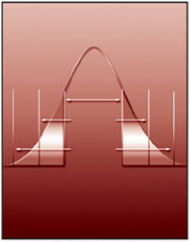


Introduction to Probability and Its Applications

THIRD EDITION

Richard L. Scheaffer
Linda J. Young

ADVANCED SERIES

This page intentionally left blank

THIRD EDITION

Introduction to Probability and Its Applications

RICHARD L. SCHEAFFER

University of Florida

LINDA J. YOUNG

University of Florida

BROOKS/COLE
CENGAGE Learning

Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

BROOKS/COLE
CENGAGE Learning

Introduction to Probability and Its Applications, Third Edition
Richard L. Scheaffer
Linda J. Young

To my wife, Nancy, for her unswerving patience and understanding over the many years that textbook writing has cut into our personal time together.
Richard L. Scheaffer

To my parents, Marvin and Wilma Cornette, who have always encouraged me to pursue my interests.
Linda J. Young

Publisher: Richard Stratton

Senior Sponsoring Editor: Molly Taylor

Associate Editor: Daniel Seibert

Editorial Assistant: Shaylin Walsh

Associate Media Editor: Catie Ronquillo

Senior Marketing Manager: Greta Kleinert

Marketing Coordinator: Erica O'Connell

Marketing Communications Manager: Mary Anne Payumo

Content Project Manager: Susan Miscio

Art Director: Linda Helcher

Senior Print Buyer: Diane Gibbons

Senior Rights Acquisition Account Manager, Text: Timothy Sisler

Production Service: S4Carlisle Publishing Services

Rights Acquisition Account Manager, Images: Don Schlotman

Photo Researcher: Jamie Jankowski

Cover Designer: Denise Davidson

Compositor: S4Carlisle Publishing Services

Library of Congress Control Number: 2009922034

ISBN-13: 978-0-534-38671-9
ISBN-10: 0-534-38671-7

Brooks/Cole
20 Channel Center Street
Boston, MA 02210
USA

Cengage Learning products are represented in Canada by Nelson Education, Ltd.

Printed in the United States of America
1 2 3 4 5 6 7 13 12 11 10 09

PREFACE

This continues to be a book designed for a one-semester course in probability for students with a solid knowledge of integral calculus. Development of the theory is mixed with discussion of the practical uses of probability. Numerous examples and problems, many involving real data, are provided for practice calculating probabilities in a variety of settings. Some problems, placed toward the end of an exercise set, allow the student to extend the theory. The level of the presentation has been kept as accessible as possible, and problem sets carefully progress from routine to more difficult to suit students of varying academic backgrounds. The text should provide a solid background for students going on to more advanced courses in probability and statistics and, at the same time, provide a working knowledge of probability for those who must apply it in engineering or the sciences.

When I began using the second edition of the text, I enjoyed the way that real-life applications were used to motivate and illustrate the many ways that probability affects all of our lives. In this third edition, every effort has been made to keep this focus on the great utility of probability for solving real-world problems. Numerous examples and problems showing the diversity of applications, especially in the biological sciences, have been added to complement those from the social and engineering sciences present in the second edition. The existing problems, often based on national data, such as surveys from the U.S. Census, were updated with the most current data available.

The discussion within the text has been expanded in a number of places, and a few additional topics are provided. A fuller discussion of sets as they relate to probability is included. An increased emphasis on the importance of whether or not sampling is with replacement and whether order matters is now in the sections on counting methods. Discrete multivariate distributions and transformations for discrete distributions have been added. For continuous random variables, the two-dimensional transformation method is now included. This should give a flavor for the types of changes that have been made. Short historical notes have been added to a number of sections.

The increased diversity of applications has led to a change in the parameterizations of the geometric and the negative binomial. These distributions are often good models for count data. Because the nonnegative integers are needed for most count data, the distributions have been defined as the number of failures until the first or rth success instead of the number of trials until that success as in the second edition.

Applets have been developed to illustrate some basic concepts and to allow for easy computations. These interactivities are referenced in the text and in some exercises and provide optional means for students to develop their understanding of these key concepts through active exploration. These and other resources are available on the text's companion Web site at http://www.cengage.com/statistics/scheaffer.

A Student Solutions Manual *(ISBN-10: 0-495-82974-9; ISBN-13: 978-0-495-82974-4)*, containing full worked solutions to all odd-numbered exercises in the text, is available for students to purchase. Complete instructor-only solutions to both the odd- and even-numbered exercises are available to adopters in password-protected online format by signing up for an account on Cengage Learning's Solution Builder service at http://www.cengage.com/solutionbuilder.

Finally, I would like to express thanks to some of the people who have helped make this edition possible. First, I felt honored that Dick Scheaffer gave me this opportunity. He encouraged me to make the changes that I thought would strengthen the book but, at the same time, was always ready to discuss ideas when I was not sure how best to proceed. The students in my classes provided good comments on what they liked and did not like. Ann Watkins used the text for a couple of semesters. Her feedback was detailed and caused two chapters to be completely rewritten, resulting in a stronger text. Yasar Yesiclay and Arthur Berg each used the text for a semester and identified additional weaknesses that we have corrected. The following reviewers provided valuable suggestions for improvements to the third edition manuscript: Jonathan Culkins, Carnegie Mellon University; Paul Roback, St. Olaf College; and Frederi Viens, Purdue University.

Others were instrumental in moving this third edition from the final draft to print, and I appreciate and want to recognize their efforts. Meghan Brennan did a detailed job authoring complete solutions for all exercises in the text. Araceli S. Popen copyedited the complete manuscript. Frank Wilson and Sue Steele provided a through accuracy check of all material in the text. Ben Langton developed the interactive applets. Susan Miscio and Sangeetha Sambasivam Rajan kept the project moving from copy editing through publication. Carolyn Crockett served as editor for this book in its early stages, and Daniel Seibert and Molly Taylor served as editors in the final stages.

Finally, I thank my husband Jerry who has shown great patience as I've spent many weekends and evenings working on this project.

Linda J. Young

CONTENTS

1

Probability in the World Around Us

1.1 Why Study Probability?

We live in an information society. We are confronted—in fact, inundated—with quantitative information at all levels of endeavor. Charts, graphs, rates, percentages, averages, forecasts, and trend lines are an inescapable part of our everyday lives. They affect small decisions we make every day, such as whether to wear a coat or to take an umbrella when we leave the house in the morning. They affect larger decisions on health, citizenship, parenthood, jobs, financial concerns, and many other important matters. Today, an informed person must have some facility for dealing with data and making intelligent decisions based on quantitative arguments that involve uncertainty or chance.

We live in a scientific age. We are confronted with arguments that demand logical, scientific reasoning, even if we are not trained scientists. We must be able to make our way successfully through a maze of reported "facts," in order to separate credible conclusions from specious ones. We must be able to weigh intelligently such issues as the evidence on the causes of cancer, the effects of pollutants on the environment, and the likely results of eating genetically modified plants and animals.

We live amidst burgeoning technology. We are confronted with a job market that demands scientific and technological skills, and students must be trained to deal with the tools of this technology productively, efficiently, and correctly. Much of this new technology is concerned with information processing and dissemination, and proper use of this technology requires probabilistic skills. These skills are in demand in engineering, business, and computer science for jobs involving market research, product development and testing, economic forecasting, credit research, quality control, reliability, business management, and data management, to name just a few.

Few results in the natural or social sciences are known absolutely. Most are reported in terms of chances or probabilities: the chance of rain tomorrow, the chance of your getting home from school or work safely, the chance of your living past 60 years of age, the chance of contracting (or recovering from) a certain disease, the chance of inheriting a certain trait, the chance of your annual income exceeding \$60,000 in two years, the chance of winning an election. Today's adults must obtain some knowledge of probability and must be able to tie probabilistic concepts to real scientific investigations if they are to understand science and the world around them.

This book provides an introduction to probability that is both mathematical, in the sense that the underlying theory is developed from axioms, and practical, in the sense that applications to real-world problems are discussed. The material is designed to provide a strong basis in probability for students who may go on to deeper studies of statistics, mathematics, engineering, business, or the physical and biological sciences; at the same time, it should provide a basis for practical decision making in the face of uncertainty.

1.2 Deterministic and Probabilistic Models

1.2.1 Modeling Reality

It is essential that we grasp the difference between theory and reality. Theories are ideas proposed to explain phenomena in the real world. As such they are approximations or models of reality. Sometimes theories are wrong. For example, prior to Copernicus, scientists believed the theory that the sun and other planets revolved around the earth. Copernicus was the first to recognize that the earth and other planets revolved around the sun, but he believed that the orbits of the planets were circular. Thus, Copernicus's theory (or model) of the solar system was closer to, but not the same as, reality. Through the scientific process, theories are constantly refined so that they become closer to reality.

Theories are presented in verbal form in some (less quantitative) fields and as mathematical relationships in others. Thus, a theory of social change might be expressed verbally in sociology, whereas the theory of heat transfer is presented in a precise and deterministic mathematical manner in physics. Neither gives an accurate and unerring explanation for real life, however. Slight variations from the mathematically expected can be observed in heat-transfer phenomena and in other areas of physics. The deviations cannot be blamed solely on the measuring instruments (the explanation that one often hears); they are due in part to a lack of agreement between theory and reality. These differences between theory and reality led George Box (1979, 202), a famous statistician, to note: "All models are wrong, but some are useful."

In this text, we shall develop certain theoretical models of reality. We shall attempt to explain the motivation behind such a development and the uses of the resulting models. At the outset, we shall discuss the nature and importance of model building in the real world to convey a clear idea of the meaning of the term *model* and of the types of models generally encountered.

1.2.2 Deterministic Models

Suppose that we wish to measure the area covered by a lake that, for all practical purposes, has a circular shoreline. Because we know that the area A is given by $A = \pi r^2$, where r is the radius, we attempt to measure the radius (perhaps by averaging a number of measurements taken at various points), and then we substitute the value obtained into the formula. The formula $A = \pi r^2$, as used here, constitutes a *deterministic model*. It is *deterministic* because, once the radius is known, the area is assumed to be known. It is a *model* of reality because the true border of the lake has some irregularities and therefore does not form a true circle. Even though the planar object in question is not exactly a circle, the model identifies a useful relationship between the area and the radius that makes approximate measurements of area easy to calculate. Of course, the model becomes poorer and poorer as the shape of the figure deviates more and more from that of a circle until, eventually, it ceases to be of value and a new model must take over.

Another deterministic model is Ohm's law, $I = V/R$, which states that electric current I is directly proportional to the voltage V and inversely proportional to the resistance R in a circuit. Once the voltage and the resistance are known, the current can be determined. If we investigated many circuits with identical voltages and resistances, we might find that the current measurements differed by small amounts from circuit to circuit, owing to inaccuracies in the measuring equipment or other uncontrolled influences. Nevertheless, any such discrepancies are negligible, and Ohm's law thus provides a useful deterministic model of reality.

1.2.3 Probabilistic Models

Contrast the two preceding situations with the problem of tossing a balanced coin and observing the upper face. No matter how many measurements we may make on the coin before it is tossed, we cannot predict with absolute accuracy whether the coin will come up heads or tails. However, it is reasonable to assume that, if many identical tosses are made, approximately $1/2$ will result in outcomes of heads; that is, we cannot predict the outcomes of the next toss, but we can predict what will happen in the long run. We sometimes convey this long-run information by saying that the "chance" or "probability" of heads on a single toss is $1/2$. This probability statement is actually a formulation of a probabilistic model of reality. *Probabilistic models* are useful in describing experiments that give rise to random, or chance, outcomes. In some situations, such as the tossing of an unbalanced coin, preliminary experimentation must be conducted before realistic probabilities can be assigned to the outcomes. It is possible, however, to construct fairly accurate probabilistic models for many real-world phenomena. Such models are useful in varied applications, such as in describing the movement of particles in physics (Brownian motion), in explaining the changes in the deer population within a region, and in predicting the profits for a corporation during some future quarter.

1.3 Applications in Probability

We shall now consider two uses of probability theory. Both involve an underlying probabilistic model, but the first hypothesizes a model and then uses this model for practical purposes, whereas the second deals with the more basic question of whether the hypothesized model is in fact a correct one.

Suppose that we attempt to model the random behavior of the arrival times and lengths of service for patients at a medical clinic. Such a mathematical function would be useful in describing the physical layout of the building and in helping us determine how many physicians are needed to service the facility. Thus, this use of probability assumes that the probabilistic model is known and offers a good characterization of the real system. The model is then employed to enable us to infer the behavior of one or more variables. The inferences will be correct—or nearly correct—if the assumptions that governed construction of the model were correct.

The problem of choosing the correct model introduces the second use of probability theory, and this use reverses the reasoning procedure just described. Assume that we do not know the probabilistic mechanism governing the behavior of arrival and service times at the clinic. We might then observe a hospital's emergency room and acquire a sample of arrival and service times for emergency room patients. Based on the sample data, inferences can be drawn about the nature of the underlying probabilistic mechanism; this type of application is known as *statistical inference*. This book deals mostly with problems of the first type but, on occasion, it makes use of data as a basis for model formulation. Ideally, readers will go on to take a formal course in statistical inference later in their academic studies.

Consider the problem of replacing the light bulbs in a particular socket in a factory. A bulb is to be replaced either at failure or at a specific age T, whichever comes first. Suppose that the cost c_1 of replacing a failed bulb is greater than the cost c_2 of replacing a bulb at age T. This may be true because in-service failures disrupt the factory, whereas scheduled replacements do not. A simple probabilistic model enables us to conclude that the average replacement cost C_a per unit time, in the long run, is approximately

$$C_a = \frac{1}{\mu}[c_1(\text{Probability of an in-service failure}) \\ + c_2(\text{Probability of a planned replacement})]$$

where μ denotes the average service time per bulb. The average cost is a function of T. If μ and the indicated probabilities can be obtained from the model, a value of T can be chosen to minimize this function. Problems of this type are discussed more fully in Chapter 9.

Biological populations are often characterized by birth rates, death rates, and a probabilistic model that relates the size of the population at a given time to these rates. One simple model allows us to show that a population has a high probability of becoming extinct even if the birth and death rates are equal. Only if the birth rate exceeds the death rate might the population exist indefinitely.

Again referring to biological populations, models have been developed to explain the diffusion of a population across a geographic area. One such model concludes that the square root of the area covered by a population is linearly related to the length of time the population has been in existence. This relationship has been shown to hold reasonably well for many varieties of plants and animals.

Probabilistic models like those mentioned give scientists a wealth of information for explaining and controlling natural phenomena. Much of this information is intuitively clear, such as the fact that connecting identical components in series reduces the system's expected life length compared to that of a single component, whereas parallel connections increase the system's expected life length. But many results of probabilistic models offer new insights into natural phenomena—such as the fact that, if a person has a 50:50 chance of winning on any one trial of a gambling game, the excess of wins over losses will tend to stay either positive or negative for long periods, given that the trials are independent. (That is, the difference between number of wins and number of losses does not fluctuate rapidly from positive to negative.)

1.4 A Brief Historical Note

The study of probability has its origins in games of chance, which have been played throughout recorded history and, no doubt, during prehistoric times as well. The astragalus, a bone that lies above the talus (heel bone), was used in various board games in Egypt (c. 3500 B.C.). A game called "hounds and jackals" by the excavators of Egyptian tombs apparently used astragali in the same manner as dice. The hounds and jackals were moved on the board according to the results of throwing the astragali. Homer (c. 900) reported that Patroclus became angry with his opponent while playing a game based on astragali and nearly killed him (David 1955, Folks 1981).

The ancient Greeks used the knucklebones of sheep or goats, as illustrated in Figure 1.1, to make astragali for games of chance. The Romans used the term *tali*, which is the Latin name for knucklebones. They made tali from brass, silver, gold, ivory, marble, wood, bone, bronze, glass, terracotta, and precious gems. When tossed, the astragali, or tali, would land on one of four sides. The most popular game resembled modern dice.

FIGURE **1.1**
Astragali as used in ancient Greece.

Leicestershire Museums Service

FIGURE 1.2
Bronze Roman die made in the first through the third centuries A.D. Note that opposite sides are 1-6, 2-3, and 4-5.

Prisma/Ancient Art & Architecture Collection Ltd.

Gambling was so popular in Roman times that laws had to be passed to regulate it. The Church's stern opposition followed and continues today. Yet gaming and gambling have flourished among all classes of people. Most early games involved the astragalus. Dice and cards were used later. Dice dating from the beginning of the third millennium are the earliest found so far. A die made from well-fired buff potter and found in Iraq has the opposite points in consecutive order: 2 opposite 3, 4 opposite 5, and 6 opposite 1 (Figure 1.2). Die with opposite faces totaling 7 must have evolved about 1400 B.C. (David 1955, Folks 1981).

In 1494, Fra Luca Pacciolo published a mathematical discussion of the problem of points, that is, the problem of how to divide equitably the stakes between two players when a game is interrupted before its conclusion. However, it is generally agreed that a major impetus to the formal study of probability was provided by the Chevalier de Méré when he posed a problem of points to the famous mathematician Blaise Pascal (1623–1662). The question was along the following lines: To win a particular game of chance, a gambler must throw a 6 with a die; he has eight throws in which to do it. If he has no success on the first three throws, and the game is thereupon ended prematurely, how much of the stake is rightfully his? Pascal cast this problem in probabilistic terms and engaged in extensive correspondence with another French mathematician, Pierre de Fermat (1608–1665), about its solution. This correspondence began the formal mathematical development of probability theory.

In addition to the use of games, chance mechanisms have been used to divine the will of the gods. In 1737, John Wesley sought guidance by drawing lots to decide whether or not to marry. Even today, a group will sometimes draw straws with the agreement that the person drawing the short straw will do the unpleasant task at hand.

Notice that so far we have been talking about "games of chance" and the "chance" of certain events occurring. According to the *Oxford English Dictionary*, the use of "chance" to convey probabilistic ideas dates back to the turn of the fourteenth century. Although the term "probability" can be traced back to the midsixteenth century, de Moivre related the two terms in the opening proposition of *The Doctrine of Chances: or, a Method of Calculating the Probability of Events in Play* (1718): "The Probability of an Event is greater or less, according to the number of Chances by which it may happen, compared with the whole number of Chances by which it may happen or fail." That is, chances are counted and used to determine probabilities.

Scientists in the eighteenth and nineteenth centuries continued to develop the probability theory and recognized its usefulness in solving important problems in science. Yet probability was lacking a solid foundation on which to build a major branch of mathematical research until 1933 when Russian-born A. N. Kolmogorov presented the three axioms of probability. Although the three axioms are intuitive ones, as we shall see in the next chapter, they provided the needed foundation on which to build the modern study of probability. Although games of chance continue to illustrate interesting applications of probability, applications of probability theory have spread to virtually every corner of scientific research.

1.5 A Look Ahead

This text is concerned with the theory and applications of probability as a model of reality. We shall postulate theoretical frequency distributions for populations and develop a theory of probability in a precise mathematical manner. The net result will be a theoretical or mathematical model for acquiring and utilizing information in real life. It will not be an exact representation of nature, but this should not disturb us. Like other theories, its utility should be gauged by its ability to assist us in understanding nature and in solving problems in the real world. Whether we are assessing how the likelihood an individual has health insurance changes (or does not change) with the person's age; counting the numbers of insects on plants in a cotton field; predicting the time until the next hurricane; or exploring the relationship between the use of "soda," "coke," or "pop" and the region of the country, probability theory can be used to provide insight into many areas of our daily lives.

REFERENCES

Box, G. E. P. 1979. Robustness in scientific model building. In *Robustness in statistics*, ed. R. L. Launer and G. N. Wilkinson, 201–236. New York: Academic Press.

David, F. N. 1955. Studies in the history of probability and statistics. *Biometrika* 52:1–15.

de Moivre, A. 1718. *The doctrine of chances, or, a method of calculating the probability of events in play*. London: W. Pearson.

Folks, J. L. 1981. *Ideas of statistics*. New York: John Wiley & Sons.

2

Foundations of Probability

2.1 Randomness

2.1.1 Randomness with Known Structure

At the start of a football game, a balanced coin is flipped in the air to decide which team will receive the ball first. What is the chance that the coin will land heads up? Most of us would say that this chance or probability is 0.5, or something very close to that. But, what is the meaning of this number 0.5? If the coin is flipped 10 times, will it come up heads exactly 5 times? Upon deeper reflection, we recognize that 10 flips need not result in exactly 5 heads; but in repeated flipping, the coin should land heads up approximately $1/2$ of the time. From there on, the reasoning begins to get fuzzier. Will 50 flips result in exactly 25 heads? Will 1000 flips result in exactly 500 heads? Not necessarily, but the fraction of heads should be close to $1/2$ after "many" flips of the coin. So the 0.5 is regarded as a *long-run* or *limiting relative frequency* as the number of flips gets large.

John Kerrich (1964) was an English mathematician visiting Copenhagen when World War II began. Two days before he was to leave for England the Germans invaded Denmark. He spent the rest of the war interned in a prison camp in Jutland. To pass the time, Kerrich conducted a series of experiments in probability theory. In one, he flipped a coin 10,000 times, keeping a careful tally of the number of heads. After 10 flips he had 4 heads, which is a relative frequency of 0.4; after 100 flips he had 44 heads (0.44); after 1000 flips he had 502 heads (0.502); and after 10,000 flips he had 5067 heads (0.5067). From 1000 to 10,000 flips, the relative frequency of heads remained very close to 0.5, although the recorded relative frequency at 10,000 was slightly farther from 0.5 than it was at 1000.

In the long run, Kerrich obtained a relative frequency of heads close to 0.5. For that reason, the number 0.5 can be called the **probability** of obtaining a head on the flip of a balanced coin. Another way of expressing this result is to say that Kerrich **expected** to see about 5000 heads among the outcomes of his 10,000 flips. He actually came close to his expectations and so have others who have repeated the coin-flipping study. This idea of a stabilizing relative frequency after many trials lies at the heart of random behavior.

Sometimes, instead of actually flipping the coin, the flips can be simulated. The simulation can be conducted in numerous ways. One approach would be to use a table of random digits (such as Table 1 in the appendix) or random digits generated by computer. These random digits are produced according to the following model. Think of 10 equal-sized chips, numbered from 0 to 9, with one number per chip, and thoroughly mixed in a box. Without looking, someone reaches into the box and pulls out a chip, recording the number on the chip. That process constitutes a single draw of a random digit. Putting the chip back in the box, mixing the chips, and then drawing another chip produces a second random digit. A random number table or a string of random digits from a computer is the result of hundreds of such draws, each from the same group of thoroughly mixed chips.

To simulate the flip of a coin using random digits, some of the digits must be associated with a *head* and some with a *tail.* Because we believe the chance of obtaining a head is the same as that of getting a tail, we would assign half of the 0 to 9 digits to represent a head; the other half would represent a tail. Whether we assign the even numbers to be heads and the odds to be tails or the digits 0 to 4 to be heads and those that are 5 to 9 to be tails or some other assignment does not matter as long as it is clearly determined which of the five digits will represent heads and which will represent tails.

Once the digits have been assigned to heads and tails we are ready to simulate a flip. Suppose the even numbers represent a head and the odd numbers represent a tail. To simulate 10 flips of the coin, we need to have 10 random digits; suppose they are as follows:

0 7 9 7 2 1 3 8 7 6

These would produce the corresponding sequence of simulated coin flips:

H T T T H T T H T H

From these 10 simulated coin flips, the estimated probability of heads is $4/10 = 0.4$. One advantage of simulation is that the process is often more rapid than actually conducting the study, which, in this case, is flipping the coin. This allows many flips to be simulated, and it facilitates multiple simulations.

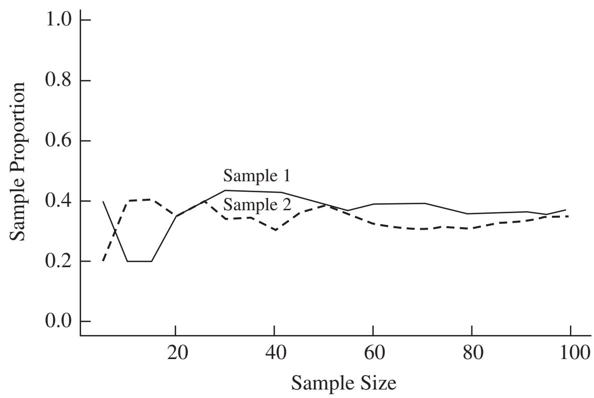
The *Relative Frequency* applet can be used to simulate experiments that have two possible outcomes and for which the probability of the outcome of interest is assumed known. Use the applet to simulate 100 flips of a fair coin. *Note*: Because we expect about half of the flips to result in a head on the upper face, we anticipate that the probability of success p will be 0.5. Repeat 100 flips of the coin without closing the window. Notice that the outcome of the first 100 flips is still visible in light gray in the background. Compare the relative frequencies of observing a head as a function of the number of coins flipped. Now run many simulations of flipping a fair coin 100 times.

You should observe two important features. First, the sample paths oscillate greatly for small numbers of observations and then settle down around a value close to 0.5. Second, the variation between the sample fractions is quite large for small sample sizes and quite small for the larger sample sizes. From these data, we can approximate the probability of observing a head on a single flip of the coin, recognizing that the approximation tends to get better as the sample size increases.

In the previous example, half of the digits between 0 and 9 were assigned to heads and the other half to tails because we believe that the two outcomes have the same chance of occurring. Suppose now we want to determine the probability of selecting a digit that is a multiple of 3 from a random number table? We can find this probability, at least approximately, by selecting digits from a random number table and counting the number of multiples of 3 (3s, 6s, and 9s). Figure 2.1 shows the results of two different attempts at doing this with 100 digits each. The relative frequencies are recorded as a function of sample size.

As with the coin-flipping simulations, the sample paths show great changes for small numbers of observations and then settle down as the numbers of observations increase, and the two sample paths differ more for small sample sizes and less as the sample size increases.

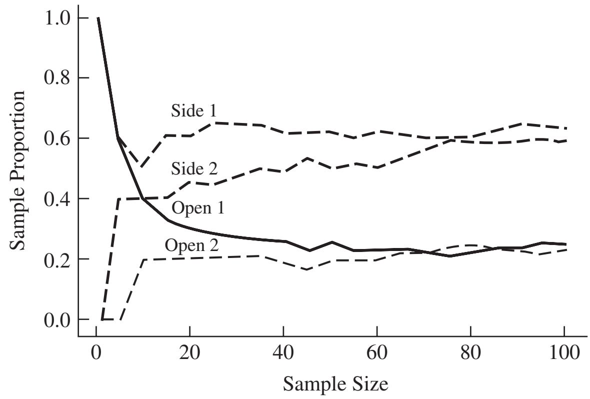
FIGURE 2.1
Proportion of multiples of 3.



2.1.2 Randomness with Unknown Structure

For the two examples discussed in the preceding subsection, we knew what the resulting long-run relative frequency should be. Now, we will consider an example for which the result is less than obvious before the data are collected. A standard paper cup, with the open end slightly larger than the closed end, was tossed in the air and allowed to land on the floor. The goal was to approximate the probability of the cup's landing on the open end. (You might want to generate your own data here

FIGURE **2.2**
Proportions of cup landings.



before looking at the results presented next.) After 100 tosses on each of two trials, the sample paths looked like those shown in Figure 2.2.

Notice that the pattern observed in the other two examples occurs here, too; both sample paths seem to stabilize around 0.2, and the variability between the paths is much greater for small sample sizes than for large ones. We can now say that a tossed cup's probability of landing on its open end is approximately 0.2; in 100 tosses of this cup, we would expect to see it land on its open end about 20 times. Figure 2.2 also includes two sample paths for the outcome "landed on side," which seem to stabilize at slightly less than 0.6. Thus, the basic notion of probability as a long-run relative frequency works here just as well as it did in the cases for which we had a good theoretical idea about what the relative frequency should be!

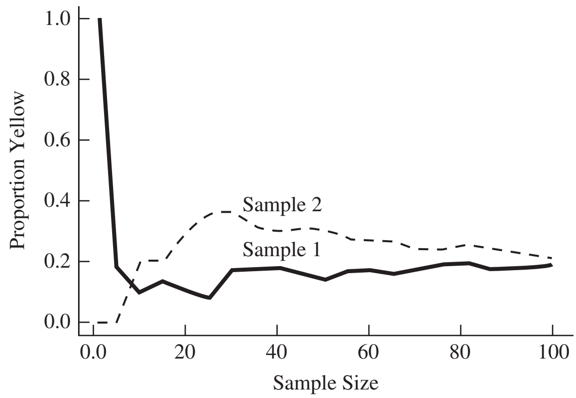
Notice that the approximate probabilities for the cup toss depend on the dimension of the cup, the material of which the cup is made, and the manner of tossing the cup. Another experiment with a different cup should generate different approximate probabilities.

Note that simulation is not helpful in this setting. Because we do not have a theoretical idea about what the relative frequency of the cup landing on its end should be, we do not know what proportion of the random digits should be assigned to the outcome of interest, nor can we specify the probability of success in the *Relative Frequency* applet.

2.1.3 Sampling a Finite Universe

In the preceding subsection, there was no limit to the number of times that the coin could have been flipped, the random digits could have been drawn, or the paper cup could have been tossed. The set of possible sample values in such cases is both infinite and conceptual; that is, it does not exist on a list somewhere. (A list of random digits may exist, but it could always be made larger.) Now, however, consider a jar of differently colored beads that sits on my desk. I plan to take samples of various sizes from this jar with a goal of estimating the proportion of yellow beads in the jar. After mixing the beads, I take a sample of 5, then 5 more (to make a total of 10), and so on, up to a total of 100 beads. The beads sampled are *not* returned to the jar before the next group is selected. Will the relative frequency of yellow beads stabilize here, as it

FIGURE 2.3
Proportion of yellow beads.



did for the infinite conceptual universes sampled earlier? Figure 2.3 shows two actual sample paths for this type of sampling from a jar containing more than 600 beads.

The same properties observed earlier in the infinite cases do, indeed, hold here as well. Both sample paths vary greatly for small sample sizes but stabilize around 0.2 as the sample size gets larger. That leads us to conclude that the probability of randomly sampling a yellow bead is about 0.2; a sample of 100 beads can be expected to contain about 20 yellows (or equivalently, the jar can be expected to contain more than 120 yellows). Actually, the jar contains 19.6% yellow beads, so the sampling probabilities came very close to the true proportion.

In general, as long as the selection mechanism remains random and consistent for all selections, the sample proportion for a certain specified event will eventually stabilize at a specific value that can be called the probability for the event in question.

If n is the number of trials of an experiment (such as the number of flips of a coin), one might define the probability of an event E (such as observing an outcome of heads) by

$$P(E) = \lim_{n\to\infty} \frac{\text{Number of times } E \text{ occurs}}{n}.$$

But will this limit always converge? If so, can we ever determine what it will converge to without actually conducting the experiment many times? For these and other reasons, this is not an acceptable mathematical definition of probability, although it is a property that should hold in some sense. Another definition must be found that allows such a limiting result to be proved as a consequence. This is done in Section 2.3.

Exercises

2.1 Conduct numerous simulations of flipping a fair coin 1000 times using the *Relative Frequency* applet. Then, in a different window, conduct numerous simulations of 1000 flips of a biased coin for which the probability of a head is 0.9. The observed proportion of heads varies greatly initially and then settles around the true probability in both cases. Does there appear to be a difference in how rapidly the proportion settles down for the probabilities of 0.5 and 0.9?

2.2 Conduct numerous simulations of flipping a fair coin 1000 times using the *Relative Frequency* applet. Then, in a different window, conduct numerous simulations of 1000 flips of a biased coin for which the probability of a head is 0.1. The observed proportion of heads varies greatly initially and then settles around the true probability in both cases. Does there appear to be a difference in how rapidly the proportion settles down for the probabilities of 0.1 and 0.5?

2.3 Based on the results in exercises 2.1 and 2.2 and perhaps additional simulations, speculate on the behavior of the rate at which an observed proportion approaches the true probability of a head as that probability ranges between 0 and 1.

2.4 Verify your conclusions in 2.3 using $p = 0.75$.

2.2 Sample Space and Events

Consider again the process of flipping a coin. Given all of the conditions associated with flipping the coin (distance from the floor when flipped, height of the flip, number of spins, air currents, etc.), the outcome can be completely determined by the laws of physics. Yet in repeated flips, under what appears to be identical conditions, the coin lands *head* about half the time and *tail* the other half of the time. We can imagine an infinite population of all possible coin flips. Each coin flip in this population can be categorized as being a *head* or a *tail*, and the random process of flipping the coin allows us to make a meaningful statement about the probability of *head* on the next flip. As illustrated here, probability requires three elements: a target population (conceptual or real) from which observable outcomes are obtained, meaningful categorizations of these outcomes, and a random mechanism for generating outcomes.

To change the setting a bit, suppose that a regular six-sided die is rolled onto a table and the number on the upper face is observed. This is a probabilistic situation because the number that shows on the upper face cannot be determined in advance. We shall analyze the components of the situation and arrive at a definition of probability that permits us to model mathematically what happens in die rolls as well as in many similar situations.

First, we might roll the die several times to collect data on possible outcomes. This data-generating phase, called a *random experiment*, allows us to see the nature of the possible outcomes, which we can list in a **sample space**.

DEFINITION **2.1**

A **sample space** S is a set that includes all possible outcomes for a random experiment listed in a mutually exclusive and exhaustive way. ▪

The phrase *mutually exclusive* means that the outcomes of the set do not overlap. The set $S^* = \{1, 2, 3, 4, 5, 6, \text{even}, \text{odd}\}$ is not an appropriate sample space for the die-roll experiment because the outcome of two dots on the upper face is represented by “2” and “even” in this set; that is, the outcomes are not mutually exclusive. The term *exhaustive* in Definition 2.1 means that the list contains all possible outcomes. The set $S^{\sim} = \{1, 2, 3, 4, 5\}$ is not an appropriate sample space for the experiment of rolling a die because “6,” which is one of the possible outcomes, is not represented in the sample space.

For the die roll, we could identify a sample space of

$$S_1 = \{1, 2, 3, 4, 5, 6\}$$

where the integers indicate the possible numbers of dots on the upper face, or of

$$S_2 = \{\text{even}, \text{odd}\}.$$

Both S_1 and S_2 satisfy Definition 2.1, but S_1 seems the better choice because it maintains a higher level of detail. S_2 has three possible upper-face outcomes in each listed element, whereas S_1 has only one possible outcome per element.

As another example, suppose that a nurse is measuring the height of a patient. (This measurement process constitutes the experiment; the patient is a member of the target population.) The sample space for this experiment could be given as

$$S_3 = \{1, 2, 3, \ldots, 50, 51, 52, \ldots, 70, 71, 72, \ldots\}$$

if the height is rounded to the closest integer number of inches. On the other hand, a more appropriate sample space could be

$$S_4 = \{x|x > 0\}$$

which is read, "the set of all real numbers x such that $x > 0$." Whether S_3 or S_4 should be used in a particular problem depends on the nature of the measurement process. If decimals are to be used, we need S_4. If only integers are to be used, S_3 will suffice. The point is that sample spaces for a particular experiment are not unique and must be selected to provide all pertinent information for a given situation.

In an experiment, we are generally interested in a particular outcome or set of outcomes; these are the events of interest.

DEFINITION **2.2**

An **event** is any subset of a sample space. ▪

Let us go back to our first example, the roll of a die. Suppose that player A can have the first turn at a board game if he or she rolls a 6. Therefore, the event "roll a 6" is important to that player. Other possible events of interest in the die-rolling experiment are "roll an even number," "roll a number greater than 4," and so on. Definition 2.2 holds as stated for any sample space that has a finite or a countable number of outcomes. Some subsets must be ruled out if a sample space covers a continuum of real numbers, as S_4 (given earlier) does, but any subset likely to occur in practice can be called an **event.** Because the null set (the set with no outcomes and often denoted by ϕ) is a subset of every set, the null set is an event associated with every sample space. Similarly, because every set is a subset of itself, the sample space is an event in every experiment. If a set has n elements, it has 2^n possible subsets. Thus, if a sample space has n outcomes, it has 2^n possible events.

If the outcome of an experiment is contained in the event E, then the event E is said to have *occurred*. Events are usually denoted with capital letters at the beginning of the alphabet.

For the die-rolling experiment, suppose we are interested in the following events:

A is the event of "an even number."
B is the event of "an odd number."
C is the event of "a number greater than 4."
D is the event of "a number greater than 8."
E_1 is the event of "observe a 1."
E_i is the event of "observe an integer *i*."

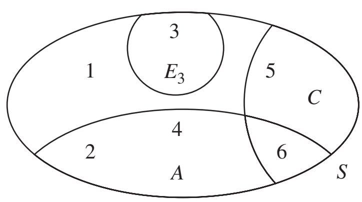
Then, if $S = \{1,\ 2,\ 3,\ 4,\ 5,\ 6\}$,

$$\begin{aligned} A &= \{2, 4, 6\} \\ B &= \{1, 3, 5\} \\ C &= \{5, 6\} \\ D &= \phi \\ E_1 &= \{1\} \\ E_i &= \{i\}, \quad i = 1,\ 2,\ 3,\ 4,\ 5,\ 6. \end{aligned}$$

Notice that we would not have been able to fully define the events C, E_1, or E_i if we had used the sample space S_2.

Now suppose we roll a die, and we observe a 4 on the upper face. Which of the above events have occurred? Of course, E_4 has occurred because 4 is the only outcome in that event. However, the event *A* has also occurred because 4 is an outcome in that event. This may seem counterintuitive at first. After all, neither 2 nor 6 were observed, and they are also outcomes in this event. The following example may clarify this concept. Suppose you are playing a game in which you win if, after a die is rolled, the number of dots observed on the upper face is even; that is, you win if event *A* occurs. If a 4 is rolled, then you win. It is not necessary (or even possible) to get the other even numbers (2 and 6) on that one roll; it is enough to get any one of the three even numbers. Thus, if a 4 appears on the upper face, the event *A* is said to have occurred. Similarly, the event *S* (the sample space) has occurred. *Note*: For an event to occur, every outcome in the event need not be an observed result from the experiment; it is only necessary for the outcome to be an outcome of the event. Because *S* is the set of all possible outcomes, it is sometimes called the *sure event*; it will surely occur when the experiment is conducted. ϕ will never occur and is sometimes called the *impossible event* or the *null event* because it has no element that could be an outcome of the experiment.

FIGURE **2.4**
Venn diagram for a die roll.



Sample spaces and events often can be conveniently displayed in Venn diagrams. Some events for the die-rolling experiment are shown in Figure 2.4.

EXAMPLE 2.1 A small company has two automobiles in its car pool. Each automobile may break down (only once) or not break down on any given day. An experiment consists of counting the number of automobile breakdowns to occur on a randomly selected day.

1 Specify a sample space for the experiment.

2 List all possible events.

3 Suppose one car breaks down on a randomly selected day. List the event(s) in part (2) that has (have) occurred.

Solution **1** Because only two cars are in the car pool, neither, one, or both may break down on a randomly selected day. That is, $S = \{0, 1, 2\}$.

2 Recall that if a sample space has n outcomes, it has 2^n possible events. Here S has three outcomes, so there are $2^3 = 8$ events: ϕ, $\{0\}$, $\{1\}$, $\{2\}$, $\{0, 1\}$, $\{0, 2\}$, $\{1, 2\}$, and S.

3 If one car breaks down, any event with the outcome 1 as a member will have occurred; that is, $\{1\}$, $\{0, 1\}$, $\{1, 2\}$, and S. Remember: Every outcome of an event need not be observed for the event to occur; if any outcome in an event is observed, the event has occurred.

■

Because sample spaces and events are sets, we need to be familiar with set notation when working with them so let us review the set notation that we will use. For the die-rolling experiment, recall $S = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 4, 6\}$, $B = \{1, 3, 5\}$, and $C = \{5, 6\}$. Every event is a subset of the sample space S. For event A in the die-rolling study, this would be denoted by $A \subset S$, signifying that every outcome in A is also an outcome in S; that is, A is "contained in" S. We denote the fact that outcome 2 is an outcome in A by $2 \in A$. Two events, A and B, are equal if they are each subsets of the other, that is $A \subset B$ and $B \subset A$; that is, every outcome in A is an outcome in B, and every outcome in B is also an outcome in A.

Sometimes we want to know whether at least one of two or more events has occurred. Suppose we have two events, A and B, of a sample space S, and we want to know whether event A or event B, or both, have occurred. This new event, denoted by $A \cup B$, consists of all outcomes (elements) in A or in B, or in both. Thus the event $A \cup B$ will occur if A occurs, B occurs, or both occur. The event $A \cup B$ is called the *union* of the event A and the event B. For example, for the die-rolling experiment,

$$A \cup B = \{1, 2, 3, 4, 5, 6\} = S$$

and

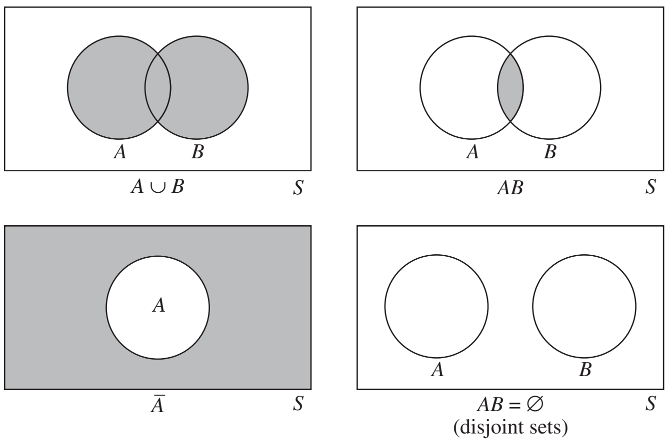
$$A \cup C = \{2, 4, 5, 6\}.$$

At times, we are interested in whether or not two events have occurred. If we have two events, A and B, of a sample space, $A \cap B$ (or merely AB) is the new event consisting of outcomes (elements) in A *and* in B. That is, the event AB occurs only if both A and B occur. The event AB is called the *intersection* of the set A and the set B. For the die-rolling experiment, $BC = \{5\}$ occurs only if a 5 is observed on the upper face after the roll of the die because it is the only outcome in both B and C. The events A and B have no outcomes in common, $AB = \phi$; that is, it is not possible for both A and B to occur on a single roll of a die.

The complement of the event A with respect to sample space S is the set of all outcomes in S that are not in A; the complement of A is denoted by $\bar{A}$. The event $\bar{A}$ will occur if and only if A does not occur. For the die-rolling experiment, $\bar{A} = \{1, 3, 5\} = B$. Notice that $\bar{S} = \phi$ and $\bar{\phi} = S$; that is, the sure event and the null event are complements of each other. Two events A and B are said to be mutually exclusive, or disjoint, if they have no outcomes in common, that is, $A \cap B = AB = \phi$. A and B are mutually exclusive events for the die-rolling experiment.

Venn diagrams can be used to portray graphically the concepts of union, intersection, complement, and disjoint sets (Figure 2.5).

FIGURE **2.5**
Venn diagrams of set relations.



We can easily see from Figure 2.5 that

$$A \cup \bar{A} = S$$

for any event A.

The operations of forming unions, intersections, and complements of events obey certain rules that are similar to the rules of algebra, including the commutative, associative, and distributive laws:

Commutative laws: $A \cup B = B \cup A$ and $AB = BA$

Associative laws: $(A \cup B) \cup C = A \cup (B \cup C)$ and $(AB)C = A(BC)$

Distributive laws: $A(B \cup C) = AB \cup AC$ and
$A \cup (BC) = (A \cup B) \cap (A \cup C)$

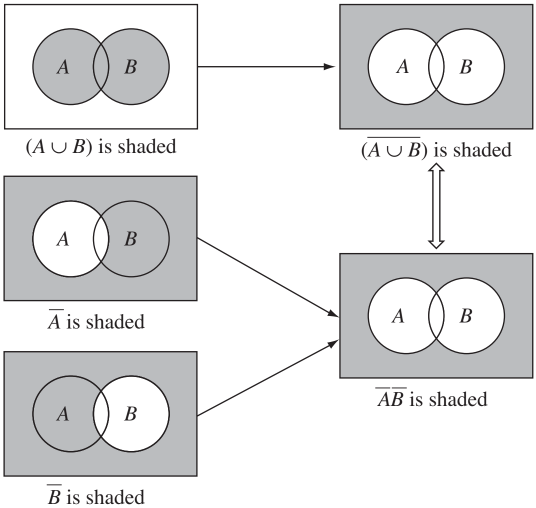


FIGURE **2.6** Showing equality of DeMorgan's first law for two sets using Venn diagrams.

DeMorgan's laws: $\overline{A \cup B} = \bar{A}\bar{B}$ or $\overline{\left(\bigcup_{i=1}^{n} A_i\right)} = \bigcap_{i=1}^{n} \bar{A}_i$

$\overline{A \cap B} = \bar{A} \cup \bar{B}$ or $\overline{\left(\bigcap_{i=1}^{n} A_i\right)} = \bigcup_{i=1}^{n} \bar{A}_i$

These laws and other relationships may be verified by showing that any outcome that is contained in the event to the left of the equality sign is also contained in the event to the right of the sign, and vice versa. Though not a rigorous proof, one way of verifying a relationship is by using Venn diagrams. For example, the first of DeMorgan's laws for two sets may be verified by the sequence of diagrams in Figure 2.6.

EXAMPLE **2.2** After a hurricane, 50% of the residents in a particular Florida county were without electricity, 47% were without water, and 38% were without telephone service. Although 20% of the residents had all three utilities, 10% were without all three, 12% were without electricity and water but still had a working telephone, and 4% were without electricity and a working telephone but still had water. A county resident is randomly selected and whether or not the person was without water, electricity, or telephone service after the hurricane is recorded. Express each of the following events in set notation and find the percent of county residents represented by each.

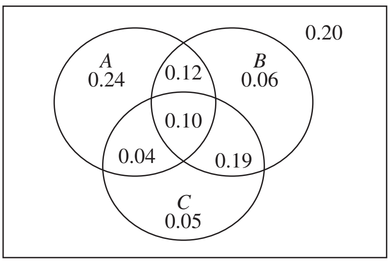
1 The selected county resident was without electricity only but had water and telephone service.

2 The selected county resident was without only one of the utilities but had the other two.

3 The selected county resident was without exactly two of the utilities.

4 The selected county resident was without at least one of the utilities.

Solution In problems of this type, it is best to start with a Venn diagram. Let A, B, and C be the events that a county resident has no electricity, no water, and no telephone service, respectively. Notice that $\bar{A}$ represents the event that a resident has electricity; care needs to be taken not to confuse A, B, and C with their complements. Now, ABC is the event of a county resident not having any of the three utilities, which is 10% of the residents in the county. $A \cup B \cup C$ is the event of a county resident having lost at least one of the utilities so $\overline{A \cup B \cup C}$ is the event of a county resident having all three utilities after the hurricane, which is 20% of the residents in the county. We are also given that 12% were without electricity and water but still had a working telephone ($AB\bar{C}$) and that 4% were without electricity and a working telephone but still had water ($A\bar{B}C$). Because we know that 50% are without electricity (A), 12% are without electricity and water (AB), and 4% are without electricity and telephone service (AC), we can use subtraction to determine that 24% lost electricity but still had water and a working telephone ($A\bar{B}\bar{C}$). To complete filling in percentages in the Venn diagram, we must determine ($\bar{A}BC$). The proportion in $B \cup C$ equals 100% minus the percent with all three utilities minus the percent without only electricity ($100 - 20 - 24 = 56\%$); however, the proportions in B and C total $47 + 38 = 85\%$. The difference, $85 - 56 = 29\%$ represents the proportion having neither water nor a working telephone. Subtracting the 10% without all three utilities leaves 19% without water or a working telephone but with electricity. With this information, the proportions in each part of the Venn diagram have been determined.



Once the Venn diagram is correctly filled out, then the questions become easy to answer.

1. This event is $A\bar{B}\bar{C}$, which represents 24% of the county's residents.
2. $(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C)$ is the event that represents a county resident losing one utility but having the other two, and $24 + 6 + 5 = 35\%$ of the county's residents had this experience.
3. The event of a county resident being without two of the utilities but with the third is $(AB\bar{C}) \cup (A\bar{B}C) \cup (\bar{A}BC)$, which constitutes $12 + 4 + 19 = 35\%$ of the county's residents.
4. $A \cup B \cup C$ is the event representing the county resident who has lost at least one utility, and $100 - 20 = 24 + 12 + 10 + 6 + 19 + 5 + 4 = 80\%$ had that experience.

■

Exercises

2.5 An experiment consists of observing the gender of a baby when it is born.

- **a** List the outcomes in the sample space.
- **b** List all possible events.
- **c** If the baby is a girl, which event(s) in (b) has (have) occurred?

2.6 In an experiment, the number of insects on a randomly selected plant is recorded.

- **a** Specify the sample space for this experiment.
- **b** Give the set A representing the event of at most three insects being on the plant.
- **c** Explain why each of the possible outcomes in the sample space is not equally likely.

2.7 Of 30 laptop computers available in a supply room, 12 have a wireless card, 8 have a CD/DVD burner, and 14 have neither. In an experiment, a laptop is randomly selected from the supply room and whether or not it has a wireless card or a CD/DVD burner recorded. Using W to denote the event that the selected laptop has a wireless card and B to denote the event that the selected laptop has a CD/DVD burner, symbolically denote the following events and find the number of laptop computers represented by each.

- **a** The selected laptop has both a wireless card and a CD/DVD burner.
- **b** The selected laptop has either a wireless card or a CD/DVD burner.
- **c** The selected laptop has a CD/DVD burner only.
- **d** The selected laptop has either a wireless card or a CD/DVD burner but not both.

2.8 In an isolated housing area consisting of 50 households, the residents are allowed to have at most one dog and at most one cat per household. Currently 25 households have a dog, 7 have both a cat and a dog, and 18 have neither a cat nor a dog. An experiment consists of randomly selecting a household and observing whether that household has a cat or a dog, or both. Using D to denote the event that the selected household has a dog and C to denote the event that the selected household has a cat, symbolically denote the following events and identify the number of households represented by each.

- **a** The selected household has a dog but no cat.
- **b** The selected household has a cat.
- **c** The selected household has a cat but no dog.
- **d** The selected household has a cat or a dog but not both.

2.9 Four species of fish—black molly, guppy, goldfish, and neon—are available in a fish store. A child has been told that she may choose any two of these fish species for her aquarium. Once she makes the choice, several fish of the two selected species will be purchased.

- **a** Give the sample space for this experiment.
- **b** Let A denote the event that the child's selections include a black molly. How many outcomes are in A?
- **c** Let B, C, and D denote the events that the selections include a guppy, goldfish, and neon, respectively. Symbolically denote the event of selections containing a guppy, but no goldfish or neon, and identify the number of outcomes in the set.
- **d** List the outcomes in $\bar{A}$, $AB \cup CD$, $\overline{AB \cup CD}$, and $\overline{(A \cup C)(B \cup D)}$.

2.10 Five applicants (Jim, Don, Mary, Sue, and Nancy) are available for two identical jobs. A supervisor selects two applicants to fill these jobs.

- **a** Give the sample space associated with this experiment.
- **b** Let A denote the event that at least one male is selected. How many outcomes are in A?
- **c** Let B denote the event that *exactly* one male is selected. How many outcomes are in B?
- **d** Write the event that two males are selected in terms of A and B.
- **e** List the outcomes in $\bar{A}$, AB, $A \cup B$, and $\overline{AB}$.

2.11 Although there is some controversy, tongue rolling (the ability to roll the edges of the tongue into a loop) is generally thought to be a genetic trait. Another genetic trait is earlobe attachment. Some people have the earlobe attached to the side of the head; others have earlobes that hang freely. In a class of 60 students, 13 could not roll their tongue, 17 had attached earlobes, and 10 could roll their tongues and had attached earlobes. A student is randomly selected from the class, and whether or not the student can roll his or her tongue and whether or not his or her earlobe is attached are observed. Let T represent the event that the selected student can roll his or her tongue and E represent the event that the student has attached earlobes. Symbolically denote the following events and identify the number of students represented by each.

- **a** The student can roll his or her tongue.
- **b** The student can neither roll his or her tongue nor has attached earlobes.
- **c** The student has attached earlobes but cannot roll his or her tongue.
- **d** The student can roll his or her tongue or has attached earlobes but not both.

2.12 On a large college campus, 84% of the students report drinking alcohol within the past month, 33% report using some type of tobacco product within the past month, and 86% report using at least one of these two substances during the past month. In an experiment, whether or not a randomly selected student has drunk alcohol or used a tobacco product during the past month is recorded. Using set notation, symbolically denote each of the following events and then determine the proportion of the student body represented by each.

- **a** The selected student has both drunk alcohol and used a tobacco product during the past month.
- **b** The selected student has abstained from both alcohol and tobacco products during the past month.
- **c** The selected student has either consumed alcohol or used a tobacco product during the past month but not both.

2.13 On a large college campus, the students are able to get free copies of the school newspaper, the local newspaper, and a national paper. Eighty-two percent of the students read at least one of the papers, 42% read only the school newspaper, 18% read only the local paper, 6% read only the national paper, and 1% read all three papers. Fifty-two percent read the school newspaper, and 5% read the school and national newspapers. A randomly selected student is asked whether he or she reads the school, local, or national paper. Express each of the following events in set notation and find the percentage of students represented by each.

- **a** The student does not read any paper.
- **b** The student reads the local paper.
- **c** The student reads exactly two papers.
- **d** The student reads at least two papers.

2.14 A study of children 15 to 18 years of age in the United States considered the availability of various types of media in the child's bedroom. It was found that 68% have a television, 92% have a CD/tape player, and 32% have a computer in their bedroom. Only 4% do not have a television, a CD/tape player, or a computer. Nine percent have a computer and a CD/tape player but no television. Forty-four percent have a television and a CD/tape player but no computer, and 1% have a television and a computer but no CD/tape player. A U.S. child, aged 15 to 18, is selected, and whether or not the child has a television, a CD/tape player, or a computer in his or her room is observed. Express each of the following events in set notation and find the percentage of children, ages 15 to 18, represented by each.

- **a** The child has a television, CD/tape player, and computer in his or her bedroom.
- **b** The child has a television but no CD/tape player or computer in his or her bedroom.
- **c** The child has a CD/tape player but no television or computer in his or her bedroom.
- **d** The child has at least two of the media types (television, CD/tape player, computer) in his or her bedroom.

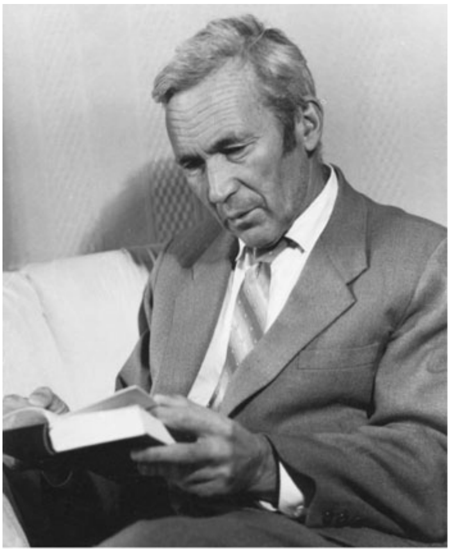
2.15 Use Venn diagrams to verify the distributive laws.

2.16 Use Venn diagrams to verify DeMorgan's second law.

2.3 Definition of Probability

In the last section, we discussed how to establish a sample space and how to list appropriate events for an experiment. The next step is to define a probability for these events. As discussed in Chapter 1, people have been interested in probability since ancient times. Much of this early interest stemmed from games of chance and thus was not condoned by the Church. As a consequence, people considered various probability problems, but a formal mathematical foundation for probability was lacking. That changed when, in 1933, at the age of 30, Russian-born Andrey N. Kolmogorov (Figure 2.7) set forth the three axioms of probability in *Foundations of the Calculus of Probabilities,* which was published in German.

FIGURE **2.7**
Andrey N. Kolmogorov (1903–1987).



Sovfoto

The three axioms proposed by Kolmogorov are intuitive ones. We have already seen that the intuitive idea of probability is related to relative frequency of occurrence. When rolled, a regular die should have an even number of dots on the upper face about $1/2$ of the time and a 3 about $1/6$ of the time. All probabilities should be fractions between 0 and 1, inclusive. One of the integers 1, 2, 3, 4, 5, or 6 must occur every time the die is rolled, so the total probability associated with the sample space must be 1. In repeated rolls of the die, if a 1 occurs $1/6$ of the time and a 2 occurs $1/6$ of the time, then a 1 or 2 must occur $1/6 + 1/6 = 1/3$ of the time. Because relative frequencies for mutually exclusive events can be added, so must the associated probabilities. These considerations lead to the following definition.

DEFINITION 2.3

Suppose that a random experiment has associated with it a sample space S. A **probability** is a numerically valued function that assigns a number $P(A)$ to every event A so that the following axioms hold:

1 $P(A) \geq 0$

2 $P(S) = 1$

3 If $A_1, A_2, \ldots,$ is a sequence of mutually exclusive events (that is, a sequence in which $A_i A_j = \phi$ for any $i \neq j$), then

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

■

Some basic properties of probability follow immediately from these three axioms, and we will look at several of them. The first one is that

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

To see this, let $A_1 = S$ and $A_i = \phi,\ i = 2, 3, 4, \ldots$ Then $S = \bigcup_{i=1}^{\infty} A_i$ and $A_i A_j = \phi$ for any $i \neq j$. From axioms 2 and 3, we have

$$1 = P(S) = P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) = P(S) + \sum_{i=2}^{\infty} P(\phi).$$

Thus,

$$\sum_{i=2}^{\infty} P(\phi) = 0$$

and because, by axiom 1, every probability must be nonnegative,

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

From axiom 3 and the fact that the probability of the null event is 0, it follows that if A and B are mutually exclusive events,

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

To see this, let $A_1 = A$, $A_2 = B$, and $A_i = \phi$, $i = 3, 4, 5, \ldots,$ in axiom 3 above. This result is what we intuitively proposed in the addition of relative frequencies in the die-rolling example discussed earlier.

What happens if the events A and B are not disjoint as illustrated in the last diagram in Figure 2.5? Notice by looking at the Venn diagram that if we add $P(A)$

and $P(B)$, we have added the probability associated with $A \cap B$ twice and would need to subtract this probability to obtain $P(A \cup B)$; that is,

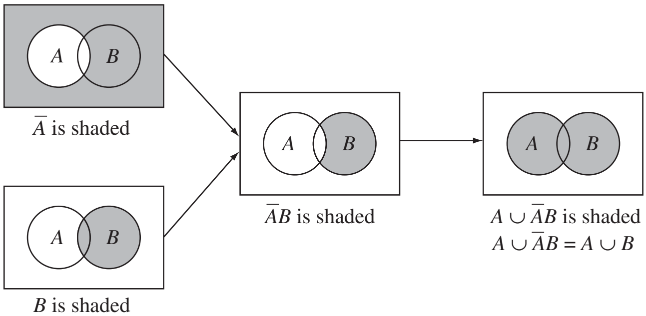
$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

To more formally verify the above, we first write $A \cup B$ as the union of two disjoint sets:

$$A \cup B = A \cup \bar{A}B$$

To verify that this relationship is correct, see the Venn diagrams in Figure 2.8.

FIGURE 2.8
Showing $A \cup B = A \cup \bar{A}B$ using Venn diagrams.



Now, because A and $\bar{A}B$ are disjoint sets, we have

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

We can also write $B = AB \cup \bar{A}B$ where AB and $\bar{A}B$ are disjoint sets. Therefore,

$$P(B) = P(AB \cup \bar{A}B) = P(AB) + P(\bar{A}B),$$

which may be rewritten as

$$P(\bar{A}B) = P(B) - P(AB).$$

By combining the above, we have

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

Another elementary property of probability follows immediately:

$$\text{If } A \subset B, \text{ then } P(A) \leq P(B).$$

Again notice that, because $A \subset B$, $AB = A$, and we can write B as $B = A \cup \bar{A}B$. Because A and $\bar{A}B$ are disjoint sets, we have

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

Now by axiom 1, we know that $P(\bar{A}B) \geq 0$ so we must add something that is at least 0, but may be larger, to $P(A)$ to get $P(B)$; that is, $P(A) \leq P(B)$.

Consider any event A associated with sample space S. By the definition of an event, we know that $A \subset S$ for any event A so $P(A) \leq P(S)$. From axiom 2, we know that $P(S) = 1$. Therefore, $P(A) \leq 1$. This, with axiom 1, gives us that for any event $A, 0 \leq P(A) \leq 1$.

$$0 \leq P(A) \leq 1.$$

The fact that probabilities are always between 0 and 1, inclusive, is an important fact to remember; any reported value outside this range cannot be correct.

Further, we can show that

$$P(\bar{A}) = 1 - P(A).$$

Because A and $\bar{A}$ are disjoint and $A \cup \bar{A} = S$,

$$1 = P(S) = P(A \cup \bar{A}) = P(A) + P(\bar{A}).$$

From axiom 2, $P(S) = 1$ so, using the above, $P(\bar{A}) = 1 - P(A)$.

Suppose we now have n events, $E_1, E_2, \ldots, E_n$, and we want to find the $P(\overset{n}{\underset{i=1}{\cup}} E_i)$. Using the fact that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ and induction, we can prove what is known as the Inclusion-Exclusion Principle, stated in Theorem 2.1.

THEOREM **2.1** **Inclusion-Exclusion Principle:**

$$\begin{aligned} P\left(\overset{n}{\underset{i=1}{\cup}} E_i\right) &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \cdots \\ &\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r}) + \cdots \\ &\quad + (-1)^{n+1} P(E_1 E_2 \cdots E_n) \end{aligned}$$

Proof The proof is left as an exercise. ▪

The definition of probability tells us only the axioms that a probability function must obey; it does not tell us what numbers to assign to specific events. The assignment of probabilities to events is usually based on empirical evidence or on careful thought about the experiment. If a die is balanced, we could roll it a few times to see whether the upper faces all seem equally likely to occur. Or we could simply assume that this result would be obtained and assign a probability of $1/6$ to each of the six outcomes in S; that is $P(E_i) = 1/6,\ i = 1, 2, \ldots, 6$. Once we have done this, the model is complete; by axiom 3, we can now find the probability of any event. For example, for the events defined on page 15 and Figure 2.4,

$$\begin{aligned} P(A) &= P(E_2 + E_4 + E_6) \\ &= P(E_2) + P(E_4) + P(E_6) \\ &= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2} \end{aligned}$$

and

$$\begin{aligned} P(C) &= P(E_5 + E_6) \\ &= P(E_5) + P(E_6) \\ &= \frac{1}{6} + \frac{1}{6} = \frac{1}{3}. \end{aligned}$$

Definition 2.3, in combination with the actual assignment of probabilities to events, provides a probabilistic model for an experiment. If $P(E_i) = 1/6$ is used in the die-rolling experiment, we can assess the suitability of the model by examining how closely the long-run relative frequencies for each outcome match the numbers predicted by the theory underlying the model. If the die is balanced, the model should be quite accurate in telling us what we can expect to happen. If the die is not balanced, the model will fit poorly with the actual data obtained and other probabilities should be substituted for the $P(E_i)$. Throughout the remainder of this book, we shall develop many specific models based on the underlying definition of probability and discuss practical situations in which they work well. None is perfect, but many are adequate for describing real-world phenomena.

EXAMPLE 2.3 A farmer has decided to plant a new variety of corn on some of his land, and he has narrowed his choice to one of three varieties, which are numbered 1, 2, and 3. All three varieties have produced good yields in variety trials. Which corn variety produces the greatest yield depends on the weather. The optimal conditions for each are equally likely to occur, and none does poorly when the weather is not optimal. Being unable to choose, the farmer writes the name of each variety on a piece of paper, mixes the pieces, and blindly selects one. The variety that is selected is purchased and planted. Let E_i denote the event that variety i is selected ($i = 1, 2, 3$), let A denote the event that variety 2 or 3 is selected, and let B denote the event that variety 3 is *not* selected. Find the probabilities of E_i, A, and B.

Solution The sample space is $S = \{1, 2, 3\}$. The events E_1, E_2, and E_3 are the one-element events, that is, $E_1 = \{1\}, E_2 = \{2\}, E_3 = \{3\}$, and $E_1 \cup E_2 \cup E_3 = S$. Thus, if we assign appropriate probabilities to these events, the probability of any other event can easily be found.

Because one number (variety) is picked at random from the three numbers (varieties) available, it is intuitively reasonable to assign a probability of $1/3$ to each E_i:

$$P(E_1) = P(E_2) = P(E_3) = \frac{1}{3}$$

In other words, we find no reason to suspect that one variety has a greater likelihood of being selected than any of the others. Now,

$$A = E_2 \cup E_3$$

and by axiom 3 of Definition 2.3,

$$P(A) = P(E_2 \cup E_3) = P(E_2) + P(E_3) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

Similarly,

$$B = E_1 \cup E_2$$

and therefore,

$$P(B) = P(E_1 \cup E_2) = P(E_1) + P(E_2) = \frac{1}{3} + \frac{1}{3} = \frac{2}{3}.$$

Notice that different probability models could have been selected for the sample space connected with this experiment, but only this model is reasonable under the assumption that the varieties are all equally likely to be selected. The terms *blindly* and *at random* are interpreted as imposing equal probability of observing each of the finite number of points in the sample space. ▪

The examples discussed so far have assigned equal probabilities to the elements of a sample space, but this is not always the case. If you have one quarter and one penny in your pocket and you pull out the first coin you touch, the quarter may have a higher probability of being chosen because of its larger size.

Often the probabilities assigned to events are based on experimental evidence or observational studies that yield relative frequency data on the events of interest. The data provide only approximations to the true probabilities, but these approximations often are quite good and usually are the only information we have on the events of interest.

EXAMPLE **2.4** The National Crime Victimization Survey (NCVS) is an ongoing survey of a national representative sample of residential households. Twice each year, data are obtained from about 49,000 households, consisting of approximately 100,000 persons, on the frequency, characteristics, and consequences of victimization in the United States. Results for 1995 and 2002 are given in Table 2.1.

Answer the following questions.

1 For a randomly selected college student in 1995, what is the probability that the person was a victim of simple assault that year?

2 For a randomly selected college student in 2002, what is the probability that the person was a victim of simple assault that year?

3 For a randomly selected nonstudent in 2002 who was in the 18 to 24 age group, what is the probability that the person was not a victim of a violent crime in 2002?

TABLE **2.1**
Violent victimization rates of college students and nonstudents, by type of crime, 1995–2002. Rates per 1000 persons ages 18–24.

	1995		2002	
	College Students	**Non-Students**	**College Students**	**Non-Students**
Violent Crime	87.7	101.6	40.6	56.1
Rape/Sexual Assault	4.3	4.4	3.3	4.1
Robbery	8.4	12.1	2.9*	6.8
Aggravated Assault	14.5	22.2	9.1	13.2
Simple Assault	60.5	62.8	25.3	32.0
Serious Violent Crime	27.3	38.8	15.3	24.1

*Based on 10 or fewer sample cases.
Source: U.S. Department of Justice. 2005. Bureau of Justice Statistics Special Report: Violent victimization of college students, 1995–2002. NCJ 206836. 7 pages.

4 Notice that in Table 2.1, the category "Violent Crime" has been broken into five categories: Rape/Sexual Assault, Robbery, Aggravated Assault, Simple Assault, and Serious Violent Crime. Why do the rates in the five subcategories not total to the Violent Crime rate?

5 Why was the * added to the 2.9 cases per 1000 persons for college students in 2005?

Solution

1 The estimated rate of simple assault of college students in 1995 is 60.5 cases per 1000 persons. The probability that the randomly selected 1995 college student was a victim of simple assault during that year is about $60.5/1000 = 0.0605$.

2 Using the same reasoning as in part 1, the probability of the randomly selected 2002 college person being a victim of simple assault during that year is about $25.3/1000 = 0.0253$.

3 The probability that the randomly selected nonstudent who was aged 18 to 24 during 2002 was a victim of a violent crime during that year is about $56.1/1000 = 0.0561$. Thus, the probability that the person was not a victim of violent crime is about $1 - 0.0561 = 0.9439$.

4 Notice that if we total the rates for rape/sexual assault, robbery, aggravated assault, simple assault, and serious violent crimes for noncollege persons aged 18 to 24 in 1995, we have $4.4 + 12.1 + 22.2 + 62.8 + 38.8 = 140.3$, which is a total much higher than the 101.6 per 1000 reported for the rate of violent crimes. The reason is that the subcategories are not mutually exclusive; that is, a person might be a victim of both a sexual assault and an aggravated assault during the same year. This would be recorded within each subcategory but would be considered as only one person experiencing a single violent crime in the overall category. A formal approach to properly accounting for these overlaps is discussed in Chapter 3.

5 All of the rates in the table are estimates based on the number of persons interviewed within each category and the numbers of cases reported. *Note 1*: Even though about 100,000 people were interviewed, a much smaller number of people in the 18 to 24 age group were interviewed, and these were broken into two groups, college students and nonstudents. When the number of cases becomes small, the estimate can change quite dramatically with each additional case or

with a reduction of a case; that is, the estimate is not stable. The * in the table is a warning that the estimated rate, and thus the estimated probability, may not be very precise. *Note 2*: When computing probabilities from data, we should keep in mind that these are unlikely to be exact. Even the United States Census in which every "whole person" is to be counted suffers from an undercount because some groups of people are difficult to count. When the sample sizes are large, as in the Census or this example, the observed proportions are "close" to the true probabilities. Measures of how close the estimates are to the probabilities are topics for the study of statistical inference. In this text, we use the estimates for the probabilities.

■

Exercises

2.17 As in Exercise 2.9, four species of fish—black molly, guppy, goldfish, and neon—are available in a fish store:. A child has been told that she may choose any two of these fish species for her aquarium. Then several fish of each of the selected species will be purchased. Suppose that she is equally likely to choose each fish species. Find the probability of each of the following events.

- **a** She does not choose a guppy.
- **b** She chooses a guppy and a neon.
- **c** She has either a black molly or a neon but not both.

2.18 As in Exercise 2.10, five applicants (Jim, Don, Mary, Sue, and Nancy) are available for two identical jobs. Because all applicants are equally qualified, a supervisor randomly selects two applicants to fill these jobs. Find the probability of each of the following events.

- **a** Both males are selected.
- **b** At least one male is selected.
- **c** At least one female is selected.

2.19 As discussed in Exercise 2.11, in a class of 60 students, 13 could not roll their tongue, 17 had attached earlobes, and 10 could roll their tongues and had attached earlobes. A student is randomly selected from this class. Find the probability that the selected student:

- **a** can roll his or her tongue.
- **b** could not roll his or her tongue and had attached earlobes.
- **c** could either roll his or her tongue or had attached earlobes but not both.

2.20 A manufacturing company has two retail outlets. It is known that 30% of all potential customers buy products from outlet 1 alone, 40% buy from outlet 2 alone, and 10% buy from both 1 and 2. Let A denote the event that a potential customer, randomly chosen, buys from outlet 1, and let B denote the event that the customer buys from outlet 2. Suppose a potential customer is chosen at random. For each of the following events, represent the event symbolically and then find its probability.

- **a** The customer buys from outlet 1.
- **b** The customer does not buy from outlet 2.
- **c** The customer does not buy from outlet 1 or does not buy from outlet 2.
- **d** The customer does not buy from outlet 1 and does not buy from outlet 2.

2.21 Among donors at a blood center, 1 in 2 gave type O^+ blood, 1 in 11 gave O^-, 1 in 4 gave A^+, and 1 in 20 gave A^-. What is the probability that the first person who shows up tomorrow to donate blood has the following blood type?

a Type O^+

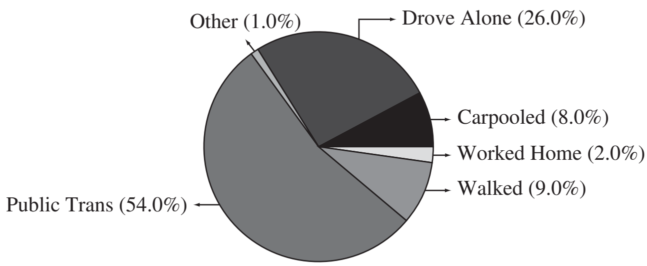
b Type O

c Type A

d Either type A^- or O^-

e Neither type A^+ nor O^+

2.22 Information on modes of transportation for workers 16 years of age or older living in New York is shown in the accompanying chart. If a New York worker is selected at random, find the probability that (s)he was transported by the following means:



Source: U.S. Census Bureau.

a By driving alone

b By public transportation

c By commuting (that is, the person did not work at home)

d By modes that do not involve a motorized vehicle (assume "Other" does *not* include a motorized vehicle)

2.23 Hydraulic assemblies for landing gear produced by an aircraft rework facility are inspected for defects. History shows that 10% have defects in the shafts, 8% have defects in the bushings, and 2% have defects in both the shafts and the bushings. If one such assembly is randomly chosen, find the probability that it has the following characteristics:

a Only a bushing defect

b A shaft or bushing defect

c Only one of the two types of defects

d No defects in either shafts or bushings

2.24 Of the 1.6 million computer systems designers in the United Sates (as of 2005), 26.6% are female. With respect to ethnicity, 73.7% are White, 6.1% are Black or African American, 15.9% are Asian, and 4.3% Hispanic or Latino.

Of the 0.251 million working in veterinarian services, 76.1% are female, 92.5% are White, 1.8% are Black or African American, 0.8% are Asian, and 4.9% are Hispanic or Latino.

a Construct a meaningful table to compare the numbers of computer systems designers and veterinarian services workers by sex.

b Construct a meaningful table to compare the numbers of computer systems designers and veterinarian services workers by ethnic group.

2.25 Consider the information given in Exercise 2.24. Suppose that 1000 computer systems designers are interviewed in a national poll on computer science education.

a How many would you expect to be male?

b How many would you expect to be Asian?

2.26 Consider the information given in Exercise 2.24. In a poll of randomly selected veterinarian services workers, it is desired to get responses from approximately 50 Blacks or African Americans. How many veterinarian services workers should be polled?

2.27 For events A and B, using the axioms of probability, prove that $P(AB) \leq P(A)$.

2.28 Using the axioms of probability, prove Bonferroni's inequality: For events A and B, $P(AB) \geq P(A) + P(B) - 1$.

2.29 For events A and B, using the axioms of probability, prove that $P(A\bar{B}) = P(A) - P(AB)$.

2.30 Using the axioms of probability, show that for events A, B, and C,

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

2.31 Suppose that an experiment is performed n times. For any event A of the experiment, let $n(A)$ denote the number of times that event A occurs. The relative frequency definition of probability would propose that $P(A) = n(A)/n$. Prove that this definition satisfies the three axioms of probability.

2.32 For the n events, $E_1, E_2, \ldots, E_n$, use induction to prove that

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} E_i\right) &= \sum_{i=1}^{n} P(E_i) - \sum_{i_1 < i_2} P(E_{i_1} E_{i_2}) + \cdots \\ &\quad + (-1)^{r+1} \sum_{i_1 < i_2 < \cdots < i_r} P(E_{i_1} E_{i_2} \cdots E_{i_r}) + \cdots + (-1)^{n+1} P(E_1 E_2 \cdots E_n). \end{aligned}$$

2.33 Verify Bonferroni's inequality given in Exercise 2.28 using Venn diagrams.

2.34 Verify the result in Exercise 2.30 using Venn diagrams.

2.4 Counting Rules Useful in Probability

Let us now look at the die-rolling experiment from a slightly different perspective. Because the six outcomes should be equally likely for a balanced die, the probability of A, an even number, is

$$P(A) = \frac{3}{6} = \frac{\text{Number of outcomes favorable to } A}{\text{Total number of equally likely outcomes}}.$$

This *relative frequency definition* of probability will work for any experiment that results in a finite sample space with *equally likely* outcomes. Thus, it is important to be able to count the number of possible outcomes for an experiment. Unfortunately, the number of outcomes for an experiment can easily become quite large, and counting them is difficult unless one knows a few counting rules. Four such rules are presented as theorems in this section, and a fifth rule is discussed in the next section.

Suppose that a quality control inspector examines two manufactured items selected from a production line. Item 1 can be defective or nondefective, as can item 2. How many outcomes are possible for this experiment? In this case it is easy to list them. Using D_i to denote that the ith item is defective and N_i to denote that the ith item is not defective, the possible outcomes are

$$D_1D_2, \quad D_1N_2, \quad N_1D_2, \quad N_1N_2.$$

FIGURE 2.9
Possible outcomes for inspecting two items (D_i denotes that the ith item is defective; N_i denotes that the ith item is not defective).

		Second Item	
		D_2	N_2
First Item	D_1	D_1D_2	D_1N_2
	N_1	N_1D_2	N_1N_2

These four outcomes can be placed in a two-way table, as in Figure 2.9, to help clarify that the four outcomes arise from the fact that the first item has two possible outcomes and the second item has two possible outcomes—and hence, the experiment of looking at both items has $2 \times 2 = 4$ outcomes. This is an example of the Fundamental Principle of Counting, given as Theorem 2.2.

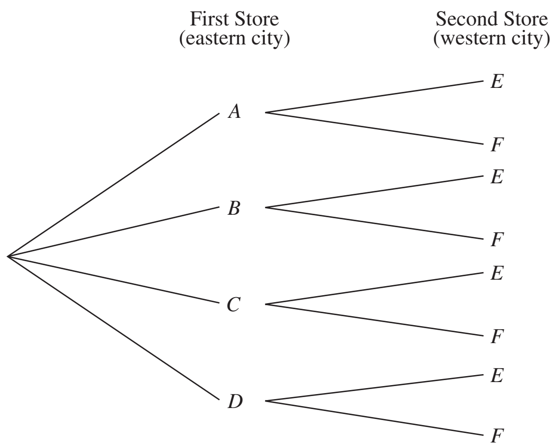
THEOREM 2.2

Fundamental Principle of Counting: If the first task of an experiment can result in n_1 possible outcomes and, for each such outcome, the second task can result in n_2 possible outcomes, then there are n_1n_2 possible outcomes for the two tasks together. ▪

The **Fundamental Principle of Counting** extends to more tasks in a sequence. If, for example, three items are inspected and each of these could be defective or not defective, there would be $2 \times 2 \times 2 = 8$ possible outcomes.

Tree diagrams are also helpful in verifying the Fundamental Principle of Counting and for listing outcomes of experiments. Suppose that a national retail chain is deciding where to build two new stores—one in an eastern city and one in a western city. Four potential locations in the eastern city and two in the western city have been identified. Thus there are $n_1n_2 = 4(2) = 8$ possibilities for locating the two stores. Figure 2.10 lists these possibilities on a tree diagram.

