her unwavering support and encouragement of my efforts in writing this edition.

Joe McKean

Acknowledgments for the Global Edition

Pearson would like to thank and acknowledge the following people for their work on this Global Edition.

Contributors

Polina Dolmatova, American University of Central Asia
Tsung Fei Khang, University of Malaya
Eric A. L. LI, The University of Hong Kong
Choung Min Ng, University of Malaya
Neelesh S. Upadhye, Indian Institute of Technology Madras

Reviewers

Samprit Chakrabarti, ICFAI Business School Kolkata
Rubén Garcia Berasategui, Jakarta International College
Aneesh Kumar K., Mahatma Gandhi College, Iritty
Jesper Rydén, Swedish University of Agricultural Sciences
Pooja Sengupta, International Management Institute Kolkata
Neelesh S. Upadhye, Indian Institute of Technology Madras

This page intentionally left blank

Chapter 1

Probability and Distributions

1.1 Introduction

In this section, we intuitively discuss the concepts of a probability model which we formalize in Secton 1.3 Many kinds of investigations may be characterized in part by the fact that repeated experimentation, under essentially the same conditions, is more or less standard procedure. For instance, in medical research, interest may center on the effect of a drug that is to be administered; or an economist may be concerned with the prices of three specified commodities at various time intervals; or an agronomist may wish to study the effect that a chemical fertilizer has on the yield of a cereal grain. The only way in which an investigator can elicit information about any such phenomenon is to perform the experiment. Each experiment terminates with an *outcome*. But it is characteristic of these experiments that the outcome cannot be predicted with certainty prior to the experiment.

Suppose that we have such an experiment, but the experiment is of such a nature that a collection of every possible outcome can be described prior to its performance. If this kind of experiment can be repeated under the same conditions, it is called a **random experiment**, and the collection of every possible outcome is called the experimental space or the **sample space**. We denote the sample space by $\mathcal{C}$.

Example 1.1.1. In the toss of a coin, let the outcome tails be denoted by T and let the outcome heads be denoted by H. If we assume that the coin may be repeatedly tossed under the same conditions, then the toss of this coin is an example of a random experiment in which the outcome is one of the two symbols T or H; that is, the sample space is the collection of these two symbols. For this example, then, $\mathcal{C} = \{H, T\}$. ■

Example 1.1.2. In the cast of one red die and one white die, let the outcome be the ordered pair (number of spots up on the red die, number of spots up on the white die). If we assume that these two dice may be repeatedly cast under the same conditions, then the cast of this pair of dice is a random experiment. The sample space consists of the 36 ordered pairs: $\mathcal{C} = \{(1,1), \ldots, (1,6), (2,1), \ldots, (2,6), \ldots, (6,6)\}$. ■

We generally use small Roman letters for the elements of $\mathcal{C}$ such as a, b, or c. Often for an experiment, we are interested in the chances of certain subsets of elements of the sample space occurring. Subsets of $\mathcal{C}$ are often called **events** and are generally denoted by capitol Roman letters such as $A, B,$ or C. If the experiment results in an element in an event A, we say the event A has occurred. We are interested in the chances that an event occurs. For instance, in Example 1.1.1 we may be interested in the chances of getting heads; i.e., the chances of the event $A = \{H\}$ occurring. In the second example, we may be interested in the occurrence of the sum of the upfaces of the dice being "7" or "11;" that is, in the occurrence of the event $A = \{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1),(5,6),(6,5)\}$.

Now conceive of our having made N repeated performances of the random experiment. Then we can count the number f of times (the **frequency**) that the event A actually occurred throughout the N performances. The ratio f/N is called the **relative frequency** of the event A in these N experiments. A relative frequency is usually quite erratic for small values of N, as you can discover by tossing a coin. But as N increases, experience indicates that we associate with the event A a number, say p, that is equal or approximately equal to that number about which the relative frequency seems to stabilize. If we do this, then the number p can be interpreted as that number which, in future performances of the experiment, the relative frequency of the event A will either equal or approximate. Thus, although we *cannot* predict the outcome of a random experiment, we *can*, for a large value of N, predict approximately the relative frequency with which the outcome will be in A. The number p associated with the event A is given various names. Sometimes it is called the *probability* that the outcome of the random experiment is in A; sometimes it is called the *probability* of the event A; and sometimes it is called the *probability measure* of A. The context usually suggests an appropriate choice of terminology.

Example 1.1.3. Let $\mathcal{C}$ denote the sample space of Example 1.1.2 and let B be the collection of every ordered pair of $\mathcal{C}$ for which the sum of the pair is equal to seven. Thus $B = \{(1,6),(2,5),(3,4),(4,3),(5,2)(6,1)\}$. Suppose that the dice are cast $N = 400$ times and let f denote the frequency of a sum of seven. Suppose that 400 casts result in $f = 60$. Then the relative frequency with which the outcome was in B is $f/N = \frac{60}{400} = 0.15$. Thus we might associate with B a number p that is close to 0.15, and p would be called the probability of the event B. ∎

Remark 1.1.1. The preceding interpretation of probability is sometimes referred to as the *relative frequency approach*, and it obviously depends upon the fact that an experiment can be repeated under essentially identical conditions. However, many persons extend probability to other situations by treating it as a rational measure of belief. For example, the statement $p = \frac{2}{5}$ for an event A would mean to them that their *personal* or *subjective* probability of the event A is equal to $\frac{2}{5}$. Hence, if they are not opposed to gambling, this could be interpreted as a willingness on their part to bet on the outcome of A so that the two possible payoffs are in the ratio $p/(1-p) = \frac{2}{5}/\frac{3}{5} = \frac{2}{3}$. Moreover, if they truly believe that $p = \frac{2}{5}$ is correct, they would be willing to accept either side of the bet: (a) win 3 units if A occurs and lose 2 if it does not occur, or (b) win 2 units if A does not occur and lose 3 if it does. However, since the mathematical properties of probability given in Section 1.3 are consistent with either of these interpretations, the subsequent mathematical development does not depend upon which approach is used. ■

The primary purpose of having a mathematical theory of statistics is to provide mathematical models for random experiments. Once a model for such an experiment has been provided and the theory worked out in detail, the statistician may, within this framework, make inferences (that is, draw conclusions) about the random experiment. The construction of such a model requires a theory of probability. One of the more logically satisfying theories of probability is that based on the concepts of sets and functions of sets. These concepts are introduced in Section 1.2.

1.2 Sets

The concept of a *set* or a *collection* of objects is usually left undefined. However, a particular set can be described so that there is no misunderstanding as to what collection of objects is under consideration. For example, the set of the first 10 positive integers is sufficiently well described to make clear that the numbers $\frac{3}{4}$ and 14 are not in the set, while the number 3 is in the set. If an object belongs to a set, it is said to be an *element* of the set. For example, if C denotes the set of real numbers x for which $0 \le x \le 1$, then $\frac{3}{4}$ is an element of the set C. The fact that $\frac{3}{4}$ is an element of the set C is indicated by writing $\frac{3}{4} \in C$. More generally, $c \in C$ means that c is an element of the set C.

The sets that concern us are frequently *sets of numbers*. However, the language of sets of *points* proves somewhat more convenient than that of sets of numbers. Accordingly, we briefly indicate how we use this terminology. In analytic geometry considerable emphasis is placed on the fact that to each point on a line (on which an origin and a unit point have been selected) there corresponds one and only one number, say x; and that to each number x there corresponds one and only one point on the line. This one-to-one correspondence between the numbers and points on a line enables us to speak, without misunderstanding, of the "point x" instead of the "number x." Furthermore, with a plane rectangular coordinate system and with x and y numbers, to each symbol (x, y) there corresponds one and only one point in the plane; and to each point in the plane there corresponds but one such symbol. Here again, we may speak of the "point (x, y)," meaning the "ordered number pair x and y." This convenient language can be used when we have a rectangular coordinate system in a space of three or more dimensions. Thus the "point $(x_1, x_2, \ldots, x_n)$" means the numbers $x_1, x_2, \ldots, x_n$ in the order stated. Accordingly, in describing our sets, we frequently speak of a set of points (a set whose elements are points), being careful, of course, to describe the set so as to avoid any ambiguity. The notation $C = \{x : 0 \le x \le 1\}$ is read "C is the one-dimensional set of points x for which $0 \le x \le 1$." Similarly, $C = \{(x, y) : 0 \le x \le 1, 0 \le y \le 1\}$ can be read "C is the two-dimensional set of points (x, y) that are interior to, or on the boundary of, a square with opposite vertices at $(0, 0)$ and $(1, 1)$."

We say a set C is **countable** if C is finite or has as many elements as there are positive integers. For example, the sets $C_1 = \{1, 2, \ldots, 100\}$ and $C_2 = \{1, 3, 5, 7, \ldots\}$

are countable sets. The interval of real numbers $(0, 1]$, though, is not countable.

1.2.1 Review of Set Theory

As in Section 1.1, let $\mathcal{C}$ denote the sample space for the experiment. Recall that events are subsets of $\mathcal{C}$. We use the words event and subset interchangeably in this section. An elementary algebra of sets will prove quite useful for our purposes. We now review this algebra below along with illustrative examples. For illustration, we also make use of **Venn diagrams**. Consider the collection of Venn diagrams in Figure 1.2.1. The interior of the rectangle in each plot represents the sample space C. The shaded region in Panel (a) represents the event A.

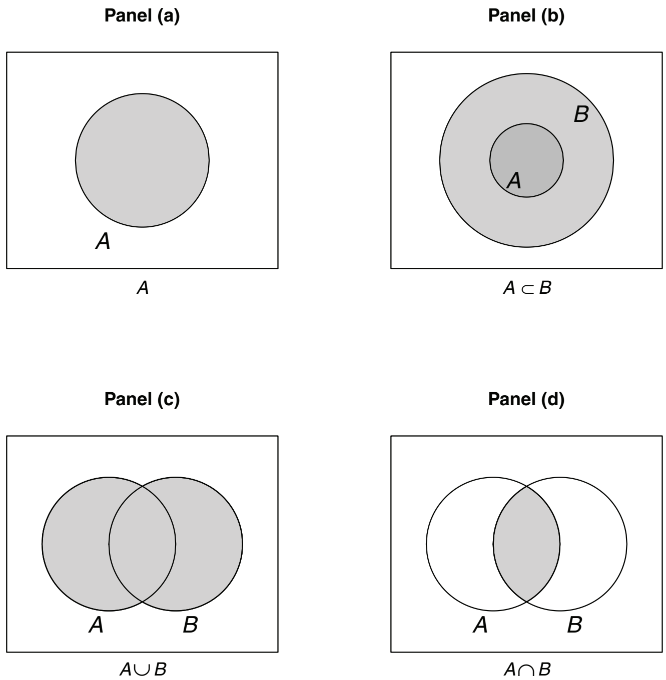

Figure 1.2.1: A series of Venn diagrams. The sample space C is represented by the interior of the rectangle in each plot. Panel (a) depicts the event A; Panel (b) depicts $A \subset B$; Panel (c) depicts $A \cup B$; and Panel (d) depicts $A \cap B$.

We first define the complement of an event A.

Definition 1.2.1. *The* **complement** *of an event A is the set of all elements in C which are not in A. We denote the complement of A by A^c. That is, $A^c = \{x \in \mathcal{C} : x \notin A\}$.*

The complement of A is represented by the white space in the Venn diagram in Panel (a) of Figure 1.2.1.

The empty set is the event with no elements in it. It is denoted by ϕ. Note that $\mathcal{C}^c = \phi$ and $\phi^c = \mathcal{C}$. The next definition defines when one event is a subset of another.

Definition 1.2.2. *If each element of a set A is also an element of set B, the set A is called a* **subset** *of the set B. This is indicated by writing $A \subset B$. If $A \subset B$ and also $B \subset A$, the two sets have the same elements, and this is indicated by writing $A = B$.*

Panel (b) of Figure 1.2.1 depicts $A \subset B$.

The event A or B is defined as follows:

Definition 1.2.3. *Let A and B be events. Then the* **union** *of A and B is the set of all elements that are in A or in B or in both A and B. The union of A and B is denoted by $A \cup B$*

Panel (c) of Figure 1.2.1 shows $A \cup B$.

The event that both A and B occur is defined by,

Definition 1.2.4. *Let A and B be events. Then the* **intersection** *of A and B is the set of all elements that are in both A and B. The intersection of A and B is denoted by $A \cap B$*

Panel (d) of Figure 1.2.1 illustrates $A \cap B$.

Two events are **disjoint** if they have no elements in common. More formally we define

Definition 1.2.5. *Let A and B be events. Then A and B are* **disjoint** *if $A \cap B = \phi$*

If A and B are disjoint, then we say $A \cup B$ forms a **disjoint union.** The next two examples illustrate these concepts.

Example 1.2.1. Suppose we have a spinner with the numbers 1 through 10 on it. The experiment is to spin the spinner and record the number spun. Then $\mathcal{C} = \{1, 2, \ldots, 10\}$. Define the events A, B, and C by $A = \{1, 2\}$, $B = \{2, 3, 4\}$, and $C = \{3, 4, 5, 6\}$, respectively.

$$A^c = \{3, 4, \ldots, 10\}; \quad A \cup B = \{1, 2, 3, 4\}; \quad A \cap B = \{2\}$$
$$A \cap C = \phi; \quad B \cap C = \{3, 4\}; \quad B \cap C \subset B; \quad B \cap C \subset C$$
$$A \cup (B \cap C) = \{1, 2\} \cup \{3, 4\} = \{1, 2, 3, 4\} \tag{1.2.1}$$
$$(A \cup B) \cap (A \cup C) = \{1, 2, 3, 4\} \cap \{1, 2, 3, 4, 5, 6\} = \{1, 2, 3, 4\} \tag{1.2.2}$$

The reader should verify these results. ■

Example 1.2.2. For this example, suppose the experiment is to select a real number in the open interval $(0, 5)$; hence, the sample space is $\mathcal{C} = (0, 5)$. Let $A = (1, 3)$,

$B = (2,4)$, and $C = [3,4.5)$.

$$A \cup B = (1,4); \quad A \cap B = (2,3); \quad B \cap C = [3,4)$$
$$A \cap (B \cup C) = (1,3) \cap (2,4.5) = (2,3) \tag{1.2.3}$$
$$(A \cap B) \cup (A \cap C) = (2,3) \cup \phi = (2,3) \tag{1.2.4}$$

A sketch of the real number line between 0 and 5 helps to verify these results. ■

Expressions (1.2.1)–(1.2.2) and (1.2.3)–(1.2.4) are illustrations of general **distributive laws**. For any sets A, B, and C,

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C) \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned} \tag{1.2.5}$$

These follow directly from set theory. To verify each identity, sketch Venn diagrams of both sides.

The next two identities are collectively known as **DeMorgan's Laws**. For any sets A and B,

$$(A \cap B)^c = A^c \cup B^c \tag{1.2.6}$$
$$(A \cup B)^c = A^c \cap B^c. \tag{1.2.7}$$

For instance, in Example 1.2.1,

$$(A \cup B)^c = \{1,2,3,4\}^c = \{5,6,\ldots,10\} = \{3,4,\ldots,10\} \cap \{\{1,5,6,\ldots,10\} = A^c \cap B^c;$$

while, from Example 1.2.2,

$$(A \cap B)^c = (2,3)^c = (0,2] \cup [3,5) = [(0,1] \cup [3,5)] \cup [(0,2] \cup [4,5)] = A^c \cup B^c.$$

As the last expression suggests, it is easy to extend unions and intersections to more than two sets. If $A_1, A_2, \ldots, A_n$ are any sets, we define

$$A_1 \cup A_2 \cup \cdots \cup A_n = \{x : x \in A_i, \text{ for some } i = 1,2,\ldots,n\} \tag{1.2.8}$$
$$A_1 \cap A_2 \cap \cdots \cap A_n = \{x : x \in A_i, \text{ for all } i = 1,2,\ldots,n\}. \tag{1.2.9}$$

We often abbreviative these by $\cup_{i=1}^n A_i$ and $\cap_{i=1}^n A_i$, respectively. Expressions for countable unions and intersections follow directly; that is, if $A_1, A_2, \ldots, A_n \ldots$ is a sequence of sets then

$$A_1 \cup A_2 \cup \cdots = \{x : x \in A_n, \text{ for some } n = 1,2,\ldots\} = \cup_{n=1}^\infty A_n \tag{1.2.10}$$
$$A_1 \cap A_2 \cap \cdots = \{x : x \in A_n, \text{ for all } n = 1,2,\ldots\} = \cap_{n=1}^\infty A_n. \tag{1.2.11}$$

The next two examples illustrate these ideas.

Example 1.2.3. Suppose $\mathcal{C} = \{1,2,3,\ldots\}$. If $A_n = \{1,3,\ldots,2n-1\}$ and $B_n = \{n, n+1, \ldots\}$, for $n = 1,2,3,\ldots$, then

$$\cup_{n=1}^\infty A_n = \{1,3,5,\ldots\}; \quad \cap_{n=1}^\infty A_n = \{1\}; \tag{1.2.12}$$
$$\cup_{n=1}^\infty B_n = \mathcal{C}; \quad \cap_{n=1}^\infty B_n = \phi. \quad ■ \tag{1.2.13}$$

Example 1.2.4. Suppose $\mathcal{C}$ is the interval of real numbers $(0, 5)$. Suppose $C_n = (1 - n^{-1}, 2 + n^{-1})$ and $D_n = (n^{-1}, 3 - n^{-1})$, for $n = 1, 2, 3, \ldots$. Then

$$\cup_{n=1}^{\infty} C_n = (0, 3); \quad \cap_{n=1}^{\infty} C_n = [1, 2] \tag{1.2.14}$$

$$\cup_{n=1}^{\infty} D_n = (0, 3); \quad \cap_{n=1}^{\infty} D_n = (1, 2). \; \blacksquare \tag{1.2.15}$$

We occassionally have sequences of sets that are **monotone**. They are of two types. We say a sequence of sets $\{A_n\}$ is **nondecreasing, (nested upward)**, if $A_n \subset A_{n+1}$ for $n = 1, 2, 3, \ldots$. For such a sequence, we define

$$\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n. \tag{1.2.16}$$

The sequence of sets $A_n = \{1, 3, \ldots, 2n - 1\}$ of Example 1.2.3 is such a sequence. So in this case, we write $\lim_{n\to\infty} A_n = \{1, 3, 5, \ldots\}$. The sequence of sets $\{D_n\}$ of Example 1.2.4 is also a nondecreasing suquence of sets.

The second type of monotone sets consists of the **nonincreasing, (nested downward)** sequences. A sequence of sets $\{A_n\}$ is **nonincreasing**, if $A_n \supset A_{n+1}$ for $n = 1, 2, 3, \ldots$. In this case, we define

$$\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n. \tag{1.2.17}$$

The sequences of sets $\{B_n\}$ and $\{C_n\}$ of Examples 1.2.3 and 1.2.4, respectively, are examples of nonincreasing sequences of sets.

1.2.2 Set Functions

Many of the functions used in calculus and in this book are functions that map real numbers into real numbers. We are concerned also with functions that map sets into real numbers. Such functions are naturally called functions of a set or, more simply, **set functions**. Next we give some examples of set functions and evaluate them for certain simple sets.

Example 1.2.5. Let $\mathcal{C} = R$, the set of real numbers. For a subset A in $\mathcal{C}$, let $Q(A)$ be equal to the number of points in A that correspond to positive integers. Then $Q(A)$ is a set function of the set A. Thus, if $A = \{x : 0 < x < 5\}$, then $Q(A) = 4$; if $A = \{-2, -1\}$, then $Q(A) = 0$; and if $A = \{x : -\infty < x < 6\}$, then $Q(A) = 5$. $\blacksquare$

Example 1.2.6. Let $\mathcal{C} = R^2$. For a subset A of $\mathcal{C}$, let $Q(A)$ be the area of A if A has a finite area; otherwise, let $Q(A)$ be undefined. Thus, if $A = \{(x, y) : x^2 + y^2 \leq 1\}$, then $Q(A) = \pi$; if $A = \{(0, 0), (1, 1), (0, 1)\}$, then $Q(A) = 0$; and if $A = \{(x, y) : 0 \leq x, 0 \leq y, x + y \leq 1\}$, then $Q(A) = \frac{1}{2}$. $\blacksquare$

Often our set functions are defined in terms of sums or integrals.[1] With this in mind, we introduce the following notation. The symbol

$$\int_A f(x)\, dx$$

[1]Please see Chapters 2 and 3 of *Mathematical Comments*, at site noted in the Preface, for a review of sums and integrals

means the ordinary (Riemann) integral of $f(x)$ over a prescribed one-dimensional set A and the symbol

$$\iint_A g(x,y)\,dxdy$$

means the Riemann integral of $g(x,y)$ over a prescribed two-dimensional set A. This notation can be extended to integrals over n dimensions. To be sure, unless these sets A and these functions $f(x)$ and $g(x,y)$ are chosen with care, the integrals frequently fail to exist. Similarly, the symbol

$$\sum_A f(x)$$

means the sum extended over all $x \in A$ and the symbol

$$\sum\sum_A g(x,y)$$

means the sum extended over all $(x,y) \in A$. As with integration, this notation extends to sums over n dimensions.