FIGURE 2.10
Possible locations for two stores (A, B, C, D denote eastern cities; E, F denote western cities).



The tree diagram in Figure 2.10 has features common to all tree diagrams. The beginning of the diagram is a point, which is generally on the extreme left. The branches emanating from this point represent the possible outcomes for the *first* task without any consideration to the outcomes of the second task. From the end of each of the first task branches, a second set of branches represents the possible outcomes for the second task given the outcome of the first one. Because there were two tasks (selecting the eastern city's store location and the western city's store location), there are two levels of branching. Also notice that each combination of branches represents a possible outcome in the sample space. For example, the top branch in both tasks represents the outcome of store location A in the eastern city and store location E in the western city.

The Fundamental Principle of Counting (Theorem 2.2) helps only in identifying the number of elements in a sample space for an experiment. We must still assign probabilities to these elements to complete our probabilistic model. This is done in Example 2.5 for the site selection problem.

EXAMPLE 2.5 In connection with the national retail chain that plans to build two new stores, the eight possible combinations of locations are as shown in Figure 2.10. If all eight choices are equally likely (that is, if one of the pairs of cities is selected at random), find the probability that City E is selected.

Solution City E can get selected in four different ways, because four possible locations in the eastern city may be paired with it. Let A, B, C, D, and E represent the events that locations A, B, C, D, and E, respectively, are chosen. Then

$$E = (AE) \cup (BE) \cup (CE) \cup (DE).$$

Each of the eight outcomes has a probability of $1/8$, because they are assumed to be equally likely. Because the events (AE), (BE), (CE), and (DE) are mutually exclusive and each consists of one outcome,

$$\begin{aligned} P(E) &= P(AE) + P(BE) + P(CE) + P(DE) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8} = \frac{1}{2}. \end{aligned}$$

■

EXAMPLE 2.6 Five cans of paint (numbered 1 through 5) were delivered to a professional painter. Unknown to her, some of the cans (1 and 2) are satin finish and the remaining cans (3, 4, and 5) are glossy finish. Suppose she selects two cans at random for a particular job. Let A denote the event that the painter selects the two cans of satin-finish paint, and let B denote the event that the two cans have different finishes (one of satin and one of glossy). Find $P(A)$ and $P(B)$.

Solution We can see from the tree diagram in Figure 2.11 that this experiment has 20 possible outcomes. Because there are five ways to choose the first can and four ways to choose the second can (because there are only four cans left after the first is selected), the Fundamental Principle of Counting also gives that there are $5(4) = 20$ possible outcomes. These 20 one-element events may be represented in the form $\{(1, 2)\}$, $\{(1, 3)\}$, and so forth, where the order of the pairs indicates the order of selection. Because the cans are randomly selected, each of the 20 outcomes has a probability of $1/20$. Thus,

$$P(A) = P(\{(1,2)\} \cup \{(2,1)\}) = \frac{2}{20} = 0.1$$

because the probability of the union of disjoint events is equal to the sum of the probabilities of the events in the union. Similarly,

$$\begin{aligned} P(B) &= P(\{(1,3)\} \cup \{(1,4)\} \cup \{(1,5)\} \cup \{(2,3)\} \cup \{(2,4)\} \cup \{(2,5)\} \cup \\ &\qquad \{(3,1)\} \cup \{(3,2)\} \cup \{(4,1)\} \cup \{(4,2)\} \cup \{(5,1)\} \cup \{(5,2)\}) \\ &= \frac{12}{20} = 0.6. \end{aligned}$$

The Fundamental Principle of Counting is often used to develop other counting rules, and we will consider four of these. Some companies allow a new customer to choose the last four digits of a telephone number. Because each of the 10 digits could be chosen for each position, there are $10 \times 10 \times 10 \times 10 = 10^4 = 10{,}000$ possible choices.

FIGURE **2.11** Outcomes for experiment in Example 2.6.

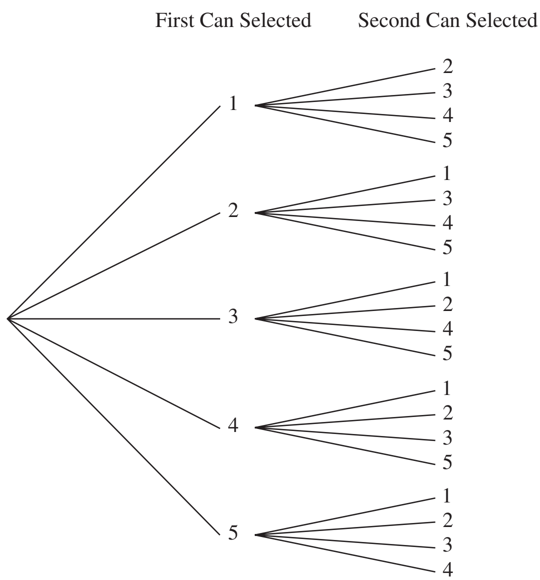


FIGURE **2.12**
Counting the number of ways to select r items from n.

	Order Is Important	Order Is Not Important
With Replacement	n^r	
Without Replacement		

It is important to note a couple of things about the telephone example. First, *order* is important. It is not enough to know which digits occur in the number, the digits must also be dialed in the correct order. Second, the same digit can occur more than once; that is, the four digits are chosen *with replacement*. Whether or not order is important and whether selection is with or without replacement affects the manner in which we count the possible outcomes. We can imagine the 2×2 box in Figure 2.12. The counting technique we use for each square differs.

The Fundamental Principle of Counting may be used to fill in some of the squares. When selecting the last four digits of the telephone number, there were 10^4 possible outcomes. In general, if we have n objects, we select r of these with replacement, and order is important, then there are $n \times n \times n \times \cdots \times n = n^r$ possible outcomes. This corresponds to the upper left square in Figure 2.12.

Suppose that the customer decides that she does not want to use any digit more than once. Of course, order is still important. The number of possible choices for the last four digits of the telephone number is $10 \times 9 \times 8 \times 7 = 5040$. As another illustration, suppose that, from among three pilots, a crew of two is to be selected to form a pilot-copilot team. To count the number of ways that this can be done, observe that the pilot's seat can be filled in three ways. Once that seat is filled the copilot's seat can be filled in two ways. Thus, there are $3 \times 2 = 6$ ways of forming the team. Here the selection is without replacement because the same person cannot be both pilot and copilot. Order is important because a different crew results if the pilot and copilot exchange seats. Each of these is an example of a *permutation* and corresponds to the lower left square in Figure 2.13. The general use of permutations is given in Theorem 2.3.

FIGURE **2.13**
Counting the number of ways to select r items from n when order is important.

	Order Is Important	Order Is Not Important
With Replacement	n^r	
Without Replacement	$P_r^n = \dfrac{n!}{(n-r)!}$	

THEOREM **2.3**

Permutations. The number of ordered arrangements or permutations P_r^n of r objects selected from n distinct objects ($r \le n$) is given by

$$P_r^n = n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!}$$

Proof The basic idea of a permutation can be thought of as a process of filling r slots in a line, with one object in each slot, by drawing these objects one at a time from a pool of n distinct objects. The first slot can be filled in n ways, but the second can only be filled in $(n - 1)$ ways after the first is filled. Thus, by the Fundamental Principle of Counting, the first two slots can be filled in $n(n - 1)$ ways. Extending this reasoning to r slots, the number of ways of filling all r slots is

$$n(n-1)\cdots(n-r+1) = \frac{n!}{(n-r)!} = P_r^n.$$

Hence the theorem is proved. ▪

Example 2.7 illustrates the use of Theorem 2.3. ▪

EXAMPLE 2.7 A small company has 12 account managers. Three potential customers have been identified and each customer has quite different needs. The company's director decides to send an account manager to visit each of the potential customers and considers the customers' needs in making his selection. How many ways are there for him to assign three different account managers to make the contacts?

Solution Because three different account managers are to make the contacts, the choices are made without replacement. The order in which the account managers are assigned to the customers matters because the customers' needs differ, and this is considered in making the assignments. Thus, the number of ways to assign 3 of the 12 account managers to the customers is the permutations of 3 objects selected from 12 objects, that is,

$$P_3^{12} = \frac{12!}{9!} = 12 \times 11 \times 10 = 1320.$$

The director has 1320 possible assignments! We note here that $0! = 1$ by definition. ▪

Thus far we have only discussed the cases where order is important, but sometimes order is not important. Suppose we are playing bridge and are dealt 13 cards. The order in which the cards are dealt does not affect the final hand; that is, it does not matter whether we get the ace of spades in the first or last or some other round of the deal as long as we get the ace of spades. If order is not important, then we still must consider whether the selection is made with or without replacement. In the case of

FIGURE **2.14**
Counting the number of ways to select r items from n.

	Order Is Important	Order Is Not Important
With Replacement	n^r	
Without Replacement	$P_r^n = \frac{n!}{(n-r)!}$	$C_r^n = \binom{n}{r}$

being dealt 13 cards for bridge, selection is made without replacement. Thus we are interested only in the number of *combinations*, the number of subsets of a certain size, which can be selected from the set of n objects. The use of combinations corresponds to the lower right square in Figure 2.14 and is stated generally in Theorem 2.4.

THEOREM **2.4**

Combinations. The number of distinct subsets or combinations of size r that can be selected from n distinct objects ($r \leq n$) is given by

$$C_r^n = \binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

Proof

The number of ordered subsets of size r, selected from n distinct objects, is given by P_r^n. The number of unordered subsets of size r is denoted by $\binom{n}{r}$. Because any particular set of r objects can be ordered among themselves in $P_r^n = r!$ ways, it follows that

$$\begin{aligned}\binom{n}{r} r! &= P_r^n \\ \binom{n}{r} &= \frac{1}{r!} P_r^n \\ &= \frac{n!}{r!\,(n-r)!} \\ &= C_r^n.\end{aligned}$$

■

EXAMPLE **2.8**

Most states conduct lotteries as a means of raising revenue. In Florida's lottery, a player selects six numbers from 1 to 53. For each drawing, balls numbered from 1 to 53 are placed in a hopper. Six balls are drawn from the hopper at random and without replacement. To win the jackpot, all six of the player's numbers must match those drawn in any order. How many winning numbers are possible?

Solution Balls are drawn at random and without replacement. Because it is enough that the numbers match in any order, the order in which the balls are drawn does not matter. Thus, the number of possible winning numbers is the combination of 6 objects selected from 53, that is,

$$\binom{53}{6} = \frac{53!}{6!\,(53-6)!} = \frac{53 \times 52 \times 51 \times 50 \times 49 \times 48}{6 \times 5 \times 4 \times 3 \times 2 \times 1} = 22,957,480.$$

Because great care is taken to be sure that each of the 22,957,480 possible winning numbers are equally likely, the probability that a player who buys one ticket wins is $\frac{1}{22,957,480}$. ▪

EXAMPLE **2.9** Twenty-six states participate in the Powerball lottery. In this lottery, a player selects five numbers between 1 and 53 and a Powerball number between 1 and 42. For each drawing, five balls are drawn at random and without replacement from a hopper with 53 white balls numbered 1 to 53. A sixth ball is drawn from a second hopper with 42 red balls numbered 1 to 42. To win the jackpot, the five numbers selected by the player must match those of the five white balls drawn, and the player's Powerball number must match the number on the red ball drawn from the hopper. How many possible winning numbers are there? Is there a greater probability of winning Florida's lottery or the Powerball if one buys a single ticket?

Solution Similar to Example 2.8, the five white balls are drawn at random without replacement, and it is enough to match the numbers in any order. The number of ways to choose the 5 numbers from the 53 is the combinations of 5 objects chosen from 53:

$$\binom{53}{5} = \frac{53!}{5!\,(53-5)!} = \frac{53 \times 52 \times 51 \times 50 \times 49}{5 \times 4 \times 3 \times 2 \times 1} = 2,869,685$$

There are 42 ways to select a Powerball. We can think correctly of this as the combinations of selecting 1 object from 42:

$$\binom{42}{1} = \frac{42!}{1!\,(42-1)!} = 42$$

Because a player must match the numbers on the five white balls and on the red Powerball, the number of possible winning numbers is, by the Fundamental Principle of Counting, the product of the number of ways to select the five white and the number of ways to select the one red:

$$\binom{53}{5}\binom{42}{1} = 2,869,685 \times 42 = 120,536,770$$

The probability of winning the Powerball lottery is $\dfrac{1}{120{,}536{,}770}$, which is less than the probability of winning Florida's lottery. ■

EXAMPLE 2.10 A department in a company has 12 members: 8 males and 4 females. To gain greater insight into the employees' views of various benefits, the human resources office plans to form a focus group from members of this department. Five departmental members will be selected at random from the department's members. What is the probability that the focus group will only have males? What is the probability that the focus group will have two males and three females?

Solution The selection of focus group members will be without replacement. (It makes no sense to have the same person take two positions in the group.) The order in which focus group members is selected is not important; it is the final set of five members that matters. Thus, the number of ways to select the five focus group members from the 12 departmental members is the combinations of 5 objects selected from 12:

$$\binom{12}{5} = \frac{12!}{5!\,(12-5)!} = \frac{12 \times 11 \times 10 \times 9 \times 8}{5 \times 4 \times 3 \times 2 \times 1} = 792$$

The number of ways to select five males from the eight men in the department is $\binom{8}{5} = \frac{8!}{5!\,(8-5)!} = 56$, and the number of ways to choose zero females from the four in the department is $\binom{4}{0} = \frac{4!}{0!\,(4-0)!} = 1$. Then, by the Fundamental Principle of Counting, the number of ways to select five males and no females from the department's members is $\binom{8}{5}\binom{4}{0} = 56 \times 1 = 56$. The probability of having five males and no females in the focus group is $\frac{\binom{8}{5}\binom{4}{0}}{\binom{12}{5}} = \frac{56}{792} = \frac{7}{99}$. Using similar reasoning, the probability of the focus group having two males and three females is $\frac{\binom{8}{2}\binom{4}{3}}{\binom{12}{5}} = \frac{28 \times 4}{792} = \frac{112}{792} = \frac{14}{99}$. The probability of two males and three females is twice that of five males and no females. ■

We have one more case to consider in Figure 2.14, but let us first consider the following. Suppose a die is rolled 60 times and the number of dots on the upper face is recorded after each roll. In how many ways could we get exactly 10 each of 1, 2, 3, 4, 5, and 6? If we had 60 unique objects and we were interested in how many ways to order these objects, then there would be 60! possible sequences. However, the 10 1s are not distinguishable. That is, we can exchange any two outcomes that have a single dot on the upper face and are recorded as 1s, and the sequence is the same. Similarly, the 2s are not distinguishable from one another, nor are the 3s, the 4s, the 5s, or the 6s. This must be considered in our counting process. Theorem 2.5 discusses how to properly count in these situations.

THEOREM **2.5**

Partitions. The number of ways of partitioning n distinct objects into k groups containing $n_1, n_2, \ldots, n_k$ objects, respectively, is

$$\frac{n!}{n_1!\,n_2!\cdots n_k!}$$

where

$$\sum_{i=1}^{k} n_i = n.$$

Proof

The partitioning of n objects into k groups can be done by first selecting a subset of size n_1 from the n objects, then selecting a subset of size n_2 from the $n - n_1$ objects that remain, and so on until all groups are filled. The number of ways of doing this is

$$\begin{aligned}\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-\cdots-n_{k-1}}{n_k} &= \left(\frac{n!}{n_1!\,(n-n_1)!}\right)\left(\frac{(n-n_1)!}{n_2!\,(n-n_1-n_2)!}\right)\cdots\left(\frac{(n-n_1-\cdots-n_{k-1})!}{n_k!\,0!}\right)\\ &= \frac{n!}{n_1!\,n_2!\cdots n_k!}.\end{aligned}$$

■

EXAMPLE **2.11**

Suppose 10 employees are to be divided among three job assignments, with 3 employees going to job I, 4 to job II, and 3 to job III. In how many ways can the job assignments be made?

Solution

This problem involves partitioning the $n = 10$ employees into groups of size $n_1 = 3$, $n_2 = 4$, and $n_3 = 3$. This can be accomplished in

$$\frac{n!}{n_1!\,n_2!\,n_3!} = \frac{10!}{3!\,4!\,3!} = \frac{10 \times 9 \times 8 \times 7 \times 6 \times 5}{3 \times 2 \times 1 \times 3 \times 2 \times 1} = 4200$$

ways. What happened to the 4! in the denominator? It was used to cancel out $4 \times 3 \times 2 \times 1$ associated with 10! in the numerator. ■

EXAMPLE **2.12** In the setting of Example 2.11, suppose that three employees of a certain ethnic group all get assigned to job I. Assuming that they are the only employees among the 10 under consideration who belong to this ethnic group, what is the probability of this happening under a random assignment of employees to jobs?

Solution We have seen in Example 2.11 that there are 4200 ways of assigning the 10 workers to the three jobs. The event of interest assigns three specified employees to job I. This can be done in only one way. It remains for us to determine how many ways the other seven employees can be assigned to jobs II and III, which is

$$\frac{7!}{4!\,3!} = \frac{7 \times 6 \times 5}{3 \times 2 \times 1} = 35.$$

Thus, the chance of randomly assigning three specified workers to job I is

$$\frac{35}{4200} = \frac{1}{120}$$

which is very small, indeed! ▪

Determining whether or not order is important is critical to counting. It is not enough to determine that order does not matter for the particular application of interest. The random mechanism or sampling process that generates the outcome must be considered. To illustrate, suppose a fair coin is flipped three times and the number of heads observed. Is order important or not? If order is important, eight outcomes (with a head denoted by "H" and a tail by "T") are possible (TTT, TTH, THT, HTT, THH, HTH, HHT, HHH) and it is clear that the probability of each outcome is $1/8$. If order is not important, the outcome of one head could be denoted by HTT and the outcome of two heads by HHT; we would not be interested in which flip resulted in a head, only in how many heads resulted. Thus, if order is not important, four outcomes are possible: TTT, HTT, HHT, and HHH. In Figure 2.15, the outcomes in the ordered and unordered sample spaces as well as the probabilities are shown.

FIGURE **2.15** Unordered and ordered sample spaces from flipping a coin three times.

Unordered	TTT	HTT	HHT	HHH
Ordered	TTT	HTT, THT, TTH	HHT, HTH, THH	HHH
Probabilities	1/8	3/8	3/8	1/8

Notice that even though order is not important, we have attached the probability of $3/8$ to the event of one head for both the ordered and unordered outcomes. Why is the probability of one head not $1/4$ for the unordered case? Although interest lies in how many heads were observed, we cannot ignore the random process by which those heads were obtained. Suppose Person A observes the coin being flipped, but Person B does not. Person A writes down whether a head or a tail occurred on each flip. Because the order is recorded, Person A would assign probabilities as in Figure 2.15. Person B should assign the same probabilities even though he did not have the opportunity to observe the process. If instead the unordered outcomes were listed on four separate pieces of paper, mixed thoroughly and one was drawn at random, the probability of $1/4$ would be assigned to each of the unordered outcomes in Figure 2.15. Obviously, this is a very different process from flipping a coin three times. A key to determining whether or not order is important is an understanding of the random process giving rise to the outcomes.

Our ability to compute probabilities is greatly enhanced by our ability to combine the axioms of probability and the resulting relationships with one or more of the counting rules within a single problem. This is illustrated by the following examples.

EXAMPLE 2.13 The Birthday Problem. Suppose n people are in a room.

1 What is the probability that no two of them have the same birthday?

2 How many people must be in the room for the probability that at least two of the n people have the same birthday to be greater than $1/2$?

Solution

1 Each person celebrates his or her birthday on one of the 365 days in a year. (We are ignoring the possibility of someone's birthday being on February 29.) Because n people are in the room and each has 365 possibilities for a birthday, there are 365^n possible outcomes. The number of possible outcomes for which no 2 have the same birthday is $365(364)(363)\ldots(365-n+1)$. Assuming that each outcome is equally likely, the probability that no 2 people in the room have the same birthday is $365(364)(363)\ldots(365-n+1)/365^n$. *Note*: The probability is 0 if $n = 366$. This is intuitively correct because once we have more than 365 people at least 2 must share the same birthday. Thus, the formula we have provided for the probability is appropriate for $n \leq 366$.

2 If we want the probability that at least two of the n people have the same birthday to be greater than $1/2$, then we want the probability that no two have the same birthday to be less than $1/2$. It is interesting to find that an n as small as 23 results in a probability of no two having the same birthday that is 0.492, which is less than $1/2$. The probability that at least two share the same birthday grows rapidly with n. When $n = 50$, the probability of at least two sharing the same birthday exceeds 0.97, and it exceeds 0.999 when $n = 80$.

■

EXAMPLE **2.14** A poker hand consists of five cards. If all of the cards are from the same suit but are not in consecutive order, we say that the hand is a flush. For instance, if we have five clubs that are not in consecutive order (such as 2, 4, 5, 10, J), then we have a flush. What is the probability of a flush but not a straight flush?

Solution We begin by determining how many possible poker hands there are. A deck of cards has 52 cards, and we select 5 to form a poker hand so there are $\binom{52}{5}$ possible poker hands. We assume that each of these is equally likely. There are $\binom{4}{1}$ ways to choose one of the four suits from which we will get the flush. There are $\binom{13}{5}$ ways to choose 5 cards from the selected suit. However, for some of these sets of five cards, the cards are in consecutive order; such a hand is called a straight flush. How many straight flushes are possible for each suit? Think about lining up all the cards of a particular suit in order with the ace appearing on either end because it can represent the low or the high card: ace, 2, 3, 4, 5, 6, 7, 8, 9, 10, jack, queen, king, ace. Now, the lowest five values represent a straight (ace, 2, 3, 4, 5). The straight could begin with a 2 (2, 3, 4, 5, 6). If we continue, we see that the largest card that could begin a suit is a 10 (10, jack, queen, king, ace). Thus, there are 10 possible starting points, and consequently 10 possible straights, within each suit. Now, there are $\binom{13}{5} - 10$ possible flushes within each suit and $\binom{4}{1}\left(\binom{13}{5} - 10\right)$ possible flushes. The probability of a flush is the number of possible flushes divided by the total number of possible hands of five cards or

$$\frac{\binom{4}{1}\left(\binom{13}{5} - 10\right)}{\binom{52}{5}}.$$

■

EXAMPLE **2.15** Pierre-Remond Montmort considered several matching problems such as this one. An absent-minded secretary prepared five letters and envelopes to send to five different people. Then he randomly placed letters in the envelopes. A match occurs if the letter and its envelope are addressed to the same person. What is the probability of the following events?

1 All five letters and envelopes match.

2 At least one of the five letters and envelopes match.

3 Four of the letters and envelopes match.

Solution Suppose the envelopes are lined up on the secretary's desk. The first letter can be placed in any one of the five envelopes, the second letter can be placed in any one of the remaining four envelopes, and so on. Thus, the total number of ways to assign the letters to the envelopes is $5! = 60$.

1 The only way for all five letters and envelopes to match is for each letter to be placed in its envelope. The probability of this occurring is $1/60$.

2 Let A denote the event that at least one letter is placed in the correct envelope. Let E_i denote the event that letter i is placed in the correct envelope for $i = 1, 2, 3, 4, 5$. At least one letter matches occurs if letter 1 is placed in the correct envelope or letter 2 is placed in the correct envelope, . . . , or letter 5 is placed in the correct envelope; that is, $A = \bigcup_{i=1}^{5} E_i$. Now, by the Inclusion-Exclusion Principle,

$$P(A) = P\left(\bigcup_{i=1}^{5} E_i\right) = \sum_{i=1}^{5} P(E_i) - \sum_{i_1<i_2} P(E_{i_1}E_{i_2}) + \cdots$$
$$+ (-1)^{r+1} \sum_{i_1<i_2<\cdots<i_5} P(E_{i_1}E_{i_2}\cdots E_{i_r}) + \cdots$$
$$+ (-1)^6 P(E_1E_2\cdots E_5).$$

The event that exactly r letters are placed in the correct envelopes, denoted by $E_{i_1}E_{i_2}\cdots E_{i_r}$, can occur in any of $(5-r)!$ ways. To see this, we first realize that there is only one way for each of the r letters to be placed in their respective envelopes. For the remaining $(5-r)$ letters, the first can be placed in any of the remaining $(5-r)$ envelopes, the second can be placed in any of the remaining $(5-r-1)$ envelopes, and so on. Thus, assuming that all $5!$ possible outcomes are equally likely, we have that

$$P\left(E_{i_1}E_{i_2}\cdots E_{i_r}\right) = \frac{(5-r)!}{5!}.$$

Now, notice that there are $\binom{5}{r}$ terms in $\sum_{i_1<i_2<\cdots<i_5} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ so that

$$\sum_{i_1<i_2<\cdots<i_5} P(E_{i_1}E_{i_2}\cdots E_{i_r}) = \binom{5}{r}\frac{(5-r)!}{5!} = \frac{5!\,(5-r)!}{r!\,(5-r)!\,5!} = \frac{1}{r!}.$$

Thus,

$$P(A) = P\left(\bigcup_{i=1}^{5} E_i\right) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \frac{1}{5!} = 0.6417.$$

The probability that at least one letter is placed in the correct envelope is 0.6417.

3 If four letters are in the correct envelopes, the only remaining letter must match the remaining envelope. So, the probability of exactly four matches is 0.

■

Exercises

2.35 A commercial building is designed with two entrances: entrance I and entrance II. Two customers arrive (separately) and enter the building.

a Give the sample space for this observational experiment.

b Assuming that all outcomes in part (a) are equally likely, find the probability that both customers use door I.

c Assuming that all outcomes in part (a) are equally likely, find the probability that the customers use different doors.

2.36 A corporation has two construction contracts that are to be assigned to one or more of four firms bidding for them. (It is possible for one firm to receive both contracts.)

a List the possible outcomes for the assignment of contracts to the firms.

b If all outcomes are equally likely, find the probability that both contracts will go to the same firm.

c Under the assumptions of part (b), find the probability that one specific firm—say, firm A—will get at least one contract.

2.37 A business traveler has five shirts, four pairs of pants, and two jackets in his closet. How many possible outfits, consisting of a shirt, a pair of pants, and a jacket, could he wear?

2.38 On the way to work each day, Jane drops her children off at their elementary school. She has found three ways to get from home to the elementary school and six ways to get from the elementary school to work. How many ways does she have to get from home to work, assuming that she must go by the elementary school?

2.39 An agricultural researcher wishes to investigate the effects of three variables—variety, fertilizer, and pesticide—on corn yield for a particular area in Iowa. The researcher plans to use four varieties, three levels of fertilizer, and two methods of pesticide application. How many combinations of the three variables are to be studied, given that all possible combinations are of interest?

2.40 Eight applicants have applied for two jobs. How many ways can the jobs be filled if the following additional information is known?

a The first person chosen receives a higher salary than the second.

b There are no differences between the jobs.

2.41 A personnel director for a corporation has hired 10 new engineers. If four distinctly different positions are open at a particular plant, in how many ways can the director fill the positions?

2.42 A lab process consists of four steps that can be performed in any sequence. The complexity of each task differs. The lab director decides to conduct a study to determine the order of performing the tasks that will result in the lowest number of errors. If he wants to consider all possible orderings of the tasks, how many sequences will be studied?

2.43 Suppose a class of 8 boys and 7 girls are to attend a theatrical performance, and the teacher obtains 15 tickets (one for each student) in a row. How many ways are there to order the students under the following conditions?

a The children are randomly assigned seats.

b Boys and girls are alternated so that boys sit by girls and girls sit by boys.

c All boys sit together, and all girls sit together.

2.44 In a certain state, license plates have six characters that may include letters and numerals. How many different license plates can be produced if:

a letters and numerals can be repeated?

b each letter and numeral can be used at most once?

c the license plate must have a letter as its first character and each letter or numeral can be used at most once?

2.45 A fleet of 10 taxis is to be divided among three airports, A, B, and C, with 5 going to A, 3 going to B, and 2 to C. In how many ways can this be done?

2.46 A teacher has 24 students in a classroom. For a group project, he decides to divide the students into four groups of equal size. In how many ways can this be done?

2.47 Among six plasma televisions received by a home electronics firm in one day, two are defective. If two plasma televisions are selected for sale, what is the probability that both will be nondefective? (Assume that the two selected for sale are chosen in such a way that every possible sample of size two has the same probability of being selected.)

2.48 A package of eight light bulbs contains three defective bulbs. If two bulbs are randomly selected for use, find the probability that neither one is defective.

2.49 An assembly operation for a computer circuit board consists of four operations, which can be performed in any order.

a In how many ways can the assembly operation be performed?

b One of the operations involves soldering wire to a microchip. If all possible assembly orderings are equally likely, what is the probability that the soldering operation comes first or second?

2.50 A student is given a true-false test with 10 questions. If she gets seven or more correct, she passes. If she is guessing, what is the probability of passing the test?

2.51 A cereal manufacturer is evaluating a new design for the box of its leading product. Ten people are shown the new design and the old design. Each is asked which one he or she likes better. If each person does not have a preference and simply selects one of the two designs at random, what is the probability of each of the following events?

a All 10 select the new design.

b More than half select the new design.

2.52 A teacher randomly assigns the 20 students in her class to study groups, each with 5 students. If there are 10 boys and 10 girls, what is the probability that two of the study groups have all boys and two have all girls?

2.53 Suppose that 15 employees are to be divided among 4 job assignments, with 3 going to job I, 4 going to job II, 6 going to job III, and 2 going to job IV. Suppose that job I is the least favorite among all employees and that three employees of a certain ethnic group all get assigned to job I. Assuming that they are the only employees among the 15 under consideration who belong to this ethnic group, what is the probability of this happening under a random assignment of employees to jobs?

2.54 A firm places three orders for supplies among five different distributors. Each order is randomly assigned to one of the distributors, and a distributor may receive multiple orders. Find the probabilities of the following events.

a All orders go to different distributors.

b All orders go to the same distributor.

c Exactly two of the three orders go to one particular distributor.

2.55 In a storage facility for an automobile parts firm, four boxes are labeled for a particular part. If 36 of these parts are placed in the boxes at random, what is the probability that all parts are in the same box?

2.56 Six employees of a firm are ranked from 1 to 6 in their abilities to fix problems with desktop computers. Three of these employees are randomly selected to service three desktop computers. If all possible choices of three (out of the six) are equally likely, find the probabilities of the following events.

a The employee ranked number 1 is selected.

b The bottom three employees (4, 5, and 6) are selected.

c The highest-ranked employee among those selected has rank 3 or lower.

d The employees ranked 5 and 6 are selected.

2.57 Nine impact wrenches are to be divided evenly among three assembly lines.

a In how many ways can this be done?

b Two of the wrenches are used and seven are new. What is the probability that a particular line (line A) gets both used wrenches?

2.58 Antoine Gambaud, nicknamed “the Chevalier de Méré,” gambled frequently to increase his wealth. He bet on a roll of a die that at least one 6 would appear during a total of four rolls. From past experience, he knew that he was more successful than not with this game of chance. What is the probability that he would win?

2.59 The Chevalier de Méré grew tired of the game described in Exercise 2.58 so he decided to change the game. He bet that he would get a total of 12, or a double 6, on 24 rolls of two dice. Soon he realized that his old approach to the game resulted in more money. What is the probability that he would win in this new game?

2.60 In poker, each player is dealt five cards. What is the probability of obtaining the following on the initial deal?

a Royal flush (ace, king, queen, jack, 10, all in the same suit)

b Straight flush (five cards of the same suit in sequence, such as 3, 4, 5, 6, 7, all spades)

c Four of a kind (four cards of the same rank, such as four queens or four 10s)

d Full house (three cards of one rank and two cards of another rank, such as three jacks and two 4s)

e Straight (five cards of mixed suits in sequence)

f Three of a kind (three cards of the same rank plus two other cards. Note that the two other cards must be of different ranks, or the hand would be a full house.)

g Two pair (two sets of two cards of equal rank and another card). Note that the two pair must be of different ranks, or the hand would be four of a kind. Further the fifth card must be of a different rank from that of either pair, or the hand would be a full house.

h One pair (two cards of equal rank and three other cards. Note that the three other cards must be of different ranks, or the hand would be two pair or a full house.)

2.61 For a certain style of new automobile, the colors blue, white, black, and green are in equal demand. Three successive orders are placed for automobiles of this style. Find the probabilities of the following events.

a One blue, one white, and one green are ordered.

b Two blues are ordered.

c At least one black is ordered.

d Exactly two of the orders are for the same color.

2.62 Four couples go to dinner together. The waiter seats the men randomly on one side of the table and the women randomly on the other side of the table. What is the probability that all four couples are seated across from one another?

2.63 Suppose that N people at a restaurant check their coats. The attendant loses the tickets so at the end of the evening he hands each of the N people a coat at random from among those that were checked. What is the probability that:

a exactly r get the proper coat?

b none get the proper coat?

2.64 For Exercise 2.63, if $N \longrightarrow \infty$, what is the limiting probability that:

a none get the proper coat?

b exactly r get the proper coat?

2.65 Show that $\binom{n}{r} = \binom{n-1}{r-1} + \binom{n-1}{r}$, where $1 \leq r \leq n$.

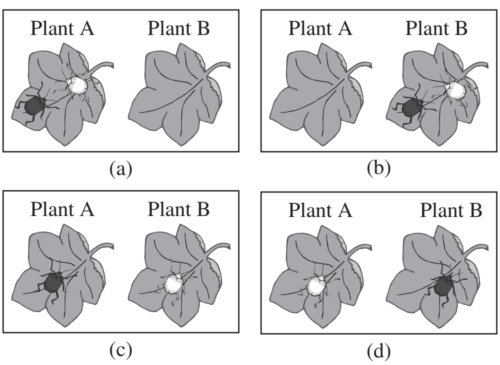
2.66 Capture-recapture studies are common in ecology. One form of the study is conducted as follows. Suppose we have a population of N deer in a study area. Initially n deer from this population are captured, marked so that they can be identified as having been captured, and returned to the population. After the deer are allowed to mix together, m deer are captured from the population and the number k of these deer having marks from the first capture is observed. Assuming that the first and second captures can be considered random selections from the population and that no deer have either entered or left the study area during the sampling period, what is the probability of observing k marked deer in the second sample of m deer?

2.5 More Counting Rules Useful in Probability

In the last section, we considered counting rules that were useful when order is important and, if sampling is without replacement, when order is not important. We still need to consider the case where order is not important and sampling is with replacement to complete the square in Figure 2.14. First, suppose that an entomologist is interested in the number of ways to arrange r insects on n plants. He decides to begin by considering the number of ways that each of two insects could be arranged on two plants for habitation. The plants may be distinguished by their position, but what about the insects? First, suppose the insects are distinguishable. There are $2^2 = 4$ possible arrangements of the insects on the plants as shown in Figure 2.16. Note that in the figure, one insect is black and other is white, so that they may be distinguished.

Now suppose that the insects are not distinguishable. Which insect is on a plant is not important; what matters is the number of insects on each plant in the final arrangement. Thus, we have the possible arrangements in Figure 2.17. Notice that the insects are no longer different colors because they are not distinguishable so that there is only one arrangement of one insect on each plant instead of the two when the insects are distinguishable (see Figure 2.16).

FIGURE **2.16**
Arrangement of two insects on two plants where the insects are distinguishable.



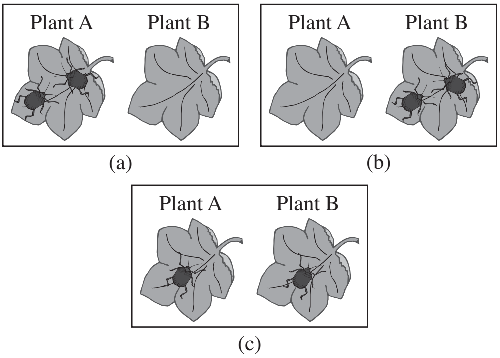


FIGURE **2.17**
Arrangement of two insects on two plants where the insects are indistinguishable.

Listing all of the possibilities of arranging insects on plants soon becomes too labor intensive as the numbers of insects and plants increase. However, let us consider one more case. Suppose the entomologist wants to determine the number of arrangements of three insects on five plants. If the insects are distinguishable, the number of possible arrangements is easily determined to be $5^3 = 125$. The number of arrangements if the insects are not distinguishable is not as evident. In Figure 2.18, we represent the plants by |s and insects by 0s. Notice that we need six |s to represent the five plants. For this particular arrangement, there would be one insect on the first plant, none on the second, two on the third, and none on the fourth and fifth.

FIGURE **2.18**
Arrangement of insects on plants where the plants are delineated by |s and the insects by 0s.

$$| 0 | | 0 0 | | |$$

The number of ways to arrange the three insects on the five plants is the same as the number of distinguishable ways to arrange the |s and 0s. The first and the last symbols must be a | because they form the outer boundaries of the plants. That leaves $5 + 3 - 1 = 7$ |s and 0s that can be rearranged to represent different arrangements of the insects on the plants. Because the |s are not distinguishable from each other and the 0s are not distinguishable from each other, we have

$$\frac{(5+3-1)!}{(5-1)!\,3!} = \binom{5+3-1}{3} = 35$$

distinguishable arrangements of the three insects on the five plants. *Note*: The random mechanism leading to the insects being on the plants has not been considered here. We are only interested in the number of spatial patterns of the insects on the plants, given that the insects are indistinguishable. This example is generalized in Theorem 2.5.

THEOREM 2.6 The number of ways of making r selections from n objects when selection is made with replacement and order is not important is

$$\binom{n+r-1}{r}.$$

Proof As in Figure 2.17, we delineate the n objects using $(n+1)|$s, and each selection is denoted by a 0 between two $|$s. Notice that there must be $(n-1)$ dividers to form the n bins. The number of ways to arrange the $(n-1)$ dividers and the r 0s, where the dividers are indistinguishable and the 0s are indistinguishable, is

$$\frac{(n+r-1)!}{(n-1)!\,r!} = \binom{n+r-1}{r}.$$

We can now complete the table in Figure 2.19. ■

FIGURE 2.19 Counting the number of ways to select r items from n in four different circumstances.

	Order Is Important	Order Is Not Important
With Replacement	n^r	$\binom{n+r-1}{r}$
Without Replacement	$P_r^n = \frac{n!}{(n-r)!}$	$C_r^n = \frac{n!}{(n-r)!\,r!} = \binom{n}{r}$

EXAMPLE 2.16 Two lotteries were described in Examples 2.8 and 2.9. In another lottery, a player selects six numbers between 1 and 44. The same number may be chosen more than once. For each drawing, 6 balls are drawn at random and with replacement from a hopper with 44 white balls numbered 1 to 44. Sufficient time is allowed between selections of a ball for the previously selected ball to be mixed with the others. To win the jackpot, all six of the player's numbers must match those drawn in any order. How many winning numbers are possible?

Solution Because the chosen ball is returned to the hopper after each draw, sampling is with replacement. To win, it is only necessary that a player must match the six numbers, but order is not important. Thus, the number of possible winning numbers is

$$\binom{44+6-1}{6} = \frac{49\times 48\times 47\times 46\times 45\times 44}{6\times 5\times 4\times 3\times 2\times 1} = 13{,}983{,}816.$$

■

One interesting application of the counting rules we have just developed is in assessing the probabilities of runs of like items in a sequence of random events. Let us examine this idea in terms of monitoring the quality of an industrial process in which manufactured items are produced at a regular rate throughout the day. Periodically, items are inspected; a D is recorded for a defective item, and a G is recorded for a good (nondefective) item. A typical daily record might look like this:

G D G G D G G G G D D G

for 12 inspected items.

If the Ds bunch together noticeably, one might conclude that the rate of producing defective items is not constant across the day. For example, a record such as

G G G G G G G G D D D D

might suggest that workers get careless toward the end of the day or that machinery used in the process got badly out of tune.

A key difference between these patterns involves the number of runs of like items. The first shows three runs of Ds and four runs of Gs for a total of seven runs. The second shows one G-run and one D-run for a total of two runs. If the rate of producing defects remains constant during the day, all possible arrangements of m Ds and n Gs, in a total of $(m + n)$ trials are equally likely. There are

$$\binom{m+n}{m} = \binom{m+n}{n}$$

such arrangements.

Suppose that the number of D-runs and the number of G-runs both equal k. The number of ways dividing m Ds into k groups can be determined by considering the placement of $(k - 1)$ bars into the spaces between runs of Ds:

D | D D | D D . . . D D D | D D.

This can be done in

$$\binom{m-1}{k-1}$$

ways. Similarly, n Gs can be divided into k groups in

$$\binom{n-1}{k-1}$$

ways. Because both of these divisions must occur together, there are a total of

$$\binom{m-1}{k-1}\binom{n-1}{k-1}$$

ways of producing k D-runs and k G-runs. When these two sets are merged, we could begin with either a D or a G; so the probability of getting $2k$ runs becomes

$$P(2k \text{ runs}) = \frac{2\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{m}}.$$

The number of D-runs can be one more (or one less) than the number of G-runs. To get $(k+1)$ D-runs and k G-runs, we must begin the sequence with a D and end it with a D. To get k D-runs and $(k+1)$ G-runs, we must begin and end the sequence with a G. Thus,

$$P(2k+1 \text{ runs}) = \frac{\binom{m-1}{k}\binom{n-1}{k-1} + \binom{m-1}{k-1}\binom{n-1}{k}}{\binom{m+n}{m}}.$$

EXAMPLE 2.17 A basketball player shot 44 free throws, making 39 and missing 5.

1 What is the probability that these hits and misses occurred in only two runs?

2 A sportswriter, observing that the player's five misses occurred in the last five free throws of the season, remarked that the player had been "unable to hit late in the season." Do you agree with this assessment?

Solution **1** The probability of having only two runs is given by

$$P(2 \text{ runs}) = \frac{2\binom{38}{0}\binom{4}{0}}{\binom{44}{39}} = \frac{2 \times 1 \times 1}{1,086,008} = \frac{1}{543,004}$$

assuming that all arrangements of 39 hits and 5 misses are equally likely. (What does that assumption say about the player's ability to hit a free throw?)

2 The probability of having one run of making 39 free throws followed by one run of missing 5 free throws, assuming equally likely arrangements, is

$$\frac{\binom{38}{0}\binom{4}{0}}{\binom{44}{39}} = \frac{1}{1,086,008}.$$

(Why do we not need to multiply by 2 in this case?)

If the player had a constant ability to hit free throws across the season, the chance of all five of his missed free throws coming at the end of the season would be very small. Thus the sportswriter could be correct. Are there other explanations?

■

Exercises

2.67 In a set of dominoes, each piece is marked with two numbers, one on each end. The pieces are symmetrical so that the two numbers are unordered. That is, the combination (2, 6) and (6, 2) are not distinguishable. How many pieces can be formed using the numbers $1, 2, \ldots, 12$?

2.68 Bose and Einstein determined that the distribution of particles among phase cells could be well modeled by assigning equal probability to distinguishable patterns (particles are indistinguishable, but phase cells are distinguishable), giving rise to "Bose-Einstein statistics." For a system of 8 particles and 10 phase cells, what is the probability that all particles will be in the same phase cell?

2.69 A baseball player is said to be having a hitting streak if he has gotten on base, by hitting the ball or walking, in a "large" number of consecutive times at bat. Suppose a batter has batted a large number of times during the season and that his batting average is 300; that is, he has hit 30% of his times at bat.

a What is the probability that the batter has a streak of six hits?

b How many hits must he get for you to consider this batter to have had an unusually long streak? Justify your choice.

2.70 A quality improvement plan calls for daily inspection of 10 items from a production process with the items periodically sampled throughout the day. To see whether a "clumping" of the defects seems to be occurring, inspectors count the total number of runs of defectives and nondefectives. Would you suspect a nonrandom arrangement if, among four defectives and eight nondefectives, the following numbers of runs occurred?

a 4

b 3

c 2

2.71 A group of 10 individuals purchased student tickets to attend a play. On the night of the play, three forgot their student IDs. It is known that the ticket takers randomly check IDs for 10% of the student tickets. There are two ticket takers on the night of the play. What is the probability that all members of the group get in without having any of the students without IDs being detected if they enter as follows?

a All 10 go to the same ticket taker.

b Five go to one ticket taker and the other five go to the other ticket taker.

2.72 Among 10 people traveling in a group, 2 have outdated passports. It is known that inspectors will check the passports of 20% of the people in any group passing their desks. The group can go as a whole (all 10) to one desk or can split into two groups of 5 and use two different desks. How should members of the group arrange themselves to maximize the probability of getting by the inspectors without having the outdated passports detected?

2.6 Summary

Data are the key to making sound and objective decisions, and *randomness* is the key to obtaining good data. The importance of randomness derives from the fact that the relative frequencies of random events tend to stabilize in the long run; this long-run relative frequency is called *probability*.

A more formal definition of probability allows for its use as a mathematical modeling tool to help explain and anticipate outcomes of events not yet seen. *Complements* and rules for *addition* and *multiplication* of probabilities are essential parts of the modeling process.

The relative frequency notion of probability and the resulting rules governing probability calculations have direct parallels in the analysis of frequency data recorded in tables. To correctly count the number of possible outcomes of an experiment, one must carefully consider whether or not order is important and whether sampling is with or without replacement.

Supplementary Exercises

2.73 A vehicle arriving at an intersection can turn left, turn right, or continue straight ahead. An experiment consists of observing the movement of one vehicle at this intersection.

- **a** List the outcomes in the sample space.
- **b** List all possible events.
- **c** If the vehicle turns right at the intersection, which event(s) in part (b) has (have) occurred.
- **d** Find the probability that the vehicle turns, assuming that each outcome in part (a) is equally likely.

2.74 When a patrolwoman monitors the speed of traffic using a radar gun, she may take any one of three possible actions for each car. She can let the car continue without stopping. She can stop the car and either issue a warning or issue a ticket. An experiment consists of observing the action of a patrolman after she records the speed of a car using the radar gun.

- **a** List the outcomes in the sample space.
- **b** List all possible events.
- **c** If she issues a warning to the driver of a car, which event(s) in part (b) has (have) occurred.
- **d** Attach probabilities to the outcomes in the sample space if the probability of the patrolman issuing a ticket is twice that of issuing a warning and a third of that of allowing the car to continue on.
- **e** Find the probability that the patrolwoman stops the car under the probabilistic model of part (d).

2.75 After a person interviews for admission to a medical school, he may be admitted, wait-listed, or denied admission. An experiment consists of observing the outcome for a person who interviews for medical school.

- **a** List the outcomes in the sample space.
- **b** List all possible events.
- **c** If the person is admitted, which event(s) in part (b) has (have) occurred.
- **d** Attach probabilities to the outcomes in the sample space if the probability of being wait-listed is half that of being admitted and three times as many applicants are denied than are admitted.

2.76 A hydraulic rework shop in a factory turned out seven rebuilt pumps today. Suppose that three pumps are still defective. Two of the seven are selected for thorough testing and then classified as defective or not defective.

- **a** List the outcomes for this experiment.
- **b** Let A be the event that the selection includes one defective. List the outcomes in A.
- **c** Based on the information given, assign probabilities to the outcomes and find $P(A)$.

2.77 An experiment consists of rolling a pair of dice.

- **a** Use the combinatorial theorems to determine the number of outcomes in the sample space S.
- **b** Find the probability that the sum of the numbers appearing on the dice is equal to 7.

2.78 Show that $\binom{3}{0} + \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 2^3$. Note that, in general, $\sum_{i=0}^{n} \binom{n}{i} = 2^n$.

2.79 A beach resort has 25 jet skis for guests to rent. Of these, 12 are two-person skis, 14 have turbo packs, and 7 are both for two persons and have turbo packs. Let T be the event that a jet ski, randomly chosen, is a two-person ski, and let P be the event that the ski has a turbo pack. A jet ski is chosen at random for rental. Find the probability for each of the following events.

a The jet ski is for two persons and has turbo packs.
b The jet ski is not for two persons but has turbo packs.
c The jet ski is for two persons but does not have turbo packs.

2.80 A particularly intense peewee basketball coach decides to place his players into groups based on their skills. The team consists of 12 players of which the coach classifies 4 as three-point specialists, 6 as defensive standouts, and 5 as neither three-point specialists nor defensive standouts. A player is chosen at random. Let T denote the event that a three-point specialist is selected and D denote the event that a defensive standout is selected. Find the probability of the following events.

a The selected player is a three-point specialist and a defensive standout.
b The selected player is a defensive standout but not a three-point specialist.
c The selected player is either a three-point specialist or a defensive standout.

2.81 In how many ways can a committee of three be selected from among 10 people?

2.82 An inspector must perform eight tests on a randomly selected keyboard coming off an assembly line. The sequence in which the tests are conducted is important because the time lost between tests varies. If an efficiency expert wanted to study all possible sequences to find the one that requires the minimum length of time, how many sequences would he include in his study?

2.83 In how many ways can a slate of 4 officers (president, vice-president, secretary, treasurer) be selected from among 20 people?

2.84 The Senate consists of 100 senators, 2 from each of the 50 states. A Senate committee of six is to be formed.

a How many ways are there to form the committee from the 100 senators?
b If at most one senator from each state can serve on the committee, how many ways are there to form the committee?

2.85 In an upcoming election, 12 states will elect a new governor. Each state has only two candidates, a Democrat and a Republican, running for governor. Also, in each state, the polls indicate that each candidate is equally likely to win. Find the probabilities associated with the following events.

a Democrats win all 12 gubernatorial elections.
b Half of the elections are won by Democrats and the other half by Republicans.

2.86 Two cards are randomly selected from a 52-card deck. What is the probability that the draw will yield an ace and a face card?

2.87 For an upcoming civil trial, 25 people are called for jury duty. Of these, 15 are men and 10 are women. Six people are needed for the jury.

a How many ways are there to select the jury from the people available?
b If the jury members are selected at random, what is the probability that three men and three women will be on the jury?
c If the jury members are selected at random, what is the probability that at least one man will be on the jury?

2.88 An elevator starts with five passengers and stops on eight floors. If each passenger is equally likely to get off on any floor, what is the probability that no two passengers will get off on the same floor?

2.89 At a European summit, four Germans, three Italians, and six Frenchmen are seated randomly on the front row. Find the probability that all members from the same country will be seated together.

2.90 Suppose n people attend a party and that every person shakes hands with every other person in attendance. How many handshakes will there be?

2.91 A deck of 52 cards is thoroughly shuffled so that the order of the cards in the deck is random. Then the cards are dealt to 4 bridge players; each player receives 13 cards. What is the probability that each player has all cards from a single suit; that is, one player has all hearts, one has all spades, one has all diamonds, and one has all clubs?

2.92 For a particular office telephone, seven telephone calls are randomly received in a week, and these calls are randomly distributed among the seven days. What is the probability that no call will be received on exactly one day?

REFERENCE

Kerrich, J. 1964. *An experimental introduction to the theory of probability*. Copenhagen: J. Jorgenson.

3

Conditional Probability and Independence

3.1 Conditional Probability

As we saw in Chapter 2, when conducting an experiment, we are often interested in the probabilities of two or more events. At times, some partial information about the outcome of an experiment is available, and we want to take advantage of this information when calculating probabilities. Two examples are used to illustrate this idea. First, consider the employment data in Table 3.1.

TABLE **3.1**
Civilian Labor Force in the United States, 2004 (Figures in Thousands).

Education	Employed	Unemployed	Total
Less than a high school diploma	11,408	1062	12,470
High school graduate, no college	35,944	1890	37,834
Some college, no degree	21,284	1014	22,298
Associate degree	11,693	447	12,141
Bachelor's degree and higher	39,293	1098	40,390
Total	119,622	5511	125,133

Note: Figures are for noninstitutionalized civilians who are at least 25 years of age.
Source: U.S. Bureau of Labor Statistics.

TABLE 3.2
2004 Employment Rate by Education.

Education	Employed	Unemployed
Less than a high school diploma	91.5%	8.5%
High school graduate, no college	95.0%	5.0%
Some college, no degree	95.5%	4.5%
Associate degree	96.3%	3.7%
Bachelor's degree and higher	97.3%	2.7%

Source: United States Bureau of Labor Statistics.

A common summary of these data is the "unemployment rate," which is the percentage of unemployed workers, given by

$$\frac{5{,}511{,}000}{125{,}133{,}000}(100) = 4.4.$$

(This figure does not take into account persons no longer actively seeking work.) But, the overall unemployment rate does not tell us anything about the association between employment and education. To get at this question, we must calculate unemployment rates separately for each education category (each row of the table). Narrowing the focus to a single row is often referred to as *conditioning* on the row factor.

The conditional relative frequencies for the data of Table 3.1 are given in Table 3.2. It is apparent that unemployment is associated to some extent with educational level; categories of less education have higher unemployment rates. The conditional relative frequencies relate directly to **conditional probability**. If a national poll samples 1000 people from the national labor force, the expected percentage of unemployed workers it would find (in 2004) is about 4.4% of 1000 (1000(5511)/125,133)—that is, 44 individuals. If, however, the 1000 people all have four or more years of college education, the expected percentage of unemployed workers drops to 2.7% or 27 people.

EXAMPLE 3.1

Projected percentages of workers in the labor force for 2014 are shown in Table 3.3. How do the relative frequencies for the four ethnic groups compare between women and men?

TABLE 3.3
Projected Percentage of Workers in 2014.

	Men	Women	Total
White	43%	37%	80%
Black	6%	6%	12%
Asian	3%	3%	6%
Other	1%	1%	2%
Total	53%	47%	100%

Source: United States Bureau of Labor Statistics.

Solution Even though the data are expressed as percentages rather than as frequencies, the relative frequencies can still be computed. The total number of men represents 53% of the population, while the number of white men represents 43% of the population. Therefore, (43/53) represents the proportion of whites among men. Proceeding similarly across the other categories produces the two conditional distributions (one for

TABLE 3.4
Projected Percentage of Workers in 2014, by Gender.

	Men	Women
White	81%	79%
Black	11%	13%
Asian	6%	6%
Other	2%	2%
Total	100%	100%

Source: United States Census Bureau.

men and one for women), shown in Table 3.4. Notice that the proportion of each ethnic group is about the same for men and women; the proportion of an ethnic group changes little, if any, with gender. ▪

Look back at Table 3.1. If we randomly select a person from the civilian labor force, the probability that the person will be unemployed is 0.044. If we randomly select a person from the civilian workforce who has less than a high school diploma, the probability that the person is unemployed is 0.085. One consequence of having additional information is that the *sample space is reduced*. When randomly selecting a person who has less than a high school diploma, the sample space is restricted from all people in the civilian workforce to only those with less than a high school diploma. As another illustration, suppose we consider the probability that a family with two children has two girls. The sample space is

$$S = \{BB,\ BG,\ GB,\ GG\}$$

where B represents a boy and G represents a girl. The order of the letters represents the birth order so BG represents a family with the older child being a boy and the younger a girl. Because each outcome in the sample space is equally likely, the probability of two girls is $1/4$.

If we are told that a family has at least one girl, what is the probability that the family has two girls? The sample space $= \{BB, BG, GB, GG\}$ is no longer appropriate because BB is not a possible outcome if the family has at least one girl. Instead, using the information provided, we have the reduced sample space

$$S_R = \{BG,\ GB,\ GG\}.$$

Because the outcomes in this reduced sample space are equally likely, the probability of a family having two girls given that they have at least one girl is $1/3$. (*Note:* It is important to realize that we are given that at least one child is a girl and not that the oldest [or youngest] child is a girl. If we are told which child is a girl, the sample space is reduced more than if we are told that at least one is a girl, and the probability of two girls would then be $1/2$.)

To formalize the preceding discussion, let A be the event that a family has two girls and B be the event that a family has at least one girl. We have $P(A \text{ given } B)$—written as $P(A|B)$—is $1/3$. Notice that

$$P(A|B) = \frac{1}{3} = \frac{P(AB)}{P(B)} = \frac{1/4}{3/4}.$$

This relationship motivates Definition 3.1.

DEFINITION **3.1**

If A and B are any two events, then the **conditional probability** of A given B, denoted by $P(A|B)$, is

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

provided that $P(B) > 0$. ▪

Notice that the equation for conditional probability may be rewritten in terms of $P(AB)$ as follows:

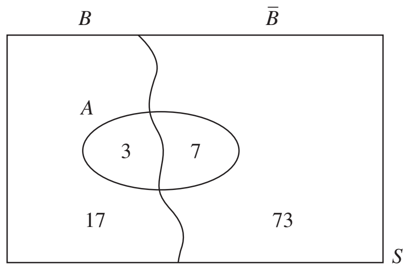
$$P(AB) = P(A|B)P(B)$$

or

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

Conditioning can be represented in Venn diagrams as well. Of 100 students who completed an introductory statistics course, 20 were business majors. Further, 10 students received As in the course, and three of these were business majors. These facts are easily displayed on a Venn diagram, such as Figure 3.1, where A represents students who received As and B represents business majors.

FIGURE **3.1**
Partitioning of the Sample Space and the Event A.



Then, by Definition 3.1,

$$P(A|B) = \frac{3}{20} = \frac{P(AB)}{P(B)} = \frac{3/100}{20/100}.$$

EXAMPLE **3.2**

There are four batteries, and one is defective. Two are to be selected at random for use on a particular day. Find the probability that the second battery selected is not defective, given that the first was not defective.

Solution

Let N_i denote the event that the ith battery selected is nondefective. We want to find $P(N_2|N_1)$. From Definition 3.1, we have

$$P(N_2|N_1) = \frac{P(N_1N_2)}{P(N_1)}.$$

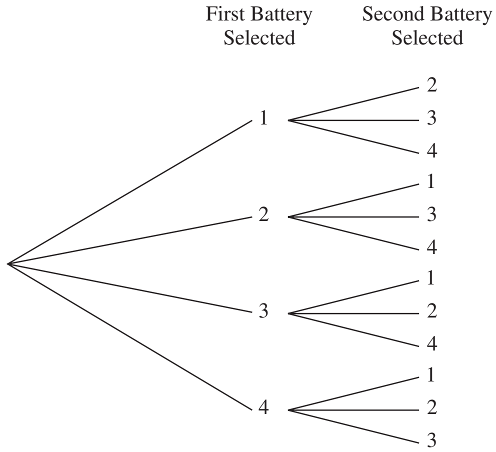


FIGURE **3.2**
Outcomes of Experiment in Example 3.2.

The tree diagram associated with the experiment of selecting two batteries from among four, one of which (say the fourth one) is defective, is displayed in Figure 3.2. Of the 12 possible outcomes, we can see that event N_1 contains 9 of these outcomes, and N_1N_2 contains 6. Thus, because the 12 outcomes are equally likely,

$$P(N_2|N_1) = \frac{P(N_1N_2)}{P(N_1)} = \frac{6/12}{9/12} = \frac{6}{9} = \frac{2}{3}.$$

Does this answer seem intuitively reasonable? ▪

Conditional probabilities satisfy the three axioms of probability (Definition 3.1), as can easily be seen. First, because $AB \subset B$, then $P(AB) \leq P(B)$. Also, $P(AB) \geq 0$, so

$$0 \leq P(A|B) = \frac{P(AB)}{P(B)} \leq 1.$$

Second,

$$P(S|B) = \frac{P(SB)}{P(B)} = \frac{P(B)}{P(B)} = 1.$$

Third, if $A_1, A_2, \ldots$ are mutually exclusive events, then so are $A_1B, A_2B, \ldots$; and

$$P\left(\bigcup_{i=1}^{\infty} A_i \middle| B\right) = \frac{P\left(\left(\bigcup_{i=1}^{\infty} A_i\right) B\right)}{P(B)}$$

$$= \frac{P\left(\bigcup_{i=1}^{\infty}(A_iB)\right)}{P(B)} = \frac{\sum_{i=1}^{\infty}P(A_iB)}{P(B)}$$

$$= \sum_{i=1}^{\infty}\frac{P(A_iB)}{P(B)} = \sum_{i=1}^{\infty}P(A_i|B).$$

Conditional probability plays a key role in many practical applications of probability. In these applications, important conditional probabilities are often drastically affected by seemingly small changes in the basic information from which the probabilities are derived. The following discussion of a medical application of probability illustrates the point.

A screening test indicates the presence or absence of a particular disease; such tests are often used by physicians to detect diseases. Virtually all screening tests, however, have levels of error associated with their use. Two different kinds of errors are possible: The test could indicate that a person has the disease when he or she actually does not (false positive), or it could fail to show that a person has the disease when he or she actually does have it (false negative). Measures of the probability of *not* making one of these errors are conditional probabilities called sensitivity and specificity. *Sensitivity* is the probability a person selected randomly from among those who have the disease will have a positive test. *Specificity* is the probability that a person selected randomly from among those who do not have the disease will have a negative test.

The following diagram helps in defining and interpreting these measures, where the + indicates the presence of the disease under study and the − indicates the absence of the disease. The true diagnosis may never be known, but often it can be determined by more intensive follow-up tests.

		True Diagnosis		
		+	−	Sum
Test	+	a	b	$a + b$
Result	−	c	d	$c + d$
Sum		$a + c$	$b + d$	$a + b + c + d = n$

In this scenario, n people are tested and the test results indicate that $a + b$ of them have the disease. Of these, a really have the disease and b do not (false positives). Of the $c + d$ who test negative, c actually do have the disease (false negatives). Using these labels,

$$\text{Sensitivity} = \frac{a}{a+c}$$

which represents the conditional probability of having a positive test, given that the person has the disease; and

$$\text{Specificity} = \frac{d}{b+d}$$

which represents the conditional probability of having a negative test, given that the person does not have the disease.

Obviously, a good test should have values for both sensitivity and specificity that are close to 1. If sensitivity is close to 1, then c (the number of false negatives) must be small. If specificity is close to 1, then b (the number of false positives) must be small. Even when sensitivity and specificity are both close to 1, a screening test can produce misleading results if it is not carefully applied. To illustrate this, let us look at one other important measure, the *predictive value* of a test, which is given by

$$\text{Predictive value} = \frac{a}{a+b}.$$

The *predictive value* is the conditional probability that a randomly selected person actually has the disease, given that he or she tested positive. Clearly, a good test should have a high predictive value, but this is not always possible—even for highly sensitive and specific tests. The reason that all three measures may not always be close to 1 simultaneously is that the predictive value is affected by the *prevalence* of the disease (that is, the proportion of the population under study that actually has the disease). We can show this with examples of three numerical situations (given next as diagrams I, II, and III).

			True Diagnosis		
			+	−	Sum
I.	Test	+	90	10	100
	Result	−	10	90	100
	Sum		100	100	200

			+	−	Sum
II.	Test	+	90	100	190
	Result	−	10	900	910
	Sum		100	1000	1100

			+	−	Sum
III.	Test	+	90	1000	1090
	Result	−	10	9000	9010
	Sum		100	10,000	10,100

Among the 200 people under study in diagram I, 100 have the disease (a prevalence rate of 50%). The sensitivity and the specificity of the test are each equal to 0.90, and the predictive value is $90/100 = 0.90$. This is a good situation; the test is a good one.

In diagram II, the prevalence rate changes to 100/1100, or 9%. Even though the sensitivity and specificity values are both still 0.90, the predictive value has dropped to $90/190 = 0.47$. In diagram III, the prevalence rate is 100/10,100, or about 1%, and the predictive value has dropped farther to 0.08. Thus, only 8% of those tested positive actually have the disease, even though the test has high sensitivity and high specificity. What does this imply about the use of screening tests on a large population in which the prevalence rate for the disease being studied is low? Assessing the answer to this question involves taking a careful look at conditional probabilities.

EXAMPLE 3.3 Nucleic acid amplification tests (NAATs) are generally agreed to be better than non-NAATs for diagnosing the presence of *Chlamydia trachomatis,* the most prevalent sexually transmitted disease. The ligase chain reaction (LCR) test is one such test. In a large study, the sensitivity and specificity of LCR for women were assessed. Following are the results:

LCR		Tissue Culture +	Tissue Culture −	Sum
Test	+	139	84	223
Results	−	13	1,896	1,909
Sum		152	1,980	2,132

Source: Hadgu (1999).

Assuming that the tissue culture is exact (a "gold standard") and that the women in the study constitute a random sample of women in the United States, answer the following questions.

a What is the prevalence of *Chlamydia trachomatis*?
b What is the sensitivity of LCR?
c What is the specificity of LCR?
d What is the predictive value of LCR?

Solution **a** The prevalence is

$$\text{Prevalence} = \frac{152}{2132} = 0.071.$$

An estimated 7.1% of women have *Chlamydia trachomatis.*

b The sensitivity is estimated to be

$$\text{Sensitivity} = \frac{139}{152} = 0.914.$$

That is, the sensitivity is estimated to be 91.4%.

c The specificity is estimated to be

$$\text{Specificity} = \frac{1896}{1980} = 0.958.$$

The specificity is 95.8%.

d The predictive probability of LCR is

$$\text{Predictive value} = \frac{139}{223} = 0.623.$$

That is, the predictive value is estimated to be 62.3%.

Note: One of the challenges with this approach is that we have assumed that the tissue culture is 100% accurate, and this is not the case. Various approaches have been suggested to adjust the estimates to account for this lack of accuracy in the gold standard. A second assumption is that the women in the sample are a random sample from the population. This may or may not be a valid assumption. Even if it is valid, we have not looked at all women so that we have estimates, and not the true values, of the quantities of interest. This is always the case when one is working with a sample from a population and not the whole population. Because the sample size in this study is large, we believe that the estimates will be close to the true values of interest.

Exercises

3.1 The likelihood of a fatal vehicular crash is affected by numerous factors. The fatal crashes by speed limit and land use during 2004 are given in the table that follows.

	Land Use		
Speed Limit (mph)	**Rural**	**Urban**	**Total**
≤ 30	944	2,929	3,873
> 30 and ≤ 40	1,951	4,463	6,414
> 40 and ≤ 50	3,496	3,559	7,055
55	9,646	2,121	11,767
≥ 60	5,484	2,347	7,831
No statutory limit	92	31	123
Total	21,613	15,450	37,063

Source: U.S. Department of Transportation (2005).

Suppose a 2004 fatal crash is selected at random. What is the probability that it occurred

a in a rural area?

b in an area with a speed limit of no more than 50 mph?

c in a rural area, given that the speed limit was no more than 40 mph?

d in an urban area, given that the speed limit was no more than 40 mph?

3.2 The National Survey on Drug Use and Health is conducted annually to provide data on drug use in the United States. The results of a study of those who have smoked a cigarette within the past year for persons aged 26 and older are displayed, by educational level, in the following table.

Education	Smoked a Cigarette Within Past Year	Have Not Smoked a Cigarette Within Past Year	Total
< High School Diploma	10,393	19,472	29,865
High School Graduate	17,798	39,247	57,045
Some College	13,463	30,969	44,432
College Graduate	8,320	43,357	51,677
Total	49,974	133,045	183,019

Source: Substance Abuse & Mental Health Services Administration, Office of Applied Studies (2005).

Answer the following questions based on the information given in the preceding table.

a What is the probability that a randomly selected person has smoked a cigarette within the past year?

b What is the probability that a randomly selected person who has not completed high school smoked a cigarette in the past year?

c What is the probability that a randomly selected person who has smoked a cigarette in the past year has completed college?

d If a person has had at least some college (could be a college graduate), is it more or less likely that she or he has smoked a cigarette in the past year as compared to a person with at most a high school diploma? Justify your answer.

3.3 The numbers of workers, in thousands, in the U.S. workforce in 2004 are shown in the next table.

Age (in years)	Men	Women	Total
16 to 24	11,673	10,595	22,268
25 to 44	37,337	31,028	68,365
45 to 64	27,182	24,589	51,771
65 and Older	2,787	2,211	4,998
Total	78,979	68,423	147,402

Source: U.S. Bureau of Labor Statistics.

Answer the following questions based on the information given in the preceding table.

a What is the probability that a randomly selected worker is a male who is at least 65 years of age?

b What is the probability that a randomly selected worker is a female?

c What is the probability that a randomly selected worker between 16 and 24 years old is a male?

d What is the probability that a randomly selected female worker is between 25 and 64 years of age?

3.4 The numbers (in thousands), by gender and marital status, of the 2004 U.S. population aged 15 and over are shown in the table that follows.

	Never Married	Married	Widowed	Divorced	Totals
Females	23,655	63,282	11,141	12,804	110,882
Males	29,561	62,483	2,641	8,956	103,641
Totals	53,216	125,765	13,782	21,760	214,523

Source: Statistical Abstract of the United States (2006).

Answer the following questions based on the information given in the preceding table.

a What is the probability that a randomly selected person has never been married?

b What is the probability that a randomly selected female is a widow?

c What is the probability that a randomly selected divorced person is a male?

d Is it more likely for males or females to be widowed? Justify your answer.

3.5 The probability that Trevor eats breakfast and gets to work on time is 0.2. The probability that he eats breakfast is 0.4. If Trevor eats breakfast, what is the probability that he is on time for work?

3.6 The probability that Elise studies for a science test and passes it is 0.8. The probability that she studies is 0.9. If Elise studies, what is the probability that she will pass the science test?

3.7 According to the National Center for Statistics and Analysis, in 2004, 28% of the drivers involved in fatal crashes were between 20 and 29 years of age. Further, 39% of the 20- to 29-year-old drivers involved in fatal crashes had a blood alcohol level of at least 0.01. In what percentage of fatal crashes were the drivers between 20 and 29 years of age and found to have a blood alcohol level above 0.01?

3.8 Jessica does not have any cat food and plans to buy some on her way home from work. However, she tends to be forgetful. Even though she passes by the grocery store, the probability that she will remember to stop is 0.6. If she remembers to stop, Jessica will decide to pick up other things that she also needs. The probability that she will include the cat food among her purchases is 0.5. What is the probability that Jessica will have cat food to give her cat when she gets home?

3.9 In a particular community, 70% of the voters are Democrats, and 30% are Republicans. Thirty percent of the Republican voters and 60% of the Democratic voters favor the incumbent. What is the probability that a randomly selected voter from this community is a Republican who favors the incumbent?

3.10 A purchasing office is to assign a contract for copier paper and another contract for DVD + RWs to any one of three firms bidding for these contracts. (Any one firm could receive both contracts.) Find the probabilities of the following events.

a Firm I receives both contracts.

b Firm I received a contract, given that both contracts do not go to the same firm.

c Firm I receives the contract for paper, given that it does not receive the contract for the DVD + RWs.

d What assumptions have you made?

3.11 An incoming lot of silicon wafers is to be inspected for defective ones by an engineer in a microchip manufacturing plant. Suppose that in a tray containing 20 wafers 4 are defective. Two wafers are to be selected randomly for inspection. Find the probabilities of the following events.

a Neither is defective.

b At least one of the two is defective.

c Neither is defective, given that at least one is not defective.

3.12 In the setting of Exercise 3.11, answer the same three questions, assuming this time that only 2 among the 20 wafers are defective.

3.13 The frequency of twin births has been increasing in recent years so that in 2000 about 3% of all births were twins. Twins may be either identical or fraternal. The probability of a birth being that of identical twins has remained fairly constant at 0.4%. What proportion of twin births is identical?

3.14 By using the definition of conditional probability, show that

$$P(ABC) = P(A)P(B|A)P(C|AB).$$

3.15 Children and adults with sore throats are often tested for strep throat. If untreated, strep throat can lead to rheumatic fever. The traditional method for assessing whether or not someone has strep throat is a culture. Because the results of the culture take a day to obtain, more rapid tests are often used. The Biostat A Optical Immunoassay (Strep A OIA), first developed in the early 1990s, is one of the rapid tests. Heiter and Bourbeau (1995) conducted a study in which the results of both the culture and Strep A OIA were obtained for 801 patients who potentially had strep throat. The results are in the following table.

	Number of Results			
Test	**True Positive**	**False Positive**	**True Negative**	**False Negative**
Culture	239	7	555	0
Strep A OIA	225	21	526	29

Answer the following questions based on the information given in the preceding table.

a What is the sensitivity of the culture?

b What is the specificity of the culture?

c What is the predictive value of the culture?

3.16 Using the results presented in the table in Exercise 3.15, answer the following questions.

a What is the sensitivity of the Strep A OIA?

b What is the specificity of the Strep A OIA?

c What is the predictive value of the Strep A OIA?

3.17 In a certain village, the eldest daughter is expected to tend the family's flock of sheep. Suppose that every family in this large village has two children. What proportion of all daughters are eldest daughters? What assumptions did you make?

3.18 Recall the dice game played by the Chevalier de Méré described in Exercise 2.58. Suppose a player bets that at least one 6 would appear during a total of four rolls. However, after the first two rolls, neither of which resulted in a 6, she had to leave the game. What proportion of her bet should be returned to her for this to be a fair game? (A game is fair if the amount paid to play is equal to the probability of winning multiplied by the amount that could be won.)

3.19 Students in a college program have two opportunities to pass an exam required for graduation. The probability that a student passes the test the first time is 0.8. For those who fail the first time, the probability of passing the test the second time is 0.6.

a Find the probability that a randomly selected student passes the test.

b If the student passes the test, what is the probability that she or he did so on the first try?

3.20 In a version of five-card-stud poker, each player is dealt five cards. Three of the five are dealt face down so that only the player can see them; the other two are dealt face up. Two players are dealt cards. One player has two jacks showing; the other has a 2 and a 10. What is the probability that the first player has a full house?

3.21 Let A and B be events in a sample space with positive probability. Prove that $P(B|A) > P(B)$ if and only if $P(A|B) > P(A)$.

3.22 A survey is to be taken, and one of the questions to be asked is, "Have you ever smoked marijuana?" If the respondent has smoked marijuana, she may be concerned that a truthful response would lead to prosecution and thus respond "no," leading to an inaccurate estimate of the proportion in the population who has smoked marijuana. The randomized response design was developed to provide better information for such questions. Here the respondent may be instructed to flip a coin, not allowing the interviewer to see the outcome. If the coin lands with the "head" up, the respondent is to respond "yes"; otherwise, the respondent answers the question truthfully (when the coin lands with the "tail" side up). Let p be the proportion of people in the population who have smoked marijuana, and suppose that 0.7 of the people respond "yes." What proportion of the population would be estimated to have smoked marijuana?

3.2 Independence

Probabilities are usually sensitive to the conditioning information. Sometimes, however, a probability does not change when conditioning information is supplied. If the extra information provided by knowing that an event B has occurred does not change the probability of A—that is, if $P(A|B) = P(A)$—then events A and B are said to be **independent.** Because

$$P(A|B) = \frac{P(AB)}{P(B)},$$

the condition $P(A|B) = P(A)$ is equivalent to

$$\frac{P(AB)}{P(B)} = P(A)$$

or

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

DEFINITION **3.2**

Two events A and B are said to be **independent** if

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

This is equivalent to stating that

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

if the conditional probabilities exist. ▪

Sometimes a conditional probability is known, and we want to find the probability of an intersection. By rearranging the terms in the definition of conditional probability and considering the definition of independence, we obtain the **Multiplicative Rule.**

THEOREM **3.1**

Multiplicative Rule. If A and B are any two events, then

$$\begin{aligned} P(AB) &= P(A)P(B|A) \\ &= P(B)P(A|B) \end{aligned}$$

If A and B are independent, then

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

▪

From Table 3.1, we found the unemployment rate during 2004 to be 4.4%. The conditional distributions of being employed given educational level and of being unemployed given educational level were presented in Table 3.2 and are displayed again in the next table. Notice again that the probability of being unemployed decreases with educational level. As an example, if we randomly select an individual from those who have completed at least a bachelor's degree, the probability of that person being unemployed is 0.027, which is well below the 0.04 probability associated with the whole population. Because the probability of being unemployed changes with educational level, unemployment and educational level are *not* independent; they are dependent.

Education	Employed	Unemployed
Less than a high school diploma	91.5%	8.5%
High school graduate, no college	95.0%	5.0%
Some college, no degree	95.5%	4.5%
Associate degree	96.3%	3.7%
Bachelor's degree and higher	97.3%	2.7%

Source: United States Bureau of Labor Statistics.

Now consider the conditional distributions of the projected percentage of workers in each ethnic group by 2014 for men and women that were first presented in Table 3.4 and are displayed again next. Here the two conditional distributions are very close to each other. Knowing whether a worker is male or female provides little, if any, additional information. If these distributions are exactly equal to the values in the table, then ethnicity and gender are not independent because the probability that a randomly selected male worker is white is not the same as the probability that a randomly selected female worker will be white. However, because these are projections, one may argue that the small differences in these two distributions could be due to the error of the projections and that ethnicity and gender could be independent.

	Men	Women
White	81%	79%
Black	11%	13%
Asian	6%	6%
Other	2%	2%
Total	100%	100%

Source: United States Census Bureau.

EXAMPLE 3.4 Suppose that a foreman must select one worker from a pool of four available workers (numbered 1, 2, 3, and 4) for a special job. He selects the worker by mixing the four names and randomly selecting one. Let A denote the event that worker 1 or 2 is selected, let B denote the event that worker 1 or 3 is selected, and let C denote the event that worker 1 is selected. Are A and B independent? Are A and C independent?

Solution Because the name is selected at random, a reasonable assumption for the probabilistic model is to assign a probability of $1/4$ to each individual worker. Then $P(A) = 1/2$,

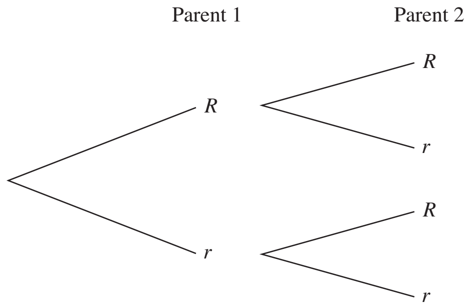
$P(B) = 1/2$, and $P(C) = 1/4$. Because the intersection AB contains only worker 1, $P(AB) = 1/4$. Now $P(AB) = 1/4 = P(A)P(B)$, so A and B are independent. Because AC also contains only worker 1, $P(AC) = 1/4$. But $P(AC) = 1/4 \neq 1/8 = P(A)P(C)$, so A and C are not independent. A and C are said to be *dependent* because the fact that C occurs changes the probability that A occurs. ▪

Most situations in which independence issues arise are not like the one portrayed in Example 3.4, where events were well defined and we merely calculated probabilities to check the definition. Often independence is assumed for two events in order to calculate their joint probability. For example, let A denote the event that machine A does not break down today, and let B denote the event that machine B does not break down today. $P(A)$ and $P(B)$ can be approximated from the repair records of the machines. How do we find $P(AB)$, the probability that neither machine breaks down today? If we assume independence, $P(AB) = P(A)P(B)$—a straightforward calculation. If we do not assume independence, however, we cannot calculate $P(AB)$ unless we form a model for their dependence structure or collect data on their joint performance. Is independence a reasonable assumption? It may be if the operation of one machine is not affected by the other; but it may not be if the machines share the same room, the same power supply, or the same job foreman. Thus, independence is often used as a simplifying assumption and may not hold precisely in all cases where it is assumed. Remember, probabilistic models are simply models; they do not always precisely mirror reality. But all branches of science make simplifying assumptions when developing their models, whether these are probabilistic or deterministic. The genetics application that follows is an example of using the simplifying assumption of independence.

Genetics is one of the most active fields of current scientific research. A unit of inheritance is a gene, which transmits chemical information that is expressed as a trait, such as color or size. Each individual plant or animal has many genes. In many familiar organisms, two genes for each trait are present in each individual. These paired genes, both governing the same trait, are called alleles. The two allelic genes in any one individual may be alike (homozygous) or different (heterozygous). When two individuals mate, each parent contributes one of his or her genes from each allele. In the simplest probabilistic model, the probability of each gene from an allele being passed to the offspring is $1/2$, and the two parents contribute alleles independently of each other.

In 1856, Gregor Mendel, a monk, began a series of inheritance studies using peas. He studied seven traits, each determined by a single allele. One trait was whether the peas were round (R) or wrinkled (r). He began with peas that either had two round genes (RR) or two wrinkled genes (rr). When he crossed the homozygous round peas (RR) with the homozygous wrinkled peas (rr), all offspring were round peas! This led Mendel to conclude that the round gene was dominant over the wrinkled gene, which is recessive; that is, peas that are heterozygous (Rr) for this trait are round because the R gene dominates the recessive r gene. He then crossed these heterozygous (Rr) peas with each other. Assuming that each parent is equally likely to contribute either gene (R or r) and that the gene one parent contributes is independent of the gene contributed by the other parent, the tree diagram in Figure 3.3 presents the possible outcomes.

FIGURE 3.3
The outcomes from crossing two peas that are heterozygous for round.



From the tree diagram, we see that $1/4$ of the peas are expected to be homozygous round (RR), $1/4$ are expected to be homozygous wrinkled (rr), and $1/2$ are expected to be heterozygous (Rr) and thus express the dominant trait of round. Thus $3/4$ of the peas are expected to be round and $1/4$ wrinkled. For one study, Mendel obtained 433 and 133 round and wrinkled peas, respectively. His observed proportion of round peas was

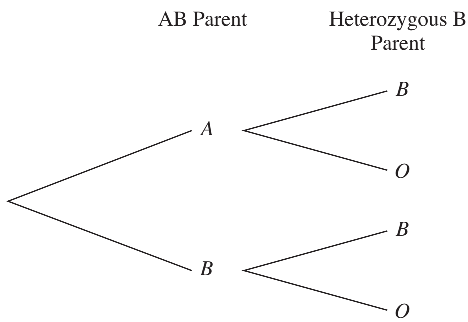
$$\frac{433}{566} = 0.765,$$

which is very close to the predicted 0.75. Mendel first presented his results in 1865, but it was not until the twentieth century that scientists verified the existence of genes.

EXAMPLE 3.5 Blood type, the best known of the blood factors, is determined by a single allele. Each person has blood type A, B, AB, or O. Type O represents the absence of a factor and is recessive to factors A and B. Thus, a person with type A blood may be either homozygous (AA) or heterozygous (AO) for this allele; similarly, a person with type B blood may be either homozygous (BB) or heterozygous (BO). Type AB occurs if a person is given an A factor by one parent and a B factor by the other parent. To have type O blood, an individual must be homozygous O (OO). Suppose a couple is preparing to have a child. One parent has blood type AB, and the other is heterozygous B. What are the possible blood types the child will have and what is the probability of each?

Solution First, we will use a tree diagram to help us determine all the options (Figure 3.4).

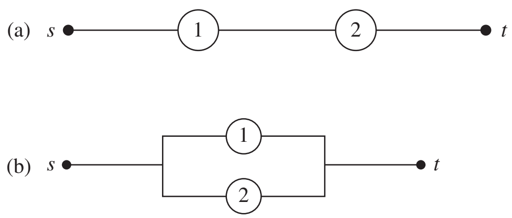
FIGURE 3.4



Notice that the first set of branches represents the gene given by the parent with type AB blood. Because we assume that each gene is equally likely to be given, the probability is 0.5 that the parent gives an A factor to the child, and that it is 0.5 that the parent gives a B factor to the child. Similarly, as represented by the second set of branches, the second parent will give either a B factor or no factor (O), each with probability 0.5. Thus, the four possible outcomes (AB, AO, BB, and BO) are equally likely. The probability that the child will have type B blood is 0.5 because BB and BO are both expressed as type B. The probabilities of type AB and type A (AO) are each 0.25. ▪

Relays in electrical circuits are often assumed to work (or fail) independently of each other. These relays may be set up in series (Figure 3.5[a]), in parallel (Figure 3.5[b]), or in some combination of series and parallel. For current to flow through a relay, it must be closed. A switch is used to open or close a relay. The circuit functions if current can flow through it.

FIGURE **3.5**
Circuits in series (a) or in parallel (b).



EXAMPLE **3.6** A section of an electrical circuit has two relays in parallel, as shown in Figure 3.5(b). The relays operate independently, and when a switch is thrown, each will close properly with a probability of 0.8. If both relays are open, find the probability that the current will flow from s to t when the switch is thrown.

Solution Let O denote an open relay, and let C denote a closed relay. The four outcomes from this experiment are shown in the following diagram.

$$\begin{array}{ccll} & & \text{Relay} & \text{Relay} \\ & & \quad 1 & \quad 2 \\ E_1 & = & \{(O, & O)\} \\ E_2 & = & \{(O, & C)\} \\ E_3 & = & \{(C, & O)\} \\ E_4 & = & \{(C, & C)\} \end{array}$$

The probability that a relay closes is given to be $P(C) = 0.8$. Therefore, the probability that a relay remains open is $P(O) = 1 - P(C) = 0.2$. Because the relays operate independently, we can find the probabilities for each of these outcomes as follows:

$$\begin{aligned}
P(E_1) &= P(O)P(O) = (0.2)(0.2) = 0.04 \\
P(E_2) &= P(O)P(C) = (0.2)(0.8) = 0.16 \\
P(E_3) &= P(C)P(O) = (0.8)(0.2) = 0.16 \\
P(E_4) &= P(C)P(C) = (0.8)(0.8) = 0.64
\end{aligned}$$

If A denotes the event that current flows from s to t, then

$$A = E_2 \cup E_3 \cup E_4$$

or, the event that the current does not flow from s to t is

$$\bar{A} = E_1.$$

That is, at least one of the relays must close in order for current to flow. Thus,

$$\begin{aligned}
P(A) &= 1 - P(\bar{A}) \\
&= 1 - P(E_1) \\
&= 1 - 0.04 \\
&= 0.96
\end{aligned}$$

which is the same as $P(E_2) + P(E_3) + P(E_4)$. ■

When applying probability theory in real-life applications, the assumptions that are made can have a dramatic effect on the conclusions and the consequences of those conclusions as illustrated by the next example.

EXAMPLE 3.7 Sally and Stephen Clark lived in Cheshire, England. In 1996, their child Christopher, aged 11 weeks, died in his sleep. The death was certified as natural because there was sign of an infection but not evidence of lack of care. About a year later their child Harry, aged 8 weeks, died in his sleep. Four weeks later the couple was arrested. Eventually Sally Clark was tried for murdering both children. At her trial, Sir Roy Meadow, a pediatrician, spoke as an expert for the prosecution. He claimed that the probability of two babies in the same, affluent, nonsmoking family dying of cot death (referred to as crib death in the United States) is 1 in 73 million. Meadow reached his conclusion based on an analysis of data from the Confidential Enquiry for Stillbirths and Deaths in Infancy (CESDI), a study of deaths of babies in five regions of England from 1993 through 1996. Based on the CESDI, it is estimated that the probability that a randomly selected baby dies from cot death is 1 in 1300. If the family is affluent and nonsmoking (as the Clarks were), the probability that a baby dies of cot death is 1 in 8500. If the occurrence of cot death is independent from baby to baby, then the probability of two randomly selected babies from an affluent, nonsmoking family dying of cot death is about 1 in 73 million as reported by Sir Meadow. What, if anything, is wrong with this reasoning?

Solution The key assumption is the independence in the occurrence of cot death from baby to baby. However, Christopher and Harry were siblings and would not represent a random selection of two babies from the population. Is the occurrence of cot death independent for siblings? In a subsequent analysis of the CESDI data, Professor Ray Hill of Salford University concluded that such deaths are not independent. He estimated that siblings of children who die of cot death are 10 to 22 times more likely than average to die the same way. If the probability of a first cot death within a family is 1 in 1303, the probability of a second one is then between 1 in 60 and 1 in 130. Using an intermediate value of 1 in 100, the probability of two cot deaths within a family is estimated to be

$$\frac{1}{1303}\left(\frac{1}{100}\right) = \frac{1}{130{,}000}.$$

About 650,000 children are born each year in England and Wales so we would expect about $650{,}000(1/130{,}000) = 5$ families to lose a second child to cot death each year. This is consistent with public records of such events.

What happened to Sally Clark? With little forensic data, the jury's decision was based primarily on Meadow's probabilistic argument. Sally Clark was convicted of murder and sent to prison in 1999 although she continued to claim her innocence. In 2002, evidence emerged that Harry had a *Staphylococcus aureus* infection that had spread to his cerebrospinal fluid. Sally Clark was released in January 2003. However, she never recovered from the loss of her children and subsequent incarceration, dying in 2007 at the age of 42. ▪

Exercises

3.23 The percentages, by age and gender categories, of the 2004 U.S. population aged 16 and over are shown in the following table.

Age (in years)	Men	Women	Total
16 to 24	18	17	35
25 to 44	15	14	29
45 to 64	12	12	24
65 and Older	5	7	12
Total	50	50	100

Source: United States Census Bureau.

a Find the conditional distribution of age group for men.

b Find the conditional distribution of age group for women.

c Is age independent of gender? Justify your answer.

3.24 The numbers (in thousands), by gender and marital status, of the 2004 U.S. population aged 15 and over are shown in the next table.

	Never Married	Married	Widowed	Divorced	Totals
Females	23,655.0	63,282.0	11,141.0	12,804.0	110,883.0
Males	29,561.0	62,483.0	2,641.0	8,956.0	103,641.0
Totals	53,216.0	125,765.0	13,782.0	21,760.0	214,524.0

Source: U.S. Census Bureau (2006).

Answer the following questions based on the information given in the preceding table.

a Find the conditional distribution of marital status for females.

b Find the conditional distribution of marital status for males.

c Is marital status independent of gender? Is it necessary to find both conditional distributions to answer this question? Justify your answers.

3.25 Use the table in Exercise 3.1 to do the following.

a Find the conditional distribution of speed limit, given a fatal crash occurred in a rural area.

b Find the conditional distribution of speed limit, given a fatal crash occurred in an urban area.

c Does the distribution of speed limit differ with respect to the land use of the area of a fatal crash? Justify your answer.

3.26 Using the data from Exercise 3.2, answer the following questions.

a Find the conditional distribution of education for those who have smoked a cigarette within the past year.

b Find the conditional distribution of education for those who have not smoked a cigarette within the past year.

c Is education independent of whether or not the person had smoked a cigarette within the past year? Justify your answer.

3.27 A proficiency examination for a certain skill was given to 100 employees of a firm. Forty of the employees were men. Sixty of the employees passed the examination (by scoring above a preset level for satisfactory performance.) The breakdown of test results among men and women is shown in the accompanying diagram.

	Male	Female	Total
Pass	24	36	60
Fail	16	24	40
Total	40	60	100

Suppose that an employee is selected at random from among the 100 who took the examination.

a Find the probability that the employee passed, given that he was a man.

b Find the probability that the employee was a man, given that a passing grade was received.

c Is the event of passing the exam independent of the event of gender? Justify your answer.

3.28 Two dice are rolled and the dots on the upper faces observed. Is the event of observing an 8 independent of the event of rolling doubles? Justify your answer.

3.29 A card is drawn from a standard deck of 52 cards. Is the event of drawing a jack independent of the event of drawing a heart? Justify your answer.

3.30 A cow without horns is said to be "polled." Horns are dominant to poll; that is, if a calf gets a horn gene from one parent and a poll gene from the other parent, the calf will grow horns. A horned bull had a mother who was poll.

a If the bull is bred to a poll cow, what is the probability that the calf will grow horns?

b If the bull is bred to a horned cow that also had a mother who was poll, what is the probability that the calf will grow horns?

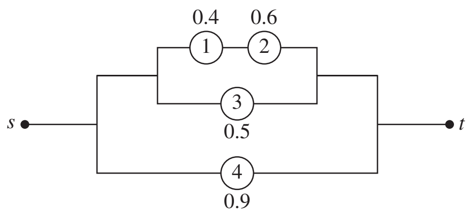
c If the bull is bred to a horned cow that had parents who were both homozygous for horns, what is the probability that the calf will grow horns?

3.31 A couple plans to have a child and wants to know what blood type the baby will have. One parent is heterozygous A, and the other is heterozygous B. What blood types are possible for the child, and what is the probability of each?

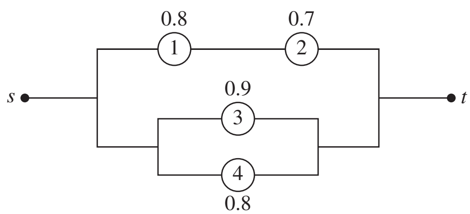
3.32 An electrical circuit has two relays in series. The probability that each relay closes when the switch is thrown is 0.9. Assuming that the relays operate independently of one another, what is the probability that current will flow through the circuit when the switch is thrown?

3.33 An electrical circuit has two relays in parallel. The probability that each relay closes when the switch is thrown is 0.9. If current flows through the circuit, what is the probability that both switches closed?

3.34 A portion of an electrical circuit is displayed next. The switches operate independently of each other, and the probability that each relay closes when the switch is thrown is displayed by the switch. What is the probability that current will flow from s to t when the switch is thrown?



3.35 A portion of an electrical circuit is displayed next. The switches operate independently of each other, and the probability that each relay closes when the switch is thrown is displayed by the switch. What is the probability that current will flow from s to t when the switch is thrown?



3.36 If A and B are independent events each with positive probability, prove that they cannot be mutually exclusive.

3.37 If A and B are independent events, show that $\bar{A}$ and $\bar{B}$ are independent.

3.38 Suppose A and B are independent events. For an event C such that $P(C) > 0$, prove that the event of A given C is independent of the event of B given C.

3.39 A box contains M balls, of which W are white. A sample of n balls is drawn at random and without replacement. Let A_j, where $j = 1, 2, \ldots, n$, denote the event that the ball drawn on the jth draw is white. Let B_k denote the event that the sample of balls contains exactly k white balls.

a Find the probability of A_j.

b Show that $P(A_j|B_k) = k/n$.

c Would the probability in part (b) change if the sampling was done with replacement?

3.40 In the game of craps, two dice are rolled. If the first roll is a 7 or an 11, the player wins. If the first roll is a 2, 3, or 12, the player loses. If any other outcome is observed on the first roll, the player wins if that outcome is rolled again before a 7 is rolled; otherwise, he loses. What is the probability of winning this game?

3.41 Sarah and Rachel play a series of games. The probability that Sarah wins a game is p, which is constant from game to game; the probability that Rachel wins a game is $1-p$. The outcome of each game is independent of the other games. Play stops when the total number of games won by one player is two greater than that of the other player and the player with the greater number of total wins is the series winner.

a Find the probability that a total of four games are played.

b Find the probability that Sarah wins the series.

3.42 Gambler's Ruin Problem. A gambler repeatedly plays a game for which his probabilities of winning a dollar is p and of losing a dollar is $1-p$ each time he plays. He plans to play the game repeatedly until he is ruined (loses all of his money) or has N dollars. If the gambler begins with k dollars, what is the probability that he is ruined?

3.3 Theorem of Total Probability and Bayes' Rule

Sometimes, it is possible to partition an event, say A, into the union of two or more mutually exclusive events. To partition the event A, we begin by partitioning the sample space S. Events $B_1, B_2, \ldots, B_k$ are said to partition a sample space S if the following two conditions are satisfied:

1 $B_iB_j = \phi$ for any pair i and j (Recall: ϕ denotes the null or impossible event.)

2 $B_1 \cup B_2 \cup \cdots \cup B_k = S$

For example, the set of tires in an auto assembly warehouse may be partitioned according to suppliers, or employees of a firm may be partitioned according to level of education. A partition for the case $k = 2$ is illustrated in Figure 3.6. Suppose we are interested in the probability of event A. The key idea with regard to a partition consists of observing that the event A can be written as the union of mutually exclusive events AB_1 and AB_2; that is,

$$A = AB_1 \cup AB_2.$$

And, thus,

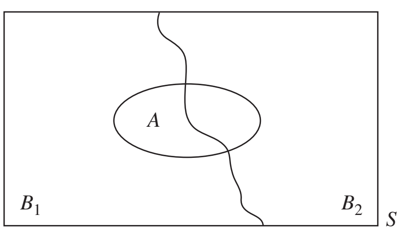
$$P(A) = P(AB_1) + P(AB_2).$$

If conditional probabilities $P(A|B_1)$ and $P(A|B_2)$ are known, then $P(A)$ can be found by writing

$$P(A) = P(B_1)P(A|B_1) + P(B_2)P(A|B_2).$$

This result is known as the Theorem of Total Probability and is restated in Theorem 3.2.

FIGURE **3.6**
Partition of S into B_1 and B_2.



THEOREM **3.2**

Theorem of Total Probability: If $B_1, B_2, \ldots, B_k$ is a collection of mutually exclusive and exhaustive events, then for any event A,

$$P(A) = \sum_{i=1}^{k} P(B_i)P(A|B_i).$$

■

EXAMPLE **3.8**

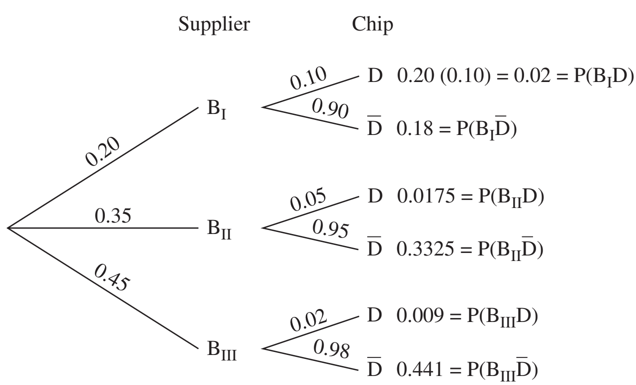
A company buys microchips from three suppliers—I, II, and III. Supplier I has a record of providing microchips that contain 10% defectives; Supplier II has a defective rate of 5%; and Supplier III has a defective rate of 2%. Suppose 20%, 35%, and 45% of the current supply came from Suppliers I, II, and III, respectively. If a microchip is selected at random from this supply, what is the probability that it is defective?

Solution

Let B_i denote the event that a microchip comes from Supplier i, where i = I, II, or III. (Notice that B_I, B_{II}, and B_{III} form a partition of the sample space for the experiment of selecting one microchip.) Let D denote the event that the selected microchip is defective. In problems of this type, it is best to write down everything that is known in symbols. Because we know what proportion of the microchips come from each supplier, we know $P(B_I) = 0.20$, $P(B_{II}) = 0.35$, and $P(B_{III}) = 0.45$. Given the supplier, we also know the probability that a randomly selected microchip is defective; that is, $P(D|B_I) = 0.10$, $P(D|B_{II}) = 0.05$, and $P(D|B_{III}) = 0.02$. This same information can be presented in a tree diagram, as in Figure 3.7.

Notice that in Figure 3.7 the probabilities for the second set of branches depend on the first branch with which they are associated; that is, they are conditional probabilities. These can be used to find the probability that a randomly selected part is from a specific supplier and that it is defective (or not defective). As an example, the probability that a randomly selected part is from Supplier I and defective may be

FIGURE **3.7**



found by multiplying $P(B_I) \times P(D|B_I)$, and this is just the product of the values on the uppermost branches.

Once we have identified the information provided either using symbols or a tree, we can easily find the probability that a randomly selected part is defective. Using symbols and the Law of Total Probability, we have

$$\begin{aligned} P(D) &= P(B_I)P(D|B_I) + P(B_{II})P(D|B_{II}) + P(B_{III})P(D|B_{III}) \\ &= 0.20(0.10) + 0.35(0.05) + 0.45(0.02) \\ &= 0.02 + 0.0175 + 0.009 = 0.0465. \end{aligned}$$

Alternatively, working with the tree diagram, we take the sum of the probabilities associated with branches terminating with a defective part: $0.02 + 0.0175 + 0.009 = 0.0465$. The approaches using symbols and the tree are equivalent and naturally lead to the same result. ▪

Suppose the events $B_1, B_2, \ldots, B_k$ partition the sample space S, and A is another event. In this setting, investigators frequently want to find probabilities of the form $P(B_i|A)$, which can be written as

$$\begin{aligned} P(B_i|A) &= \frac{P(B_iA)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^{k} P(B_k)P(A|B_k)}. \end{aligned}$$

This result is Bayes' rule, of which Theorem 3.3 is a general statement.

THEOREM **3.3**

Bayes' Rule. If the events $B_1, B_2, \ldots, B_k$ form a partition of the sample space S, and A is any event in S, then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{k} P(A|B_j)P(B_j)}.$$

Proof

From the definition of conditional probability and the multiplication theorem,

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{P(A)}.$$

The result follows by using the Theorem of Total Probability to rewrite the denominator. ▪

EXAMPLE 3.9 Consider again the information in Example 3.8. If a randomly selected microchip is defective, what is the probability that it came from Supplier B_{II}?

Solution Again, let D be the event that a microchip is defective, and let B_i be the event that the microchip came from Supplier i. In Example 3.8, we were given that $P(B_{II}) = 0.35$ and $P(D|B_{II}) = 0.05$, and we determined that $P(D) = 0.0465$. Then, by Bayes' rule, we have the probability that a randomly selected chip is from Supplier II and, given that it is defective, is

$$P(B_{II}|D) = \frac{P(D|B_{II})P(B_{II})}{P(D)} = \frac{0.05(0.35)}{0.0465} = 0.376.$$

■

Exercises

3.43 John flies frequently and likes to upgrade his seat to first class. He has determined that if he checks in for his flight at least 2 hours early, the probability that he will get the upgrade is 0.8; otherwise, the probability that he will get the upgrade is 0.3. With his busy schedule, he checks in at least 2 hours before his flight only 40% of the time. What is the probability that for a randomly selected trip John will be able to upgrade to first class?

3.44 A diagnostic test for a certain disease has 95% sensitivity and 95% specificity. Only 1% of the population has the disease in question. If the diagnostic test reports that a person chosen at random from the population tests positive, what is the conditional probability that the person does, in fact, have the disease? Are you surprised by the size of the answer? Do you consider this diagnostic test reliable?

3.45 Two methods, A and B, are available for teaching a certain industrial skill. The failure rate is 30% for method A and 10% for method B. Method B is more expensive, however, and hence is used only 20% of the time. (Method A is used the other 80% of the time.) A worker is taught the skill by one of the two methods, but he fails to learn it correctly. What is the probability that he was taught by using method A?

3.46 A mumps vaccine was licensed in the United States in 1967. The American Committee on Immunization Practices recommended routine vaccination of children 12 months and older in 1977. Most children receive an MMR (measles-mumps-rubella) vaccination at 15 months. In recent years, there have been mumps outbreaks on some high school and college campuses, leading some colleges to either highly recommend or require a second mumps vaccination before entering. The Centers for Disease Control and Prevention (CDC) report that a single mumps vaccine is 80% protective and two vaccines are 90% protective from mumps. (Eighty percent protective means that, if a vaccinated person is exposed to mumps, the probability of acquiring the disease is $1 - 0.80 = 0.20$.) On one university campus, all students have at least one vaccination. A second vaccination is strongly recommended but not required. Sixty percent of the students have the second vaccination. A mumps outbreak occurs on campus, and all students are exposed. What is the probability that a randomly selected student with mumps had the second vaccination?

3.47 In 2003, approximately 0.38% of the U.S. population had HIV/AIDS. Of these, it was estimated that 24.8% were not aware they have the disease. What is the probability that a randomly selected person who does not know whether or not he or she has the disease will actually have it? What assumption(s), if any, did you have to make?

3.48 "Pop," "soda," and "coke" are three terms that are used to refer to carbonated soda drinks. The frequency each is used varies across the United States according to the table that follows.

	Term Used for Carbonated Soda Drinks			
Region	**Pop**	**Soda**	**Coke**	**Total**
Pacific	0.15	0.71	0.14	1.0
Rocky Mountains	0.61	0.31	0.08	1.0
Southwest	0.12	0.23	0.65	1.0
Midwest	0.70	0.26	0.04	1.0
Northeast	0.30	0.68	0.02	1.0
Southeast	0.18	0.43	0.39	1.0

Source: http://www.popvssoda.com

At a large university, 4%, 10%, 6%, 18%, 28%, and 34% of the students are from the Pacific, Rocky Mountains, Southwest, Midwest, Northeast, and Southeast, respectively.

a What is the probability that a randomly selected student from this university is from the Pacific region and uses the term "soda"?

b What is the probability that a randomly selected student from this university uses "pop" when referring to carbonated beverages?

c A student is selected at random from among the students who use the term "coke." What is the probability that she or he is from the Southeast?

3.49 According to the U.S. Census Bureau, the 2005 poverty rate for households with female heads (no husband present) and children under age 6 was 48.3%; it was 24.0% for household with male heads (no wife present) and children under age 6; and it was 8.7% for married-couple homes with children under age 6. For households with children under age 6, 22.2% have female heads, 5.5% have male heads, and 72.3% are married-couple homes.

a What is the probability that a randomly selected household with a child under age 6 lived in poverty in 2005?

b What is the probability that a randomly selected household with a child under age 6 who was living in poverty in 2005 was a married-couple home?

c Is whether a household is in poverty or not independent of the type of head of household? Justify your answer.

3.50 During May 2006, the Gallup Organization took a poll of 1000 adults, aged 18 and older, in which they asked the following question: "Which comes closest to describing you?" Each individual was given these options for response: (1) You are convinced that God exists; (2) You think God probably exists, but you have a little doubt; (3) You think God probably exists, but you have a lot of doubt; (4) You think God probably does not exist, but you are not sure; or (5) You are convinced that God does not exist. Of those with a high school degree or less, 92% said they were "certain" or had "little doubt" that God exists. For individuals with some college, college graduates, and those with postgraduate education, the percentages of those who were "certain" or had a "little doubt" were 90%, 85%, and 77%, respectively. According to the U.S. Census Bureau, the percentages of people in the United States with no more than a high school degree, some college, college graduate, and postgraduate education are 49%, 29%, 15%, and 7%, respectively.

a What is the probability that a randomly selected adult from the U.S. population will be a college graduate and be "certain" or have "little doubt" that God exists?

b What is the probability that a randomly selected adult from the U.S. population will be "certain" or have "little doubt" that God exists?

c What is the probability that a randomly selected adult who is "certain" or has "little doubt" that God exists has no more than a high school education?

3.51 A single multiple-choice question has n choices, only one of which is correct. A student knows the answer with probability p. If the student does not know the answer, he or she guesses randomly. Find the conditional probability that the student knew the answer, given that the question was answered correctly.

3.52 After packing k boxes (numbered $1, 2, \ldots, k$) of m items each, workers discovered that one defective item had slipped in among the km items packed. In an attempt to find the defective item, they randomly sample n items from each box and examine these.

- **a** Find the probability that the defective item is in box i. What assumption is necessary for your answer to be valid?
- **b** Find the probability that the defective item is found in box 1, given that it was actually put in box 1.
- **c** Find the unconditional probability that the defective item is not found in box 1.
- **d** Find the conditional probability that the defective item is in box 1, given that it was not found in box 1.
- **e** Find the conditional probability that the defective item is in box 2, given that it was not found in box 1.
- **f** Comment on the behavior of these probabilities as $n \longrightarrow m$; as $n \longrightarrow 0$.

3.53 A genetic carrier is a person who carries one gene for a specified recessive trait. Because the person is heterozygous (has only one gene) for that trait, she does not express that trait but, when mated with another carrier, could produce offspring who do. Based on family history, June has a 50-50 chance of being a carrier of a rare genetic disease. If she is a carrier, each of her children has a 50-50 chance of having the disease. If June has had three children, none of whom have the disease, what is the probability that she is a carrier?

3.54 On the television show *Let's Make a Deal*, participants were shown three doors, each of which had a prize behind it. One of the prizes was very nice; the other two were not. A participant was asked to choose one of the doors. Then the game show host, Monty Hall, would open one of the other two doors, revealing one of the two poor prizes. The participant was then given the option of either keeping the door that was originally chosen or switching for the other closed door. Should the contestant switch? Justify your answer.

3.4 Odds, Odds Ratios, and Relative Risk

"What are the odds that our team will win today?" This is a common way of talking about events whose unknown outcomes have probabilistic interpretations. The *odds in favor* of an event A is the ratio of the probability of A to the probability of $\bar{A}$; that is,

$$\text{Odds in favor of } A = \frac{P(A)}{P(\bar{A})}.$$

The odds in favor of a balanced coin's coming up heads when flipped is $P(H)/P(T) = (1/2)/(1/2) = 1$, often written as 1:1 (one to one). Odds are not just a matter of betting and sports. They are a serious component of the analysis of frequency data, especially when researchers are comparing categorical variables in two-way frequency tables.

The Physicians' Health Study on the effects of aspirin on heart attacks randomly assigned over 22,000 male physicians to either the "aspirin" or the "placebo" arm of the study. The data on myocardial infarctions (MIs) are summarized in Table 3.5.

TABLE 3.5
Results of the Physicians' Health Study.

	MI	No MI	Total
Aspirin	139	10,898	11,037
Placebo	239	10,795	11,034
Total	378	21,683	22,071

Source: Steering Committee of the Physicians' Health Study Research Group (1989).
Note: Because studies have shown that people often respond positively to treatment, whether it is effective or not, a placebo treatment is generally given to the control group. A placebo treatment is one made to resemble the treatment of interest but having no active ingredient. For this study, a tablet appearing to be an aspirin, but containing no active ingredient, was given to participants assigned to the nontreatment or control group.

For the aspirin group, the odds in favor of suffering an MI are

$$\frac{P(\text{MI})}{P(\overline{\text{MI}})} = \frac{139/11{,}037}{10{,}898/11{,}037} = \frac{139}{10{,}898} = 0.013.$$

For the placebo group, the odds in favor of MI are

$$\frac{P(\text{MI})}{P(\overline{\text{MI}})} = \frac{239/11{,}034}{10{,}795/11{,}034} = \frac{239}{10{,}796} = 0.022.$$

The preceding results show that the odds of a heart attack with the placebo is considerably higher than the risk with aspirin. More specifically, the ratio of the two odds is called the *odds ratio*:

$$\begin{aligned}\text{Odds ratio of MI} &= \frac{\text{Odds of MI with aspirin}}{\text{Odds of MI without aspirin}}\\ &= \frac{0.013}{0.022}\\ &= 0.59\end{aligned}$$

Thus, the odds of suffering an MI for an individual in the aspirin group is 59% of the odds for an individual in the placebo group.

Odds ratios form a very useful single-number summary of the frequencies in a 2×2 (two-way) frequency table. In fact, the odds ratio has a simpler form for any 2×2 table, which can be written generically as

	Yes	No
A	a	b
B	c	d

The odds in favor of A are a/b, and the odds in favor of B are c/d. Therefore, the odds ratio is simply

$$\frac{a/b}{c/d} = \frac{ad}{bc}$$

which is the ratio of the products of the diagonal elements.

For a randomized clinical trial such as the Physicians' Health Study, the *relative risk* is more commonly used to compare groups than the odds ratio. The *relative risk* compares the probability (instead of the odds) of MI in each group; that is, the relative risk is the ratio of the probability of an event (MI) in the treatment (aspirin) group to the probability of an event in the placebo group:

$$\begin{aligned}\text{Relative risk of MI} &= \frac{P(\text{MI}\,|\,\text{aspirin})}{P(\text{MI}\,|\,\text{placebo})}\\ &= \frac{139/11{,}037}{239/11{,}034}\\ &= 0.58.\end{aligned}$$

That is, if a person takes a daily aspirin, the risk of having an MI is 58% of the MI risk of a person who is not taking a daily aspirin. The relative risk is easier to interpret and more consistent with how people think than the odds ratio. The relative risk cannot be computed for all studies. Here the two measures are very close to each other. Although this often occurs, the two may be quite different, especially for medium to large probabilities.

EXAMPLE **3.10** The Physicians' Health Study included only men, and the results clearly indicated that taking a low dose of aspirin reduced the risk of MI. In 2005 the results of the Women's Health Study, shown in Table 3.6, were published. This study randomized almost 40,000 women, ages 45 and older, to either aspirin or placebo and followed the women for 10 years.

TABLE **3.6** Results of the Women's Health Study.

	MI	No MI	Total
Aspirin	198	19,736	19,934
Placebo	193	19,749	19,942
Total	391	39,485	39,876

Source: Ridler et al. (2005).

1 Find the odds of MI for the aspirin group.

2 Find the odds of MI for the placebo (nonaspirin) group.

3 Find the odds ratio of MI for the aspirin and placebo groups.

4 Find the relative risk of MI for the aspirin and placebo groups.

Solution **1** The odds for the aspirin group is

$$\frac{P(\text{MI})}{P(\overline{\text{MI}})} = \frac{198/19{,}934}{19{,}736/19{,}934} = \frac{198}{19{,}736} = 0.010032.$$

2 The odds for the nonaspirin (placebo) group is

$$\frac{P(\text{MI})}{P(\overline{\text{MI}})} = \frac{193/19{,}942}{19{,}749/19{,}942} = \frac{193}{19{,}749} = 0.009977.$$

3 The odds ratio of MI for the aspirin group compared to the nonaspirin group is

$$\text{Odds ratio of MI} = \frac{\text{Odds of MI with aspirin}}{\text{Odds of MI without aspirin}}$$
$$= \frac{0.010332}{0.009977}$$
$$= 1.01.$$

4 The relative risk of MI for the aspirin group compared to the nonaspirin group is

$$\text{Relative risk of MI} = \frac{P(\text{MI} \mid \text{aspirin})}{P(\text{MI} \mid \text{without aspirin})}$$
$$= \frac{198/19{,}934}{193/19{,}942}$$
$$= 1.03.$$

When comparing two treatment groups, a relative risk of one indicates that there is no difference in the risks for the two groups. For this study, the estimated odds ratio is 1.01, and the estimated relative risk is 1.03, which are values close to one. In fact, the observed odds of MI are slightly higher for the aspirin group than for the placebo group. This result led to a report, published in 2005, indicating that a low-dose aspirin regime is not effective for reducing MI for women. ▪

Exercises

3.55 From the results of the Physicians' Health Study (discussed earlier in this section), an important factor in MIs seems to be the cholesterol level. The data in the accompanying table identify the number of MIs over the number in the cholesterol category for each arm of the study.

Cholesterol Level (mg per 100 mL)	Aspirin Group	Placebo Group
≤ 159	2/382	9/406
160–209	12/1587	37/1511
210–259	26/1435	43/1444
≥ 260	14/582	23/570

a Did the randomization in the study seem to do a good job of balancing the cholesterol levels between the two groups? Explain.

b Construct a 2×2 table of aspirin versus placebo MI response for each of the four cholesterol levels. Reduce the data in each table to the odds ratio.

c Compare the four odds ratios you found in part (b). Comment on the relationship between the effect of aspirin on heart attacks and the different cholesterol levels. Do you see why odds ratios are handy tools for summarizing data in a 2×2 table?

3.56 Is a defendant's race associated with his or her chance of receiving the death penalty? This controversial issue has been studied by many researchers. One important data set was collected on 326 cases in which the defendant was convicted of homicide. The death penalty was imposed on 36 of these cases. The accompanying table shows the defendant's race, the homicide victim's race, and whether or not the death penalty was imposed.

	White Defendant			Black Defendant	
	Death Penalty	No Death Penalty		Death Penalty	No Death Penalty
White Victim	19	132	White Victim	11	52
Black Victim	0	9	Black Victim	6	97

Source: Radelet (1981).

a Construct a single 2 × 2 table showing penalty versus defendant's race across all homicide victims. Calculate the odds ratio and interpret it.

b Decompose the table in part (a) into two 2 × 2 tables of penalty versus defendant's race, one for white homicide victims and one for black homicide victims. Calculate the odds ratio for each table and interpret each one.

c Do you see any inconsistency between the results of part (a) and the results of part (b)? Can you explain the apparent paradox?

3.57 Approximately 20% of adults become infected with human papillomavirus virus type 16 (HPV-16). Although most infections are benign, some progress to cervical cancer in women. A randomized clinical trial was conducted to determine whether the use of an HPV vaccine was effective for women (Koutsky et al., 2006). In this study 2392 women, aged 16 to 23, were randomly assigned to receive three doses of HPV-16 vaccine or three doses of placebo (no active ingredient). Some women had HPV-16 infections or other cervical abnormalities when they entered the study; others developed the infection before they received all three shots. These women (859 in all) were excluded when the researchers calculated the vaccine's effectiveness. Of the remaining 1533 women, 768 received the HPV-16 vaccine and 761 received the placebo. The women were followed for a median 17.4 months after receiving the third shot. During the study, 41 developed HPV-16 infection; all were in the placebo group.

a From the information given, construct a table displaying the information given about the placebo and vaccine groups.

b Find the odds of HPV-16 infection for the placebo group.

c Find the odds of HPV-16 infection for the vaccine group.

d Compute the odds ratio.

3.58 The *Titanic* was a large luxury ocean liner that was declared to be an "unsinkable ship." During its maiden voyage across the Atlantic Ocean, it hit an iceberg and sank on April 14, 1912. Large numbers of people lost their lives. The economic status of the passengers has been roughly grouped according to whether they were traveling first class, second class, or third class. The crew has been reported separately. Although the exact numbers are still a matter of debate, one report (Dawson 1995) of the numbers of those who did and did not survive, by economic status and gender, is displayed in the table that follows.

	Population Exposed to Risk		Number of Deaths	
Economic Status	Male	Female	Male	Female
First Class	180	145	118	4
Second Class	179	106	154	13
Third Class	510	196	422	106
Crew	862	23	670	3

a Find the odds of a male in first class dying on the *Titanic*.
b Find the odds of a male in third class dying on the *Titanic*.
c Find the odds ratio of a male in first class relative to a male in third class dying on the *Titanic*.
d Find the odds of a female in first class dying on the *Titanic*.
e Find the odds ratio of a male in first class to a female in the same class dying on the *Titanic*.

3.5 Summary

In conducting an experiment, knowledge about whether or not one event has occurred may provide information on whether or not a second event has occurred. If so, the two events are dependent. Conditional probability is important when modeling dependent events. If the occurrence of the first event provides no information on whether or not the second event occurs, the two are independent. By taking advantage of the Theorem of Total Probability, Bayes' rule is useful for computing some conditional probabilities.

Supplementary Exercises

3.59 Volunteers provide important services to the community. For September 2004, the numbers of males and females in the U.S. population and the numbers of those who volunteer by age category are displayed in the next table.

	Males		Females	
Age (in years)	**Population**	**Volunteers**	**Population**	**Volunteers**
16 to 19	8,245	2,072	8,001	2,702
20 to 24	10,146	1,727	10,084	2,320
25 to 34	19,383	3,956	19,593	6,090
35 to 44	21,232	6,068	21,936	8,714
45 to 54	20,255	5,917	21,187	7,667
55 to 64	14,033	3,869	15,173	4,915
65 and over	14,727	3,402	19,946	5,122
Total	108,021	27,011	115,920	37,530

Source: U.S. Census Bureau.

a What is the probability that a randomly selected person, aged at least 16 years, is *not* a volunteer?
b What is the probability that a randomly selected male, aged at least 16 years, is a volunteer?
c What is the probability that a randomly selected volunteer, aged at least 16 years, is a male?
d What is the probability that a randomly selected female, aged at least 65 years, is a volunteer?

3.60 Consider again the table about the numbers of volunteers in Exercise 3.59. Use that information to answer the following.

a Find the conditional distribution of age for the female volunteers.
b Find the conditional distribution of age for the male volunteers.
c Do you think the age of volunteers is independent of gender? Justify your answer.

3.61 Tobacco use is considered to be the leading cause of preventable death and disease in the United States. The National Survey on Drug Use and Health provides information on drug use in the United States. Following is a table showing the numbers of persons, by age category and gender, who have and have not smoked a cigarette within the past month.

Numbers (in Thousands) Who Have and Have Not Smoked a Cigarette Within the Past Month, by Age and Gender (Numbers in Thousands)

	Female		Male	
Age Category	**Tobacco User**	**Non-Tobacco**	**Tobacco User**	**Non-Tobacco**
12–17	1,545	10,793	1,453	11,423
18–25	5,690	10,331	7,041	9,131
26 or Older	21,453	74,220	23,784	63,651
Total	28,688	95,344	32,278	84,205

Source: Substance Abuse & Mental Health Services Administration, Office of Applied Studies (2005).

a What is the probability that a randomly selected person in the United States smoked a cigarette during the past month?

b What is the probability that a randomly selected person aged 12 to 17 smoked a cigarette during the past month?

c What is the probability that a randomly selected male smoked a cigarette within the past month?

d What is the probability that a randomly selected female smoked a cigarette in the past month?

3.62 In a survey, the primary question of interest is, "Have you ever left the scene of an accident?" A randomized response design is used but instead of flipping a coin as in Exercise 3.22, the respondent is told to roll two dice. If the sum of the two dice is 2, 3, or 4, the respondent is to answer "yes." If the sum is 5, 6, 7, 8, 9, or 10, the respondent is to answer according to the truth. If the sum is 11 or 12, the respondent is to answer "no." Let p be the proportion in the population who has left the scene of an accident. The proportion responding "yes" is 0.4. What is an estimate of the proportion of people in the population who have left the scene of an accident?

3.63 Use the data on cigarette smoking by age and gender in Exercise 3.61 to answer the following.

a Find the distribution of age for females who had smoked a cigarette during the past month.

b Find the distribution of age for males who had smoked a cigarette in the past month.

c For females, compare the conditional distributions of age for cigarette and noncigarette smokers.

d For males, compare the conditional distributions of age for cigarette and noncigarette smokers.

e For males, is age independent of whether or not a person has smoked a cigarette in the past month? Are they independent for females? Justify your answers.

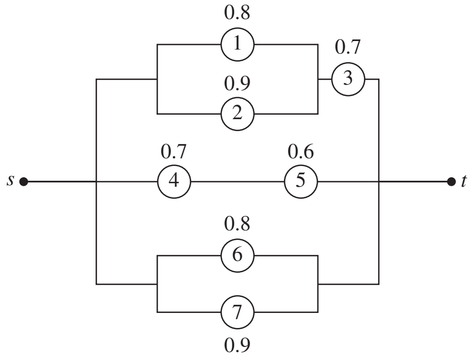
3.64 Suppose that the probability of exposure to the flu during an epidemic is 0.7. For adults under age 65, the effectiveness of the vaccine being used during that flu season is 80%; that is, if a vaccinated adult under the age of 65 is exposed to the flu, the probability of *not* catching the flu is 0.80. If an unvaccinated adult less than 65 years old is exposed to the flu, the probability is 0.90 that the individual acquires the flu. Two persons—one inoculated and one not—can perform a highly specialized task in a business. Assume that they are not at the same location, are not in contact with the same people, and cannot give each other the flu. What is the probability that at least one will get the flu?

3.65 Two gamblers bet \$1 each on successive flips of a coin. Each has a bankroll of \$6.

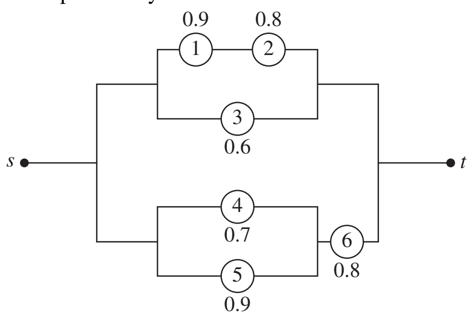
a What is the probability that they break even after six flips of the coin?

b What is the probability that one particular player (say, Jones) wins all the money on the eighth flip of the coin?

3.66 A portion of an electrical circuit is displayed next. The switches operate independently of each other, and the probability that each relay closes when the switch is thrown is displayed by the switch. What is the probability that current will flow from s to t when the switch is thrown?



3.67 A portion of an electrical circuit is displayed next. The switches operate independently of each other, and the probability that each relay closes when the switch is thrown is displayed by the switch. What is the probability that current will flow from s to t when the switch is thrown?



3.68 An accident victim will die unless he receives, in the next 10 minutes, a transfusion of type A Rh+ blood, which can be supplied by a single donor. The medical team requires 2 minutes to type a prospective donor's blood and 2 minutes more to complete the transfer of blood. A large number of untyped donors are available, 40% of whom have type A Rh+ blood. What is the probability that the accident victim will be saved, if only one blood-typing kit and numerous donors are available?

3.69 Suppose that the streets of a city are laid out in a grid, with streets running north-south and east-west. Consider the following scheme for patrolling an area of 16 blocks by 16 blocks. A patrolman commences walking at the intersection in the center of the area. At the corner of each block, he randomly elects to go north, south, east, or west.

a What is the probability that he will reach the boundary of his patrol area by the time he walks the first eight blocks?

b What is the probability that he will return to the starting point after walking exactly four blocks?

3.70 Prostate-specific antigen (PSA) is the most commonly used marker for the detection of prostate cancer. Its sensitivity is 0.80, and its specificity is 0.59. About 0.07% of the males in the United States are diagnosed with prostate cancer each year.

a What is the probability that someone with prostate cancer will have a negative result (indicating no cancer) when tested using PSA?

b What is the probability that someone without prostate cancer will have a positive result (indicating cancer) when tested using PSA?

3.71 Refer to Exercise 3.70. Suppose 100,000 men who have not previously been diagnosed with prostate cancer are tested for the disease using PSA.

a How many would you expect to have a true positive test?

b How many would you expect to have a true negative test?

c What is the predictive value of the test? (Hint: Construct a 2×2 table as in Table 3.5).

3.72 The *Titanic* was first discussed in Exercise 3.58. Following is a table showing the relationship between economic status and age with respect to the survivors from the *Titanic*.

	Population Exposed to Risk		Number of Deaths	
Economic Status	**Adult**	**Child**	**Adult**	**Child**
First and Second Class	580	30	289	0
Third Class	627	79	476	52
Crew	885	0	673	0

a Find the odds of an adult in first or second class perishing on the *Titanic.*

b Find the odds of an adult in third class perishing on the *Titanic*.

c Find the odds ratio of adults in first or second class to those in third class perishing on the *Titanic*.

d Find the odds of a child in third class perishing on the *Titanic*.

e Find the odds ratio of adults in first or second class to children in third class perishing on the *Titanic*.

3.73 Consider two mutually exclusive events A and B such that $P(A) > 0$ and $P(B) > 0$. Are A and B independent? Give a proof for your answer.

3.74 Show that for three events, A, B, and C,

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

3.75 If A and B are independent events, show that A and $\bar{B}$ are also independent.

3.76 Three events, A, B, and C, are said to be independent if the following equalities hold:

$$P(AB) = P(A)P(B)$$

$$P(AC) = P(A)P(C)$$

$$P(BC) = P(B)P(C)$$

$$P(ABC) = P(A)P(B)P(C)$$

Suppose that a balanced coin is independently tossed two times. Events A, B, and C are defined as follows:

A: Heads comes up on the first toss.

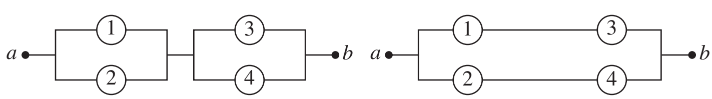
B: Heads comes up on the second toss.

C: Both tosses yield the same outcome.

Are A, B, and C independent? Justify your answer.

3.77 A line from a to b has midpoint c. A point is chosen at random on the line and marked x. (The fact that the point x was chosen at random implies that x is equally likely to fall in any subinterval of fixed length l.) Find the probability that the line segments ax, bx, and ac can be joined to form a triangle.

3.78 Relays in a section of an electrical circuit operate independently, and each one has a probability of 0.8 of closing properly when a switch is thrown. The following two designs, each involving four relays, are presented for a section of a new circuit. Which design has the higher probability of permitting current to flow from a to b when the switch is thrown?



A B

REFERENCES

Dawson, R. J. MacG. 1995. The “Unusual Episode” data revisited. *Journal of Statistics Education*, 3:3.

Hadgu, A. 1999. Discrepant analysis: A biased and an unscientific method for estimating test sensitivity and specificity. *Journal of Clinical Epidemiology* 52:1231–37.

Heiter, B. J., and P. P. Bourbeau. 1995. Comparison of two rapid streptococcal antigen detection assays with culture for diagnosis of streptococcal pharyngitis. *Journal of Clinical Microbiology* 33:1408–10.

Koutsky, L. A., K. A. Ault, C. M. Wheeler, D. R. Brown, E. Barr, F. B. Alvarez, L. M. Chiacchierini, and K. U. Jansen. 2006. A controlled trial of a human papillomavirus type 16 vaccine. *New England Journal of Medicine* 347:1645–51.

Radelet, M. 1981. Racial characteristics and imposition of the death penalty. *American Sociological Review* 46:918–27.

Ridler, P. M., N. R. Cook, I-M. Lee, D. Gordan, J. M. Gaziano, J. E. Manson, C. H. Hennekens, and J. E. Buring. 2005. A randomized trial of low-dose aspirin in the primary prevention of cardiovascular disease in women. *New England Journal of Medicine* 352:1293–1304.

Substance Abuse & Mental Health Services Administration, Office of Applied Studies. 2005. *Results from the 2004 National Survey on Drug Use and Health: Detailed Tables*. Rockville, Maryland: SAMSHA.

Steering Committee of the Physicians’ Health Study Research Group. 1989. Final report on the aspirin component of the ongoing Physicians’ Health Study. *New England Journal of Medicine* 321:129–35.

U.S. Census Bureau. 2006. *2006 Statistical abstract of the United States*. Washington, D.C.: U.S. Census Bureau.

U.S. Department of Transportation. 2005. *Traffic safety facts 2004*. National Highway Traffic Administration National Center for Statistics and Analysis. DOT HS 809 919. Washington, DC: U.S. Department of Transportation.

4

Discrete Probability Distributions

4.1 Random Variables and Their Probability Distributions

Most of the experiments we encounter generate outcomes that can be interpreted in terms of real numbers, such as heights of children, numbers of voters favoring various candidates, tensile strength of wires, and numbers of accidents at specified intersections. These numerical outcomes, whose values can change from experiment to experiment, are called *random variables*. We will look at an illustrative example of a random variable before we attempt a more formal definition.

A section of an electrical circuit has two relays, numbered 1 and 2, operating in parallel. The current will flow when a switch is thrown if either one or both of the relays close. The probability that a relay will close properly is 0.8, and the probability is the same for each relay. We assume that the relays operate independently. Let E_i denote the event that relay i closes properly when the switch is thrown. Then $P(E_i) = 0.8$.

When the switch is thrown, a numerical outcome of some interest to the operator of this system is X, which is the number of relays that close properly. Now, X can take on only three possible values, because the number of relays that close must be 0, 1, or 2. We can find the probabilities associated with these values of X by relating them to the underlying events E_i. Thus, we have

$$\begin{aligned} P(X = 0) &= P(\bar{E}_1\bar{E}_2) \\ &= P(\bar{E}_1)P(\bar{E}_2) \\ &= 0.2(0.2) \\ &= 0.04 \end{aligned}$$

because $X = 0$ means that neither relay closes and the relays operate independently. Similarly,

$$\begin{aligned} P(X = 1) &= P(E_1\bar{E}_2 \cup \bar{E}_1E_2) \\ &= P(E_1\bar{E}_2) + P(\bar{E}_1E_2) \\ &= P(E_1)P(\bar{E}_2) + P(\bar{E}_1)P(E_2) \\ &= 0.8(0.2) + 0.2(0.8) \\ &= 0.32 \end{aligned}$$

and

$$\begin{aligned} P(X = 2) &= P(E_1E_2) \\ &= P(E_1)P(E_2) \\ &= 0.8(0.8) \\ &= 0.64. \end{aligned}$$

The values of X, along with their probabilities, are more useful for keeping track of the operation of this system than are the underlying events E_i, because the number of properly closing relays is the key to whether the system will work. The current will flow if X is equal to at least 1, and this event has probability

$$\begin{aligned} P(X \geq 1) &= P(X = 1 \text{ or } X = 2) \\ &= P(X = 1) + P(X = 2) \\ &= 0.32 + 0.64 \\ &= 0.96. \end{aligned}$$

Notice that we have mapped the outcomes of an experiment into a set of three meaningful real numbers and have attached a probability to each. Such situations provide the motivation for Definitions 4.1 and 4.2.

DEFINITION **4.1**

A **random variable** is a real-valued function whose domain is a sample space. ▪

Random variables are denoted by uppercase letters, usually toward the end of the alphabet, such as X, Y, and Z. The actual values that random variables can assume are denoted by lowercase letters, such as x, y, and z. We can then talk about the "probability that X takes on the value x," $P(X = x)$, which is denoted by $p(x)$.

In the relay example, the random variable X has only three possible values, and it is a relatively simple matter to assign probabilities to these values. Such a random variable is called *discrete*.

DEFINITION **4.2**

A random variable X is said to be **discrete** if it can take on only a finite number—or a countably infinite number—of possible values x. The **probability function** of X, denoted by $p(x)$, assigns probability to each value x of X so that the following conditions are satisfied:

1 $P(X = x) = p(x) \geq 0$.

2 $\sum_x P(X = x) = 1$, where the sum is over all possible values of x.

■

The probability function is sometimes called the **probability mass function** of X to denote the idea that a mass of probability is associated with values for discrete points.

It is often convenient to list the probabilities for a discrete random variable in a table. With X defined as the number of closed relays in the problem just discussed, the table is as follows:

x	$p(x)$
0	0.04
1	0.32
2	0.64
Total	1.00

This listing constitutes one way of representing the probability distribution of X.

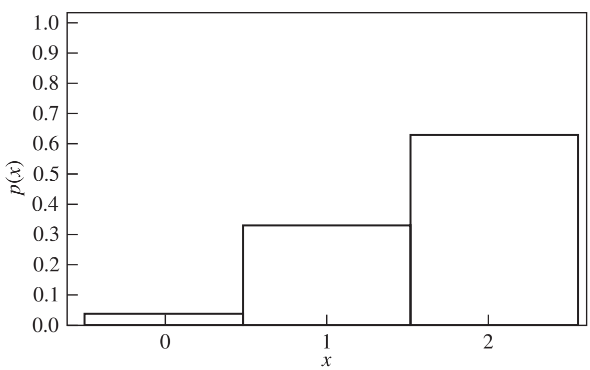
Notice that the probability function $p(x)$ satisfies two properties:

1 $0 \leq p(x) \leq 1$ for any x.

2 $\sum_x p(x) = 1$, where the sum is over all possible values of x.

In general, a function is a probability function if and only if the two preceding conditions are satisfied. Bar graphs are used to display the probability functions for discrete random variables. The probability distribution of the number of closed relays discussed earlier is shown in Figure 4.1.

FIGURE **4.1**
Graph of a probability mass function.



Functional forms for some probability functions that have been useful for modeling real-life data are given in later sections. We now illustrate another method for arriving at a tabular presentation of a discrete probability distribution.

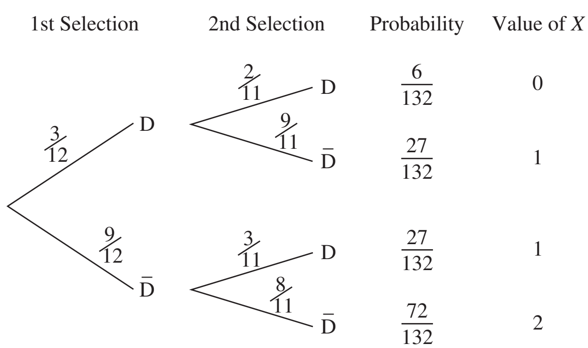
EXAMPLE 4.1 A local video store periodically puts its used movies in a bin and offers to sell them to customers at a reduced price. Twelve copies of a popular movie have just been added to the bin, but three of these are defective. A customer randomly selects two of the copies for gifts. Let X be the number of defective movies the customer purchased. Find the probability function for X and graph the function.

Solution The experiment consists of two selections, each of which can result in one of two outcomes. Let D_i denote the event that the ith movie selected is defective; thus, $\bar{D}_i$ denotes the event that it is good. The probability of selecting two good movies ($X = 0$) is

$$P(\bar{D}_1\bar{D}_2) = P(\bar{D} \text{ on 1st})P(\bar{D} \text{ on 2nd} \mid \bar{D} \text{ on 1st}).$$

The multiplicative law of probability is used, and the probability for the second selection depends on what happened on the first selection. Other possibilities for outcomes will result in other values of X. These outcomes are conveniently listed on the tree in Figure 4.2. The probabilities for the various selections are given on the branches of the tree.

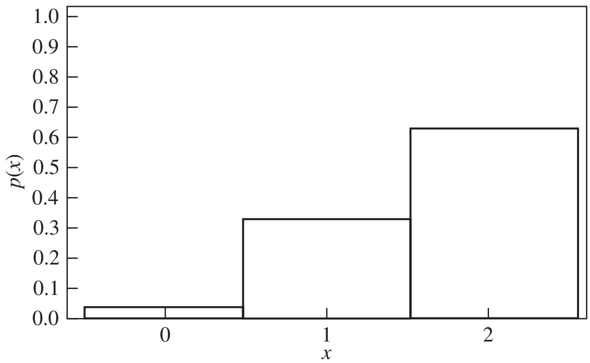
FIGURE 4.2
Outcomes for Example 4.1.



Clearly, X has three possible outcomes with probabilities as follows:

x	$p(x)$
0	$\frac{6}{132}$
1	$\frac{54}{132}$
2	$\frac{72}{132}$
Total	1.00

The probabilities are graphed in the figure that follows.



Try to envision this concept extended to more selections from the bin with various numbers of defectives. ▪

We sometimes study the behavior of random variables by looking at the *cumulative* probabilities; that is, for any random variable X, we may look at $P(X \leq b)$ for any real number b. This is the cumulative probability for X evaluated at b. Thus, we can define a function $F(b)$ as

$$F(b) = P(X \leq b).$$

DEFINITION **4.3**

The **distribution function** $F(b)$ for a random variable X is

$$F(b) = P(X \leq b)$$

if X is discrete,

$$F(b) = \sum_{x=-\infty}^{b} p(x)$$

where $p(x)$ is the probability function.

The distribution function is often called the **cumulative distribution function** (CDF). ▪

The random variable X, denoting the number of relays that close properly (as defined at the beginning of this section), has a probability distribution given by

$$P(X = 0) = 0.04$$

$$P(X = 1) = 0.32$$

$$P(X = 2) = 0.64.$$

Because positive probability is associated only for $x = 0,\ 1$, or 2, the distribution function changes values only at those points. For values of b at least 1, but less than 2, the $P(X \le b) = P(X \le 1)$. For example, we can see that

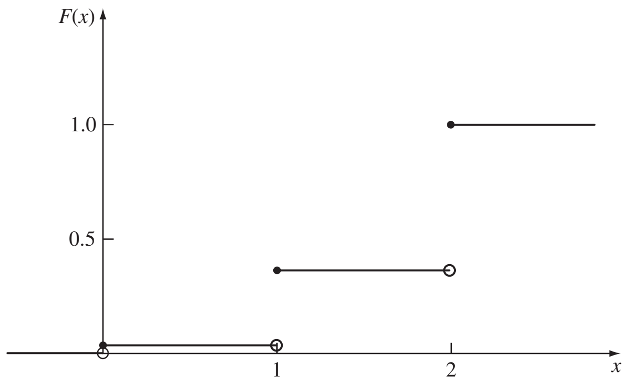
$$P(X \le 1.5) = P(X \le 1.9) = P(X \le 1) = 0.36.$$

The distribution function for this random variable then has the form

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.04, & 0 \le x < 1 \\ 0.36, & 1 \le x < 2 \\ 1, & x \ge 2 \end{cases}$$

The function is graphed in Figure 4.3.

FIGURE 4.3
Distribution function.



Notice that the distribution function is a step function and is defined for all real numbers. This is true for all discrete random variables. The distribution function is discontinuous at points of positive probability. Because the outcomes 0, 1, and 2 have positive probability associated with them, the distribution function is discontinuous at those points. The change in the value of the function at a point (the height of the step) is the probability associated with that value x. Because the outcome of 2 is the most probable ($p(2) = 0.64$), the height of the "step" at this point is the largest. Although the function has points of discontinuity, it is right-hand continuous at all points. To see this, consider $X = 1$. As we approach 1 from the right, we have $\lim_{h \to 0^+} F(1 + h) = 0.36 = F(1)$; that is, the distribution function F is right-hand continuous. However, if we approach 1 from the left, we find $\lim_{h \to 0^-} F(1 + h) = 0.04 \ne 0.36 = F(1)$, giving us that F is not left-hand continuous. Because a function must be both left- and right-hand continuous to be continuous, F is not continuous at $X = 1$.

In general, a distribution function is defined for the whole real line. Every distribution function must satisfy four properties; similarly, any function satisfying the following four properties is a distribution function.

1 $\lim_{x \to -\infty} F(x) = 0$

2 $\lim_{x \to \infty} F(x) = 1$

3 The distribution function is a nondecreasing function; that is, if $a < b$, $F(a) \leq F(b)$. The distribution function can remain constant, but it cannot decrease as we increase from a to b.

4 The distribution function is right-hand continuous; that is, $\lim_{h \to 0^+} F(x+h) = F(x)$.

We have already seen that given a probability mass function we can determine the distribution function. For any distribution function, we can also determine the probability function.

EXAMPLE 4.2 A large university uses some of the student fees to offer free use of its health center to all students. Let X be the number of times that a randomly selected student visits the health center during a semester. Based on historical data, the distribution function of X is given next.

$$F(x) = \begin{cases} 0, & x < 0 \\ 0.6, & 0 \leq x < 1 \\ 0.8, & 1 \leq x < 2 \\ 0.95, & 2 \leq x < 3 \\ 1, & x \geq 3 \end{cases}$$

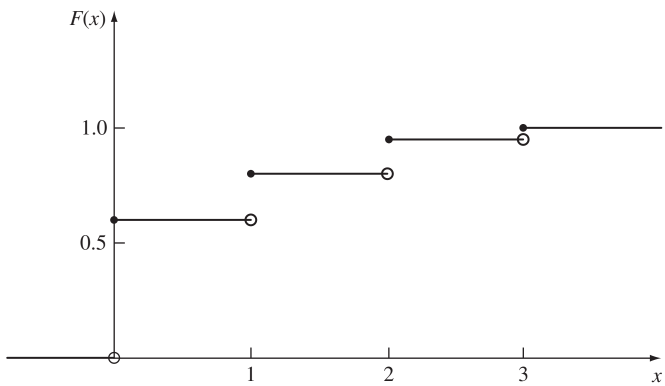
For the preceding function,

1 Graph F.

2 Verify that F is a distribution function.

3 Find the probability function associated with F.

Solution 1



2 To verify that F is a distribution function, we must confirm that the function satisfies the four conditions of a distribution function.

1 Because F is zero for all values x less than 0, $\lim_{x \to -\infty} F(x) = 0$.

2 Similarly F is one for all values of x that are 3 or greater; therefore, $\lim_{x \to +\infty} F(x) = 1$.

3 The function F is nondecreasing. There are many points for which it is not increasing, but as x increases, $F(x)$ either remains constant or increases; it never decreases.

4 The function is discontinuous at four points: 0, 1, 2, and 3. At each of these points, F is right-hand continuous. As an example, for $X = 2$, $\lim_{h \to 0^+} F(2 + h) = 0.95 = F(2)$.

Because F satisfies the four conditions, it is a distribution function.

3 The points of positive probability occur at the points of discontinuity: 0, 1, 2, and 3. Further, the probability is the height of the "jump" at that point. This gives us the following probabilities.

x	$p(x)$
0	$0.6 - 0 = 0.6$
1	$0.8 - 0.6 = 0.2$
2	$0.95 - 0.8 = 0.15$
3	$1 - 0.95 = 0.05$

■

Exercises

4.1 Circuit boards from two assembly lines set up to produce identical boards are mixed in one storage tray. As inspectors examine the boards, they find that it is difficult to determine whether a board comes from line A or line B. A probabilistic assessment of this question is often helpful. Suppose that the storage tray contains 10 circuit boards of which six came from line A and four from line B. An inspector selects two of these identical-looking boards for inspection. He is interested in X, the number of inspected boards from line A.

a Find the probability function for X.
b Graph the probability function of X.
c Find the distribution function of X.
d Graph the distribution function of X.

4.2 Among 12 applicants for an open position, 7 are women and 5 are men. Suppose that three applicants are randomly selected from the applicant pool for final interviews. Let X be the number of female applicants among the final three.

a Find the probability function for X.
b Graph the probability function of X.
c Find the distribution function of X.
d Graph the distribution function of X.

4.3 The median annual income for heads of households in a certain city is \$44,000. Four such heads of household are randomly selected for inclusion in an opinion poll. Let X be the number (out of the four) who have annual incomes below \$44,000.

a Find the probability distribution of X.
b Graph the probability distribution of X.

c Find the distribution function of X.

d Graph the distribution function of X.

e Is it unusual to see all four below \$44,000 in this type of poll? (What is the probability of this event?)

4.4 In 2005, Derrek Lee led the National Baseball League with a 0.335 batting average, meaning that he got a hit on 33.5% of his official times at bat. In a typical game, he had three official at bats.

a Find the probability distribution for X, the number of hits Lee got in a typical game.

b What assumptions are involved in the answer? Are the assumptions reasonable?

c Is it unusual for a good hitter to go 0 for 3 in one game?

4.5 A commercial building has four entrances, numbered I, II, III, and IV. Three people enter the building at 9:00 a.m. Let X denote the number of people who select entrance I. Assuming that the people choose entrances independently and at random, find the probability distribution for X. Were any additional assumptions necessary for your answer?

4.6 In 2002, 33.9% of all fires were structure fires. Of these, 78% were residential fires. The causes of structure fire and the numbers of fires during 2002 for each cause are displayed in the next table. Suppose that four independent structure fires are reported in 1 day, and let X denote the number, out of the four, that are caused by cooking.

Cause of Fire	Number of Fires
Cooking	29,706
Chimney Fires	8,638
Incinerator	284
Fuel Burner	3,226
Commercial Compactor	246
Trash/Rubbish	9,906
Total	52,006

a Find the probability distribution for X in tabular form.

b Find the probability that at least one of the four fires was caused by cooking.

4.7 Daily sales records for a car dealership show that it will sell 0, 1, 2, or 3 cars, with probabilities as listed:

Number of Sales	0	1	2	3
Probability	0.5	0.3	0.15	0.05

a Find the probability distribution for X, the number of sales in a 2-day period, assuming that the sales are independent from day to day.

b Find the probability that two or more sales are made in the next 2 days.

4.8 Four microchips, two of which are defective, are to be placed in a computer. Two of the four chips are randomly selected for inspection before the computer is assembled. Let X denote the number of defective chips found among the two inspected.

a Find the probability distribution for X.

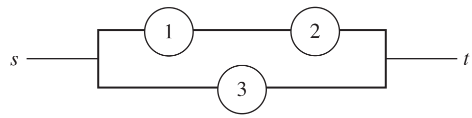
b Find the probability that no more than one of the two inspected chips was defective.

4.9 Of the people who enter a blood bank to donate blood, 1 in 3 has type O^+ blood, and 1 in 20 has type O^- blood. For the next three people entering the blood bank, let X denote the number with O^+ blood, and let Y denote the number with O^- blood. Assume the independence among the people with respect to blood type.

a Find the probability distributions of X and of Y.

b Find the probability distribution of $X + Y$, which is the number of people with type O blood.

4.10 When turned on, each of the three switches in the accompanying diagram works properly with probability 0.9. If a switch is working properly, current can flow through it when it is turned on. Find the probability distribution for X, which is the number of closed paths from s to t, when all three switches are turned on.



4.11 At a miniature golf course, players record the strokes required to make each hole. If the ball is not in the hole after five strokes, the player is to pick up the ball and record six strokes. The owner is concerned about the flow of players at hole 7. (She thinks that players tend to get backed up at that hole.) She has determined that the distribution function of X, the number of strokes a player takes to complete hole 7 to be the following:

$$F(x) = \begin{cases} 0, & x < 1 \\ 0.05, & 1 \le x < 2 \\ 0.15, & 2 \le x < 3 \\ 0.35, & 3 \le x < 4 \\ 0.65, & 4 \le x < 5 \\ 0.85, & 5 \le x < 6 \\ 1, & x \ge 6 \end{cases}$$

a Graph the distribution function of X.
b Find the probability function of X.
c Graph the probability function of X.
d Based on (a) through (c), are the owner's concerns substantiated? Justify your answer.

4.12 Observers have noticed that the distribution function of X, which is the number of commercial vehicles that cross a certain toll bridge during a minute, is as follows:

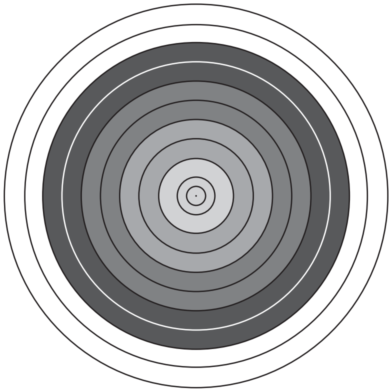
$$F(x) = \begin{cases} 0, & x < 0 \\ 0.20, & 0 \le x < 1 \\ 0.50, & 1 \le x < 2 \\ 0.85, & 2 \le x < 4 \\ 1, & x \ge 4 \end{cases}$$

a Graph the distribution function of X.
b Find the probability function of X.
c Graph the probability function of X.

4.13 A woman has seven keys on a key ring, one of which fits the door she wants to unlock. She randomly selects a key and tries it. If it does not unlock the door, she randomly selects another key from those remaining and tries to unlock the door with it. She continues in this manner until the door is unlocked. Let X be the number of keys she tries before unlocking the door, counting the key that actually worked. Find the probability function of X.

4.14 An absent-minded secretary prepares an envelope for each of five letters to different people. He then randomly puts the letters into the envelopes. Find the probability function for the number of letters that are placed in the correct envelope.

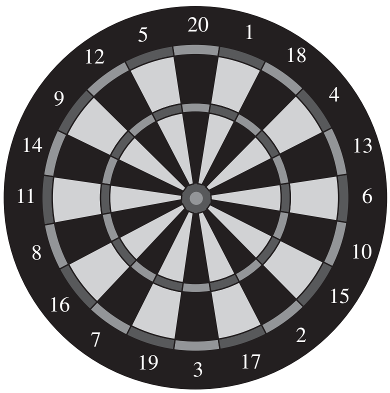
4.15 In archery, a common target is as shown in the picture to the right. Targets are marked with 10 evenly spaced concentric rings, which have score values from 1 through 10 assigned to them. In addition, there is an inner 10 ring, sometimes called the X ring. This becomes the 10 ring at indoor compound competitions. Outdoors it serves as a tiebreaker, with the archer scoring the most Xs winning. Suppose we have an outdoor competition using an Olympic size target, which has a diameter of 122 cm. When a particular archer shoots an arrow, it is equally likely to fall anywhere on the target, but it will hit the target.



a Find the probability function of the score obtained from the archer shooting one arrow at the target.

b If a target with a 40 cm diameter is used and the same assumptions hold, how does the probability function change?

4.16 A group of children is learning to play darts. Keeping score has been a challenge so they decided to modify the rules for scoring. As is standard in darts, the red bull's eye in the center of the target is worth 50 points, and the small green ring surrounding the bull's eye is worth 25 points. A dart falling in the white pie-shaped regions is scored 5 points, and one landing in the black pie-shaped regions is worth 10 points. The score is doubled if the dart falls in the outer small ring, and it is tripled if it falls in the inner small ring. No points are scored if the dart falls outside the outer small ring. The standard dimensions for a dartboard are recorded in the table that follows. Assume that the wire used to form the sections has no width.



When one of the children throws a dart, it always hits the board, but the dart is equally likely to hit on any point on the board. Under the modified scoring system, find the probability function for the score obtained from a single throw of the dart.

Double and triple ring inside width dimensions	8 mm
Bull inside diameter	12.7 mm
Semicenter inside diameter	31.8 mm
Outside edge of double wire to center bull	170 mm
Outside edge of treble wire to center bull	107 mm
Outside edge of double wire to outside edge of double wire	340 mm
Overall dartboard diameter	451 mm

Source: http://www.dartscanada.com

4.2 Expected Values of Random Variables

Because a probability can be thought of as the long-run relative frequency of occurrence for an event, a probability distribution can be interpreted as showing the long-run relative frequency of occurrences for numerical outcomes associated with an experiment. Suppose, for example, that you and a friend are matching balanced coins. Each of you flips a coin. If the upper faces match, you win \$1; if they do not match, you lose \$1 (your friend wins \$1). The probability of a match is 0.5 and in the long run you should win about half of the time. Thus, a relative frequency distribution of your winnings should look like the one shown in Figure 4.4. The -1 under the left most bar indicates a loss of \$1 by you.

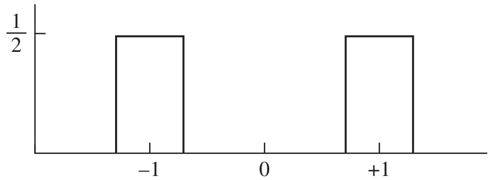


FIGURE **4.4**
Relative frequency of winnings.

On average, how much will you win per game in the long run? If Figure 4.4 presents a correct display of your winnings, you win -1 (lose a dollar) half of the time and $+1$ half of the time, for an average of

$$(-1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{2}\right) = 0.$$

This average is sometimes called your expected winnings per game, or the **expected value** of your winnings. (A game that has an expected value of winnings of 0 is called a *fair game*.) The general definition of expected value is given in Definition 4.4.

DEFINITION **4.4**

The **expected value** of a discrete random variable X with probability distribution $p(x)$ is given by

$$E(X) = \sum_x xp(x).$$

(The sum is over all values of x for which $p(x) > 0$.)

We sometimes use the notation

$$E(X) = \mu$$

for this equivalence.

Note: For the expected value of a discrete random variable X to exist, the sum above must converge absolutely; that is, $\sum_x |x|\, p(x) < \infty$. We talk about an expectation only when it is assumed to exist. ■

Now payday has arrived, and you and your friend up the stakes to \$10 per game of matching coins. You now win -10 or $+10$ with equal probability. Your expected winnings per game is

$$(-10)\left(\frac{1}{2}\right) + (10)\left(\frac{1}{2}\right) = 0,$$

and the game is still fair. The new stakes can be thought of as a function of the old in the sense that if X represents your winnings per game when you were playing for \$1, then $10X$ represents your winnings per game when you play for \$10.

Such functions of random variables arise often. If we are interested not in X but some function of X, say $g(X)$, we often want to know the mean of $g(X)$. Fortunately, if we know the distribution of X, we can find the mean of $g(X)$ without first deriving the distribution of $g(X)$. This idea is formalized in Theorem 4.1.

THEOREM **4.1**

If X is a discrete random variable with probability distribution $p(x)$ and if $g(x)$ is any real-valued function of X, then

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

(The proof of this theorem will not be given.) ■

You and your friend decide to complicate the payoff rules to the coin-matching game by agreeing to let you win \$1 if the match is tails and \$2 if the match is heads. You still lose \$1 if the coins do not match. Quickly you see that this is not a fair game, because your expected winnings are

$$(-1)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{4}\right) = 0.25.$$

You compensate for this by agreeing to pay your friend \$1.50 if the coins do not match. Then, your expected winnings per game are

$$(-1.5)\left(\frac{1}{2}\right) + (1)\left(\frac{1}{4}\right) + (2)\left(\frac{1}{4}\right) = 0,$$

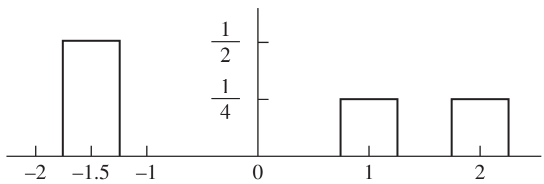


FIGURE **4.5**
Relative frequency of winnings.

and the game is again fair. What is the difference between this game and the original one, in which all payoffs were \$1? The difference certainly cannot be explained by the expected value, because both games are fair. You can win more, but also lose more, with the new payoffs, and the difference between the two games can be explained to some extent by the increased variability of your winnings across many games. This increased variability can be seen in Figure 4.5, which displays the relative frequency for your winnings in the new game; the winnings are more spread out than they are in Figure 4.4. Formally, variation is often measured by the variance and by a related quantity called the standard deviation.

DEFINITION **4.5**

The **variance** of a random variable X with expected value μ is given by

$$V(X) = E\left[(X - \mu)^2\right].$$

We sometimes use the notation

$$E\left[(X - \mu)^2\right] = \sigma^2$$

for this equivalence. ▪

Notice that the **variance** can be thought of as the average squared distance between the values of X and the expected value μ. Thus, the units associated with σ^2 are the square of the units of measurement for X. The smallest value that σ^2 can assume is zero. The variance is zero when all the probability is concentrated at a single point, that is, when X takes on a constant value with probability 1. The variance becomes larger as the points with positive probability spread out more.

The **standard deviation** is a measure of variation that maintains the original units of measure, as opposed to the squared units associated with the variance.