The first example is for a set function defined on sums involving a **geometric series**. As pointed out in Example 2.3.1 of *Mathematical Comments*,[2] if $|a| < 1$, then the following series converges to $1/(1-a)$:

$$\sum_{n=0}^{\infty} a^n = \frac{1}{1-a}, \quad \text{if } |a| < 1. \tag{1.2.18}$$

Example 1.2.7. Let $\mathcal{C}$ be the set of all nonnegative integers and let A be a subset of $\mathcal{C}$. Define the set function Q by

$$Q(A) = \sum_{n \in A} \left(\frac{2}{3}\right)^n. \tag{1.2.19}$$

It follows from (1.2.18) that $Q(\mathcal{C}) = 3$. If $A = \{1,2,3\}$ then $Q(A) = 38/27$. Suppose $B = \{1,3,5,\ldots\}$ is the set of all odd positive integers. The computation of $Q(B)$ is given next. This derivation consists of rewriting the series so that (1.2.18) can be applied. Frequently, we perform such derivations in this book.

$$\begin{aligned} Q(B) &= \sum_{n\in B}\left(\frac{2}{3}\right)^n = \sum_{n=0}^{\infty}\left(\frac{2}{3}\right)^{2n+1} \\ &= \frac{2}{3}\sum_{n=0}^{\infty}\left[\left(\frac{2}{3}\right)^2\right]^n = \frac{2}{3}\frac{1}{1-(4/9)} = \frac{6}{5} \quad \blacksquare \end{aligned}$$

In the next example, the set function is defined in terms of an integral involving the exponential function $f(x) = e^{-x}$.

[2]Downloadable at site noted in the Preface

Example 1.2.8. Let $\mathcal{C}$ be the interval of positive real numbers, i.e., $\mathcal{C} = (0, \infty)$. Let A be a subset of $\mathcal{C}$. Define the set function Q by

$$Q(A) = \int_A e^{-x}\,dx, \tag{1.2.20}$$

provided the integral exists. The reader should work through the following integrations:

$$
\begin{aligned}
Q[(1,3)] &= \int_1^3 e^{-x}\,dx = -e^{-x}\Big|_1^3 = e^{-1} - e^{-3} \doteq 0.318 \\
Q[(5 \text{ and } \infty)] &= \int_1^3 e^{-x}\,dx = -e^{-x}\Big|_5^\infty = e^{-5} \doteq 0.007 \\
Q[(1,3) \cup [3,5)] &= \int_1^5 e^{-x}\,dx = \int_1^3 e^{-x}\,dx + \int_3^5 e^{-x}\,dx = Q[(1,3)] + Q([3,5)] \\
Q(\mathcal{C}) &= \int_0^\infty e^{-x}\,dx = 1. \quad \blacksquare
\end{aligned}
$$

Our final example, involves an n dimensional integral.

Example 1.2.9. Let $\mathcal{C} = R^n$. For A in $\mathcal{C}$ define the set function

$$Q(A) = \int \cdots \underset{A}{\int} dx_1 dx_2 \cdots dx_n,$$

provided the integral exists. For example, if $A = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1 \le x_2, 0 \le x_i \le 1, \text{ for } i = 2, 3, \ldots, \text{n}\}$, then upon expressing the multiple integral as an iterated integral[3] we obtain

$$
\begin{aligned}
Q(A) &= \int_0^1 \left[\int_0^{x_2} dx_1\right] dx_2 \bullet \prod_{i=3}^n \left[\int_0^1 dx_i\right] \\
&= \frac{x_2^2}{2}\Big|_0^1 \bullet 1 = \frac{1}{2}.
\end{aligned}
$$

If $B = \{(x_1, x_2, \ldots, x_n) : 0 \le x_1 \le x_2 \le \cdots \le x_n \le 1\}$, then

$$
\begin{aligned}
Q(B) &= \int_0^1 \left[\int_0^{x_n} \cdots \left[\int_0^{x_3} \left[\int_0^{x_2} dx_1\right] dx_2\right] \cdots dx_{n-1}\right] dx_n \\
&= \frac{1}{n!},
\end{aligned}
$$

where $n! = n(n-1)\cdots 3 \cdot 2 \cdot 1$. $\blacksquare$

[3]For a discussion of multiple integrals in terms of iterated integrals, see Chapter 3 of *Mathematical Comments*.

EXERCISES

1.2.1. Find the union $C_1 \cup C_2$ and the intersection $C_1 \cap C_2$ of the two sets C_1 and C_2, where

(a) $C_1 = \{2, 3, 5, 7\}$, $C_2 = \{1, 3, 5\}$.

(b) $C_1 = \{x : 0 \le x \le 3\}$, $C_2 = \{x : 2 < x < 4\}$.

(c) $C_1 = \{(x, y) : 0 < x < 1, 0 < y < 3\}$, $C_2 = \{(x, y) : 0 < x < 2, 2 \le y < 3\}$.

1.2.2. Find the complement C^c of the set C with respect to the space $\mathcal{C}$ if

(a) $\mathcal{C} = \{x : 0 < x < 2\}$, $C = \left\{x : 0 < x < \frac{2}{3}\right\}$.

(b) $\mathcal{C} = \left\{(x, y, z) : x^2 + 2y^2 + 3z^2 \le 4\right\}$, $C = \left\{(x, y, z) : x^2 + 2y^2 + 3z^2 < 4\right\}$.

(c) $\mathcal{C} = \left\{(x, y) : x^2 + y^2 \le 1\right\}$, $C = \{(x, y) : |x| + |y| < 1\}$.

1.2.3. List all possible arrangements of the four letters l, a, m, and b. Let C_1 be the collection of the arrangements in which b is in the first position. Let C_2 be the collection of the arrangements in which a is in the third position. Find the union and the intersection of C_1 and C_2.

1.2.4. Concerning DeMorgan's Laws (1.2.6) and (1.2.7):

(a) Use Venn diagrams to verify the laws.

(b) Show that the laws are true.

(c) Generalize the laws to countable unions and intersections.

1.2.5. By the use of Venn diagrams, in which the space $\mathcal{C}$ is the set of points enclosed by a rectangle containing the circles C_1, C_2, and C_3, compare the following sets. These laws are called the **distributive laws**.

(a) $C_1 \cap (C_2 \cup C_3)$ and $(C_1 \cap C_2) \cup (C_1 \cap C_3)$.

(b) $C_1 \cup (C_2 \cap C_3)$ and $(C_1 \cup C_2) \cap (C_1 \cup C_3)$.

1.2.6. Show that the following sequences of sets, $\{C_k\}$, are nondecreasing, (1.2.16), then find $\lim_{k\to\infty} C_k$.

(a) $C_k = \{x : 1/k \le x \le 3 - 1/k\}$, $k = 1, 2, 3, \ldots$.

(b) $C_k = \{(x, y) : 1/k \le x^2 + y^2 \le 4 - 1/k\}$, $k = 1, 2, 3, \ldots$.

1.2.7. Show that the following sequences of sets, $\{C_k\}$, are nonincreasing, (1.2.17), then find $\lim_{k\to\infty} C_k$.

(a) $C_k = \{x : 2 - 1/k < x \le 2\}$, $k = 1, 2, 3, \ldots$.

(b) $C_k = \{x : 2 < x \le 2 + 1/k\}$, $k = 1, 2, 3, \ldots$.

(c) $C_k = \{(x, y) : 0 \leq x^2 + y^2 \leq 1/k\}$, $k = 1, 2, 3, \ldots$.

1.2.8. For every one-dimensional set C, define the function $Q(C) = \sum_C f(x)$, where $f(x) = \left(\frac{3}{4}\right)\left(\frac{1}{4}\right)^x$, $x = 0, 1, 2, \ldots$, zero elsewhere. If $C_1 = \{x : x = 0, 2, 4\}$ and $C_2 = \{x : x = 0, 1, 2, \ldots\}$, find $Q(C_1)$ and $Q(C_2)$.
Hint: Recall that $S_n = a + ar + \cdots + ar^{n-1} = a(1 - r^n)/(1 - r)$ and, hence, it follows that $\lim_{n\to\infty} S_n = a/(1 - r)$ provided that $|r| < 1$.

1.2.9. For every one-dimensional set C for which the integral exists, let $Q(C) = \int_C f(x)\, dx$, where $f(x) = \frac{3}{4}(1-x^2)$, $-1 < x < 1$, zero elsewhere; otherwise, let $Q(C)$ be undefined. If $C_1 = \{x : -\frac{1}{3} < x < \frac{1}{3}\}$, $C_2 = \{0\}$, and $C_3 = \{x : -1 < x < 5\}$, find $Q(C_1)$, $Q(C_2)$, and $Q(C_3)$.

1.2.10. For every two-dimensional set C contained in R^2 for which the integral exists, let $Q(C) = \int\int_C (x^2 + y^2)\, dxdy$. If $C_1 = \{(x, y) : -1 \leq x \leq 1, -1 \leq y \leq 1\}$, $C_2 = \{(x, y) : -1 \leq x = y \leq 1\}$, and $C_3 = \{(x, y) : x^2 + y^2 \leq 1\}$, find $Q(C_1), Q(C_2)$, and $Q(C_3)$.

1.2.11. Let $\mathcal{C}$ denote the set of points that are interior to, or on the boundary of, a square with opposite vertices at the points (0, 0) and (1, 1). Let $Q(C) = \int\int_C dy\, dx$.

(a) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < y/2 < x < 1/2\}$, compute $Q(C)$.

(b) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < x < 1, x + y = 1\}$, compute $Q(C)$.

(c) If $C \subset \mathcal{C}$ is the set $\{(x, y) : 0 < x/2 < y \leq x + 1/4 < 1\}$, compute $Q(C)$.

1.2.12. Let $\mathcal{C}$ be the set of points interior to or on the boundary of a cube with edge of length 1. Moreover, say that the cube is in the first octant with one vertex at the point $(0, 0, 0)$ and an opposite vertex at the point $(1, 1, 1)$. Let $Q(C) = \int\int\int_C dxdydz$.

(a) If $C \subset \mathcal{C}$ is the set $\{(x, y, z) : 0 < x < y < z < 1\}$, compute $Q(C)$.

(b) If C is the subset $\{(x, y, z) : 0 < x = y = z < 1\}$, compute $Q(C)$.

1.2.13. Let C denote the set $\{(x, y, z) : x^2 + y^2 + z^2 \leq 1\}$. Using spherical coordinates, evaluate

$$Q(C) = \int\int\int_C \sqrt{x^2 + y^2 + z^2}\, dxdydz.$$

1.2.14. To join a certain club, a person must be either a statistician or a mathematician or both. Of the 35 members in this club, 25 are statisticians and 17 are mathematicians. How many persons in the club are both a statistician and a mathematician?

1.2.15. After a hard-fought football game, it was reported that, of the 11 starting players, 7 hurt a hip, 5 hurt an arm, 7 hurt a knee, 3 hurt both a hip and an arm, 3 hurt both a hip and a knee, 2 hurt both an arm and a knee, and 1 hurt all three. Comment on the accuracy of the report.

1.3 The Probability Set Function

Given an experiment, let $\mathcal{C}$ denote the sample space of all possible outcomes. As discussed in Section 1.1, we are interested in assigning probabilities to events, i.e., subsets of $\mathcal{C}$. What should be our collection of events? If $\mathcal{C}$ is a finite set, then we could take the set of all subsets as this collection. For infinite sample spaces, though, with assignment of probabilities in mind, this poses mathematical technicalities that are better left to a course in probability theory. We assume that in all cases, the collection of events is sufficiently rich to include all possible events of interest and is closed under complements and countable unions of these events. Using DeMorgan's Laws, (1.2.6)–(1.2.7), the collection is then also closed under countable intersections. We denote this collection of events by $\mathcal{B}$. Technically, such a collection of events is called a **$\boldsymbol{\sigma}$-field** of subsets.

Now that we have a sample space, $\mathcal{C}$, and our collection of events, $\mathcal{B}$, we can define the third component in our probability space, namely a probability set function. In order to motivate its definition, we consider the relative frequency approach to probability.

Remark 1.3.1. The definition of probability consists of three axioms which we motivate by the following three intuitive properties of relative frequency. Let $\mathcal{C}$ be a sample space and let $A \subset \mathcal{C}$. Suppose we repeat the experiment N times. Then the relative frequency of A is $f_A = \#\{A\}/N$, where $\#\{A\}$ denotes the number of times A occurred in the N repetitions. Note that $f_A \geq 0$ and $f_{\mathcal{C}} = 1$. These are the first two properties. For the third, suppose that A_1 and A_2 are disjoint events. Then $f_{A_1 \cup A_2} = f_{A_1} + f_{A_2}$. These three properties of relative frequencies form the axioms of a probability, except that the third axiom is in terms of countable unions. As with the axioms of probability, the readers should check that the theorems we prove below about probabilities agree with their intuition of relative frequency. ■

Definition 1.3.1 (Probability)**.** *Let $\mathcal{C}$ be a sample space and let $\mathcal{B}$ be the set of events. Let P be a real-valued function defined on $\mathcal{B}$. Then P is a* **probability set function** *if P satisfies the following three conditions:*

1. *$P(A) \geq 0$, for all $A \in \mathcal{B}$.*

2. *$P(\mathcal{C}) = 1$.*

3. *If $\{A_n\}$ is a sequence of events in $\mathcal{B}$ and $A_m \cap A_n = \phi$ for all $m \neq n$, then*

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

A collection of events whose members are pairwise disjoint, as in (3), is said to be a **mutually exclusive** collection and its union is often referred to as a **disjoint union**. The collection is further said to be **exhaustive** if the union of its events is the sample space, in which case $\sum_{n=1}^{\infty} P(A_n) = 1$. We often say that a mutually exclusive and exhaustive collection of events forms a **partition** of $\mathcal{C}$.

A probability set function tells us how the probability is distributed over the set of events, $\mathcal{B}$. In this sense we speak of a distribution of probability. We often drop the word "set" and refer to P as a probability function.

The following theorems give us some other properties of a probability set function. In the statement of each of these theorems, $P(A)$ is taken, tacitly, to be a probability set function defined on the collection of events $\mathcal{B}$ of a sample space $\mathcal{C}$.

Theorem 1.3.1. *For each event $A \in \mathcal{B}$, $P(A) = 1 - P(A^c)$.*

Proof: We have $\mathcal{C} = A \cup A^c$ and $A \cap A^c = \phi$. Thus, from (2) and (3) of Definition 1.3.1, it follows that

$$1 = P(A) + P(A^c),$$

which is the desired result. ■

Theorem 1.3.2. *The probability of the null set is zero; that is, $P(\phi) = 0$.*

Proof: In Theorem 1.3.1, take $A = \phi$ so that $A^c = \mathcal{C}$. Accordingly, we have

$$P(\phi) = 1 - P(\mathcal{C}) = 1 - 1 = 0$$

and the theorem is proved. ■

Theorem 1.3.3. *If A and B are events such that $A \subset B$, then $P(A) \leq P(B)$.*

Proof: Now $B = A \cup (A^c \cap B)$ and $A \cap (A^c \cap B) = \phi$. Hence, from (3) of Definition 1.3.1,

$$P(B) = P(A) + P(A^c \cap B).$$

From (1) of Definition 1.3.1, $P(A^c \cap B) \geq 0$. Hence, $P(B) \geq P(A)$. ■

Theorem 1.3.4. *For each $A \in \mathcal{B}$, $0 \leq P(A) \leq 1$.*

Proof: Since $\phi \subset A \subset \mathcal{C}$, we have by Theorem 1.3.3 that

$$P(\phi) \leq P(A) \leq P(\mathcal{C}) \quad \text{or} \quad 0 \leq P(A) \leq 1,$$

the desired result. ■

Part (3) of the definition of probability says that $P(A \cup B) = P(A) + P(B)$ if A and B are disjoint, i.e., $A \cap B = \phi$. The next theorem gives the rule for any two events regardless if they are disjoint or not.

Theorem 1.3.5. *If A and B are events in $\mathcal{C}$, then*

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

Proof: Each of the sets $A \cup B$ and B can be represented, respectively, as a union of nonintersecting sets as follows:

$$A \cup B = A \cup (A^c \cap B) \quad \text{and} \quad B = (A \cap B) \cup (A^c \cap B). \tag{1.3.1}$$

That these identities hold for all sets A and B follows from set theory. Also, the Venn diagrams of Figure 1.3.1 offer a verification of them.

Thus, from (3) of Definition 1.3.1,

$$P(A \cup B) = P(A) + P(A^c \cap B)$$

and

$$P(B) = P(A \cap B) + P(A^c \cap B).$$

If the second of these equations is solved for $P(A^c \cap B)$ and this result is substituted in the first equation, we obtain

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

This completes the proof. ■

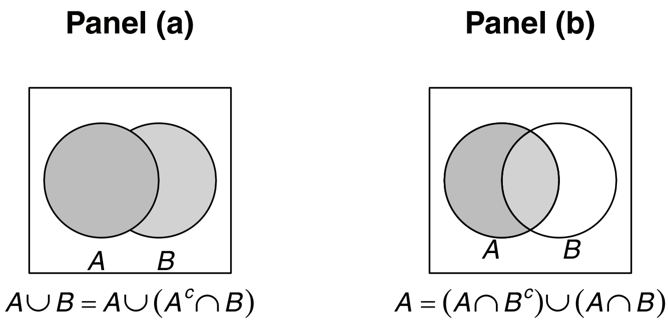

Figure 1.3.1: Venn diagrams depicting the two disjoint unions given in expression (1.3.1). Panel (a) depicts the first disjoint union while Panel (b) shows the second disjoint union.

Example 1.3.1. Let $\mathcal{C}$ denote the sample space of Example 1.1.2. Let the probability set function assign a probability of $\frac{1}{36}$ to each of the 36 points in $\mathcal{C}$; that is, the dice are fair. If $C_1 = \{(1,1),(2,1),(3,1),(4,1),(5,1)\}$ and $C_2 = \{(1,2),(2,2),(3,2)\}$, then $P(C_1) = \frac{5}{36}$, $P(C_2) = \frac{3}{36}$, $P(C_1 \cup C_2) = \frac{8}{36}$, and $P(C_1 \cap C_2) = 0$. ■

Example 1.3.2. Two coins are to be tossed and the outcome is the ordered pair (face on the first coin, face on the second coin). Thus the sample space may be represented as $\mathcal{C} = \{(H,H),(H,T),(T,H),(T,T)\}$. Let the probability set function assign a probability of $\frac{1}{4}$ to each element of $\mathcal{C}$. Let $C_1 = \{(H,H),(H,T)\}$ and $C_2 = \{(H,H),(T,H)\}$. Then $P(C_1) = P(C_2) = \frac{1}{2}$, $P(C_1 \cap C_2) = \frac{1}{4}$, and, in accordance with Theorem 1.3.5, $P(C_1 \cup C_2) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4}$. ■

For a finite sample space, we can generate probabilities as follows. Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite set of m elements. Let $p_1, p_2, \ldots, p_m$ be fractions such that

$$0 \le p_i \le 1 \text{ for } i = 1, 2, \ldots, m \text{ and } \textstyle\sum_{i=1}^m p_i = 1. \tag{1.3.2}$$

Suppose we define P by

$$P(A) = \sum_{x_i \in A} p_i, \quad \text{for all subsets } A \text{ of } \mathcal{C}. \tag{1.3.3}$$

Then $P(A) \ge 0$ and $P(\mathcal{C}) = 1$. Further, it follows that $P(A \cup B) = P(A) + P(B)$ when $A \cap B = \phi$. Therefore, P is a probability on $\mathcal{C}$. For illustration, each of the following four assignments forms a probability on $\mathcal{C} = \{1, 2, \ldots, 6\}$. For each, we also compute $P(A)$ for the event $A = \{1, 6\}$.

$$\begin{aligned}
&p_1 = p_2 = \cdots = p_6 = \frac{1}{6}; \quad P(A) = \frac{1}{3}. &(1.3.4)\\
&p_1 = p_2 = 0.1, p_3 = p_4 = p_5 = p_6 = 0.2; \quad P(A) = 0.3.\\
&p_i = \frac{i}{21}, \quad i = 1, 2, \ldots, 6; \quad P(A) = \frac{7}{21}.\\
&p_1 = \frac{3}{\pi}, p_2 = 1 - \frac{3}{\pi}, p_3 = p_4 = p_5 = p_6 = 0.0; \quad P(A) = \frac{3}{\pi}.
\end{aligned}$$

Note that the individual probabilities for the first probability set function, (1.3.4), are the same. This is an example of the equilikely case which we now formally define.

Definition 1.3.2 (Equilikely Case)**.** *Let $\mathcal{C} = \{x_1, x_2, \ldots, x_m\}$ be a finite sample space. Let $p_i = 1/m$ for all $i = 1, 2, \ldots, m$ and for all subsets A of $\mathcal{C}$ define*

$$P(A) = \sum_{x_i \in A} \frac{1}{m} = \frac{\#(A)}{m},$$

where $\#(A)$ denotes the number of elements in A. Then P is a probability on $\mathcal{C}$ and it is refereed to as the **equilikely case**. ■

Equilikely cases are frequently probability models of interest. Examples include: the flip of a fair coin; five cards drawn from a well shuffled deck of 52 cards; a spin of a fair spinner with the numbers 1 through 36 on it; and the upfaces of the roll of a pair of balanced dice. For each of these experiments, as stated in the definition, we only need to know the number of elements in an event to compute the probability of that event. For example, a card player may be interested in the probability of getting a pair (two of a kind) in a hand of five cards dealt from a well shuffled deck of 52 cards. To compute this probability, we need to know the number of five card hands and the number of such hands which contain a pair. Because the equilikely case is often of interest, we next develop some counting rules which can be used to compute the probabilities of events of interest.

1.3.1 Counting Rules

We discuss three counting rules that are usually discussed in an elementary algebra course.

The first rule is called the mn-**rule** (m times n-rule), which is also called the **multiplication rule**. Let $A = \{x_1, x_2, \ldots, x_m\}$ be a set of m elements and let $B = \{y_1, y_2, \ldots, y_n\}$ be a set of n elements. Then there are mn ordered pairs, (x_i, y_j), $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$, of elements, the first from A and the second from B. Informally, we often speak of ways, here. For example there are five roads (ways) between cities I and II and there are ten roads (ways) between cities II and III. Hence, there are $5 * 10 = 50$ ways to get from city I to city III by going from city I to city II and then from city II to city III. This rule extends immediately to more than two sets. For instance, suppose in a certain state that driver license plates have the pattern of three letters followed by three numbers. Then there are $26^3 * 10^3$ possible license plates in this state.

Next, let A be a set with n elements. Suppose we are interested in k-tuples whose components are elements of A. Then by the extended mn rule, there are $n \cdot n \cdots n = n^k$ such k-tuples whose components are elements of A. Next, suppose $k \leq n$ and we are interested in k-tuples whose components are distinct (no repeats) elements of A. There are n elements from which to choose for the first component, $n-1$ for the second component, ..., $n-(k-1)$ for the kth. Hence, by the mn rule, there are $n(n-1)\cdots(n-(k-1))$ such k-tuples with distinct elements. We call each such k-tuple a **permutation** and use the symbol P_k^n to denote the number of k permutations taken from a set of n elements. This number of permutations, P_k^n is our second counting rule. We can rewrite it as

$$P_k^n = n(n-1)\cdots(n-(k-1)) = \frac{n!}{(n-k)!}. \tag{1.3.5}$$

Example 1.3.3 (Birthday Problem)**.** Suppose there are n people in a room. Assume that $n < 365$ and that the people are unrelated in any way. Find the probability of the event A that at least 2 people have the same birthday. For convenience, assign the numbers 1 though n to the people in the room. Then use n-tuples to denote the birthdays of the first person through the nth person in the room. Using the mn-rule, there are 365^n possible birthday n-tuples for these n people. This is the number of elements in the sample space. Now assume that birthdays are equilikely to occur on any of the 365 days. Hence, each of these n-tuples has probability 365^{-n}. Notice that the complement of A is the event that all the birthdays in the room are distinct; that is, the number of n-tuples in A^c is P_n^{365}. Thus, the probability of A is

$$P(A) = 1 - \frac{P_n^{365}}{365^n}.$$

For instance, if $n = 2$ then $P(A) = 1 - (365 * 364)/(365^2) = 0.0027$. This formula is not easy to compute by hand. The following R function[4] computes the $P(A)$ for the input n and it can be downloaded at the sites mentioned in the Preface.

[4]An R primer for the course is found in Appendix B.

```
bday = function(n){ bday = 1; nm1 = n - 1
    for(j in 1:nm1){bday = bday*((365-j)/365)}
    bday <- 1 - bday; return(bday)}
```

Assuming that the file `bday.R` contains this function, here is the R segment computing $P(A)$ for $n = 10$:

```
> source("bday.R")
> bday(10)
[1] 0.1169482
```

■

For our last counting rule, as with permutations, we are drawing from a set A of n elements. Now, suppose order is not important, so instead of counting the number of permutations we want to count the number of subsets of k elements taken from A. We use the symbol $\binom{n}{k}$ to denote the total number of these subsets. Consider a subset of k elements from A. By the permutation rule it generates $P_k^k = k(k-1)\cdots 1 = k!$ permutations. Furthermore, all these permutations are distinct from the permutations generated by other subsets of k elements from A. Finally, each permutation of k distinct elements drawn from A must be generated by one of these subsets. Hence, we have shown that $P_k^n = \binom{n}{k}k!$; that is,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.3.6}$$

We often use the terminology combinations instead of subsets. So we say that there are $\binom{n}{k}$ **combinations** of k things taken from a set of n things. Another common symbol for $\binom{n}{k}$ is C_k^n.

It is interesting to note that if we expand the binomial series,

$$(a+b)^n = (a+b)(a+b)\cdots(a+b),$$

we get

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}, \tag{1.3.7}$$

because we can select the k factors from which to take a in $\binom{n}{k}$ ways. So $\binom{n}{k}$ is also referred to as a **binomial coefficient**.

Example 1.3.4 (Poker Hands)**.** Let a card be drawn at random from an ordinary deck of 52 playing cards that has been well shuffled. The sample space $\mathcal{C}$ consists of 52 elements, each element represents one and only one of the 52 cards. Because the deck has been well shuffled, it is reasonable to assume that each of these outcomes has the same probability $\frac{1}{52}$. Accordingly, if E_1 is the set of outcomes that are spades, $P(E_1) = \frac{13}{52} = \frac{1}{4}$ because there are 13 spades in the deck; that is, $\frac{1}{4}$ is the probability of drawing a card that is a spade. If E_2 is the set of outcomes that are kings, $P(E_2) = \frac{4}{52} = \frac{1}{13}$ because there are 4 kings in the deck; that is, $\frac{1}{13}$ is the probability of drawing a card that is a king. These computations are very easy because there are no difficulties in the determination of the number of elements in each event.

However, instead of drawing only one card, suppose that five cards are taken, at random and without replacement, from this deck; i.e, a five card poker hand. In this instance, order is not important. So a hand is a subset of five elements drawn from a set of 52 elements. Hence, by (1.3.6) there are $\binom{52}{5}$ poker hands. If the deck is well shuffled, each hand should be equilikely; i.e., each hand has probability $1/\binom{52}{5}$. We can now compute the probabilities of some interesting poker hands. Let E_1 be the event of a flush, all five cards of the same suit. There are $\binom{4}{1} = 4$ suits to choose for the flush and in each suit there are $\binom{13}{5}$ possible hands; hence, using the multiplication rule, the probability of getting a flush is

$$P(E_1) = \frac{\binom{4}{1}\binom{13}{5}}{\binom{52}{5}} = \frac{4 \cdot 1287}{2598960} = 0.00198.$$

Real poker players note that this includes the probability of obtaining a straight flush.

Next, consider the probability of the event E_2 of getting exactly three of a kind, (the other two cards are distinct and are of different kinds). Choose the kind for the three, in $\binom{13}{1}$ ways; choose the three, in $\binom{4}{3}$ ways; choose the other two kinds, in $\binom{12}{2}$ ways; and choose one card from each of these last two kinds, in $\binom{4}{1}\binom{4}{1}$ ways. Hence the probability of exactly three of a kind is

$$P(E_2) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2}{\binom{52}{5}} = 0.0211.$$

Now suppose that E_3 is the set of outcomes in which exactly three cards are kings and exactly two cards are queens. Select the kings, in $\binom{4}{3}$ ways, and select the queens, in $\binom{4}{2}$ ways. Hence, the probability of E_3 is

$$P(E_3) = \binom{4}{3}\binom{4}{2}\Big/\binom{52}{5} = 0.0000093.$$

The event E_3 is an example of a full house: three of one kind and two of another kind. Exercise 1.3.19 asks for the determination of the probability of a full house. ■

1.3.2 Additional Properties of Probability

We end this section with several additional properties of probability which prove useful in the sequel. Recall in Exercise 1.2.6 we said that a sequence of events $\{C_n\}$ is a nondecreasing sequence if $C_n \subset C_{n+1}$, for all n, in which case we wrote $\lim_{n\to\infty} C_n = \cup_{n=1}^{\infty} C_n$. Consider $\lim_{n\to\infty} P(C_n)$. The question is: can we legitimately interchange the limit and P? As the following theorem shows, the answer is yes. The result also holds for a decreasing sequence of events. Because of this interchange, this theorem is sometimes referred to as the continuity theorem of probability.

Theorem 1.3.6. *Let $\{C_n\}$ be a nondecreasing sequence of events. Then*

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcup_{n=1}^{\infty} C_n\right). \tag{1.3.8}$$

Let $\{C_n\}$ be a decreasing sequence of events. Then

$$\lim_{n\to\infty} P(C_n) = P(\lim_{n\to\infty} C_n) = P\left(\bigcap_{n=1}^{\infty} C_n\right). \tag{1.3.9}$$

Proof. We prove the result (1.3.8) and leave the second result as Exercise 1.3.20. Define the sets, called rings, as $R_1 = C_1$ and, for $n > 1$, $R_n = C_n \cap C_{n-1}^c$. It follows that $\bigcup_{n=1}^{\infty} C_n = \bigcup_{n=1}^{\infty} R_n$ and that $R_m \cap R_n = \phi$, for $m \neq n$. Also, $P(R_n) = P(C_n) - P(C_{n-1})$. Applying the third axiom of probability yields the following string of equalities:

$$\begin{aligned} P\left[\lim_{n\to\infty} C_n\right] &= P\left(\bigcup_{n=1}^{\infty} C_n\right) = P\left(\bigcup_{n=1}^{\infty} R_n\right) = \sum_{n=1}^{\infty} P(R_n) = \lim_{n\to\infty} \sum_{j=1}^{n} P(R_j) \\ &= \lim_{n\to\infty}\left\{P(C_1)+\sum_{j=2}^{n}[P(C_j) - P(C_{j-1})]\right\} = \lim_{n\to\infty} P(C_n). \end{aligned} \tag{1.3.10}$$

This is the desired result. ∎

Another useful result for arbitrary unions is given by

Theorem 1.3.7 (Boole's Inequality). *Let $\{C_n\}$ be an arbitrary sequence of events. Then*

$$P\left(\bigcup_{n=1}^{\infty} C_n\right) \leq \sum_{n=1}^{\infty} P(C_n). \tag{1.3.11}$$

Proof: Let $D_n = \bigcup_{i=1}^{n} C_i$. Then $\{D_n\}$ is an increasing sequence of events that go up to $\bigcup_{n=1}^{\infty} C_n$. Also, for all j, $D_j = D_{j-1} \cup C_j$. Hence, by Theorem 1.3.5,

$$P(D_j) \leq P(D_{j-1}) + P(C_j),$$

that is,

$$P(D_j) - P(D_{j-1}) \leq P(C_j).$$

In this case, the C_is are replaced by the D_is in expression (1.3.10). Hence, using the above inequality in this expression and the fact that $P(C_1) = P(D_1)$, we have

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} C_n\right) &= P\left(\bigcup_{n=1}^{\infty} D_n\right) = \lim_{n\to\infty}\left\{P(D_1) + \sum_{j=2}^{n}[P(D_j) - P(D_{j-1})]\right\} \\ &\leq \lim_{n\to\infty} \sum_{j=1}^{n} P(C_j) = \sum_{n=1}^{\infty} P(C_n). \quad \blacksquare \end{aligned}$$

Theorem 1.3.5 gave a general additive law of probability for the union of two events. As the next remark shows, this can be extended to an additive law for an arbitrary union.

Remark 1.3.2 (Inclusion Exclusion Formula)**.** It is easy to show (Exercise 1.3.9) that

$$P(C_1 \cup C_2 \cup C_3) = p_1 - p_2 + p_3,$$

where

$$\begin{aligned} p_1 &= P(C_1) + P(C_2) + P(C_3) \\ p_2 &= P(C_1 \cap C_2) + P(C_1 \cap C_3) + P(C_2 \cap C_3) \\ p_3 &= P(C_1 \cap C_2 \cap C_3). \end{aligned} \tag{1.3.12}$$

This can be generalized to the **inclusion exclusion formula**:

$$P(C_1 \cup C_2 \cup \cdots \cup C_k) = p_1 - p_2 + p_3 - \cdots + (-1)^{k+1} p_k, \tag{1.3.13}$$

where p_i equals the sum of the probabilities of all possible intersections involving i sets.

When $k = 3$, it follows that $p_1 \geq p_2 \geq p_3$, but more generally $p_1 \geq p_2 \geq \cdots \geq p_k$. As shown in Theorem 1.3.7,

$$p_1 = P(C_1) + P(C_2) + \cdots + P(C_k) \geq P(C_1 \cup C_2 \cup \cdots \cup C_k).$$

For $k = 2$, we have

$$1 \geq P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2),$$

which gives **Bonferroni's inequality**,

$$P(C_1 \cap C_2) \geq P(C_1) + P(C_2) - 1, \tag{1.3.14}$$

that is only useful when $P(C_1)$ and $P(C_2)$ are large. The inclusion exclusion formula provides other inequalities that are useful, such as

$$p_1 \geq P(C_1 \cup C_2 \cup \cdots \cup C_k) \geq p_1 - p_2$$

and

$$p_1 - p_2 + p_3 \geq P(C_1 \cup C_2 \cup \cdots \cup C_k) \geq p_1 - p_2 + p_3 - p_4. \quad \blacksquare$$

EXERCISES

1.3.1. A positive integer from one to six is to be chosen by casting a die. Thus the elements c of the sample space $\mathcal{C}$ are 1, 2, 3, 4, 5, 6. Suppose $C_1 = \{1, 3, 5\}$ and $C_2 = \{2, 4, 6\}$. If the probability set function P assigns a probability of 1/6 to each of the elements of $\mathcal{C}$, compute $P(C_1)$, $P(C_2)$, $P(C_1 \cap C_2)$, and $P(C_1 \cup C_2)$.

1.3.2. A random experiment consists of drawing a card from an ordinary deck of 52 playing cards. Let the probability set function P assign a probability of $\frac{1}{52}$ to each of the 52 possible outcomes. Let C_1 denote the collection of the 13 hearts and let C_2 denote the collection of the 4 kings. Compute $P(C_1)$, $P(C_2)$, $P(C_1 \cap C_2)$, and $P(C_1 \cup C_2)$.

1.3.3. A coin is to be tossed as many times as necessary to turn up one head. Thus the elements c of the sample space $\mathcal{C}$ are H, TH, TTH, $TTTH$, and so forth. Let the probability set function P assign to these elements the respective probabilities $\frac{1}{2}$, $\frac{1}{4}$, $\frac{1}{8}$, $\frac{1}{16}$, and so forth. Show that $P(\mathcal{C}) = 1$. Let $C_1 = \{c : c \text{ is } H, TTH, \text{or } TTTTH\}$. Compute $P(C_1)$. Next, suppose that $C_2 = \{c : c \text{ is } H, TH, TTH, \text{or } TTTH\}$. Compute $P(C_2)$, $P(C_1 \cap C_2)$, and $P(C_1 \cup C_2)$.

1.3.4. If the sample space is $\mathcal{C} = C_1 \cup C_2$ and if $P(C_1) = 0.65$ and $P(C_2) = 0.75$, find $P(C_1 \cap C_2)$.

1.3.5. Let the sample space be $\mathcal{C} = \{c : 0 < c < \infty\}$. Let $C \subset \mathcal{C}$ be defined by $C = \{c : 0 < c < 10\}$ and take $P(C) = \int_C \frac{1}{2} e^{-x/2} dx$. Show that $P(\mathcal{C}) = 1$. Evaluate $P(C)$, $P(C^c)$, and $P(C \cap C^c)$.

1.3.6. If the sample space is $\mathcal{C} = \{c : -\infty < c < \infty\}$ and if $C \subset \mathcal{C}$ is a set for which the integral $\int_C e^{-|3x|} dx$ exists, show that this set function is not a probability set function. What constant do we multiply the integrand by to make it a probability set function?

1.3.7. If C_1 and C_2 are subsets of the sample space $\mathcal{C}$, show that

$$P(C_1 \cap C_2) \leq P(C_1) \leq P(C_1 \cup C_2) \leq P(C_1) + P(C_2).$$

1.3.8. Let C_1, C_2, and C_3 be three mutually disjoint subsets of the sample space $\mathcal{C}$. Find $P[(C_1 \cup C_2) \cap C_3]$ and $P(C_1^c \cup C_2^c)$.

1.3.9. Consider Remark 1.3.2.

(a) If C_1, C_2, and C_3 are subsets of $\mathcal{C}$, show that

$$\begin{aligned} P(C_1 \cup C_2 \cup C_3) &= P(C_1) + P(C_2) + P(C_3) - P(C_1 \cap C_2) \\ &\quad - P(C_1 \cap C_3) - P(C_2 \cap C_3) + P(C_1 \cap C_2 \cap C_3). \end{aligned}$$

(b) Now prove the general inclusion exclusion formula given by the expression (1.3.13).

Remark 1.3.3. In order to solve Exercises (1.3.10)–(1.3.19), certain reasonable assumptions must be made. ■

1.3.10. A bowl contains 16 chips, of which 6 are red, 7 are white, and 3 are blue. If four chips are taken at random and without replacement, find the probability that: (a) each of the four chips is red; (b) none of the four chips is red; (c) there is at least one chip of each color.

1.3.11. A person has purchased 10 of 1000 tickets sold in a certain raffle. To determine the five prize winners, five tickets are to be drawn at random and without replacement. Compute the probability that this person wins at least one prize.
Hint: First compute the probability that the person does not win a prize.

1.3.12. Compute the probability of being dealt at random and without replacement a 13-card bridge hand consisting of: (a) 6 spades, 4 hearts, 2 diamonds, and 1 club; (b) 13 cards of the same suit.

1.3.13. Three distinct integers are chosen at random from the first 20 positive integers. Compute the probability that: (a) their sum is even; (b) their product is even.

1.3.14. There are five red chips and three blue chips in a bowl. The red chips are numbered 1, 2, 3, 4, 5, respectively, and the blue chips are numbered 1, 2, 3, respectively. If two chips are to be drawn at random and without replacement, find the probability that these chips have either the same number or the same color.

1.3.15. In a lot of 50 light bulbs, there are 2 bad bulbs. An inspector examines five bulbs, which are selected at random and without replacement.

(a) Find the probability of at least one defective bulb among the five.

(b) How many bulbs should be examined so that the probability of finding at least one bad bulb exceeds $\frac{1}{2}$?

1.3.16. For the birthday problem, Example 1.3.3, use the given R function `bday` to determine the value of n so that $p(n) \geq 0.5$ and $p(n-1) < 0.5$, where $p(n)$ is the probability that at least two people in the room of n people have the same birthday.

1.3.17. If $C_1, \ldots, C_k$ are k events in the sample space $\mathcal{C}$, show that the probability that at least one of the events occurs is one minus the probability that none of them occur; i.e.,

$$P(C_1 \cup \cdots \cup C_k) = 1 - P(C_1^c \cap \cdots \cap C_k^c). \tag{1.3.15}$$

1.3.18. A secretary types three letters and the three corresponding envelopes. In a hurry, he places at random one letter in each envelope. What is the probability that at least one letter is in the correct envelope? *Hint:* Let C_i be the event that the ith letter is in the correct envelope. Expand $P(C_1 \cup C_2 \cup C_3)$ to determine the probability.

1.3.19. Consider poker hands drawn from a well-shuffled deck as described in Example 1.3.4. Determine the probability of a full house, i.e, three of one kind and two of another.

1.3.20. Prove expression (1.3.9).

1.3.21. Suppose the experiment is to choose a real number at random in the interval $(0, 1)$. For any subinterval $(a, b) \subset (0, 1)$, it seems reasonable to assign the probability $P[(a, b)] = b - a$; i.e., the probability of selecting the point from a subinterval is directly proportional to the length of the subinterval. If this is the case, choose an appropriate sequence of subintervals and use expression (1.3.9) to show that $P[\{a\}] = 0$, for all $a \in (0, 1)$.