DEFINITION **4.6**

The **standard deviation** of a random variable X is the square root of the variance and is given by

$$\sigma = \sqrt{\sigma^2} = \sqrt{E[(X - \mu)^2]}. \quad ▪$$

For the game represented in Figure 4.4, the variance of your winnings (with $\mu = 0$) is

$$\begin{aligned}\sigma^2 &= E\left[(X - \mu)^2\right] \\ &= (-1)^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{2}\right) = 1.\end{aligned}$$

It follows that $\sigma = 1$ as well. For the game represented in Figure 4.5, the variance of your winnings is

$$\begin{aligned}\sigma^2 &= (-1.5)^2\left(\frac{1}{2}\right) + 1^2\left(\frac{1}{4}\right) + 2^2\left(\frac{1}{4}\right) \\ &= 2.375\end{aligned}$$

and the standard deviation is

$$\sigma = \sqrt{\sigma^2} = \sqrt{2.375} = 1.54.$$

Which game would you rather play?

The standard deviation can be thought of as the size of a "typical" deviation between an observed outcome and the expected value. For the situation displayed in Figure 4.4, each outcome (-1 or $+1$) deviates by precisely one standard deviation from the expected value. For the situation described in Figure 4.5, the positive values average 1.5 units from the expected value of 0 (as do the negative values), and so 1.5 units is approximately one standard deviation here.

The mean and the standard deviation often yield a useful summary of the probability distribution for a random variable that can assume many values. An illustration is provided by the age distribution of the U.S. population for 2000 and 2100 (projected, as shown in Table 4.1).

Age is actually a continuous measurement but because it is reported in categories we can treat it as a discrete random variable for purposes of approximating its key function. To move from continuous age intervals to discrete age classes, we assign

TABLE **4.1**
Age Distribution in 2000 and 2100 (Projected).

Age Interval	Age Midpoint	2000	2100
Under 5	3	6.9	6.3
5–9	8	7.3	6.2
10–19	15	14.4	12.8
20–29	25	13.3	12.3
30–39	35	15.5	12.0
40–49	45	15.3	11.6
50–59	55	10.8	10.8
60–69	65	7.3	9.8
70–79	75	5.9	8.3
80 and over	90	3.3	9.9

Source: U.S. Census Bureau.

each interval the value of its midpoint (rounded). Thus, the data in Table 4.1 are interpreted as showing that 6.9% of the 2000 population were around 3 years of age and that 11.6% of the 2100 population is anticipated to be around 45 years of age. (The open intervals at the upper end were stopped at 100 for convenience.)

Interpreting the percentages as probabilities, we see that the mean age for 2000 is approximated by

$$\begin{aligned}\mu &= \sum_x xp(x)\\ &= 3(0.069) + 8(0.073) + 15(0.144) + \cdots + 90(0.033)\\ &= 36.6.\end{aligned}$$

(How does this compare with the median age for 2000, as approximated from Table 4.1?)

For 2100, the mean age is approximated by

$$\begin{aligned}\mu &= \sum_x xp(x)\\ &= 3(0.063) + 8(0.062) + 15(0.128) + \cdots + 90(0.099)\\ &= 42.5.\end{aligned}$$

Over the projected period, the mean age increases rather markedly (as does the median age).

The variations in the two age distributions can be approximated by the standard deviations. For 2000, this is

$$\begin{aligned}\sigma &= \sqrt{\sum_x (x-\mu)^2 p(x)}\\ &= \sqrt{(3-36.6)^2(0.069) + (8-36.6)^2(0.073) + (15-36.6)^2(0.144) + \cdots + (90-36.6)^2(0.033)}\\ &= 22.6.\end{aligned}$$

A similar calculation for the 2100 data yields $\sigma = 26.3$. These results are summarized in Table 4.2.

Not only is the population getting older on average, but its variability is increasing. What are some of the implications of these trends?

We now provide other examples and extensions of these basic results.

TABLE **4.2**
Age Distribution of U.S. Population Summary.

	2000	**2100**
Mean	36.6	42.5
Standard Deviation	22.6	26.3

EXAMPLE 4.3 A department supervisor is considering purchasing a photocopy machine. One consideration is how often the machine will need repairs. Let X denote the number of repairs during a year. Based on past performance, the distribution of X is shown as follows:

Number of repairs, x	0	1	2	3
$p(x)$	0.2	0.3	0.4	0.1

1 What is the expected number of repairs during a year?

2 What are the variance and standard deviation of the number of repairs during a year?

Solution **1** From Definition 4.4, we see that

$$\begin{aligned} E(X) &= \sum_x xp(x) \\ &= 0(0.2) + 1(0.3) + 2(0.4) + 3(0.1) \\ &= 1.4. \end{aligned}$$

The photocopy machine will need to be repaired an average of 1.4 times per year.

2 From Definition 4.5, we see that

$$\begin{aligned} V(X) &= \sum_x (x-\mu)^2 p(x) \\ &= (0-1.4)^2(0.2) + (1-1.4)^2(0.3) + (2-1.4)^2(0.4) + (3-1.4)^2(0.1) \\ &= 0.84. \end{aligned}$$

The variance of the number of repairs is 0.84 times^2, and the standard deviation is $\sqrt{0.84} = 0.92$ times per year. ■

Often, we are interested in some function of X, say $g(X)$, that is a linear function. When that is the case, the calculations of expected value and variance are especially simple. Our work in manipulating expected values can be greatly facilitated by making use of the two results of Theorem 4.2.

THEOREM 4.2 For any random variable X and constants a and b,

1 $E(aX + b) = aE(X) + b$.

2 $V(aX + b) = a^2 V(X)$.

Proof We sketch a proof of this theorem for a discrete random variable X having a probability distribution given by $p(x)$. By Theorem 4.1,

$$\begin{aligned} E(aX+b) &= \sum_x (ax+b)p(x) \\ &= \sum_x [(ax)p(x) + bp(x)] \\ &= \sum_x axp(x) + \sum_x bp(x) \\ &= a\sum_x xp(x) + b\sum_x p(x) \\ &= aE(X) + b. \end{aligned}$$

Notice that $\sum_x p(x)$ must equal unity. Also, by Definition 4.5,

$$\begin{aligned} V(aX+b) &= E[(aX+b) - E(aX+b)]^2 \\ &= E[aX+b-(aE(X)+b)]^2 \\ &= E[aX - aE(X)]^2 \\ &= E\left[a^2(x-E(X))^2\right] \\ &= a^2E\left[(X-E(X))^2\right] \\ &= a^2V(X). \quad ▪ \end{aligned}$$

Note: Consider the implications of Theorem 4.2. Adding a constant b to a random variable shifts the mean b units, but it does not affect the variance. Multiplying by a constant a affects both the mean and the variance.

An important special case of Theorem 4.2 involves establishing a "standardized" variable. If X has mean μ and standard deviation σ, then the "standardized" form of X is given by

$$Y = \frac{X-\mu}{\sigma}.$$

Employing Theorem 4.2, one can show that $E(Y) = 0$ and $V(Y) = 1$. This idea is used often in later chapters.

We illustrate the use of these results in the following example.

EXAMPLE **4.4** The department supervisor in Example 4.3 wants to consider the cost of maintenance before purchasing the photocopy machine. The cost of maintenance consists of the expense of a service agreement and the cost of repairs. The service agreement can be purchased for \$200. With the agreement, the cost of each repair is \$50. Find the mean, variance, and standard deviation of the annual costs of repair for the photocopy machine.

Solution Recall that the X of Example 4.3 is the annual number of photocopy machine repairs. The annual cost of the maintenance contract is $50X + 200$. By Theorem 4.2, we have

$$\begin{aligned} E(50X + 200) &= 50E(X) + 200 \\ &= 50(1.4) + 200 \\ &= 270. \end{aligned}$$

Thus, the manager could anticipate the average annual cost of maintenance of the photocopy machine to be \$270.

Also, by Theorem 4.2,

$$\begin{aligned} V(50X + 200) &= 50^2 V(X) \\ &= 50^2(0.84) \\ &= 2100. \end{aligned}$$

The variance of the annual cost of maintenance of the photocopy machine is 2100 dollars^2, and the standard deviation is \$45.83. ▪

Determining the variance by Definition 4.5 is not computationally efficient. Theorem 4.2 leads us to a more efficient formula for computing the variance, as given in Theorem 4.3.

THEOREM 4.3 If X is a random variable with mean μ, then

$$V(X) = E\left(X^2\right) - \mu^2.$$

Proof Starting with the definition of variance, we have

$$\begin{aligned} V(X) &= E\left[(X - \mu)^2\right] \\ &= E\left(X^2 - 2X\mu + \mu^2\right) \\ &= E\left(X^2\right) - E\left(2X\mu\right) + E\left(\mu^2\right) \\ &= E\left(X^2\right) - 2\mu^2 + \mu^2 \\ &= E\left(X^2\right) - \mu^2. \quad ▪ \end{aligned}$$

EXAMPLE 4.5 Use the result of Theorem 4.3 to compute the variance of X as given in Example 4.3.

Solution In Example 4.3, X had a probability distribution given by

x	0	1	2	3
$p(x)$	0.2	0.3	0.4	0.1

and we found that $E(X) = 1.4$. Now,

$$\begin{aligned} E(X^2) &= \sum_x x^2 p(x) \\ &= 0^2(0.2) + 1^2(0.3) + 2^2(0.4) + 3^2(0.1) \\ &= 2.8. \end{aligned}$$

By Theorem 4.3,

$$\begin{aligned} V(X) &= E(X^2) - \mu^2 \\ &= 2.8 - (1.4)^2 \\ &= 0.84. \quad \blacksquare \end{aligned}$$

We have computed means and variances for a number of probability distributions and noted that these two quantities give us some useful information on the center and spread of the probability mass. Now suppose that we know only the mean and the variance for a probability distribution. Can we say anything specific about probabilities for certain intervals about the mean? The answer is "yes." Tchebysheff's Theorem is a useful result of the relationship among mean, standard deviation, and relative frequency and is given in Theorem 4.4.

The inequality in the statement of the theorem is equivalent to

$$P(\mu - k\sigma < X < \mu + k\sigma) \geq 1 - \frac{1}{k^2}.$$

To interpret this result, let $k = 2$, for example. Then the interval from $\mu - 2\sigma$ to $\mu + 2\sigma$ must contain at least $1 - 1/k^2 = 1 - 1/4 = 3/4$ of the random variable X's probability. We consider more specific illustrations in the following two examples.

THEOREM 4.4 **Tchebysheff's Theorem.** Let X be a random variable with mean μ and variance σ^2. Then for any positive k,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}.$$

Proof We begin with the definition of $V(X)$ and then make substitutions in the sum defining this quantity. Now,

$$\begin{aligned} V(X) &= \sigma^2 \\ &= \sum_{-\infty}^{\infty} (x-\mu)^2 p(x) \\ &= \sum_{-\infty}^{\mu-k\sigma} (x-\mu)^2 p(x) + \sum_{\mu-k\sigma}^{\mu+k\sigma} (x-\mu)^2 p(x) + \sum_{\mu+k\sigma}^{\infty} (x-\mu)^2 p(x). \end{aligned}$$

(The first sum stops at the largest value of x smaller than $\mu - k\sigma$, and the third sum begins at the smallest value of x larger than $\mu + k\sigma$; the middle sum collects the remaining terms.) Observe that the middle sum is never negative; and for both of the outside sums,

$$(x-\mu)^2 \geq k^2\sigma^2.$$

Eliminating the middle sum and substituting $k^2\sigma^2$ for $(x-\mu)^2$ in the other two, we get

$$\sigma^2 \geq \sum_{-\infty}^{\mu-k\sigma} k^2\sigma^2 p(x) + \sum_{\mu+k\sigma}^{\infty} k^2\sigma^2 p(x)$$

or

$$\sigma^2 \geq k^2\sigma^2 \left[\sum_{-\infty}^{\mu-k\sigma} p(x) + \sum_{\mu+k\sigma}^{\infty} p(x) \right]$$

or

$$\sigma^2 \geq k^2\sigma^2 P\left(|X-\mu| \geq k\sigma\right).$$

It follows that

$$P\left(|X-\mu| \geq k\sigma\right) \leq \frac{1}{k^2}$$

or

$$P\left(|X-\mu| < k\sigma\right) \geq 1 - \frac{1}{k^2}. \quad \blacksquare$$

EXAMPLE 4.6 The daily production of electric motors at a certain factory averaged 120 with a standard deviation of 10.

1 What can be said about the fraction of days on which the production level falls between 100 and 140?

2 Find the shortest interval certain to contain at least 90% of the daily production levels.

Solution **1** The interval from 100 to 140 represents $\mu - 2\sigma$ to $\mu + 2\sigma$, with $\mu = 120$ and $\sigma = 10$. Thus, $k = 2$ and

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

At least 75% of all days, therefore, will have a total production value that falls in this interval. (This percentage could be closer to 95% if the daily production figures show a mound-shaped, symmetric relative frequency distribution.)

2 To find k, we must set $\left(1 - 1/k^2\right)$ equal to 0.9 and solve for k; that is,

$$\begin{aligned} 1 - \frac{1}{k^2} &= 0.9 \\ \frac{1}{k^2} &= 0.1 \\ k^2 &= 10 \\ k &= \sqrt{10} \\ &= 3.16. \end{aligned}$$

The interval

$$\mu - 3.16\sigma \text{ to } \mu + 3.16\sigma$$

or

$$120 - 3.16(10) \text{ to } 120 + 3.16(10)$$

or

$$88.4 \text{ to } 151.6$$

will then contain at least 90% of the daily production levels. ▪

EXAMPLE **4.7** The annual cost of maintenance for a certain photocopy machine has a mean of \$270 and a variance of \$2100 (see Example 4.4). The manager wants to budget enough for maintenance that he is unlikely to go over the budgeted amount. He is considering budgeting \$400 for maintenance. How often will the maintenance cost exceed this amount?

Solution First, we must find the distance between the mean and 400 in terms of the standard deviation of the distribution of costs. We have

$$\frac{400 - \mu}{\sqrt{\sigma^2}} = \frac{400 - 270}{\sqrt{2100}} = \frac{130}{45.8} = 2.84.$$

Thus, 400 is 2.84 standard deviations above the mean. Letting $k = 2.84$ in Theorem 4.4, we can find the interval

$$\mu - 2.84\sigma \quad \text{to} \quad \mu + 2.84\sigma$$

or

$$270 - 2.84(45.8) \quad \text{to} \quad 270 + 2.84(45.8)$$

or

$$140 \quad \text{to} \quad 400$$

must contain at least

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(2.84)^2} = 1 - 0.12 = 0.88$$

of the probability. Because the annual cost is \$200 plus \$50 for each repair, the annual cost cannot be less than \$200. Thus, at most 0.12 of the probability mass can exceed \$400; that is, the cost cannot exceed \$400 more than 12% of the time. ■

Note: It cannot be assumed that half of the values outside the interval are above the upper limit and half are below the lower limit. This is anticipated for some special cases but not in general. In this example, it is not even possible.

EXAMPLE 4.8 Suppose the random variable X has the probability mass function given in the next table.

x	-1	0	1
$p(x)$	$1/8$	$3/4$	$1/8$

Evaluate Tchebysheff's inequality for $k = 2$.

Solution First, we find the mean of X.

$$\mu = \sum_x xp(x) = (-1)\left(\frac{1}{8}\right) + 0\left(\frac{3}{4}\right) + 1\left(\frac{1}{8}\right) = 0$$

Then

$$E(X^2) = \sum_x x^2p(x) = (-1)^2\left(\frac{1}{8}\right) + 0^2\left(\frac{3}{4}\right) + 1^2\left(\frac{1}{8}\right) = \frac{1}{4}$$

and

$$\sigma^2 = E(X^2) - \mu^2 = \frac{1}{4} - 0 = \frac{1}{4}$$

Thus, the standard deviation of X is

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{4}} = \frac{1}{2}.$$

Now, for X, the probability that X is within 2σ of μ is

$$\begin{aligned} P(|X - \mu| < 2\sigma) &= P\left(|X - \mu| < 2\left(\frac{1}{2}\right)\right) \\ &= P(|X - \mu| < 1) \\ &= P(X = 0) \\ &= \frac{3}{4}. \end{aligned}$$

By Tchebysheff's Theorem, the probability that any random variable X is within 2σ of μ is

$$P(|X - \mu| < 2\sigma) \geq 1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = \frac{3}{4}.$$

Therefore, for this particular random variable X, equality holds in Tchebysheff's Theorem. Thus, without knowing more than the mean and variance of a distribution, one cannot obtain a shorter interval than that given by the bounds of this theorem. ■

Exercises

4.17 You are to pay \$1 to play a game that consists of drawing one ticket at random from a box of numbered tickets. You win the amount (in dollars) of the number on the ticket you draw. The following two boxes of numbered tickets are available.

I. $\boxed{0, 1, 2}$ II. $\boxed{0, 0, 0\ 1, 4}$

a Find the expected value and variance of your net gain per play with box I.

b Repeat part (a) for box II.

c Given that you have decided to play, which box would you choose and why?

4.18 The size distribution of U.S. families is shown in the following table.

Number of Persons	Percentage
1	25.7
2	32.2
3	16.9
4	15.0
5	6.9
6	2.2
7 or more	1.1

a Calculate the mean and the standard deviation of family size. Are these exact values or approximations?

b How does the mean family size compare to the median family size?

4.19 The next table shows the estimated number of AIDS cases in the United States by age group.

Numbers of AIDS Cases in the U.S. during 2004

Age	Number of Cases
Under 14	108
15 to 19	326
20 to 24	1,788
25 to 29	3,576
30 to 34	4,786
35 to 39	8,031
40 to 44	8,747
45 to 49	6,245
50 to 54	3,932
55 to 59	2,079
60 to 64	996
65 or older	901
Total	41,515

Source: U.S. Centers for Disease Control and Prevention.

Let X denote the age of a person with AIDS.

a Using the midpoint of the interval to represent the age of all individuals in that age category, find the approximate probability distribution for X.

b Approximate the mean and the standard deviation of this age distribution.

c How does the mean age compare to the approximate median age?

4.20 How old are our drivers? The accompanying table gives the age distribution of licensed drivers in the United States. Describe this age distribution in terms of median, mean, and standard deviation.

Licensed U.S. Drivers in 2004

Age	Number (in millions)
19 and under	9.3
20–24	16.9
25–29	17.4
30–34	18.7
35–39	19.4
40–44	21.3
45–49	20.7
50–54	18.4
55–59	15.8
60–64	11.9
65–69	9.0
70–74	7.4
75–79	6.1
80–84	4.1
85 and over	2.5
Total	198.9

Source: U.S. Department of Transportation.

4.21 Who commits the crimes in the United States? Although this is a very complex question, one way to address it is to look at the age distribution of those who commit violent crimes. This is presented in the table that follows. Describe the distribution in terms of median, mean, and standard deviation.

Age	Percent of Violent Crimes
14 and Under	5.1
15–19	19.7
20–24	20.2
25–29	13.6
30–34	12.0
35–39	10.8
40–44	8.7
45–49	5.1
50–54	2.5
55–59	1.2
60–64	0.6
65 and Older	0.5

Bureau of Justice Statistics.

4.22 A fisherman is restricted to catching at most two red grouper per day when fishing in the Gulf of Mexico. A field agent for the wildlife commission often inspects the day's catch for boats as they come to shore near his base. He has found that the number of red grouper caught has the following distribution.

Number of Grouper	0	1	2
Probability	0.2	0.7	0.1

Assuming that these records are representative of red grouper daily catches in the Gulf, find the expected value, the variance, and the standard deviation for the individual daily catch of red grouper.

4.23 Approximately 10% of the glass bottles coming off a production line have serious defects in the glass. Two bottles are randomly selected for inspection. Find the expected value and the variance of the number of inspected bottles with serious defects.

4.24 Two construction contracts are to be randomly assigned to one or more of three firms—I, II, and III. A firm may receive more than one contract. Each contract has a potential profit of $90,000.

a Find the expected potential profit for firm I.

b Find the expected potential profit for firms I and II together.

4.25 Two balanced coins are flipped. What are the expected value and variance of the number of heads observed?

4.26 In a promotional effort, new customers are encouraged to enter an online sweepstakes. To play, the new customer picks 9 numbers between 1 and 50, inclusive. At the end of the promotional period, 9 numbers from 1 to 50, inclusive, are drawn without replacement from a hopper. If the customer's 9 numbers match all of those drawn (without concern for order), the customer wins $5,000,000.

a What is the probability that a randomly selected new customer wins the $5,000,000?

b What is the expected value and variance of the winnings?

c If a new customer had to mail in the picked numbers, assuming that the cost of postage and handling is $0.50, what is the expected value and variance of the winnings?

4.27 The number of equipment breakdowns in a manufacturing plant is closely monitored by the supervisor of operations, because it is critical to the production process. The number averages five per week with a standard deviation of 0.8 per week.

a Find an interval that includes at least 90% of the weekly figures for the number of breakdowns.

b The supervisor promises that the number of breakdowns will rarely exceed 8 in a 1-week period. Is the director safe in making this claim? Why?

4.28 Keeping an adequate supply of spare parts on hand is an important function of the parts department of a large electronics firm. The monthly demand for 100-gigabyte hard drives for personal computers was studied for some months and found to average 28 with a standard deviation of 4. How many 100-gigabyte hard drives should be stocked at the beginning of each month to ensure that the demand will exceed the supply with a probability of less than 0.10?

4.29 An important feature of golf cart batteries is the number of minutes they will perform before needing to be recharged. A certain manufacturer advertises batteries that will run, under a 75 amp discharge test, for an average of 125 minutes with a standard deviation of 5 minutes.

a Find an interval that contains at least 90% of the performance periods for batteries of this type.

b Would you expect many batteries to die out in less than 100 minutes? Why?

4.30 Costs of equipment maintenance are an important part of a firm's budget. Each visit by a field representative to check out a malfunction in a certain machine used in a manufacturing process is \$65, and the parts cost, on average, about \$125 to correct each malfunction. In this large plant, the expected number of these machine malfunctions is approximately five per month, and the standard deviation of the number of malfunctions is two.

a Find the expected value and standard deviation of the monthly cost of visits by the field representative.

b How much should the firm budget per month to ensure that the costs of these visits are covered at least 75% of the time?

4.31 Four possible winning tickets for a lottery—DAC-6732, MKK-1972, LOJ-8221, and JPM-1182—are given to you. You will win a grand prize of \$1,000,000 if one of your tickets has the same letters and numbers as the winning ticket, and the letters and numbers do not have to be in the same order. All you have to do is mail the coupon back; no purchase is required. From the structure of the numbers you have received, it is obvious that the entire list consists of all the permutations of three letters from the alphabet followed by four digits and that the same letter or number may be used more than once in any lottery number. Is the coupon worth mailing back given that postage would cost 50 cents?

4.32 A man has seven keys on a key ring, one of which fits the door he wants to unlock. He randomly selects a key and tries it. If it does not unlock the door, he randomly selects another key from those remaining and tries to unlock the door with it. He continues in this manner until the door is unlocked. Let X be the number of keys he tries before unlocking the door, counting the key that actually worked. Find the expected value and the standard deviation of X.

4.33 A merchant stocks a certain perishable item. He knows that on any given day he will have a demand for two, three, or four of these items, with probabilities 0.2, 0.3, and 0.5, respectively. He buys the items for \$1 each and sells them for \$1.20 each. Any items left at the end of the day represent a total loss. How many items should the merchant stock to maximize his expected daily profit?

4.34 The National Basketball Association conducts a draft lottery for the teams who do not make the playoffs (those with the poorest records). From 1990 to 1993, the 11 teams participating in the lottery received the picks according to the following process. Ranking the 11 teams from best to worst in terms of win-loss records, balls are put in a hopper for each team, with the number of balls equaling the rank. (So, the team with the worst record gets 11 balls and the team with the best record, among these 11 teams, gets one ball.) Then balls are drawn at random from the hopper. The team whose ball is drawn first gets the first pick in the draft. If any other ball from this team is selected in a subsequent draw, it is ignored. The team whose ball is drawn second gets the second pick, and any other ball drawn from this team is ignored. A third ball is drawn, and the team associated with this ball gets the third pick. The remaining 4th to 11th picks are then assigned to teams in reverse order of their win-loss record. So, the team who has the worst record and who does not yet have an assigned draft pick after the lottery gets the fourth pick. For this scheme, answer the following questions.

a What is the probability that the worst team gets the first pick?

b What is the expected value and variance of the draft pick of the worst team?

4.35 Consider Exercise 4.34 again.

a What is the probability that the best (of the 11 worst) team gets the first draft pick?

b What is the expected value and variance of the draft pick of this best team?

4.36 In 1994, the National Basketball Association changed the process of the draft lottery so that the poorest teams had a much higher probability of getting an early pick. As of 2008, 16 teams qualify for the playoffs and 14 participate in the draft. To determine the winner, 14 ping pong balls numbered 1 to 14 are placed in a hopper and thoroughly mixed. Then four balls are randomly selected from the hopper. The team holding the same four numbers, in any order, is awarded the pick. As an example, if balls 3-5-1-2 were drawn, the team holding this combination of numbers (1-2-3-5) wins the pick. (If 11-12-13-14 is drawn, a new set of balls is drawn.) The balls are returned to the hopper and another set of four balls is used to identify the team receiving the second pick, ignoring any combinations that belong to the team receiving the first pick. Similarly, the team receiving the third pick is identified. The remaining 11 teams are awarded the remaining picks in order of their rankings. Ranking the teams from worst (1) to best (14), the number of combinations awarded each team in the lottery is shown in the table that follows. Let X be the pick the worst team gets from this lottery.

Rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Combinations	250	199	156	119	88	63	43	28	17	11	8	7	6	5

a Find the probability function of X.

b Find the expected value and variance of the pick of the worst team.

4.37 Four couples go to dinner together. The waiter seats the men randomly on one side of the table and the women randomly on the other side of the table. Find the expected value and variance of the number of couples who are seated across from each other.

4.38 A family has an old watchdog that is hard of hearing. If a burglar comes, the probability that the dog will hear him is 0.6. If the dog does hear the burglar, the probability that he will bark and awaken the family is 0.8. Suppose burglars come to this family's house on three separate occasions. Let X be the number of times (out of the three attempted burglaries) the dog barks and awakens the family. Find the mean and standard deviation of X.

4.39 Consider Exercise 4.15. Find the mean and standard deviation of the score the archer obtains:

a from shooting one arrow.

b from shooting five arrows.

4.40 Consider Exercise 4.16. Find the mean and standard deviation of the score a child obtains:

a from throwing one dart.

b from throwing two darts.

4.41 Red-green color blindness is a genetic disorder. Transmission of the disorder is sex-linked because it is only carried on the X chromosome. Females have two X chromosomes, and males have one X and one Y chromosome. For a male to be color-blind, he must get an X chromosome from his mother that carries the color-blindness gene. For a female to be color-blind, she must receive X chromosomes from both her mother and father who carry the color-blindness gene. If a female has the gene on only one of the two X chromosomes, she is not color-blind, but she is a carrier because she can pass the trait on to her children. Suppose a male who is not color-blind and a female who is a carrier for color blindness plan to marry and have three children. What is the expected value and standard deviation for the number of their future children who will be color-blind?

4.42 For a discrete random variable X taking on values $0,\ 1,\ 2,\ \ldots,$ show that $E(X) = \sum_{n=0}^{\infty} P(X > n)$.

At this point, it may seem that every problem has its own unique probability distribution, and that we must start from basics to construct such a distribution each time a new problem comes up. Fortunately, this is not the case. Certain basic probability distributions can be developed as models for a large number of practical problems. In the remainder of this chapter, we shall consider some fundamental discrete distributions, looking at the theoretical assumptions that underlie these distributions as well as at the means, variances, and applications of the distributions.

4.3 The Bernoulli Distribution

Numerous experiments have two possible outcomes. If an item is selected from the assembly line and inspected, it is either defective or not defective. A piece of fruit is either damaged or not damaged. A cow is either pregnant or not pregnant. A child is either female or male. Such experiments are called *Bernoulli trials* after the Swiss mathematician Jacob Bernoulli.

For simplicity, suppose one outcome of a Bernoulli trial is identified to be a success and the other a failure. Define the random variable X as follows:

$$\begin{aligned} X &= 1, \text{ if the outcome of the trial is a success} \\ &= 0, \text{ if the outcome of the trial is a failure.} \end{aligned}$$

If the probability of observing a success is p, the probability of observing a failure is $1 - p$. The probability distribution of X, then, is given by

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0, 1$$

where $p(x)$ denotes the probability that $X = x$. Such a random variable is said to have a *Bernoulli distribution* or to represent the outcome of a single Bernoulli trial. A general formula for $p(x)$ identifies a family of distributions indexed by certain constants called parameters. For the Bernoulli distribution, the probability of success, p, is the only parameter.

Suppose that we repeatedly observe the outcomes of random experiments of this type, recording a value of X for each outcome. What average of X should we expect to see? By Definition 4.4, the expected value of X is given by

$$\begin{aligned} E(X) &= \sum_x xp(x) \\ &= 0p(0) + 1(p(1) \\ &= 0(1-p) + 1(p) = p. \end{aligned}$$

Thus, if we inspect a single item from an assembly line and 10% of the items are defective, we should expect to observe an average of 0.1 defective items per item inspected. (In other words, we should expect to see 1 defective item for every 10 items inspected.)

For the Bernoulli random variable X, the variance (see Theorem 4.3) is

$$\begin{aligned}
V(X) &= E(X^2) - [E(X)]^2 \\
&= \sum_x x^2 p(x) - p^2 \\
&= 0^2(1-p) + 1^2(p) - p^2 \\
&= p - p^2 = p(1-p).
\end{aligned}$$

Seldom is one interested in observing only one outcome of a Bernoulli trial. However, the Bernoulli random variable will be used as a building block to form other probability distributions, such as the binomial distribution of Section 4.4. The properties of the Bernoulli distribution are summarized below.

The Bernoulli Distribution

$$p(x) = p^x(1-p)^{1-x}, \quad x = 0,\ 1 \quad \text{for } 0 < p < 1$$

$$E(X) = p \quad \text{and} \quad V(X) = p(1-p)$$

4.4 The Binomial Distribution

4.4.1 Probability Function

Suppose we conduct n independent Bernoulli trials, each with a probability p of success. Let the random variable X be the number of successes in the n trials. The distribution of X is called the binomial distribution. As an illustration, instead of inspecting a single item, as we do with a Bernoulli random variable, suppose that we now independently inspect n items and record values for $Y_1, Y_2, \ldots, Y_n$, where $Y_i = 1$ if the ith inspected item is defective and $Y_i = 0$, otherwise. The sum of the Y_i's,

$$X = \sum_{i=1}^{n} Y_i$$

denotes the number of defectives among the n sampled items.

We can easily find the probability distribution for X under the assumption that $P(Y_i = 1) = p$, where p remains constant over all trials. For the sake of simplicity, let us look at the specific case of $n = 3$. The random variable X can then take on four possible values: 0, 1, 2, and 3. For X to be 0, all three Y_i values must be 0. Thus,

$$\begin{aligned}
P(X = 0) &= P(Y_1 = 0,\ Y_2 = 0,\ Y_3 = 0) \\
&= P(Y_1 = 0)P(Y_2 = 0)P(Y_3 = 0) \\
&= (1-p)^3.
\end{aligned}$$

Now if $X = 1$, then exactly one value of Y_i is 1 and the other two are 0. The one defective could occur on any of the three trials; thus,

$$
\begin{aligned}
P(X = 1) &= P[(Y_1 = 1,\ Y_2 = 0,\ Y_3 = 0) \cup (Y_1 = 0,\ Y_2 = 1,\ Y_3 = 0) \\
&\qquad \cup (Y_1 = 0,\ Y_2 = 0,\ Y_3 = 1)] \\
&= P(Y_1 = 1,\ Y_2 = 0,\ Y_3 = 0) + P(Y_1 = 0,\ Y_2 = 1,\ Y_3 = 0) \\
&\qquad + P(Y_1 = 0,\ Y_2 = 0,\ Y_3 = 1)
\end{aligned}
$$

(because the three possibilities are mutually exclusive)

$$
\begin{aligned}
&= P(Y_1 = 1)P(Y_2 = 0)P(Y_3 = 0) + P(Y_1 = 0)P(Y_2 = 1)P(Y_3 = 0) \\
&\qquad + P(Y_1 = 0)P(Y_2 = 0)P(Y_3 = 1)
\end{aligned}
$$

(by independence of the Y_i's)

$$
\begin{aligned}
&= p(1-p)^2 + p(1-p)^2 + p(1-p)^2 \\
&= 3p(1-p)^2.
\end{aligned}
$$

Notice that the probability of each specific outcome is the same, $p(1-p)^2$.

For $X = 2$, two values of Y_i must be 1 and one must be 0, which can occur in three mutually exclusive ways. Hence,

$$
\begin{aligned}
P(X = 2) &= P[(Y_1 = 1,\ Y_2 = 1,\ Y_3 = 0) \cup (Y_1 = 1,\ Y_2 = 0,\ Y_3 = 1) \\
&\qquad \cup (Y_1 = 0,\ Y_2 = 1,\ Y_3 = 1)] \\
&= P(Y_1 = 1)P(Y_2 = 1)P(Y_3 = 0) + P(Y_1 = 1)P(Y_2 = 0)P(Y_3 = 1) \\
&\qquad + P(Y_1 = 0)P(Y_2 = 1)P(Y_3 = 1) \\
&= p^2(1-p) + p^2(1-p) + p^2(1-p) \\
&= 3p^2(1-p).
\end{aligned}
$$

The event $X = 3$ can occur only if all values of X_i are 1, so

$$
\begin{aligned}
P(X = 3) &= P(Y_1 = 1,\ Y_2 = 1,\ Y_3 = 1) \\
&= P(Y_1 = 1)P(Y_2 = 1)P(Y_3 = 1) \\
&= p^3.
\end{aligned}
$$

Notice that the coefficient in each of the expressions for $P(X = x)$ is the number of ways of selecting x positions in sequence in which to place 1s. Because there are three positions in the sequence, this number amounts to $\binom{3}{x}$. Thus we can write

$$
P(X = x) = \binom{3}{x} p^x (1-p)^{3-x}, \qquad x = 0,\ 1,\ 2,\ 3, \qquad \text{when } n = 3.
$$

For general values of n, the probability that X will take on a specific value—say, x—is given by the term $p^x(1-p)^{n-x}$ multiplied by the number of possible outcomes that result in exactly x defectives being observed. This number, which represents the number of possible ways of selecting x positions for defectives in the n possible positions of the sequence, is given by

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where $n! = n(n-1)\ldots 1$ and $0! = 1$. Thus, in general, the probability mass function for the binomial distribution is

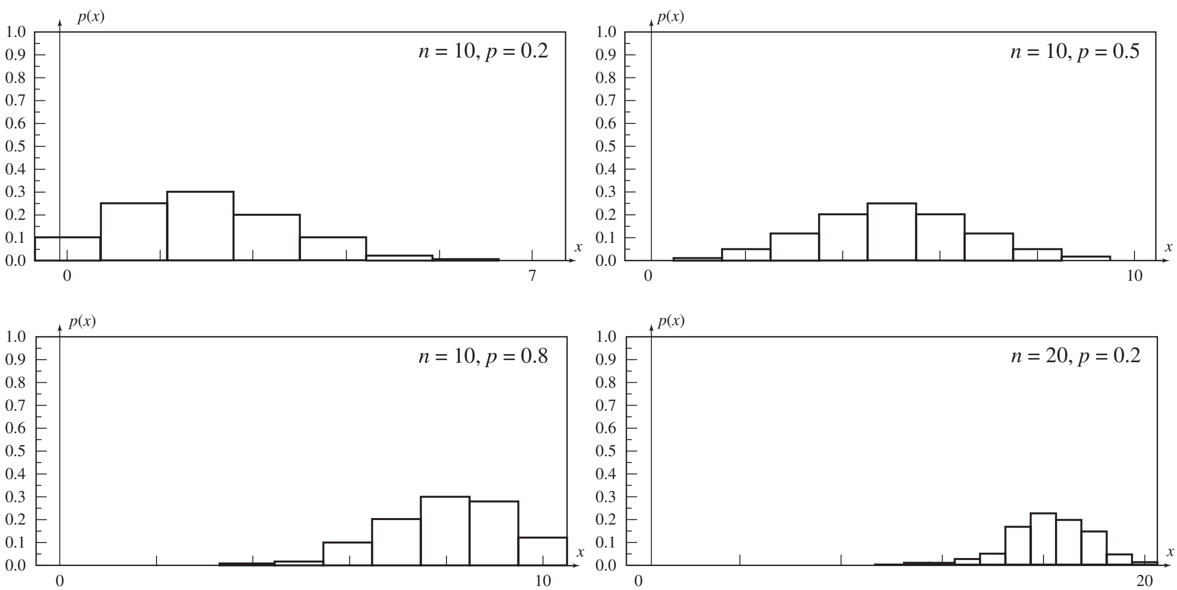
$$P(X = x) = p(x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0,\ 1,\ 2,\ \ldots,\ n.$$

Once n and p are specified we can completely determine the probability function for the binomial distribution; hence, the parameters of the binomial distribution are n and p.

The shape of the binomial distribution is affected by both parameters n and p. If $p = 0.5$, the distribution is symmetric. If $p < 0.5$, the distribution is skewed right, becoming less skewed as n increases. Similarly, if $p > 0.5$, the distribution is skewed left and becomes less skewed as n increases (Figure 4.6). You can explore the shape of the binomial distribution using the *Discrete Distributions* applet. When $n = 1$,

$$p(x) = \binom{1}{x} p^x(1-p)^{1-x} = p^x(1-p)^{1-x}, \qquad x = 0,\ 1$$

FIGURE **4.6** Binomial probabilities.



the probability function of the Bernoulli distribution. Thus, the Bernoulli distribution is a special case of the binomial distribution with $n = 1$.

Notice that the binomial probability function satisfies the two conditions of a probability function. First, probabilities are nonnegative. Second, the sum of the probabilities is one, which can be verified using the binomial theorem:

$$\begin{aligned}\sum_x p(x) &= \sum_{x=0}^{n} \binom{n}{y} p^y (1-p)^{n-y} \\ &= (p + (1-p))^n \\ &= 1\end{aligned}$$

Although we have used $1 - p$ to denote the probability of failure, $q = 1 - p$ is a common notation that we will use here and in later sections.

To summarize, a random variable X possesses a binomial distribution if the following conditions are satisfied:

1 The experiment consists of a fixed number n of identical trials.
2 Each trial can result in one of only two possible outcomes, called success or failure; that is, each trial is a Bernoulli trial.
3 The probability of success p is constant from trial to trial.
4 The trials are independent.
5 X is defined to be the number of successes among the n trials.

Many experimental situations involve random variables that can be adequately modeled by the binomial distribution. In addition to the number of defectives in a sample of n items, examples include the number of employees who favor a certain retirement policy out of n employees interviewed, the number of pistons in an eight-cylinder engine that are misfiring, and the number of electronic systems sold this week out of the n that were manufactured.

EXAMPLE **4.9** Suppose that 10% of a large lot of apples are damaged. If four apples are randomly sampled from the lot, find the probability that exactly one apple is damaged. Find the probability that at least one apple in the sample of four is defective.

Solution We assume that the four trials are independent and that the probability of observing a damaged apple is the same (0.1) for each trial. This would be approximately true if the lot indeed is *large*. (If the lot contains only a few apples, removing one apple would substantially change the probability of observing a damaged apple on the second draw.) Thus, the binomial distribution provides a reasonable model for this experiment, and we have (with X denoting the number of defectives)

$$\begin{aligned}p(1) &= \binom{4}{1} (0.1)^1 (0.9)^3 \\ &= 0.2916.\end{aligned}$$

To find $P(X \geq 1)$, we observe that

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) = 1 - p(0) \\ &= 1 - \binom{4}{0}(0.1)^0(0.9)^4 \\ &= 1 - (0.9)^4 \\ &= 0.3439. \quad \blacksquare \end{aligned}$$

Discrete distributions, like the binomial, can arise in situations where the underlying problem involves a continuous (that is, nondiscrete) random variable. The following example provides an illustration.

EXAMPLE 4.10 In a study of life lengths for a certain battery for laptop computers, researchers found that the probability that a battery life Y will exceed 5 hours is 0.12. If three such batteries are in use in independent laptops, find the probability that only one of the batteries will last 5 hours or more.

Solution Letting X denote the number of batteries lasting 5 hours or more, we can reasonably assume X to have a binomial distribution with $p = 0.12$. Hence,

$$P(X = 1) = p(1) = \binom{3}{1}(0.12)^1(0.88)^2 = 0.279. \quad \blacksquare$$

4.4.2 Mean and Variance

There are numerous ways to find $E(X)$ and $V(X)$ for a binomially distributed random variable X. We might use the basic definition and compute

$$\begin{aligned} E(X) &= \sum_x x p(x) \\ &= \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x} \end{aligned}$$

but direct evaluation of this expression is a bit tricky. Another approach is to make use of the results on linear functions of random variables, which is presented in Chapter 6. We shall see in Chapter 6 that, because the binomial X arose as a sum of independent

Bernoulli random variables $Y_1, Y_2, \ldots, Y_n$,

$$\begin{aligned}E(X) &= E\left[\sum_{i=1}^{n} Y_i\right] \\ &= \sum_{i=0}^{n} E(Y_i) \\ &= \sum_{i=1}^{n} p \\ &= np\end{aligned}$$

and

$$V(X) = \sum_{i=1}^{n} V(Y_i) = \sum_{i=1}^{n} p(1-p) = np(1-p).$$

EXAMPLE 4.11 Referring to Example 4.9, suppose that a customer is the one who randomly selects and then purchases the four apples. If an apple is damaged, the customer will complain. To keep the customers satisfied, the store has a policy of replacing any damaged item (here the apple) and giving the customer a coupon for future purchases. The cost of this program has, through time, been found to be $C = 0.5X^2$, where X denotes the number of defective apples in the purchase of four. Find the expected cost of the program when a customer randomly selects four apples from the lot.

Solution We know that

$$E(C) = E(0.5X^2) = 0.5E(X^2)$$

and it now remains for us to find $E(X^2)$. From Theorem 4.3,

$$V(X) = E(X-\mu)^2 = E(X^2) - \mu^2.$$

Because $V(X) = np(1-p)$ and $\mu = E(X) = np$, we see that

$$\begin{aligned}E(X^2) &= V(X) + \mu^2 \\ &= np(1-p) + (np)^2.\end{aligned}$$

For Example 4.9, $p = 0.1$ and $n = 4$; hence,

$$\begin{aligned}E(C) = 0.5E(X^2) &= 0.5[np(1-p) + (np)^2] \\ &= 0.5[4(0.1)(0.9) + (4)^2(0.1)^2] \\ &= 0.26.\end{aligned}$$

If the costs were originally expressed in dollars, we could expect to pay an average of \$0.26 when a customer purchases four apples. ▪

4.4.3 History and Applications

The binomial expansion can be written as

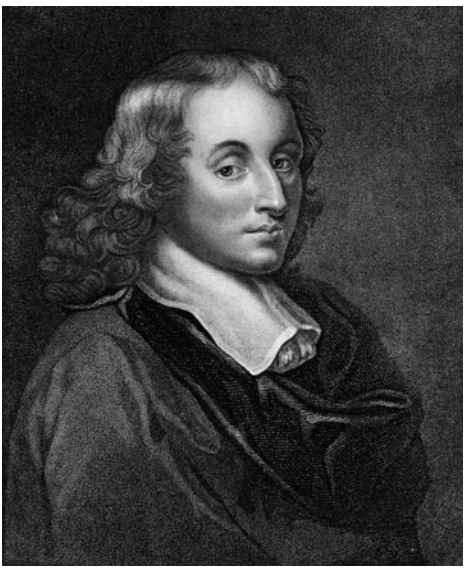
$$(a+b)^n = \sum_{x=0}^{n} \binom{n}{x} a^x b^{n-x}.$$

If $a = p$, where $0 < p < 1$, and $b = 1 - p$, we see that the terms on the left are the probabilities of the binomial distribution. Long ago it was found that the binomial coefficients, $\binom{n}{x}$, could be generated from Pascal's triangle (Figures 4.7 and 4.8).

To construct the triangle, the first two rows are created, consisting of 1s. Subsequent rows have the outside entries as 1s; each of the interior numbers is the sum of the numbers immediately to the left and to the right on the row above.

According to David (1955), the Chinese writer Chu Shih-chieh published the arithmetical triangle of binomial coefficients in 1303, referring to it as an ancient method. The triangle seems to have been discovered and rediscovered several times. Michael Stifel published the binomial coefficients in 1544 (Eves 1969). Pascal's name seems to have become firmly attached to the arithmetical triangle, becoming known as Pascal's triangle at about 1665, although a triangle-type array was also given by Bernoulli in the 1713 *Ars Conjectandi*.

FIGURE **4.7**
Blaise Pascal (1623–1662).



North Wind/North Wind Picture Archives

FIGURE 4.8
Pascal's triangle.

$$\begin{array}{ccccccccc} & & & & 1 & & & & \\ & & & 1 & & 1 & & & \\ & & 1 & & 2 & & 1 & & \\ & 1 & & 3 & & 3 & & 1 & \\ 1 & & 4 & & 6 & & 4 & & 1 \end{array}$$

Jacob Bernoulli (1654–1705) is generally credited for establishing the binomial distribution for use in probability (Figure 4.9) (Folks 1981, Stigler 1986). Although his father planned for him to become a minister, Bernoulli became interested in and began to pursue mathematics. By 1684, Bernoulli and his younger brother John had developed differential calculus from hints and solutions provided by Leibniz. However, the brothers became rivals, corresponding in later years only by print. Jacob Bernoulli would pose a problem in a journal. His brother John would provide an answer in the same issue, and Jacob would respond, again in print, that John had made an error.

FIGURE 4.9
Jacob Bernoulli (1654–1705).

When Bernoulli died of a "slow fever" on August 16, 1705, he left behind numerous unpublished, and some uncompleted, works. The most important of these was on probability. He had worked over a period of about 20 years prior to his death on the determination of chance, and it was this work, *Ars Conjectandi* (The Art of Conjecturing), that his nephew published in 1713. In this work, he used the binomial expansion to address probability problems, presented his theory of permutations and combinations, developed the Bernoulli numbers, and provided the weak law of large numbers for Bernoulli trials.

As was the case with many of the early works in probability, the early developments of the binomial distribution resulted from efforts to address questions relating to games of chance. Subsequently, problems in astronomy, the social sciences, insurance, meteorology, and medicine are but a few of those that have been addressed using this distribution. Polls are frequently reported in the newspaper and on the radio and television. The binomial distribution is used to determine how many people to survey and how to present the results. Whenever an event has two possible outcomes and n such events are to be observed, the binomial distribution is generally the first model considered. This has led it to be widely used in quality control of manufacturing processes.

Determining the probabilities of the binomial distribution quickly becomes too complex to be done quickly by hand. Many calculators and software programs have built-in functions for this purpose. The *Discrete Distributions* applet may be used to compute probabilities for the binomial distribution as well as those for the other discrete distributions discussed in this book. Table 2 in the Appendix gives cumulative binomial probabilities for selected values of n and p. The entries in the table are values of

$$\sum_{x=0}^{a} p(x) = \sum_{x=0}^{a} \binom{n}{x} p^x (1-p)^{n-x}, \qquad a = 0, 1, \ldots, n-1.$$

The following example illustrates the use of both the applet and the table.

EXAMPLE 4.12 An industrial firm supplies 10 manufacturing plants with a certain chemical. The probability that any one firm will call in an order on a given day is 0.2, and this probability is the same for all 10 plants. Find the probability that, on the given day, the number of plants calling in orders is as follows;

1 At most 3

2 At least 3

3 Exactly 3

Solution Let X denote the number of plants that call in orders on the day in question. If the plants order independently, then X can be modeled to have a binomial distribution with $p = 0.2$.

1 We then have

$$\begin{aligned} P(X \leq 3) &= \sum_{x=0}^{3} p(x) \\ &= \sum_{x=0}^{3} \binom{10}{x} (0.2)^x (0.8)^{10-x} \\ &= 0.879. \end{aligned}$$

To use the *Discrete Distributions* applet, we select the binomial distribution and set $n = 10$, $p = 0.2$, and x "less than or equal to" 3. The area of the probability mass function associated with $X \le 3$ is shaded, and the probability is displayed. To use the binomial tables, we turn to Table 2(b) in the Appendix. Note that we use 2(b) because it corresponds to $n = 10$. The probability corresponds to the entry in column $p = 0.2$ and row $k = 3$. Both provide the same result.

2 Notice that

$$\begin{aligned} P(X \ge 3) &= 1 - P(X \le 2) \\ &= 1 - \sum_{x=0}^{2} \binom{10}{x} (0.2)^x (0.8)^{10-x} \\ &= 1 - 0.678 = 0.322. \end{aligned}$$

Here we took advantage of the fact that positive probability only occurs at integer values so that, for example, $P(X = 2.5) = 0$.

3 Observe that

$$\begin{aligned} P(X = 3) &= P(X \le 3) - P(X \le 2) \\ &= 0.879 - 0.678 = 0.201 \end{aligned}$$

from the results just established. ▪

The examples used to this point have specified n and p in order to calculate probabilities or expected values. Sometimes, however, it is necessary to choose n so as to achieve a specified probability. Example 4.13 illustrates the point.

EXAMPLE 4.13 Every hospital has backup generators for critical systems should the electricity go out. Independent but identical backup generators are installed so that the probability that at least one system will operate correctly when called on is no less than 0.99. Let n denote the number of backup generators in a hospital. How large must n be to achieve the specified probability of at least one generator operating, if

1 $p = 0.95$?

2 $p = 0.8$?

Solution Let X denote the number of correctly operating generators. If the generators are identical and independent, X has a binomial distribution. Thus,

$$\begin{aligned} P(X \ge 1) &= 1 - P(X = 0) \\ &= 1 - \binom{n}{0} p^0 (1-p)^n \\ &= 1 - (1-p)^n. \end{aligned}$$

The conditions specify that n must be such that $P(X \geq 1) = 0.99$ or more.

1 When $p = 0.95$,

$$P(X \geq 1) = 1 - (1 - 0.95)^n \geq 0.99$$

results in

$$1 - (0.05)^n \geq 0.99$$

or

$$(0.05)^n \leq 1 - 0.99 = 0.01$$

so $n = 2$; that is, installing two backup generators will satisfy the specifications.

2 When $p = 0.80$,

$$P(X \geq 1) = 1 - (1 - 0.8)^n \geq 0.99$$

results in

$$(0.2)^n \leq 0.01.$$

Now $(0.2)^2 = 0.04$, and $(0.2)^3 = 0.008$, so we must go to $n = 3$ systems to ensure that

$$P(X \geq 1) = 1 - (0.2)^3 = 0.992 > 0.99.$$

Note: We cannot achieve the 0.99 probability exactly, because X can assume only integer values. ▪

EXAMPLE 4.14 Virtually any process can be improved by the use of statistics, including the law. A much-publicized case that involved a debate about probability was the Collins case, which began in 1964. An incident of purse snatching in the Los Angeles area led to the arrest of Michael and Janet Collins. At their trial, an "expert" presented the following probabilities on characteristics possessed by the couple seen running from the crime. The chance that a couple had all of these characteristics together is 1 in 12 million. Because the Collinses had all of the specified characteristics, they must be guilty. What, if anything, is wrong with this line of reasoning?

Man with beard	$\frac{1}{10}$
Blond woman	$\frac{1}{4}$
Yellow car	$\frac{1}{10}$
Woman with ponytail	$\frac{1}{10}$
Man with mustache	$\frac{1}{3}$
Interracial couple	$\frac{1}{1000}$

Solution First, no background data are offered to support the probabilities used. Second, the six events are not independent of one another and, therefore, the probabilities cannot be multiplied. Third, and most interesting, the wrong question is being addressed. The question of interest is not "What is the probability of finding a couple with these characteristics?" Because one such couple has been found (the Collinses), the proper question is: "What is the probability that *another* such couple exists, given that we found one?" Here is where the binomial distribution comes into play. In the binomial model, let

$n =$ Number of couples who could have committed the crime
$p =$ Probability that any one couple possesses the six listed characteristics
$x =$ Number of couples who possess the six characteristics

From the binomial distribution, we know that

$$\begin{aligned} P(X = 0) &= (1 - p)^n \\ P(X = 1) &= np(1 - p)^{n-1} \\ P(X \geq 1) &= 1 - (1 - p)^n. \end{aligned}$$

Then, the answer to the conditional question posed is

$$\begin{aligned} P(X > 1|X \geq 1) &= \frac{P[(X > 1) \cap (X \geq 1)]}{P(X \geq 1)} \\ &= \frac{P(X > 1)}{P(X \geq 1)} \\ &= \frac{1 - (1 - p)^n - np(1 - p)^{n-1}}{1 - (1 - p)^n}. \end{aligned}$$

Substituting $p = 1/12$ million and $n = 12$ million, which are plausible but not well-justified guesses, we get

$$P(X > 1|X \geq 1) = 0.42$$

so the probability of seeing another such couple, given that we have already seen one, is much larger than the probability of seeing such a couple in the first place. This holds true even if the numbers are dramatically changed. For instance, if n is reduced to 1 million, the conditional probability becomes 0.05, which is still much larger than $1/12$ million.

The important lessons illustrated here are that the correct probability question is sometimes difficult to determine and that conditional probabilities are *very* sensitive to conditions. ▪

We shall soon move on to a discussion of other discrete random variables, but the binomial distribution, which is summarized next, is used frequently throughout the remainder of the text.

The Binomial Distribution

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, 2, \ldots, n \quad \text{for } 0 \le p \le 1$$

$$E(X) = np \qquad V(X) = np(1-p)$$

Exercises

4.43 Let X denote a random variable that has a binomial distribution with $p = 0.3$ and $n = 5$. Find the following values.

a $P(X = 3)$
b $P(X \le 3)$
c $P(X \ge 3)$
d $E(X)$
e $V(X)$

4.44 Let X denote a random variable that has a binomial distribution with $p = 0.6$ and $n = 25$. Use your calculator, Table 2 in the Appendix, or the *Discrete Distributions* applet to evaluate the following probabilities.

a $P(X \le 10)$
b $P(X \ge 15)$
c $P(X = 10)$

4.45 A machine that fills milk cartons underfills a certain proportion p. If 50 cartons are randomly selected from the output of this machine, find the probability that no more than 2 cartons are underfilled when:

a $p = 0.05$.
b $p = 0.1$.

4.46 When testing insecticides, the amount of the chemical when given all at once will result in the death of 50% of the population is called the LD50, where LD stands for lethal dose. If 40 insects are placed in separate Petri dishes and treated with an insecticide dosage of LD50, find the probabilities of the following events.

a Exactly 20 survive
b At most 15 survive
c At least 20 survive
d Does it matter whether the insects are placed in separate Petri dishes as they were here or in one large Petri dish? Justify your answer.
e Find the number expected to survive out of 40.
f Find the variance and standard deviation of the number of survivors out of 40.

4.47 Among persons donating blood to a clinic, 85% have Rh^+ blood (that is, the Rhesus factor is present in their blood.) Six people donate blood at the clinic on a particular day.

a Find the probability that at least one of the five does not have the Rh factor.
b Find the probability that at most four of the six have Rh^+ blood.
c The clinic needs six Rh^+ donors on a certain day. How many people must donate blood to have the probability of obtaining blood from at least six Rh^+ donors over 0.95?

4.48 During the 2002 Survey of Business Owners (SBO), it was found that the numbers of female-owned, male-owned, and jointly male- and female-owned businesses were 6.5, 13.2, and 2.7 million, respectively. Among four randomly selected businesses, find the probabilities of the following events.

a All four had a female but no male owner.

b One of the four was either owned or co-owned by a male.

c None of the four were jointly owned by a female and a male.

4.49 Goranson and Hall (1980) explain that the probability of detecting a crack in an airplane wing is the product of p_1, the probability of inspecting a plane with a wing crack; p_2, the probability of inspecting the wing in which the crack is located; and p_3, the probability of detecting the damage.

a What assumptions justify the multiplication of these probabilities?

b Suppose that $p_1 = 0.9$, $p_2 = 0.8$, and $p_3 = 0.5$ for a certain fleet of planes. If three planes are inspected from this fleet, find the probability that a wing crack will be detected in at least one of them.

4.50 Each day a large animal clinic schedules 10 horses to be tested for a common respiratory disease. The cost of each test is \$80. The probability of a horse having the disease is 0.1. If the horse has the disease, treatment costs \$500.

a What is the probability that at least one horse will be diagnosed with the disease on a randomly selected day?

b What is the expected daily revenue that the clinic earns from testing horses for the disease and treating those that are sick?

4.51 Tay-Sachs is a rare genetic disease that results in a fatty substance called ganglioside G_{M2} building up in tissues and nerve cells in the brain, eventually leading to death. The disease is recessive. If a person inherits a gene for Tay-Sachs from one parent but not the other, then he or she shows no signs of the disease but could pass the gene to an offspring. Thus a person who is heterozygous with a Tay-Sachs gene is called a carrier for Tay-Sachs. Suppose two people who are carriers for Tay-Sachs wed and have children.

a What is the probability that a child of these two people will neither have the disease nor be a carrier?

b If the couple has five children, what is the probability that none will have the disease or be a carrier?

4.52 The efficacy of the mumps vaccine is about 80%; that is, 80% of those receiving the mumps vaccine will not contract the disease when exposed. Assume that each person's response to the mumps is independent of another person's response. Find the probability that at least one exposed person will get the mumps if n are exposed where:

a $n = 2$. **b** $n = 4$.

4.53 Refer to Exercise 4.52.

a How many vaccinated people must be exposed to the mumps before the probability that at least one person will contract the disease is at least 0.95?

b In 2006, an outbreak of mumps in Iowa resulted in 605 suspected, probable, and confirmed cases. Given broad exposure in this state of 2.9 million people, do you find this number to be excessively large? Justify your answer.

4.54 A complex electronic system is built with a certain number of backup components in its subsystems. One subsystem has four identical components, each with a probability of 0.15 of failing in less than 1000 hours. The subsystem will operate if any two or more of the four components are operating. Assuming that the components operate independently, find the probabilities of the following events.

a Exactly two of the four components last longer than 1000 hours.

b The subsystem operates for longer than 1000 hours.

4.55 A firm sells four items randomly selected from a large lot that is known to contain 12% defectives. Let X denote the number of defectives among the four sold. The purchaser of the items will return the defectives for repair, and the repair cost is given by

$$C = 2X^2 + X + 3.$$

Find the expected repair cost.

4.56 In a study, dogs were trained to detect the presence of bladder cancer by smelling urine (Willis, et al., 2004). During training, each dog was presented with urine specimens from healthy people, those from people with bladder cancer, and those from people who are sick with unrelated diseases. The dog was trained to lie down by any urine specimen from a person with bladder cancer. Once training was completed, each dog was presented with seven urine specimens, only one of which came from a person with bladder cancer. The specimen that the dog laid down beside was recorded. Each dog took the test nine times. Six dogs were tested.

- **a** One dog had only 1 success in 9. What is the probability of the dog having at least this much success if it cannot detect the presence of bladder cancer by smelling a person's urine?
- **b** Two dogs correctly identified the bladder cancer specimen on 5 of the 9 trials. If neither were able to detect the presence of bladder cancer by smelling a person's urine, what is the probability that both dogs correctly detected the bladder specimen on at least 5 of the 9 trials?

4.57 Fifteen freestanding ranges with smooth tops are available for sale in a wholesale appliance dealer's warehouse. The ranges sell for \$550 each, but a double-your-money-back guarantee is in effect for any defective range that a customer might purchase. Find the expected net revenue for the seller if the probability of any one range being defective is 0.06. (Assume that the quality of any one range is independent of the quality of the others.)

4.58 From a large lot of memory chips for use in personal computers, n are to be sampled by a potential buyer, and the number of defectives X is to be observed. If at least one defective is observed in the sample of n, the entire lot is to be rejected. Find n so that the probability of detecting at least one defective is approximately 0.95 if the following percentages are correct.

- **a** Of the lot of memory chips, 10% are defective.
- **b** Of the lot of memory chips, 5% are defective.

4.59 Consider an insurance company that sells hurricane insurance. In the policy the company agrees to pay the amount A if a hurricane damages the policyholder's home. The company determines that the probability that a hurricane damages the policyholder's home is p. How much should the company charge the policyholder if they want the expected profit from selling the insurance policy to be 5% of A?

4.60 In a certain state, the sentence of each person who has been convicted of a crime that could potentially lead to the death penalty is set by a panel of three judges. The judges determine whether the person will serve a life sentence or receive the death penalty by majority vote. Each judge acts independently. Suppose that when the person is guilty of the crime, each judge votes for the death penalty 40% of the time, whereas if the person is actually innocent each judge votes for the death penalty 5% of the time. One percent of the people convicted of such a crime are innocent. Out of 100 randomly selected cases, what is the expected value and standard deviation of the number of innocent people receiving the death penalty?

4.61 A large food company relies on taste panels to decide whether to go forward with a proposed new product. A taste panel consists of 10 people who are tested independently of one another. Each person is presented with the new product and the standard for that product. (They are blinded to which is the new product and which is the standard one.) They are asked to identify the one they prefer. The company moves forward with the new product if 80% or more of the panel prefer it. Unknown to the company, for the new product under consideration, 30% of all people cannot tell the difference between the two and so will randomly choose one of them. Of the remaining people, 90% prefer the new product. What is the probability that the company will move forward with the new product?

4.62 A particular type of machinery can breakdown from one of two possible types of failures. To assess whether a breakdown is due to the first type costs C_1 dollars. If that type caused the breakdown, the repair costs R_1 dollars. Similarly, the cost of assessing whether the breakdown is due to the second type is C_2 dollars and, if it is a failure of the second type, the repair cost is R_2 dollars. The probability that the first type of failure caused the breakdown is p so $(1 - p)$ is the probability that the second type of failure caused the breakdown. If one type of failure is checked and found not to be the cause of the breakdown, the other type of failure must still be checked before a repair can be done. Under what conditions on p, C_1, C_2, R_1, and R_2 should the first type of failure be checked first, as opposed to the second type of failure being checked first, to minimize the expected total cost of repair.

4.63 A farmer hires a consultant to tell him whether or not he needs to spray his cotton crop to control insects. In any given year, the probability that treatment is necessary to prevent serious economic loss is 0.6. If treatment is necessary, the consultant recommends treatment 99% of the time. If treatment is not truly necessary, the consultant recommends treatment 40% of the time. The farmer always follows the recommendation.

a What is the probability that the farmer will spray 3 years straight?

b The cost of spraying is C. Without considering the cost of spraying, the farmer's profit is P_1 if no treatment is needed or if treatment is given (whether or not it is needed) and P_2 if treatment is needed but not given. Find the expected value of profit, taking into consideration the cost of spraying over the 3 years.

4.64 It is known that 5% of the population has disease A, which can be discovered by means of a blood test. Suppose that N (a large number) people are tested. This can be done in two ways:

1 Each person is tested separately.

2 The blood samples of k people are pooled together and analyzed. (Assume that $N = nk$, with n an integer.) If the test is negative, all of the persons in the pool are healthy (i.e., just this one test is needed). If the test is positive, each of the k persons must be tested separately (i.e., a total of $k + 1$ tests are needed).

a For fixed k, what is the expected number of tests needed in method (2)?

b Find the value for k that will minimize the expected number of tests in method (2).

c How many tests does part (b) save in comparison with part (a)?

4.5 The Geometric Distribution

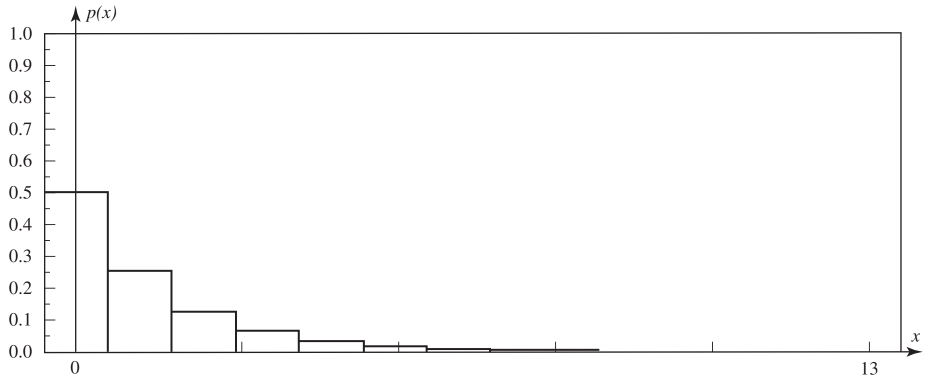
4.5.1 Probability Function

Suppose that a series of test firings of a rocket engine can be represented by a sequence of independent Bernoulli random variables, with $Y_i = 1$ if the ith trial results in a successful firing and with $Y_i = 0$, otherwise. Assume that the probability of a successful firing is constant for the trials, and let this probability be denoted by p. For this problem, we might be interested in the number of failures prior to the trial on which the first successful firing occurs. If X denotes the number of failures prior to the first success, then

$$\begin{aligned}
P(X = x) = p(x) &= P(Y_1 = 0, Y_2 = 0, \ldots, Y_x = 0, Y_{x+1} = 1)\\
&= P(Y_1 = 0)P(Y_2 = 0)\cdots P(Y_x = 0)P(Y_{x+1} = 1)\\
&= (1-p)(1-p)\cdots(1-p)p\\
&= (1-p)^x p,\\
&= q^x p \qquad x = 0,\ 1,\ 2,\ \ldots
\end{aligned}$$

because of the independence of the trials. This formula is referred to as the geometric probability distribution. Notice that this random variable can take on a countably infinite number of possible values. In addition,

FIGURE 4.10 Geometric distribution probability function with $p = 0.5$.



$$
\begin{aligned}
P(X = x) &= q^x p \\
&= q\left[q^{x-1}p\right] \\
&= qP(X = x - 1) \\
&< P(X = x - 1), \qquad x = 1,\ 2,\ \ldots
\end{aligned}
$$

That is, each succeeding probability is less than the previous one (Figure 4.10). The *Discrete Distributions* applet can be used to obtain plots and probabilities of the geometric distribution.

In addition to the rocket-firing example just given, other situations may result in a random variable whose probability can be modeled by a geometric distribution: the number of customers contacted before the first sale is made; the number of times a child is exposed to the measles before contracting the disease; and the number of automobiles going through a radar check before the first speeder is detected.

The following example illustrates the use of the geometric distribution.

EXAMPLE 4.15 A recruiting firm finds that 20% of the applicants for a particular sales position are fluent in both English and Spanish. Applicants are selected at random from the pool and interviewed sequentially. Find the probability that five applicants are interviewed before finding the first applicant who is fluent in both English and Spanish.

Solution Each applicant either is or is not fluent in English and Spanish, so the interview of an applicant corresponds to a Bernoulli trial. The probability of finding a suitable applicant will remain relatively constant from trial to trial if the pool of applicants is reasonably large. Because applicants will be interviewed until the first one fluent in English and Spanish is found, the geometric distribution is appropriate. Let $X =$ the number of *unqualified* applicants prior to the first qualified one. If five unqualified applicants are interviewed before finding the first applicant who is fluent in English and Spanish, we want to find the probability that $X = 5$. Thus,

$$\begin{aligned} P(X = 5) &= p(5) = (0.8)^5(0.2) \\ &= 0.066. \quad \blacksquare \end{aligned}$$

The name of the geometric distribution comes from the geometric series its probabilities represent. Properties of the geometric series are useful when finding the probabilities of the geometric distribution. For example, the sum of a geometric series is

$$\sum_{x=0}^{\infty} t^x = \frac{1}{1-t}$$

for $|t| < 1$. The partial sum of a geometric series is

$$\sum_{x=0}^{n} t^x = \frac{1-t^{n+1}}{1-t}.$$

Using the sum of a geometric series, we can show the geometric probabilities sum to 1:

$$\begin{aligned} \sum_x p(x) &= \sum_{x=0}^{\infty} (1-p)^x p \\ &= p \sum_{x=0}^{\infty} (1-p)^x \\ &= p \frac{1}{1-(1-p)} \\ &= 1 \end{aligned}$$

Similarly, using the partial sum of a geometric series, we can find the functional form of the geometric distribution function. For any integer $x \geq 0$,

$$\begin{aligned} F(x) = P(X \leq x) &= \sum_{t=0}^{x} q^t p \\ &= p \sum_{t=0}^{x} q^t \\ &= p \frac{1-q^{x+1}}{1-q} \end{aligned}$$

$$= p\frac{1-q^{x+1}}{p}$$
$$= 1 - q^{x+1}.$$

Using the distribution function, we have, for any integer $x \geq 0$,

$$P(X \geq x) = 1 - F(x-1) = 1 - (1 - q^x) = q^x.$$

4.5.2 Mean and Variance

From the basic definition,

$$\begin{aligned} E(X) &= \sum_y xp(x) = \sum_{x=0}^{\infty} xpq^x \\ &= p\sum_{x=0}^{\infty} xq^x \\ &= p[0 + q + 2q^2 + 3q^3 + \cdots] \\ &= pq[1 + 2q + 3q^2 + \cdots] \end{aligned}$$

The infinite series can be split up into a triangular array of series as follows:

$$\begin{aligned} E(X) = pq\,[1 + q + q^2 + \cdots \\ +q + q^2 + \cdots \\ +q^2 + \cdots \\ + \cdots] \end{aligned}$$

Each line on the right side is an infinite, decreasing geometric progression with common ratio q. Recall that $a + ax + ax^2 + \cdots = a\big/(1-x)$ if $|x| < 1$. Thus, the first line inside the bracket sums to $1/(1-q) = 1/p$; the second, to q/p; the third, to q^2/p; and so on. On accumulating these totals, we then have

$$\begin{aligned} E(X) &= pq\left[\frac{1}{p} + \frac{q}{p} + \frac{q^2}{p} + \cdots\right] \\ &= q\,[1 + q + q^2 + \cdots] \\ &= \frac{q}{1-q} \\ &= \frac{q}{p}. \end{aligned}$$

This answer for $E(X)$ should seem intuitively realistic. For example, if 10% of a certain lot of items are defective, and if an inspector looks at randomly selected items one at a time, she should expect to find nine good items before finding the first defective one.

The variance of the geometric distribution is derived in Section 4.9 and in Chapter 6. The result, however, is

$$V(X) = \frac{q}{p^2}.$$

EXAMPLE 4.16 Referring to Example 4.15, let X denote the number of unqualified applicants interviewed prior to the first qualified one. Suppose that the first applicant who is fluent in both English and Spanish is offered the position, and the applicant accepts. Suppose each interview costs \$125.

1 Find the expected value and the variance of the total cost of interviewing until the job is filled.

2 Within what interval should this cost be expected to fall?

Solution **1** Because $(X+1)$ is the number of the trial on which the interviewing process ends, the total cost of interviewing is $C = 125(X + 1) = 125X + 125$. Now,

$$\begin{aligned} E(C) &= 125E(X) + 125 \\ &= 125\left(\frac{q}{p}\right) + 125 \\ &= 125\left(\frac{0.8}{0.2}\right) + 125 \\ &= 625 \end{aligned}$$

and

$$\begin{aligned} V(C) &= (125)^2 V(X) \\ &= 125\left(\frac{q}{p^2}\right) \\ &= 125\left(\frac{0.8}{(0.2)^2}\right) \\ &= 2500. \end{aligned}$$

The expected cost of interviewing is \$625, and the standard deviation of the cost is $\sqrt{V(C)} = \sqrt{2500}$ =\$50.

2 Tchebysheff's Theorem (see Section 4.2) says that the cost C will lie within two standard deviations of its mean at least 75% of the time. Thus, it is quite likely that the cost will be between

$$625 - 2(50) \text{ and } 625 + 2(50)$$

or

$$525 \text{ and } 725.$$

The total cost of such an interviewing process is likely to cost between \$525 and \$725. ▪

The geometric is the only discrete distribution that has the **memoryless property**. By this we mean that, if we have observed j straight failures, then the probability of observing at least k more failures (at least $j + k$ total failures) before a success is the same as if we were just beginning and wanted to determine the probability of observing at least k failures prior to the first success; that is, for integers j and k greater than 0,

$$P(X \geq j + k | X \geq j) = P(X \geq k).$$

To verify this, we use the properties of conditional probabilities.

$$\begin{aligned}
P(X \geq j + k | X \geq j) &= \frac{P\left((X \geq j + k) \cap (X \geq j)\right)}{P(X > j)} \\
&= \frac{P(X \geq j + k)}{P(X \geq j)} \\
&= \frac{q^{j+k}}{q^j} \\
&= q^k \\
&= P(X \geq k).
\end{aligned}$$

EXAMPLE **4.17** Referring once again to Example 4.15, suppose that 10 applicants have been interviewed and no person fluent in both English and Spanish has been identified. What is the probability that 15 unqualified applicants will be interviewed before finding the first applicant who is fluent in English and Spanish?

Solution By the memoryless property, the probability that 15 unqualified applicants will be interviewed before finding an applicant who is fluent in English and Spanish, given that the first 10 are not qualified, is equal to the probability of finding the first qualified candidate after interviewing 5 unqualified applicants. Again, let X denote the number of unqualified applicants interviewed prior to the first candidate who is fluent in English and Spanish. Thus,

$$\begin{aligned}
P(X = 15 | X \geq 10) &= \frac{P\left((X = 15) \cap (X \geq 10)\right)}{P(X \geq 10)} \\
&= \frac{P(X = 15)}{P(X \geq 10)}
\end{aligned}$$

$$\begin{aligned} &= \frac{pq^{15}}{q^{10}} \\ &= pq^5 \\ &= P(X = 5). \end{aligned}$$

That is, if we know that 10 unqualified applicants have been interviewed, then the probability that the first qualified applicant will be the 15th person interviewed is the same as the probability that the first qualified applicant would be the 5th person interviewed. ▪

4.5.3 An Alternate Parameterization: Number of Trials versus Number of Failures

For applications such as the ones discussed thus far, the geometric random variable X is often defined as the number of *trials* required to obtain the first success instead of the number of *failures* prior to the first success. If X is the number of trials, the probability function is

$$p(x) = p(1-p)^{x-1}, \quad x = 1,\ 2,\ 3,\ \ldots \qquad \text{for } 0 \le p \le 1.$$

This form of the probability function is often used in calculators and computer software. Notice that it is easy to go between the two formulations. If we are working with the number of failures, we simply add one to the value of the random variable when using software based on trials.

Why do we use the definition of the geometric distribution based on the failures prior to the first success instead of the trials required to obtain the first success? The geometric distribution serves as a model for a number of applications that are not associated with Bernoulli trials. Count data, such as the number of insects on a plant or the number of weeds within a square foot area, may be well modeled by the geometric distribution. In these applications, it is common to observe a count of 0, and we need a probability function that allows for that possibility. An example of such an application follows.

EXAMPLE 4.18 The number of weeds within a randomly selected square meter of a pasture has been found to be well modeled using the geometric distribution. For a given pasture, the number of weeds per square meter averages 0.5. What is the probability that no weeds will be found in a randomly selected square meter of this pasture?

Solution In this application it does not make sense to talk about Bernoulli trials and the probability of success. Instead, we have counts of $0,\ 1,\ 2,\ \ldots$ Let X denote the number of weeds in a randomly selected square meter of the pasture. To find the probability that $X = 0$ (no weeds in the square meter), we must first determine the value of the geometric parameter p. This may be done by equating the mean of the geometric distribution to the mean of the weed counts; that is,

$$E(X) = \frac{1-p}{p} = 0.5.$$

Solving for p, we find $p = 2/3$. Then the probability of finding no weeds in a randomly selected square meter is

$$P(X = 0) = p = \frac{2}{3}.$$

There is only a probability of $1/3$ of seeing one or more weeds in a randomly selected square meter. ▪

The Geometric Distribution

$$p(x) = p(1-p)^x, \quad x = 0, 1, 2, \ldots \qquad \text{for } 0 \le p \le 1$$

$$E(X) = \frac{q}{p} \qquad V(X) = \frac{q}{p^2}$$

4.6 The Negative Binomial Distribution

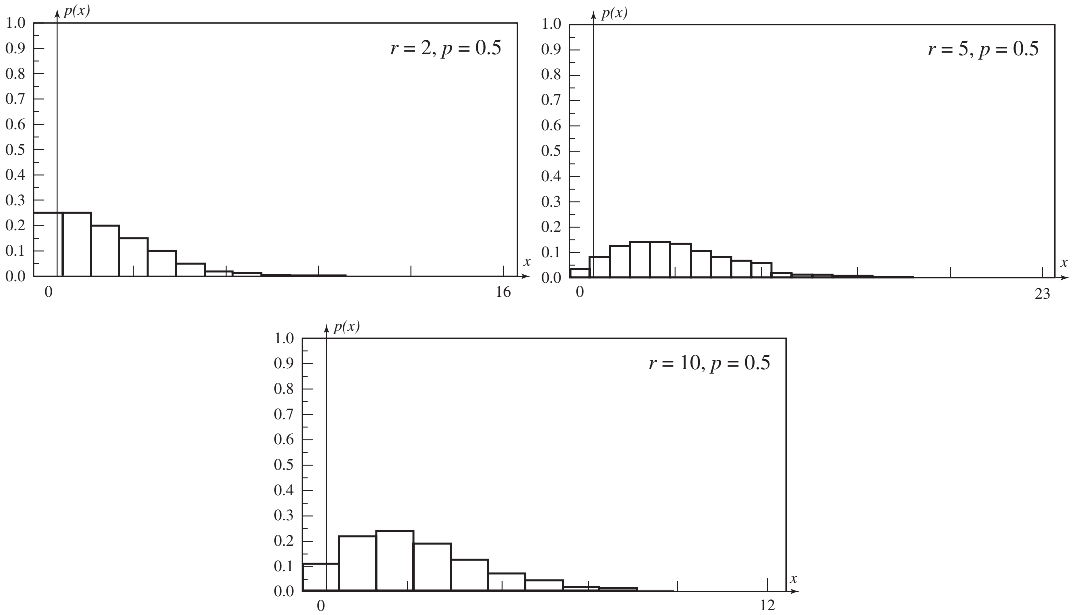
4.6.1 Probability Function

In section 4.5, we saw that the geometric distribution models the probabilistic behavior of the number of failures prior to the *first success* in a sequence of independent Bernoulli trials. But what if we were interested in the number of failures prior to the second success, or the third success, or (in general) the rth success? The distribution governing probabilistic behavior in these cases is called the *negative binomial distribution*.