1.3.22. Consider the events C_1, C_2, C_3.

(a) Suppose C_1, C_2, C_3 are mutually exclusive events. If $P(C_i) = p_i$, $i = 1, 2, 3$, what is the restriction on the sum $p_1 + p_2 + p_3$?

(b) In the notation of part (a), if $p_1 = 4/10$, $p_2 = 3/10$, and $p_3 = 5/10$, are C_1, C_2, C_3 mutually exclusive?

For the last two exercises it is assumed that the reader is familiar with σ-fields.

1.3.23. Suppose $\mathcal{D}$ is a nonempty collection of subsets of $\mathcal{C}$. Consider the collection of events

$$\mathcal{B} = \cap\{\mathcal{E} : \mathcal{D} \subset \mathcal{E} \text{ and } \mathcal{E} \text{ is a } \sigma\text{-field}\}.$$

Note that $\phi \in \mathcal{B}$ because it is in each σ-field, and, hence, in particular, it is in each σ-field $\mathcal{E} \supset \mathcal{D}$. Continue in this way to show that $\mathcal{B}$ is a σ-field.

1.3.24. Let $\mathcal{C} = R$, where R is the set of all real numbers. Let $\mathcal{I}$ be the set of all open intervals in R. The Borel σ-field on the real line is given by

$$\mathcal{B}_0 = \cap\{\mathcal{E} : \mathcal{I} \subset \mathcal{E} \text{ and } \mathcal{E} \text{ is a } \sigma\text{-field}\}.$$

By definition, $\mathcal{B}_0$ contains the open intervals. Because $[a, \infty) = (-\infty, a)^c$ and $\mathcal{B}_0$ is closed under complements, it contains all intervals of the form $[a, \infty)$, for $a \in R$. Continue in this way and show that $\mathcal{B}_0$ contains all the closed and half-open intervals of real numbers.

1.4 Conditional Probability and Independence

In some random experiments, we are interested only in those outcomes that are elements of a subset A of the sample space $\mathcal{C}$. This means, for our purposes, that the sample space is effectively the subset A. We are now confronted with the problem of defining a probability set function with A as the "new" sample space.

Let the probability set function $P(A)$ be defined on the sample space $\mathcal{C}$ and let A be a subset of $\mathcal{C}$ such that $P(A) > 0$. We agree to consider only those outcomes of the random experiment that are elements of A; in essence, then, we take A to be a sample space. Let B be another subset of $\mathcal{C}$. How, relative to the new sample space A, do we want to define the probability of the event B? Once defined, this probability is called the *conditional probability* of the event B, relative to the hypothesis of the event A, or, more briefly, the conditional probability of B, given A. Such a conditional probability is denoted by the symbol $P(B|A)$. The "|" in this symbol is usually read as "given." We now return to the question that was raised about the definition of this symbol. Since A is now the sample space, the only elements of B that concern us are those, if any, that are also elements of A, that is, the elements of $A \cap B$. It seems desirable, then, to define the symbol $P(B|A)$ in such a way that

$$P(A|A) = 1 \quad \text{and} \quad P(B|A) = P(A \cap B|A).$$

Moreover, from a relative frequency point of view, it would seem logically inconsistent if we did not require that the ratio of the probabilities of the events $A \cap B$ and A, relative to the space A, be the same as the ratio of the probabilities of these events relative to the space $\mathcal{C}$; that is, we should have

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

These three desirable conditions imply that the relation conditional probability is reasonably defined as

Definition 1.4.1 (Conditional Probability)**.** *Let B and A be events with $P(A) > 0$. Then we defined the* **conditional probability** *of B given A as*

$$P(B|A) = \frac{P(A \cap B)}{P(A)}. \quad \blacksquare \tag{1.4.1}$$

Moreover, we have

1. $P(B|A) \geq 0$.
2. $P(A|A) = 1$.
3. $P\left(\cup_{n=1}^{\infty} B_n|A\right) = \sum_{n=1}^{\infty} P(B_n|A)$, provided that $B_1, B_2, \ldots$ are mutually exclusive events.

Properties (1) and (2) are evident. For Property (3), suppose the sequence of events $B_1, B_2, \ldots$ is mutually exclusive. It follows that also $(B_n \cap A) \cap (B_m \cap A) = \phi$, $n \neq m$. Using this and the first of the distributive laws (1.2.5) for countable unions, we have

$$\begin{aligned} P(\cup_{n=1}^{\infty} B_n|A) &= \frac{P[\cup_{n=1}^{\infty}(B_n \cap A)]}{P(A)} \\ &= \sum_{n=1}^{\infty} \frac{P[B_n \cap A]}{P(A)} \\ &= \sum_{n=1}^{\infty} P[B_n|A]. \end{aligned}$$

Properties (1)–(3) are precisely the conditions that a probability set function must satisfy. Accordingly, $P(B|A)$ is a probability set function, defined for subsets of A. It may be called the conditional probability set function, relative to the hypothesis A, or the conditional probability set function, given A. It should be noted that this conditional probability set function, given A, is defined at this time only when $P(A) > 0$.

Example 1.4.1. A hand of five cards is to be dealt at random without replacement from an ordinary deck of 52 playing cards. The conditional probability of an all-spade hand (B), relative to the hypothesis that there are at least four spades in the hand (A), is, since $A \cap B = B$,

$$P(B|A) = \frac{P(B)}{P(A)} = \frac{\binom{13}{5}/\binom{52}{5}}{\left[\binom{13}{4}\binom{39}{1} + \binom{13}{5}\right]/\binom{52}{5}}$$
$$= \frac{\binom{13}{5}}{\binom{13}{4}\binom{39}{1} + \binom{13}{5}} = 0.0441.$$

Note that this is not the same as drawing for a spade to complete a flush in draw poker; see Exercise 1.4.3. ■

From the definition of the conditional probability set function, we observe that

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

This relation is frequently called the **multiplication rule** for probabilities. Sometimes, after considering the nature of the random experiment, it is possible to make reasonable assumptions so that both $P(A)$ and $P(B|A)$ can be assigned. Then $P(A \cap B)$ can be computed under these assumptions. This is illustrated in Examples 1.4.2 and 1.4.3.

Example 1.4.2. A bowl contains eight chips. Three of the chips are red and the remaining five are blue. Two chips are to be drawn successively, at random and without replacement. We want to compute the probability that the first draw results in a red chip (A) and that the second draw results in a blue chip (B). It is reasonable to assign the following probabilities:

$$P(A) = \tfrac{3}{8} \quad \text{and} \quad P(B|A) = \tfrac{5}{7}.$$

Thus, under these assignments, we have $P(A \cap B) = (\frac{3}{8})(\frac{5}{7}) = \frac{15}{56} = 0.2679$. ■

Example 1.4.3. From an ordinary deck of playing cards, cards are to be drawn successively, at random and without replacement. The probability that the third spade appears on the sixth draw is computed as follows. Let A be the event of two spades in the first five draws and let B be the event of a spade on the sixth draw. Thus the probability that we wish to compute is $P(A \cap B)$. It is reasonable to take

$$P(A) = \frac{\binom{13}{2}\binom{39}{3}}{\binom{52}{5}} = 0.2743 \quad \text{and} \quad P(B|A) = \frac{11}{47} = 0.2340.$$

The desired probability $P(A \cap B)$ is then the product of these two numbers, which to four places is 0.0642. ■

The multiplication rule can be extended to three or more events. In the case of three events, we have, by using the multiplication rule for two events,

$$P(A \cap B \cap C) = P[(A \cap B) \cap C]$$
$$= P(A \cap B)P(C|A \cap B).$$

But $P(A \cap B) = P(A)P(B|A)$. Hence, provided $P(A \cap B) > 0$,

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

This procedure can be used to extend the multiplication rule to four or more events. The general formula for k events can be proved by mathematical induction.

Example 1.4.4. Four cards are to be dealt successively, at random and without replacement, from an ordinary deck of playing cards. The probability of receiving a spade, a heart, a diamond, and a club, in that order, is $(\frac{13}{52})(\frac{13}{51})(\frac{13}{50})(\frac{13}{49}) = 0.0044$. This follows from the extension of the multiplication rule. ∎

Consider k mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_k$ such that $P(A_i) > 0$, $i = 1, 2, \ldots, k$; i.e., $A_1, A_2, \ldots, A_k$ form a partition of $\mathcal{C}$. Here the events $A_1, A_2, \ldots, A_k$ do *not* need to be equally likely. Let B be another event such that $P(B) > 0$. Thus B occurs with one and only one of the events $A_1, A_2, \ldots, A_k$; that is,

$$\begin{aligned} B &= B \cap (A_1 \cup A_2 \cup \cdots A_k) \\ &= (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_k). \end{aligned}$$

Since $B \cap A_i$, $i = 1, 2, \ldots, k$, are mutually exclusive, we have

$$P(B) = P(B \cap A_1) + P(B \cap A_2) + \cdots + P(B \cap A_k).$$

However, $P(B \cap A_i) = P(A_i)P(B|A_i)$, $i = 1, 2, \ldots, k$; so

$$\begin{aligned} P(B) &= P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_k)P(B|A_k) \\ &= \sum_{i=1}^{k} P(A_i)P(B|A_i). \end{aligned} \tag{1.4.2}$$

This result is sometimes called the **law of total probability** and it leads to the following important theorem.

Theorem 1.4.1 (Bayes)**.** *Let $A_1, A_2, \ldots, A_k$ be events such that $P(A_i) > 0$, $i = 1, 2, \ldots, k$. Assume further that $A_1, A_2, \ldots, A_k$ form a partition of the sample space $\mathcal{C}$. Let B be any event. Then*

$$P(A_j|B) = \frac{P(A_j)P(B|A_j)}{\sum_{i=1}^{k} P(A_i)P(B|A_i)}, \tag{1.4.3}$$

Proof: Based on the definition of conditional probability, we have

$$P(A_j|B) = \frac{P(B \cap A_j)}{P(B)} = \frac{P(A_j)P(B|A_j)}{P(B)}.$$

The result then follows by the law of total probability, (1.4.2). ∎

This theorem is the well-known **Bayes' Theorem**. This permits us to calculate the conditional probability of A_j, given B, from the probabilities of $A_1, A_2, \ldots, A_k$ and the conditional probabilities of B, given A_i, $i = 1, 2, \ldots, k$. The next three examples illustrate the usefulness of Bayes Theorem to determine probabilities.

Example 1.4.5. Say it is known that bowl A_1 contains three red and seven blue chips and bowl A_2 contains eight red and two blue chips. All chips are identical in size and shape. A die is cast and bowl A_1 is selected if five or six spots show on the side that is up; otherwise, bowl A_2 is selected. For this situation, it seems reasonable to assign $P(A_1) = \frac{2}{6}$ and $P(A_2) = \frac{4}{6}$. The selected bowl is handed to another person and one chip is taken at random. Say that this chip is red, an event which we denote by B. By considering the contents of the bowls, it is reasonable to assign the conditional probabilities $P(B|A_1) = \frac{3}{10}$ and $P(B|A_2) = \frac{8}{10}$. Thus the conditional probability of bowl A_1, given that a red chip is drawn, is

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1)P(B|A_1)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)} \\ &= \frac{(\frac{2}{6})(\frac{3}{10})}{(\frac{2}{6})(\frac{3}{10}) + (\frac{4}{6})(\frac{8}{10})} = \frac{3}{19}. \end{aligned}$$

In a similar manner, we have $P(A_2|B) = \frac{16}{19}$. ■

In Example 1.4.5, the probabilities $P(A_1) = \frac{2}{6}$ and $P(A_2) = \frac{4}{6}$ are called **prior probabilities** of A_1 and A_2, respectively, because they are known to be due to the random mechanism used to select the bowls. After the chip is taken and is observed to be red, the conditional probabilities $P(A_1|B) = \frac{3}{19}$ and $P(A_2|B) = \frac{16}{19}$ are called **posterior probabilities**. Since A_2 has a larger proportion of red chips than does A_1, it appeals to one's intuition that $P(A_2|B)$ should be larger than $P(A_2)$ and, of course, $P(A_1|B)$ should be smaller than $P(A_1)$. That is, intuitively the chances of having bowl A_2 are better once that a red chip is observed than before a chip is taken. Bayes' theorem provides a method of determining exactly what those probabilities are.

Example 1.4.6. Three plants, A_1, A_2, and A_3, produce respectively, 10%, 50%, and 40% of a company's output. Although plant A_1 is a small plant, its manager believes in high quality and only 1% of its products are defective. The other two, A_2 and A_3, are worse and produce items that are 3% and 4% defective, respectively. All products are sent to a central warehouse. One item is selected at random and observed to be defective, say event B. The conditional probability that it comes from plant A_1 is found as follows. It is natural to assign the respective prior probabilities of getting an item from the plants as $P(A_1) = 0.1$, $P(A_2) = 0.5$, and $P(A_3) = 0.4$, while the conditional probabilities of defective items are $P(B|A_1) = 0.01$, $P(B|A_2) = 0.03$, and $P(B|A_3) = 0.04$. Thus the posterior probability of A_1, given a defective, is

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{(0.10)(0.01)}{(0.1)(0.01) + (0.5)(0.03) + (0.4)(0.04)} = \frac{1}{32}.$$

This is much smaller than the prior probability $P(A_1) = \frac{1}{10}$. This is as it should be because the fact that the item is defective decreases the chances that it comes from the high-quality plant A_1. ■

Example 1.4.7. Suppose we want to investigate the percentage of abused children in a certain population. The events of interest are: a child is abused (A) and its complement a child is not abused ($N = A^c$). For the purposes of this example, we assume that $P(A) = 0.01$ and, hence, $P(N) = 0.99$. The classification as to whether a child is abused or not is based upon a doctor's examination. Because doctors are not perfect, they sometimes classify an abused child (A) as one that is not abused (N_D, where N_D means classified as not abused by a doctor). On the other hand, doctors sometimes classify a nonabused child (N) as abused (A_D). Suppose these error rates of misclassification are $P(N_D \mid A) = 0.04$ and $P(A_D \mid N) = 0.05$; thus the probabilities of correct decisions are $P(A_D \mid A) = 0.96$ and $P(N_D \mid N) = 0.95$. Let us compute the probability that a child taken at random is classified as abused by a doctor. Because this can happen in two ways, $A \cap A_D$ or $N \cap A_D$, we have

$$P(A_D) = P(A_D \mid A)P(A) + P(A_D \mid N)P(N) = (0.96)(0.01) + (0.05)(0.99) = 0.0591,$$

which is quite high relative to the probability of an abused child, 0.01. Further, the probability that a child is abused when the doctor classified the child as abused is

$$P(A \mid A_D) = \frac{P(A \cap A_D)}{P(A_D)} = \frac{(0.96)(0.01)}{0.0591} = 0.1624,$$

which is quite low. In the same way, the probability that a child is not abused when the doctor classified the child as abused is 0.8376, which is quite high. The reason that these probabilities are so poor at recording the true situation is that the doctors' error rates are so high relative to the fraction 0.01 of the population that is abused. An investigation such as this would, hopefully, lead to better training of doctors for classifying abused children. See also Exercise 1.4.17. ■

1.4.1 Independence

Sometimes it happens that the occurrence of event A does not change the probability of event B; that is, when $P(A) > 0$,

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

In this case, we say that the events A and B are *independent*. Moreover, the multiplication rule becomes

$$P(A \cap B) = P(A)P(B|A) = P(A)P(B). \tag{1.4.4}$$

This, in turn, implies, when $P(B) > 0$, that

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

Note that if $P(A) > 0$ and $P(B) > 0$, then by the above discussion, independence is equivalent to

$$P(A \cap B) = P(A)P(B). \tag{1.4.5}$$

What if either $P(A) = 0$ or $P(B) = 0$? In either case, the right side of (1.4.5) is 0. However, the left side is 0 also because $A \cap B \subset A$ and $A \cap B \subset B$. Hence, we take Equation (1.4.5) as our formal definition of independence; that is,

Definition 1.4.2. *Let A and B be two events. We say that A and B are* **independent** *if $P(A \cap B) = P(A)P(B)$.* ■

Suppose A and B are independent events. Then the following three pairs of events are independent: A^c and B, A and B^c, and A^c and B^c. We show the first and leave the other two to the exercises; see Exercise 1.4.11. Using the disjoint union, $B = (A^c \cap B) \cup (A \cap B)$, we have

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = [1 - P(A)]P(B) = P(A^c)P(B). \tag{1.4.6}$$

Hence, A^c and B are also independent.

Remark 1.4.1. Events that are *independent* are sometimes called *statistically independent*, *stochastically independent*, or *independent in a probability sense*. In most instances, we use *independent* without a modifier if there is no possibility of misunderstanding. ■

Example 1.4.8. A red die and a white die are cast in such a way that the numbers of spots on the two sides that are up are independent events. If A represents a four on the red die and B represents a three on the white die, with an equally likely assumption for each side, we assign $P(A) = \frac{1}{6}$ and $P(B) = \frac{1}{6}$. Thus, from independence, the probability of the ordered pair (red = 4, white = 3) is

$$P[(4,3)] = (\tfrac{1}{6})(\tfrac{1}{6}) = \tfrac{1}{36}.$$

The probability that the sum of the up spots of the two dice equals seven is

$$\begin{aligned} &P[(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)] \\ &\quad = \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) + \left(\tfrac{1}{6}\right)\left(\tfrac{1}{6}\right) = \tfrac{6}{36}. \end{aligned}$$

In a similar manner, it is easy to show that the probabilities of the sums of the upfaces $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$ are, respectively,

$$\tfrac{1}{36}, \tfrac{2}{36}, \tfrac{3}{36}, \tfrac{4}{36}, \tfrac{5}{36}, \tfrac{6}{36}, \tfrac{5}{36}, \tfrac{4}{36}, \tfrac{3}{36}, \tfrac{2}{36}, \tfrac{1}{36}. \quad ■$$

Suppose now that we have three events, A_1, A_2, and A_3. We say that they are **mutually independent** if and only if they are *pairwise independent*:

$$\begin{aligned} &P(A_1 \cap A_3) = P(A_1)P(A_3), \quad P(A_1 \cap A_2) = P(A_1)P(A_2), \\ &P(A_2 \cap A_3) = P(A_2)P(A_3), \end{aligned}$$

and

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3).$$

More generally, the n events $A_1, A_2, \ldots, A_n$ are **mutually independent** if and only if for every collection of k of these events, $2 \leq k \leq n$, and for every permutation $d_1, d_2, \ldots, d_k$ of $1, 2, \ldots, k$,

$$P(A_{d_1} \cap A_{d_2} \cap \cdots \cap A_{d_k}) = P(A_{d_1})P(A_{d_2}) \cdots P(A_{d_k}).$$

In particular, if $A_1, A_2, \ldots, A_n$ are mutually independent, then

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

Also, as with two sets, many combinations of these events and their complements are independent, such as

1. The events A_1^c and $A_2 \cup A_3^c \cup A_4$ are independent,
2. The events $A_1 \cup A_2^c$, A_3^c and $A_4 \cap A_5^c$ are mutually independent.

If there is no possibility of misunderstanding, *independent* is often used without the modifier *mutually* when considering more than two events.

Example 1.4.9. Pairwise independence does not imply mutual independence. As an example, suppose we twice spin a fair spinner with the numbers 1, 2, 3, and 4. Let A_1 be the event that the sum of the numbers spun is 5, let A_2 be the event that the first number spun is a 1, and let A_3 be the event that the second number spun is a 4. Then $P(A_i) = 1/4$, $i = 1, 2, 3$, and for $i \neq j$, $P(A_i \cap A_j) = 1/16$. So the three events are pairwise independent. But $A_1 \cap A_2 \cap A_3$ is the event that $(1, 4)$ is spun, which has probability $1/16 \neq 1/64 = P(A_1)P(A_2)P(A_3)$. Hence the events A_1, A_2, and A_3 are not mutually independent. ■

We often perform a sequence of random experiments in such a way that the events associated with one of them are independent of the events associated with the others. For convenience, we refer to these events as as outcomes of *independent experiments*, meaning that the respective events are independent. Thus we often refer to independent flips of a coin or independent casts of a die or, more generally, independent trials of some given random experiment.

Example 1.4.10. A coin is flipped independently several times. Let the event A_i represent a head (H) on the ith toss; thus A_i^c represents a tail (T). Assume that A_i and A_i^c are equally likely; that is, $P(A_i) = P(A_i^c) = \frac{1}{2}$. Thus the probability of an ordered sequence like HHTH is, from independence,

$$P(A_1 \cap A_2 \cap A_3^c \cap A_4) = P(A_1)P(A_2)P(A_3^c)P(A_4) = (\tfrac{1}{2})^4 = \tfrac{1}{16}.$$

Similarly, the probability of observing the first head on the third flip is

$$P(A_1^c \cap A_2^c \cap A_3) = P(A_1^c)P(A_2^c)P(A_3) = (\tfrac{1}{2})^3 = \tfrac{1}{8}.$$

Also, the probability of getting at least one head on four flips is

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3 \cup A_4) &= 1 - P[(A_1 \cup A_2 \cup A_3 \cup A_4)^c] \\ &= 1 - P(A_1^c \cap A_2^c \cap A_3^c \cap A_4^c) \\ &= 1 - \left(\tfrac{1}{2}\right)^4 = \tfrac{15}{16}. \end{aligned}$$

See Exercise 1.4.13 to justify this last probability. ■

Example 1.4.11. A computer system is built so that if component K_1 fails, it is bypassed and K_2 is used. If K_2 fails, then K_3 is used. Suppose that the probability that K_1 fails is 0.01, that K_2 fails is 0.03, and that K_3 fails is 0.08. Moreover, we can assume that the failures are mutually independent events. Then the probability of failure of the system is

$$(0.01)(0.03)(0.08) = 0.000024,$$

as all three components would have to fail. Hence, the probability that the system does not fail is $1 - 0.000024 = 0.999976$. ■

1.4.2 Simulations

Many of the exercises at the end of this section are designed to aid the reader in his/her understanding of the concepts of conditional probability and independence. With diligence and patience, the reader will derive the exact answer. Many real life problems, though, are too complicated to allow for exact derivation. In such cases, scientists often turn to computer simulations to estimate the answer. As an example, suppose for an experiment, we want to obtain $P(A)$ for some event A. A program is written that performs one trial (one simulation) of the experiment and it records whether or not A occurs. We then obtain n independent simulations (runs) of the program. Denote by $\hat{p}_n$ the proportion of these n simulations in which A occurred. Then $\hat{p}_n$ is our estimate of the $P(A)$. Besides the estimation of $P(A)$, we also obtain an error of estimation given by $1.96 * \sqrt{\hat{p}_n(1-\hat{p}_n)/n}$. As we discuss theoretically in Chapter 4, we are 95% confident that $P(A)$ lies in the interval

$$\hat{p}_n \pm 1.96\sqrt{\frac{\hat{p}_n(1-\hat{p}_n)}{n}}. \tag{1.4.7}$$

In Chapter 4, we call this interval a 95% **confidence interval** for $P(A)$. For now, we make use of this confidence interval for our simulations.

Example 1.4.12. As an example, consider the game:

> Person A tosses a coin and then person B rolls a die. This is repeated independently until a head or one of the numbers $1, 2, 3, 4$ appears, at which time the game is stopped. Person A wins with the head and B wins with one of the numbers $1, 2, 3, 4$. Compute the probability $P(A)$ that person A wins the game.

For an exact derivation, notice that it is implicit in the statement A wins the game that the game is completed. Using abbreviated notation, the game is completed if H or $T\{1, \ldots, 4\}$ occurs. Using independence, the probability that A wins is thus the conditional probability $(1/2)/[(1/2) + (1/2)(4/6)] = 3/5$.

The following R function, `abgame`, simulates the problem. This function can be downloaded and sourced at the site discussed in the Preface. The first line of the program sets up the draws for persons A and B, respectively. The second line sets up a flag for the while loop and the returning values, `Awin` and `Bwin` are initialized at 0. The command `sample(rngA,1,pr=pA)` draws a sample of size 1 from `rngA` with pmf `pA`. Each execution of the while loop returns one complete game. Further, the executions are independent of one another.

```
abgame <- function(){
   rngA <- c(0,1);  pA <- rep(1/2,2); rngB <- 1:6; pB <- rep(1/6,6)
   ic <- 0; Awin <- 0; Bwin <- 0
   while(ic == 0){
       x <- sample(rngA,1,pr=pA)
       if(x==1){
           ic <- 1; Awin <- 1
       } else {
           y <- sample(rngB,1,pr=pB)
           if(y <= 4){ic <- 1; Bwin <- 1}
       }
   }
return(c(Awin,Bwin))
}
```

Notice that one and only one of `Awin` or `Bwin` receives the value 1 depending on whether or not A or B wins. The next R segment simulates the game 10,000 times and computes the estimate that A wins along with the error of estimation.

```
ind <- 0; nsims <- 10000
for(i in 1:nsims){
     seeA <- abgame  ()
     if(seeA[1] == 1){ind <- ind + 1}
}
estpA <- ind/nsims
err <- 1.96*sqrt(estpA*(1-estpA)/nsims)
estpA; err
```

An execution of this code resulted in `estpA = 0.6001` and `err = 0.0096`. As noted above the probability that A wins is 0.6 which is in the interval 0.6001 ± 0.0096. As discussed in Chapter 4, we expect this to occur 95% of the time when using such a confidence interval. ■

EXERCISES

1.4.1. If $P(A_1) > 0$ and if $A_2, A_3, A_4, \ldots$ are mutually disjoint sets, show that

$$P(A_2 \cup A_3 \cup \cdots | A_1) = P(A_2|A_1) + P(A_3|A_1) + \cdots .$$

1.4.2. Assume that $P(A_1 \cap A_2 \cap A_3) > 0$. Prove that

$$P(A_1 \cap A_2 \cap A_3 \cap A_4) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3).$$

1.4.3. Suppose we are playing draw poker. We are dealt (from a well-shuffled deck) five cards, which contain four spades and another card of a different suit. We decide to discard the card of a different suit and draw one card from the remaining cards to complete a flush in spades (all five cards spades). Determine the probability of completing the flush.

1.4.4. From a well-shuffled deck of ordinary playing cards, four cards are turned over one at a time without replacement. What is the probability that the spades and red cards alternate?

1.4.5. Twelve marbles are drawn at random and without replacement from an urn containing four red, four blue, four yellow, and 24 green marbles of the same shapes and sizes. Find the conditional probability that there are at least three red marbles drawn given that at least two red marbles are drawn.

1.4.6. A drawer contains 10 different pairs of socks. If five socks are taken at random and without replacement, compute the probability that there is at least one matching pair among these five socks. *Hint*: Compute the probability that there is not a matching pair.

1.4.7. A pair of dice is cast until either the sum of seven or eight appears.

(a) Show that the probability of a seven before an eight is 6/11.

(b) Next, this pair of dice is cast until a seven appears twice or until each of a six and eight has appeared at least once. Show that the probability of the six and eight occurring before two sevens is 0.546.

1.4.8. In a certain factory, machines I, II, and III are all producing springs of the same length. Machines I, II, and III produce 3%, 2%, and 5% defective springs, respectively. Of the total production of springs in the factory, Machine I produces 55%, Machine II produces 30%, and Machine III produces 15%.

(a) If one spring is selected at random from the total springs produced in a given day, determine the probability that it is defective.

(b) Given that the selected spring is defective, find the conditional probability that it was produced by Machine II.

1.4.9. Bowl I contains four red chips and six blue chips. Six of these 10 chips are selected at random and without replacement and put in bowl II, which was originally empty. One chip is then drawn at random from bowl II. Given that this chip is blue, find the conditional probability that two red chips and four blue chips are transferred from bowl I to bowl II.

1.4.10. In an office there are two boxes of thumb drives: Box A_1 contains five 100 GB drives and five 500 GB drives, and box A_2 contains six 100 GB drives and four 500 GB drives. A person is handed a box at random with prior probabilities $P(A_1) = \frac{1}{4}$ and $(A_2) = \frac{3}{4}$, possibly due to the boxes' respective locations. A drive is then selected at random and the event B occurs if it is a 500 GB drive. Using an equally likely assumption for each drive in the selected box, compute $P(A_1|B)$ and $P(A_2|B)$.

1.4.11. Suppose A and B are independent events. In expression (1.4.6) we showed that A^c and B are independent events. Show similarly that the following pairs of events are also independent: (a) A and B^c and (b) A^c and B^c.

1.4.12. Let C_1 and C_2 be independent events with $P(C_1) = 0.2$ and $P(C_2) = 0.5$ Compute (a) $P(C_1 \cap C_2)$, (b) $P(C_1 \cup C_2)$, and (c) $P(C_1 \cup C_2^c)$.

1.4.13. Generalize Exercise 1.2.5 to obtain

$$(C_1 \cup C_2 \cup \cdots \cup C_k)^c = C_1^c \cap C_2^c \cap \cdots \cap C_k^c.$$

Say that $C_1, C_2, \ldots, C_k$ are independent events that have respective probabilities $p_1, p_2, \ldots, p_k$. Argue that the probability of at least one of $C_1, C_2, \ldots, C_k$ is equal to

$$1 - (1 - p_1)(1 - p_2) \cdots (1 - p_k).$$

1.4.14. Each of four persons fires one shot at a target. Let C_k denote the event that the target is hit by person k, $k = 1, 2, 3, 4$. If C_1, C_2, C_3, C_4 are independent and if $P(C_1) = 0.6$, $P(C_2) = P(C_3) = 0.9$, and $P(C_4) = 0.3$, compute the probability that (a) all of them hit the target; (b) exactly one hits the target; (c) no one hits the target; (d) at least one hits the target.

1.4.15. A bowl contains three red (R) balls and seven white (W) balls of exactly the same size and shape. Select balls successively at random and with replacement so that the events of white on the first trial, white on the second, and so on, can be assumed to be independent. In four trials, make certain assumptions and compute the probabilities of the following ordered sequences: (a) WWRW; (b) RWWW; (c) WWWR; and (d) WRWW. Compute the probability of exactly one red ball in the four trials.

1.4.16. A coin is tossed two independent times, each resulting in a tail (T) or a head (H). The sample space consists of four ordered pairs: TT, TH, HT, HH. Making certain assumptions, compute the probability of each of these ordered pairs. What is the probability of at least one head?

1.4.17. For Example 1.4.7, obtain the following probabilities. Explain what they mean in terms of the problem.

(a) $P(N_D)$.

(b) $P(N \mid A_D)$.

(c) $P(A \mid N_D)$.

(d) $P(N \mid N_D)$.

1.4.18. A die is cast independently until the first 6 appears. If the casting stops on an odd number of times, Bob wins; otherwise, Joe wins.

(a) Assuming the die is fair, what is the probability that Bob wins?

(b) Let p denote the probability of a 6. Show that the game favors Bob, for all p, $0 < p < 1$.

1.4.19. Balls are drawn at random and with replacement from a box of 20 red and 30 green balls until a red ball appears.

(a) What is the probability that at least four draws are necessary?

(b) Same as part (a), except the balls are drawn without replacement.

1.4.20. A person answers each of two multiple choice questions at random. If there are four possible choices on each question, what is the conditional probability that both answers are correct given that at least one is correct?

1.4.21. Suppose a fair 6-sided die is rolled six independent times. A match occurs if side i is observed on the ith trial, $i = 1, \ldots, 6$.

(a) What is the probability of at least one match on the six rolls? *Hint:* Let C_i be the event of a match on the ith trial and use Exercise 1.4.13 to determine the desired probability.

(b) Extend part (a) to a fair n-sided die with n independent rolls. Then determine the limit of the probability as $n \to \infty$.

1.4.22. Players A and B play a sequence of independent games. Player A throws a die first and wins on a "six." If he fails, B throws and wins on a "five" or "six." If he fails, A throws and wins on a "four," "five," or "six." And so on. Find the probability of each player winning the sequence.

1.4.23. Let C_1, C_2, C_3 be independent events with probabilities 1/3, 2/5, 3/7, respectively. Compute $P(C_1 \cup C_2 \cup C_3)$.

1.4.24. From a bowl containing two red, four white, and six blue chips, select three at random and without replacement. Compute the conditional probability of zero red, one white, and two blue chips given that there are at least two blue chips in this sample of three chips.

1.4.25. Let the three mutually independent events C_1, C_2, and C_3 be such that $P(C_1) = P(C_2) = \frac{1}{3}$ and $P(C_3) = \frac{1}{4}$. Find $P[(C_1^c \cap C_2^c) \cup C_3]$.

1.4.26. Each bag in a large box contains 20 tulip bulbs. It is known that 70% of the bags contain bulbs for 8 red and 12 yellow tulips, while the remaining 30% of the bags contain bulbs for 10 red and 10 yellow tulips. A bag is selected at random and a bulb taken at random from this bag is planted.

(a) What is the probability that it will be a yellow tulip?

(b) Given that it is yellow, what is the conditional probability it comes from a bag that contained 8 red and 12 yellow bulbs?

1.4.27. The following game is played. The player randomly draws from the set of integers $\{1, 2, \ldots, 20\}$. Let x denote the number drawn. Next the player draws at random from the set $\{x, \ldots, 25\}$. If on this second draw, he draws a number greater than 21 he wins; otherwise, he loses.

(a) Determine the sum that gives the probability that the player wins.

(b) Write and run a line of R code that computes the probability that the player wins.

(c) Write an R function that simulates the game and returns whether or not the player wins.

(d) Do 10,000 simulations of your program in Part (c). Obtain the estimate and confidence interval, (1.4.7), for the probability that the player wins. Does your interval trap the true probability?

1.4.28. A bowl contains 9 chips numbered $1, 2, \ldots, 9$, respectively. Five chips are drawn at random, one at a time, and without replacement. What is the probability that two odd-numbered chips are drawn and they occur on even-numbered draws?

1.4.29. A person bets 1 dollar to b dollars that he can draw two cards from an ordinary deck of cards without replacement and that they will be of the same suit. Find b so that the bet is fair.

1.4.30 (Monte Hall Problem)**.** Suppose there are three curtains. Behind one curtain there is a nice prize, while behind the other two there are worthless prizes. A contestant selects one curtain at random, and then Monte Hall opens one of the other two curtains to reveal a worthless prize. Hall then expresses the willingness to trade the curtain that the contestant has chosen for the other curtain that has not been opened. Should the contestant switch curtains or stick with the one that she has? To answer the question, determine the probability that she wins the prize if she switches.

1.4.31. A French nobleman, Chevalier de Méré, had asked a famous mathematician, Pascal, to explain why the following two probabilities were different (the difference had been noted from playing the game many times): (1) at least one six in four independent casts of a six-sided die; (2) at least a pair of sixes in 24 independent casts of a pair of dice. From proportions it seemed to de Méré that the probabilities should be the same. Compute the probabilities of (1) and (2).

1.4.32. Hunters A and B shoot at a target; the probabilities of hitting the target are p_1 and p_2, respectively. Assuming independence, can p_1 and p_2 be selected so that

$$P(\text{zero hits}) = P(\text{one hit}) = P(\text{two hits})?$$

1.4.33. At the beginning of a study of individuals, 10% were classified as heavy smokers, 25% were classified as light smokers, and 65% were classified as nonsmokers. In the five-year study, it was determined that the death rates of the heavy and light smokers were six and three times that of the nonsmokers, respectively. A randomly selected participant died over the five-year period: calculate the probability that the participant was a nonsmoker.

1.4.34. A chemist wishes to detect an impurity in a certain compound that she is making. There is a test that detects an impurity with probability 0.80; however, this test indicates that an impurity is there when it is not about 2% of the time. The chemist produces compounds with the impurity about 25% of the time. A compound is selected at random from the chemist's output. The test indicates that an impurity is present. What is the conditional probability that the compound actually has the impurity?

1.5 Random Variables

The reader perceives that a sample space $\mathcal{C}$ may be tedious to describe if the elements of $\mathcal{C}$ are not numbers. We now discuss how we may formulate a rule, or a set of rules, by which the elements c of $\mathcal{C}$ may be represented by numbers. We begin the discussion with a very simple example. Let the random experiment be the toss of a coin and let the sample space associated with the experiment be $\mathcal{C} = \{H, T\}$, where H and T represent heads and tails, respectively. Let X be a function such that $X(T) = 0$ and $X(H) = 1$. Thus X is a real-valued function defined on the sample space $\mathcal{C}$ which takes us from the sample space $\mathcal{C}$ to a space of real numbers $\mathcal{D} = \{0, 1\}$. We now formulate the definition of a random variable and its space.

Definition 1.5.1. *Consider a random experiment with a sample space $\mathcal{C}$. A function X, which assigns to each element $c \in \mathcal{C}$ one and only one number $X(c) = x$, is called a* **random variable**. *The* **space** *or* **range** *of X is the set of real numbers $\mathcal{D} = \{x : x = X(c), c \in \mathcal{C}\}$.* ■

In this text, $\mathcal{D}$ generally is a countable set or an interval of real numbers. We call random variables of the first type **discrete** random variables, while we call those of the second type **continuous** random variables. In this section, we present examples of discrete and continuous random variables and then in the next two sections we discuss them separately.

Given a random variable X, its range $\mathcal{D}$ becomes the sample space of interest. Besides inducing the sample space $\mathcal{D}$, X also induces a probability which we call the **distribution** of X.

Consider first the case where X is a discrete random variable with a finite space $\mathcal{D} = \{d_1, \ldots, d_m\}$. The only events of interest in the new sample space $\mathcal{D}$ are subsets of $\mathcal{D}$. The induced probability distribution of X is also clear. Define the function $p_X(d_i)$ on $\mathcal{D}$ by

$$p_X(d_i) = P[\{c : X(c) = d_i\}], \quad \text{for } i = 1, \ldots, m. \tag{1.5.1}$$

In the next section, we formally define $p_X(d_i)$ as the **probability mass function** (**pmf**) of X. Then the induced probability distribution, $P_X(\cdot)$, of X is

$$P_X(D) = \sum_{d_i \in D} p_X(d_i), \quad D \subset \mathcal{D}.$$

As Exercise 1.5.11 shows, $P_X(D)$ is a probability on $\mathcal{D}$. An example is helpful here.

Example 1.5.1 (First Roll in the Game of Craps)**.** Let X be the sum of the upfaces on a roll of a pair of fair 6-sided dice, each with the numbers 1 through 6 on it. The sample space is $\mathcal{C} = \{(i,j) : 1 \le i,j \le 6\}$. Because the dice are fair, $P[\{(i,j)\}] = 1/36$. The random variable X is $X(i,j) = i + j$. The space of X is $\mathcal{D} = \{2,\ldots,12\}$. By enumeration, the pmf of X is given by

Range value x	2	3	4	5	6	7	8	9	10	11	12
Probability $p_X(x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

To illustrate the computation of probabilities concerning X, suppose $B_1 = \{x : x = 7, 11\}$ and $B_2 = \{x : x = 2, 3, 12\}$. Then, using the values of $p_X(x)$ given in the table,

$$\begin{aligned} P_X(B_1) &= \sum_{x \in B_1} p_X(x) = \frac{6}{36} + \frac{2}{36} = \frac{8}{36} \\ P_X(B_2) &= \sum_{x \in B_2} p_X(x) = \frac{1}{36} + \frac{2}{36} + \frac{1}{36} = \frac{4}{36}. \quad \blacksquare \end{aligned}$$

The second case is when X is a continuous random variable. In this case, $\mathcal{D}$ is an interval of real numbers. In practice, continuous random variables are often measurements. For example, the weight of an adult is modeled by a continuous random variable. Here we would not be interested in the probability that a person weighs exactly 200 pounds, but we may be interested in the probability that a person weighs over 200 pounds. Generally, for the continuous random variables, the simple events of interest are intervals. We can usually determine a nonnegative function $f_X(x)$ such that for any interval of real numbers $(a,b) \in \mathcal{D}$, the induced probability distribution of X, $P_X(\cdot)$, is defined as

$$P_X[(a,b)] = P[\{c \in \mathcal{C} : a < X(c) < b\}] = \int_a^b f_X(x)\,dx; \tag{1.5.2}$$

that is, the probability that X falls between a and b is the area under the curve $y = f_X(x)$ between a and b. Besides $f_X(x) \ge 0$, we also require that $P_X(\mathcal{D}) = \int_{\mathcal{D}} f_X(x)\,dx = 1$ (total area under the curve over the sample space of X is 1). There are some technical issues in defining events in general for the space $\mathcal{D}$; however, it can be shown that $P_X(D)$ is a probability on $\mathcal{D}$; see Exercise 1.5.11. The function f_X is formally defined as the **probability density function** (**pdf**) of X in Section 1.7. An example is in order.