Let X denote the number of failures prior to the rth success in a sequence of independent Bernoulli trials, with p denoting the common probability of success. We can derive the distribution of X from known facts. Now,

$$\begin{aligned} P(X = x) &= P(\text{1st } (x+r-1) \text{ trials contain } (r-1) \text{ successes and the} \\ &\qquad (x+r)\text{th trial is a success}) \\ &= P[\text{1st } (x+r-1) \text{ trials contain } (r-1) \text{ successes}] \\ &\qquad \times P[(x-r)\text{th trial is a success}]. \end{aligned}$$

FIGURE 4.11 Form of the negative binomial distribution.



Because the trials are independent, the joint probability can be written as a product of probabilities. The first probability statement is identical to the one that results in a binomial model; and hence,

$$\begin{aligned} P(X = x) &= p(x) \\ &= \binom{x+r-1}{r-1} p^{r-1}(1-p)^x \times p \\ &= \binom{x+r-1}{r-1} p^r q^x, \quad x = 0,\ 1,\ \ldots \end{aligned}$$

Notice that, if $r = 1$, we have the geometric distribution. Thus, the geometric distribution is a special case of the negative binomial distribution. The negative binomial is a quite flexible model. Its shape ranges from one being highly skewed to the right when r is small to one that is relatively symmetric as r becomes large and p small. The *Discrete Distributions* applet can be used to explore how the probability mass function changes as the parameters change, as may be seen in Figure 4.11.

EXAMPLE **4.19** As in Example 4.15, 20% of the applicants for a certain sales position are fluent in English and Spanish. Suppose that four jobs requiring fluency in English and Spanish are open. Find the probability that two unqualified applicants are interviewed before finding the fourth qualified applicant, if the applicants are interviewed sequentially and at random.

Solution Again we assume independent trials, with 0.2 being the probability of finding a qualified candidate on any one trial. Let X denote the number of unqualified applicants interviewed prior to interviewing the fourth applicant who is fluent in English and Spanish. X can reasonably be assumed to have a negative binomial distribution, so

$$\begin{aligned} P(X = 2) = p(2) &= \binom{5}{3}(0.2)^4(0.8)^2 \\ &= 10(0.2)^4(0.8)^2 \\ &= 0.01. \quad ■ \end{aligned}$$

4.6.2 Mean and Variance

The expected value, or mean, and the variance for the negative binomially distributed random variable X can easily be found by analogy with the geometric distribution. Recall that X denotes the number of failures prior to the rth success. Let W_1 denote the number of failures prior to the first success; let W_2 denote the number of failures between the first success and the second success; let W_3 denote the number of failures between the second success and the third success; and so forth. The results of the trials can then be represented as follows (where F stands for failure and S represents a success):

$$\underbrace{FF\cdots F}_{W_1}\underline{S},\ \underbrace{FF\cdots}_{W_2}\underline{S},\ \underbrace{FF\cdots F}_{W_3}\underline{S}$$

Clearly, $X = \sum_{i=1}^{r} W_i$, where the W_i values are independent and each has a geometric distribution. Thus, by results to be derived in Chapter 6,

$$E(X) = \sum_{i=1}^{r} E(W_i) = \sum_{i=1}^{r}\left(\frac{q}{p}\right) = \frac{rq}{p}$$

and

$$V(X) = \sum_{i=1}^{r} V(W_i) = \sum_{i=1}^{r}\left(\frac{q}{p^2}\right) = \frac{rq}{p^2}.$$

4.6.3 An Alternate Parameterization: Number of Trials versus Number of Failures

As with the geometric distribution, the negative binomial is often defined as the number of *trials* required to obtain the rth success instead of the number of *failures* prior to the rth success. The probability function is then written as

$$p(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r,\ r+1,\ r+2,\ \ldots \qquad \text{for } 0 \le p \le 1.$$

The negative binomial distribution is often used to model count data, such as the number of accidents in a year, the number of trees in a plot, or the number of insects on a plant. For these applications, it is essential to have positive probability associated with a count of zero. For this reason, we define the negative binomial as the number of failures prior to the rth success. However, it is easy to go between the two definitions as illustrated in the next example.

EXAMPLE 4.20 A swimming pool repairperson has three check valves in stock. Ten percent of the service calls require a check valve. What is the expected number and standard deviation of the number of service calls she will make before running out of check valves?

Solution Let X denote the number of service calls that *do not* require a check valve that the repairperson will make before using the last check valve. Assuming that each service call is independent from the others and that the probability of needing a check valve is constant for each service call, the negative binomial is a reasonable model with $r = 3$ and $p = 0.1$. The total number of service calls made to use the three check valves is $C = X + 3$. Now,

$$\begin{aligned} E(C) &= E(X) + 3 \\ &= \frac{3(0.9)}{0.1} + 3 \\ &= 30 \end{aligned}$$

and

$$\begin{aligned} V(C) &= V(X) \\ &= \frac{3(0.9)}{0.1^2} \\ &= 270. \end{aligned}$$

On average, the repairperson will make 30 calls before running out of check valves. The standard deviation of the number of calls is $\sqrt{270} = 16.4$ calls. ▪

EXAMPLE 4.21 Barnacles often attach to hulls of ships. Their presence speeds corrosion and increases drag resistance, leading to reduced speed and maneuverability. Let X denote the number of barnacles on a randomly selected square meter of a ship hull. For a particular shipyard, the mean and variance of X are 0.5 and 0.625, respectively. Find the probability that at least one barnacle will be on a randomly selected square meter of a ship hull.

Solution To find the probability of at least one barnacle on a randomly selected square meter of a ship hull, we must first determine the value of the parameters of the distribution for this shipyard. We can do this by equating the mean and variance of the negative binomial distribution to those associated with the shipyard; that is,

$$E(X) = \frac{r(1-p)}{p} = 0.5$$

and

$$V(X) = \frac{r(1-p)}{p^2} = 0.625.$$

We now have a system of two equations in two unknowns, and we can solve for the two unknowns. This is simplified because

$$V(X) = \frac{E(X)}{p}$$

so

$$0.625 = \frac{0.5}{p}.$$

Solving for p, we obtain $p = 0.8$. Then r is found to be 2. Now, the probability of at least one barnacle is

$$P(X \geq 1) = 1 - P(X = 0) = 1 - (0.8)^2 = 0.36. \quad ■$$

4.6.4 History and Applications

William S. Gosset (1876–1937) studied mathematics and chemistry at New College Oxford before joining the Arthur Guinness Son and Company in 1899 (Figure 4.12). At the brewery, he worked on a variety of mathematical and statistical problems that arose in the brewing process, publishing the results of his efforts under the pen name "Student." He encountered the negative binomial while working with the distributions of yeast cells counted with a haemocytometer (1907). Gosset reasoned that, if the liquid in which the cells were suspended was properly mixed, then a given particle had an equal chance of falling on any unit area of the hemocytometer. Thus, he was working with the binomial distribution, and he focused on estimating the parameters n and p. To his surprise, in two of his four series, the estimated variance exceeded the mean, resulting in negative estimates of n and p. Nevertheless, these "negative" binomials fit his data well.

FIGURE 4.12
William S. Gosset.

He noted that this may have occurred due to a tendency of the yeast cells "to stick together in groups which was not altogether abolished even by vigorous shaking" (p. 357).

Several other cases appeared in the literature during the early 1900s where estimation of the binomial parameters resulted in negative values of n and p. This phenomenon was explained to some extent by arguing that for large n and small p, the variability of the estimators would cause some negative estimates to be observed. Whitaker (1915) investigated the validity of this claim. In addition to Student's work, she reviewed that of Mortara, who dealt with deaths due to chronic alcoholism, and that of Bortkewitsch, who studied suicides of children in Prussia, suicides of women in German states, accidental deaths in trade societies, and deaths from the kick of a horse in Prussian army corps. Whitaker found it highly unlikely that all negative estimates of p and n could be explained by variability. She, therefore, suggested that a new interpretation was needed with the negative binomial distribution.

Although we have motivated the negative binomial distribution as being the number of failures prior to the rth success in independent Bernoulli trials, several other models have been described that gives rise to this distribution (Boswell and Patil 1970). The Pólya distribution and the Pascal distribution have been other names for the negative binomial distribution. The Pólya distribution was motivated from an urn model and generally refers to the special case where r is a positive integer, although r may be any positive number for the more general negative binomial distribution.

Because the negative binomial distribution has been derived in so many different ways, it has also been presented using different parameterizations. As discussed earlier, some define the negative binomial as the number of *trials* required to get the first success. Because we must have at least r trials, the points of positive probability begin with r, not 0, as we have noted earlier. The negative binomial has been used extensively to model the number of organisms within a sampling unit. For these applications, the parameters are often taken to be r and the mean μ, instead of r and p, because the mean μ is of primary interest in such studies. In these applications, k is often used instead of r to denote one of the parameters. For these reasons, it is important to read carefully how the negative binomial random variable is defined when going from one source to another.

The negative binomial distribution has been applied in many fields, including accident statistics, population counts, psychological data, and communications. Some of these applications are highlighted in the exercises.

The Negative Binomial Distribution

$$p(x) = \binom{x+r-1}{r-1} p^r (1-p)^x, \qquad x = 0,\ 1,\ 2,\ \ldots \qquad \text{for } 0 \le p \le 1$$

$$E(X) = \frac{rq}{p} \qquad\qquad V(X) = \frac{rq}{p^2}$$

Exercises

4.65 Let X denote a random variable that has a geometric distribution with a probability of success on any trial denoted by p. Let $p = 0.1$.

a Find $P(X \geq 2)$.
b Find $P(X \geq 4 | X \geq 2)$

4.66 Let X denote a negative binomial random variable, with $p = 0.6$. Find $P(X \geq 3)$ for the following values of r.

a $r = 2$
b $r = 4$

4.67 Suppose that 10% of the engines manufactured on a certain assembly line are defective. If engines are randomly selected one at a time and tested, find the probability that exactly two defective engines will be tested before a good engine is found.

4.68 Referring to Exercise 4.67, given that the first two engines are defective, find the probability that at least two more defectives are tested before the first nondefective engine is found.

4.69 Referring to Exercise 4.67, find the probability that the 5th nondefective engine will be found as follows:

a When the 7th engine is tested
b When the 10th engine is tested

4.70 Referring to Exercise 4.67, find the mean and the variance of the number of defectives tested before the following events occur.

a The first nondefective engine is found.
b The third nondefective engine is found.

4.71 Greenbugs are pests in oats. If their populations get too high, the crop will be destroyed. When recording the number of greenbugs on randomly selected seedling oat plants, the counts have been found to be modeled well by the geometric distribution. Suppose the average number of greenbugs on a seedling oat plant is 0.25. Find the probability that a randomly selected plant has:

a no greenbugs.
b two greenbugs.
c at least one greenbug.

4.72 The employees of a firm that does asbestos cleanup are being tested for indications of asbestos in their lungs. The firm is asked to send four employees who have positive indications of asbestos on to a medical center for further testing. If 40% of the employees have positive indications of asbestos in their lungs, find the probability that six employees who do not have asbestos in their lungs must be tested before finding the four who do have asbestos in their lungs.

4.73 Referring to Exercise 4.72, if each test costs $40, find the expected value and the variance of the total cost of conducting the tests to locate four positives. Is it highly likely that the cost of completing these tests will exceed $650?

4.74 People with O^- blood are called universal donors because they may give blood to anyone without risking incompatibility due blood type factors (A and B) or to the Rh factor. Of the persons donating blood at a clinic, 9% have O^- blood. Find the probabilities of the following events.

a The first O^- donor is found after blood typing five people who were not O^-.

b The second O^- donor is the sixth donor of the day.

4.75 A geological study indicates that an exploratory oil well drilled in a certain region should strike oil with probability 0.25. Find the probabilities of the following events.

a The first strike of oil comes after drilling three dry (nonproductive) wells.

b Three dry wells are hit before obtaining the third strike of oil.

c What assumptions must be true for your answers to be correct?

4.76 In the setting of Exercise 4.75, suppose that a company wants to set up three producing wells. Find the expected value and the variance of the number of wells that must be drilled to find three successful ones. (Hint: First find the expected value and variance of the number of dry wells that will be drilled before finding the three successful ones.)

4.77 A large lot of tires contains 5% defectives. Four tires are to be chosen from the lot and placed on a car.

a Find the probability that two defectives are found before four good ones.

b Find the expected value and the variance of the number of selections that must be made to get four good tires. (Hint: First find the expected value and variance of the number of defective tires that will be selected before finding the four good ones.)

4.78 An interviewer is given a list of potential people she can interview. Suppose that the interviewer needs to interview five people and that each person independently agrees to be interviewed with probability 0.6. Let X be the number of people she must ask to be interviewed to obtain her necessary number of interviews.

a What is the probability that she will be able to obtain the five people by asking no more than seven people?

b What is the expected value and variance of the number of people she must ask to interview five people?

4.79 A car salesman is told that he must make three sales each day. The salesman believes that if he visits with a customer the probability that the customer will purchase a car is 0.2.

a What is the probability that the salesman will have to visit with at least five customers to make three sales?

b What is the expected number of customers that the salesman must visit with to make his daily sales goal?

4.80 The number of cotton fleahoppers (a pest) on a cotton plant has been found to be modeled well using a negative binomial distribution with $r = 2$. Suppose the average number of cotton fleahoppers on plants in a certain cotton field is one. Find the probability that a randomly selected cotton plant from that field has the following number of fleahoppers.

a No cotton fleahopper

b Five cotton fleahoppers

c At least one cotton fleahopper

4.81 The number of thunderstorm days in a year has been modeled using a negative binomial model (Sakamoto 1973). A thunderstorm day is defined as a day during which at least one thunderstorm cloud (cumulonimbus) occurs accompanied by lightning and thunder. It may or may not be accompanied by strong gusts of wind, rain, or hail. For one such site, the mean and variance of the number of thunderstorm days are 24 days and 40 days^2, respectively. For a randomly selected year, complete the following.

a Find the values for r and p.

b Find the probability that there are no thunderstorm days during a year.

c Find the probability of at least two thunderstorm days during a year.

4.82 Refer again to Exercise 4.81. Someone who is considering moving to the area is concerned about the number of thunderstorm days in a year. He wants assurances that there will be only a 10% chance of the number of thunderstorm days exceeding a specified number of days. Find the number of days that you may properly use in making this assurance.

4.83 John and Fred agree to play a series of tennis games. The first one to win three games is declared the overall winner. Suppose that John is a stronger tennis player than Fred. So the probability that John wins each game is 0.6, and the outcome of each game is independent of the outcomes of the other games.

a Find the probability that John wins the series in i games, for $i = 3, 4, 5$.

b Compare the probability that John wins with the probability that he would win if they played a win two-out-of-three series.

4.84 A child wins a video game with probability p, and the outcome on any one game is independent of that on any other game. He decides to play four games. However, if he loses the fourth game, he will continue to play until he wins a game.

a Find the expected number of games that the child will play.

b Find the expected number of games that the child will win.

4.85 This problem is known as the Banach Match Problem. A pipe-smoking mathematician always carries two matchboxes, one in his right-hand pocket and one in his left-hand pocket. Each time he needs a match he is equally likely to take it from either pocket. The mathematician discovers that one of his matchboxes is empty. If it is assumed that both matchboxes initially contained N matches, what is the probability that there are exactly k matches in the other box, $k = 0, 1, 2, \ldots, N$?

4.86 Refer to Exercise 4.85. Suppose that instead of N matches in each box, the left-hand pocket originally has N_1 matches and the one in the right-hand pocket originally has N_2 matches. What is the probability that there are exactly k matches in the other box, $k = 0, 1, 2, \ldots, N_1$?

4.87 Prove that the negative binomial probability function sums to 1.

4.88 Let X be a negative binomial random variable with parameters r and p, and let Y be a binomial random variable with parameters n and p. Show that

$$P(X > n) = P(Y < r).$$

4.7 The Poisson Distribution

4.7.1 Probability Function

A number of probability distributions come about through limiting arguments applied to other distributions. One useful distribution of this type is called the *Poisson distribution*.

Consider the development of a probabilistic model for the number of accidents that occur at a particular highway intersection in a period of 1 week. We can think of the time interval as being split up into n subintervals such that

$$P(\text{One accident in a subinterval}) = p$$
$$P(\text{No accidents in a subinterval}) = 1 - p.$$

Here we are assuming that the same value of p holds for all subintervals, and that the probability of more than one accident occurring in any one subinterval is zero. If the occurrence of accidents can be regarded as independent from subinterval to subinterval, the total number of accidents in the time period, which equals the total number of subintervals that contain one accident, will have a binomial distribution.

Although there is no unique way to choose the subintervals—and we, therefore, know neither n nor p—it seems reasonable to assume that, as n increases, p should decrease. Thus, we want to look at the limit of the binomial probability distribution as $n \to \infty$ and $p \to 0$. To get something interesting, we take the limit under the restriction that the mean (np in the binomial case) remains constant at a value we call λ.

Now, with $np = \lambda$ or $p = \lambda/n$, we have

$$\begin{aligned}
&\lim_{n\to\infty} \binom{n}{x} \left(\frac{\lambda}{n}\right)^x \left(1 - \frac{\lambda}{n}\right)^{n-x} \\
&\quad = \lim_{n\to\infty} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \frac{n(n-1)\cdots(n-x+1)}{n^x} \left(1 - \frac{\lambda}{n}\right)^{-x} \\
&\quad = \lim_{n\to\infty} \frac{\lambda^x}{x!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-x} \left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\cdots\left(1 - \frac{x-1}{n}\right)
\end{aligned}$$

Noting that

$$\lim_{n\to\infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

and that all other terms involving n tend toward unity, we have the limiting distribution

$$p(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \qquad x = 0,\ 1,\ 2,\ \ldots \qquad \text{for } \lambda > 0$$

which is called the *Poisson distribution with parameter* λ. Recall that λ denotes the mean number of occurrences in one time period (a week, for the example under consideration); hence, if t nonoverlapping time periods were considered, the mean would be λt. Based on this derivation, the Poisson distribution is often referred to as the distribution of rare events.

The Poisson distribution is sometimes used to approximate binomial probabilities. As the preceding derivation would imply, the approximation improves as n increases and/or p decreases. The *Approximations to Distributions* applet may be used to explore how the closeness of the approximation to the true binomial probabilities changes with n and p. We will return to this approximation in Chapter 8.

The Poisson distribution can be used to model counts in areas or volumes as well as in time. For example, we may use this distribution to model the number of flaws in a square yard of textile, the number of bacteria colonies in a cubic centimeter of water, or the number of times a machine fails in the course of a workday. We illustrate the use of the Poisson distribution in the following example.

EXAMPLE 4.22 During business hours, the number of calls passing through a particular cellular relay system averages five per minute.

1 Find the probability that no call will pass through the relay system during a given minute.

2 Find the probability that no call will pass through the relay system during a 2-minute period.

3 Find the probability that three calls will pass through the relay system during a 2-minute period.

Solution **1** If calls tend to occur independently of one another, and if they occur at a constant rate over time, the Poisson model provides an adequate presentation of the probabilities. Thus,

$$p(0) = \frac{5^0}{0!}e^{-5} = e^{-5} = 0.007.$$

2 If the mean number of calls in 1 minute is 5, the mean number of calls in 2 minutes is 10. The Poisson model should still provide an adequate presentation of the probabilities. Thus,

$$p(0) = \frac{10^0}{0!}e^{-10} = e^{-10} = 0.00005.$$

Because the probability that a call being received in any time period is independent of a call being received in any other time period, the event of getting 0 calls in the first minute is independent of the event of getting 0 calls in the second time period. Therefore, it is not surprising that the probability of not getting a call in a 2-minute time period is the product of the probabilities associated with these two events.

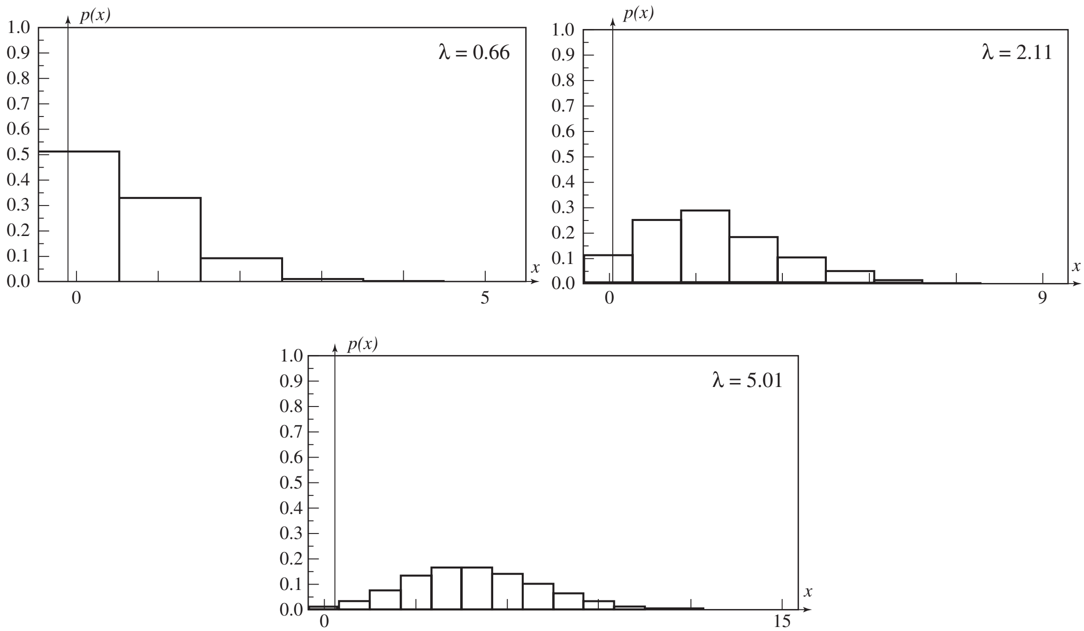
3 Again, using a Poisson model with a mean of 10 for a 2-minute period, we have

$$p(3) = \frac{10^3}{3!}e^{-10} = e^{-10} = 0.0076.$$

Notice that we could consider all possible pairs of outcomes for the first minute and the second minute that would result in three calls: (3, 0), (2, 1), (1, 2), (0, 3). However, it is simpler to recognize that the form of the distribution remains Poisson for the longer time period and only the mean needs to be adjusted. ▪

The shape of the Poisson changes from highly skewed when λ is small to fairly symmetric when λ is large (Figure 4.13). Calculators and functions in computer

FIGURE **4.13** Form of the Poisson distribution.



software often can be used to find the probabilities from the probability mass function and the cumulative distribution function. The *Discrete Distributions* applet may be used to graph the Poisson distribution, as shown in Figure 4.13, and to compute probabilities. Table 3 of the Appendix may also be used to find probabilities associated with Poisson random variables. It gives values for cumulative Poisson probabilities of the form

$$\sum_{x=0}^{a} e^{-\lambda}\frac{\lambda^x}{x!}, \qquad a = 0,\ 1,\ 2, \ldots$$

for selected values of λ.

The following example illustrates the use of the *Discrete Distributions* applet and Table 3 in the Appendix.

EXAMPLE **4.23** Refer to Example 4.22. Find the probability of each of the following events.

1 No more than four calls in the given minute
2 At least four calls in the given minute
3 Exactly four calls in the given minute

Solution Let X denote the number of calls in the given minute. From the *Discrete Distributions* applet or Table 3 in the Appendix:

1 $P(X \le 4) = \sum_{x=0}^{4} \frac{(5)^x}{x!} e^{-5} = 0.440.$

2
$$\begin{aligned} P(X \ge 4) &= 1 - P(X \le 3) \\ &= 1 - \sum_{x=0}^{3} \frac{(5)^x}{x!} e^{-5} \\ &= 1 - 0.265 = 0.735. \end{aligned}$$

3
$$\begin{aligned} P(X = 4) &= P(X \le 4) - P(X \le 3) \\ &= 0.440 - 0.265 \\ &= 0.175. \quad \blacksquare \end{aligned}$$

4.7.2 Mean and Variance

We can intuitively determine what the mean and the variance of a Poisson distribution should be by recalling the mean and the variance of a binomial distribution and the relationship between the two distributions. A binomial distribution has mean np and variance $np(1-p) = np - (np)p$. Now, if n gets large and p becomes small but $np = \lambda$ remains constant, the variance $np - (np)p = \lambda - \lambda p$ should tend toward λ. In fact, the Poisson distribution does have both its mean and its variance equal to λ.

The mean of the Poisson distribution can easily be derived formally if one remembers a simple Taylor series expansion of e^x—namely,

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

Then,

$$\begin{aligned} E(X) &= \sum_{y} p(x) \\ &= \sum_{x=0}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \sum_{x=1}^{\infty} x \frac{\lambda^x}{x!} e^{-\lambda} \\ &= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!} \\ &= \lambda e^{-\lambda} \left(1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots \right) \\ &= \lambda e^{-\lambda} e^{\lambda} \\ &= \lambda. \end{aligned}$$

The formal derivation of the fact that

$$V(X) = \lambda$$

is left as an exercise (see Exercise 4.105).

EXAMPLE 4.24 The manager of an industrial plant is planning to buy a new machine of either type A or type B. For each day's operation, the number of repairs X that machine A requires is a Poisson random variable with mean $0.10t$, where t denotes the time (in hours) of daily operation. The number of daily repairs Y for machine B is a Poisson random variable with mean $0.12t$. The daily cost of operating A is $C_A(t) = 20t + 40X^2$; for B, the cost is $C_B(t) = 16t + 40Y^2$. Assume that the repairs take negligible time and that each night the machines are to be cleaned so that they operate like new machines at the start of each day. Which machine minimizes the expected daily cost for the following times of daily operation?

1 10 hours

2 20 hours

Solution The expected cost for machine A is

$$\begin{aligned} E\left[C_A(t)\right] &= 20t + 40E\left(X^2\right) \\ &= 20t + 40\left[V(X) + (E(X))^2\right] \\ &= 20t + 40\left[0.1t + 0.01t^2\right] \\ &= 24t + 0.4t^2. \end{aligned}$$

Similarly,

$$\begin{aligned} E\left[C_B(t)\right] &= 16t + 40E\left(Y^2\right) \\ &= 16t + 40\left[V(Y) + (E(Y))^2\right] \\ &= 16t + 40\left[0.12t + 0.0144t^2\right] \\ &= 20.8t + 0.576t^2. \end{aligned}$$

1 If the machines operate 10 hours each day,

$$E\left[C_A(10)\right] = 24(10) + 0.4(10)^2 = 280$$

and

$$E\left[C_B(10)\right] = 20.8(10) + 0.576(10)^2 = 265.60$$

which results in the choice of machine B.

2 If the machines operate 20 hours each day,

$$E\left[C_A(20)\right] = 24(20) + 0.4(20)^2 = 640$$

and

$$E\left[C_B(20)\right] = 20.8(20) + 0.576(20)^2 = 646.40$$

which results in the choice of machine A. ▪

4.7.3 History and Applications

Siméon-Denis Poisson (1781–1840), examiner and professor at the École Polytechnique of Paris for nearly 40 years, wrote over 300 papers in the fields of mathematics, physics, and astronomy (Figure 4.14). His most important works were a series of papers on definite integrals and his advances on Fourier series, providing a foundation for later work in this area by Dirichlet and Riemann. However, it was his derivation of the exponential limit of the binomial distribution, much as we saw above, for which he is best known in probability and statistics. The derivation was given no special emphasis. Cournot republished it in 1843 with calculations demonstrating the effectiveness of the approximation. Although De Moivre had presented the exponential limit of the binomial distribution in the first edition of *The Doctrine of Chances*, published in 1718, this distribution became known as the Poisson distribution.

FIGURE **4.14**
Siméon-Denis Poisson (1781–1840).

Bettmann/CORBIS

TABLE 4.3 Deaths of Prussian Cavalrymen Due to Kick by a Horse.

Number of Cavalrymen Killed during a Year	Frequency	Relative Frequency	Theoretical Probability
0	109	0.545	0.544
1	65	0.325	0.331
2	22	0.110	0.101
3	3	0.015	0.021
4	1	0.005	0.003

The Poisson distribution was rediscovered by von Bortkiewicz in 1898. He tabulated the number of cavalrymen in the Prussian army who died from a kick from a horse during a year. To see how the Poisson distribution applies, first suppose that a cavalryman can either be killed by a horsekick during a year or not. Further suppose that the chance of this rare event is the same for all soldiers and that soldiers have independent chances of being killed. Thus, the number of cavalrymen killed during a year is a binomial random variable. However, the probability of being killed, p, is very small and the number of cavalrymen (trials) is very large. Therefore, the Poisson limit is a reasonable model for the data. A comparison of the observed and theoretical relative frequencies is shown in Table 4.3. Notice that the two agree well, indicating the Poisson is an adequate model for these data.

The Poisson distribution can be used as an approximation for the binomial (large n and small p). When count data are observed, the Poisson model is often the first model considered. If the estimates of mean and variance differ significantly so that this property of the Poisson is not reasonable for a particular application, then one turns to other discrete distributions, such as the binomial (or negative binomial) for which the variance is less (greater) than the mean. The number of radioactive particles emitted in a given time period, number of telephone calls received in a given time period, number of equipment failures in a given time period, number of defects in a specified length of wire, and number of insects in a specified volume of soil are some of the many types of data that have been modeled using the Poisson distribution.

The Poisson Distribution

$$p(x) = \frac{\lambda^x}{x!}e^{-\lambda}, \qquad x = 0,\ 1,\ 2,\ \ldots \qquad \text{for } \lambda > 0$$

$$E(X) = \lambda \qquad\qquad V(X) = \lambda$$

Exercises

4.89 Let X denote a random variable that has a Poisson distribution with mean $\lambda = 4$. Find the following probabilities.

a $P(X = 5)$

b $P(X < 5)$

c $P(X \geq 5)$

d $P(X \geq 5 | X \geq 2)$

4.90 The number of calls coming into a hotel's reservation center averages three per minute.

a Find the probability that no calls will arrive in a given 1-minute period.

b Find the probability that at least two calls will arrive in a given 1-minute period.

c Find the probability that at least two calls will arrive in a given 2-minute period.

4.91 The Meteorology Department of the University of Hawaii modeled the number of hurricanes coming within 250 nautical miles of Honolulu during a year using a Poisson distribution with a mean of 0.45. Using this model, determine the probabilities of the following events.

a At least one hurricane will come within 250 nautical miles of Honolulu during the next year.

b At most four hurricanes will come within 250 nautical miles of Honolulu during the next year.

4.92 A certain type of copper wire has a mean number of 0.5 flaws per meter.

a Justify using the Poisson distribution as a model for the number of flaws in a certain length of this wire.

b Find the probability of having at least one flaw in a meter length of the copper wire.

c Find the probability of having at least one flaw in a 5-meter length of the copper wire.

4.93 Customer arrivals at a checkout counter in a department store have a Poisson distribution with an average of seven per hour. For a given hour, find the probabilities of the following events.

a Exactly seven customers arrive.

b No more than two customers arrive.

c At least two customers arrive.

4.94 Referring to Exercise 4.92, the cost of repairing the flaw in the copper wire is $8 per flaw. Find the mean and the standard deviation of the distribution of repair costs for a 10-meter length of wire in question.

4.95 Referring to Exercise 4.93, suppose that it takes approximately 10 minutes to service each customer. Assume that an unlimited number of servers are available so that no customer has to wait for service.

a Find the mean and the variance of the total service time connected to the customer arrivals for 1 hour.

b Is total service time highly likely to exceed 200 minutes?

4.96 Referring to Exercise 4.93, find the probabilities that exactly five customers will arrive in the following 2-hour periods.

a Between 2:00 p.m. and 4:00 p.m. (one continuous 2-hour period)

b Between 1:00 p.m. and 2:00 p.m. and between 3:00 p.m. and 4:00 p.m. (two separate 1-hour periods for a total of 2 hours)

4.97 The number of grasshoppers per square meter of rangeland is often well modeled using the Poisson distribution. Suppose the mean number of grasshoppers in a specified region that has been grazed by cattle is 0.5 grasshoppers per square meter. Find the probabilities of the following events.

a Five or more grasshoppers in a randomly selected square meter in this region

b No grasshoppers in a randomly selected square meter in this region

c At least one grasshopper in a randomly selected square meter in this region

4.98 The number of particles emitted by a radioactive source is generally well modeled by the Poisson distribution. If the average number of particles emitted by the source in an hour is four, find the following probabilities.

a The number of emitted particles in a given hour is at least 6.

b The number of emitted particles in a given hour will be at most 3.

c No particles will be emitted in a given 24-hour period.

4.99 Chu (2003) studied the number of goals scored during the 232 World Cup soccer games played from 1990 to 2002 and found them to be well modeled by the Poisson distribution. Only goals scored during the 90 minutes of regulation play were considered. The average number of goals scored each game was 2.5. Assuming that this mean continues to hold for other World Cup games, find the probabilities associated with the following events.

a At least six goals are scored during the 90 minutes of regulation play in a randomly selected game during the next World Cup.

b No goals are scored during the 90 minutes of regulations play in a randomly selected game during the next World Cup.

4.100 The number of fatalities due to shark attack during a year is modeled using a Poisson distribution. The International Shark Attack File (ISAF) investigates shark-human interactions worldwide. Internationally, an average of 4.4 fatalities per year occurred during a 5-year period. Assuming that this mean remains constant for the next 5 years, find the probabilities of the following events.

a No shark fatalities will be recorded in a given year.

b Sharks will cause at least six human deaths in a given year.

c No shark fatalities will occur during the 5-year period.

d At most 12 shark fatalities will occur during the 5-year period.

4.101 Schmuland (2001) explored the use of the Poisson distribution to model the number of goals the hockey star, Wayne Gretzky, scored during a game as an Edmonton Oiler. Gretzky played 696 games with the following distribution of the number of goals scored:

Points	0	1	2	3	4	5	6	7	8	9
Number of Games	69	155	171	143	79	57	14	6	2	0

a Find the average number of goals scored per game.

b Using the average found in part (a) and assuming that Gretzky's goals were scored according to a Poisson distribution, find the expected number of games in which $0, 1, 2, \ldots, \geq 9$ goals were scored.

c Compare the actual numbers of games with the expected numbers found in part (b). Does the Poisson seem to be a reasonable model?

4.102 The number of bacteria colonies of a certain type in samples of polluted water has a Poisson distribution with a mean of two per cubic centimeter.

a If four 1-cubic-centimeter samples of this water are independently selected, find the probability that at least one sample will contain one or more bacteria colonies.

b How many 1-cubic-centimeter samples should be selected to establish a probability of approximately 0.95 of containing at least one bacteria colony?

4.103 The number of days that a patient stays in a particular hospital is recorded. It is believed that the distribution of a patient's length of stay should be Poisson, but no stays of 0 lengths are recorded. Thus, let Y denote the length of stay and let

$$P(Y = y) = P(X = y|X > 0)$$

where X is a Poisson random variable with mean λ.

a Find the probability function of Y.

b Find the expected length of stay of patients at this hospital.

4.104 A food manufacturer uses an extruder (a machine that produces bite-size foods, like cookies and many snack foods) that has a revenue-producing value to the firm of \$300 per hour when it is in operation. However, the extruder breaks down an average of twice every 10 hours of operation. If X denotes the number of breakdowns during the time of operation, the revenue generated by the machine is given by

$$R = 300t - 75X^2$$

where t denotes hours of operations. The extruder is shut down for routine maintenance on a regular schedule, and it operates like a new machine after this maintenance. Find the optimal maintenance interval t_0 to maximize the expected revenue between shutdowns.

4.105 Let X have a Poisson distribution with mean λ. Find $E[X(X\ -1)]$, and use the result to show that $V(X) = \lambda$.

4.106 Let X be a Poisson random variable with mean λ. Find the value of λ that maximizes $P(X = k)$ for $k \geq 0$.

4.107 Let X be a Poisson random variable with mean λ. Show that

$$E\left(X^n\right) = \lambda\left[E(X+1)^{n-1}\right].$$

Use this result to find $E\left(X^2\right)$ and $V(X)$.

4.108 Let X be a Poisson random variable with mean λ. Find the probability that X is even in closed form (with no summation sign).

4.8 The Hypergeometric Distribution

4.8.1 The Probability Function

The distributions already discussed in this chapter have as their basic building block a series of *independent* Bernoulli trials. The examples, such as sampling from large lots, depict situations in which the trials of the experiment generate, for all practical purposes, independent outcomes. But suppose that we have a relatively small lot consisting of N items, of which k are defective. If two items are sampled sequentially, the outcomes for the second draw is significantly influenced by what happened on the first draw, provided that the first item drawn remains out of the lot. A new distribution must be developed to handle this situation involving *dependent* trials.

In general, suppose that a lot consists of N items, of which k are of one type (called *successes*) and $N - k$ are of another type (called *failures*). Suppose that n items are sampled randomly and sequentially from the lot, and suppose that none of the sampled items is replaced. (This is called *sampling without replacement*.) Let $Y_i = 1$ if the ith draw results in a success, and let $Y_i = 0$ otherwise, where $i = 1, 2, \ldots, n$. Let X denote the total number of successes among the n sampled items. To develop the probability distribution for X, let us start by looking at a special case for $X = x$. One way for successes to occur is to have

$$Y_1 = 1, \quad Y_2 = 1, \ldots, \quad Y_x = 1, \quad Y_{x+1} = 0, \ldots, \quad Y_n = 0.$$

We know that

$$P(Y_1 = 1, Y_2 = 1) = P(Y_1 = 1)P(Y_2 = 1 | Y_1 = 1)$$

and this result can be extended to give

$$\begin{aligned} &P(Y_1 = 1, Y_2 = 1, \ldots, Y_x = 1, Y_{x+1} = 0, \ldots, Y_n = 0) \\ &\quad = P(Y_1 = 1)P(Y_2 = 1 | Y_1 = 1)P(Y_3 = 1 | Y_2 = 1, Y_1 = 1)\cdots \\ &\qquad P(Y_n = 0 | Y_{n-1} = 0, \ldots, Y_{x+1} = 0, Y_x = 1, \ldots, Y_1 = 1). \end{aligned}$$

Now,

$$P(Y_1 = 1) = \frac{k}{N}$$

if the item is randomly selected; and similarly,

$$P(Y_2 = 1|Y_1 = 1) = \frac{k-1}{N-1}$$

because, at this point, one of the k successes has been removed. Using this idea repeatedly, we see that

$$\begin{aligned} &P(Y_1 = 1, Y_2 = 1, \dots, Y_x = 1, Y_{x+1} = 0, \dots, Y_n = 0) \\ &\quad = \left(\frac{k}{N}\right)\left(\frac{k-1}{N-1}\right)\cdots\left(\frac{k-x+1}{N-x+1}\right) \times \left(\frac{N-k}{N-x}\right)\cdots\left(\frac{N-k-n+y\backslash x+1}{N-n+1}\right)\end{aligned}$$

provided that $x \leq k$. A more compact way to write the preceding expression is to employ factorials, arriving at the formula

$$\frac{\dfrac{k!}{(k-x)!} \times \dfrac{(N-k)!}{(N-k-n+x)!}}{\dfrac{N!}{(N-n)!}}.$$

(The reader can check the equivalence of the two expressions.)

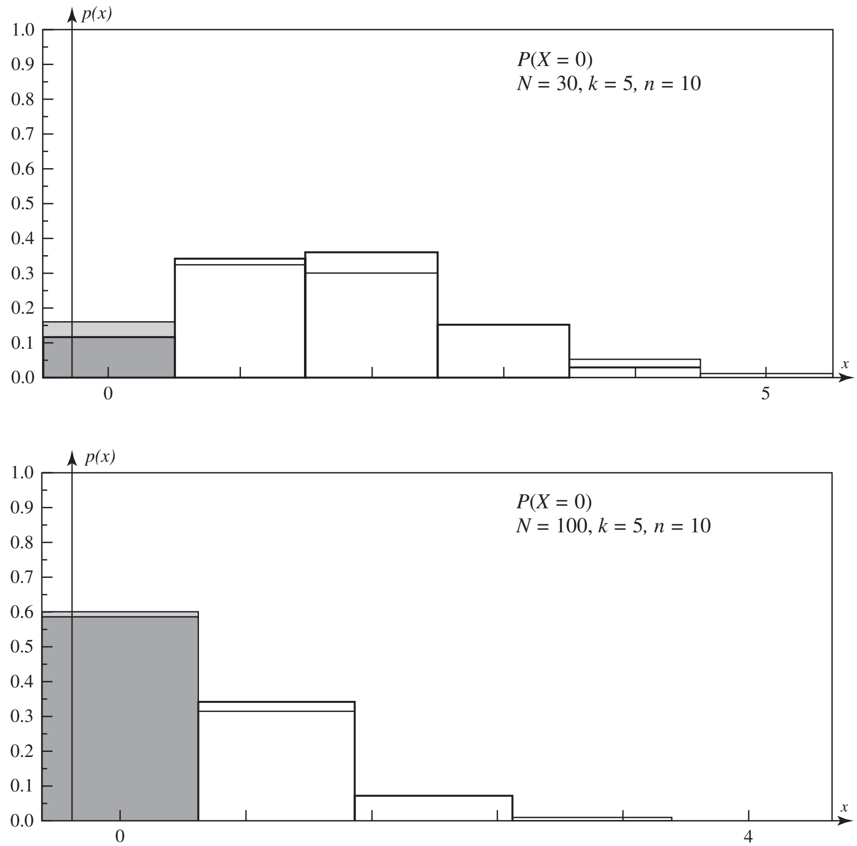
Any specified arrangement of x successes and $(n-x)$ failures will have the same probability as the one just derived for all successes followed by all failures; the terms will merely be rearranged. Thus, to find $P(X = x)$, we need only count how many of these arrangements are possible. Just as in the binomial case, the number of such arrangements is $\binom{n}{x}$. Hence, we have

$$\begin{aligned} P(X = x) &= \binom{n}{x} \frac{\frac{k!}{(k-x)!} \times \frac{(N-k)!}{(N-k-n+x)!}}{\frac{N!}{(N-n)!}} \\ &= \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}.\end{aligned}$$

Of course, $0 \leq x \leq k \leq N$ and $0 \leq x \leq n \leq N$. This formula is referred to as the *hypergeometric probability distribution*. Notice that it arises from a situation quite similar to the binomial, except that the trials here are *dependent*.

Finding the probabilities associated with the hypergeometric distribution can be computationally intensive, especially as N and n increase. When N is large, the probability of success changes little with each random selection from the population, and the binomial distribution can approximate the hypergeometric probabilities well. Because most populations are finite, many of the binomial applications in which sampling is without replacement involve finite populations so the distribution is actually hypergeometric. Yet, the population is so large that the difference in the hypergeometric and binomial probabilities is negligible. The *Approximations to Distributions* applet may be used to explore using the binomial distribution to approximate hypergeometric probabilities. As an example, for $N = 30$, $k = 5$, and $n = 10$, the Poisson

FIGURE 4.15
The binomial approximation to the hypergeometric distribution.



approximation of the probability of no success is 0.16150, which is not very close to the exact probability of 0.10869. However, if we increase N to 100 and keep $n = 10$ and $k = 5$, 0.59873, the Poisson approximation of the probability of zero is very close to 0.58375, the exact hypergeometric probability (Figure 4.15). The binomial approximation to the hypergeometric generally works well when $n/N < 0.10$.

Calculators and computer software are often valuable tools in determining hypergeometric probabilities. The shape of the hypergeometric distribution can be explored using the *Discrete Distributions* applet.

EXAMPLE 4.25 Two positions are open in a company. Ten men and five women have applied for a job at this company, and all are equally qualified for either position. The manager randomly hires two people from the applicant pool to fill the positions. What is the probability that a man and a woman were chosen.

Solution If the selections are made at random, and if X denotes the number of men selected, then the hypergeometric distribution would provide a good model for the behavior of X. Hence,

$$P(X = 1) = p(1) = \frac{\binom{10}{1}\binom{5}{1}}{\binom{15}{2}} = \frac{(10)(5)}{\left(\frac{15(14)}{2}\right)} = \frac{10}{21}.$$

Here, $N = 15,\ k = 10,\ n = 2$, and $x = 1$. ▪

4.8.2 Mean and Variance

The techniques needed to derive the mean and the variance of the hypergeometric distribution is given in Chapter 6. The results are

$$E(X) = n\left(\frac{k}{N}\right)$$
$$V(X) = n\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right).$$

Because the probability of selecting a success on one draw is k/n, the mean of the hypergeometric distribution has the same form as the mean of the binomial distribution. Likewise, the variance of the hypergeometric matches the variance of the binomial, multiplied by $(N - n)/(N - 1)$, which is a correction factor for dependent samples.

EXAMPLE 4.26 In an assembly-line production of industrial robots, gearbox assemblies can be installed in 1 minute each, if the holes have been properly drilled in the boxes, and in 10 minutes each, if the holes must be redrilled. There are 20 gearboxes in stock, and 2 of these have improperly drilled holes. From the 20 gearboxes available, 5 are selected randomly for installation in the next 5 robots in line.

1 Find the probability that all five gearboxes will fit properly.

2 Find the expected value, the variance, and the standard deviation of the time it will take to install these five gearboxes.

Solution **1** In this problem, $N = 20$; and the number of nonconforming boxes is $k = 2$, according to the manufacturer's usual standards. Let X denote the number of nonconforming boxes (i.e., the number with improperly drilled holes) in the sample of five. Then,

$$P(X = 0) = \frac{\binom{2}{0}\binom{18}{5}}{\binom{20}{5}}$$

$$= \frac{(1)(8568)}{15{,}504}$$
$$= 0.55.$$

The total time T taken to install the boxes (in minutes) is

$$\begin{aligned} T &= 10X + (5 - X) \\ &= 9X + 5 \end{aligned}$$

because each of the X nonconforming boxes takes 10 minutes to install, and the others take only 1 minute.

2 To find $E(T)$ and $V(T)$, we first need to calculate $E(X)$ and $V(X)$:

$$E(X) = n\left(\frac{k}{N}\right) = 5\left(\frac{2}{20}\right) = 0.5$$

and

$$\begin{aligned} V(X) &= n\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right) \\ &= 5(0.1)(1 - 0.1)\left(\frac{20-5}{20-1}\right) \\ &= 0.355. \end{aligned}$$

It follows that

$$\begin{aligned} E(T) &= 9E(X) + 5 \\ &= 9(0.5) + 5 \\ &= 9.5 \end{aligned}$$

and

$$\begin{aligned} V(T) &= (9)^2 V(X) \\ &= 81(0.355) \\ &= 28.755. \end{aligned}$$

Thus, installation time should average 9.5 minutes with a standard deviation of $\sqrt{28.755} = 5.4$ minutes. ▪

4.8.3 History and Applications

Although they did not use the term *hypergeometric distribution*, Bernoulli and de Moivre used the distribution to solve some of the probability problems they encountered. In 1899, Karl Pearson (Figure 4.16) discussed using the "hypergeometrical series" to model data. But it was not until 1936 that the term *hypergeometric distribution* actually appeared in the literature.

FIGURE 4.16
Karl Pearson.

The primary application of the hypergeometric distribution is in the study of finite populations. Although most populations are finite, many are large enough so that the probabilities are relatively stable as units are drawn. However, as the fraction of the population sampled becomes larger, the probabilities begin to change significantly with each new unit selected for the sample. The hypergeometric distribution has been used extensively in discrimination cases, quality control, and surveys.

The Hypergeometric Distribution

$$P(X = x) = \frac{\binom{k}{x}\binom{N-k}{n-x}}{\binom{N}{n}}, \quad x = 0,\ 1,\ \ldots, k \text{ with } \binom{b}{a} = 0 \text{ if } a > b$$

$$E(X) = n\left(\frac{k}{N}\right) \quad \text{and} \quad V(X) = n\left(\frac{k}{N}\right)\left(1 - \frac{k}{N}\right)\left(\frac{N-n}{N-1}\right)$$

Exercises

4.109 From a box containing five white and four red balls, two balls are selected at random without replacement. Find the probabilities of the following events.

a Exactly one white ball is selected.

b At least one white ball is selected.

c Two white balls are selected, given that at least one white ball is selected.

d The second ball drawn is white.

4.110 A small firm has 25 employees, of whom 8 are single and the other 17 are married. The owner is contemplating a change in insurance coverage, and she randomly selects five people to get their opinions.

a What is the probability that she only talks to single people?

b What is the probability that she only talks to married people?

c What is the probability that two single and three married people are in the sample?

4.111 An auditor checking the accounting practices of a firm samples 4 accounts from an accounts receivable list of 12. Find the probability that the auditor sees at least one past-due account under the following conditions.

a There are 2 such accounts among the 12.

b There are 6 such accounts among the 12.

c There are 8 such accounts among the 12.

4.112 The pool of qualified jurors called for a high-profile case has 12 whites, 9 blacks, 4 Hispanics, and 2 Asians. From these, 12 will be selected to serve on the jury. Assume that all qualified jurors meet the criteria for serving. Find the probabilities of the following events.

a No white is on the jury.

b All nine blacks serve on the jury.

c No Hispanics or Asians serve on the jury.

4.113 A foreman has 10 employees, and the company has just told him that he must terminate 4 of them. Of the 10 employees, 3 belong to a minority ethnic group. The foreman selected all three minority employees (plus one other) to terminate. The minority employees then protested to the union steward that they were discriminated against by the foreman. The foreman claimed that the selection had been completely random. What do you think? Justify your answer.

4.114 A student has a drawer with 20 AA batteries. However, the student does not realize that three of the batteries have lost their charge (will not work). She realizes that the batteries on her calculator are no longer working, and she is in a hurry to get to a test. She grabs two batteries from the drawer at random to replace the two batteries in her calculator. What is the probability that she will get two good batteries so that her calculator will work during the test?

4.115 A corporation has a pool of six firms (four of which are local) from which they can purchase certain supplies. If three firms are randomly selected without replacement, find the probabilities of the following events.

a All three selected firms are local.

b At least one selected firm is not local.

4.116 An eight-cylinder automobile engine has two misfiring spark plugs. If all four plugs are removed from one side of the engine, what is the probability that the two misfiring plugs are among them?

4.117 Two assembly lines (I and II) have the same rate of defectives in their production of voltage regulators. Five regulators are sampled from each line and tested. Among the total of 10 tested regulators, 4 are defective. Find the probability that exactly two of the defectives came from line I.

4.118 A group of four men and six women are faced with an extremely difficult and unpleasant task requiring three people. They decide to draw straws to see who must do the job. Eight long and two short straws are made. A person outside the group is asked to hold the straws, and each person in the group selects one. Find the probability of the following events.

a Both short straws are drawn by men.

b Both short straws are drawn by women.

c The short straws are drawn by a man and a woman.

4.119 Specifications call for a type of thermistor to test out at between 900 and 1000 ohms at 25°C. Ten thermistors are available, and three of these are to be selected for use. Let X denote the number among the three that do not conform to specifications. Find the probability distribution for X (in tabular form) if the following conditions prevail.

a The 10 contain 2 thermistors not conforming to specifications.

b The 10 contain 4 thermistors not conforming to specifications.

4.120 A company has 10 personal computers (PCs) in its warehouse. Although all are new and still in boxes, four do not currently function properly. One of the company's offices requests 5 PCs, and the warehouse foreman selects 5 from the stock of 10 and ships them to the requesting office. What is the probability that all five of the PCs are not defective?

4.121 Referring to Exercise 4.120, the office requesting the PCs returns the defective ones for repair.

a If it costs \$80 to repair each PC, find the mean and the variance of the distribution of the total repair cost.

b By Tchebysheff's Theorem, in what interval should you expect, with a probability of at least 0.95, the repair costs of these five PCs to lie?

4.122 The following problem is based on probability calculations that Shuster (1991) performed for a court case. In an earlier cocaine "bust," Florida police seized 496 packets alleged to be cocaine. Of the 496 packets, 4 were tested, and all 4 were found to contain cocaine. For this case, the police randomly selected two of the remaining packets and, posing as drug dealers, sold them to the defendant. Between the time of the sell and the time of the arrest, the defendant disposed of the two packets. The defense attorney argued that the defendant's packets had not been tested and that those packets could have been negative for cocaine. The attorney suggested that the packets could have been switched in the evidence room, or that only some of the original $N = 496$ packets contained cocaine.

a If the original 496 packets had k cocaine packets and $M = 496 - k$ noncocaine packets, show that the probability of selecting four cocaine packets and then two noncocaine packets, which is the probability that the defendant is innocent of buying cocaine, is

$$\frac{\binom{k}{4}\binom{M}{2}}{\binom{k+M}{4}\binom{k+M-4}{2}}.$$

b Show that the maximum probability in part (a) occurs when $k = 331$ and $M = 165$. [Hint: Let $Q(k, M)$ represent the probability in part (a). Find the values of k and M for which $Q(k+1,\ M-1)/Q(k,\ M) > 1$.] This probability is 0.022.

4.123 Lot acceptance sampling procedures for an electronics manufacturing firm call for sampling n items from a lot of N items and accepting the lot if $X \le c$, where X is the number of nonconforming items in the sample. For an incoming lot of 20 transistors, 5 are to be sampled. Find the probability of accepting the lot if $c = 1$ and the actual number of nonconforming transistors in the lot are as follows:

a 0 **b** 1 **c** 2 **d** 3

4.124 A 10-acre area has N raccoons. Ten of these raccoons were captured, marked so they could be recognized, and then released. After 5 days, 20 raccoons were captured. Let X denote the number of those captured on the second occasion that was marked during the first sampling occasion. Suppose that captures at both time points can be treated as random selections from the population and that the same N raccoons were in the area on both sampling occasions (no additions or deletions).

a If $N = 30$, what is the probability that no more than 5 of those captured during the second sampling period were marked during the first sampling occasion?

b If eight raccoons in the second sample were marked from having been caught in the first one, what value of N would result in the probability of this happening being the largest?

4.9 The Moment-Generating Function

We saw in earlier sections that, if $g(X)$ is a function of a random variable X, with probability distribution given by $p(x)$, then

$$E[g(X)] = \sum_{y} g(x)p(x).$$

A special function with many theoretical uses in probability theory is the expected value of e^{tX}, for a random variable X, and this expected value is called the moment-generating function (mgf). The definition of a mgf is given in Definition 4.7.

DEFINITION **4.7**

The **moment-generating function** (mgf) of a random variable is denoted by $M(t)$ and defined to be

$$M(t) = E(e^{tX}). \quad \blacksquare$$

The expected values of powers of a random variable are often called moments. Thus, $E(X)$ is the first moment of X, and $E(X^2)$ is the second moment of X. One use for the mgfs is that, in fact, it does generate moments of X. When $M(t)$ exists, it is differentiable in a neighborhood of the origin $t = 0$, and the derivatives may be taken inside the expectation. Thus

$$\begin{aligned} M^{(1)}(t) &= \frac{dM(t)}{dt} \\ &= \frac{d}{dt}E(e^{tX}) \\ &= E\left[\frac{d}{dt}e^{tX}\right] \\ &= E\left[Xe^{tX}\right]. \end{aligned}$$

Now, if we set $t = 0$, we have

$$M^{(1)}(0) = E(X).$$

Going on to the second derivative,

$$M^{(2)}(t) = E(X^2 e^{tX})$$

and

$$M^{(2)}(0) = E(X^2)$$

In general,

$$M^{(k)}(0) = E(X^k).$$

It often is easier to evaluate $M(t)$ and its derivatives than to find the moments of the random variable directly. Other theoretical uses of the mgf are discussed in later chapters.

EXAMPLE 4.27 Evaluate the mgf for the geometric distribution, and use it to find the mean and the variance of this distribution.

Solution For the geometric variable X, we have

$$\begin{aligned}M(t) &= E(e^{tX}) \\ &= \sum_{x=0}^{\infty} e^{tx} pq^x \\ &= p \sum_{x=0}^{\infty} \left(qe^t\right)^x \\ &= p\left[1 + qe^t + \left(qe^t\right)^2 + \cdots\right] \\ &= p\left(\frac{1}{1-qe^t}\right) \\ &= \frac{p}{1-qe^t}\end{aligned}$$

because the series is geometric with a common ratio of qe^t. Note: For the series to converge and the mgf to exist, we must have $qe^t < 1$, which is the case if $t < \ln(1/q)$. It is important for the mgf to exist for t in a neighborhood about 0, and it does here.

To evaluate the mean, we have

$$\begin{aligned}M^{(1)}(t) &= \frac{0 + pqe^t}{\left(1-qe^t\right)^2} \\ &= \frac{pqe^t}{\left(1-qe^t\right)^2}\end{aligned}$$

and

$$E(X) = M^{(1)}(0) = \frac{pq}{(1-q)^2} = \frac{q}{p}.$$

To evaluate the variance, we first need

$$E\left(X^2\right) = M^{(2)}(t).$$

Now,

$$\begin{aligned}M^{(2)}(t) &= \frac{\left(1-qe^t\right)^2 pqe^t - pqe^t(2)\left(1-qe^t\right)(-1)qe^t}{\left(1-qe^t\right)^4} \\ &= \frac{pqe^t + pq^2e^{2t}}{\left(1-qe^t\right)^3}\end{aligned}$$

and

$$E\left(X^2\right) = M^{(2)}(0) = \frac{pq + pq^2}{(1-q)^3} = \frac{q(1+q)}{p^2}.$$

Hence,

$$\begin{aligned} V(X) &= E\left(X^2\right) - [E(X)]^2 \\ &= \frac{q(1+q)}{p^2} - \left(\frac{q}{p}\right)^2 \\ &= \frac{q}{p^2}. \quad \blacksquare \end{aligned}$$

As we have seen in this section, the mgf is aptly named because it can be used to find the moments of a probability distribution. Sometimes, finding the mean and variance using the mgf is much easier than it would be to find them using the probability density function as suggested in their definitions. Mgfs have other useful properties. As an example they can be used to find the probability distributions of random variables. Additional properties are discussed in detail in Chapters 5 and 6, but one such property is given in Exercise 4.137.

4.10 The Probability-Generating Function

In an important class of discrete random variables, X takes integral values ($X = 0, 1, 2, 3, \ldots$) and, consequently, represents a count. The binomial, geometric, hypergeometric, and Poisson random variables all fall in this class. The following examples present practical situations involving integral-valued random variables. One, tied to the theory of queues (waiting lines), is concerned with the number of persons (or objects) awaiting service at a particular point in time. Understanding the behavior of this random variable is important in designing manufacturing plants where production consists of a sequence of operations, each of which requires a different length of time to complete. An insufficient number of service stations for a particular production operation can result in a bottleneck—the forming of a queue of products waiting to be serviced—which slows down the entire manufacturing operation. Queuing theory is also important in determining the number of checkout counters needed for a supermarket and in designing hospitals and clinics.

Integer-valued random variables are extremely important in studies of population growth, too. For example, epidemiologists are interested in the growth of bacterial populations and also in the growth of the number of persons afflicted by a particular disease. The number of elements in each of these populations is an integral-valued random variable.

A mathematical device that is very useful in finding the probability distributions and other properties of integral-valued random variables is the probability-generating function $P(t)$, which is defined in Definition 4.8.

DEFINITION 4.8

The **probability-generating function** of a random variable is denoted by $P(t)$ and is defined to be

$$P(t) = E(t^X). \quad ▪$$

If X is an integer-valued random variable, with

$$P(X = i) = p_i, \quad i = 0, 1, 2, \ldots$$

Then

$$P(t) = E(t^X) = p_0 + p_1 t + p_2 t^2 + \cdots$$

The reason for calling $P(t)$ a probability–generating function is clear when we compare $P(t)$ with the mgf $M(t)$. Particularly, the coefficient of t^i in $P(t)$ is the probability p_i. If we know $P(t)$ and can expand it into a series, we can determine $p(x)$ as the coefficient of t^x. Repeated differentiation of $P(t)$ yields *factorial moments* for the random variable X.

DEFINITION 4.9

The kth **factorial moment** for a random variable X is defined to be

$$\mu_{[k]} = E[X(X-1)(X-2)\cdots(X-k+1)]$$

where k is a positive integer. ▪

When a probability-generating function exists, it can be differentiated in a neighborhood of $t = 1$. Thus, with

$$P(t) = E(t^X)$$

we have

$$\begin{aligned} P^{(1)}(t) &= \frac{dP(t)}{dt} \\ &= \frac{d}{dt}E(t^X) \\ &= E\left[\frac{d}{dt}t^X\right] \\ &= E\left[Xt^{X-1}\right]. \end{aligned}$$

Setting $t = 1$, we have

$$P^{(1)}(1) = E(X).$$

Similarly,

$$P^{(2)}(t) = E[X(X-1)t^{X-2}]$$

and

$$P^{(2)}(1) = E[X(X-1)].$$

In general,

$$P^{(k)}(t) = E[X(X-1)\cdots(X-k+1)t^{X-k}]$$

and

$$\begin{aligned} P^{(k)}(1) &= E[X(X-1)\cdots(X-k+1)] \\ &= \mu_{[k]}. \end{aligned}$$

EXAMPLE 4.28 Find the probability-generating function for the geometric random variable, and use this function to find the mean.

Solution

$$\begin{aligned} P(t) = E\left(t^X\right) &= \sum_{x=0}^{\infty} t^x p q^x \\ &= p\sum_{x=0}^{\infty} (qt)^x \\ &= \frac{p}{1-qt} \end{aligned}$$

where $qt < 1$ for the series to converge. Now,

$$P^{(1)}(t) = \frac{d}{dt}\left(\frac{p}{1-qt}\right) = \frac{pq}{(1-qt)^2}$$

Setting $t = 1$,

$$P^{(1)}(1) = \frac{pq}{(1-q)^2} = \frac{q}{p}$$

which is the mean of a geometric random variable. ▪

Because we already have the mgf to assist us in finding the moment of a random variable, we might ask how knowing $P(t)$ can help us. The answer is that in some instances it may be exceedingly difficult to find $M(t)$ but easy to find $P(t)$. Alternatively, $M(t)$ may be easier to work with in a particular setting. Thus, $P(t)$ simply provides an additional tool for finding the moments of a random variable. It may or may not be useful in a given situation.

Finding the moments of a random variable is not the major use of the probability-generating function. Its primary application is in deriving the probability function (and hence the probability distribution) for related integral-valued random variables. For these applications, see Feller (1968), Parzen (1992), and Section 7.7.

Exercises

4.125 Find the moment-generating function for the Bernoulli random variable.

4.126 Show that the moment-generating function for the binomial random variable is given by

$$M(t) = \left(pe^t + q\right)^n.$$

4.127 Show that the moment-generating function for the Poisson random variable with mean λ is given by

$$M(t) = e^{\lambda(e^t - 1)}.$$

4.128 Show that the moment-generating function for the negative binomial random variable is given by

$$M(t) = \left(\frac{p}{1 - qe^t}\right)^r.$$

4.129 Derive the mean and variance of the Bernoulli random variable using the moment generating function derived in Exercise 4.125.

4.130 Derive the mean and variance of the binomial random variable using the moment-generating function derived in Exercise 4.126.

4.131 Derive the mean and variance of the Poisson random variable using the moment-generating function derived in Exercise 4.127.

4.132 Derive the mean and variance of the negative binomial random variable using the moment-generating function derived in Exercise 4.128.

4.133 Derive the probability-generating function of a Poisson random variable with parameter λ.

4.134 Derive the probability-generating function of the binomial random variable of n trials with probability of success p.

4.135 Using the probability-generating function derived in Exercise 4.133, find the first and second factorial moments of a Poisson random variable. From the first two factorial moments, find the mean and variance of the Poisson random variable.

4.136 Find the first and second factorial moments of the binomial random variable in Exercise 4.134. Using the first two factorial moments, find the mean and variance of the Poisson random.

4.137 If X is a random variable with moment-generating function $M(t)$, and Y is a function of X given by $Y = aX + b$, show that the moment-generating function for Y is $e^{tb}M(at)$.

4.138 Use the result of Exercise 4.137 to show that

$$E(Y) = aE(X) + b$$

and

$$V(Y) = a^2 V(X).$$

4.11 Markov Chains

Consider a system that can be in any of a finite number of states. Assume that the system moves from state to state according to some prescribed probability law. The system, for example, could record weather conditions from day to day, with the possible states being clear, partly cloudy, and cloudy. Observing conditions over a long period would allow one to find the probability of its being clear tomorrow given that it is partly cloudy today.

Let X_i denote the state of the system at time point i, and let the possible states be denoted by $S_1, \ldots, S_m$, for a finite integer m. We are interested not in the elapsed time between transitions from one time state to another, but only in the states and the probabilities of going from one state to another—that is, in the *transition probabilities*. We assume that

$$P(X_i = S_k \mid X_{i-1} = S_j) = p_{jk}$$

where p_{jk} is the transition probability from S_j to S_k; and this probability is independent of i. Thus, the transition probabilities depend not on the time points, but only on the states. The event $(X_i = S_k | X_{i-1} = S_j)$ is assumed to be independent of the past history of the process. Such a process is called a Markov chain with stationary transition probabilities. The transition probabilities can conveniently be displayed in a matrix:

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{bmatrix}$$

Let X_0 denote the starting state of the system, with probabilities given by

$$P_k^{(0)} = P(X_0 = S_k)$$

and let the probability of being in state S_k after n steps be given by $p_k^{(n)}$. These probabilities are conveniently displayed by vectors:

$$\boldsymbol{p}^{(0)} = [p_1^{(0)}, p_2^{(0)}, \ldots, p_m^{(0)}]$$

and

$$\boldsymbol{p}^{(n)} = [p_1^{(n)}, p_2^{(n)}, \ldots, p_m^{(n)}]$$

To see how $p^{(0)}$ and $p^{(1)}$ are related, consider a Markov chain with only two states, so that

$$\boldsymbol{P} = \begin{bmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{bmatrix}.$$

There are two ways to get to state 1 after one step: Either the chain starts in state 1 and stays there, or the chain starts in state 2 and then moves to state 1 in one step. Thus,

$$\boldsymbol{p}_1^{(1)} = p_1^{(0)} p_{11} + p_2^{(0)} p_{21}.$$

Similarly,

$$p_2^{(1)} = p_1^{(0)} p_{12} + p_2^{(0)} p_{22}.$$

In terms of matrix multiplication,

$$\boldsymbol{p}^{(1)} = \boldsymbol{p}^{(0)} \boldsymbol{P}$$

and, in general,

$$\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)} \boldsymbol{P}.$$

$\boldsymbol{P}$ is said to be regular if some power of $\boldsymbol{P}(\boldsymbol{P}^n$ for some $n)$ has all positive entries. Thus, one can get from state S_j to state S_k, eventually, for any pair (j, k). (Notice that the condition of regularity rules out certain chains that periodically return to certain states.) If $\boldsymbol{P}$ is regular, the chain has a stationary (or equilibrium) distribution that gives the probabilities of its being in the respective states after many transitions have evolved. In other words, $p_j^{(n)}$ must have a limit π_j, as $n \to \infty$. Suppose that such limits exist; then $\pi = (\pi_1, \ldots, \pi_n)$ must satisfy

$$\pi = \pi \boldsymbol{P}$$

because $\boldsymbol{p}^{(n)} = \boldsymbol{p}^{(n-1)} \boldsymbol{P}$ will have the same limit as $\boldsymbol{p}^{(n-1)}$.

EXAMPLE 4.29 A supermarket stocks three brands of coffee—A, B, and C—and customers switch from brand to brand according to the transition matrix

$$\boldsymbol{P} = \begin{bmatrix} 3/4 & 1/4 & 0 \\ 0 & 2/3 & 1/3 \\ 1/4 & 1/4 & 1/2 \end{bmatrix}$$

where S_1 corresponds to a purchase of brand A, S_2 to brand B, and S_3 to brand C; that is, 3/4 of the customers buying brand A also buy brand A the next time they purchase coffee, whereas 1/4 of these customers switch to brand B.

1 Find the probability that a customer who buys brand A today will again purchase brand A 2 weeks from today, assuming that he or she purchases coffee once a week.

2 In the long run, what fractions of customers purchase the respective brands?

Solution **1** Assuming that the customer is chosen at random, his or her transition probabilities are given by $\boldsymbol{P}$. The given information indicates that $\boldsymbol{p}^{(0)} = (1, 0, 0)$; that is, the customer starts with a purchase of brand A. Then

$$\boldsymbol{p}^{(1)} = \boldsymbol{p}^{(0)} P = \left(\frac{3}{4}, \frac{1}{4}, 0\right)$$

gives the probabilities for the next week's purchase. The probabilities for 2 weeks from now are given by

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(1)} \boldsymbol{P} = \left(\frac{9}{16}, \frac{17}{48}, \frac{1}{12}\right).$$

That is, the chance of the customer's purchasing brand A 2 weeks from now is only 9/16.

2 The answer to the long-run frequency ratio is given by π, which is the stationary distribution. The equation

$$\boldsymbol{\pi} = \boldsymbol{\pi P}$$

yields the system of equations

$$\pi_1 = \left(\frac{3}{4}\right)\pi_1 + \left(\frac{1}{4}\right)\pi_3$$
$$\pi_2 = \left(\frac{1}{4}\right)\pi_1 + \left(\frac{2}{3}\right)\pi_2 + \left(\frac{1}{4}\right)\pi_3$$
$$\pi_3 = \left(\frac{1}{3}\right)\pi_2 + \left(\frac{1}{2}\right)\pi_3$$

Combining these equations with the fact that $\pi_1 + \pi_2 + \pi_3 = 1$ yields

$$\boldsymbol{\pi} = \left(\frac{2}{7}, \frac{3}{7}, \frac{2}{7}\right).$$

Thus, the store should stock more brand B coffee than either brand A or brand C. ▪

EXAMPLE 4.30 Markov chains are used in the study of probabilities connected to genetic models. Recall from Section 3.2 that genes come in pairs. For any trait governed by a pair of genes, an individual may have genes that are homozygous dominant (GG), heterozygous (Gg), or homozygous recessive (gg). Each offspring inherits one gene of a pair from each parent, at random and independently.

Suppose that an individual of unknown genetic makeup is mated with a heterozygous individual. Set up a transition matrix to describe the possible states of a resulting offspring and their probabilities. What will happen to the genetic makeup of the offspring after many generations of mating with a heterozygous individual?

Solution If the unknown is homozygous dominant (GG) and is mated with a heterozygous individual (Gg), the offspring has a probability of 1/2 of being homozygous dominant and a probability of 1/2 of being heterozygous. If two heterozygous individuals are mated, the offspring may be homozygous dominant, heterozygous, or homozygous recessive with probabilities 1/4, 1/2, and 1/4, respectively. If the unknown is homozygous recessive (gg), the offspring of it and a heterozygous individual has a probability of 1/2 of being homozygous recessive and a probability of 1/2 of being heterozygous. Following along these lines, a transition matrix from the unknown parent to an offspring is given by

$$\boldsymbol{P} = \begin{array}{c} \\ d \\ h \\ r \end{array} \begin{array}{c} \begin{array}{ccc} d & h & r \end{array} \\ \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix} \end{array}$$

where d, h, and r represent homozygous dominant, heterozygous, and homozygous recessive, respectively.

The matrix $\boldsymbol{P}^2$ has all positive entries ($\boldsymbol{P}$ is regular); and hence a stationary distribution exists. From the matrix equation

$$\boldsymbol{\pi} = \boldsymbol{\pi} \boldsymbol{P}$$

we obtain

$$\pi_1 = \left(\frac{1}{2}\right)\pi_1 + \left(\frac{1}{2}\right)\pi_2$$

$$\pi_2 = \left(\frac{1}{2}\right)\pi_1 + \left(\frac{1}{2}\right)\pi_2 + \left(\frac{1}{2}\right)\pi_3$$

$$\pi_3 = \left(\frac{1}{2}\right)\pi_2 + \left(\frac{1}{2}\right)\pi_3.$$

Because $\pi_1 + \pi_2 + \pi_3 = 1$, the second equation yields $\pi_2 = \frac{1}{2}$. It is then easy to establish that $\pi_1 = \pi_3 = \frac{1}{4}$. Thus,

$$\pi = \left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right).$$

No matter what the genetic makeup of the unknown parent happened to be, the ratio of homozygous dominant to heterozygous to homozygous recessive offspring among its descendants, after many generations of mating with heterozygous individuals, should be 1:2:1. ▪

An interesting example of a transition matrix that is not regular is formed by a Markov chain with absorbing states. A state S_i is said to be absorbing if $p_{ii} = 1$ and $p_{ij} = 0$ for $j \neq i$. That is, once the system is in state S_i, it cannot leave it. The transition matrix for such a chain can always be arranged in a standard form, with the absorbing states listed first. For example, suppose that a chain has five states, of which two are absorbing. Then P can be written as

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} \end{bmatrix} = \begin{bmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{R} & \boldsymbol{Q} \end{bmatrix}.$$

where $\boldsymbol{I}$ is a 2×2 identity matrix and $\boldsymbol{0}$ is a matrix of zeroes. Such a transition matrix is not regular. Many interesting properties of these chains can be expressed in terms of $\boldsymbol{R}$ and $\boldsymbol{Q}$ (see Kemeny and Snell 1983).

The following discussion is restricted to the case in which $\boldsymbol{R}$ and $\boldsymbol{Q}$ are such that it is possible to get to an absorbing state from every other state eventually. In that case, the Markov chain eventually will end up in an absorbing state. Questions of interest then involve the expected number of steps to absorption and the probability of absorption in the various absorbing states.

Let m_{ij} denote the expected (or mean) number of times the system is in state S_j, given that it started in S_i, for nonabsorbing states S_i and S_j. From S_i, the system could go to an absorbing state in one step, or it could go to a nonabsorbing state—say S_k—and eventually be absorbed from there. Thus, m_{ij} must satisfy

$$m_{ij} = \partial_{ij} + \sum_k p_{ik} m_{kj}.$$

Where the summation is over all nonabsorbing states and

$$\partial_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}$$

The term ∂_{ij} accounts for the fact that if the chain goes to an absorbing state in one step, it was in state S_i one time.

If we denote the matrix of m_{ij} terms by $\boldsymbol{M}$, the preceding equation can then be generalized to

$$\boldsymbol{M} = \boldsymbol{I} + \boldsymbol{QM}$$

or

$$\boldsymbol{M} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$$

Matrix operations, such as inversion, are not discussed here. (The equations can be solved directly if matrix operations are unfamiliar to the reader.)

The expected number of steps to absorption from the nonabsorbing starting state S_i, is denoted by m_i and given simply by

$$m_i = \sum_k m_{ik}$$

again summing over nonabsorbing states.

Turning now to the probability of absorption into the various absorbing states, we let a_{ij} denote the probability of the system's being absorbed in state S_j, given that it started in state S_i for nonabsorbing S_i and *absorbing* S_j. Repeating the preceding argument, the system could move to S_j in one step, or it could move to a nonabsorbing state S_k and be absorbed from there. Thus, a_{ij} satisfies

$$a_{ij} = p_{ij} + \sum_k p_{ik} a_{kj}$$

where the summation occurs over the nonabsorbing states. If we denote the matrix of a_{ij} terms by A, the preceding equation then generalizes to

$$\boldsymbol{A} = \boldsymbol{R} + \boldsymbol{Q}\boldsymbol{A}$$

or

$$\begin{aligned} \boldsymbol{A} &= (\boldsymbol{I} - \boldsymbol{Q})^{-1}\boldsymbol{R} \\ &= \boldsymbol{M}\boldsymbol{R} \end{aligned}$$

The following example illustrates the computations.

EXAMPLE 4.31 A manager of one section of a plant has different employees working at level I and at level II. New employees may enter his section at either level. At the end of each year, the performance of each employee is evaluated; employees can be reassigned to their level I or II jobs, terminated, or promoted to level III, in which case they never go back to I or II. The manager can keep track of employee movement as a Markov chain. The absorbing states are termination (S_1) and employment at level III (S_2); the nonabsorbing states are employment at level I (S_3) and employment at level II (S_4). Records over a long period indicate that the following is a reasonable assignment of probabilities:

$$\boldsymbol{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0.2 & 0.1 & 0.2 & 0.5 \\ 0.1 & 0.3 & 0.1 & 0.5 \end{bmatrix}$$

Thus, if an employee enters at a level I job, the probability is 0.5 that she will jump to level II work at the end of the year, but the probability is 0.2 that she will be terminated.

1 Find the expected number of evaluations an employee must go through in this section before either being terminated or promoted to level III.

2 Find the probabilities of being terminated or promoted to level III eventually.

Solution **1** For the $\boldsymbol{P}$ matrix,

$$\boldsymbol{R} = \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.3 \end{bmatrix}$$

and

$$Q = \begin{bmatrix} 0.2 & 0.5 \\ 0.1 & 0.5 \end{bmatrix}.$$

Thus,

$$I - Q = \begin{bmatrix} 0.8 & -0.5 \\ -0.1 & 0.5 \end{bmatrix}$$

and

$$M = (I - Q)^{-1} = \begin{bmatrix} 10/7 & 10/7 \\ 2/7 & 16/7 \end{bmatrix} = \begin{bmatrix} m_{33} & m_{34} \\ m_{43} & m_{44} \end{bmatrix}.$$

It follows that

$$m_3 = \frac{20}{7} \quad \text{and} \quad m_4 = \frac{18}{7}.$$

In other words, a new employee in this section can expect to remain there through 20/7 evaluation periods if she enters at level I, whereas she can expect to remain there through 18/7 evaluations if she enters at level II.

2 The fact that

$$A = MR = \begin{bmatrix} 3/7 & 4/7 \\ 2/7 & 5/7 \end{bmatrix} = \begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix}$$

implies that an employee entering at level I has a probability of 4/7 of reaching level III, whereas an employee entering at level II has a probability of 5/7 of reaching level III. The probabilities of termination at levels I and II are therefore 3/7 and 2/7, respectively. ▪

EXAMPLE 4.32 Continuing the genetics problem in Example 4.30, suppose that an individual of unknown genetic makeup is mated with a known homozygous dominant (GG) individual. The matrix of transition probabilities for the first-generation offspring then becomes

$$P = \begin{array}{c} \\ d \\ h \\ r \end{array} \begin{array}{c} \begin{array}{ccc} d & h & r \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

which has one absorbing state. If the offspring of each generation marry a homozygous dominant individual, find the mean number of generations until all offspring become dominant.

Solution In the notation used earlier,

$$Q = \begin{bmatrix} \frac{1}{2} & 0 \\ 1 & 0 \end{bmatrix}$$

$$I - Q = \begin{bmatrix} \frac{1}{2} & 0 \\ -1 & 1 \end{bmatrix}$$

and

$$(I - Q)^{-1} = \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix} = M.$$

Thus, if the unknown is heterozygous, we should expect all offspring to be dominant after two generations. If the unknown is homozygous recessive, we should expect all offspring to be dominant after three generations. Notice that

$$A = MR = \begin{bmatrix} 2 & 0 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

which simply indicates that we are guaranteed to reach the fully dominant state eventually, no matter what the genetic makeup of the unknown parent may be. ▪

Exercises

4.139 A certain city prides itself on having sunny days. If it rains one day, there is a 90% chance that it will be sunny the next day. If it is sunny one day, there is a 30% chance that it will rain the following day. (Assume that there are only sunny or rainy days.) Does the city have sunny days most of the time? In the long run, what fraction of all days is sunny?