Example 1.5.2. For an example of a continuous random variable, consider the following simple experiment: choose a real number at random from the interval $(0,1)$. Let X be the number chosen. In this case the space of X is $\mathcal{D} = (0,1)$. It is not obvious as it was in the last example what the induced probability P_X is. But there are some intuitive probabilities. For instance, because the number is chosen at random, it is reasonable to assign

$$P_X[(a,b)] = b - a, \text{ for } 0 < a < b < 1. \tag{1.5.3}$$

It follows that the pdf of X is

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.5.4}$$

For example, the probability that X is less than an eighth or greater than seven eighths is

$$P\left[\left\{X < \frac{1}{8}\right\} \cup \left\{X > \frac{7}{8}\right\}\right] = \int_0^{\frac{1}{8}} dx + \int_{\frac{7}{8}}^1 dx = \frac{1}{4}.$$

Notice that a discrete probability model is not a possibility for this experiment. For any point a, $0 < a < 1$, we can choose n_0 so large such that $0 < a - n_0^{-1} < a < a + n_0^{-1} < 1$, i.e., $\{a\} \subset (a - n_0^{-1}, a + n_0^{-1})$. Hence,

$$P(X = a) \le P\left(a - \frac{1}{n} < X < a + \frac{1}{n}\right) = \frac{2}{n}, \text{ for all } n \ge n_0. \tag{1.5.5}$$

Since $2/n \to 0$ as $n \to \infty$ and a is arbitrary, we conclude that $P(X = a) = 0$ for all $a \in (0,1)$. Hence, the reasonable pdf, (1.5.4), for this model excludes a discrete probability model. ■

Remark 1.5.1. In equations (1.5.1) and (1.5.2), the subscript X on p_X and f_X identifies the pmf and pdf, respectively, with the random variable. We often use this notation, especially when there are several random variables in the discussion. On the other hand, if the identity of the random variable is clear, then we often suppress the subscripts. ■

The pmf of a discrete random variable and the pdf of a continuous random variable are quite different entities. The distribution function, though, uniquely determines the probability distribution of a random variable. It is defined by:

Definition 1.5.2 (Cumulative Distribution Function)**.** *Let X be a random variable. Then its* **cumulative distribution function** *(cdf) is defined by $F_X(x)$, where*

$$F_X(x) = P_X((-\infty, x]) = P(\{c \in \mathcal{C} : X(c) \le x\}). \tag{1.5.6}$$

As above, we shorten $P(\{c \in \mathcal{C} : X(c) \le x\})$ to $P(X \le x)$. Also, $F_X(x)$ is often called simply the distribution function (df). However, in this text, we use the modifier *cumulative* as $F_X(x)$ accumulates the probabilities less than or equal to x.

The next example discusses a cdf for a discrete random variable.

Example 1.5.3. Suppose we roll a fair die with the numbers 1 through 6 on it. Let X be the upface of the roll. Then the space of X is $\{1, 2, \ldots, 6\}$ and its pmf is $p_X(i) = 1/6$, for $i = 1, 2, \ldots, 6$. If $x < 1$, then $F_X(x) = 0$. If $1 \le x < 2$, then $F_X(x) = 1/6$. Continuing this way, we see that the cdf of X is an increasing step function which steps up by $p_X(i)$ at each i in the space of X. The graph of F_X is given by Figure 1.5.1. Note that if we are given the cdf, then we can determine the pmf of X. ■

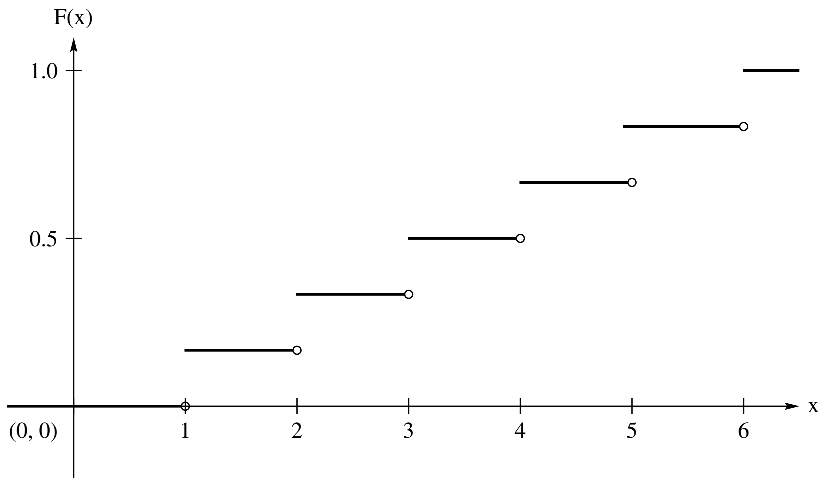

Figure 1.5.1: Distribution function for Example 1.5.3.

The following example discusses the cdf for the continuous random variable discussed in Example 1.5.2.

Example 1.5.4 (Continuation of Example 1.5.2)**.** Recall that X denotes a real number chosen at random between 0 and 1. We now obtain the cdf of X. First, if $x < 0$, then $P(X \leq x) = 0$. Next, if $x \geq 1$, then $P(X \leq x) = 1$. Finally, if $0 < x < 1$, it follows from expression (1.5.3) that $P(X \leq x) = P(0 < X \leq x) = x - 0 = x$. Hence the cdf of X is

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1. \end{cases} \tag{1.5.7}$$

A sketch of the cdf of X is given in Figure 1.5.2. Note, however, the connection between $F_X(x)$ and the pdf for this experiment $f_X(x)$, given in Example 1.5.2, is

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad \text{for all } x \in R,$$

and $\frac{d}{dx}F_X(x) = f_X(x)$, for all $x \in R$, except for $x = 0$ and $x = 1$. ■

Let X and Y be two random variables. We say that X and Y are **equal in distribution** and write $X \stackrel{D}{=} Y$ if and only if $F_X(x) = F_Y(x)$, for all $x \in R$. It is important to note while X and Y may be equal in distribution, they may be quite different. For instance, in the last example define the random variable Y as $Y = 1 - X$. Then $Y \neq X$. But the space of Y is the interval $(0, 1)$, the same as X. Further, the cdf of Y is 0 for $y < 0$; 1 for $y \geq 1$; and for $0 \leq y < 1$, it is

$$F_Y(y) = P(Y \leq y) = P(1 - X \leq y) = P(X \geq 1 - y) = 1 - (1 - y) = y.$$

Hence, Y has the same cdf as X, i.e., $Y \stackrel{D}{=} X$, but $Y \neq X$.

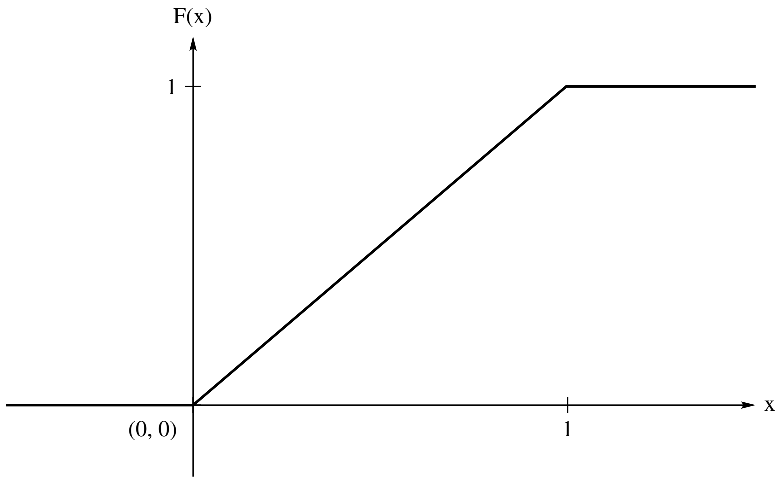

Figure 1.5.2: Distribution function for Example 1.5.4.

The cdfs displayed in Figures 1.5.1 and 1.5.2 show increasing functions with lower limits 0 and upper limits 1. In both figures, the cdfs are at least right continuous. As the next theorem proves, these properties are true in general for cdfs.

Theorem 1.5.1. *Let X be a random variable with cumulative distribution function $F(x)$. Then*

(a) *For all a and b, if $a < b$, then $F(a) \leq F(b)$ (F is nondecreasing).*

(b) $\lim_{x \to -\infty} F(x) = 0$ *(the lower limit of F is 0).*

(c) $\lim_{x \to \infty} F(x) = 1$ *(the upper limit of F is 1).*

(d) $\lim_{x \downarrow x_0} F(x) = F(x_0)$ *(F is right continuous).*

Proof: We prove parts (a) and (d) and leave parts (b) and (c) for Exercise 1.5.10.
Part (a): Because $a < b$, we have $\{X \leq a\} \subset \{X \leq b\}$. The result then follows from the monotonicity of P; see Theorem 1.3.3.
Part (d): Let $\{x_n\}$ be any sequence of real numbers such that $x_n \downarrow x_0$. Let $C_n = \{X \leq x_n\}$. Then the sequence of sets $\{C_n\}$ is decreasing and $\cap_{n=1}^{\infty} C_n = \{X \leq x_0\}$. Hence, by Theorem 1.3.6,

$$\lim_{n \to \infty} F(x_n) = P\left(\bigcap_{n=1}^{\infty} C_n \right) = F(x_0),$$

which is the desired result. ■

The next theorem is helpful in evaluating probabilities using cdfs.

Theorem 1.5.2. *Let X be a random variable with the cdf F_X. Then for $a < b$, $P[a < X \leq b] = F_X(b) - F_X(a)$.*

Proof: Note that

$$\{-\infty < X \le b\} = \{-\infty < X \le a\} \cup \{a < X \le b\}.$$

The proof of the result follows immediately because the union on the right side of this equation is a disjoint union. ■

Example 1.5.5. Let X be the lifetime in years of a mechanical part. Assume that X has the cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-x} & 0 \le x. \end{cases}$$

The pdf of X, $\frac{d}{dx}F_X(x)$, is

$$f_X(x) = \begin{cases} e^{-x} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Actually the derivative does not exist at $x = 0$, but in the continuous case the next theorem (1.5.3) shows that $P(X = 0) = 0$ and we can assign $f_X(0) = 0$ without changing the probabilities concerning X. The probability that a part has a lifetime between one and three years is given by

$$P(1 < X \le 3) = F_X(3) - F_X(1) = \int_1^3 e^{-x}\, dx.$$

That is, the probability can be found by $F_X(3) - F_X(1)$ or evaluating the integral. In either case, it equals $e^{-1} - e^{-3} = 0.318$. ■

Theorem 1.5.1 shows that cdfs are right continuous and monotone. Such functions can be shown to have only a countable number of discontinuities. As the next theorem shows, the discontinuities of a cdf have mass; that is, if x is a point of discontinuity of F_X, then we have $P(X = x) > 0$.

Theorem 1.5.3. *For any random variable,*

$$P[X = x] = F_X(x) - F_X(x-), \tag{1.5.8}$$

for all $x \in R$, where $F_X(x-) = \lim_{z\uparrow x} F_X(z)$.

Proof: For any $x \in R$, we have

$$\{x\} = \bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right];$$

that is, $\{x\}$ is the limit of a decreasing sequence of sets. Hence, by Theorem 1.3.6,

$$\begin{aligned} P[X = x] &= P\left[\bigcap_{n=1}^{\infty}\left\{x - \frac{1}{n} < X \le x\right\}\right] \\ &= \lim_{n\to\infty} P\left[x - \frac{1}{n} < X \le x\right] \\ &= \lim_{n\to\infty} [F_X(x) - F_X(x - (1/n))] \\ &= F_X(x) - F_X(x-), \end{aligned}$$

which is the desired result. ∎

Example 1.5.6. Let X have the discontinuous cdf

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x/2 & 0 \le x < 1 \\ 1 & 1 \le x. \end{cases}$$

Then

$$P(-1 < X \le 1/2) = F_X(1/2) - F_X(-1) = \frac{1}{4} - 0 = \frac{1}{4}$$

and

$$P(X = 1) = F_X(1) - F_X(1-) = 1 - \frac{1}{2} = \frac{1}{2}.$$

The value 1/2 equals the value of the step of F_X at $x = 1$. ∎

Since the total probability associated with a random variable X of the discrete type with pmf $p_X(x)$ or of the continuous type with pdf $f_X(x)$ is 1, then it must be true that

$$\textstyle\sum_{x\in\mathcal{D}} p_X(x) = 1 \text{ and } \int_{\mathcal{D}} f_X(x)\,dx = 1,$$

where $\mathcal{D}$ is the space of X. As the next two examples show, we can use this property to determine the pmf or pdf if we know the pmf or pdf down to a constant of proportionality.

Example 1.5.7. Suppose X has the pmf

$$p_X(x) = \begin{cases} cx & x = 1, 2, \ldots, 10 \\ 0 & \text{elsewhere,} \end{cases}$$

for an appropriate constant c. Then

$$1 = \sum_{x=1}^{10} p_X(x) = \sum_{x=1}^{10} cx = c(1 + 2 + \cdots + 10) = 55c,$$

and, hence, $c = 1/55$. ∎

Example 1.5.8. Suppose X has the pdf

$$f_X(x) = \begin{cases} cx^3 & 0 < x < 2 \\ 0 & \text{elsewhere,} \end{cases}$$

for a constant c. Then

$$1 = \int_0^2 cx^3\,dx = c\left[\frac{x^4}{4}\right]_0^2 = 4c,$$

and, hence, $c = 1/4$. For illustration of the computation of a probability involving X, we have

$$P\left(\frac{1}{4} < X < 1\right) = \int_{1/4}^{1} \frac{x^3}{4}\,dx = \frac{255}{4096} = 0.06226. \quad ∎$$

EXERCISES

1.5.1. Let a card be selected from 30 cards numbered 1 to 30. The outcome c is one of these 30 numbers. Let $X(c) = 3$ if c is a prime number, let $X(c) = 2$ if c is a multiple of 4, let $X(c) = 1$ if c is a multiple of 9, and let $X(c) = 0$ otherwise. Suppose that P assigns a probability of $\frac{1}{30}$ to each outcome c. Describe the induced probability $P_X(D)$ on the space $\mathcal{D} = \{0, 1, 2, 3\}$ of the random variable X.

1.5.2. For each of the following, find the constant c so that $p(x)$ satisfies the condition of being a pmf of one random variable X.

(a) $p(x) = c\left(\frac{1}{4}\right)^x$, $x = 1, 2, 3, \ldots$, zero elsewhere.

(b) $p(x) = cx^2$, $x = 0, 1, 3, 5, 7$, zero elsewhere.

1.5.3. Let $p_X(x) = x/10$, $x = 0, 1, 2, 3, 4$, zero elsewhere, be the pmf of X. Find $P(X = 2 \text{ or } 3)$, $P\left(\frac{3}{2} < X < \frac{7}{2}\right)$, and $P(2 \le X \le 3)$.

1.5.4. Let $p_X(x)$ be the pmf of a random variable X. Find the cdf $F(x)$ of X and sketch its graph along with that of $p_X(x)$ if:

(a) $p_X(x) = 1,\ x = 0$, zero elsewhere.

(b) $p_X(x) = \frac{1}{3},\ x = -1, 0, 1$, zero elsewhere.

(c) $p_X(x) = x/15,\ x = 1, 2, 3, 4, 5$, zero elsewhere.

1.5.5. Let us select five cards at random and without replacement from an ordinary deck of playing cards.

(a) Find the pmf of X, the number of hearts in the five cards.

(b) Determine $P(X \le 1)$.

1.5.6. Let the probability set function of the random variable X be $P_X(D) = \int_D f(x)dx$, where $f(x) = x/8$, for $x \in \mathcal{D} = \{x : 0 < x < 4\}$. Define the events $D_1 = \{x : 0 < x < 2\}$ and $D_2 = \{x : 3 \le x < 4\}$. Compute $P_X(D_1)$, $P_X(D_2)$, and $P_X(D_1 \cup D_2)$.

1.5.7. Let the space of the random variable X be $\mathcal{D} = \{x : 1 < x < 2\}$. If $D_1 = \{x : 1 < x \le \frac{4}{3}\}$ and $D_2 = \{x : \frac{4}{3} < x < 2\}$, find $P_X(D_2)$ if $P_X(D_1) = \frac{1}{3}$.

1.5.8. Suppose the random variable X has the cdf

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{x+2}{4} & -1 \le x < 1 \\ 1 & 1 \le x. \end{cases}$$

Write an R function to sketch the graph of $F(x)$. Use your graph to obtain the probabilities: (a) $P(-\frac{1}{2} < X \le \frac{1}{2})$; (b) $P(X = 0)$; (c) $P(X = 1)$; (d) $P(2 < X \le 3)$.

1.5.9. Consider an urn that contains slips of paper each with one of the numbers $1, 2, \ldots, 50$ on it. Suppose there are i slips with the number i on it for $i = 1, 2, \ldots, 50$. For example, there are 25 slips of paper with the number 25. Assume that the slips are identical except for the numbers. Suppose one slip is drawn at random. Let X be the number on the slip.

(a) Show that X has the pmf $p(x) = x/1275$, $x = 1, 2, \ldots, 50$, zero elsewhere.

(b) Compute $P(X \leq 30)$.

(c) Show that the cdf of X is $F(x) = [x]([x] + 1)/2550$, for $1 \leq x \leq 50$, where $[x]$ is the greatest integer in x.

1.5.10. Prove parts (b) and (c) of Theorem 1.5.1.

1.5.11. Let X be a random variable with space $\mathcal{D}$. For $D \subset \mathcal{D}$, recall that the probability induced by X is $P_X(D) = P[\{c : X(c) \in D\}]$. Show that $P_X(D)$ is a probability by showing the following:

(a) $P_X(\mathcal{D}) = 1$.

(b) $P_X(D) \geq 0$.

(c) For a sequence of sets $\{D_n\}$ in $\mathcal{D}$, show that

$$\{c : X(c) \in \cup_n D_n\} = \cup_n \{c : X(c) \in D_n\}.$$

(d) Use part (c) to show that if $\{D_n\}$ is sequence of mutually exclusive events, then

$$P_X\left(\cup_{n=1}^{\infty} D_n\right) = \sum_{n=1}^{\infty} P_X(D_n).$$

Remark 1.5.2. In a probability theory course, we would show that the σ-field (collection of events) for $\mathcal{D}$ is the smallest σ-field which contains all the open intervals of real numbers; see Exercise 1.3.24. Such a collection of events is sufficiently rich for discrete and continuous random variables. ■

1.6 Discrete Random Variables

The first example of a random variable encountered in the last section was an example of a discrete random variable, which is defined next.

Definition 1.6.1 (Discrete Random Variable)**.** *We say a random variable is a* **discrete random variable** *if its space is either finite or countable.*

Example 1.6.1. Consider a sequence of independent flips of a coin, each resulting in a head (H) or a tail (T). Moreover, on each flip, we assume that H and T are equally likely; that is, $P(H) = P(T) = \frac{1}{2}$. The sample space $\mathcal{C}$ consists of sequences like TTHTHHT$\cdots$. Let the random variable X equal the number of flips needed to obtain the first head. Hence, $X(\text{TTHTHHT}\cdots) = 3$. Clearly, the space of X is $\mathcal{D} = \{1, 2, 3, 4, \ldots\}$. We see that $X = 1$ when the sequence begins with an H and thus $P(X = 1) = \frac{1}{2}$. Likewise, $X = 2$ when the sequence begins with TH, which has probability $P(X = 2) = (\frac{1}{2})(\frac{1}{2}) = \frac{1}{4}$ from the independence. More generally, if $X = x$, where $x = 1, 2, 3, 4, \ldots$, there must be a string of $x - 1$ tails followed by a head; that is, $\text{TT}\cdots\text{TH}$, where there are $x - 1$ tails in $\text{TT}\cdots\text{T}$. Thus, from independence, we have a geometric sequence of probabilities, namely,

$$P(X = x) = \left(\frac{1}{2}\right)^{x-1}\left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{x}, \quad x = 1, 2, 3, \ldots, \tag{1.6.1}$$

the space of which is countable. An interesting event is that the first head appears on an odd number of flips; i.e., $X \in \{1, 3, 5, \ldots\}$. The probability of this event is

$$P[X \in \{1, 3, 5, \ldots\}] = \sum_{x=1}^{\infty}\left(\frac{1}{2}\right)^{2x-1} = \frac{1}{2}\sum_{x=1}^{\infty}\left(\frac{1}{4}\right)^{x-1} = \frac{1/2}{1 - (1/4)} = \frac{2}{3}. \quad \blacksquare$$

As the last example suggests, probabilities concerning a discrete random variable can be obtained in terms of the probabilities $P(X = x)$, for $x \in \mathcal{D}$. These probabilities determine an important function, which we define as

Definition 1.6.2 (Probability Mass Function (pmf))**.** *Let X be a discrete random variable with space $\mathcal{D}$. The* **probability mass function** *(pmf) of X is given by*

$$p_X(x) = P[X = x], \quad \textit{for } x \in \mathcal{D}. \tag{1.6.2}$$

Note that pmfs satisfy the following two properties:

$$\text{(i) } 0 \le p_X(x) \le 1\ , x \in \mathcal{D}, \text{ and (ii) } \textstyle\sum_{x\in\mathcal{D}} p_X(x) = 1. \tag{1.6.3}$$

In a more advanced class it can be shown that if a function satisfies properties (i) and (ii) for a discrete set $\mathcal{D}$, then this function uniquely determines the distribution of a random variable.