4.140 Suppose a car rental agency in a large city has three locations: a downtown location (labeled A), an airport location (labeled B), and a hotel location (labeled C). The agency has a group of delivery drivers to serve all three locations. Of the calls to the downtown location, 30% are delivered in the downtown area, 30% are delivered to the airport, and 40% are delivered to the hotel. Of the calls to the airport location, 40% are delivered in the downtown area, 40% are delivered to the airport, and 20% are delivered to the hotel. Of the calls to the hotel location, 50% are delivered in the downtown area, 30% are delivered to the airport area, and 20% are delivered to the hotel area. After making a delivery, a driver goes to the nearest location to make the next delivery. This way, the location of a specific driver is determined only by his or her previous location.

a Give the transition matrix.

b Find the probability that a driver who begins in the downtown location will be at the hotel after two deliveries.

c In the long run, what fraction of the total number of stops does a driver make at each of the three locations?

4.141 For a Markov chain, show that $P^{(n)} = P^{(n-1)}P$.

4.142 Suppose that a particle moves in unit steps along a straight line. At each step, the particle either remains where it is, moves one step to the right, or moves one step to the left. The line along which the particle moves has barriers at 0 and at b, a positive integer, and the particle only moves between these barriers; it is absorbed if it lands on either barrier. Now, suppose that the particle moves to the right with probability p and to the left with probability $1 - p = q$.

a Set up the general form of the transition matrix for a particle in this system.

b For the case $b = 3$ and $p \neq q$, show that the absorption probabilities are as follows:

$$a_{10} = \frac{\left(\frac{q}{p}\right) - \left(\frac{q}{p}\right)^3}{1 - \left(\frac{q}{p}\right)^3}$$

$$a_{20} = \frac{\left(\frac{q}{p}\right)^2 - \left(\frac{q}{p}\right)^3}{1 - \left(\frac{q}{p}\right)^3}$$

c In general, it can be shown that

$$a_{10} = \frac{\left(\frac{q}{p}\right)^1 - \left(\frac{q}{p}\right)^b}{1 - \left(\frac{q}{p}\right)^b}, \quad p \neq q.$$

By taking the limit of a_{j0} as $p \to 1/2$, show that

$$a_{j0} = \frac{b - j}{b}, \quad p = q.$$

d For the case $b = 3$, find an expression for the mean time to absorption from state j with $p \neq q$. Can you generalize this result?

4.143 Suppose that n white balls and n black balls are placed in two urns so that each urn contains n balls. A ball is randomly selected from each urn and placed in the opposite urn. (This is one possible model for the diffusion of gases.)

a Number the urns 1 and 2. The state of the system is the number of black balls in urn 1. Show that the transition probabilities are given by the following quantities:

$$p_{jj-1} = \left(\frac{j}{n}\right)^2, \quad j > 0$$

$$p_{jj} = \frac{2j(n-j)}{n^2}$$

$$p_{jj+1} = \left(\frac{n-j}{n}\right)^2, \quad j < n$$

$$p_{jk} = 0, \quad \text{otherwise}$$

b After many transitions, show that the stationary distribution is satisfied by

$$p_j = \frac{\binom{n}{j}^2}{\binom{2n}{n}}.$$

Give an intuitive argument as to why this looks like a reasonable answer.

4.144 Suppose that two friends, A and B, toss a balanced coin. If the coin comes up heads, A wins \$1 from B. If it comes up tails, B wins \$1 from A. The game ends only when one player has all the other's money. If A starts with \$1 and B with \$3, find the expected duration of the game and the probability that A will win.

4.12 Summary

The outcomes of interest in most investigations involving random events are numerical. The simplest numbered outcomes to model are counts, such as the number of nonconforming parts in a shipment, the number of sunny days in a month, or the number of water samples that contain a pollutant. One amazing result of probability theory is the fact that a small number of theoretical distributions can cover a wide array of applications. Six of the most useful discrete probability distributions are introduced in this chapter.

The *Bernoulli* random variable is simply an indicator random variable; it uses a numerical code to indicate the presence or absence of a characteristic.

The *binomial* random variable counts the number of "successes" among a fixed number n of independent events, each with the same probability p of success.

The *geometric* random variable counts the number of failures one obtains prior to the first "success" when sequentially conducting independent Bernoulli trials, each with the probability p of success.

The *negative binomial* random variable counts the number of failures observed when Bernoulli trials, each with the probability p of success, are conducted sequentially until the rth success is obtained. The negative binomial may be derived from several other models and is often considered as a model for discrete count data if the variance exceeds the mean.

The *Poisson* random variable arises from counts in a restricted domain of time, area, or volume and is most useful for counting fairly rare outcomes.

The *hypergeometric* random variable counts the number of successes in sampling from a finite population, which makes the sequential seletions dependent on one another.

These theoretical distributions serve as models for real data that might arise in our quest to improve a process. Each involves assumptions that should be checked carefully before the distribution is applied.

Supplementary Exercises

4.145 Construct probability histograms for the binomial probability distribution for $n = 7$, and $p = 0.2$, 0.5, and 0.8. Notice the symmetry for $p = 0.5$ and the direction of skewness for $p = 0.2$ and $p = 0.8$.

4.146 Construct a probability histogram for the binomial probability distribution for $n = 20$ and $p = 0.5$. Notice that almost all of the probability fall in the interval $5 \le x \le 15$.

4.147 The probability that a single-field radar set will detect an enemy plane is 0.8. Assume that the sets operate independently of each other.

a If we have five radar sets, what is the probability that exactly four sets will detect the plane?

b What is the probability that at least one set will detect the plane?

4.148 Suppose that the four engines of a commercial aircraft were arranged to operate independently and that the probability of in-flight failure of a single engine is 0.01. What are the probabilities that on a given flight the following events occur?

a No failures are observed.

b No more than one failure is observed.

c All engines fail.

4.149 Sampling for defectives from among large lots of a manufactured product yields a number of defectives X that follows a binomial probability distribution. A sampling plan consists of specifying the number n of items to be included in a sample and an acceptance number a. The lot is accepted if $X \leq a$, and it is rejected if $X > a$. Let p denote the proportion of defectives in the lot. For $n = 5$ and $a = 0$, calculate the probability of lot acceptance if the following lot proportions of defectives exist.

a $p = 0$

b $p = 0.3$

c $p = 1.0$

d $p = 0.1$

e $p = 0.5$

A graph showing the probability of lot acceptance as a function of lot fraction defective is called the operating characteristic curve for the sample plan. Construct this curve for the plan $n = 5,\ a = 0$. Notice that a sampling plan is an example of statistical inference. Accepting or rejecting a lot based on information contained in the sample is equivalent to concluding that the lot is either good or bad, respectively. "Good" implies that a low fraction of items are defective and, therefore, the lot is suitable for shipment.

4.150 Refer to Exercise 4.149. A quality control engineer wishes to study two alternative sample plans: $n = 5,\ a = 1$; and $n = 25,\ a = 5$. On a sheet of graph paper, construct the operating characteristic curves for both plans; make use of acceptance probabilities at $p = 0.05$, 0.10, 0.20, 0.30, and 0.40 in each case.

a If you were a seller producing lots whose fraction of defective items ranged from $p = 0$ to $p = 0.10$, which of the two sampling plans would you prefer?

b If you were a buyer wishing to be protected against accepting lots with a fraction of defective items that exceeds $p = 0.30$, which of the two sampling plans would you prefer?

4.151 At an archaeological site that was an ancient swamp, the bones from 22 mastodon skeletons have been unearthed. The bones do not show any sign of disease or malformation. It is thought that these animals wandered into a deep area of the swamp and became trapped in the swamp bottom. The 22 right hind femur bones (thigh bones) were located and 5 of these right hind femurs are to be randomly selected without replacement for DNA testing to determine sex. Let X be the number out of the five selected right hind femurs that are from males.

a Suppose that the group of 22 mastodons whose remains were found in the swamp had been made up of 11 males and 11 females. What is the probability that all five in the sample to be tested are male?

b The DNA testing revealed that all five femurs tested were from males. Based on this result and your answer from part (a), do you think that males and females were equally represented in the group of 22 mastodons stuck in the swamp? Explain.

4.152 For a certain section of a pine forest, the number X of diseased trees per acre has a Poisson distribution with mean $\lambda = 10$. To treat the trees, spraying equipment must be rented for \$150. The diseased trees are sprayed with an insecticide at a cost of \$5 per tree. Let C denote the total spraying cost for a randomly selected acre.

a Find the expected value and the standard deviation for C.

b Within what interval would you expect C to lie with a probability of at least 0.80?

4.153 In checking river water samples for bacteria, a researcher places water in a culture medium to grow colonies of certain bacteria, if present. The number of colonies per dish averages 15 for water samples from a certain river.

a Find the probability that the next dish observed will have at least 10 colonies.

b Find the mean and the standard deviation of the number of colonies per dish.

c Without calculating exact Poisson probabilities, find an interval in which at least 75% of the colony count measurements should lie.

4.154 The number of vehicles passing a specified point on a highway averages eight per minute.

a Find the probability that at least 15 vehicles will pass this point in the next minute.

b Find the probability that at least 15 vehicles will pass this point in the next 2 minutes.

c What assumptions must you make for your answers in parts (a) and part (b) to be valid?

4.155 A production line produces a variable number N of items each day. Suppose that each item produced has the same probability p of not conforming to manufacturing standards. If N has a Poisson distribution with mean λ, then the number of nonconforming items in one day's production X has a Poisson distribution with mean λp. The average number of resistors produced by a facility in 1 day has a Poisson distribution,with a mean of 100. Typically, 5% of the resistors produced do not meet specifications.

a Find the expected number of resistors that will not meet specifications on a given day.

b Find the probability that all resistors will meet the specifications on a given day.

c Find the probability that more than two resistors will fail to meet specifications on a given day.

4.156 The mean number of customers arriving in a bank during a randomly selected hour is four. The bank manager is considering reducing the number of tellers, but she wants to be sure that lines do not get too long. She decides that if no more than two customers come in during a 15-minute period, two tellers (instead of the current three) will be sufficient.

a What is the probability no more than two customers will come in the bank during a randomly selected 15-minute period?

b What is the probability that more than two customers will come in during two consecutive 15-minute periods?

c The manager records the number of customers coming in during 15-minute time periods until she observes a time period during which more than two customers arrive. Eight time periods have been recorded, each with two or fewer customers arriving in each. What is the probability that more than 14 time periods will be observed before having more than two customers arriving during a timed 15-minute period.

4.157 Three men flip coins to see who pays for coffee. If all three match (all heads or all tails), they flip again. Otherwise, the "odd man" pays for coffee.

a What is the probability that they will need to flip the coins more than once?

b What is the probability that they will need to flip the coins more than three times?

c Suppose the men have flipped the coins three times and they matched all three times. What is the probability that they will need to flip the coins more than three more times (more than six times total)?

4.158 A certain type of bacteria cell divides at a constant rate λ over time. Thus, the probability that a particular cell will divide in a small interval of time t is approximately λt. Given that a population starts out at time zero with k cells of this type, and cell divisions are independent of one another, the size of the population at time t, $X(t)$, has the probability distribution

$$P[X(t) = n] = \binom{n-1}{k-1} e^{-\lambda k t}\left(1 - e^{-\lambda t}\right)^{n-k}, \quad n = k, k+1, \ldots$$

a Find the expected value of $X(t)$ in terms of λ and t.

b If, for a certain type of bacteria cell, $\lambda = 0.1$ per second, and the population starts out with two cells at time zero, find the expected population size after 5 seconds.

4.159 In a certain region, the probability of at least one child contracting malaria during a given week is 0.2. Find the average number of weeks with no cases during a 4-week period, assuming that a person contracting malaria is independent of another person contracting malaria.

4.160 The probability that any one vehicle will turn left at a particular intersection is 0.2. The left-turn lane at this intersection has room for three vehicles. If five vehicles arrive at this intersection while the light is red, find the probability that the left-turn lane will hold all of the vehicles that want to turn left.

4.161 Referring to Exercise 4.160, find the probability that six cars must arrive at the intersection while the light is red to fill up the left-turn lane.

4.162 For any probability $p(x)$, $\sum_x p(x) = 1$ if the sum is taken over all possible values x that the random variable in question can assume. Show that this is true for the following distributions.

a The binomial distribution

b The geometric distribution

c The Poisson distribution

4.163 During World War I, the British government established the Industrial Fatigue Research Board (IFRB), later known as the Industrial Health Research Board (IHRB) (Haight 2001). The board was created because of concern for the large number of accidental deaths and injuries in the British war production industries. One of the data sets they considered was the number of accidents experienced by women working on 6-inch shells during the period February 13, 1918 to March 20, 1918. These are displayed in the table that follows. Thus, 447 women had no accidents during this time period, but 2 had at least 5 accidents.

Number of Accidents	Frequency Observed
0	447
1	132
2	42
3	21
4	3
5 or more	2

a Find the average number of accidents a woman had during this time period. (Assume that all observations in the category "5 or more" are exactly 5.)

b After the work of von Bortkiewicz (see Section 4.7.3), the Poisson distribution had been applied to a large number of random phenomena and, with few exceptions, had been found to describe the data well. This had led the Poisson distribution to be called the "random distribution," a term that is still found in the literature. Thus, the mathematicians at the IFRB began by modeling these data using the Poisson distribution. Find the expected number of women having 0, 1, 2, 3, 4, and ≥ 5 accidents using the mean found in part (a) and the Poisson distribution. How well do you think this model describes the data?

c Greenwood and Woods (1919) suggested fitting a negative binomial distribution to these data. Find the expected number of licensed drivers having 0, 1, 2, 3, 4, and ≥ 5 accidents using the mean found in part (a) and the geometric (negative binomial with $r = 1$) distribution. How well do you think this model describes the data?

Historical note: Researchers were puzzled as to why the negative binomial fit better than the Poisson distribution until, in 1920, Greenwood and Yule suggested the following model. Suppose that the probability any given licensed driver will have an accident is distributed according to a Poisson distribution with mean λ. However, λ varies from woman to woman according to a gamma distribution (see Chapter 5). Then the number of accidents would have a negative binomial distribution. The value of λ associated with a licensed driver was called his or her "accident proneness."

4.164 The Department of Transportation's Federal Auto Insurance and Compensation Study was based on a random sample of 7,842 California licensed drivers (Ferreira 1971). The number of accidents in which each was involved from November 1959 to February 1968 was determined. The summary results are given in the next table.

Number of Accidents	Frequency Observed
0	5147
1	1849
2	595
3	167
4	54
5	14
6 or more	6

a Find the average number of accidents a California licensed driver had during this time period. (Assume that all observations in the category "6 or more" are exactly 6.)

b Find the expected number of drivers being involved in 0, 1, 2, 3, 4, 5 and ≥ 6 accidents using the mean found in part (a) and the Poisson distribution. How well do you think this model describes the data?

c Find the expected number of women having 0, 1, 2, 3, 4, 5 and ≥ 6 accidents using the mean found in part (a) and the geometric (negative binomial with $r = 1$) distribution. How well do you think this model describes the data?

Note: In both this and Exercise 4.136, a better fit of the negative binomial can be found by using a noninteger value for r. Although we continue to restrict our attention to only integer values, the negative binomial distribution is well defined for any real value $r > 0$.

4.165 The supply office for a large construction firm has three welding units of Brand A in stock. If a welding unit is requested, the probability is 0.7 that the request will be for this particular brand. On a typical day, five requests for welding units come to the office.

a Find the probability that all three Brand A units will be in use on that day.

b If the supply office also stocks three welding units that are not Brand A, find the probability that exactly one of these units will be left immediately after the third Brand A unit is requested.

4.166 In the game Lotto 6-49, six numbers are randomly chosen without replacement from 1 to 49. A player who matches all six numbers in any order wins the jackpot.

a What is the probability of winning any given jackpot with one game ticket?

b If a game ticket costs \$1.00, what are the expected winnings from playing Lotto 6-49 once.

c Suppose a person buys one Lotto 6-49 ticket each week for a hundred years. Assuming all years have 52 weeks, what is the probability of winning at least one jackpot during this time? (Hint: Use a Poisson approximation.)

d Given the setting in part (c), what are the expected winnings over 100 years?

4.167 The probability of a customer's arriving at a grocery service counter in any one second equals 0.1. Assume that customers arrive in a random stream and, hence, that the arrival at any one second is independent of any other arrival. Also assume that at most one customer can arrive during any one second.

a Find the probability that the first arrival will occur during the third 1-second interval.

b Find the probability that the first arrival will not occur until at least the third 1-second interval.

c Find the probability that no arrivals will occur in the first 5 seconds.

d Find the probability that at least three people will arrive in the first 5 seconds.

4.168 Of a population of consumers, 60% are reputed to prefer a particular brand, A, of toothpaste. If a group of consumers is interviewed, find the probability of the following events.

a Exactly five people are interviewed before encountering a consumer who prefers brand A.

b At least five people are interviewed before encountering a consumer who prefers brand A.

4.169 The mean number of automobiles entering a mountain tunnel per 2-minute periods is one. If an excessive number of cars enter the tunnel during a brief period, the result is a hazardous situation.

a Find the probability that the number of automobiles entering the tunnel during a 2-minute period exceeds three.

b Assume that the tunnel is observed during 10 2-minute intervals, thus giving 10 independent observations, $X_1, X_2, \ldots X_{10}$, on a Poisson random variable. Find the probability that $X > 3$ during at least 1 of the 10 2-minute intervals.

4.170 Suppose that 10% of a brand of Mp3 players will fail before their guarantee has expired. Suppose 1000 players are sold this month, and let X denote the number that will not fail during the guarantee period.

a Find the expected value and variance of X.

b Within what limit would X be expected to fail with a probability of 0.90? (Hint: Use Tchebysheff's Theorem.)

4.171 **a** Consider a binomial experiment for $n = 20$ and $p = 0.05$. Calculate the binomial probabilities for $X = 0, 1, 2, 3$, and 4.

b Calculate the same probabilities, but this time use the Poisson approximation with $\lambda = np$. Compare the two results.

4.172 The manufacturer of a low-calorie dairy drink wishes to compare the taste appeal of a new formula (B) with that of the standard formula (A). Each of four judges is given three glasses in random order: two containing formula A and the other containing formula B. Each judge is asked to choose which glass he most enjoyed. Suppose that the two formulas are equally attractive. Let X be the number of judges stating a preference for the new formula.

a Find the probability function for X.

b What is the probability that at least three of the four judges will state a preference for the new formula?

c Find the expected value of X.

d Find the variance of X.

4.173 Show that the hypergeometric probability function approaches the binomial in the limit as $N \to \infty$ and as $p = r/N$ remains constant; that is, show that

$$\lim_{\substack{N\to\infty \\ r\to\infty}} = \frac{\binom{r}{x}\binom{N-r}{n-r}}{\binom{N}{n}} = \binom{n}{x} p^x q^{n-x}$$

for constant $p = r/N$.

4.174 A lot of $N = 100$ industrial products contains 40 defectives. Let X be the number of defectives in a random sample of size 20. Find $p(10)$ using the following distributions.

a The hypergeometric probability distribution

b The binomial probability distribution

Is N large enough so that the binomial probability function provides a good approximation to the hypergeometric probability function?

REFERENCES

Bortkewitsch, L. von. 1898. *Das Gasetz der kleinen Zahlen*. Leipzig: G. Teubner.

Boswell, M. T., and G. P. Patil. 1970. Chance mechanisms generating the negative binomial distribution. In *Random counts in scientific work*, ed. G. P. Patil, vol 1. University Park: Pennsylvania State University.

Chu, S. 2003. Using soccer goals to motivate the Poisson process. *INFORMS Transactions on Education* 3, no. 2. http://archive.ite.journal.informs.org/Vol3No2/Chu/.

David, F. N. 1955. Studies in the history of probability and statistics. *Biometrika* 52:1–15.

Eves, H. 1969. *Introduction to the history of mathematics*. 3rd ed. New York: Holt, Rinehart and Winston.

Feller, W. 1968. *An introduction to probability theory and its applications, Vol.1, 3rd ed*. New York: Wiley.

Ferreira, J., Jr. 1971. Designing equitable merit rating plans. Working Paper #OR-057-76, Operations Research Center, Massachusetts Institute of Technology.

Folks, J. L. 1981. *Ideas of statistics*. New York: John Wiley & Sons.

Goranson, U. G., and J. Hall. 1980. Airworthiness of long life jet transport structures. *Aeronautical Journal* 84:374–85.

Greenwood, M., and H. M. Woods. 1919. *The incidence of industrial accidents upon individuals with specific reference to multiple accidents*. Industrial Fatigue Research Board. Report No. 4, London.

Greenwood, M., and C. V. Yule. 1920. An inquiry into the nature of frequency distributions representative of multiple happenings, with particular reference to the occurrence of multiple attacks of disease or repeated accidents. *Journal of the Royal Statistical Society* 89:255–79.

Haight, F. A. 2001. *Accident proneness: The history of an idea*. Technical Report #UCIITS-WP-01-4. Institute of Transportation Studies, University of California, Irvine.

Kemeny, J. G., and J. L. Snell. 1983. *Finite Markov chains*. New York: Springer-Verlag.

Kennedy, W., and J. Gentle. 1980. *Statistical computing*. New York: Marcel Dekker.

Parzen, E. 1992. *Modern probability theory and its applications*. New York: Wiley-Interscience.

Sakamoto, C. M. 1973. Application of the Poisson and negative binomial models to thunderstorms and hail days probabilities in Nevada. *Monthly Weather Review* 101:350–55.

Shuster, J. J. 1991. The statistician in a reverse cocaine sting. *The American Statistician* 45: 123–124.

Schmuland, B. 2001. *Shark attacks and the Poisson approximation, π in the sky*, no. 4. http://www.pims.math.ca/pi/issue4/page12-14.pdf.

Stigler, S. M. 1986. *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press.

Student. 1907. On the error of counting with a haemocytometer. *Biometrika* 5:351–60.

Whitaker, L. 1915. On the Poisson law of large numbers. *Biometrika* 10:36–71.

Willis, C. M., S. M. Church, C. M. Guest, W. A. Cook, N. McCarthy, A. J. Bransbury, M. R. T. Church, and J. C. T. Church. 2004. Olfactory detection of human bladder cancer by dogs: proof of principle study. *British Medical Journal* **329**: 712. doi:10.1136/bmj.329.7468.712

Yang, M. C. K., and D. H. Robinson. 1986. *Understanding and learning statistics by computer*. Philadelphia: World Scientific.

5

Continuous Probability Distributions

5.1 Continuous Random Variables and Their Probability Distributions

All of the random variables discussed in Chapter 4 are discrete, assuming only a finite number or a countably infinite number of values. However, many of the random variables seen in practice have more than a countable collection of possible values. Weights of adult patients coming to a clinic may be anywhere from, say, 80 to 300 pounds. Diameters of machined rods from a certain industrial process range from 1.2 to 1.5 centimeters. Proportions of impurities in ore samples may run from 0.10 to 0.80. These random variables can take on any value in an interval of real numbers. That is not to say that every value in the interval can be found in the sample data if one looks long enough; one may never observe a patient weighing exactly 182.38 pounds. Yet no value can be ruled out as a possible observation; one might encounter a patient weighing 182.38 pounds, so this number must be considered in the set of possible outcomes. Because random variables of this type have a continuum of possible values, they are called continuous random variables. Probability distributions for *continuous random variables* are developed in this chapter, and the basic ideas are presented in the context of an experiment on life lengths.

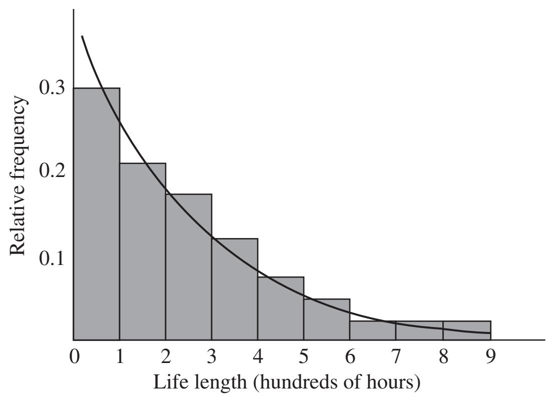
An experimenter is measuring the life length X of a transistor. In this case, X can assume an infinite number of possible values. We cannot assign a positive probability to each possible outcome of the experiment because no matter how small we might make the individual probabilities, they would sum to a value greater than one when accumulated over the entire sample space. However, we can assign positive probabilities to intervals of real numbers in a manner consistent with the axioms of probability. To introduce the basic ideas involved here, let us consider a specific example in some detail.

Suppose that we have measured the life lengths of 50 batteries of a certain type that were selected from a larger population of such batteries. The observed life lengths are as given in Table 5.1. The relative frequency histogram for these data (Figure 5.1) shows clearly that most of the life lengths are near zero, and the frequency drops off rather smoothly as we look at longer life lengths. Here, 32% of the 50 observations fall into the first subinterval (0–1), and another 22% fall into the second (1–2). There is a decline in frequency as we proceed across the subintervals until the last subinterval (8–9) contains a single observation.

TABLE **5.1**
Life Lengths of Batteries (in Hundreds of Hours).

0.406	0.685	4.778	1.725	8.223
2.343	1.401	1.507	0.294	2.230
0.538	0.234	4.025	3.323	2.920
5.088	1.458	1.064	0.774	0.761
5.587	0.517	3.246	2.330	1.064
2.563	0.511	3.246	2.330	1.064
0.023	0.225	1.514	3.214	3.810
3.334	2.325	0.333	7.514	0.968
3.491	2.921	1.624	0.334	4.490
1.267	1.702	2.634	1.849	0.186

FIGURE **5.1**
Relative frequency histogram of data from Table 5.1.



Not only does this sample relative frequency histogram allow us to picture how the sample behaves, but it also gives us some insight into a possible probabilistic model for the random variable X. The histogram of Figure 5.1 looks as though it could be approximated quite closely by a negative exponential curve. The particular function

$$f(x) = \frac{1}{2}e^{-x/2}, \quad x > 0$$

is sketched through the histogram in Figure 5.1 and seems to fit reasonably well. Thus, we could take this function as a mathematical model for the behavior of the random variable X. If we want to use a battery of this type in the future, we might want to know the probability that it will last longer than 400 hours. This probability can be approximated by the area under the curve to the right of the value 4—that is, by

$$\int_4^\infty \frac{1}{2}e^{-x/2}\,\mathrm{d}x = 0.135.$$

Notice that this figure is quite close to the observed sample fraction of lifetimes that exceed 4—namely, $(8/50) = 0.16$. One might suggest that because the sample fraction 0.16 is available, we do not really need the model. However, the model would give more satisfactory answers for other questions than could otherwise be obtained. For example, suppose that we are interested in the probability that X is greater than 9. Then, the model suggests the answer

$$\int_9^\infty \frac{1}{2}e^{-x/2}\,\mathrm{d}x = 0.011$$

whereas the sample shows no observations in excess of 9. These are quite simple examples, in fact, and we shall see many examples of more involved questions for which a model is essential.

Why did we choose the exponential function as a model here? Would some others not do just as well? The choice of a model is a fundamental problem, and we shall spend considerable time in later sections delving into theoretical and practical reasons for these choices. In this preliminary discussion, we merely examine some models that look as though they might do the job.

The function $f(x)$, which models the relative frequency behavior of X, is called the **probability density function.**

DEFINITION 5.1

A random variable X is said to be **continuous** if there is a function $f(x)$, called the **probability density function,** such that

1 $f(x) \geq 0, \quad$ for all x

2 $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$

3 $P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$ ▪

Notice that for a continuous random variable X,

$$P(X = a) = \int_a^a f(x)\,\mathrm{d}x = 0$$

for any specific value a. The need to assign zero probability to any specific value should not disturb us, because X can assume an infinite number of possible values. For example, given all the possible lengths of life of a battery, what is the probability that the battery we are using will last exactly 497.392 hours? Assigning probability zero to this event does not rule out 497.392 as a possible length, but it does imply that the chance of observing this particular length is extremely small.

One consequence of assigning a probability of zero to any specific value is that for a random variable X and constants a and b such that $a < b$,

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

With discrete distributions, probability is associated with specific values, and we must be very careful when working with the probability associated with an interval to specify whether or not the interval's endpoints are included. For continuous distributions, the probability of observing any specific value is zero, positive probability is only associated with intervals, and so the probability of an interval does not depend on whether or not the endpoints are included.

EXAMPLE 5.1 The random variable X of the life lengths of batteries discussed earlier is associated with a probability density function of the form

$$f(x) = \begin{cases} \frac{1}{2}e^{-x/2}, & \text{for } x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that the life of a particular battery of this type is less than 200 or greater than 400 hours.

Solution Let A denote the event that X is less than 2, and let B denote the event that X is greater than 4. Then, because A and B are mutually exclusive,

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) \\ &= \int_0^2 \frac{1}{2}e^{-x/2}\,dx + \int_4^\infty \frac{1}{2}e^{-x/2}\,dx \\ &= (1 - e^{-1}) + (e^{-2}) \\ &= 1 - 0.368 + 0.135 \\ &= 0.767. \quad \blacksquare \end{aligned}$$

EXAMPLE 5.2 Refer to Example 5.1. Find the probability that a battery of this type lasts more than 300 hours, given that it already has been in use for more than 200 hours.

Solution We are interested in $P(X > 3 \mid X > 2)$; and by the definition of conditional probability,

$$P(X > 3 \mid X > 2) = \frac{P(X > 3)}{P(X > 2)}$$

because the intersection of the events $(X > 3)$ and $(X > 2)$ is the event $(X > 3)$.

$$\frac{P(X > 3)}{P(X > 2)} = \frac{\int_3^\infty \frac{1}{2} e^{-x/2}\,dx}{\int_2^\infty \frac{1}{2} e^{-x/2}\,dx} = \frac{e^{-3/2}}{e^{-1}} = e^{-1/2} = 0.606. \quad \blacksquare$$

Sometimes it is convenient to look at cumulative probabilities of the form $P(X \le b)$. To do this, we use the **distribution function** or cumulative distribution function, which we discussed for discrete distributions in Section 4.1. For continuous distributions we have the following definition:

DEFINITION 5.2 The **distribution function** for a random variable X is defined as

$$F(b) = P(X \le b).$$

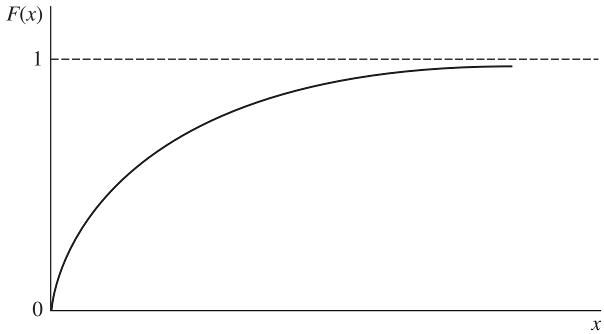
If X is continuous with probability density function $f(x)$, then

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

Notice that $F'(x) = f(x)$. $\blacksquare$

In the battery example, X has a probability density function given by

$$f(x) = \begin{cases} \frac{1}{2} e^{-x/2}\,dx, & \text{for } x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$



F I G U R E **5.2**
Distribution function for a continuous random variable.

Thus,

$$
\begin{aligned}
F(b) &= P(X \le b) \\
&= \int_0^b \frac{1}{2} e^{-x/2}\,\mathrm{d}x \\
&= -e^{-x/2} \,|_a^b \\
&= \begin{cases} 1 - e^{-b/2}, & b > 0 \\ 0, & b \le 0. \end{cases}
\end{aligned}
$$

The function is shown graphically in Figure 5.2.

Notice that the distribution function for a continuous random variable is absolutely continuous over the whole real line. This is in contrast to the distribution function of a discrete random variable, which is a step function. However, whether discrete or continuous, any distribution function must satisfy the four properties of a distribution function:

1 $\lim_{x \to -\infty} F(x) = 0$

2 $\lim_{x \to \infty} F(x) = 1$

3 The distribution function is a nondecreasing function; that is, if $a < b, F(a) \le F(b)$. The distribution function can remain constant, but it cannot decrease as we increase from a to b.

4 The distribution function is right-hand continuous; that is, $\lim_{h \to 0^+} F(x+h) = F(x)$.

For a distribution function of a continuous random variable, not only is the distribution function right-hand continuous as specified in (4), it is left-hand continuous and, thus, absolutely continuous (see Figures 5.2 and 5.3). We may also have a random variable that has discrete probabilities at some points and probability associated with some intervals. In this case, the distribution function will have some points of discontinuity (steps) at values of the variable with discrete probabilities and continuous increases over the intervals that have positive probability. These are discussed more fully in Section 5.11.

As with discrete distribution functions, the probability density function of a continuous random variable may be derived from the distribution function. Just as the

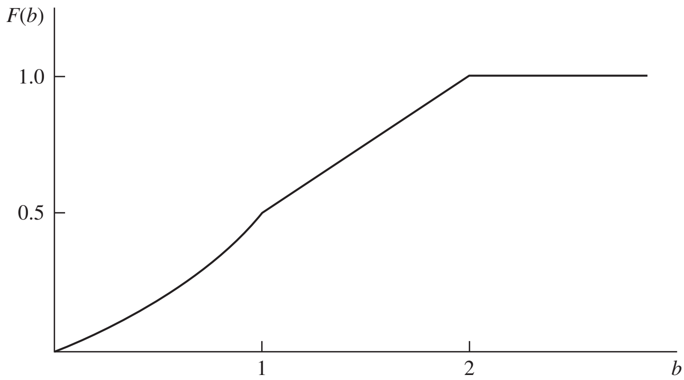


FIGURE 5.3 Distribution function of a continuous random variable.

distribution function can be obtained from the probability density function of a continuous random variable through integration, the probability density may be found by differentiating the distribution function. Specifically, if X is a continuous random variable,

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}(F(x)), \quad x \in \Re.$$

EXAMPLE 5.3 The distribution function of the random variable X, the time (in months) from the diagnosis age until death for one population of patients with AIDS, is as follows:

$$F(x) = 1 - e^{-0.03x^{1.2}}, \quad x > 0$$

1 Find the probability that a randomly selected person from this population survives at least 12 months.

2 Find the probability density function of X.

Solution **1** The probability of surviving at least 12 months is

$$\begin{aligned} P(X \geq 12) &= 1 - P(X \leq 12) \\ &= 1 - F(12) \\ &= 1 - \left(1 - e^{-0.03(12)^{1.2}}\right) \\ &= 1 - 0.45 \\ &= 0.55. \end{aligned}$$

45% of this population will survive more than a year from the time of the diagnosis of AIDS.

2 The probability density function of X is

$$f(x) = \frac{\mathrm{d}}{\mathrm{d}x}(F(x)) = \begin{cases} 0, & x < 0 \\ 0.036\,x^{0.2}e^{-0.03x^{1.2}}, & x \geq 0. \end{cases}$$

Notice that both the probability density function and the distribution function are defined for all real values of X. ▪

Exercises

5.1 For each of the following situations, define an appropriate random variable and state whether it is continuous or discrete.

a An entomologist observes the distance that insects move after emerging from pupation.
b A neonatologist records how many stem cells differentiate into brain cells.
c A toxicologist measures the proportion of toxins in the water.
d An ichthyologist measures the length of fish.

5.2 For each of the following situations, define an appropriate random variable and state whether it is continuous or discrete.

a An astronomer observes the number of stars in a quadrant of the sky.
b A physician records the diastolic blood pressure of a study participant.
c A forester measures the diameter at breast height (dbh) of trees.
d A pathologist is looking for signs of disease in blood samples.

5.3 Suppose that a random variable X has a probability density function given by

$$f(x) = \begin{cases} \frac{x^2}{3}, & -1 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

a Find the probability that $-1 < X < 1$.
b Find the probability that $1 < X < 3$.
c Find the probability that $X \leq 1$ given $X \leq 1.5$.
d Find the distribution function of X.

5.4 The weekly repair cost, X, for a certain machine has a probability density function given by

$$f(x) = \begin{cases} cx(1-x), & 0 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

with measurements in \$100s.

a Find the value of c that makes this function a valid probability density function.
b Find and sketch the distribution function of X.
c What is the probability that repair costs will exceed \$75 during a week?
d What is the probability that the repair costs will exceed \$75 during a week given that they will exceed \$50?

5.5 The distance X between trees in a given forest has a probability density function given by

$$f(x) = \begin{cases} ce^{-x/10}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

with measurement in feet.

a Find the value of c that makes this function a valid probability density function.

b Find and sketch the distribution function of X.

c What is the probability that the distance from a randomly selected tree to its nearest neighbor is at least 15 feet?

d What is the probability that the distance from a randomly selected tree to its nearest neighbor is at least 20 feet given that it is at least 5 feet?

5.6 The effectiveness of solar-energy heating units depends on the amount of radiation available from the sun. During a typical October, daily total solar radiation in Tampa, Florida, approximately follows the following probability density function (units are hundreds of calories).

$$f(x) = \begin{cases} \frac{3}{32}(x-2)(6-x), & 2 \le x \le 6 \\ 0, & \text{otherwise} \end{cases}$$

Find the probability that solar radiation will exceed 400 calories on a typical October day.

5.7 The distribution function of a random variable X is as follows:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^3}{2}, & 0 \le x < 1 \\ \frac{x}{2}, & 1 \le x \le 2 \\ 1, & x > 2 \end{cases}$$

a Graph the distribution function.

b Find the probability that X is between 0.25 and 0.75.

c Find the probability density function of X.

d Graph the probability density function of X.

5.8 Jerry is always early for appointments, arriving between 10 minutes early to exactly on time. The distribution function associated with X, the number of minutes early he arrives, is as follows:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{40}, & 0 \le x \le 4 \\ \frac{20x - x^2 - 40}{60}, & 4 \le x \le 10 \\ 1, & x > 10 \end{cases}$$

a Graph the distribution function.

b Find the probability that Jerry arrives at least 5 minutes early.

c Find the probability density function of X.

d Graph the probability density function of X.

5.9 A firm has been monitoring its total daily telephone usage. The daily use of time conforms to the following probability density function (measured in hours).

$$f(x) = \begin{cases} \frac{3}{64}x^2(4-x), & 0 \le x \le 4 \\ 0, & \text{otherwise} \end{cases}$$

a Graph the probability density function.

b Find the distribution function $F(x)$ for daily telephone usage X.

c Find the probability that the time telephone usage by the firm will exceed 2 hours for a randomly selected day.

d The current budget of the firm covers only 3 hours of daily telephone usage. How often will the budgeted figure be exceeded?

5.10 The pH level, a measure of acidity, is important in studies of acid rain. For a certain lake, baseline measurements of acidity are made so that any changes caused by acid rain can be noted. The pH for water samples from the lake is a random variable X with probability density function

$$f(x) = \begin{cases} \frac{3}{8}(7-x)^2, & 5 \le x \le 7 \\ 0, & \text{otherwise.} \end{cases}$$

a Sketch the curve of $f(x)$.

b Find the distribution function $F(x)$ for X.

c Find the probability that the pH for a water sample from this lake will be less than 6.

d Find the probability that the pH of a water sample from this lake will be less than 5.5 given that it is known to be less than 6.

5.11 The proportion of time during a 40-hour workweek that an industrial robot was in operation was measured for a large number of weeks. The measurements can be modeled by the probability density function

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

If X denotes the proportion of time this robot will be in operation during a coming week, find the following.

a $P(X > 1/2)$

b $P(X > 1/2 \mid X > 1/4)$

c $P(X > 1/4 \mid X > 1/2)$

d $F(x)$. Graph this function. Is $F(x)$ continuous?

5.12 The proportion of impurities by weight X in certain copper ore samples is a random variable having a probability density function of

$$f(x) = \begin{cases} 12x^2(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

If four such samples are independently selected, find the probabilities of the following events.

a Exactly one sample has a proportion of impurities exceeding 0.5.

b At least one sample has a proportion of impurities exceeding 0.5.

5.2 Expected Values of Continuous Random Variables

As in the discrete case, we often want to summarize the information contained in a continuous variable's probability distribution by calculating **expected values** for the random variable and for certain functions of the random variable.

DEFINITION **5.3**

The **expected value** of a continuous random variable X that has a probability density function $f(x)$ is given by

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

Note: We assume the absolute convergence of all integrals so that the expectations exist. ▪

For functions of random variables, we have the following theorem.

THEOREM **5.1**

If X is a continuous random variable with probability distribution $f(x)$, and if $g(x)$ is any real-valued function of X, then

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x.$$

The proof of Theorem 5.1 is not given here. ▪

The definitions of **variance** and of *standard deviation* given in Definitions 4.5 and 4.6 and the properties given in Theorems 4.2 and 4.3 hold for the continuous case as well.

DEFINITION **5.4**

For a random variable X with probability density function $f(x)$, the **variance** of X is given by

$$\begin{aligned} V(X) &= E\left[(X-\mu)^2\right] \\ &= \int_{-\infty}^{+\infty} (x-\mu)^2 f(x)\,\mathrm{d}x \\ &= E\left(X^2\right) - \mu^2 \end{aligned}$$

where $\mu = E(X)$. ▪

For constants a and b,

$$E(aX + b) = aE(X) + b$$

and

$$V(aX + b) = a^2 V(X).$$

We illustrate the expectations of continuous random variables in the following two examples.

EXAMPLE **5.4** For a given teller in a bank, let X denote the proportion of time, out of a 40-hour workweek, that he is directly serving customers. Suppose that X has a probability density function given by

$$f(x) = \begin{cases} 3x^2, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

1 Find the mean proportion of time during a 40-hour workweek the teller directly serves customers.

2 Find the variance of the proportion of time during a 40-hour workweek the teller directly serves customers.

3 Find an interval that, for 75% of the weeks, contains the proportion of time that the teller spends directly serving customers.

Solution **1** From Definition 5.3,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \\ &= \int_0^1 x\left(3x^2\right)\mathrm{d}x \\ &= \int_0^1 3x^3\,\mathrm{d}x \\ &= 3\left[\frac{x^4}{4}\right]_0^1 \\ &= \frac{3}{4} \\ &= 0.75. \end{aligned}$$

Thus, on average, the teller spends 75% of his time each week directly serving customers.

2 To compute $V(X)$, we first find $E(X^2)$:

$$\begin{aligned} E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \\ &= \int_0^1 x^2\left(3x^2\right)\,\mathrm{d}x \end{aligned}$$

$$= \int_0^1 3x^4 \, dx$$

$$= 3\left[\frac{x^5}{5}\right]_0^1$$

$$= \frac{3}{5}$$

$$= 0.60.$$

Then,

$$\begin{aligned} V(X) &= E(X^2) - [E(X)]^2 \\ &= 0.60 - (0.75)^2 \\ &= 0.60 - 0.5625 \\ &= 0.0375. \end{aligned}$$

3 An interval that captures the proportion of time that the teller spends directly serving customers for 75% of the weeks can be constructed in an infinite number of ways. Here we construct one so that for the other 25% of the weeks, half of the weeks the proportion he spends directly helping customers is less than the lower limit, and the other half of the weeks the proportion of time he spends directly helping customers is more than the upper limit. That is, we want to find a and b so that $P(X < a) = 0.125$ and $P(X > b) = 0.125$. Now,

$$P(X < a) = \int_0^a 3x^2 \, dx = x^3 \mid_0^a = a^3.$$

Because $a^3 = 0.125$, $a = 0.5$. Similarly,

$$P(X > b) = \int_b^1 3x^2 \, dx = x^3 \mid_b^1 = 1 - b^3.$$

Because $1 - b^3 = 0.125$, we find that $b = 0.956$. That is, for 75% of the weeks, the teller spends between 50% and 95.6% of his time directly serving customers. ▪

EXAMPLE **5.5** The weekly demand X, in hundreds of gallons, for propane at a certain supply station has a density function given by

$$f(x) = \begin{cases} \frac{x}{4}, & 0 \le x \le 2 \\ \frac{1}{2}, & 2 < x \le 3 \\ 0, & \text{elsewhere.} \end{cases}$$

It takes \$50 per week to maintain the supply station. Propane is purchased for \$270 per hundred gallons and redistributed by the supply station for \$1.75 per gallon.

1 Find the expected weekly demand.

2 Find the expected weekly profit.

Solution **1**

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\,dx \\ &= \int_0^2 x\left(\frac{x}{4}\right) dx + \int_2^3 x\left(\frac{1}{2}\right) dx \\ &= \int_0^2 \left(\frac{1}{4}\right) x^2\,dx + \int_2^3 \left(\frac{1}{2}\right) x\,dx \\ &= \left(\frac{1}{4}\right)\left[\frac{x^3}{3}\right]_0^2 + \left(\frac{1}{2}\right)\left[\frac{x^2}{2}\right]_2^3 \\ &= \frac{1}{12}(8) + \left(\frac{1}{4}\right)(9-4) \\ &= \frac{2}{3} + \frac{5}{4} \\ &= 1.92. \end{aligned}$$

On average, the weekly demand for propane will be 192 gallons at this supply center.

2 The propane is purchased for \$270 per hundred gallons and sold for \$175 per hundred gallons (\$1.75 × 100), yielding a profit of \$95 per \$100 gallons sold. The weekly profit P is

$$P = 95X - 50.$$

Therefore, the expected profit is

$$\begin{aligned}E(P) &= E(95X - 50)\\ &= 95E(X) - 50\\ &= 95(1.92) - 50\\ &= 132.40.\end{aligned}$$

That is, on average, the supply station has a profit of \$132.40 per week. ▪

Tchebysheff's Theorem (Theorem 4.4) holds for continuous random variables, just as it does for discrete ones. Thus, if X is continuous, with a mean of μ and a standard deviation of σ,

$$P(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

for any positive number k. Because Tchebysheff's Theorem provides a bound on the probability and is not exact, we choose to compute the exact probabilities when it is reasonable to do so, as in Example 5.4. However, sometimes we do not know the exact form of the distribution, or the probabilities are difficult to compute for a particular distribution. In these cases, Tchebysheff's Theorem is useful. We illustrate its use in the next example.

EXAMPLE 5.6 The weekly amount X spent for chemicals by a certain firm has a mean of \$1565 and a variance of \$428. Within what interval should these weekly costs for chemicals be expected to lie in at least 75% of the time?

Solution To find an interval guaranteed to contain at least 75% of the probability mass for X, we specify

$$1 - \frac{1}{k^2} = 0.75$$

which yields

$$\begin{aligned}\frac{1}{k^2} &= 0.25\\ k^2 &= \frac{1}{0.25}\\ &= 4\\ k &= 2.\end{aligned}$$

Thus, the interval $\mu - 2\sigma$ to $\mu + 2\sigma$ will contain at least 75% of the probability. This interval is given by

$$1565 - 2\sqrt{428} \quad \text{to} \quad 1565 + 2\sqrt{428}$$
$$1565 - 41.38 \quad \text{to} \quad 1565 + 41.38$$
$$1523.62 \quad \text{to} \quad 1606.38.$$

75% of the weekly chemical costs will be between \$1523.62 and \$1606.38. ▪

If Tchebysheff's Theorem had been used in Example 5.4, we would have found that in at least 75% of the weeks, the teller spends between 36% and 100% of his time helping customers. Notice that this interval is much broader than the exact interval (50% to 95.6%) we found in Example 5.4. To find the exact interval, we had to know and use the distribution of the random variable. Tchebysheff's Theorem gives the same result for all distributions with the same mean and variance, generally leading to a broader interval, sometimes a much broader interval. Therefore, if we know the form of the distribution, we prefer the exact interval.

The expected value of the random variable X can be found using the distribution function without first finding the probability density function. That is, for any nonnegative continuous random variable with distribution function $F(x)$ and finite mean $E(X)$,

$$E(X) = \int_0^\infty [1 - F(x)]\,dx.$$

The proof of this will be left as an exercise.

Exercises

5.13 The effectiveness of solar-energy heating units depends on the amount of radiation available from the sun. During a typical October, daily total solar radiation in Tampa, Florida, approximately follows the following probability density function (units are hundreds of calories).

$$f(x) = \begin{cases} \frac{3}{32}(x-2)(6-x), & 2 \le x \le 6 \\ 0, & \text{otherwise.} \end{cases}$$

Find the mean, variance, and standard deviation of the distribution of the daily total solar radiation in Tampa in October.

5.14 The "on" temperature of a thermostatically controlled switch for an air-conditioning system is set at 72°, but the actual temperature X at which the switch turns on is a random variable having the probability density function

$$f(x) = \begin{cases} \frac{1}{2}, & 71 \le x \le 73 \\ 0, & \text{otherwise.} \end{cases}$$

Find the mean and standard deviation of the distribution of the temperatures at which the switch turns on.

5.15 The weekly repair cost, X, for a certain machine has a probability density function given by

$$f(x) = \begin{cases} 6x(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

with measurements in \$100s.

a Find the mean and variance of the distribution of repair costs.

b Find an interval within which these weekly repair costs should lie at least 75% of the time using Tchebysheff's Theorem.

c Find an interval within which these weekly repair costs lie exactly 75% of the time with exactly half of those not lying in the interval above the upper limit and the other half below the lower limit. Compare this interval to the one obtained in part (b).

5.16 In a genetics study, it was found that the distance X between mutations in a certain strand of DNA had a probability density function given by

$$f(x) = \begin{cases} \frac{1}{5}e^{-x/5}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

with measurements in kb.

a Find the mean and variance of the distribution of distances between mutations in this strand of DNA.

b Find an interval within which these distances between mutations should lie at least 75% of the time using Tchebysheff's Theorem.

c Find an interval within which these distances between mutations lie exactly 75% of the time with exactly half of those not lying in the interval above the upper limit and the other half lying below the lower limit. Compare this interval to the one obtained in part (b).

5.17 A firm has been monitoring its total daily telephone usage. The daily use of time conforms to the following probability density function (measured in hours).

$$f(x) = \begin{cases} \frac{3}{64}x^2(4-x), & 0 \le x \le 4 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the mean, variance, and standard deviation of the distribution of the firm's daily telephone usage.

b Find an interval in which the daily telephone usage should lie at least 75% of the time.

5.18 The pH level, a measure of acidity, is important in studies of acid rain. For a certain lake, baseline measurements of acidity are made so that any changes caused by acid rain can be noted. The pH for water samples from the lake is a random variable X, with probability density function

$$f(x) = \begin{cases} \frac{3}{8}(7-x)^2, & 5 \le x \le 7 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the mean, variance, and standard deviation of the distribution of the pH of water in this lake.

b Find an exact interval within which 80% of the pH measurements must lie with half of the other 20% of the pH measurements lying below the lower limit and the other half lying above the upper limit.

c Would you expect to see a pH measurement greater than 6.5 very often? Why?

5.19 The proportion of time during a 40-hour workweek that an industrial robot was in operation was measured for a large number of weeks where X denotes the proportion of time this robot will be in operation during a coming week. The measurements can be modeled by the probability density function

$$f(x) = \begin{cases} 2x, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the mean and variance of the distribution of the proportions of time that the robot operates during a week.

b For the robot under study, the profit Y for a week is given by $Y = 400X - 80$. Find $E(Y)$ and $V(Y)$.

c Find an interval in which the profit should lie for at least 75% of the weeks that the robot is in use.

5.20 The proportion of impurities by weight X in certain copper ore samples is a random variable having a probability density function of

$$f(x) = \begin{cases} 12x^2(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the mean and standard deviation of the distribution of the proportions of impurities in these copper ore samples.

b The value Y of 100 pounds of copper ore is $Y = 200(1-X)$ dollars. Find the mean and standard deviation of the value of 100 pounds of copper from this sample.

5.21 The distribution function of the random variable X, the time (in years) from the time a machine is serviced until it breaks down, is as follows:

$$F(x) = 1 - e^{-4x}, \quad x > 0.$$

Find the mean time until the machine breaks down after service.

5.22 Jerry is always early for appointments, arriving between 10 minutes early to exactly on time. The distribution function associated with X, the number of minutes early that he arrives, is as follows:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x^2}{40}, & 0 \le x \le 4 \\ \frac{20x - x^2 - 40}{60}, & 4 \le x \le 10 \\ 1, & x > 10. \end{cases}$$

Find the mean number of minutes that Jerry is early for appointments.

5.23 A gas station owner has a daily demand of $100X$ gallons of gas. (Note that $X = 0.5$ represents a demand for 50 gallons, and $X = 1$ represents a demand for 100 gallons.) The distribution of daily demand is as follows:

$$f(x) = \begin{cases} \frac{3}{2}\left(1 - x + \frac{x^2}{4}\right), & 0 \le x \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

The owner's profit is \$10 for each 100 gallons sold (10 cents per gallon) if $X \le 1$, and \$15 per 100 gallons if $X > 1$. Find the retailer's expected profit for any given day.

5.24 A retail grocer has a daily demand X for a certain food sold by the pound such that X (measured in hundreds of pounds) has a probability density function of

$$f(x) = \begin{cases} 3x^2, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The grocer, who cannot stock more than 100 pounds, wants to order $100k$ pounds of food on a certain day. He buys the food at 10 cents per pound and sells it at 15 cents per pound. What value of k will maximize his expected daily profit? (There is no salvage value for food not sold.)

5.25 The density function of X is given by

$$f(x) = \begin{cases} ax + bx^2, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

$E(X) = 1/9$. Find a and b.

5.26 Let X be a nonnegative, continuous random variable with distribution function $F(x)$. Prove that

$$E(X) = \int_0^\infty [1 - F(x)]\,\mathrm{d}x.$$

Hint: Use integration by parts.

5.3 The Uniform Distribution

5.3.1 Probability Density Function

We now move from a general discussion of continuous random variables to discussions of specific models that have been found useful in practice. Consider an experiment that consists of observing events in a certain time frame, such as buses arriving at a bus stop or telephone calls coming into a switchboard during a specified period. Suppose that we know that one such event has occurred in the time interval (a, b): A bus arrived between 8:00 and 8:10. It may then be of interest to place a probability distribution on the actual time of occurrence of the event under observation, which we denote by X. A very simple model assumes that X is equally likely to lie in any small subinterval—say, of length d—no matter where that subinterval lies within (a, b). This assumption leads to the uniform probability distribution, which has the probability density function given by

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{elsewhere.} \end{cases}$$

This density function is graphed in Figure 5.4.

The distribution function for a uniformly distributed X is given by

$$F(x) = \begin{cases} 0, & x < a \\ \int_a^x \frac{1}{b-a}\,\mathrm{d}t = \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b. \end{cases}$$

A graph of the distribution function is shown in Figure 5.5.

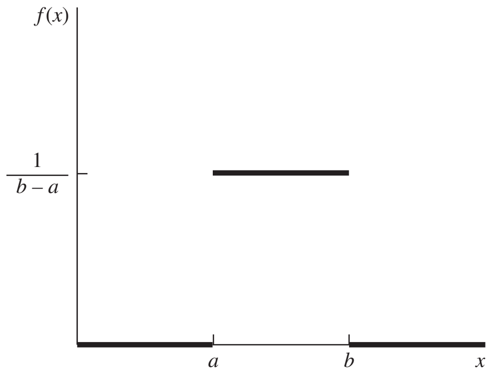


FIGURE **5.4** The probability density function of a uniform random variable.

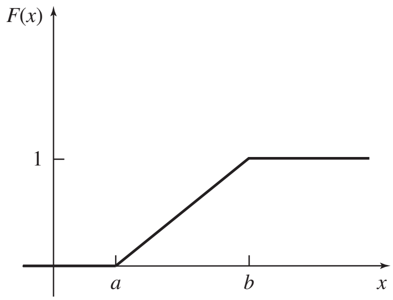


FIGURE **5.5** Distribution function of a uniform random variable.

If we consider a subinterval $(c, c + d)$ contained entirely within (a, b),

$$\begin{aligned} P(c \leq X \leq c + d) &= P(X \leq c + d) - P(X \leq c) \\ &= F(c + d) - F(c) \\ &= \frac{(c + d) - a}{b - a} - \frac{c - a}{b - a} \\ &= \frac{d}{b - a}. \end{aligned}$$

Notice that this probability does not depend on the subinterval's location c, but only on its length d.

A relationship exists between the uniform distribution and the Poisson distribution, which was introduced in Section 4.7. Suppose that the number of events that occur in an interval—say, (0, t)—has a Poisson distribution. If exactly one of these events is known to have occurred in the interval (a, b) with $a \geq 0$ and $a \leq t$, then the conditional probability distribution of the actual time of occurrence for this event (given that it has occurred) is uniform over (a, b).

5.3.2 Mean and Variance

Paralleling our approach in Chapter 4, we now look at the mean and the variance of the uniform distribution. From Definition 5.3,

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x \\ &= \int_{a}^{b} x\left(\frac{1}{b-a}\,\mathrm{d}x\right) \\ &= \left(\frac{1}{b-a}\right)\left(\frac{b^2-a^2}{2}\right) \\ &= \frac{a+b}{2}. \end{aligned}$$

This result is an intuitive one. Because the probability density function is constant over the interval, the mean value of a uniformly distribution random variable should lie at the midpoint of the interval.

Recall from Theorem 4.2 that $V(X) = E\left[(X-\mu)^2\right] = E\left(X^2\right) - \mu^2$; we have, in the uniform case,

$$\begin{aligned} E\left(X^2\right) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \\ &= \int_{a}^{b} x^2\left(\frac{1}{b-a}\,\mathrm{d}x\right) \\ &= \left(\frac{1}{b-a}\right)\left(\frac{b^3-a^3}{3}\right) \\ &= \frac{b^2+ab+a^2}{3}. \end{aligned}$$

Then,

$$\begin{aligned} V(X) &= \frac{b^2+ab+a^2}{3} - \left(\frac{a+b}{2}\right)^2 \\ &= \frac{1}{12}\left[4\left(b^2+ab+a^2\right) - 3(a+b)^2\right] \\ &= \frac{1}{12}(b-a)^2. \end{aligned}$$

This result may not be intuitive, but we see that the variance depends only on the length of the interval (a, b).

EXAMPLE 5.7 A farmer living in western Nebraska has an irrigation system to provide water for crops, primarily corn, on a large farm. Although he has thought about buying a backup pump, he has not done so. If the pump fails, delivery time X for a new pump to arrive is uniformly distributed over the interval from 1 to 4 days. The pump fails. It is a critical time in the growing season in that the yield will be greatly reduced if the crop is not watered within the next 3 days. Assuming that the pump is ordered immediately and that installation time is negligible, what is the probability that the farmer will suffer major yield loss?

Solution Let T be the time until the pump is delivered. T is uniformly distributed over the interval (1, 4). The probability of major loss is the probability that the time until delivery exceeds 3 days. Thus,

$$P(T > 3) = \int_3^4 \frac{1}{3}\,dt = \frac{1}{3}.$$

Notice that the bounds of integration go from 3 to 4. The upper bound is 4 because the probability density function is zero for all values outside the interval [1, 4]. ▪

5.3.3 History and Applications

The term *uniform distribution* appears in 1937 in *Introduction to Mathematical Probability* by J. V. Uspensky. On page 237 of this text, it is noted that "A stochastic variable is said to have uniform distribution of probability if probabilities attached to two equal intervals are equal." In practice, the distribution is generally used when every point in an interval is equally likely to occur, or at least insufficient knowledge exists to propose another model.

We now review the properties of the uniform distribution.

The Uniform Distribution

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & \text{elsewhere} \end{cases}$$

$$E(X) = \frac{a+b}{2} \quad \text{and} \quad V(X) = \frac{(b-a)^2}{12}$$

Exercises

5.27 Suppose X has a uniform distribution over the interval (a, b).

a Find $P(X > c)$ for some point c between a and b.

b If $a \leq c \leq d \leq b$, find $P(X > d \mid X > c)$.

5.28 Henry is to be at Megan's apartment at 6:30. From past experience, Megan knows that Henry will be up to 20 minutes late (never early) and that he is equally likely to arrive any time up to 20 minutes after the scheduled arrival time.

a What is the probability that Henry will be more than 15 minutes late?

b What is the mean and standard deviation of the amount of time that Megan waits for Henry?

5.29 The space shuttle has a 2-hour window during which it can launch for an upcoming mission. Launch time is uniformly distributed in the launch window. Find the probability that the launch will occur as follows:

a During the first 30 minutes of the launch window

b During the last 10 minutes of the launch window

c Within 10 minutes of the center of the launch window

5.30 If a point is randomly located in an interval (a, b), and if X denotes the distance of the point from a, then X is assumed to have a uniform distribution over (a, b). A plant efficiency expert randomly picks a spot along a 500-foot assembly line from which to observe work habits. Find the probability that the point she selects is located as follows:

a Within 25 feet of the end of the line

b Within 25 feet of the beginning of the line

c Closer to the beginning of the line than to the end of the line

5.31 A bomb is to be dropped along a 1-mile-long line that stretches across a practice target zone. The target zone's center is at the midpoint of the line. The target will be destroyed if the bomb falls within 0.1 mile on either side of the center. Find the probability that the target will be destroyed given that the bomb falls at a random location along the line.

5.32 The number of defective Blu-ray players among those produced by a large manufacturing firm follows a Poisson distribution. For a particular 8-hour day, one defective player is found.

a Find the probability that it was produced during the first hour of operation for that day.

b Find the probability that it was produced during the last hour of operation for that day.

c Given that no defective players were seen during the first 4 hours of operation, find the probability that the defective player was produced during the fifth hour.

5.33 A researcher has been observing a certain volcano for a long time. He knows that an eruption is imminent and is equally likely to occur any time in the next 24 hours.

a What is the probability that the volcano will not erupt for at least 15 hours?

b Find a time such that there is only a 10% chance that the volcano would not have erupted by that time.

5.34 In determining the range of an acoustic source by triangulation, one must accurately measure the time at which the spherical wave front arrives at a receiving sensor. According to Perruzzi and Hilliard (1984), errors in measuring these arrival times can be modeled as having uniform distributions. Suppose that measurement errors are uniformly distributed from -0.04 to $+0.05$ microsecond.

a Find the probability that a particular arrival time measurement will be in error by less than 0.01 microsecond.

b Find the mean and the variance of these measurement errors.

5.35 Arrivals of customers at a bank follow a Poisson distribution. During the first hour that the bank is open, one customer arrives at the bank.

a Find the probability that he arrives during the first 15 minutes that the bank is open.

b Find the probability that he arrives after the bank has been open 30 minutes.

5.36 To win a particular video game, one must successfully pass through three levels. The difficulty of passing through a level varies but tends to increase with level; that is, the difficulty of passing through level i is uniformly distributed on the interval $(1/i,\ i+1)$. Because she tends to improve as the game progresses, the player knows that she will pass the first level if the difficulty is less than 1.5, the second level if the difficulty is less than 2.25, and the third level if the difficulty is less than 2.75.

a What is the probability of passing through the first level?

b What is the probability of passing through the second level given that the player passed through the first level?

c What is probability of successfully passing through all three levels?

5.37 According to Zimmels (1983), the sizes of particles used in sedimentation experiments often have uniform distributions. It is important to study both the mean and the variance of particle sizes because in sedimentation with mixtures of various-size particles, the larger particles hinder the movements of the smaller particles.

Suppose that spherical particles have diameters uniformly distributed between 0.01 and 0.05 centimeter. Find the mean and the variance of the volumes of these particles. (Recall that the volume of a sphere is $(4/3)\pi r^3$.)

5.38 In ecology, the broken stick model is sometimes used to describe the allocation of environmental resources among species. For two species, assume that one species is assigned one end of the stick; the other end represents the resources for the other species. A point is randomly selected along a stick of unit length. The stick is broken at the selected point, and each species receives the proportion of the environmental resources equal to the proportion of the stick it receives. For this model, find the probability of the following events.

a The two species have equal proportions of resources.

b One species gets at least twice as much resources as the other species.

c If each species is known to have received at least 10% of the resources, what is the probability that one received at least twice as much as the other species?

5.39 In tests of stopping distance for automobiles, cars traveling 30 miles per hour before the brakes were applied tended to travel distances that appeared to be uniformly distributed between two points a and b. Find the probabilities of the following events.

a One of these automobiles, selected at random, stops closer to a than to b.

b One of these automobiles, selected at random, stops at a point where the distance to a is more than three times the distance to b.

c Suppose that three automobiles are used in the test. Find the probability that exactly one of the three travels past the midpoint between a and b.

5.40 The cycle time for trucks hauling concrete to a highway construction site is uniformly distributed over the interval from 50 to 70 minutes.

a Find the expected value and the variance for these cycle times.

b How many trucks should you expect to have to schedule for this job so that a truckload of concrete can be dumped at the site every 15 minutes?

5.4 The Exponential Distribution

5.4.1 Probability Density Function

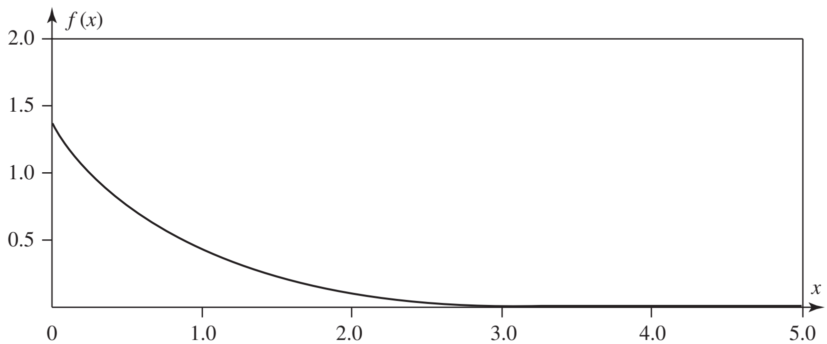
The life-length data of Section 5.1 displayed a nonuniform probabilistic behavior; the probability over intervals of constant length decreased as the intervals moved farther to the right. We saw that an exponential curve seemed to fit these data rather well, and we now discuss the exponential probability distribution in more detail. In general, the exponential density function is given by

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-x/\theta}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

where the parameter θ is a constant ($\theta > 0$) that determines the rate at which the curve decreases.

An exponential density function with $\theta = 2$ is sketched in Figure 5.1; and in general, the exponential functions have the form shown in Figure 5.6. Many random variables in engineering and the sciences can be modeled appropriately as having exponential distributions. Figure 5.7 shows two examples of relative frequency distributions for times between arrivals (interarrival times) of vehicles at a fixed point on a one-directional roadway. Both of these relative frequency histograms can be modeled quite nicely by exponential functions. Notice that the higher traffic density causes shorter interarrival times to be more frequent.

FIGURE 5.6 Exponential probability density function.



The distribution function for the exponential case has the following simple form:

$$F(x) = \begin{cases} 0, & x < 0 \\ P(X \leq x) = \int\limits_0^x \frac{1}{\theta} e^{-t/\theta}\,dt = -\,e^{-x/\theta}\big|_0^x = 1 - e^{-x/\theta}, & x \geq 0 \end{cases}$$

The distribution function of the exponential distribution has the form displayed in Figure 5.8.

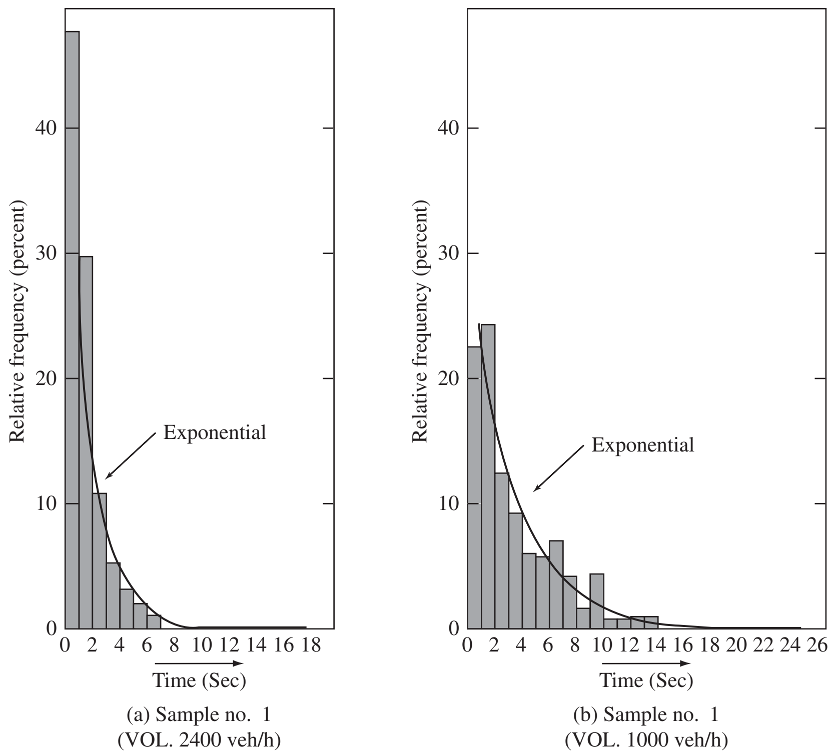


FIGURE **5.7**
Interarrival times of vehicles on a one-directional road (Mahalel and Hakkert 1983).

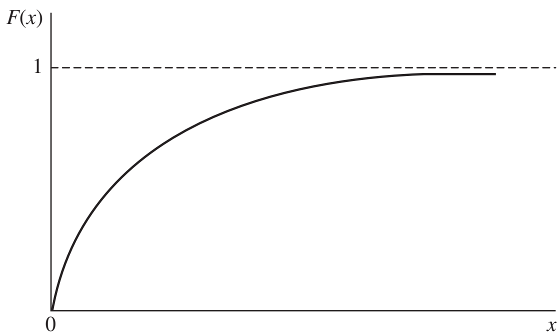


FIGURE **5.8**
Exponential distribution function.

5.4.2 Mean and Variance

Finding expected values for the exponential distribution is simplified by understanding a certain type of integral called a *gamma* (Γ) *function*. The function $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

EXAMPLE 5.8 Show that $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$.

Solution

$$\Gamma(\alpha + 1) = \int_0^\infty x^\alpha e^{-x}\, dx$$

Using integration by parts, let

$$u = x^\alpha \qquad \text{and} \qquad dv = e^{-x}\, dx$$

$$du = \alpha x^{\alpha-1} \qquad\qquad v = -e^{-x}$$

Then

$$\begin{aligned}\Gamma(\alpha + 1) &= -x^\alpha e^{-x}\big|_0^\infty + \int_0^\infty \alpha x^{\alpha-1} e^{-x}\, dx \\ &= -(0 - 0) + \alpha \int_0^\infty x^{\alpha-1} e^{-x}\, dx \\ &= \alpha\Gamma(\alpha).\end{aligned}$$

■

It follows from the above example that $\Gamma(n) = (n - 1)!$, for any positive integer n. The integral

$$\int_0^\infty x^{\alpha-1} e^{-x/\beta}\, dx,$$

for positive constants α and β can be evaluated by making the transformation $y = x/\beta$, or $x = \beta y$, where $dx = \beta dy$. We then have

$$\int_0^\infty (\beta y)^{\alpha-1} e^{-y} (\beta\, dy) = \beta^\alpha \int_0^\infty y^{\alpha-1} e^{-y}\, dy = \beta^\alpha \Gamma(\alpha).$$

It is left as an exercise to show that $\Gamma(1/2) = \sqrt{\pi}$.

Using the properties of the gamma function, we see that, for the exponential distribution,

$$\begin{aligned}E(X) &= \int_{-\infty}^\infty x f(x)\, dx \\ &= \int_0^\infty x \left(\frac{1}{\theta}\right) e^{-x/\theta}\, dx\end{aligned}$$

$$= \frac{1}{\theta} \int_0^\infty x e^{-x/\theta} \, dx$$

$$= \frac{1}{\theta} \Gamma(2) \theta^2$$

$$= \theta.$$

Thus, the parameter θ actually is the mean of the distribution.

To evaluate the variance of the exponential distribution, we start by finding

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x) \, dx$$

$$= \int_0^\infty x^2 \left(\frac{1}{\theta}\right) e^{-x/\theta} \, dx$$

$$= \frac{1}{\theta} \int_0^\infty x^2 e^{-x/\theta} \, dx$$

$$= \frac{1}{\theta} \Gamma(3) \theta^3$$

$$= 2\theta^2.$$

It follows that

$$V(X) = E\left(X^2\right) - \mu^2$$

$$= 2\theta^2 - \theta^2$$

$$= \theta^2$$

and θ becomes the standard deviation as well as the mean.

EXAMPLE **5.9** A sugar refinery has three processing plants, all of which receive raw sugar in bulk. The amount of sugar that one plant can process in one day can be modeled as having an exponential distribution with a mean of 4 tons for each of the three plants. If the plants operate independently, find the probability that exactly two of the three plants will process more than 4 tons on a given day.

Solution The probability that any given plant will process more than 4 tons on a given day, with X denoting the amount used, is

$$\begin{aligned} P(X > 4) &= \int_4^\infty f(x)\,dx \\ &= \int_4^\infty \frac{1}{4}e^{-x/4}\,dx \\ &= -e^{-x/4}\Big|_4^\infty \\ &= e^{-1} \\ &= 0.37. \end{aligned}$$

Note: $P(X > 4) \neq 0.5$, even though the mean of X is 4. The median is less than the mean, indicating that the distribution is skewed.

Knowledge of the distribution function allows us to evaluate this probability immediately as

$$\begin{aligned} P(X > 4) &= 1 - P(X \leq 4) \\ &= 1 - (1 - e^{-4/4}) \\ &= e^{-1} = 0.37. \end{aligned}$$

Assuming that the three plants operate independently, the problem is to find the probability of two successes out of three tries, where the probability of success is 0.37. This is a binomial problem, and the solution is

$$\begin{aligned} P(\text{Exactly two plants use more than 4 tons}) &= \binom{3}{2}(0.37)^2(0.63) \\ &= 3(0.37)^2(0.63) \\ &= 0.26. \quad \blacksquare \end{aligned}$$

EXAMPLE **5.10** Consider a particular plant in Example 5.9. How much raw sugar should be stocked for that plant each day so that the chance of running out of product is only 0.05?

Solution Let a denote the amount to be stocked. Because the amount to be used X has an exponential distribution,

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

We want to choose a so that

$$P(X > a) = e^{-a/4} = 0.05.$$

Solving this equation yields

$$a = 11.98. \quad \blacksquare$$

5.4.3 Properties

Recall that in Section 4.5, we learned that the geometric distribution is the discrete distribution with the memoryless property. The exponential distribution is the continuous distribution with the memoryless property. To verify this, suppose that X has an exponential distribution with parameter θ. Then,

$$\begin{aligned}
P(X > a + b \mid X > a) &= \frac{P[(X > a + b) \cap (X > a)]}{P(X > a)} \\
&= \frac{P(X > a + b)}{P(X > a)} \\
&= \frac{1 - F(a + b)}{1 - F(a)} \\
&= \frac{1 - \left(1 - e^{-(a+b)/\theta}\right)}{1 - (1 - \left(e^{-a/\theta}\right)} \\
&= e^{-b/\theta} \\
&= 1 - F(b) \\
&= P(X > b).
\end{aligned}$$

This memoryless property sometimes causes concerns about the exponential distribution's usefulness as a model. As an illustration, the length of time that a light bulb burns may be modeled with an exponential distribution. The memoryless property implies that, if a bulb has burned for 1000 hours, the probability it will burn at least 1000 more hours is the same as the probability that the bulb would burn more than 1000 hours when new. This failure to account for the deterioration of the bulb over time is the property that causes one to question the appropriateness of the exponential model for life-time data though it is still used often.

A relationship also exists between the exponential distribution and the Poisson distribution. Suppose that events are occurring in time according to a Poisson distribution with a rate of λ events per hour. Thus, in t hours, the number of events—say, Y—will have a Poisson distribution with mean value λt. Suppose that we start at time zero and ask the question, "How long do I have to wait to see the first event occur?"

Let X denote the length of time until this first event. Then,

$$\begin{aligned} P(X > t) &= P[Y = 0 \text{ on the interval } (0, t)] \\ &= \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \\ &= e^{-\lambda t} \end{aligned}$$

and

$$P(X \le t) = 1 - P(X > t) = 1 - e^{-\lambda t}.$$

We see that $P(X \le t) = F(t)$, the distribution function for X, has the form of an exponential distribution function, with $\lambda = (1/\theta)$. Upon differentiating, we see that the probability density function of X is given by

$$\begin{aligned} f(t) &= \frac{\mathrm{d}F(t)}{\mathrm{d}t} \\ &= \frac{\mathrm{d}\left(1 - e^{-\lambda t}\right)}{\mathrm{d}t} \\ &= \lambda e^{-\lambda t} \\ &= \frac{1}{\theta} e^{-t/\theta}, \quad t > 0 \end{aligned}$$

and X has an exponential distribution. Actually, we need not start at time zero, because it can be shown that the waiting time from the occurrence of any one event until the occurrence of the next has an exponential distribution for events occurring according to a Poisson distribution. Similarly, if the number of events X in a specified area has a Poisson distribution, the distance between any event and the next closest event has an exponential distribution.

5.4.4 History and Applications

Karl Pearson first used the term "negative exponential curve" in his *Contributions to the Mathematical Theory of Evolution. II. Skew Variation in Homogeneous Material*, published in 1895 (David and Edwards 2001). However, the curve and its formulation appeared as early as 1774 in a work by Laplace (Stigler 1986).

The primary application of the exponential distribution has been to model the distance, whether in time or space, between events in a Poisson process. Thus the time between the emissions of radioactive particles, the time between telephone calls, the time between equipment failures, the distance between defects of a copper wire, and the distance between soil insects are just some of the many types of data that have been modeled using the exponential distribution.

FIGURE **5.9**
Pierre-Simon Laplace (1749–1827).

The Exponential Distribution

$$f(x) = \begin{cases} \frac{1}{\theta}e^{-x/\theta}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$E(X) = \theta \quad \text{and} \quad V(X) = \theta^2$$

Exercises

5.41 The magnitudes of earthquakes recorded in a region of North America can be modeled by an exponential distribution with a mean of 2.4, as measured on the Richter scale. Find the probabilities that the next earthquake to strike this region will have the following characteristics.

- **a** It will be no more than 2.5 on the Richter scale.
- **b** It will exceed 4.0 on the Richter scale.
- **c** It will fall between 2.0 and 3.0 on the Richter scale.

5.42 Referring to Exercise 5.41, find the probability that of the next 10 earthquakes to strike the region at least 1 will exceed 5.0 on the Richter scale.

5.43 Referring to Exercise 5.41, find the following.

- **a** The variance and standard deviation of the magnitudes of earthquakes for this region
- **b** The magnitude of earthquakes that we can be assured that no more than 10% of the earthquakes will have larger magnitudes on the Richter scale

5.44 A pumping station operator observes that the demand for water at a certain hour of the day can be modeled as an exponential random variable with a mean of 100 cubic feet per second (cfs).

a Find the probability that the demand will exceed 200 cfs on a randomly selected day.

b What is the maximum water-producing capacity that the station should keep on line for this hour so that the demand will have a probability of only 0.01 of exceeding this production capacity?

5.45 Suppose the customers arrive at a certain convenience store checkout counter at a rate of two per minute.

a Find the mean and the variance of the waiting time between successive customer arrivals.

b If a clerk takes 3 minutes to serve the first customer arriving at the counter, what is the probability that at least one more customer will be waiting when the service provided to the first customer is completed?

5.46 In a particular forest, the distance between any randomly selected tree and the tree nearest to it is exponentially distributed with a mean of 40 feet.

a Find the probability that the distance from a randomly selected tree to the tree nearest to it is more than 30 feet.

b Find the probability that the distance from a randomly selected tree to the tree nearest to it is more than 80 feet given that the distance is at least 50 feet.

c Find the minimum distance that separates at least 50% of the trees from their nearest neighbor.

5.47 A roll of copper wire has flaws that occur according to a Poisson process with a rate of 1.5 flaws per meter. Find the following.

a The mean and variance of the distance between successive flaws on the wire

b The probability that the distance between a randomly selected flaw and the next flaw is at least a meter

c The probability that the distance between a randomly selected flaw and the next flaw is no more than 0.2 meter

d The probability that the distance between a randomly selected flaw and the next flaw is between 0.5 and 1.5 meters

5.48 The number of animals killed on a segment of highway is often modeled using a Poisson process (Waller, Servheen, and Peterson et al. 2005) with a mean that depends on the density of traffic on that highway. Thus, the time that it takes an animal to cross the road affects the likelihood of its being hit. For a particular segment of road, the probability of an animal that takes 1 second to cross the road getting hit is 2.3%. Considering only the time that an animal is on the road:

a find the mean times between animals getting hit for this road segment.

b find the probability that a large bear, which will take 5 seconds to cross the road, will get to the other side safely.

5.49 The interaccident times (times between accidents) for all fatal accidents on scheduled American domestic passenger airplane flights for the period 1948 to 1961 were found to follow an exponential distribution with a mean of approximately 44 days (Pyke 1965).

a If one of those accidents occurred on July 1, find the probability that another one occurred in that same (31-day) month.

b Find the variance of the interaccident times.

c What does this information suggest about the clumping of airline accidents?

5.50 The number of hurricanes coming within 250 miles of Honolulu has been modeled according to a Poisson process with a mean of 0.45 per year.

a Find the mean and variance of the time between successive hurricanes coming within 250 miles of Honolulu.

b Given that a hurricane has just occurred, what is the probability that it will be less than 3 months until the next hurricane?

c Given that a hurricane has just occurred, what is the probability that it will be at least a year until the next hurricane will be observed within 250 miles of Honolulu?

d Suppose the last hurricane was 6 months ago. What is the probability that it will be at least another 6 months before the next hurricane comes within 250 miles of Honolulu?

5.51 The breakdowns of an industrial robot follow a Poisson distribution with an average of 0.5 breakdowns per an 8-hour workday. If this robot is placed in service at the beginning of the day, find the probabilities of the following events.

a It will not break down during the day.

b It will work for at least 4 hours without breaking down.

c Does what happened the day before have any effect on your answers? Justify your answer.

5.52 Under average driving conditions, the life lengths of automobile tires of a certain brand are found to follow an exponential distribution with a mean of 30,000 miles. Find the probability that one of these tires, bought today, will last the following number of miles.

a Over 30,000 miles

b Over 30,000 miles given that it already has gone 15,000 miles

5.53 Air samples from a large city are found to have 1-hour carbon monoxide concentrations that are well modeled by an exponential distribution with a mean of 3.6 parts per million (ppm) (Zammers 1984, p. 637).

a Find the probability that a concentration will exceed 9 ppm.

b A traffic control strategy reduced the mean to 2.5 ppm. Now find the probability that a concentration will exceed 9 ppm.

5.54 The weekly rainfall totals for a section of the midwestern United States follow an exponential distribution with a mean of 1.6 inches.

a Find the probability that a randomly chosen weekly rainfall total in this section will exceed 2 inches.

b Find the probability that the weekly rainfall totals will not exceed 2 inches in either of the next 2 weeks.

5.55 Chu (2003) used the exponential distribution to model the time between goals during the 90 minutes of regulation play in the World Cup soccer games from 1990 to 2002. The mean time until the first goal was 33.7 minutes. Assuming that the average time between goals is 33.7 minutes, find the probabilities of the following events.

a The time between goals is less than 10 minutes.

b The time between goals is at least 45 minutes.

c The time between goals is between 5 and 20 minutes.

5.56 Referring to Exercise 5.55, again assume that the time between soccer goals is exponentially distributed with a mean of 33.7 minutes. Suppose that four random selections of the times between consecutive goals are made. Find the probabilities of the following events.

a All four times are less than 30 minutes.

b At least one of the four times is more than 45 minutes.

5.57 Mortgage companies sometimes have customers default on their loans. To understand the effect that it can have on the company, they often model the risk of default. In the reduced form model of credit risk, it is assumed that defaults occur according to a Poisson process with a mean of λ per year. For a particular company, $\lambda = 10$; that is, on average 10 customers default each year.

a What is the probability that the firm will have more than eight defaults in a given year?

b What is the probability that the firm will have no defaults during a given year?

c Given that there has been no default in the first half of the year, what is the probability that there will be at least eight this year?

5.58 A native thistle found in prairies of the northern plains states is distributed according to a Poisson process. For a particular prairie, the mean number of thistles per square meter is 0.5.

a Find the probability that the distance between a randomly selected thistle and its nearest neighbor (the thistle closest to it) is more than 1 meter.

b Suppose 10 thistles are randomly selected. What is the probability that all 10 of the selected thistles are more than 2 meters from their nearest neighbors?

c If instead of randomly selecting a thistle, the researcher randomly selects a point and measures the distance from the point to the nearest thistle. What is the distribution of this distance? Justify your answer.

5.59 The service times at teller windows in a bank were found to follow an exponential distribution with a mean of 3.4 minutes. A customer arrives at a window at 4:00 p.m.

a Find the probability that he will still be there at 4:02 p.m.

b Find the probability that he will still be there at 4:04 p.m. given that he was there at 4:02 p.m.

5.60 The time required to service a customer at a particular post office is exponentially distributed with a mean of 2 minutes. Cindy arrives at the post office, and a customer is currently being helped at each of the two service windows. Given that she will be served as soon as one of the current two customers leaves, what is the probability that Cindy will be the last of the three customers to leave?

5.61 In deciding how many customer service representatives to hire and in planning their schedules, a firm that markets lawnmowers studies repair times for the machines. One such study revealed that repair times have an approximately exponential distribution with a mean of 36 minutes.

a Find the probability that a randomly selected repair time will be less than 10 minutes.

b The charge for lawnmower repairs is \$60 for each half hour (or part thereof) for labor. What is the probability that a repair job will result in a charge for labor of \$120?

c In planning schedules, how much time should the firm allow for each repair to ensure that the chance of any one repair time exceeding this allowed time is only 0.01?

5.62 Explosive devices used in a mining operation cause nearly circular craters to form in a rocky surface. The radii of these craters are exponentially distributed with a mean of 10 feet. Find the mean and the variance of the area covered by such a crater.

5.63 The median m of a continuous random variable X with cumulative distribution function F is that value of m such that $F(m) = 0.5$. Find the median of an exponential distribution with mean θ. How does it compare to the mean?

5.64 Let X be an exponential random variable with mean θ. Show that

$$E(X^k) = k!\,\theta^k$$

Hint: Make use of the gamma function.

5.5 The Gamma Distribution

5.5.1 Probability Density Function

Many sets of data, of course, do not have relative frequency curves with the smooth decreasing trend found in the exponential model. It is perhaps more common to see distributions that have low probabilities for intervals close to zero, with the probability increasing for a while as the interval moves to the right (in the positive direction) and then decreasing as the interval moves out even further; that is, the relative frequency curves follow the pattern graphed as in Figure 5.10. In the case of electronic components, for example, few have very short life lengths, many have something close to an average life length, and very few have extraordinarily long life lengths.

A class of functions that serve as good models for this type of behavior is the *gamma* class. The gamma probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

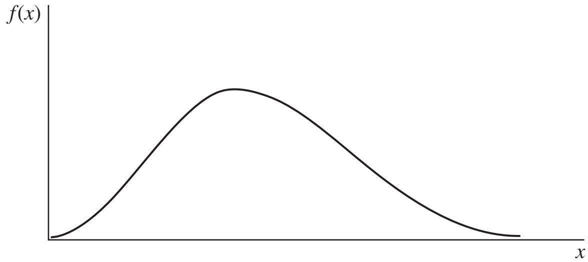


FIGURE 5.10
Common relative frequency curve.

where α and β are parameters that determine the specific shape of the curve. Notice immediately that the gamma density reduces to the exponential density when $\alpha = 1$. The parameters α and β must be positive, but they need not be integers. As we discussed in the last section, the symbol $\Gamma(\alpha)$ is defined by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x}\, dx.$$

A probability density function must integrate to one. Before showing that this is true for the gamma distribution, recall that

$$\int_0^\infty x^{\alpha-1} e^{-x/\beta}\, dx = \beta^\alpha \Gamma(\alpha).$$

It follows then

$$\begin{aligned}
\int_{-\infty}^\infty f(x)\, dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}\, dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty x^{\alpha-1} e^{-x/\beta}\, dx \\
&= \frac{1}{\Gamma(\alpha)\beta^\alpha} \beta^\alpha \Gamma(\alpha) \\
&= 1.
\end{aligned}$$

Some typical gamma densities are shown in Figure 5.11. The probabilities for the gamma distribution cannot be computed easily for all values of α and β. The *Continuous Distributions* applet can be used to graph the gamma density function and compute probabilities. Functions in calculators and computer software are also available for these computations.

An example of a real data set that closely follows a gamma distribution is shown in Figure 5.12. The data consist of 6-week summer rainfall totals for Ames, Iowa. Notice that many totals fall in the range of 2 to 8 inches, but occasionally a rainfall total goes well beyond 8 inches. Of course, no rainfall measurements can be negative.

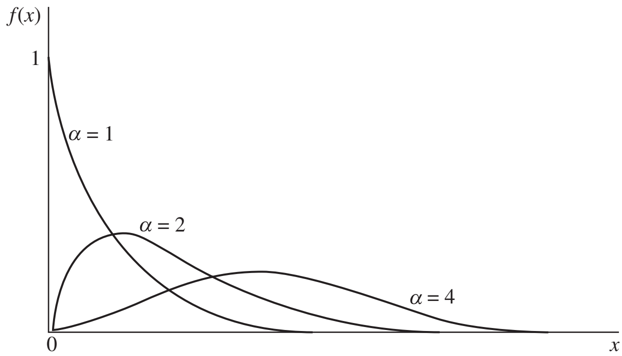


FIGURE 5.11
Gamma density function, $\beta = 1$.

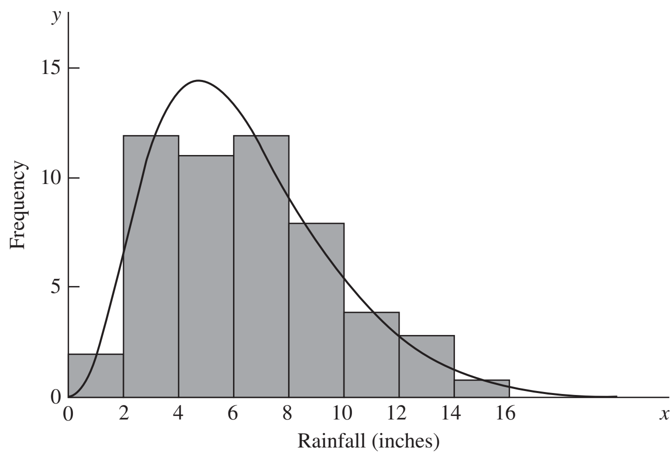


FIGURE 5.12
Summer rainfall (6-week totals) for Ames, Iowa (Barger and Thom 1949).

We have already noted that the exponential distribution is a special case of the gamma distribution with $\alpha = 1$. Another interesting relationship exists between the exponential and gamma distributions. Suppose we have light bulbs. Further assume that the time that each will burn is exponentially distributed with parameter β and that the length of life of one bulb is independent of others. The time until the αth one ceases to burn is gamma with parameters α and β. This is true whether we have n ($> \alpha$) bulbs that burn simultaneously or we burn one bulb after another until α bulbs cease to burn.

5.5.2 Mean and Variance

As might be anticipated because of the relationship of the exponential and gamma distributions, the derivations of expectations here is very similar to the exponential case of Section 5.4. We have

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

$$= \int_0^{\infty} x \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}\, dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \int_0^{\infty} x^{\alpha} e^{-x/\beta}\, dx$$

$$= \frac{1}{\Gamma(\alpha)\beta^{\alpha}} \Gamma(\alpha+1)\beta^{\alpha+1}$$

$$= \alpha\beta.$$

Similar manipulations yield $E\left(X^2\right) = \alpha(\alpha+1)\beta^2$; and hence,

$$\begin{aligned} V(X) &= E\left(X^2\right) - \mu^2 \\ &= \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2 \\ &= \alpha\beta^2. \end{aligned}$$

A simple and often used property of sums of identically distributed, independent gamma random variables will be stated, but not proved, at this point. Suppose that $X_1, X_2, \ldots, X_n$ represent independent gamma random variables with parameters α and β, as just used. If

$$Y = \sum_{i=1}^{n} X_i$$

then Y also has a gamma distribution with parameters $n\alpha$ and β. Thus, one can immediately see that

$$E(Y) = n\alpha\beta$$

and

$$V(Y) = n\alpha\beta^2.$$

EXAMPLE 5.11 A certain electronic system has a life length of X_1, which has an exponential distribution with a mean of 450 hours. The system is supported by an identical backup system that has a life length of X_2. The backup system takes over immediately when the system fails. If the systems operate independently, find the probability distribution and expected value for the total life length of the primary and backup systems.

Solution Letting Y denote the total life length, we have $Y = X_1 + X_2$, where X_1 and X_2 are independent exponential random variables, each with a mean $\beta = 450$. By the results stated earlier, Y has a gamma distribution with $\alpha = 2$ and $\beta = 450$; that is,

$$f_Y(y) = \begin{cases} \frac{1}{\Gamma(2)(450)^2} y e^{-y/450}, & y > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

The mean value is given by

$$E(Y) = \alpha\beta = 2(450) = 900$$

which is intuitively reasonable. ▪

EXAMPLE 5.12 Suppose that the length of time X needed to conduct a periodic maintenance check on a pathology lab's microscope (known from previous experience) follows a gamma distribution with $\alpha = 3$ and $\beta = 2$ (minutes). Suppose that a new repairperson requires 20 minutes to check a particular microscope. Does this time required to perform a maintenance check seem out of line with prior experience?

Solution The mean and the variance for the length of maintenance time (prior experience) are

$$\mu = \alpha\beta \qquad \text{and} \qquad \sigma^2 = \alpha\beta^2.$$

Then, for our example,

$$\mu = \alpha\beta = (3)(2) = 6$$
$$\sigma^2 = \alpha\beta^2 = (3)(2)^2 = 12$$
$$\sigma = \sqrt{12} = 3.446$$

and the observed deviation $(X - \mu)$ is $20 - 6 = 14$ minutes.

For our example, $x = 20$ minutes exceeds the mean $\mu = 6$ by $k = 14/3.46$ standard deviations. Then, from Tchebysheff's Theorem,

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

or

$$P(|X - 6| \geq 14) \leq \frac{1}{k^2} = \left(\frac{3.46}{14}\right)^2 = 0.06.$$

Notice that this probability is based on the assumption that the distribution of maintenance times has not changed from prior experience. Then observing that $P(X \geq 20$ minutes) is small, we must conclude either that our new maintenance person has encountered a machine that needs an unusually lengthy maintenance time (which occurs with low probability) or that the person is somewhat slower than previous repairers. Noting the low probability for $P(X \geq 20)$, we would be inclined to favor the latter view. ▪

5.5.3 History and Applications

In 1893, Karl Pearson (Figure 5.13) presented the first of what would become a whole family of skewed curves; this curve is now known as the gamma distribution (Stigler 1986). It was derived as an approximation to an asymmetric binomial. Pearson initially called this a "generalised form of the normal curve of an asymmetrical character." Later it became known as a Type III curve. It was not until the 1930s and 1940s that the distribution became known as the gamma distribution.

SPL/Photo Researchers, Inc.

FIGURE **5.13**
Karl Pearson (1857–1936).

The gamma distribution often provides a good model to nonnegative, skewed data. Applications include fish lengths, rainfall amounts, and survival times. Its relationship to the exponential makes it considered whenever the time or distance between two or more Poisson events is to be modeled.

The Gamma Distribution

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$E(X) = \alpha\beta \quad \text{and} \quad V(X) = \alpha\beta^2$$

Exercises

5.65 For each month, the daily rainfall (in mm) recorded at the Goztepe rainfall station in the Asian part of Istanbul from 1961 to 1990 was modeled well using a gamma distribution (Aksoy 2000). However, the parameters differed quite markedly from month to month. For September, $\alpha = 0.4$ and $\beta = 20$. Find the following.

a The mean and standard deviation of rainfall during a randomly selected September day at this station

b An interval that will include the daily rainfall for a randomly selected September day with a probability of at least 0.75

5.66 Refer again to Exercise 5.65. For June, $\alpha = 0.5$ and $\beta = 7$. Find the following.

a The mean and standard deviation of rainfall during a randomly selected June day at this station

b An interval that will include the daily rainfall for a randomly selected June day with a probability of at least 0.75

c Compare the results for September and June.

5.67 The weekly downtime X (in hours) for a certain industrial machine has approximately a gamma distribution with $\alpha = 3.5$ and $\beta = 1.5$. The loss L (in dollars) to the industrial operation as a result of this downtime is given by

$$L = 30X + 2X^2.$$

a Find the expected value and the variance of L.

b Find an interval that will contain L on approximately 89% of the weeks that the machine is in use.

5.68 Customers arrive to the checkout counter of a convenience store according to a Poisson process at a rate of two per minute. Find the mean, the variance, and the probability density function of the waiting time between the opening of the counter and the following events.

a The arrival of the second customer

b The arrival of the third customer

5.69 Suppose that two houses are to be built, each involving the completion of a certain key task. Completion of the task has an exponentially distributed time with a mean of 10 hours. Assuming that the completion times are independent for the two houses, find the expected value and the variance of the following times.

a The total time to complete both tasks

b The average time to complete the two tasks

5.70 A population often increases in size until it reaches some equilibrium abundance. However, if the population growth of a particular organism is observed for numerous populations, the populations are not all exactly the same size when they reach the equilibrium. Instead the size fluctuates about some average size. Dennis and Costantino (1988) suggested the gamma distribution as a model of the equilibrium population size. When studying the flour beetle *Tribolium castaneum*, the gamma distribution with parameters $\alpha = 5.5$ and $\beta = 5$ provided a good model for these equilibrium population sizes.

a Find the mean and variance of the equilibrium population size for the flour beetle.

b Find an interval that will include at least 75% of the equilibrium population sizes of this flour beetle.

5.71 Over a 30-minute time interval the distance that largemouth bass traveled were found to be well modeled using an exponential distribution with a mean of 20 meters (Essington and Kitchell 1999).

a Find the probability that a randomly selected largemouth bass will move more than 50 meters in 30 minutes.

b Find the probability that a randomly selected largemouth bass will move less than 10 meters in 30 minutes.

c Find the probability that a randomly selected largemouth bass will move between 20 and 60 meters in 30 minutes.

d Give the probability density function, including parameters, of the distance that a largemouth bass moves in 1 hour.

5.72 Refer to the setting in Exercise 5.71.

a Find the mean and variance of the total distance that two randomly selected largemouth bass will travel in 30 minutes.

b Give an interval that will include the total distance that two randomly selected largemouth bass will travel in 30 minutes with 75% probability.

5.73 The total sustained load on the concrete footing of a planned building is the sum of the dead load plus the occupancy load. Suppose that the dead load X_1 has a gamma distribution with $\alpha_1 = 50$ and $\beta_1 = 2$, whereas the occupancy load X_2 also has a gamma distribution but with $\alpha_2 = 20$ and $\beta_2 = 2$. (Units are in kips, or thousands of pounds.)

a Find the mean, the variance, and the probability density function of the total sustained load on the footing.

b Find a value for the sustained load that should be exceeded only with a probability of less than $1/16$.

5.74 A 40-year history of annual maximum river flows for a certain small river in the United States shows a relative frequency histogram that can be modeled by a gamma density function with $\alpha = 1.6$ and $\beta = 150$ (measurements in cubic feet per second).

a Find the mean and the standard deviation of the annual maximum river flows.

b Within what interval should the maximum annual flow be contained with a probability of at least $8/9$?

5.75 Let X be a gamma random variable with parameters α and β. For an integer $k > 0$, find the $E(X^k)$.

5.76 Let X be a gamma random variable with parameters α and β, where α is an integer. Let Y be a Poisson random variable with mean α/β. Show that

$$P(X \leq x) = P(Y \geq \alpha)$$

Hint: Use successive integration by parts and the fact that $\Gamma(\alpha) = (\alpha - 1)!$

5.6 The Normal Distribution

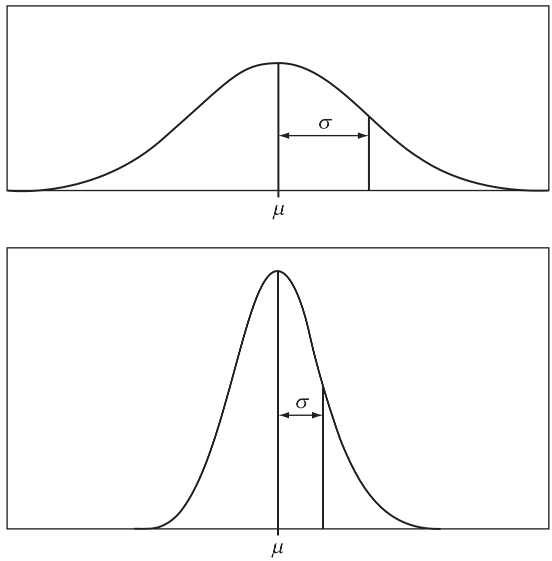
5.6.1 Normal Probability Density Function

The most widely used continuous probability distribution is referred to as the *normal distribution*. The normal probability density function has the familiar symmetric "bell" shape, as indicated in the two graphs shown in Figure 5.14. The curve is centered at the mean value μ, and its spread is measured by the standard deviation σ. These two parameters, μ and σ^2, completely determine the shape and location of the normal density function whose functional form is given by

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

The basic reason that the normal distribution works well as a model for many different types of measurements generated in real experiments is discussed in some detail in Chapter 8. For now it suffices to say that any time responses tend to be averages of independent quantities, the normal distribution quite likely will provide a reasonably good model for their relative frequency behavior. Many naturally occurring measurements tend to have relative frequency distributions that closely resemble the normal curve, probably because nature tends to "average out" the effects of the

FIGURE 5.14
Normal density functions.



many variables that relate to a particular response. For example, heights of adult American men tend to have a distribution that shows many measurements clumped closely about a mean height with relatively few very short or very tall men in the population. In other words, the relative frequency distribution is close to normal. In contrast, life lengths of biological organisms or electronic components tend to have relative frequency distributions that are not normal or even close to normal. This often is because life-length measurements are a product of "extreme" behavior, not "average" behavior. A component may fail because of one extremely severe shock rather than the average effect of many shocks. Thus, the normal distribution is not often used to model life lengths, and we will not discuss the failure rate function for this distribution.

A naturally occurring example of the normal distribution is seen in Michelson's measurements of the speed of light. A histogram of these measurements is given in Figure 5.15. The distribution is not perfectly symmetrical, but it still exhibits an approximately normal shape.

5.6.2 Mean and Variance

A very important property of the normal distribution (which is proved in Section 5.9) is that any linear function of a normally distributed random variable also is normally distributed; that is, if X has a normal distribution with a mean of μ and a variance of σ^2, and if $Y = aX + b$ for constants a and b, then Y also is normally distributed. It can easily be seen that

$$E(Y) = a\mu + b \quad \text{and} \quad V(Y) = a^2\sigma^2.$$

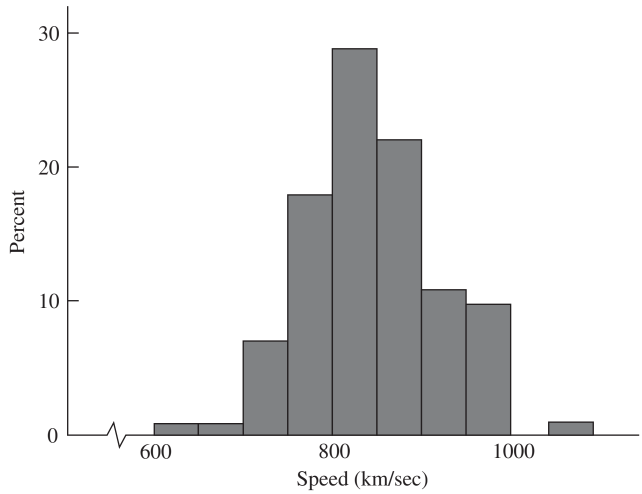


FIGURE 5.15
Michelson's (1878) 100 measures of the speed of light in air (−299,000 km/s).

Suppose that Z has a normal distribution with $\mu = 0$ and $\sigma = 1$. This random variable Z is said to have a *standard normal distribution*. Direct integration will show that $E(Z) = 0$ and $V(Z) = 1$. We have

$$\begin{aligned}
E(Z) &= \int_{-\infty}^{\infty} zf(z)\,dz \\
&= \int_{-\infty}^{\infty} z\frac{1}{\sqrt{2\pi}}e^{-z^2}\,dz \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} ze^{-z^2/2}z\,dz \\
&= \frac{1}{\sqrt{2\pi}}\left[-e^{-z^2/2}\right]_{-\infty}^{\infty} \\
&= 0.
\end{aligned}$$

Similarly,

$$\begin{aligned}
E(Z^2) &= \int_{-\infty}^{\infty} z^2\frac{1}{\sqrt{2\pi}}e^{-z^2}\,dz \\
&= \frac{1}{\sqrt{2\pi}}(2)\int_{0}^{\infty} z^2e^{-z^2/2}\,dz.
\end{aligned}$$

On our making the transformation $u = z^2$, the integral becomes

$$E(Z^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} u^{1/2}e^{-u/2}\,du$$

$$= \frac{1}{\sqrt{2\pi}} \Gamma\left(\frac{3}{2}\right) (2)^{3/2}$$
$$= \frac{1}{\sqrt{\pi}} 2\left(\frac{1}{2}\right) \Gamma\left(\frac{1}{2}\right)$$
$$= 1$$

because $\Gamma(1/2) = \sqrt{\pi}$. Therefore, $E(Z) = 0$ and $V(Z) = E(Z^2) - \mu^2 = E(Z^2) = 1$.

For any normally distributed random variable X, with parameters μ and σ^2,

$$Z = \frac{X - \mu}{\sigma}$$

will have a standard normal distribution. Then

$$X = Z\sigma + \mu,$$

$$E(X) = \sigma E(Z) + \mu = \mu$$

and

$$V(X) = \sigma^2 V(Z) = \sigma^2.$$

This shows that the parameters μ and σ^2 do, indeed, measure the mean and the variance of the distribution.

5.6.3 Calculating Normal Probabilities

Because any normally distributed random variable can be transformed into standard normal form, probabilities can be evaluated for any normal distribution simply by evaluating the appropriate standard normal integral. By inspecting any standard normal integral over the interval (a, b), it quickly becomes evident that this is no trivial task; there is no closed form solution for these integrals. The *Continuous Distribution* applet as well as numerous software packages and calculators can be used to quickly evaluate these integrals. Tables of standard normal integrals can also be used; one such table is given in Table 4 of the Appendix. Table 4 gives numerical values for

$$P(0 \leq Z \leq z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz.$$

Values of the integral are given for values of z between 0.00 and 3.09.

We shall now use Table 4 to find $P(-0.5 \leq Z \leq 1.5)$ for a standard normal variable X. Figure 5.16 helps us visualize the necessary areas.

We first must write the probability in terms of intervals to the left and to the right of zero (the mean of the distribution). This produces

$$P(-0.5 \leq Z \leq 1.5) = P(0 \leq Z \leq 1.5) + P(-0.5 \leq Z \leq 0).$$

Now $(0 \leq Z \leq 1.5) = A_1$ in Figure 5.16, and this value can be found by looking up $z = 1.5$ in Table 4. The result is $A_1 = 0.4332$. Similarly, $P(-0.5 \leq Z \leq 0) = A_2$ in

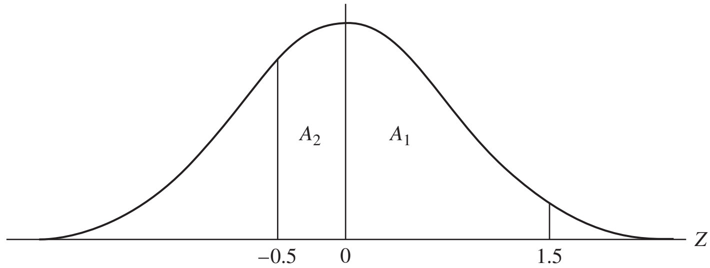


FIGURE **5.16**
Standard normal density function.

Figure 5.14, and this value can be found by looking up $z = 0.5$ in Table 4. Areas under the standard normal curve for negative z-values are equal to those for corresponding positive z-values, because the curve is symmetric around zero. We find $A_2 = 0.1915$. It follows that

$$P(-0.5 \le Z \le 1.5) = A_1 + A_2 = 0.4332 + 0.1915 = 0.6247.$$

EXAMPLE **5.13** If Z denotes a standard normal variable, find the following probabilities.

1 $P(Z \le 1.5)$
2 $P(Z \ge 1.5)$
3 $P(Z < -2)$
4 $P(-2 \le Z \le 1)$
5 Also find a value of z—say z_0—such that $P(0 \le Z \le z_0) = 0.35$.

Solution This example provides practice in reading Table 4. We see from the table that the following values are correct.

1
$$\begin{aligned} P(Z \le 1.5) &= P(Z \le 0) + P(0 \le Z \le 1.5) \\ &= 0.5 + 0.4332 \\ &= 0.9332 \end{aligned}$$

2
$$\begin{aligned} P(Z \ge 1.5) &= 1 - P(Z < 1.5) \\ &= 1 - P(Z \le 1.5) \\ &= 1 - 0.9332 \\ &= 0.0668 \end{aligned}$$

3
$$\begin{aligned} P(Z < -2) &= P(Z > 2) \\ &= 0.5 - P(0 \le Z \le 2) \\ &= 0.5 - 0.4772 \\ &= 0.0228 \end{aligned}$$

4 $$\begin{aligned} P(-2 \le Z \le 1) &= P(-2 \le Z \le 0) + P(0 \le Z \le 1) \\ &= P(0 \le Z \le 2) + P(0 \le Z \le 1) \\ &= 0.4772 + 0.3413 \\ &= 0.8185 \end{aligned}$$

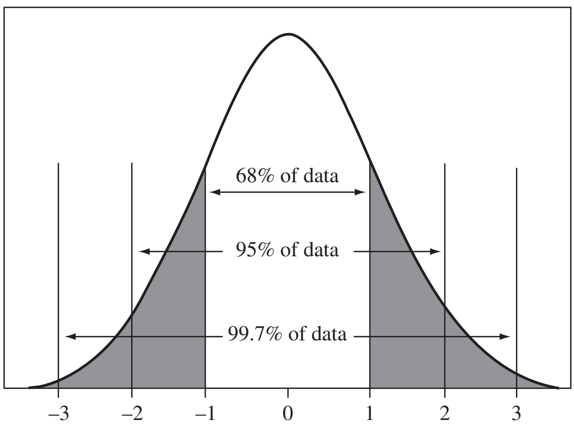
5 To find the value of z_0, we must look for the given probability of 0.35 in the body of Table.4. The closest we can come to is 0.3508, which corresponds to a z-value of 1.04. Hence $z_0 = 1.04$. We could have also used the *Continuous Distributions* applet, a calculator, or a computer software package. ▪

The empirical rule states that for any bell-shaped curve, approximately:

- 68% of the values fall within 1 standard deviation of the mean in either direction.
- 95% of the values fall within 2 standard deviations of the mean in either direction.
- 99.7% of the values fall within 3 standard deviations of the mean in either direction.

Studying the table of normal curve areas for z-scores of 1, 2, and 3 reveals how the percentages used in the empirical rule were determined. These percentages actually represent areas under the standard normal curve, as depicted in Figure 5.17. The next example illustrates how the standardization works to allow Table 4 to be used for any normally distributed random variable.

FIGURE **5.17**
Justification of the empirical rule.



EXAMPLE **5.14** A firm that manufactures and bottles apple juice has a machine that automatically fills bottles with 16 ounces of juice. (The bottle can hold up to 17 ounces.) Over a long period, the average amount dispensed into the bottle has been 16 ounces. However, there is variability in how much juice is put in each bottle; the distribution of these amounts has a standard deviation of 1 ounce. If the ounces of fill per bottle can be assumed to be normally distributed, find the probability that the machine will overflow any one bottle.

Solution A bottle will overflow if the machine attempts to put more than 17 ounces in it. Let X denote the amount of liquid (in ounces) dispensed into one bottle by the filling machine. Then X is assumed to be normally distributed with a mean of 16 and a standard deviation of 1. Hence,

$$\begin{aligned}
P(X > 17) &= P\left(\frac{X-\mu}{\sigma} > \frac{17-\mu}{\sigma}\right) \\
&= P\left(Z > \frac{17-16}{1}\right) \\
&= P(Z > 1) \\
&= 0.1587.
\end{aligned}$$

The answer can be found from Table 4, because $Z = (X - \mu)/\sigma$ has a *standard* normal distribution or by using the *Continuous Distributions* applet. ▪

EXAMPLE **5.15** Suppose that another machine similar to the one described in Example 5.14 is operating in such a way that the ounces of fill have a mean value equal to the dial setting for "amount of liquid" but also has a standard deviation of 1.2 ounces. Find the proper setting for the dial so that the 17-ounce bottle will overflow only 5% of the time. Assume that the amounts dispensed have a normal distribution.

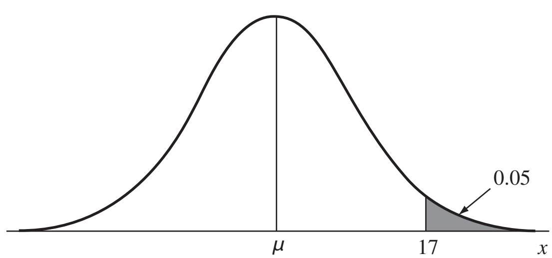
Solution Let X denote the amount of liquid dispensed; we now look for a value of μ such that

$$P(X > 17) = 0.05.$$

This is represented graphically in Figure 5.18. Now

$$\begin{aligned}
P(X > 17) &= P\left(\frac{X-\mu}{\sigma} > \frac{17-\mu}{\sigma}\right) \\
&= P\left(Z > \frac{17-\mu}{1.2}\right).
\end{aligned}$$

FIGURE **5.18**



From Table 4, we know that if

$$P(Z > z_0) = 0.05$$

then $z_0 = 1.645$. Thus, it must be that

$$\frac{17 - \mu}{1.2} = 1.645$$

and

$$\mu = 17 - 1.2(1.645) = 15.026. \quad \blacksquare$$

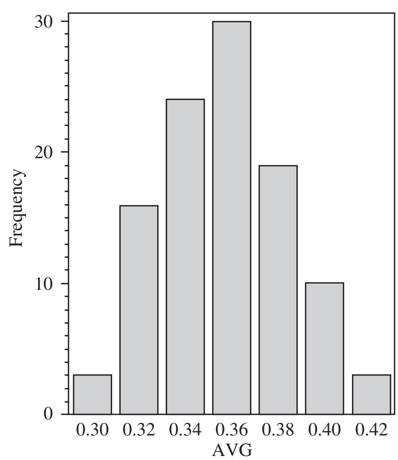
5.6.4 Applications to Real Data

The practical value of probability distributions in relation to data analysis is that the probability models help explain key features of the data succinctly and aid in the constructive use of data to help predict future outcomes. Patterns that appeared regularly in the past are expected to appear again in the future. If they do not, something of importance may have happened to disturb the process under study. Describing data patterns through their probability distribution is a very useful tool for data analysis.

EXAMPLE 5.16 The batting averages of the American League batting champions for the years 1901 through 2005 are graphed on the histogram in Figure 5.19 (*Baseball Almanac* 2006). This graph looks somewhat normal in shape, but it has a slight skew toward the high values. The mean is 0.357, and the standard deviation is 0.027 for these data.

1 Ted Williams batted 0.406 in 1941, and George Brett batted 0.390 in 1980. How would you compare these performances?

FIGURE 5.19 Batting averages: American League batting champions, 1901–2005.



2 Is there a good chance that anyone in the American League will hit over 0.400 in any 1 year?

Solution 1 Obviously, 0.406 is better than 0.390, but how much better? One way to describe how these numbers compare to each other—and how they compare to the remaining data points in the distribution—is to look at z-scores and percentile scores. For 1941, Ted Williams had a z-score of

$$z = \frac{0.406 - 0.357}{0.027} = 1.81$$

and a percentile score of

$$0.5 + 0.4649 = 0.9649.$$

For 1980, George Brett had a z-score of

$$z = \frac{0.390 - 0.357}{0.027} = 1.22$$

and a percentile score of

$$0.50 + 0.3888 = 0.8888.$$

Both are in the upper quintile and are far above average. Yet, Williams is above the 96th percentile and Brett is below the 89th. Although close, Williams's average appears significantly higher than Brett's.

2 The chance of the league leader's hitting over 0.400 in a given year can be approximated by looking at a z-score of

$$z = \frac{0.400 - 0.357}{0.027} = 1.59.$$

This translates into a probability of

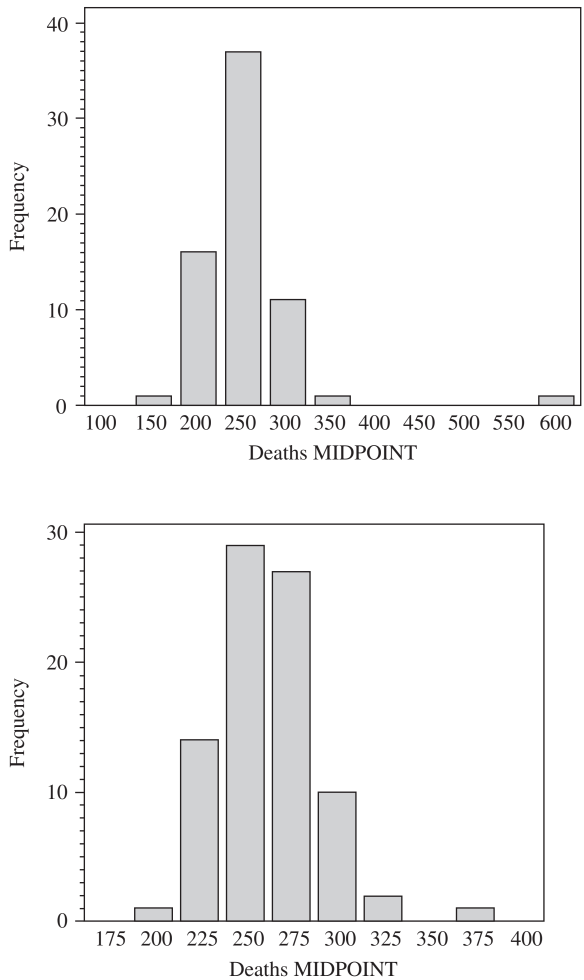
$$0.5 - 0.4441 = 0.0559$$

or about 6 chances out of 100. (This is the probability for the league leader. What would happen to this chance if *any* player were eligible for consideration?) ▪

What happens when the normal model is used to describe skewed distributions? To answer this, let's look at two data sets involving comparable variables and see how good the normal approximations are. Figure 5.20 shows histograms of cancer mortality rates for white males during the years 1998 to 2002. (These rates are expressed as deaths per 100,000 people.) The top graph shows data for the 67 counties of Florida, and the bottom graph shows data for the 84 counties of Oklahoma. Summary statistics are as follows:

State	Mean	Standard Deviation
Florida	252.9	55.3
Oklahoma	262.3	28.4

FIGURE **5.20**
Histograms of cancer mortality rates for white males in Florida (above) and Oklahoma (below) from 1998 to 2002.



On average, the states perform about the same, but the distributions are quite different. Key features of the difference can be observed by comparing the empirical relative frequencies to the theoretical relative frequencies, as shown in Table 5.2.

For Oklahoma, observation and theory are close together, although the one-standard-deviation interval and the two-standard-deviation interval, to a lesser extent, pick up a few too many data points. For Florida, observation and theory are not close. In this case, the large outlier inflates the standard deviation so that the

TABLE **5.2**
Empirical and Theoretical Normal Distribution Relative Frequencies for County Cancer Rates in Florida and Oklahoma.

Interval	Observed Proportion Florida	Oklahoma	Theoretical Proportion
$\bar{x} \pm s$	$\frac{60}{67} = 0.896$	$\frac{63}{84} = 0.75$	0.68
$\bar{x} \pm 2s$	$\frac{66}{67} = 0.985$	$\frac{81}{84} = 0.964$	0.95

one-standard-deviation interval is far too wide to agree with normal theory. The two-standard-deviation interval also has an observed relative frequency larger than expected. This is typical of the performance of relative frequencies in highly skewed situations; be careful in interpreting standard deviation under skewed conditions.

5.6.5 Quantile-Quantile (Q-Q) Plots

The normal model is popular for describing data distributions but as we have just seen, it often does not work well in instances where the data distribution is skewed. How can we recognize such instances and avoid using the normal model in situations where it would be inappropriate? One way, as we have seen, is to look carefully at histograms, dotplots, and stemplots to gauge symmetry and outliers visually. Another way, which is presented in this subsection, is to take advantage of the unique properties of z-scores for the normal case. If X has a normal (μ, σ) distribution, then

$$X = \mu + \sigma Z$$

and there is a perfect linear relationship between X and Z. Now suppose that we observe n measurements and order them so that $x_1 \le x_2 \le x_3 \le \cdots \le x_n$. The value of x_k has (k/n) values that are less than or equal to it, so it is the (k/n)th sample percentile. If the observations come from a normal distribution, x_k should approximate the (k/n)th percentile from the normal distribution and, therefore, should be linearly related to z_k (the corresponding z-score). The cancer data used in the preceding subsection (see Figure 5.18) serve to illustrate this point. The sample percentiles corresponding to each state were determined for the 25th, 50th, 75th, 95th, and 99th percentiles as shown in Table 5.3. The z-scores corresponding to these percentiles for the normal distribution are also listed in the table.

In other words, 25% of the Florida data fell on or below 220, and 95% of the Okahoma data fell on or below 311. For the normal distribution, 25% of the area will fall below a point with a z-score of -0.68.

Plots of the sample percentiles against the z-scores are shown in Figure 5.21. Such plots are called quantile-quantile or Q-Q plots. For Florida, four points fall on a line,

TABLE **5.3**
z-Scores for Florida and Oklahoma County Cancer Rates from 1998 to 2002.

Percentile	Mortality Rate Florida	Oklahoma	z-score
25th	220	241	−0.680
50th	249	259	0.000
75th	264	277	0.680
95th	319	311	1.645
99th	609	370	2.330

F I G U R E **5.21** Florida (above) and Oklahoma (below) results for five percentiles (25, 50, 75, 95, and 99).

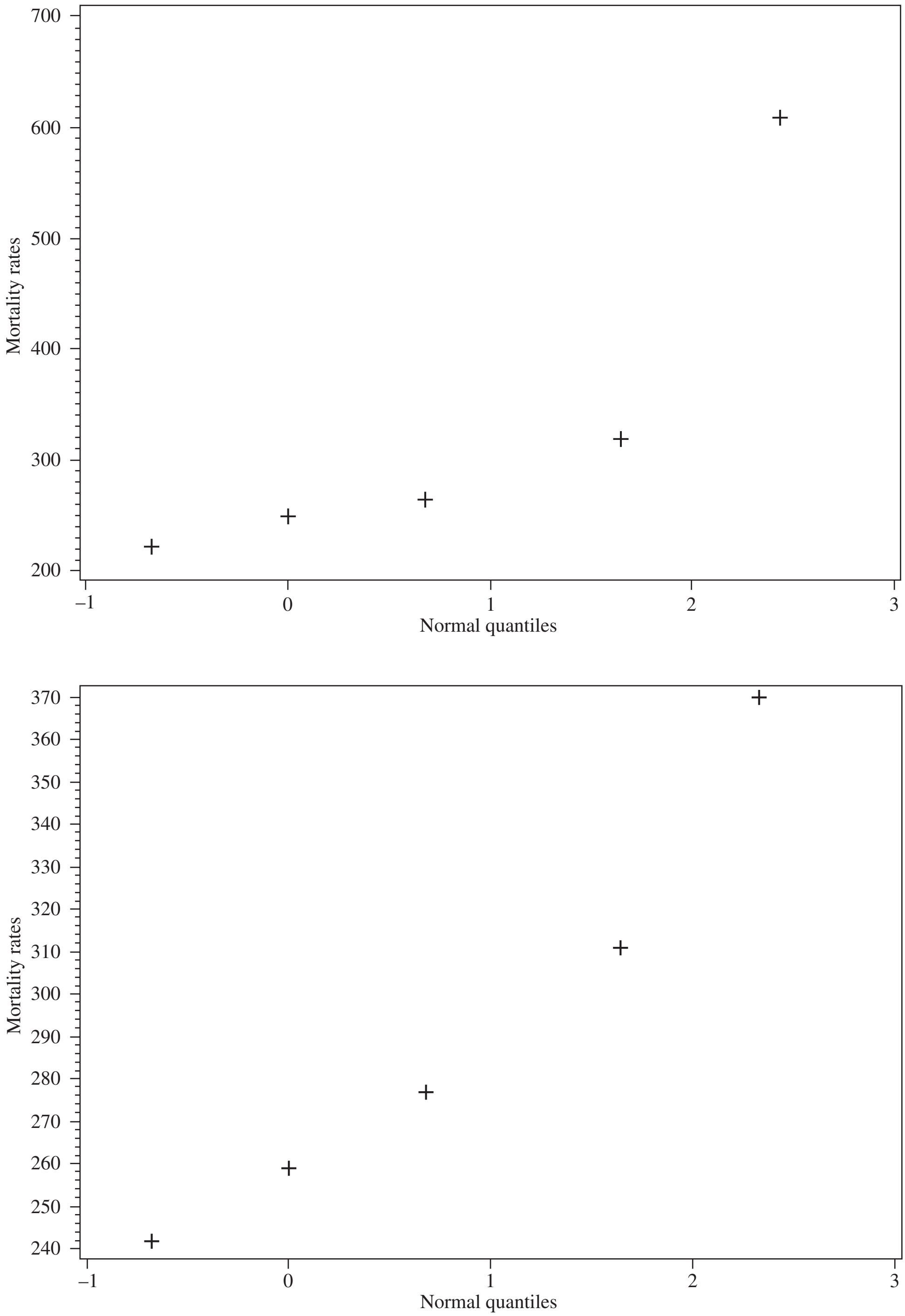
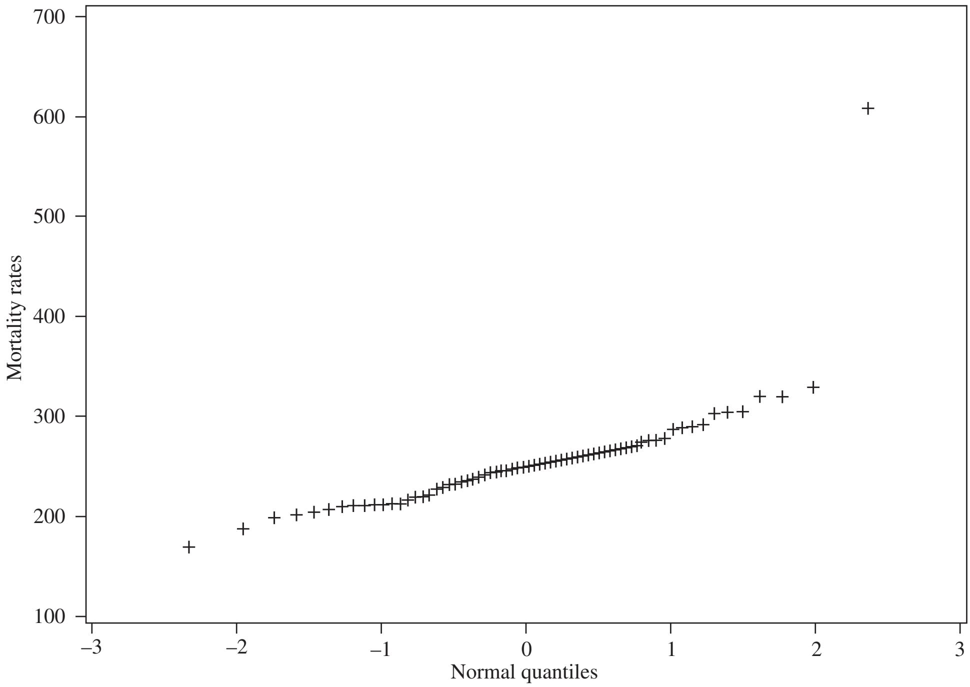
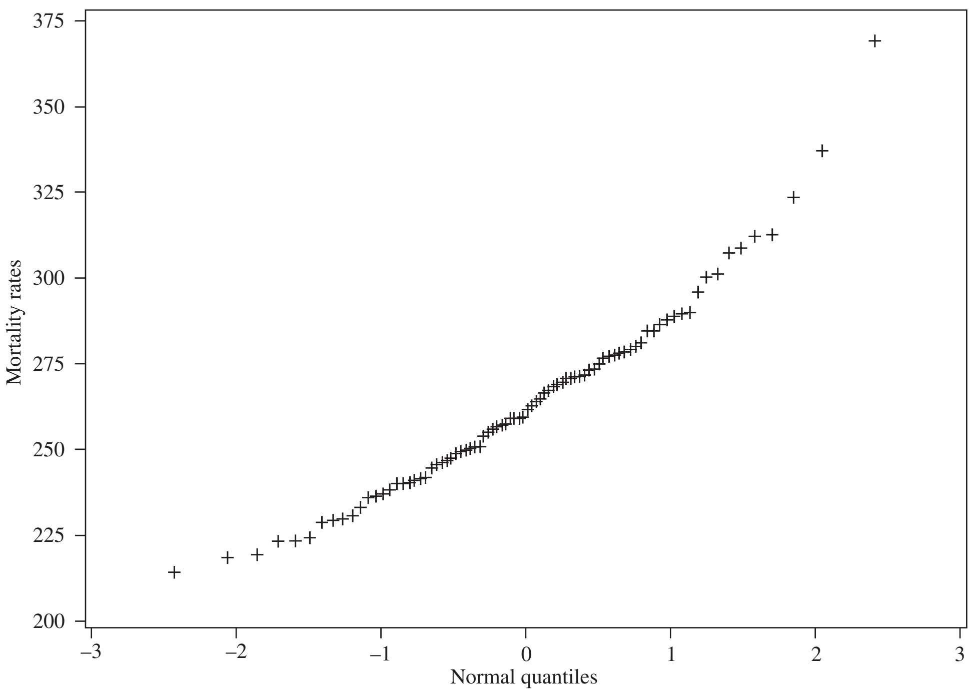


FIGURE 5.22 Q-Q plots for all of Florida (above) and Oklahoma (below) counties.





or nearly so, but the fifth point is far off the line. The observed 99th percentile is much larger than would have been expected from the normal distribution, indicating a skewness toward the larger data point(s).

For Oklahoma, the sample percentiles and z-scores fall nearly on a straight line, with a sag in the middle, indicating a reasonable, though not a really good, fit of the data to a normal distribution. The intercept of the line fit to these data is around 260, which is close to the sample mean. The slope of the line is around 40, which is below the sample standard deviation. As the fit of the normal data to the model improves, the slope of the line should become more similar to the sample standard deviation. Figure 5.22 shows the plots for all the sample data with the same result.

For small samples, it is better to think of x_k as the $(k/n+1)$th sample percentile. The reason for this is that $x_1 \leq x_2 \leq x_3 \leq \cdots \leq x_n$ actually divide the population distribution into $(n+1)$ segments, all of which are expected to process roughly equal probability masses.

EXAMPLE 5.17 In the interest of predicting future peak particulate matter values, 12 observations of this variable were obtained from various sites across the United States. The ordered measurements, in $\mu\text{g/m}^2$, are as follows:

22, 24, 24, 28, 30, 31, 33, 45, 45, 48, 51, 79

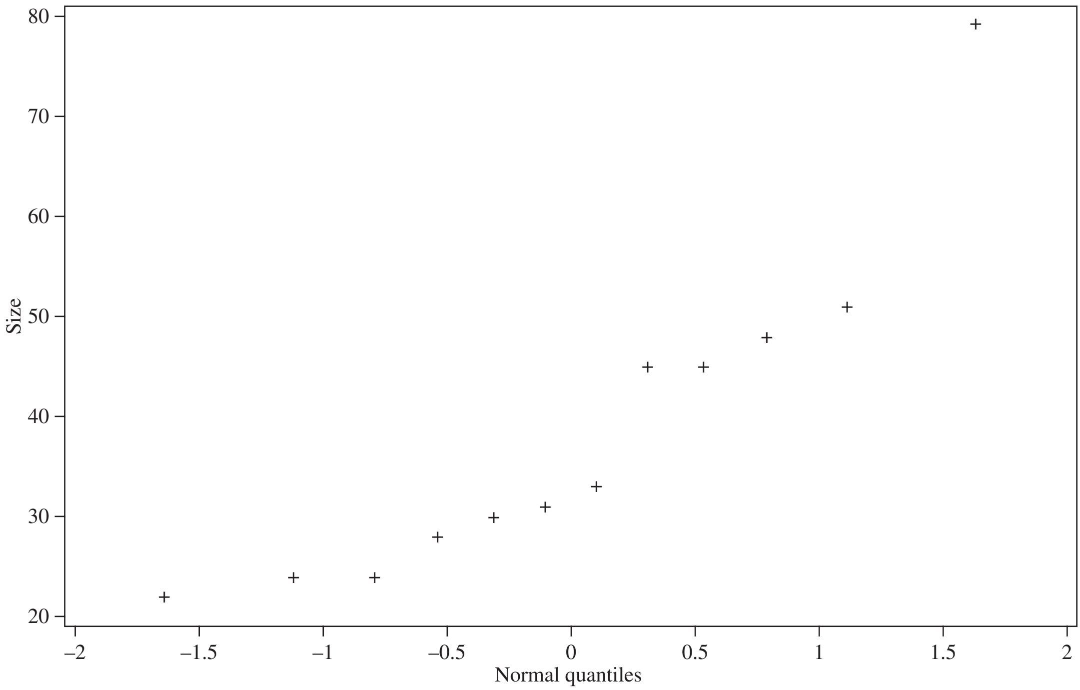
Should the normal probability model be used to anticipate peak particulate matter values in the future?

Solution Table 5.4 provides the key components of the analysis. Recall that the z-scores are the standard normal values corresponding to the percentiles listed in the $i/(n+1)$ column. For example, about 69% of the normal curve's area falls below a z-score of 0.50. The Q-Q plot of x_i versus z_i in Figure 5.23 shows that the data depart from normality at both ends: the lower values have too short a tail, and the higher values have too long a tail. The data points appear to come from a highly skewed distribution; it would be unwise to predict future values for this variable by using a normal distribution. ▪

TABLE 5.4
z-Scores for Example 5.17.

i	x_i	$i/(n+1)$	z-Score
1	22	0.077	−1.43
2	24	0.153	−1.02
3	24	0.231	−0.73
4	28	0.308	−0.50
5	30	0.385	−0.29
6	31	0.462	−0.10
7	33	0.538	0.10
8	45	0.615	0.29
9	45	0.692	0.50
10	48	0.769	0.74
11	51	0.846	1.02
12	79	0.923	1.43

FIGURE 5.23 Q-Q plot of data from Example 5.17.



Q-Q plots are cumbersome to construct by hand, especially if the data set is large, but most computer programs for statistical analysis can generate the essential parts quite easily. In Minitab, for example, issuing the NSCORES (for normal scores) command prompts Minitab to produce the z-scores corresponding to a set of sample data. Many calculators can also produce Q-Q plots for a set of data.

EXAMPLE 5.18 The heights of 20-year-old males have the cumulative relative frequency distribution given in Table 5.5.

Should the normal distribution be used to model the male height distribution? If so, what mean and what standard deviation should be used?

Solution Here the percentiles can easily be derived from the cumulative percentages simply by dividing by 100. Using an inverse normal probability function (going from probability to z-scores) on a calculator or computer or a normal area curve area table, we find the corresponding z-scores are as follows:

$$-1.881,\ -1.645,\ -1.282,\ -0.674,\ 0,\ 0.674,\ 1.282,\ 1.645,\ 1.881$$

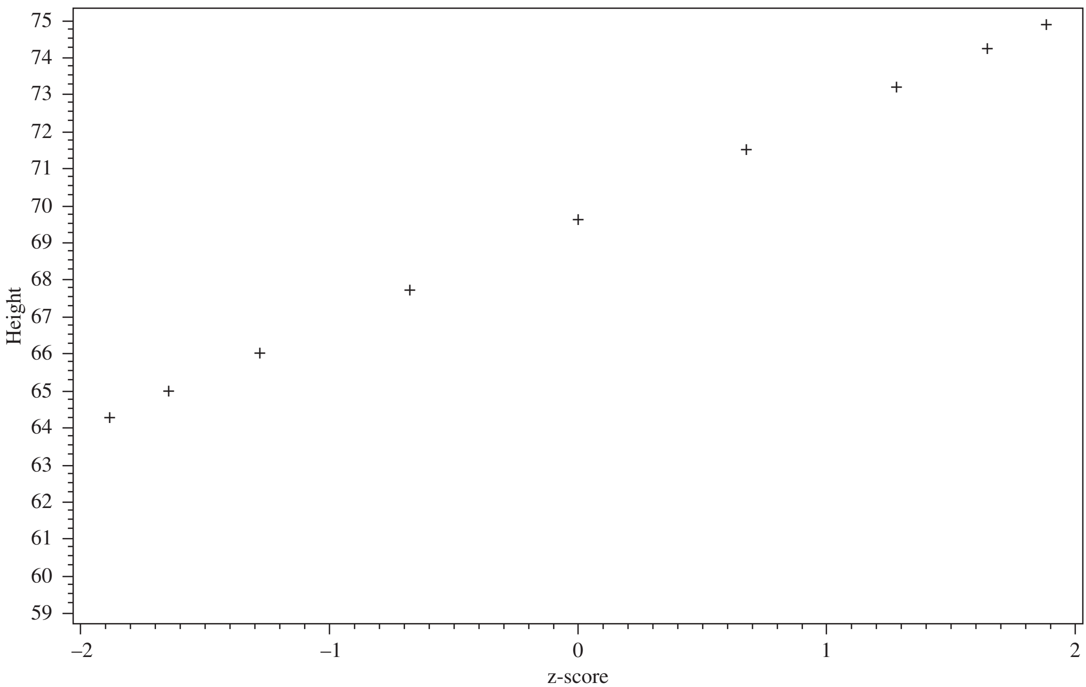
The Q-Q plot of heights versus z-scores appears in Figure 5.24.

TABLE **5.5**
Heights, in Inches, of 20-Year-Old Males in the United States.

Cumulative Percentage	Height Inches
3	64.30
5	64.98
10	66.01
25	67.73
50	69.63
75	71.52
90	73.21
95	74.22
97	74.88

Source: U.S. Centers for Disease Control.

FIGURE **5.24** Q-Q plot for data in Example 5.18.



The plot shows a nearly straight line; thus, the normal distribution can reasonably be used to model these data.

Marking a line with a straightedge through these points to approximate the line of best fit for the data yields a slope of approximately 2.8 and a y-intercept (at $z = 0$) of 69.6 inches. Thus, male heights can be modeled by a normal distribution with a mean of 69.6 inches and a standard deviation of 2.8 inches. (Can you verify the mean and the standard deviation approximations by another method?) ■

5.6.6 History

In the eighteenth century, Abraham de Moivre, a statistician who consulted with gamblers, often found a need to compute binomial probabilities (Figure 5.25). As n became larger, the computations became challenging, especially because his work predates the advent of calculators or computers. After noting that as n became large, the distribution of probabilities approach a smooth curve, de Moivre derived the equation for this curve. This formula was published on November 12, 1733, and represents the first appearance of the normal distribution in the literature.

An early application of the normal distribution was as a model of the distribution of measurement errors in astronomy. In the seventeenth century, Galileo observed that these errors tended to have a symmetric distribution and that small errors occurred with greater frequency. The formula for describing this distribution was developed independently by Adrian in 1808 and by Gauss in 1809 (Figure 5.26). They showed

FIGURE **5.25**
Abraham de Moivre (1667–1754).

FIGURE **5.26**
John Carl Friedrich Gauss (1777–1855).

that the normal distribution fit the distributions of measurement data well. Laplace had encountered the normal distribution in 1778 when he developed the Central Limit Theorem, but that is a topic for Chapter 8. Even though de Moivre's derivation preceded that of Gauss by a number of years, the normal is frequently referred to as the Gaussian distribution.

As was mentioned earlier, we shall make more use of the normal distribution in later chapters. The properties of the normal distribution are summarized next.

The Normal Distribution

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

$$E(X) = \mu \quad \text{and} \quad V(X) = \sigma^2$$

Exercises

5.77 Find the following probabilities for a standard normal random variable Z.

a $P(0 \le Z \le 0.8)$
b $P(-1.1 \le Z \le 0)$
c $P(0.5 \le Z \le 1.87)$
d $P(-1.1 \le Z \le 1.1)$
e $P(-0.8 \le Z \le 1.9)$

5.78 Find the following probabilities for a standard normal random variable Z.

a $P(Z \le 1.2)$
b $P(-1.5 \le Z)$
c $P(-1.4 \le Z \le -0.2)$
d $P(-0.25 \le Z \le 1.76)$
e $P(|Z| > 2)$

5.79 For a standard normal random variable Z, find a number z_0 such that the following probabilities are obtained.

a $P(Z \le z_0) = 0.5$
b $P(Z \le z_0) = 0.80$
c $P(Z \ge z_0) = 0.14$
d $P(Z \ge z_0) = 0.69$
e $P(-z_0 \le Z \le z_0) = 0.90$
f $P(|Z| \ge z_0) = 0.90$

5.80 For a standard normal random variable Z, find a number z_0 such that the following probabilities are obtained.

a $P(Z \le z_0) = 0.1$
b $P(Z \le z_0) = 0.78$
c $P(Z \ge z_0) = 0.05$
d $P(Z \ge z_0) = 0.82$
e $P(-z_0 \le Z \le z_0) = 0.95$
f $P(|Z| \ge z_0) = 0.50$

5.81 Find the following probabilities for a normal random variable X that has a mean of 10 and a standard deviation of 2.

a $P(10 \le X \le 12)$
b $P(8 \le X \le 10)$
c $P(11 \le X \le 14)$
d $P(9 \le X \le 11)$
e $P(7 \le Z \le 13)$

5.82 Find the following probabilities for a normal random variable X that has a mean of 15 and a standard deviation of 3.

a $P(X \leq 17)$ **b** $P(11 \leq X)$

c $P(10 \leq X \leq 14)$ **d** $P(12 \leq X \leq 20)$

5.83 For a normal random variable X with a mean of 12 and a standard deviation of 3, find a number x_0 such that the following probabilities are obtained.

a $P(X \leq x_0) = 0.5$ **b** $P(X \leq x_0) = 0.84$

c $P(X \geq x_0) = 0.23$ **d** $P(X \geq x_0) = 0.59$

5.84 For a normal random variable X with a mean of 8 and standard deviation of 2, find a number x_0 such that the following probabilities are obtained.

a $P(X \leq x_0) = 0.2$ **b** $P(X \leq x_0) = 0.72$

c $P(X \geq x_0) = 0.05$ **d** $P(X \geq x_0) = 0.95$

5.85 The weekly amount spent for maintenance and repairs in a certain company has an approximately normal distribution with a mean of \$600 and a standard deviation of \$40. If \$700 is budgeted to cover repairs for next week, what is the probability that the actual costs will exceed the budgeted amount?

5.86 In the setting of Exercise 5.85, how much should be budgeted weekly for maintenance and repairs to ensure that the probability that the budgeted amount will be exceeded in any given week is only 0.1?

5.87 A machining operation produces steel shafts where diameters have a normal distribution with a mean of 1.005 inches and a standard deviation of 0.01 inch. Specifications call for diameters to fall within the interval 1.00 ± 0.02 inches.

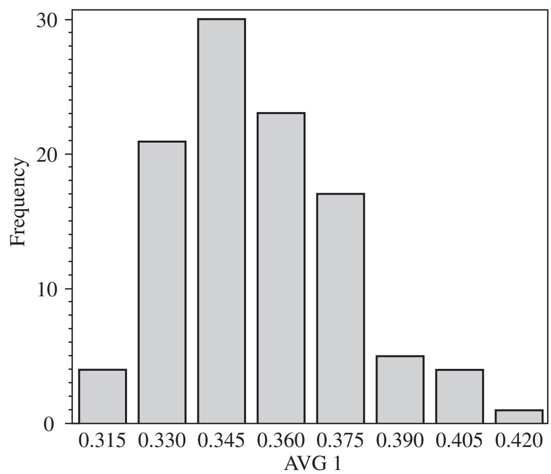
a What percentage of the output of this operation will fail to meet specifications?

b What should be the mean diameter of the shafts produced to minimize the fraction that fail to meet specifications?

5.88 The batting averages of the National League batting champions for the years 1901 through 2005 are graphed on the histogram that follows. This graph looks somewhat normal in shape, but it has a slight skew toward the high values. The mean is 0.353 and the standard deviation is 0.021 for these data.

a Barry Bonds batted 0.362 in 2004, and Tony Gwynn batted 0.394 in 1994. How would you compare these performances?

b Is there a good chance that anyone in the National League will hit over 0.400 in any 1 year?



5.89 Wires manufactured for a certain computer system are specified to have a resistance of between 0.12 and 0.14 ohm. The actual measured resistances of the wires produced by Company A have a normal probability distribution with a mean of 0.13 ohm and a standard deviation of 0.005 ohm.

a What is the probability that a randomly selected wire from Company A's production lot will meet the specifications?

b If four such wires are used in a single system and all are selected from Company A, what is the probability that all four will meet the specifications?

5.90 Refer to the setting in Exercise 5.89. Company B also produces the wires used in the computer system. The actual measured resistances of the wires produced by Company B have a normal probability distribution with a mean of 0.136 ohm and a standard deviation of 0.003 ohm. The computer firm orders 70% of the wires used in its systems from Company A and 30% from Company B.

a What is the probability that a randomly selected wire from Company B's production lot will meet the specifications?

b If four such wires are used in a single system and all are selected from Company B, what is the probability that all four will meet the specifications?

c Suppose that all four wires placed in any one computer system are all from the same company and that the computer system will fail testing if any one of the four wires does not meet specifications. A computer system is selected at random and tested. It meets specifications. What is the probability that Company A's wires were used in it?

5.91 At a temperature of 25°C, the resistances of a type of thermistor are normally distributed with a mean of 10,000 ohms and a standard deviation of 4000 ohms. The thermistors are to be sorted, and those having resistances between 8000 and 15,000 ohms are to be shipped to a vendor. What fraction of these thermistors will actually be shipped?

5.92 A vehicle driver gauges the relative speed of the next vehicle ahead by observing the speed with which the image of the width of that vehicle varies. This speed is proportional to X, the speed of variation of the angle at which the eye subtends this width. According to P. Ferrani and others (1984, 50, 51), a study of many drivers revealed X to be normally distributed with a mean of 0 and a standard deviation of 10 (10^{-4} radian per second).

a What fraction of these measurements is more than 5 units away from 0?

b What fraction is more than 10 units away from 0?

5.93 A type of capacitor has resistances that vary according to a normal distribution, with a mean of 800 megohms and a standard deviation of 200 megohms (see Nelson 1967, 261–268, for a more thorough discussion). A certain application specifies capacitors with resistances of between 900 and 1000 megohms.

a What proportion of these capacitors will meet this specification?

b If two capacitors are randomly chosen from a lot of capacitors of this type, what is the probability that both will satisfy the specifications?

5.94 When fishing off the shores of Florida, a spotted sea trout must be between 14 and 24 inches long before it can be kept; otherwise, it must be returned to the waters. In a region of the Gulf of Mexico, the lengths of spotted sea trout that are caught are normally distributed with a mean of 22 inches and a standard deviation of 4 inches. Assume that each spotted sea trout is equally likely to be caught by a fisherman.

a What is the probability that a fisherman catches a spotted sea trout within the legal limits?

b The fisherman caught a large spotted sea trout. He wants to know whether the length is in the top 5% of the lengths of spotted sea trout in that region. Find the length x for which only 5% of the spotted sea trout in this region will be larger than x.

c What is the probability that the fisherman will catch three trout outside the legal limits before catching his first legal spotted sea trout (between 14 and 24 inches)?

5.95 Sick leave time used by employees of a firm in the course of 1 month has approximately a normal distribution with a mean of 180 hours and a variance of 350 hours.

a Find the probability that the total sick leave for next month will be less than 150 hours.

b In planning schedules for next month, how much time should be budgeted for sick leave if that amount is to be exceeded with a probability of only 0.10?

5.96 Men's shirt sizes are determined by their neck sizes. Suppose that men's neck sizes are approximately normally distributed with a mean of 16.2 inches and standard deviation of 0.9 inch. A retailer sells men's shirts in sizes S, M, L, and XL, where the shirt sizes are defined in the following table.

Shirt Size	Neck Size
S	$14 \leq$ neck size < 15
M	$15 \leq$ neck size < 16
L	$16 \leq$ neck size < 17
XL	$17 \leq$ neck size < 18

a Because the retailer stocks only the sizes listed in the table, what proportion of customers will find that the retailer does not carry any shirts in their sizes?

b Calculate the proportion of men whose shirt size is M.

c Of 10 randomly selected customers, what is the probability that exactly 3 request size M?

5.97 A machine for filling cereal boxes has a standard deviation of 1 ounce in fill per box. Assume that the ounces of fill per box are normally distributed.

a What setting of the mean ounces of fill per box will allow 14-ounce boxes to overflow only 1% of the time?

b The company advertises that the box holds 12.8 ounces of cereal. Using the mean determined in part (a), what is the probability that a randomly selected box will have less than 12.8 ounces of cereal?

5.98 Referring to Exercise 5.97, suppose that the standard deviation σ is not known but can be fixed at certain levels by carefully adjusting the machine. What is the largest value of σ that will allow the actual value dispensed to fall within 1 ounce of the mean with a probability of at least 0.95?

5.99 The gestation period for a particular breed of cow is normally distributed with a mean of 280 days and a standard deviation of 3 days. A rancher who runs a cow-calf business keeps one bull in the field with the cows at all times. He sells an old bull and replaces it with a young one. A cow has a calf 288 days later. The farmer asks you which bull sired the calf. Use a probabilistic argument to respond to him.

5.100 The manager of a cultured pearl farm has received a special order for five pearls between 7 millimeters and 9 millimeters in diameter. From past experience, the manager knows that the pearls found in his oyster bed have diameters that are normally distributed with a mean of 8 millimeters and a standard deviation of 2.2 millimeters. Assume that every oyster contains one pearl.

a Determine the probability of finding a pearl of the appropriate size in an oyster selected at random.

b What is the probability that the special order can be filled after randomly selecting and opening the first five oysters?

c What is the probability that the manager will have to select more than 10 oysters at random to fill the order?

d Before the manager could give a price for the pearls, he wanted to determine his expected cost. When he begins to fill an order, he has fixed costs of \$550. He has determined that the cost of selecting an oyster, examining it for a pearl of the appropriate size, and replacing the oyster is \$60. What is the expected cost that the manager will incur in filling the order?

5.101 In the United States the heights of 20-year-old females are distributed according to the cumulative percentages listed in the accompanying table.

Cumulative Percentage	Height Inches
3	59.49
5	60.10
10	61.03
25	62.59
50	64.31
75	66.02
90	67.56
95	68.49
97	69.08

Source: U.S. Centers for Disease Control.

a Produce a Q-Q plot for these data; discuss their goodness of fit to a normal distribution.

b Approximate the mean and the standard deviation of these heights based on the Q-Q plot.

5.102 The cumulative proportions of U.S. residents of various ages are shown in the accompanying table for the years 2000 and 2050 (projected).

Cumulative Age	2000	2050
Under 5	6.9	6.3
0–9	14.2	12.5
0–19	28.6	25.3
0–29	41.9	37.6
0–39	57.4	49.6
0–49	72.7	61.2
0–59	83.5	72.0
0–69	90.8	81.8
0–79	96.7	90.1
0–120	100.0	100.0

Source: U.S. Census Bureau.

a Construct a Q-Q plot for each year. Use these plots as a basis for discussing key differences between the two age distributions.

b Each of the Q-Q plots should show some departures from normality. Explain the nature of their departures.

5.103 We have made extensive use of the fact that for a standard normal random variable z, $P(z > z_0) = P(z < -z_0)$. Prove that this is valid.

5.104 Let $f(x)$ be the probability density function of a normal random variable with a mean of μ and a standard deviation of σ. Show that the points of inflection of $f(x)$ lie at $(\mu - \sigma)$ and $(\mu + \sigma)$. That is, show that $f''(x) = 0$ when $x = \mu - \sigma$ or $x = \mu + \sigma$.

5.7 The Beta Distribution

5.7.1 Probability Density Function

Except for the uniform distribution of Section 5.3, the continuous distributions discussed thus far have had density functions that are positive over an infinite interval. It is useful to have another class of distributions to model phenomena constrained to a finite interval of possible values. One such class, the beta distributions, is valuable for modeling the probabilistic behavior of certain random variables (such as proportions) constrained to fall in the interval (0, 1). (Actually, any finite interval can be transformed into (0, 1).)

The beta distribution has the functional form

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

where the parameters α and β are positive constants. The constant term in $f(x)$ is necessary so that

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

That is,

$$\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

for positive α and β. This is a handy result to keep in mind and is left as an exercise.

The beta distribution is a rich distribution in that it can be used to model a wide range of distributional shapes. The graphs of some common beta density functions are shown in Figure 5.27. The uniform is a special case of the beta distribution with $\alpha = 1$ and $\beta = 1$. As with the gamma and normal distribution, finding probabilities is computationally challenging for many values of α and β. The *Continuous Distributions* applet, calculators, and computer software packages are valuable aids in finding these probabilities.

One important measurement in the process of sintering copper relates to the proportion of the volume that is solid rather than made up of voids. (The proportion due to voids is sometimes called the porosity of the solid.) Figure 5.28 shows a relative

FIGURE **5.27**
Beta density function.

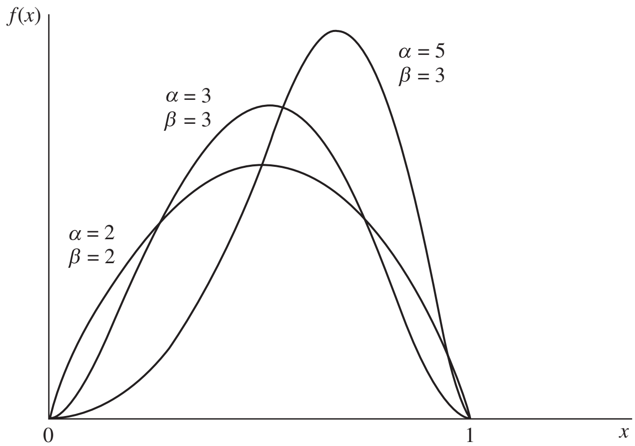
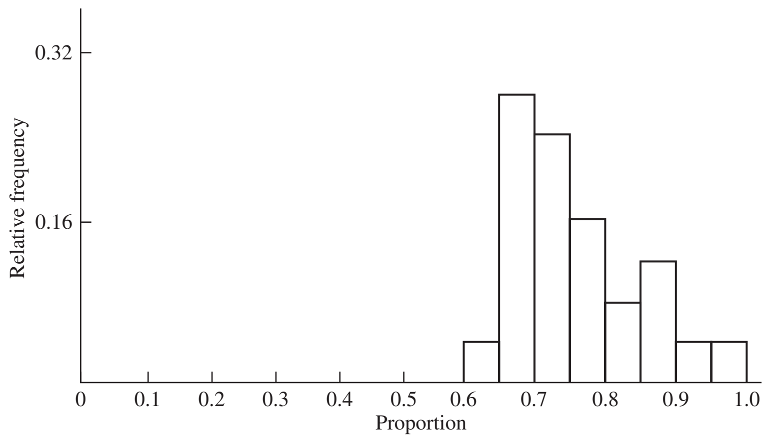


FIGURE **5.28**
Solid mass in sintered linde copper.



Source: Department of Materials Science, University of Florida.

frequency histogram of proportions of solid copper in samples drawn from a sintering process. This distribution could be modeled with a beta distribution that has a large α and a small β.

5.7.2 Mean and Variance

The expected value of a beta random variable can easily be found because

$$\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x \\
&= \int_0^1 x \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}\, \mathrm{d}x \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{\alpha}(1-x)^{\beta-1}\, \mathrm{d}x \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \times \frac{\Gamma(\alpha+1)\Gamma(\beta)}{\Gamma(\alpha+\beta+1)} \\
&= \frac{\alpha}{\alpha+\beta}
\end{aligned}$$

[Recall that $\Gamma(n+1) = n\Gamma(n)$.] Similar manipulations reveal that

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

We illustrate the use of this density function in an example.