Let X be a discrete random variable with space $\mathcal{D}$. As Theorem 1.5.3 shows, discontinuities of $F_X(x)$ define a mass; that is, if x is a point of discontinuity of F_X, then $P(X = x) > 0$. We now make a distinction between the space of a discrete random variable and these points of positive probability. We define the **support** of a discrete random variable X to be the points in the space of X which have positive probability. We often use $\mathcal{S}$ to denote the support of X. Note that $\mathcal{S} \subset \mathcal{D}$, but it may be that $\mathcal{S} = \mathcal{D}$.

Also, we can use Theorem 1.5.3 to obtain a relationship between the pmf and cdf of a discrete random variable. If $x \in \mathcal{S}$, then $p_X(x)$ is equal to the size of the discontinuity of F_X at x. If $x \notin \mathcal{S}$ then $P[X = x] = 0$ and, hence, F_X is continuous at this x.

Example 1.6.2. A lot, consisting of 100 fuses, is inspected by the following procedure. Five of these fuses are chosen at random and tested; if all five "blow" at the correct amperage, the lot is accepted. If, in fact, there are 20 defective fuses in the lot, the probability of accepting the lot is, under appropriate assumptions,

$$\frac{\binom{80}{5}}{\binom{100}{5}} = 0.31931.$$

More generally, let the random variable X be the number of defective fuses among the five that are inspected. The pmf of X is given by

$$p_X(x) = \begin{cases} \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}} & \text{for } x = 0, 1, 2, 3, 4, 5 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.6.4}$$

Clearly, the space of X is $\mathcal{D} = \{0, 1, 2, 3, 4, 5\}$, which is also its support. This is an example of a random variable of the discrete type whose distribution is an illustration of a **hypergeometric distribution**, which is formally defined in Chapter 3. Based on the above discussion, it is easy to graph the cdf of X; see Exercise 1.6.5. ■

1.6.1 Transformations

A problem often encountered in statistics is the following. We have a random variable X and we know its distribution. We are interested, though, in a random variable Y which is some **transformation** of X, say, $Y = g(X)$. In particular, we want to determine the distribution of Y. Assume X is discrete with space $\mathcal{D}_X$. Then the space of Y is $\mathcal{D}_Y = \{g(x) : x \in \mathcal{D}_X\}$. We consider two cases.

In the first case, g is one-to-one. Then, clearly, the pmf of Y is obtained as

$$p_Y(y) = P[Y = y] = P[g(X) = y] = P[X = g^{-1}(y)] = p_X(g^{-1}(y)). \tag{1.6.5}$$

Example 1.6.3. Consider the random variable X of Example 1.6.1. Recall that X was the flip number on which the first head appeared. Let Y be the number of flips before the first head. Then $Y = X - 1$. In this case, the function g is $g(x) = x - 1$, whose inverse is given by $g^{-1}(y) = y + 1$. The space of Y is $D_Y = \{0, 1, 2, \ldots\}$. The pmf of X is given by (1.6.1); hence, based on expression (1.6.5), the pmf of Y is

$$p_Y(y) = p_X(y + 1) = \left(\frac{1}{2}\right)^{y+1}, \quad \text{for } y = 0, 1, 2, \ldots. \quad ■$$

Example 1.6.4. Let X have the pmf

$$p_X(x) = \begin{cases} \frac{3!}{x!(3-x)!}\left(\frac{2}{3}\right)^x\left(\frac{1}{3}\right)^{3-x} & x = 0, 1, 2, 3 \\ 0 & \text{elsewhere.} \end{cases}$$

We seek the pmf $p_Y(y)$ of the random variable $Y = X^2$. The transformation $y = g(x) = x^2$ maps $\mathcal{D}_X = \{x : x = 0, 1, 2, 3\}$ onto $\mathcal{D}_Y = \{y : y = 0, 1, 4, 9\}$. In general, $y = x^2$ does not define a one-to-one transformation; here, however, it does, for there are no negative values of x in $\mathcal{D}_X = \{x : x = 0, 1, 2, 3\}$. That is, we have the single-valued inverse function $x = g^{-1}(y) = \sqrt{y}$ (not $-\sqrt{y}$), and so

$$p_Y(y) = p_X(\sqrt{y}) = \frac{3!}{(\sqrt{y})!(3-\sqrt{y})!}\left(\frac{2}{3}\right)^{\sqrt{y}}\left(\frac{1}{3}\right)^{3-\sqrt{y}}, \quad y = 0, 1, 4, 9. \blacksquare$$

The second case is where the transformation, $g(x)$, is not one-to-one. Instead of developing an overall rule, for most applications involving discrete random variables the pmf of Y can be obtained in a straightforward manner. We offer two examples as illustrations.

Consider the geometric random variable in Example 1.6.3. Suppose we are playing a game against the "house" (say, a gambling casino). If the first head appears on an odd number of flips, we pay the house one dollar, while if it appears on an even number of flips, we win one dollar from the house. Let Y denote our net gain. Then the space of Y is $\{-1, 1\}$. In Example 1.6.1, we showed that the probability that X is odd is $\frac{2}{3}$. Hence, the distribution of Y is given by $p_Y(-1) = 2/3$ and $p_Y(1) = 1/3$.

As a second illustration, let $Z = (X-2)^2$, where X is the geometric random variable of Example 1.6.1. Then the space of Z is $\mathcal{D}_Z = \{0, 1, 4, 9, 16, \ldots\}$. Note that $Z = 0$ if and only if $X = 2$; $Z = 1$ if and only if $X = 1$ or $X = 3$; while for the other values of the space there is a one-to-one correspondence given by $x = \sqrt{z} + 2$, for $z \in \{4, 9, 16, \ldots\}$. Hence, the pmf of Z is

$$p_Z(z) = \begin{cases} p_X(2) = \frac{1}{4} & \text{for } z = 0 \\ p_X(1) + p_X(3) = \frac{5}{8} & \text{for } z = 1 \\ p_X(\sqrt{z}+2) = \frac{1}{4}\left(\frac{1}{2}\right)^{\sqrt{z}} & \text{for } z = 4, 9, 16, \ldots. \end{cases} \tag{1.6.6}$$

For verification, the reader is asked to show in Exercise 1.6.11 that the pmf of Z sums to 1 over its space.

EXERCISES

1.6.1. Let X equal the number of heads in five independent flips of a coin. Using certain assumptions, determine the pmf of X and compute the probability that X is equal to an even number.

1.6.2. Let a bowl contain eight chips of the same size and shape. One and only one of these chips is red. Continue to draw chips from the bowl, one at a time and at random and without replacement, until the red chip is drawn.

(a) Find the pmf of X, the number of trials needed to draw the red chip.

(b) Compute $P(X > 3)$.

1.6.3. Cast a die a number of independent times until a six appears on the up side of the die.

(a) Find the pmf $p(x)$ of X, the number of casts needed to obtain that first six.

(b) Show that $\sum_{x=1}^{\infty} p(x) = 1$.

(c) Determine $P(X = 1, 3, 5, 7, \ldots)$.

(d) Find the cdf $F(x) = P(X \leq x)$.

1.6.4. Cast a die two independent times and let X equal the absolute value of the difference of the two resulting values (the numbers on the up sides). Find the pmf of X. *Hint:* It is not necessary to find a formula for the pmf.

1.6.5. For the random variable X defined in Example 1.6.2:

(a) Write an R function that returns the pmf. Note that in R, `choose(m,k)` computes $\binom{m}{k}$.

(b) Write an R function that returns the the graph of the cdf.

1.6.6. For the random variable X defined in Example 1.6.1, graph the cdf of X.

1.6.7. Let X have a pmf $p(x) = 1/4$, $x = 1, 2, 3, 4$, zero elsewhere. Find the pmf of $Y = 3X + 2$.

1.6.8. Let X have the pmf $p(x) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^x$, $x = 0, 1, 2, \ldots$, zero elsewhere. Find the pmf of $Y = X^3 + 1$.

1.6.9. Let X have the pmf $p(x) = 1/5$, $x = -2, -1, 0, 1, 2$. Find the pmf of $Y = X^2$.

1.6.10. Let X have the pmf

$$p(x) = \left(\frac{2}{3}\right)\left(\frac{1}{3}\right)^{|x|}, x = 0, -1, -2, -3, \ldots.$$

Find the pmf of $Y = X^4$.

1.6.11. Show that the function given in expression (1.6.6) is a pmf.

1.7 Continuous Random Variables

In the last section, we discussed discrete random variables. Another class of random variables important in statistical applications is the class of continuous random variables, which we define next.

Definition 1.7.1 (Continuous Random Variables)**.** *We say a random variable is a* **continuous random variable** *if its cumulative distribution function $F_X(x)$ is a continuous function for all $x \in R$.*

Recall from Theorem 1.5.3 that $P(X = x) = F_X(x) - F_X(x-)$, for any random variable X. Hence, for a continuous random variable X, there are no points of discrete mass; i.e., if X is continuous, then $P(X = x) = 0$ for all $x \in R$. Most continuous random variables are **absolutely continuous**; that is,

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt, \tag{1.7.1}$$

for some function $f_X(t)$. The function $f_X(t)$ is called a **probability density function** (pdf) of X. If $f_X(x)$ is also continuous, then the Fundamental Theorem of Calculus implies that

$$\frac{d}{dx}F_X(x) = f_X(x). \tag{1.7.2}$$

The **support** of a continuous random variable X consists of all points x such that $f_X(x) > 0$. As in the discrete case, we often denote the support of X by $\mathcal{S}$.

If X is a continuous random variable, then probabilities can be obtained by integration; i.e.,

$$P(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(t)\,dt.$$

Also, for continuous random variables,

$$P(a < X \le b) = P(a \le X \le b) = P(a \le X < b) = P(a < X < b).$$

From the definition (1.7.2), note that pdfs satisfy the two properties

$$\text{(i) } f_X(x) \ge 0 \text{ and (ii) } \textstyle\int_{-\infty}^{\infty} f_X(t)\,dt = 1. \tag{1.7.3}$$

The second property, of course, follows from $F_X(\infty) = 1$. In an advanced course in probability, it is shown that if a function satisfies the above two properties, then it is a pdf for a continuous random variable; see, for example, Tucker (1967).

Recall in Example 1.5.2 the simple experiment where a number was chosen at random from the interval $(0, 1)$. The number chosen, X, is an example of a continuous random variable. Recall that the cdf of X is $F_X(x) = x$, for $0 < x < 1$. Hence, the pdf of X is given by

$$f_X(x) = \begin{cases} 1 & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.7.4}$$

Any continuous or discrete random variable X whose pdf or pmf is constant on the support of X is said to have a **uniform** distribution; see Chapter 3 for a more formal definition.

Example 1.7.1 (Point Chosen at Random Within the Unit Circle)**.** Suppose we select a point at random in the interior of a circle of radius 1. Let X be the distance of the selected point from the origin. The sample space for the experiment is $\mathcal{C} = \{(w, y) : w^2 + y^2 < 1\}$. Because the point is chosen at random, it seems that subsets of $\mathcal{C}$ which have equal area are equilikely. Hence, the probability of the selected point lying in a set $A \subset \mathcal{C}$ is proportional to the area of A; i.e.,

$$P(A) = \frac{\text{area of } A}{\pi}.$$

For $0 < x < 1$, the event $\{X \le x\}$ is equivalent to the point lying in a circle of radius x. By this probability rule, $P(X \le x) = \pi x^2/\pi = x^2$; hence, the cdf of X is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < 1 \\ 1 & 1 \le x. \end{cases} \tag{1.7.5}$$

Taking the derivative of $F_X(x)$, we obtain the pdf of X:

$$f_X(x) = \begin{cases} 2x & 0 \leq x < 1 \\ 0 & \text{elsewhere.} \end{cases} \tag{1.7.6}$$

For illustration, the probability that the selected point falls in the ring with radii 1/4 and 1/2 is given by

$$P\left(\frac{1}{4} < X \leq \frac{1}{2}\right) = \int_{\frac{1}{4}}^{\frac{1}{2}} 2w\,dw = w^2\Big|_{\frac{1}{4}}^{\frac{1}{2}} = \frac{3}{16}. \quad \blacksquare$$

Example 1.7.2. Let the random variable be the time in seconds between incoming telephone calls at a busy switchboard. Suppose that a reasonable probability model for X is given by the pdf

$$f_X(x) = \begin{cases} \frac{1}{4}e^{-x/4} & 0 < x < \infty \\ 0 & \text{elsewhere.} \end{cases}$$

Note that f_X satisfies the two properties of a pdf, namely, (i) $f(x) \geq 0$ and (ii)

$$\int_0^\infty \frac{1}{4}e^{-x/4}\,dx = -e^{-x/4}\Big|_0^\infty = 1.$$

For illustration, the probability that the time between successive phone calls exceeds 4 seconds is given by

$$P(X > 4) = \int_4^\infty \frac{1}{4}e^{-x/4}\,dx = e^{-1} = 0.3679.$$

The pdf and the probability of interest are depicted in Figure 1.7.1. From the figure, the pdf has a long right tail and no left tail. We say that this distribution is **skewed right** or positively skewed. This is an example of a gamma distribution which is discussed in detail in Chapter 3. ■

1.7.1 Quantiles

Quantiles (percentiles) are easily interpretable characteristics of a distribution.

Definition 1.7.2 (Quantile)**.** *Let $0 < p < 1$. The* **quantile** *of order p of the distribution of a random variable X is a value ξ_p such that $P(X < \xi_p) \leq p$ and $P(X \leq \xi_p) \geq p$. It is also known as the $(100p)$th* **percentile** *of X.* ■

Examples include the **median** which is the quantile $\xi_{1/2}$. The median is also called the second quartile. It is a point in the domain of X that divides the mass of the pdf into its lower and upper halves. The first and third quartiles divide each of these halves into quarters. They are, respectively $\xi_{1/4}$ and $\xi_{3/4}$. We label these quartiles as q_1, q_2 and q_3, respectively. The difference $\text{iq} = q_3 - q_1$ is called the

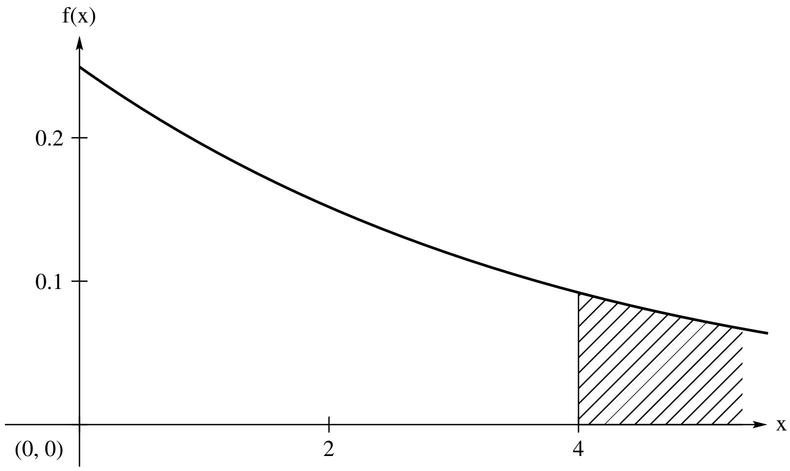

Figure 1.7.1: In Example 1.7.2, the area under the pdf to the right of 4 is $P(X > 4)$.

interquartile range of X. The median is often used as a measure of center of the distribution of X, while the interquartile range is used as a measure of **spread** or **dispersion** of the distribution of X.

Quantiles need not be unique even for continuous random variables with pdfs. For example, any point in the interval $(2, 3)$ serves as a median for the following pdf:

$$f(x) = \begin{cases} 3(1-x)(x-2) & 1 < x < 2 \\ 3(3-x)(x-4) & 3 < x < 4 \\ 0 & \text{elsewhere.} \end{cases} \qquad (1.7.7)$$

If, however, a quantile, say ξ_p, is in the support of an absolutely continuous random variable X with cdf $F_X(x)$ then ξ_p is the unique solution to the equation:

$$\xi_p = F_X^{-1}(p), \qquad (1.7.8)$$

where $F_X^{-1}(u)$ is the inverse function of $F_X(x)$. The next example serves as an illustration.

Example 1.7.3. Let X be a continuous random variable with pdf

$$f(x) = \frac{e^x}{(1+5e^x)^{1.2}}, \quad -\infty < x < \infty. \qquad (1.7.9)$$

This pdf is a member of the log F-family of ditributions which is often used in the modeling of the log of lifetime data. Note that X has the support space $(-\infty, \infty)$. The cdf of X is

$$F(x) = 1 - (1+5e^x)^{-.2} \quad -\infty < x < \infty,$$

which is confirmed immediately by showing that $F'(x) = f(x)$. For the inverse of the cdf, set $u = F(x)$ and solve for u. A few steps of algebra lead to

$$F^{-1}(u) = \log\left\{.2\left[(1-u)^{-5} - 1\right]\right\}, \quad 0 < u < 1.$$

Thus, $\xi_p = F_X^{-1}(p) = \log\left\{.2\left[(1-p)^{-5} - 1\right]\right\}$. The following three R functions can be used to compute the pdf, cdf, and inverse cdf of F, respectively. These can be downloaded at the site listed in the Preface.

```
dlogF <- function(x){exp(x)/(1+5*exp(x))^(1.2)}
plogF <- function(x){1- (1+5*exp(x))^(-.2)}
qlogF <- function(x){log(.2*((1-x)^(-5) - 1))}
```

Once the R function `qlogF` is sourced, it can be used to compute quantiles. The following is an R script which results in the computation of the three quartiles of X:

```
qlogF(.25) ; qlogF(.50); qlogF(.75)
-0.4419242; 1.824549; 5.321057
```

Figure 1.7.2 displays a plot of this pdf and its quartiles. Notice that this is another example of a skewed-right distribution; i.e., the right-tail is much longer than left-tail. In terms of the log-lifetime of mechanical parts having this distribution, it follows that 50% of the parts survive beyond 1.83 log-units and 25% of the parts live longer than 5.32 log-units. With the long-right tail, some parts attain a long life. ■

1.7.2 Transformations

Let X be a continuous random variable with a known pdf f_X. As in the discrete case, we are often interested in the distribution of a random variable Y which is some **transformation** of X, say, $Y = g(X)$. Often we can obtain the pdf of Y by first obtaining its cdf. We illustrate this with two examples.

Example 1.7.4. Let X be the random variable in Example 1.7.1. Recall that X was the distance from the origin to the random point selected in the unit circle. Suppose instead that we are interested in the square of the distance; that is, let $Y = X^2$. The support of Y is the same as that of X, namely, $\mathcal{S}_Y = (0, 1)$. What is the cdf of Y? By expression (1.7.5), the cdf of X is

$$F_X(x) = \begin{cases} 0 & x < 0 \\ x^2 & 0 \le x < 1 \\ 1 & 1 \le x. \end{cases} \tag{1.7.10}$$

Let y be in the support of Y; i.e., $0 < y < 1$. Then, using expression (1.7.10) and the fact that the support of X contains only positive numbers, the cdf of Y is

$$F_Y(y) = P(Y \le y) = P(X^2 \le y) = P(X \le \sqrt{y}) = F_X(\sqrt{y}) = \sqrt{y}^2 = y.$$

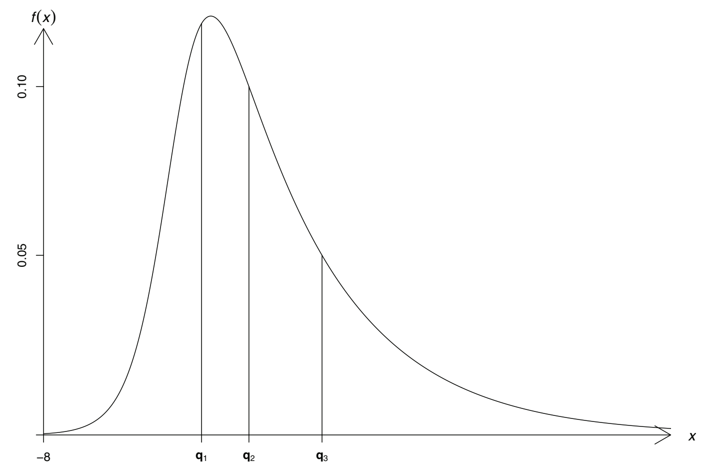

Figure 1.7.2: A graph of the pdf (1.7.9) showing the three quartiles, q_1, q_2, and q_3, of the distribution. The probability mass in each of the four sections is 1/4.

It follows that the pdf of Y is

$$f_Y(y) = \begin{cases} 1 & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases} \quad \blacksquare$$

Example 1.7.5. Let $f_X(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere, be the pdf of a random variable X. Note that X has a uniform distribution with the interval of support $(-1, 1)$. Define the random variable Y by $Y = X^2$. We wish to find the pdf of Y. If $y \geq 0$, the probability $P(Y \leq y)$ is equivalent to

$$P(X^2 \leq y) = P(-\sqrt{y} \leq X \leq \sqrt{y}).$$

Accordingly, the cdf of Y, $F_Y(y) = P(Y \leq y)$, is given by

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{2}\, dx = \sqrt{y} & 0 \leq y < 1 \\ 1 & 1 \leq y. \end{cases}$$

Hence, the pdf of Y is given by

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}} & 0 < y < 1 \\ 0 & \text{elsewhere.} \end{cases} \quad \blacksquare$$

These examples illustrate the **cumulative distribution function technique**. The transformation in Example 1.7.4 is one-to-one, and in such cases we can obtain a simple formula for the pdf of Y in terms of the pdf of X, which we record in the next theorem.

Theorem 1.7.1. *Let X be a continuous random variable with pdf $f_X(x)$ and support $\mathcal{S}_X$. Let $Y = g(X)$, where $g(x)$ is a one-to-one differentiable function, on the support of X, $\mathcal{S}_X$. Denote the inverse of g by $x = g^{-1}(y)$ and let $dx/dy = d[g^{-1}(y)]/dy$. Then the pdf of Y is given by*

$$f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|, \quad \text{for } y \in \mathcal{S}_Y\ , \tag{1.7.11}$$

where the support of Y is the set $\mathcal{S}_Y = \{y = g(x) : x \in \mathcal{S}_X\}$.

Proof: Since $g(x)$ is one-to-one and continuous, it is either strictly monotonically increasing or decreasing. Assume that it is strictly monotonically increasing, for now. The cdf of Y is given by

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = P[X \leq g^{-1}(y)] = F_X(g^{-1}(y)). \tag{1.7.12}$$

Hence, the pdf of Y is

$$f_Y(y) = \frac{d}{dy}F_Y(y) = f_X(g^{-1}(y))\frac{dx}{dy}, \tag{1.7.13}$$

where dx/dy is the derivative of the function $x = g^{-1}(y)$. In this case, because g is increasing, $dx/dy > 0$. Hence, we can write $dx/dy = |dx/dy|$.

Suppose $g(x)$ is strictly monotonically decreasing. Then (1.7.12) becomes $F_Y(y) = 1 - F_X(g^{-1}(y))$. Hence, the pdf of Y is $f_Y(y) = f_X(g^{-1}(y))(-dx/dy)$. But since g is decreasing, $dx/dy < 0$ and, hence, $-dx/dy = |dx/dy|$. Thus Equation (1.7.11) is true in both cases.[5] . ■

Henceforth, we refer to $dx/dy = (d/dy)g^{-1}(y)$ as the **Jacobian** (denoted by J) of the transformation. In most mathematical areas, $J = dx/dy$ is referred to as the Jacobian of the inverse transformation $x = g^{-1}(y)$, but in this book it is called the Jacobian of the transformation, simply for convenience.

We summarize Theorem 1.7.1 in a simple algorithm which we illustrate in the next example. Assuming that the transformation $Y = g(X)$ is one-to-one, the following steps lead to the pdf of Y:

1. Find the support of Y.

2. Solve for the inverse of the transfomation; i.e., solve for x in terms of y in $y = g(x)$, thereby obtaining $x = g^{-1}(y)$.

3. Obtain $\frac{dx}{dy}$.

4. The pdf of Y is $f_Y(y) = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|$.

[5]The proof of Theorem 1.7.1 can also be obtained by using the change-of-variable technique as discussed in Chapter 4 of *Mathematical Comments*.

Example 1.7.6. Let X have the pdf

$$f(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Consider the random variable $Y = -\log X$. Here are the steps of the above algorithm:

1. The support of $Y = -\log X$ is $(0, \infty)$.
2. If $y = -\log x$ then $x = e^{-y}$.
3. $\frac{dx}{dy} = -e^{-y}$.
4. Thus the pdf of Y is:

$$f_Y(y) = f_X\left(e^{-y}\right)\left|-e^{-y}\right| = 4(e^{-y})^3 e^{-y} = 4e^{-4y}.$$

1.7.3 Mixtures of Discrete and Continuous Type Distributions

We close this section by two examples of distributions that are not of the discrete or the continuous type.

Example 1.7.7. Let a distribution function be given by

$$F(x) = \begin{cases} 0 & x < 0 \\ \frac{x+1}{2} & 0 \leq x < 1 \\ 1 & 1 \leq x. \end{cases}$$

Then, for instance,

$$P\left(-3 < X \leq \frac{1}{2}\right) = F\left(\frac{1}{2}\right) - F(-3) = \frac{3}{4} - 0 = \frac{3}{4}$$

and

$$P(X = 0) = F(0) - F(0-) = \frac{1}{2} - 0 = \frac{1}{2}.$$

The graph of $F(x)$ is shown in Figure 1.7.3. We see that $F(x)$ is not always continuous, nor is it a step function. Accordingly, the corresponding distribution is neither of the continuous type nor of the discrete type. It may be described as a **mixture** of those types. ■

Distributions that are mixtures of the continuous and discrete type do, in fact, occur frequently in practice. For illustration, in life testing, suppose we know that the length of life, say X, exceeds the number b, but the exact value of X is unknown. This is called *censoring*. For instance, this can happen when a subject in a cancer study simply disappears; the investigator knows that the subject has lived a certain number of months, but the exact length of life is unknown. Or it might happen when an investigator does not have enough time in an investigation to observe the moments of deaths of all the animals, say rats, in some study. Censoring can also occur in the insurance industry; in particular, consider a loss with a limited-pay policy in which the top amount is exceeded but it is not known by how much.

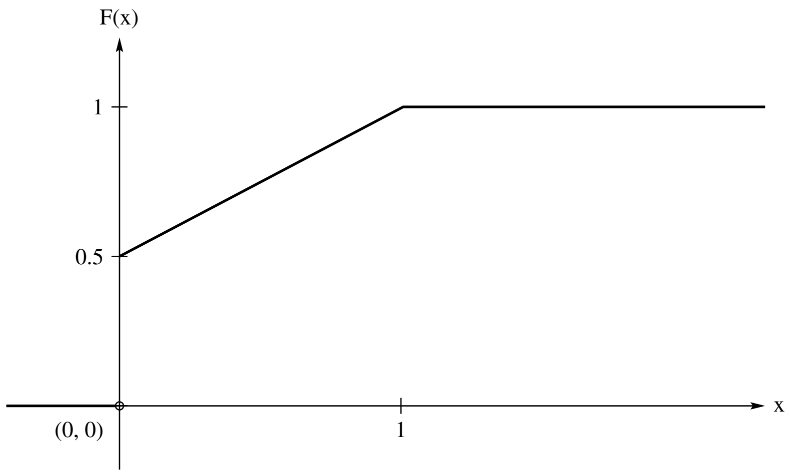

Figure 1.7.3: Graph of the cdf of Example 1.7.7.

Example 1.7.8. Reinsurance companies are concerned with large losses because they might agree, for illustration, to cover losses due to wind damages that are between \$2,000,000 and \$10,000,000. Say that X equals the size of a wind loss in millions of dollars, and suppose it has the cdf

$$F_X(x) = \begin{cases} 0 & -\infty < x < 0 \\ 1 - \left(\frac{10}{10+x}\right)^3 & 0 \le x < \infty. \end{cases}$$

If losses beyond \$10,000,000 are reported only as 10, then the cdf of this censored distribution is

$$F_Y(y) = \begin{cases} 0 & -\infty < y < 0 \\ 1 - \left(\frac{10}{10+y}\right)^3 & 0 \le y < 10, \\ 1 & 10 \le y < \infty, \end{cases}$$

which has a jump of $[10/(10+10)]^3 = \frac{1}{8}$ at $y = 10$. ■

EXERCISES

1.7.1. Let a point be selected from the sample space $\mathcal{C} = \{c : 0 < c < 5\}$. Let $C \subset \mathcal{C}$ and let the probability set function be $P(C) = \int_C \frac{1}{5} dz$. Define the random variable X to be $X(c) = c^2$. Find the cdf and the pdf of X.

1.7.2. Let the space of the random variable X be $\mathcal{C} = \{x : 0 < x < 8\}$ and let $P_X(C_1) = 7/10$, where $C_1 = \{x : 0 < x < 4\}$. Show that $P_X(C_2) \le 3/10$, where $C_2 = \{x : 5 \le x < 8\}$.

1.7.3. Let the subsets $C_1 = \{0 < x \le 1\}$ and $C_2 = \{1 < x \le \frac{3}{2}\}$ of the space $\mathcal{C} = \{x : 0 < x < 2\}$ of the random variable X be such that $P_X(C_1) = \frac{1}{2}$ and $P_X(C_2) = \frac{1}{4}$. Find $P_X(C_1 \cup C_2)$, $P_X(C_1^c)$, and $P_X(C_1^c \cap C_2^c)$.

1.7.4. Given $\int_C [1/\pi(1+x^2)]\,dx$, where $C \subset \mathcal{C} = \{x : -\infty < x < \infty\}$. Show that the integral could serve as a probability set function of a random variable X whose space is $\mathcal{C}$.

1.7.5. Let the probability set function of the random variable X be

$$P_X(C) = \int_C e^{-x} dx, \text{ where } \mathcal{C} = \{x : 0 < x < \infty\}.$$

Let $C_k = \{x : 1 \le x < 4 - 1/k\}$, $k = 1, 2, 3, \ldots$. Find the limits $\lim_{k\to\infty} C_k$ and $P_X(\lim_{k\to\infty} C_k)$. Find $P_X(C_k)$ and show that $\lim_{k\to\infty} P_X(C_k) = P_X(\lim_{k\to\infty} C_k)$.

1.7.6. For each of the following pdfs of X, find $P(|X| < 1)$ and $P(X^2 < 4)$.

(a) $f(x) = 3x^2/16$, $-2 < x < 2$, zero elsewhere.

(b) $f(x) = (x+3)/16$, $-1 < x < 3$, zero elsewhere.

1.7.7. Let $f(x) = 2/x^3$, $1 < x < \infty$, zero elsewhere, be the pdf of X. If $C_1 = \{x : 2 < x < 4\}$ and $C_2 = \{x : 3 < x < 5\}$, find $P_X(C_1 \cup C_2)$ and $P_X(C_1 \cap C_2)$.

1.7.8. A **mode** of the distribution of a random variable X is a value of x that maximizes the pdf or pmf. If there is only one such x, it is called the *mode of the distribution*. Find the mode of each of the following distributions:

(a) $p(x) = x/10$, $x = 1, 2, 3, 4$, zero elsewhere.

(b) $f(x) = x^2(3-x)/4$, $0 \le x \le 2$, zero elsewhere.

(c) $f(x) = \left(\frac{1}{9}\right) xe^{-x/3}$, $0 < x < \infty$, zero elsewhere.

1.7.9. The median and quantiles, in general, are discussed in Section 1.7.1. Find the median of each of the following distributions:

(a) $p(x) = \frac{4!}{x!(4-x)!}\left(\frac{1}{4}\right)^x\left(\frac{3}{4}\right)^{4-x}$, $x = 0, 1, 2, 3, 4$, zero elsewhere.

(b) $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere.

(c) $f(x) = \frac{1}{\pi(1+x^2)}$, $-\infty < x < \infty$.

1.7.10. Find the 0.30 quantile (30th percentile) of the distribution that has pdf $f(x) = 3x^2$, $0 < x < 1$, zero elsewhere.

1.7.11. For each of the following cdfs $F(x)$, find the pdf $f(x)$ [pmf in part **(d)**], the first quartile, and the 0.60 quantile. Also, sketch the graphs of $f(x)$ and $F(x)$. May use R to obtain the graphs. For Part(a) the code is provided.

(a) $F(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}(x)\,, -\infty < x < \infty$.

```
x<-seq(-5,5,.01); y<-.5+atan(x)/pi; y2<-1/(pi*(1+x^2))
par(mfrow=c(1,2));plot(y~x);plot(y2~x)
```

(b) $F(x) = \exp\{-e^{-x}\}$, $-\infty < x < \infty$.

(c) $F(x) = (1 + e^{-x})^{-1}$, $-\infty < x < \infty$.

(d) $F(x) = \sum_{j=1}^{x} \left(\frac{1}{2}\right)^j$.

1.7.12. Find the cdf $F(x)$ associated with each of the following probability density functions. Sketch the graphs of $f(x)$ and $F(x)$.

(a) $f(x) = 3(1 - x)^2$, $0 < x < 1$, zero elsewhere.

(b) $f(x) = 1/x^2$, $1 < x < \infty$, zero elsewhere.

(c) $f(x) = \frac{1}{3}$, $0 < x < 1$ or $2 < x < 4$, zero elsewhere.

Also, find the median and the 25th percentile of each of these distributions.

1.7.13. Consider the cdf $F(x) = 1 - e^{-2x} - 2xe^{-2x}$, $0 \le x < \infty$, zero elsewhere. Find the pdf, the mode, and the median (by numerical methods) of this distribution.

1.7.14. Let X have the pdf $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Compute the probability that X is at least 4 given that X is at least 2.

1.7.15. The random variable X is said to be **stochastically larger** than the random variable Y if

$$P(X > z) \ge P(Y > z), \tag{1.7.14}$$

for all real z, with strict inequality holding for at least one z value. Show that this requires that the cdfs enjoy the following property:

$$F_X(z) \le F_Y(z),$$

for all real z, with strict inequality holding for at least one z value.

1.7.16. Let X be a continuous random variable with support $(-\infty, \infty)$. Consider the random variable $Y = X + \Delta$, where $\Delta > 0$. Using the definition in Exercise 1.7.15, show that Y is stochastically larger than X.

1.7.17. Divide a line segment into two parts by selecting a point at random. Find the probability that the length of the larger segment is at least two times the length of the shorter segment. Assume a uniform distribution.

1.7.18. Let X be the number of gallons of ice cream that is requested at a certain store on a hot summer day. Assume that $f(x) = (600 - x)/(1.8 \times 10^5)$, $0 < x < 600$, zero elsewhere, is the pdf of X. How many gallons of ice cream should the store have on hand each of these days, so that the probability of exhausting its supply on a particular day is 0.05?

1.7.19. Find the 75th percentile of the distribution having pdf $f(x) = |x|/9$, where $-3 < x < 3$ and zero elsewhere.

1.7.20. The distribution of the random variable X in Example 1.7.3 is often used to model the log of the lifetime of a mechanical or electrical part. What about the lifetime itself? Let $Y = \exp\{X\}$.

(a) Determine the range of Y.

(b) Use the transformation technique to find the pdf of Y.

(c) Write an R function to compute this pdf and use it to obtain a graph of the pdf. Discuss the plot.

(d) Determine the 90th percentile of Y.

1.7.21. The distribution of the random variable X in Example 1.7.3 is a member of the log-F familily. Another member has the cdf

$$F(x) = \left[1 + \frac{2}{3}e^{-x}\right]^{-5/2}, \quad -\infty < x < \infty.$$

(a) Determine the corresponding pdf.

(b) Write an R function that computes this cdf. Plot the function and obtain approximations of the quartiles and median by inspection of the plot.

(c) Obtain the inverse of the cdf and confirm the percentiles in Part **(b)**.

1.7.22. Let X have the pdf $f(x) = 3x^2/8$, $0 < x < 2$, zero elsewhere. Find the pdf of $Y = X^3$.

1.7.23. If the pdf of X is $f(x) = xe^{-x^2/2}$, $0 < x < \infty$, zero elsewhere, determine the pdf of $Y = X^2$.

1.7.24. Let X have the uniform pdf $f_X(x) = \frac{1}{\pi}$, for $-\frac{\pi}{2} < x < \frac{\pi}{2}$. Find the pdf of $Y = \tan X$. This is the pdf of a **Cauchy distribution**.

1.7.25. Let X have the pdf $f(x) = \frac{4}{3}x^{1/3}$, $0 < x < 1$, zero elsewhere. Find the cdf and the pdf of $Y = -\ln X^4$.

1.7.26. Let $f(x) = \frac{1}{4}$, $-3 < x < 1$, zero elsewhere, be the pdf of X. Find the cdf and the pdf of $Y = X^2$.
Hint: Consider $P(X^2 \leq y)$ for two cases: $0 \leq y < 1$ and $1 \leq y < 9$.

1.8 Expectation of a Random Variable

In this section we introduce the expectation operator, which we use throughout the remainder of the text. For the definition, recall from calculus that absolute convergence of sums or integrals implies their convergence.

Definition 1.8.1 (Expectation)**.** *Let X be a random variable. If X is a continuous random variable with pdf $f(x)$ and*

$$\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty,$$

then the **expectation** *of X is*

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

If X is a discrete random variable with pmf $p(x)$ and

$$\sum_x |x|\, p(x) < \infty,$$

then the **expectation** *of X is*

$$E(X) = \sum_x x\, p(x).$$

Sometimes the expectation $E(X)$ is called the **mathematical expectation** of X, the **expected value** of X, or the **mean** of X. When the mean designation is used, we often denote the $E(X)$ by μ; i.e, $\mu = E(X)$.

Example 1.8.1 (Expectation of a Constant)**.** Consider a constant random variable, that is, a random variable with all its mass at a constant k. This is a discrete random variable with pmf $p(k) = 1$. We have by definition that

$$E(k) = kp(k) = k. \quad \blacksquare \tag{1.8.1}$$

Example 1.8.2. Let the random variable X of the discrete type have the pmf given by the table

x	1	2	3	4
$p(x)$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{3}{10}$	$\frac{2}{10}$

Here $p(x) = 0$ if x is not equal to one of the first four positive integers. This illustrates the fact that there is no need to have a formula to describe a pmf. We have

$$E(X) = (1)\left(\frac{4}{10}\right) + (2)\left(\frac{1}{10}\right) + (3)\left(\frac{3}{10}\right) + (4)\left(\frac{2}{10}\right) = \frac{23}{10} = 2.3. \quad \blacksquare$$

Example 1.8.3. Let the continuous random variable X have the pdf

$$f(x) = \begin{cases} 4x^3 & 0 < x < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

Then

$$E(X) = \int_0^1 x(4x^3)\, dx = \int_0^1 4x^4\, dx = \left.\frac{4x^5}{5}\right|_0^1 = \frac{4}{5}. \quad \blacksquare$$

Remark 1.8.1. The terminology of expectation or expected value has its origin in games of chance. For example, consider a game involving a spinner with the numbers 1, 2, 3 and 4 on it. Suppose the corresponding probabilities of spinning these numbers are 0.20, 0.30, 0.35, and 0.15. To begin a game, a player pays \$5 to the "house" to play. The spinner is then spun and the player "wins" the amount in the second line of the table:

Number spun x	1	2	3	4
"Wins"	\$2	\$3	\$4	\$12
$G =$ Gain	−\$3	−\$2	−\$1	\$7
$p_G(x)$	0.20	0.30	0.35	0.15

"Wins" is in quotes, since the player must pay \$5 to play. Of course, the random variable of interest is the gain to the player; i.e., G with the range as given in the third row of the table. Notice that 20% of the time the player gains −\$3; 30% of the time the player gains −\$2; 35% of the time the player gains −\$1; and 15% of the time the player gains \$7. In mathematics this sentence is expressed as

$$(-3) \times 0.20 + (-2) \times 0.30 + (-1) \times 0.35 + 7 \times 0.15 = -0.50,$$

which, of course, is $E(G)$. That is, the expected gain to the player in this game is −\$0.50. So the player expects to lose 50 cents per play. We say a game is a **fair game**, if the expected gain is 0. So this spinner game is not a fair game. ■

Let us consider a function of a random variable X. Call this function $Y = g(X)$. Because Y is a random variable, we could obtain its expectation by first finding the distribution of Y. However, as the following theorem states, we can use the distribution of X to determine the expectation of Y.

Theorem 1.8.1. *Let X be a random variable and let $Y = g(X)$ for some function g.*

(a) *Suppose X is continuous with pdf $f_X(x)$. If $\int_{-\infty}^{\infty} |g(x)| f_X(x)\, dx < \infty$, then the expectation of Y exists and it is given by*

$$E(Y) = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx. \tag{1.8.2}$$

(b) *Suppose X is discrete with pmf $p_X(x)$. Suppose the support of X is denoted by $\mathcal{S}_X$. If $\sum_{x \in \mathcal{S}_X} |g(x)| p_X(x) < \infty$, then the expectation of Y exists and it is given by*

$$E(Y) = \sum_{x \in \mathcal{S}_X} g(x) p_X(x). \tag{1.8.3}$$

Proof: We give the proof in the discrete case. The proof for the continuous case requires some advanced results in analysis; see, also, Exercise 1.8.1.