EXAMPLE 5.19 A gasoline wholesale distributor uses bulk storage tanks to hold a fixed supply. The tanks are filled every Monday. Of interest to the wholesaler is the proportion of the supply sold during the week. Over many weeks, this proportion has been observed to match fairly well a beta distribution with $\alpha = 4$ and $\beta = 2$.

1 Find the expected value of this proportion.

2 Is it highly likely that the wholesaler will sell at least 90% of the stock in a given week?

Solution **1** By the results given earlier with X denoting the proportion of the total supply sold in a given week,

$$E(X) = \frac{\alpha}{\alpha+\beta} = \frac{4}{6} = \frac{2}{3}.$$

2 Now we are interested in

$$P(X > 0.9) = \int_{0.9}^{1} \frac{\Gamma(4+2)}{\Gamma(4)\Gamma(2)} x^3(1-x)\,\mathrm{d}x$$

$$= 20\int_{0.9}^{1} \left(x^3 - x^4\right)\,\mathrm{d}x$$

$$= 20(0.004) = 0.08.$$

Therefore, it is not very likely that 90% of the stock will be sold in a given week. ▪

5.7.3 History and Applications

Reverend Thomas Bayes first encountered the beta distribution as he considered allowing the probability of success p to vary. The difficulty he encountered was in the integration process. If α and β are small, the integration is simple, but the computations quickly become challenging as the parameters get large. In 1781, Laplace determined how to derive an arbitrarily close series approximation to the integral. When Karl Pearson categorized density functions into types, his Type II curves were the symmetric beta density, and the Type III curves were the asymmetric beta density. Although the beta distribution was generally referred to as a Pearson curve of Type II or III in the English literature of the early 1900s, the texts of the 1940s adopted the name *beta distribution*, which is used here. The beta distribution is frequently used to model any data that occur over an interval because, as we noted earlier, any interval can be transformed to a (0, 1) interval. Applications include studying the time to complete a task or the proportions in a mixture.

The basic properties of the beta distribution are summarized next.

The Beta Distribution

$$f(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

$$E(X) = \frac{\alpha}{\alpha+\beta} \quad \text{and} \quad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Exercises

5.105 Suppose that X has a probability density function given by

$$f(x) = \begin{cases} kx^4(1-x)^2, & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the value of k that makes this a probability density function.

b Find $E(X)$ and $V(X)$.

5.106 The weekly repair cost, X, for a certain machine has a probability density function given by

$$f(x) = \begin{cases} cx(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

with measurements in \$100s.

a By simply looking at the probability density function, state the distribution of X, including the parameters.

b Find the value of c that makes this function a valid probability density function.

c Find and sketch the distribution function of X.

d What is the probability that repair costs will exceed \$75 during a week?

5.107 In toxicological experiments on laboratory animals, pregnant females are often exposed to a potential hazard, but the responses are recorded on each individual fetus. Variation is expected between litters as well as within a litter. For a particular hazard, the proportion of the fetuses in a randomly selected litter responding is well modeled by the following distribution:

$$f(x) = \begin{cases} 12x^2(1-x), & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$$

a What is the mean litter proportion of fetuses responding to this hazard?

b Find the probability that for a randomly selected litter the proportion of fetuses responding to the hazard will be no more than 0.2.

5.108 The proportion of impurities per batch in a certain type of industrial chemical is a random variable X that has the probability density function

$$f(x) = \begin{cases} 20x^3(1-x), & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Suppose that a batch with more than 30% impurities cannot be sold. What is the probability that a randomly selected batch cannot be sold for this reason?

b Suppose that the dollar value of each batch is given by $V = 10 - 0.75X$. Find the expected value and variance of V.

5.109 During an 8-hour shift, the proportion of time X that a sheet-metal stamping machine is down for maintenance or repairs has a beta distribution with $\alpha = 1$ and $\beta = 2$; that is,

$$f(x) = \begin{cases} 2(1-x), & \text{for } 0 < x < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

The cost (in \$100s) of this downtime in lost production and repair expenses is given by

$$C = 10 + 20X + 4X^2.$$

a Find the mean and the variance of C.

b Find an interval within which C has a probability of at least 0.75 of lying.

5.110 In trees, foliage distribution as a function of height is termed the foliage surface area density (FSAD), or needle surface area density (NSAD) for conifers, and is the foliage surface area per unit volume of space at a given height above the ground. Dividing the distance from the ground by the height of the tree rescales the height of the tree to the interval (0, 1). The NSAD has been found to be modeled well using the beta distribution (Massman 1982). For a certain old-growth Douglas fir, the mean and the variance of its NSAD are 0.75 and 0.0375, respectively.

a Find α and β.

b What proportion of the NSAD lies in the upper half of the tree?

c What proportion of the NSAD lies in the middle half of the tree?

5.111 To study the dispersal of pollutants emerging from a power plant, researchers measure the prevailing wind direction for a large number of days. The direction is measured on a scale of 0° to 360°, but by dividing each daily direction by 360 one can rescale the measurements to the interval (0, 1). These rescaled measurements X prove to follow a beta distribution with $\alpha = 4$ and $\beta = 2$. Find $E(X)$. To what angle does this mean correspond?

5.112 Errors in measuring the arrival time (in microseconds) of a wavefront from an acoustic source can sometimes be modeled by a beta distribution (Perruzzi and Hilliard 1984). Suppose that these errors have a beta distribution with $\alpha = 1$ and $\beta = 2$.

a Find the probability that a particular measurement error will be less than 0.5 microsecond.

b Find the mean and the standard deviation of these measurements.

5.113 Exposure to an airborne contaminant is defined as the time-weighted average breathing-zone concentration over an interval T (Flynn 2004). The concentration is the proportion of the volume of air breathed that has the contaminant. It has been found that this exposure is well modeled by the beta distribution. For benzene in a particular region, it was found that $\alpha = 1$ and $\beta = 344$.

a Find the mean and variance of the exposure to benzene in this region.

b What is the probability that the exposure by a randomly selected individual is less than 0.01?

5.114 The broken stick model is sometimes used to describe the allocation of environmental resources among species. For three species, assume that one species is assigned the left end of the stick; one species is associated with the middle of the stick; and the resources for the other species are represented by the right end of the stick. Two points are randomly and independently selected along a stick of unit length, and the stick is broken at those points. The available resources for each species are proportional to the length of the stick associated with that species. The proportion of resources available to the species getting the left end of the stick is then modeled using the beta distribution with $\alpha = 1$ and $\beta = 2$.

a Find the mean and variance of the proportion of resources that the species getting the left end of the stick receives.

b Find the probability that the proportion of resources received by the species getting the left end of the stick is greater than 40%.

c Find the probability that the resources received by the species getting the left end of the stick are less than 10%.

5.115 The proper blending of fine and coarse powders in copper sintering is essential for uniformity in the finished product. One way to check the blending is to select many small samples of the blended powders and to measure the weight fractions of the fine particles. These measurements should be relatively constant if good blending has been achieved.

a Suppose that the weight fractions have a beta distribution with $\alpha = \beta = 3$. Find the mean and the variance of these fractions.

b Repeat part (a) for $\alpha = \beta = 2$.

c Repeat part (a) for $\alpha = \beta = 1$.

d Which case—part (a), (b), or (c)—would exemplify the best blending? Justify your answer.

5.116 The proportion of pure iron in certain ore samples has a beta distribution with $\alpha = 3$ and $\beta = 1$.

a Find the probability that one of these samples will have more than 50% pure iron.

b Find the probability that two out of three samples will have less than 30% pure iron.

5.117 If X has a beta distribution with parameters α and β, show that

$$V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}.$$

5.118 If the random variable X has the binomial distribution with parameters n and p, prove that

$$F_X(k) = P(X \le k) = n\binom{n-1}{k}\int_0^q t^{n-k-1}(1-t)^k\,\mathrm{d}t = n\binom{n-1}{k}\mathrm{B}_p(n-k, k+1).$$

Hint:

a Show that $(n-k)\binom{n}{k} = n\binom{n-1}{k}$ and $k\binom{n}{k} = n\binom{n-1}{k-1}$.

b Use integration by parts to show that

$$\int_0^q t^{n-k-1}(1-t)^k\,\mathrm{d}t = \frac{1}{n-k}p^k q^{n-k} + \frac{k}{n-k}\int_0^q t^{n-k}(1-t)^{k-1}\,\mathrm{d}t.$$

5.8 The Weibull Distribution

5.8.1 Probability Density Function

We have observed that the gamma distribution often can serve as a probabilistic model for life lengths of components or systems. However, the failure rate function for the gamma distribution has an upper bound that limits its applicability to real systems. For this and other reasons, other distributions often provide better models for life-length data. One such distribution is the *Weibull*, which is explored in this section.

A Weibull density function has the form

$$f(x) = \begin{cases} \frac{\gamma}{\theta}x^{\gamma-1}e^{-x^\gamma/\theta}, & \text{for } x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

for positive parameters θ and γ. For $\gamma = 1$, this becomes an exponential density. For $\gamma > 1$, the functions look something like the gamma functions of Section 5.5, but they have somewhat different mathematical properties. We can integrate directly to see that

$$F(x) = \begin{cases} 0, & x < 0 \\ P(X \le x) = \int_0^x \frac{\gamma}{\theta}t^{\gamma-1}e^{-t^\gamma/\theta}\,\mathrm{d}t = -\,e^{-t^\gamma/\theta}\Big|_0^x = 1 - e^{-x^\gamma/\theta}, & x \ge 0. \end{cases}$$

Because the distribution function has a closed form (can be written without an integral), computing Weibull probabilities is much easier than it is for the gamma distribution. However, the *Continuous Distributions* applet, calculators, and computer software may also be used.

A convenient way to look at properties of the Weibull density is to use the transformation $Y = X^\gamma$. Then,

$$\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P(X^\gamma \le y) \\
&= P(X \le y^{1/\gamma}) \\
&= F_X(y^{1/\gamma}) \\
&= 1 - e^{-(y^{1/\gamma})^\gamma/\theta} \\
&= 1 - e^{-y/\theta}, \quad y > 0.
\end{aligned}$$

Hence,

$$f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \frac{1}{\theta} e^{-y/\theta}, \quad y > 0;$$

that is, Y has the familiar exponential density.

5.8.2 Mean and Variance

If we want to find $E(X)$ for a variable X that has the Weibull distribution, we take advantage of the transformation $Y = X^\gamma$ or $X = Y^{1/\gamma}$. Then

$$\begin{aligned}
E(X) &= E(Y^{1/\gamma}) \\
&= \int_0^\infty y^{1/\gamma} \frac{1}{\theta} e^{-y/\theta}\, \mathrm{d}y \\
&= \frac{1}{\theta} \int_0^\infty y^{1/\gamma} e^{-y/\theta}\, \mathrm{d}y \\
&= \frac{1}{\theta} \Gamma\left(1 + \frac{1}{\gamma}\right) \theta^{(1+1/\gamma)} \\
&= \theta^{1/\gamma} \Gamma\left(1 + \frac{1}{\gamma}\right).
\end{aligned}$$

This result follows from recognizing the integral to be of the gamma type.

If we let $\gamma = 2$ in the Weibull density, we see that $Y = X^2$ has an exponential distribution. To reverse the idea just outlined, if we start with an exponentially distributed random variable Y, then the square root of Y will have a Weibull distribution with $\gamma = 2$. We can illustrate this empirically by taking the square roots of the data from an exponential distribution given in Table 5.1. These square roots are given in

Table 5.6, and a relative frequency histogram for these data is given in Figure 5.29. Notice that the exponential form has now disappeared and that the curve given by the Weibull density with $\gamma = 2$ and $\theta = 2$ (shown in Figure 5.30) is a much more plausible model for these observations.

We illustrate the use of the Weibull distribution in the following example.

TABLE **5.6**
Square Roots of the Battery Life Lengths of Table 5.1.

0.637	0.828	2.186	1.313	2.868
1.531	1.184	1.223	0.542	1.459
0.733	0.484	2.006	1.823	1.700
2.256	1.207	1.032	0.880	0.872
2.364	1.305	1.623	1.360	0.431
1.601	0.719	1.802	1.526	1.032
0.152	0.715	1.668	2.535	0.914
1.826	0.474	1.230	1.793	0.617
1.868	1.525	0.477	2.746	0.984
1.126	1.709	1.274	0.578	2.119

FIGURE **5.29**
Relative frequency histogram for the data of Table 5.6.

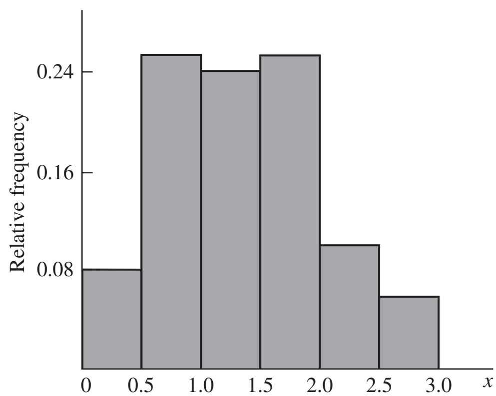
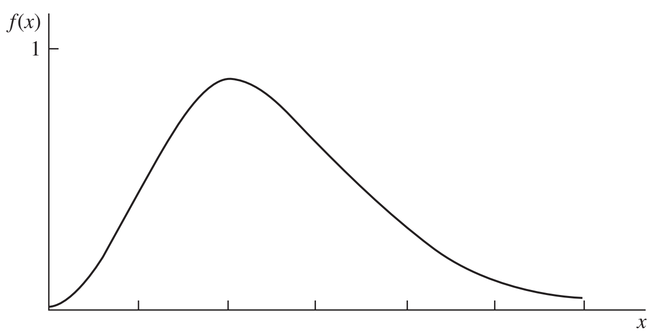


FIGURE **5.30**
Weibull density function, $\gamma = 2, \theta = 2$.



EXAMPLE 5.20 The lengths of time in service during which a certain type of thermistor produces resistances within its specifications have been observed to follow a Weibull distribution with $\theta = 50$ and $\gamma = 2$ (measurements in thousands of hours).

1 Find the probability that one of these thermistors, which is to be installed in a system today, will function properly for more than 10,000 hours.

2 Find the expected life length for thermistors of this type.

Solution The Weibull distribution has a closed-form expression for $F(x)$. Thus, if X represents the life length of the thermistor in question, then

$$\begin{aligned}
P(X > 10) &= 1 - F(10) \\
&= 1 - \left[1 - e^{-(10)^2/50}\right] \\
&= e^{-(10)^2/50} \\
&= e^{-2} \\
&= 0.14
\end{aligned}$$

because $\theta = 50$ and $\gamma = 2$. We know that

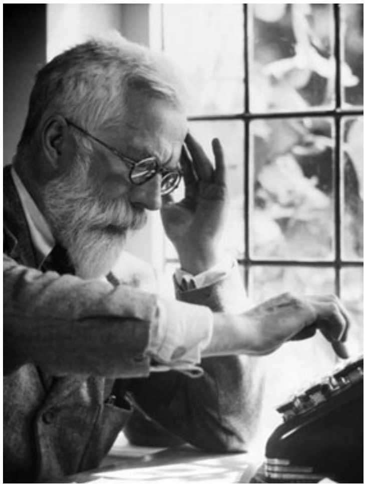
$$\begin{aligned}
E(X) &= \theta^{1/\gamma}\Gamma\left(1 + \frac{1}{\gamma}\right) \\
&= (50)^{1/2}\Gamma\left(\frac{3}{2}\right) \\
&= (50)^{1/2}\left(\frac{1}{2}\right)\Gamma\left(\frac{1}{2}\right) \\
&= (50)^{1/2}\left(\frac{1}{2}\right)\sqrt{\pi} \\
&= 6.27.
\end{aligned}$$

Thus, the average service time for these thermistors is 6270 hours. ▪

5.8.3 History and Applications to Real Data

R. A. Fisher and L. H. C. Tippett derived the Weibull distribution as a limit of an extreme value distribution in 1928 (Figure 5.31). However, in the 1930s, the great utility of the distribution for describing data was explored by the Swedish physicist Waloddi Weibull, first by modeling the strengths of materials and later in broader applications. Because he popularized it, the distribution became known as the Weibull distribution.

FIGURE **5.31**
Ronald A. Fisher (1890–1962).



A. Barrington Brown/Photo Researchers, Inc.

The Weibull distribution is versatile enough to be a good model for many distributions of data that are mound-shaped but skewed. Another advantage of the Weibull distribution over, say, the gamma is that a number of relatively straightforward techniques exist for actually fitting the model to data. One such technique is illustrated here. For the Weibull distribution,

$$P(X > x) = 1 - F(x) = e^{-x^{\gamma}/\theta}, \quad x > 0.$$

Therefore, using ln to denote natural logarithm, we have

$$\ln\left[\frac{1}{1 - F(x)}\right] = -\frac{x^{\gamma}}{\theta}$$

and

$$\ln\left(-\ln\left[\frac{1}{1 - F(x)}\right]\right) = \gamma\ln(x) - \ln(\theta).$$

For simplicity, let us call the double ln expression on the left $LF(x)$. Plotting $LF(x)$ as a function of $\ln(x)$ produces a straight line with slope γ and intercept $-\ln(\theta)$.

Now let us see how this works with real data. Subscriptions are an important asset to newspapers. The length of time people maintain their subscriptions is one consideration in valuing the newspaper both from the view of investors and advertisers. A newspaper owner wanted to learn more about the people who had failed to continue their subscriptions. Ten people from those who had discontinued their subscriptions were randomly selected from the database. In ascending order, the lengths of time each subscribed are as follows:

$$22,\ 62,\ 133,\ 205,\ 222,\ 315,\ 347,\ 403,\ 637,\ 763$$

Checking for outliers (by using a boxplot, for example) reveals that no observation is an extreme outlier so we use all of the data. Proceeding then, as in the normal case,

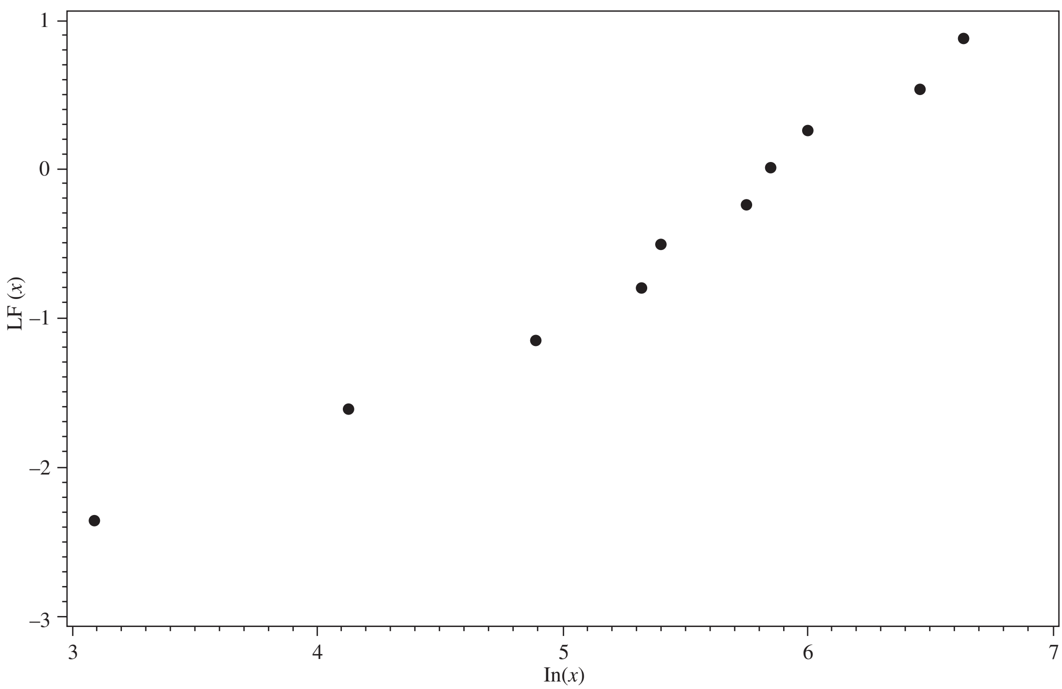
TABLE 5.7
Analysis of Particulate Matter Data.

I	*x*	ln(*x*)	*i*/(*n* + 1)	Approximate *LF*(*x*)
1	22	3.09104	0.09091	−2.35062
2	62	4.12713	0.18182	−1.60609
3	133	4.89035	0.27273	−1.14428
4	205	5.32301	0.35354	−0.79411
5	222	5.40268	0.45455	−0.50065
6	315	5.75257	0.54545	−0.23768
7	347	5.84932	0.63636	0.01153
8	403	5.99894	0.72727	0.26181
9	637	6.45677	0.81818	0.53342
10	763	6.63726	0.90909	0.87459

to produce sample percentiles as approximations to $F(x)$, we obtain the data listed in Table 5.7.

Here, $i/(n+1)$ is used to approximate $F(x_i)$ in $LF(x)$. Figure 5.32 shows the plot of $LF(x)$ versus $\ln(x)$. The points lie rather close to a straight line that has a slope of about 0.9 and a y-intercept of about -5.4. (This line can be approximated by drawing a straight line with a straightedge through the middle of the scatterplot.) Thus, the newspaper subscription times appear to be adequately modeled by a Weibull distribution, with $\gamma \approx 3.2$ and $\ln(\theta) \approx 5.4$ or $\theta \approx 221$ days. Note: This example only

FIGURE 5.32 Plot of $LF(x)$ versus $\ln(x)$.



provides insight into those who choose not to renew. It does not consider those who have continuing subscriptions.

Notice that fitting an exponential distribution to data could follow a similar procedure, with $\gamma = 1$. In this case,

$$\ln\left[\frac{1}{1 - F(x)}\right] = x/\theta.$$

So a plot of an estimate of the left side against x should reveal a straight line through the origin with a slope of $1/\theta$.

The Weibull Distribution

$$f(x) = \begin{cases} \frac{\gamma}{\theta} x^{\gamma-1} e^{-x^{\gamma}/\theta}, & \text{for } x \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

$$E(X) = \theta^{1/\gamma}\Gamma\left(1 + \frac{1}{\gamma}\right) \text{ and } V(X) = \theta^{2/\gamma}\left\{\Gamma\left(1 + \frac{2}{\gamma}\right) - \left[\Gamma\left(1 + \frac{1}{\gamma}\right)\right]^2\right\}$$

Exercises

5.119 Fatigue life (in hundreds of hours) for a certain type of bearing has approximately a Weibull distribution with $\gamma = 2$ and $\theta = 4$.

a Find the probability that a randomly selected bearing of this type will fail in less than 200 hours.

b Find the expected value of the fatigue life for these bearings.

5.120 The yearly maximum flood levels (in millions of cubic feet per second) for a certain United States river have a Weibull distribution with $\gamma = 1.5$ and $\theta = 0.6$ (see Cohen, Whitten, and Ding 1984). Find the probabilities that the maximum flood level for next year will have the following characteristics.

a It will exceed 0.6. **b** It will be less than 0.8.

5.121 The times necessary to achieve proper blending of copper powders before sintering were found to have a Weibull distribution with $\gamma = 1.1$ and $\theta = 2$ (measurements in minutes). Find the probability that proper blending in a particular case will take less than 2 minutes.

5.122 The wind speed is collected at a Hong Kong weather station at 1-minute intervals. From these data, the 60 wind speeds collected during an hour are averaged to obtain the average hourly wind speed, measured in meters/second (m/s). So 24 average hourly wind speeds are observed for each day. The distribution of the average hourly wind speeds for a year is well modeled by the Weibull distribution (Lun and Lam 2000). The average hourly wind speeds for 1997 were fit well by a Weibull distribution with $\gamma = 1.54$ and $\theta = 5.21$.

a For a randomly selected hour, find the probability that the wind speed exceeds 8 m/s.

b Find the interval containing 90% of the hourly wind speeds so that 5% of the wind speeds are below the lower limit and 5% are above the upper limit.

5.123 The ultimate tensile strength of steel wire used to wrap concrete pipe was found to have a Weibull distribution with $\gamma = 1.2$ and $\theta = 270$ (measurements in thousands of pounds). Pressure in the pipe at a certain point may require an ultimate tensile strength of at least 300,000 pounds. What is the probability that a randomly selected wire will possess this strength?

5.124 The Weibull distribution has been used to model survivorship, which is the probability of surviving to time t (Pinder, Wiener, and Smith 1978). If the random variable T denotes the age at which an organism dies, the probability of survival until time t is $P(T > t)$. For a certain population of Dall sheep, survival (in years) was found to be well modeled using a Weibull distribution with $\theta = 8.5$ and $\gamma = 2$.

a Find the mean and variance of the time that a randomly selected Dall sheep survives.

b Find the probability that a randomly selected Dall sheep survives at least a year.

c Find the age to which 50% of the Dall sheep survive.

5.125 The yield strengths of certain steel beams have a Weibull distribution with $\gamma = 2$ and $\theta = 3600$ (measurement in thousands of pounds per square inch [psi]). Two such beams are used in a construction project, which calls for yield strengths in excess of 70,000 psi. Find the probability that both beams will meet the specifications for the project.

5.126 Resistors used in an aircraft guidance system have life lengths that follow a Weibull distribution with $\gamma = 2$ and $\theta = 10$ (measurements in thousands of hours.)

a Find the probability that a randomly selected resistor of this type has a life length that exceeds 5000 hours.

b If three resistors of this type operate independently, find the probability that exactly one of the three will burn out prior to 5000 hours of use.

c Find the mean and the variance of the life length of such a resistor.

5.127 Failures in the bleed systems of jet engines were causing some concern at an airbase. It was decided to model the failure-time distribution for these systems so that future failures could be anticipated better. The following randomly identified failure times (in operating hours since installation) were observed (Abernethy et al. 1983):

1198, 884, 1251, 1249, 708, 1082, 884, 1105, 828, 1013

Does a Weibull distribution appear to be a good model for these data? If so, what values should be used for γ and θ?

5.128 The velocities of gas particles can be modeled by the Maxwell distribution with the probability density function given by

$$f(v) = \begin{cases} 4\pi \left(\frac{m}{2\pi KT}\right)^{3/2} v^2 e^{-v^2/2KT}, & \text{for } v \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

where m is the mass of the particle, K is Boltzmann's constant, and T is the absolute temperature.

a Find the mean velocity of these particles.

b The kinetic energy of a particle is given by $(1/2)mV^2$. Find the mean kinetic energy of these particles.

5.9 Reliability

One important measure of the quality of products is their reliability, or the probability of their working for a specified period. We want products—whether they be cars, television sets, or shoes—that do not break down or wear out for some definite period, and we want to know how long this time period can be expected to last. The study of

reliability is a probabilistic exercise because data or models on component lifetimes are used to predict future behavior: how long a process will operate before it fails.

In reliability studies, the underlying random variable of interest, X, is usually lifetime.

DEFINITION **5.5**

If a component has lifetime X with distribution function F, then the **reliability** of the component is

$$R(t) = P(X > t) = 1 - F(t). \quad \blacksquare$$

It follows that, for exponentially distributed lifetimes,

$$R(t) = e^{-t/\theta}, \quad t \geq 0$$

and for lifetimes following a Weibull distribution,

$$R(t) = e^{-t^{\gamma}/\theta}, \quad t \geq 0.$$

Reliability functions for gamma and normal distributions do not exist in closed form. In fact, the normal model is not often used to model life-length data because length of life tends to exhibit positively skewed behavior.

5.9.1 Hazard Rate Function

Besides probability density and distribution functions and the reliability function, another function is useful in work with life-length data. Suppose that X denotes the life length of a component with density function $f(x)$ and distribution function $F(x)$. The hazard rate function $r(t)$ is defined as

$$r(t) = \frac{f(t)}{1 - F(t)}, \quad t > 0, \quad F(t) < 1.$$

For an intuitive look at what $r(t)$ is measuring, suppose that $\mathrm{d}t$ denotes a very small interval around the point t. Then $f(t)\,\mathrm{d}t$ is approximately the probability that X will take on a value in $(t, t + \mathrm{d}t)$. Further, $1 - F(t) = P(X > t)$. Thus,

$$\begin{aligned} r(t)\,\mathrm{d}t &= \frac{f(t)\,\mathrm{d}t}{1 - F(t)} \\ &\approx P[X \in (t, t + \mathrm{d}t) \,|\, X > t]. \end{aligned}$$

In other words, $r(t)\,\mathrm{d}t$ represents the probability of failure during the time interval $(t, t + \mathrm{d}t)$ given that the component has survived up to time t.

For the exponential case,

$$r(t) = \frac{f(t)}{1 - F(t)} = \frac{\frac{1}{\theta}e^{-t/\theta}}{e^{-t/\theta}} = \frac{1}{\theta}.$$

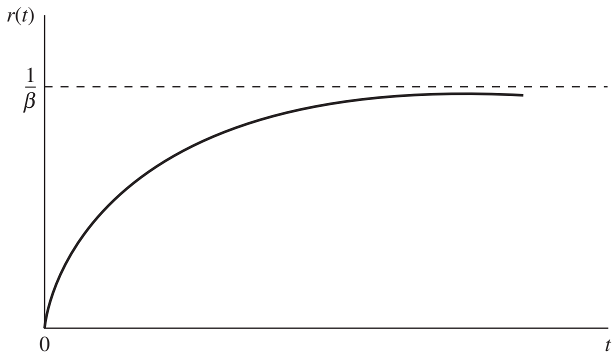


FIGURE **5.33**
Failure rate function for the gamma distribution ($\alpha > 1$).

Thus, X has a constant hazard rate. It is unlikely that many individual components have a constant hazard rate over time (most fail more frequently as they age), but it may be true of some systems that undergo regular preventive maintenance.

The hazard rate function $r(t)$ for the gamma case is not easily displayed, because $F(t)$ does not have a simple closed form. For $\alpha > 1$, however, this function will increase but always remain bounded above by $1/\beta$. A typical form is shown in Figure 5.33.

5.9.2 Series and Parallel Systems

A system is made up of a number of components, such as relays in an electrical system or check valves in a water system. The reliability of a system depends critically on how the components are networked into the system. A *series system* (Figure 5.34[a]) fails as soon as any one component fails. A *parallel system* (Figure 5.34[b]) fails only when all components have failed.

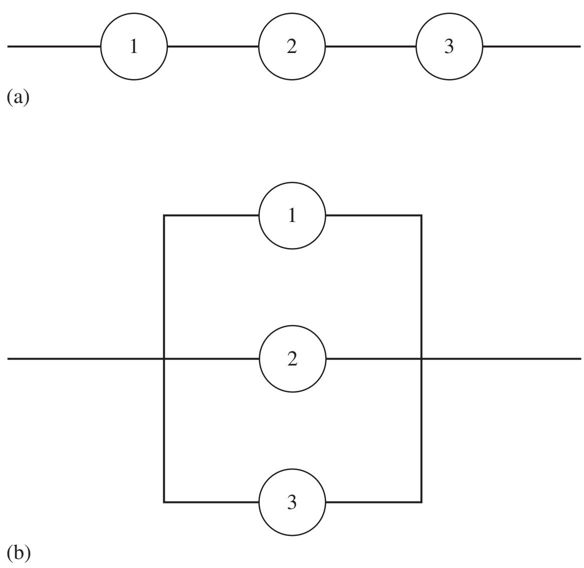


FIGURE **5.34**
Series and parallel systems.

Suppose that the system components in Figure 5.36 each have a reliability function of $R(t)$, and suppose that the components operate independently of one another. What are the system reliabilities $R_S(t)$? For the series system, life length will exceed t only if each component has a life length exceeding t. Thus,

$$R_S(t) = R(t)R(t)R(t) = [R(t)]^3 .$$

For the parallel system, life length will be less than t only if all components fail before t. Thus,

$$\begin{aligned} 1 - R_S(t) &= [1 - R(t)]\,[1 - R(t)]\,[1 - R(t)] \\ R_S(t) &= 1 - [1 - R(t)]^3 . \end{aligned}$$

Of course, most systems are combinations of components in series and components in parallel. The rules of probability must be used to evaluate system reliability in each special case, but breaking the system into series and parallel subsystems often simplifies this process.

5.9.3 Redundancy

What happens as we add components to a system? For a series of n independent components,

$$R_S(t) = [R(t)]^n$$

and because $R(t) \leq 1$, adding more components in series will just make things worse! For n components operating in parallel, however,

$$R_S(t) = 1 - [1 - R(t)]^n$$

which will increase with n. Thus, system reliability can be improved by adding backup components in parallel—a practice called redundancy.

How many components must we use to achieve a specified level of system reliability? If we have a fixed value for $R_S(t)$ in mind, and if we have independently operating parallel components, then

$$1 - R_S(t) = [1 - R(t)]^n .$$

Letting ln denote a natural logarithm, we can calculate

$$\ln\left[1 - R_S(t)\right] = n \ln\left[1 - R(t)\right]$$

and

$$n = \frac{\ln\left[1 - R_S(t)\right]}{\ln\left[1 - R(t)\right]} .$$

This is but a brief introduction to the interesting and important area of reliability. But the key to all reliability problems is a firm understanding of basic probability.

Exercises

5.129 For a component with an exponentially distributed lifetime, find the reliability up to time $t_1 + t_2$ given that the component has already lived past t_1. Why is the constant hazard rate referred to as the "memoryless" property?

5.130 For a series of n components operating independently—each with the same exponential life distribution—find an expression for $R_S(t)$. What is the mean lifetime of the system?

5.131 Let X be a Weibull random variable.

a Find the hazard function of X.

b Show that the hazard function of a Weibull random variable is increasing if $\gamma > 1$ and decreasing when $\gamma < 1$.

5.132 Refer to the results in Exercise 5.131. The lifetime X of a particular battery is well modeled by a Weibull distribution with $\gamma = 2$ and $\theta = 4$. Find the hazard function for X and interpret its meaning in the context of the problem.

5.133 A random variable X has the logistic distribution with parameters μ and β if

$$F_X(x) = \frac{1}{1 + e^{-(x-\mu)/\beta}}, \quad x \in \Re.$$

a Find the probability density function of the logistic distribution.

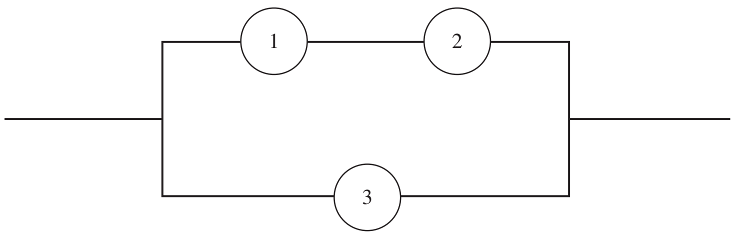
b Find the hazard function of the logistic distribution.

5.134 Show that for a continuous random variable X,

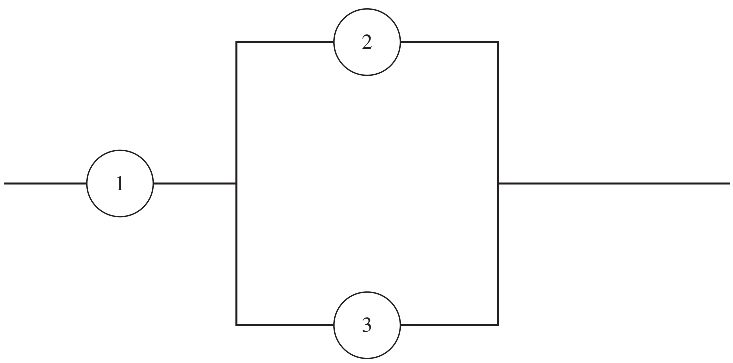
$$h_X(x) = -\frac{d}{dt} \log\left(1 - F_X(x)\right).$$

5.135 For independently operating components with identical life distributions, find the system reliability for each of the following.

(a)



(b)



Which system has the higher reliability?

5.136 Suppose that each relay in an electrical circuit has a reliability of 0.9 for a specified operating period of t hours. How could you configure a system of such relays to bring $R_S(t)$ up to 0.999?

5.10 Moment-Generating Functions for Continuous Random Variables

As in the case of discrete distributions, the moment-generating functions of continuous random variables help us find expected values and identify certain properties of probability distributions. We now undertake a short discussion of moment-generating functions for continuous random variables.

The moment-generating function of a continuous random variable X with a probability density function of $f(x)$ is given by

$$M(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f(x)\,\mathrm{d}x$$

when the integral exists.

For the exponential distribution, this becomes

$$\begin{aligned} M(t) &= \int_0^{\infty} e^{tx}\frac{1}{\theta}e^{-x/\theta}\,\mathrm{d}x \\ &= \frac{1}{\theta}\int_0^{\infty} e^{-x(1/\theta-1)}\,\mathrm{d}x \\ &= \frac{1}{\theta}\int_0^{\infty} e^{-x(1-\theta t)/\theta}\,\mathrm{d}x \\ &= \frac{1}{\theta}\Gamma(1)\left(\frac{\theta}{1-\theta t}\right) \\ &= (1-\theta t)^{-1}. \end{aligned}$$

We can now use $M(t)$ to find $E(X)$, because

$$M'(0) = E(X)$$

and for the exponential distribution we have

$$\begin{aligned} E(X) &= M'(0) \\ &= \left[-(1-\theta t)^{-2}(-\theta)\right]_{t=0} \\ &= \theta. \end{aligned}$$

Similarly, we could find $E(X^2)$ and then $V(X)$ by using the moment-generating function.

An argument analogous to the one used for the exponential distribution shows that for the gamma distribution,

$$M(t) = (1 - \beta t)^{-\alpha}.$$

From this we can see that, if X has a gamma distribution,

$$\begin{aligned}
E(X) &= M'(0) \\
&= \left[-\alpha(1 - \beta t)^{-\alpha-1}(-\beta)\right]_{t=0} \\
&= \left[\alpha\beta(1 - \beta t)^{-\alpha-1}\right]_{t=0} \\
&= \alpha\beta \\
E(X^2) &= M^{(2)}(0) \\
&= \left[\alpha\beta(-\alpha - 1)(1 - \beta t)^{-\alpha-2}(-\beta)\right]_{t=0} \\
&= \left[\alpha(\alpha + 1)\beta^2(1 - \beta t)^{-\alpha-1}\right]_{t=0} \\
&= \alpha(\alpha + 1)\beta^2 \\
E(X^3) &= M^{(3)}(0) \\
&= \left[\alpha(\alpha + 1)\beta^2(-\alpha - 2)(1 - \beta t)^{-\alpha-3}(-\beta)\right]_{t=0} \\
&= \left[\alpha(\alpha + 1)(\alpha + 2)\beta^3(1 - \beta t)^{-\alpha-3}\right]_{t=0} \\
&= \alpha(\alpha + 1)(\alpha + 2)\beta^3
\end{aligned}$$

and so on.

Moment-generating functions have two important properties:

1 If a random variable X has the moment-generating function $M_X(t)$, then $Y = aX + b$, for constants a and b, has the moment-generating function

$$M_Y(t) = e^{tb}M_X(at).$$

2 Moment-generating functions are unique; that is, two random variables that have the same moment-generating function have the same probability distributions as well.

Property 1 is easily shown (see Exercise 5.145). The proof of property 2 is beyond the scope of this textbook but we make use of it in identifying distributions, as demonstrated later.

EXAMPLE **5.21**

1 Find the moment-generating function of a normal random variable with a mean of μ and a variance of σ^2.

2 Use the properties of the moment-generating function to find the moment-generating function of a standard normal random variable.

Solution 1 Suppose X is normally distributed with a mean of μ and a variance of σ^2. One way to find the moment-generating function of X is to first find the moment-generating function of $X - \mu$. Now,

$$E\left(e^{t(X-\mu)}\right) = \int_{-\infty}^{\infty} e^{t(x-\mu)} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}\, \mathrm{d}x.$$

Letting $y = x - \mu$, we find that the integral becomes

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{ty - y^2/2\sigma^2}\, \mathrm{d}y = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}(y^2 - 2\sigma^2 ty)\right] \mathrm{d}y.$$

Completing the square in the exponent, we have

$$\begin{aligned} -\frac{1}{2\sigma^2}\left(y^2 - 2\sigma^2 ty\right) &= -\frac{1}{2\sigma^2}\left(y^2 - 2\sigma^2 ty + \sigma^4 t^2\right) + \frac{1}{2}t^2\sigma^2 \\ &= -\frac{1}{2\sigma^2}\left(y - \sigma^2 t\right)^2 + \frac{1}{2}t^2\sigma^2 \end{aligned}$$

so the integral becomes

$$e^{t^2\sigma^2/2} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2\sigma^2}\left(y - \sigma^2 t\right)^2\right] \mathrm{d}y = e^{t^2\sigma^2/2}$$

because the remaining integrand forms a normal probability density that integrates to unity. Thus, $Y = X - \mu$ has the moment-generating function given by

$$M_Y(t) = e^{t^2\sigma^2/2}.$$

Thus, by property 1 (stated earlier),

$$X = Y + \mu$$

has the moment-generating function given by

$$\begin{aligned} M_X(t) &= e^{\mu t} M_Y(t) \\ &= e^{\mu t} e^{t^2\sigma^2/2} \\ &= e^{\mu t + t^2\sigma^2/2}. \end{aligned}$$

2 Now we will find the distribution of the standardized normal random variable Z. We begin with the random variable X, which is normally distributed with a mean of μ and a variance of σ^2. Let

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma}X - \frac{\mu}{\sigma}.$$

Then, by property 1,

$$\begin{aligned} M_Z(t) &= e^{-\mu t/\sigma} e^{\mu t/\sigma + t^2\sigma^2/2\sigma^2} \\ &= e^{t^2/2}. \end{aligned}$$

The moment-generating function of Z has the form of a moment-generating function for a normal random variable with a mean of 0 and a variance of 1. Thus, by property 2, Z must have that distribution. ▪

Exercises

5.137 Derive the moment-generating function for the uniform distribution on the interval (a, b).

5.138 Suppose X has the uniform distribution on the interval (a, b).

a Using the properties of the moment-generating function, derive the moment-generating function of $Y = cX + d$.

b By looking at the moment-generating function derived in part (a), identify the distribution, including parameters of Y.

5.139 Find the moment-generating function for the random variable X with probability density function

$$f(x) = \frac{2x}{c}, \quad 0 < x < \sqrt{c}$$

for some constant $c > 0$.

5.140 Show that a gamma distribution with parameters α and β has the moment-generating function

$$M(t) = (1 - \beta t)^{-\alpha}.$$

5.141 Using the moment-generating function of the gamma distribution with parameters α and β derived in Exercise 5.140, find the mean and variance of that distribution.

5.142 Let Z denote a standard normal random variable. Find the moment-generating function of Z directly from the definition.

5.143 Let Z denote a standard normal random variable. Find the moment-generating function of Z^2. What does the uniqueness property of the moment-generating function tell you about the distribution of Z^2?

5.144 Using the moment-generating function for a normal distribution with parameters μ and σ, verify that the mean is μ and the variance is σ^2.

5.145 If a random variable X has the moment-generating function $M_X(t)$, then show that $Y = aX + b$, for constants a and b, has the moment-generating function $M_Y(t) = e^{tb}M_X(at)$.

5.146 Let Z denote a standard normal random variable. Find $cov(Z, Z^2)$.

5.11 Expectations of Discontinuous Functions and Mixed Probability Distributions

Problems in probability and statistics frequently involve functions that are partly continuous and partly discrete in one of two ways. First, we may be interested in the properties—perhaps the expectation—of a random variable $g(X)$ that is a discontinuous function of a discrete or continuous random variable X. Second, the random variable of interest may itself have a probability distribution made up of isolated points having discrete probabilities and of intervals having continuous probability. The first of these two situations is illustrated by the following example.

EXAMPLE 5.22 A certain retailer for a petroleum product sells a random amount X each day. Suppose that X (measured in hundreds of gallons) has the probability density function

$$f(x) = \begin{cases} \left(\frac{3}{8}\right) x^2, & 0 \le x \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

The retailer's profit turns out to be \$5 for each 100 gallons sold (5 cents per gallon) if $X \le 1$, and \$8 per 100 gallons if $X > 1$. Find the retailer's expected profit for any given day.

Solution Let $g(X)$ denote the retailer's daily profit. Then

$$g(x) = \begin{cases} 5X, & 0 \le X \le 1 \\ 8X, & 1 < X \le 2. \end{cases}$$

We want to find expected profit, and

$$\begin{aligned}
E[g(X)] &= \int_{-\infty}^{\infty} g(x) f(x)\, \mathrm{d}x \\
&= \int_0^1 5x\left[\left(\frac{3}{8}\right)x^2\right] \mathrm{d}x + \int_1^2 8x\left[\left(\frac{3}{8}\right)x^2\right] \mathrm{d}x \\
&= \frac{15}{(8)(4)}\left[x^4\right]_0^1 + \frac{24}{(8)(4)}\left[x^4\right]_1^2 \\
&= \frac{15}{32}(1) + \frac{24}{32}(15) \\
&= \frac{(15)(24)}{32} \\
&= 11.72.
\end{aligned}$$

Thus, the retailer can expect to profit by \$11.72 on the daily sale of this particular product. ▪

A random variable X that has some of its probability at discrete points and the remainder spread over intervals is said to have a mixed distribution. Let $F(x)$ denote a distribution function representing a mixed distribution. For practical purposes, any mixed distribution function $F(x)$ can be written uniquely as

$$F(x) = c_1 F_1(x) + c_2 F_2(x)$$

where $F_1(x)$ is a step distribution function, $F_2(x)$ is a continuous distribution function, c_1 is the accumulated probability of all discrete points, and $c_2 = 1 - c_1$ is the accumulated probability of all continuous portions. The following example offers an instance of a mixed distribution.

EXAMPLE 5.23 Let X denote the life length (in hundreds of hours) of a certain type of electronic component. These components frequently fail immediately upon insertion into a system; the probability of immediate failure is 1/4. If a component does not fail immediately, its life-length distribution has the exponential density

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the distribution function for X, and evaluate $P(X > 10)$.

Solution There is only one discrete point, $X = 0$, and this point has probability 1/4. Hence, $c_1 = 1/4$ and $c_2 = 3/4$. It follows that X is a mixture of two random variables, X_1 and X_2, where X_1 has a probability of 1 at the point 0 and X_2 has the given exponential density; that is

$$F_1(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$$

and

$$F_2(x) = \int_0^x e^{-y}\,dy$$
$$= 1 - e^{-x}, \quad x > 0.$$

Now,

$$F(x) = \left(\frac{1}{4}\right) F_1(x) + \left(\frac{3}{4}\right) F_2(x).$$

Hence,

$$\begin{aligned} P(X > 10) &= 1 - P(X \leq 10) \\ &= 1 - F(10) \\ &= 1 - \left[\frac{1}{4}(1) + \frac{3}{4}(1 - e^{-10})\right] \\ &= \left(\frac{3}{4}\right)[1 - (1 - e^{-10})] \\ &= \left(\frac{3}{4}\right)e^{-10}. \quad \blacksquare \end{aligned}$$

An easy method for finding expectations of random variables that have **mixed distributions** is given in Definition 5.6.

DEFINITION 5.6

Let X have the **mixed distribution function**

$$F(x) = c_1 F_1(x) + c_2 F_2(x).$$

And suppose that X_1 is a discrete random variable having distribution function $F_1(x)$, and X_2 is a continuous random variable having distribution function $F_2(x)$. Let $g(X)$ denote a function of X. Then

$$E[g(X)] = c_1 E[g(X_1)] + c_2 E[g(X_2)]. \quad \blacksquare$$

EXAMPLE 5.24 Find the mean and the variance of the random variable defined in Example 5.23.

Solution With all definitions remaining as given in Example 5.23, it follows that

$$E(X_1) = 0$$

and

$$E(X_2) = \int_0^\infty y e^{-y}\, dy = 1.$$

Therefore,

$$\begin{aligned}\mu &= E(X)\\ &= \left(\frac{1}{4}\right)E(X_1) + \left(\frac{3}{4}\right)E(X_2)\\ &= \frac{3}{4}.\end{aligned}$$

Also,

$$E(X_1^2) = 0$$

and

$$E(X_2^2) = \int_0^\infty y^2 e^{-y}\,\mathrm{d}y = 2.$$

Therefore,

$$\begin{aligned}E(X^2) &= \left(\frac{1}{4}\right)E(X_1^2) + \left(\frac{3}{4}\right)E(X_2^2)\\ &= \left(\frac{1}{4}\right)(0) + \left(\frac{3}{4}\right)(2)\\ &= \frac{3}{2}.\end{aligned}$$

Then,

$$\begin{aligned}V(X) &= E(X^2) - \mu^2\\ &= \frac{3}{2} - \left(\frac{3}{4}\right)^2\\ &= \frac{15}{16}.\end{aligned}$$

For any nonnegative random variable X, the mean can be expressed as

$$E(x) = \int_0^\infty [1 - F(t)]\,\mathrm{d}t$$

where $F(t)$ is the distribution function for X. To see this, write

$$\int_0^\infty [1 - F(t)]\,\mathrm{d}t = \int_0^\infty \left(\int_1^\infty f(x)\,\mathrm{d}x\right)\mathrm{d}t.$$

Upon changing the order of integration, we find that the integral becomes

$$\int_0^\infty \left(\int_0^x f(t)\,dt \right) dx = \int_0^\infty x f(x)\,dx = E(X).$$

Employing this result for the mixed distribution of Examples 5.23 and 5.24, we see that

$$1 - F(x) = \frac{3}{4} e^{-x}, \quad \text{for } x > 0. \quad \blacksquare$$

Exercises

5.147 A retail grocer has daily demand X for a certain food that is sold by the pound wholesale. Food left over at the end of the day is a total loss. The grocer buys the food for \$6 per pound and sells it for \$10 per pound. If demand is uniformly distributed over the interval 0 to 1 pound, how much of this food should the grocer order to maximize his expected daily profit?

5.148 Suppose that a distribution function has the form

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2 + 0.1, & 0 \le x < 0.5 \\ x, & 0.5 \le x < 1 \\ 1, & 1 \le x. \end{cases}$$

a Describe $F_1(x)$ and $F_2(x)$, the discrete and continuous components of $F(x)$.

b Write $F(X)$ as

$$c_1 F_1(x) + c_2 F_2(x).$$

c Sketch $F(x)$.

d Find the expected value of the random variable whose distribution function is $F(x)$.

5.149 Let X denote the length of time that a sensitive plant survives after transplant; 20% of the plants die immediately after transplant (time $= 0$). For the plants that survive transplant, the distribution of the time that the plant survives is exponentially distributed with a mean of 1. Thus, the distribution of the time that a plant lives after transplant is a mixture distribution.

a State and graph the distribution function for the discrete distribution in the mixture.

b State and graph the distribution function for the continuous distribution in the mixture.

c Find the mixture distribution function and sketch it.

5.150 The duration X of telephone calls coming through a cellular phone is a random variable whose distribution function is

$$F(x) = \begin{cases} 0, & x \le 0 \\ 1 - \frac{2}{3} e^{-x/3} - \frac{1}{3} e^{-[x/3]}, & x > 0 \end{cases}$$

where $[y]$ denotes the greatest integer less than or equal to y.

a Sketch $F(x)$.

b Find $P(X \le 6)$.

c Find $P(X > 4)$.

d Describe the discrete and continuous parts that make up $F(x)$.

e Find $E(X)$.

5.12 Summary

Practical applications of measurement, such as to improve the quality of a system, often require careful study of the behavior of observations measured on a continuum (such as time, weight, distance, or the like). Probability distributions for such measurements are characterized by *probability density functions* and *cumulative distribution functions*. The *uniform distribution* models the waiting time for an event that is equally likely to occur anywhere on a finite interval. The *exponential model* is useful for modeling the lengths of time between occurrences of random events. The *gamma model* fits the behavior of sums of exponential variables and, therefore, models the total waiting time until a number of random events have occurred. The most widely used continuous probability model is the *normal*, which exhibits the symmetric mound-shaped behavior often seen in data. If a histogram of the data is mound-shaped but skewed, the *Weibull* provides a useful and versatile model.

All models are approximations to reality, but the ones discussed here are relatively easy to use and yet provide good approximations in a variety of settings.

Supplementary Exercises

5.151 The "on" temperature of a thermostatically controlled switch for an air-conditioning system is set at 72°F, but the actual temperature X at which the switch turns on is a random variable having the probability density function

$$f(x) = \begin{cases} \frac{1}{2}, & 71 \le x \le 73 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the probability that a temperature in excess of 72°F is required to turn the switch on.

b If two such switches are used independently, find the probability that a temperature in excess of 72°F is required to turn both on.

5.152 Let X possess a density function

$$f(x) = \begin{cases} cx, & 0 \le x \le 4 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find c.

b Find F(x).

c Graph f(x) and $F(x)$.

d Use $F(x)$ from part (b) to find $P(1 \le X \le 2)$.

d Use the geometric figure for $f(x)$ from part (c) to calculate $P(1 \le X \le 2)$.

5.153 Let X possess a density function

$$f(x) = \begin{cases} x + cx^2, & 0 \le x \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find c.
b Find $F(x)$.
c Graph $f(x)$ and $F(x)$.
d Use $F(x)$ from part (b) to find $F(-1)$, $F(1/2)$, and $F(1)$.
e Find $P(1 \le X \le 2.5)$.
f Find the mean and the variance of X.

5.154 Let X possess a density function

$$f(x) = \begin{cases} 0.2, & -1 \le x \le 0 \\ 0.2 + cx, & 0 < x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find c.
b Find $F(x)$.
c Graph $f(x)$ and $F(x)$.
d Use $F(x)$ from part (b) to find $F(-1/2)$, $F(0)$, and $F(1/2)$.
e Find $P(-0.5 \le X \le 0.25)$.
f Find the mean and the variance of X.

5.155 Suppose X has the following distribution function:

$$F(x) = \begin{cases} 0, & x \le 0 \\ \frac{x^2}{27}\left(9 - 2x\right), & 0 < x \le 3 \\ 1, & x > 3. \end{cases}$$

a Verify that F is a distribution function.
b Find $f(x)$.
c Graph $f(x)$ and $F(x)$.
d Find $P(0.5 \le X \le 1.5)$.
e Find the mean and the variance of X.

5.156 Suppose X has the following distribution function:

$$F(x) = \begin{cases} 0, & x \le 0 \\ x\left(0.1 + 0.2x\right), & 0 < x \le 2 \\ 1, & x > 2. \end{cases}$$

a Verify that F is a distribution function.
b Find $f(x)$.
c Graph $f(x)$ and $F(x)$.
d Find $P(0.5 \le X \le 1.5)$.
e Find the mean and the variance of X.

5.157 The grade-point averages of a large population of college students are approximately normally distributed with a mean equal to 2.4 and a standard deviation equal to 0.5.

a What fraction of the students possesses a grade-point average in excess of 3.0?
b If students who possess a grade-point average equal to or less than 1.9 are dropped from college, what percentage of the students will be dropped?
c Suppose that three students are selected at random from the student body. What is the probability that all three possess a grade-point average in excess of 3.0?

5.158 A machine operation produces bearings whose diameters are normally distributed with a mean and a standard deviation of 3.001 and 0.001, respectively. Customer specifications require that the bearing diameters lie in the interval 3.000 ± 0.0020. Units falling outside the interval are considered scrap and must be remachined or used as stock for smaller bearings. At the current machine setting, what fraction of total production will be scrap?

5.159 Let X have the density function

$$f(x) = \begin{cases} cxe^{-2x}, & 0 \le x < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the value of c.

b Give the mean and the variance for X.

c Give the moment-generating function for X.

5.160 The yield force of a steel reinforcing bar of a certain type is found to be normally distributed with a mean of 8500 pounds and a standard deviation of 80 pounds. If three such bars are to be used on a certain project, find the probability that all three will have yield forces in excess of 8600 pounds.

5.161 An engineer has observed that the gap times between vehicles passing a certain point on a highway have an exponential distribution with a mean of 24 seconds.

a Find the probability that the next gap observed will be no longer than 1 minute.

b Find the probability density function for the sum of the next four gap times to be observed. What assumptions must be true for this answer to be correct?

5.162 The lifetime X of a certain electronic component is a random variable with density function

$$f(x) = \begin{cases} \left(\frac{1}{80}\right) e^{-x/80}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Three of these components operate independently in a piece of equipment. The equipment fails if two or more of the components fail. Find the probability that the equipment will operate for at least 150 hours without failure.

5.163 The proportion of time per day that all checkout counters in a supermarket are busy is a random variable X that has the probability density function

$$f(x) = \begin{cases} kx^2(1-x)^4, & 0 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the value of k that makes this a probability density function.

b Find the mean and the variance of X.

5.164 One major contributor to air pollution is sulfur dioxide. The following 12 peak sulfur dioxide measurements (in ppm) from randomly selected U.S. sites are: 0.003, 0.010, 0.014, 0.024, 0.024, 0.032, 0.038, 0.042, 0.043, 0.044, 0.047, 0.061

a Should the normal distribution be used as a model for these measurements?

b From a Q-Q plot for these data, approximate the mean and the standard deviation. Check the approximation by direct calculation.

c Is the Weibull distribution a good model for these measurements? Compare the fits of the normal and the Weibull distributions.

d Estimate the parameters of the Weibull distribution graphically.

5.165 Based on the 1988 National Survey of Families and Households, it was reported that men spent a mean of 18.1 hours per week doing housework. Suppose the standard deviation was 13.1 hours.

a Assuming a normal distribution, sketch the approximate distribution of the number of hours that a randomly selected man spent doing housework in 1988.

b Based on the graph in part (a), explain why the population distribution is very unlikely to be normal and why it is most likely skewed to the right.

c Assuming a normal distribution, what proportion of men spent an average of 20 hours or more on housework each week in 1988?

d What proportion of men spent less than an average of 5 hours on housework each week in 1988?

5.166 Refer to the setting in Exercise 5.165. Suppose now that we use the Weibull distribution with $\theta = 20$ and $\gamma = 2$ to model these data.

a Are the mean and variance for this Weibull distribution equal to those observed in the survey? Justify your answer.

b What proportion of men spent an average of 20 hours or more on housework each week in 1988?

c What proportion of men spent less than an average of 5 hours on housework each week in 1988?

d Compare the results in parts (b) and (c) to those obtained using the normal distribution.

5.167 If the life length X of a certain type of battery has a Weibull distribution with $\gamma = 2$ and $\theta = 3$ (with measurement in years), find the probability that such a battery will last less than 4 years given that it is now 2 years old.

5.168 The time (in hours) that a manager takes to interview an applicant has an exponential distribution with $\theta = 1/2$. Two applicants arrive at 8:00 a.m., and the interviews begin. A third applicant arrives at 8:45 a.m. What is the probability that the latecomer must wait before seeing the manager?

5.169 The weekly repair cost X for a certain machine has a probability density function given by

$$f(x) = \begin{cases} 3(1-x)^2, & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

with measurements in \$100s. How much money should be budgeted each week for repair cost to ensure that the actual cost exceeds the budgeted amount only 10% of the time?

5.170 Maximum wind-gust velocities in summer thunderstorms were found to follow a Weibull distribution with $\gamma = 200$ and $\theta = 400$ (measurements in feet per second). Engineers who design structures in areas where these thunderstorms occur are interested in finding a gust velocity that will be exceeded only with a probability of 0.01. Find such a value.

5.171 In a particular model of insect behavior, the time that an insect spends on a plant is exponentially distributed with a mean of 2 minutes.

a What is the shortest time that an insect can be on a plant and stay as long as at least 50% of all insects on the plants?

b A student is assigned to watch an insect until it flies. He wants to know the time that will be exceeded only with a probability of 0.05.

c If an insect has been observed for 2 minutes without its moving from the plant, what is the probability that it will be there for at least 2 more minutes?

5.172 The relationship between incomplete gamma integrals and sums of Poisson probabilities is given by

$$\frac{1}{\Gamma(\alpha)} \int_{\lambda}^{\infty} x^{\alpha-1} e^{-x}\, dx = \sum_{x=0}^{\alpha-1} \frac{\lambda^x e^{-\lambda}}{x!}$$

for integer values of α. If X has a gamma distribution with $\alpha = 2$ and $\beta = 1$, find $P(X > 1)$ by using this equality.

5.173 A random variable X is said to have a log-normal distribution if $Y = \ln(X)$ has a normal distribution. (The symbol ln denotes natural logarithm.) In this case, X must not be negative. The shape of the log-normal probability density function is similar to that of the gamma. The equation of the log-normal density function is

$$f(x) = \begin{cases} \frac{1}{\sigma x\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Because $\ln(x)$ is a monotonic function of x,

$$P(X \leq x) = P[\ln(X) \leq \ln(x)] = p[Y \leq \ln(x)]$$

where Y has a normal distribution with a mean of μ and a variance of σ^2. Thus probabilities in the log-normal case can be found by transforming them into probabilities in the normal case. If X has a log-normal distribution with $\mu = 4$ and $\sigma^2 = 2$, find the following probabilities.

a $P(X \leq 5)$

b $P(X > 10)$

5.174 If X has a log-normal distribution with parameters μ and variance σ^2, it can be shown that

$$E(X) = e^{\mu+\sigma^2/2}$$

and

$$V(X) = e^{2\mu+\sigma^2}\left(e^{\sigma^2} - 1\right).$$

The grains composing polycrystalline metals tend to have weights that follow a log normal distribution. For a certain type of aluminum, grain weights have a log normal distribution with $\mu = 3$ and $\sigma = 4$ (in units of 10^{-2} gram).

a Find the mean and the variance of the grain weights.

b Find an interval within which at least 75% of the grain weights should lie (use Tchebysheff's Theorem).

5.175 The lethal dose, the amount of a chemical or other toxin that will cause death, varies from individual to individual. The distribution of the lethal dose for a population is often modeled using a log-normal distribution; that is, the proportion of individuals who die with a dose X of the chemical or toxin is distributed according to the log-normal distribution. In a study of the mortality of a particular beetle, beetles were exposed to varying levels of gaseous carbon disulfide for 5 hours. The log-normal distribution with $\mu = 4 \log \text{mg/liter}$ and $\sigma = 1 \text{mg/liter}$ of carbon disulfide described the data well.

a Find the mean and the variance of the lethal dose of carbon disulfide after 5 hours of exposure for this population of beetles.

b Find an interval of carbon disulfide within which at least 75% of the beetles would perish.

5.176 Refer to Exercise 5.175. A large study explored the dose-response relation for inhalation anthrax in a certain type of monkeys. The log-normal distribution provided a good model for the data. Suppose the median lethal dose was 4200 spores. The standard deviation of the lethal dose was found to be 400 spores.

a Find the mean lethal dose. (Recall: Because the log-normal distribution is not symmetric, the median and mean are not equal.)

b Find the parameters μ and σ^2 of the corresponding normal distribution.

5.177 Let X denote a random variable whose probability density function is given by

$$f(x) = \left(\frac{1}{2}\right) e^{-|x|}, \quad -\infty < x < \infty.$$

Find the moment-generating function of X, and use it to find $E(X)$.

5.178 The life length X of a certain component in a complex electronic system is known to have an exponential density with a mean of 100 hours. The component is replaced at failure or at age 200 hours, whichever comes first.

a Find the distribution function for X, the length of time that the component is in use.
b Find $E(X)$.

5.179 We can show that the normal density function integrates to unity by showing that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2)ux^2}\,\mathrm{d}x = \frac{1}{\sqrt{u}}.$$

This, in turn, can be shown by considering the product of two such integrals,

$$\frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty} e^{-(1/2)ux^2}\,\mathrm{d}x\right)\left(\int_{-\infty}^{\infty} e^{-(1/2)uy^2}\,\mathrm{d}y\right) = \frac{1}{2\pi}\left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(1/2)u\left(x^2+y^2\right)}\,\mathrm{d}x\,\mathrm{d}y\right).$$

By transforming the expression to polar coordinates, show that the double integral is equal to $1/u$.

5.180 The function $B(\alpha, \beta)$ is defined by

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x.$$

Prove that

$$\mathrm{B}(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}.$$

Hint:

a Let $x = \sin^2\theta$, and show that

$$B(\alpha, \beta) = 2\int_0^{\pi/2} \sin^{2\alpha-1}\theta\cos^{2\beta-1}\theta\mathrm{d}\theta.$$

b Write $\Gamma(\alpha)\Gamma(\beta)$ as a double integral and transform the integral to polar coordinates.

The incomplete beta function, $B_x(\alpha, \beta)$, is defined as

$$B_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{d}t.$$

5.181 We can show that the normal density function integrates to unity, by showing that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-(1/2)uy^2}\,\mathrm{d}y = \frac{1}{\sqrt{u}}.$$

This, in turn, can be shown by considering the product of two such integrals,

$$\frac{1}{2\pi}\left(\int_{-\infty}^{\infty} e^{-(1/2)uy^2}\,\mathrm{d}y\right)\left(\int_{-\infty}^{\infty} e^{-(1/2)ux^2}\,\mathrm{d}x\right) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(1/2)u\left(x^2+y^2\right)}\,\mathrm{d}x\,\mathrm{d}y.$$

By transforming the expression to polar coordinates, show that the double integral is equal to $1/u$.

5.182 Show that $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$, by writing

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty y^{-1/2} e^{-y}\, dy,$$

making the transformation $y = (1/2)x^2$, and employing the result of Exercise 5.181.

REFERENCES

Abernethy, R. B., J. E. Breneman, C. H. Medlin, and G. L. Reinman. 1983. *Weibull analysis handbook*. U.S. Air Force AFWAL-TR-83-2079, AD #A143100.

Aksoy, H. 2000. Use of gamma distribution in hydrological analysis. *Turkish Journal of Engineering and Environmental Sciences* 24:419–28.

Barger, G. L., and H. C. S. Thom. 1949. A method for characterizing drought intensity in Iowa. *Agronomy Journal* 41:13–19.

Baseball Almanac. 2006. http://www.baseball-almanac.com/.

Chu, S. 2003. Using soccer goals to motivate the Poisson process. *INFORMS Transactions on Education* 3, no. 2. http://ite.pubs.informs.org/Vol3No2/Chu/.

Cohen A. C., B. J. Whitten, and Y. Ding. 1984. Modified moment estimation for the three-parameter Weibull distribution. *Journal of Quality Technology* 16:159–67.

David, H. A., and A. W. F. Edwards. 2001. *Annotated readings in the history of statistics*. New York: Springer.

Dennis, B., and R. F. Costantino. 1988. Analysis of steady-state populations with the gamma abundance model: Application to *Tribolium*. *Ecology* 69:1200–1213.

Essington, T. E., and J. F. Kitchell. 1999. New perspectives in the analysis of fish distributions: A case study on the spatial distribution of largemouth bass. *Canadian Journal of Fisheries and Aquatic Sciences* 56:52–60.

Ferrari, P., E. Cascetta, A. Nuzzolo, P. Treglia, and P. Olivotto. 1984. A behavioral approach to the measurement of motorway circulation, comfort and safety. *Transportation Research* 18A:46–51.

Flynn, M. R. 2004. A stochastic differential equation for exposure yields a beta distribution. *Annals of Occupational Hygiene* 48:491–97.

Folks, J. L. (1981). *Ideas of statistics*. New York: John Wiley & Sons.

Lun, I. Y. F., and J. C. Lam. 2000. A study of Weibull parameters using long-term wind observations. *Renewable Energy* 20:145–53.

Mahalel, D., and A. S. Hakkert. 1983. Traffic arrival patterns on a cross section of a multilane highway. *Transportation Research* 4:263–70.

Massman, W. J. 1982. Foliage distribution in old-growth coniferous tree canopies. *Canadian Journal of Forest Research* 12:10–17.

Michelson, A. A. 1878. Experimental determination of the velocity of light. *Proceedings of the American Association for the Advancement of Science* 27:71–77.

Nelson, W. 1964. *Modern probability theory and its applications*. New York: John Wiley.

Perruzzi, J. J., and E. J. Hilliard, Jr. 1984. Modeling time-delay measurement errors using a generalized beta density function. *The Journal of the Acoustical Society of America* 75:197–201.

Pinder, J. E., III, J. G. Wiener, and M. H. Smith. 1978. The Weibull distribution: A new method of summarizing survivorship data. *Ecology* 59:175–79.

Pyke, R. 1965. Spacings, with discussion. *Journal of the Royal Statistics Society*, series B 27:394–449.

Stigler, S. M. 1986. *The history of statistics: The measurement of uncertainty before 1900*. Cambridge, MA: The Belknap Press of Harvard University Press.

Uspensky, V. 1937. *Introduction to mathematical probability*. New York: McGraw-Hill.

Waller, J. S., C. Servheen, and D. A. Peterson. 2005. Probabilistic measure of road lethality. Paper presented at the Road Ecology Center Scholarship Repository, John Muir Institute of the Environment, University of California, Davis.

Zammers, J. 1984. Assessing the effect of transportation control strategies on carbon monoxide concentrations. *Air Pollution Control Association Journal* 34(6):637–642.

Zimmels, Y. 1983. Theory of hindered sedimentation of polydisperse mixtures. *American Institute Chemical Engineers Journal* 29:669–75.

6

Multivariate Probability Distributions

6.1 Bivariate and Marginal Probability Distributions

Chapters 4 and 5 dealt with experiments that produced a single numerical response, or random variable, of interest. We discussed, for example, the life lengths X of a battery or the strength Y of a steel casing. Often, however, we want to study the joint behavior of two random variables, such as the joint behavior of life length and casing strength for these batteries. Perhaps in such a study we can identify a region in which some combination of life length and casing strength is optimal in balancing the cost of manufacturing with customer satisfaction. To proceed with such a study, we must know how to handle a joint probability distribution. When only two random variables are involved, joint distributions are called bivariate distributions. We will discuss the bivariate case in some detail; extensions to more than two variables follow along similar lines.

Other situations in which bivariate probability distributions are important come to mind easily. A physician studies the joint behavior of pulse and exercise. An educator studies the joint behavior of grades and time devoted to study or the interrelationship of pretest and posttest scores. An economist studies the joint behavior of business volume and profits. In fact, most real problems we come across have more than one underlying random variable of interest.

On April 15, 1912, the ocean liner *Titanic* collided with an iceberg and sank. Of the 2201 people on the *Titanic*, 1490 perished. The question as to whether passenger class was related to survival of this disaster has been discussed extensively. In Table 6.1, the numbers of survivors and fatalities for four passenger classes from this incident, as reported by Dawson (1995), are recorded: first-class passengers, second-class passengers, third-class passengers, and the crew.

TABLE 6.1
Survival and Passenger Class of Passengers on the *Titanic*.

Passenger Status	Survivors	Fatalities	Total
First class	203	122	325
Second class	118	167	285
Third class	178	528	706
Crew	212	673	885
Total	711	1490	2201

As noted in earlier chapters, it is more convenient to analyze such data in terms of meaningful random variables rather than in terms of described categories. For each passenger, we want to know whether or not she or he survived and what passenger class she or he was in. An X defined as

$$X = \begin{cases} 0, & \text{if passenger survived} \\ 1, & \text{if passenger did not survive} \end{cases}$$

will keep track of the number of fatalities. The other variable of interest is passenger class. A Y defined as

$$Y = \begin{cases} 1, & \text{if passenger was in first class} \\ 2, & \text{if passenger was in second class} \\ 3, & \text{if passenger was in third class} \\ 4, & \text{if passenger was a crew member} \end{cases}$$

provides a random variable that keeps track of the number of passengers in each passenger class.

The frequencies from Table 6.1 are turned into the relative frequencies of Table 6.2 to produce the **joint probability distribution** of X and Y.

TABLE 6.2
Joint Probability Function of Survival and Passenger Class of Passengers on the *Titanic*.

		X = 0	X = 1	
Y	1	0.09	0.06	0.15
	2	0.05	0.08	0.13
	3	0.08	0.24	0.32
	4	0.10	0.30	0.40
		0.32	0.68	1.00

In general, we write

$$P(X = x, Y = y) = p(x, y)$$

and call $p(x, y)$ the **joint probability function** of (X, Y). From Table 6.2, we can see that

$$P(X = 0, Y = 4) = p(0, 4) = 0.10$$

represents the approximate probability that a randomly selected passenger is a crew member who survived. The properties of the probability function for jointly distributed random variables are similar to those for the univariate probability functions as summarized in Definition 6.1.

DEFINITION 6.1

Let X and Y be discrete random variables. The **joint probability distribution** of X and Y is given by

$$p(x, y) = P(X = x, Y = y)$$

defined for all real numbers x and y. The function $p(x, y)$ is called the joint probability function of X and Y. All joint probability functions must satisfy two properties:

1 $p(x, y) \geq 0, \qquad x, y \in \Re$

2 $\sum_x \sum_y p(x, y) = 1$

Further, any function that satisfies the two properties above is a probability function. ▪

Similarly, a distribution function is associated with every probability function.

The extension from one to two or more dimensions is an intuitive one. For two dimensions, the distribution function is defined as

$$F_{(X,Y)}(x, y) = P(X \leq x, Y \leq y), \qquad (x, y) \in \Re^2.$$

As with the univariate case, every distribution function must satisfy four conditions:

1 $\lim_{x \to -\infty} \lim_{y \to -\infty} F(x, y) = 0$

2 $\lim_{x \to \infty} \lim_{y \to \infty} F(x, y) = 1$

3 The distribution function is a nondecreasing function; that is, for each $a < b$ and $c < d, P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0$.

4 For fixed x or Y, the distribution function is right-hand continuous in the remaining variable. For example, for x fixed, $\lim_{h \to 0^+} F(x, y + h) = F(x, y)$.

It is important to check each condition as the following example illustrates.

DEFINITION 6.2

The **joint distribution function** $F(a, b)$ for a bivariate random variable (X, Y) is

$$F(a, b) = P(X \leq a, Y \leq b).$$

If X and Y are discrete,

$$F(a, b) = \sum_{x=-\infty}^{a} \sum_{y=-\infty}^{b} p(x, y)$$

where $p(x, y)$ is the joint probability function.

The distribution function is often called the cumulative distribution function (cdf). ▪

EXAMPLE 6.1 Let

$$F(x, y) = \begin{cases} 0, & x + y < 0 \\ 1, & x + y \geq 0. \end{cases}$$

Is F a distribution function? Justify your answer.

Solution The first condition is satisfied because $\lim_{x \to -\infty} F(x, y) = \lim_{y \to -\infty} F(x, y) = 0$. Similarly, $\lim_{x \to \infty} F(x, y) = \lim_{y \to \infty} F(x, y) = 1$ so the second condition is satisfied. The fourth condition is satisfied. This is verified by focusing on the points of discontinuity for the function. These occur at the line $x + y = 0$. Yet, because the function is right-hand continuous for all points on this line, it is right-hand continuous. The third condition, however, is not satisfied. To see this, notice that

$$\begin{aligned} P(-1 < X \leq 2, -1 < Y \leq 2) &= F(2, 2) - F(-1, 2) - F(2, -1) + F(-1, -1) \\ &= 1 - 1 - 1 + 0 = -1. \end{aligned}$$

Because probabilities must be positive, the above probability of -1 is not valid. Therefore, this function is not a distribution function. ▪

Although we may have information on the joint distribution of two random variables, we may still want to ask questions involving only one of them. Look again at the joint distribution of survival and passenger class displayed in Table 6.2. If we add the probabilities of surviving and not surviving in a given row (passenger class), we have the proportion of passengers in that passenger class. As an example, the probability that a randomly selected passenger will be a crew member is

$$\begin{aligned} P(Y = 4) &= P(X = 0, Y = 4) + P(X = 1,\ Y = 4) \\ &= 0.10 + 0.30 = 0.40 \end{aligned}$$

which is one of *the marginal probabilities* for Y. The probabilities for each passenger class are displayed in the right margin of Table 6.1 and constitute the **marginal probability function** of Y (passenger class). Similarly, the univariate distribution along the bottom row is the marginal distribution for X. Formally, the marginal probability functions are defined in Definition 6.3.

DEFINITION 6.3 The **marginal probability functions** of X and Y, respectively, are given by

$$p(x) = \sum_y p(x, y)$$

and

$$p(y) = \sum_x p(x, y). \quad ▪$$

EXAMPLE 6.2 Three checkout counters are in operation at a local supermarket. Two customers arrive at the counters at different times when the counters are serving no other customers. It is assumed that the customers then choose a checkout station at random and independently of one another. Let X denote the number of times counter A is selected, and let Y denote the number of times counter B is selected by the two customers.

1 Find the joint probability distribution of X and Y.

2 Find the marginal probability distributions of X and Y.

Solution **1** For convenience, let us introduce Z—the number of customers who visit counter C. Now the event $(X = 0, Y = 0)$ is equivalent to the event $(X = 0, Y = 0, Z = 2)$. It follows that

$$\begin{aligned}
P(X = 0,\ Y = 0) = P(X = 0,\ Y = 0,\ Z = 2)& \\
&= P(\text{customer I selects counter C and customer II selects counter C}) \\
&= P(\text{customer I selects counter C}) \times P(\text{customer II selects counter C}) \\
&= \frac{1}{3} \times \frac{1}{3} \\
&= \frac{1}{9}
\end{aligned}$$

because customers' choices are independent and because each customer makes a random selection from among the three available counters.

It is slightly more complicated to calculate $P(X = 1, Y = 0) = P(X = 1, Y = 0, Z = 1)$. In this event, customer I could select counter A and customer II could select counter C, or I could select C and II could select A. Thus,

$$\begin{aligned}
P(X = 1,\ Y = 0) &= P(\text{I selects A})P(\text{II selects C}) + P(\text{I selects C})P(\text{II selects A}) \\
&= \left(\frac{1}{3} \times \frac{1}{3}\right) + \left(\frac{1}{3} \times \frac{1}{3}\right) \\
&= \frac{2}{9}.
\end{aligned}$$

Similar arguments allow one to derive the results given in Table 6.3.

2 We now need to find the marginal probability functions of X and Y, the probabilities associated with the number of customers visiting each of the two counters.

TABLE 6.3 Probability Distribution Associated with Example 6.2.

		X			
		0	**1**	**2**	**Marginal Probabilities for *Y***
	0	1/9	2/9	1/9	4/9
Y	1	2/9	2/9	0	4/9
	2	1/9	0	0	1/9
Marginal Probabilities for X		4/9	4/9	1/9	1

Notice that these are also displayed in Table 6.3. For each counter the probability is 4/9 of having 0 of the 3 customers visit that counter, 4/9 of having one customer, and 1/9 of having both customers. ▪

Before we move to the bivariate continuous case, let us quickly review the situation in one dimension. If $f(x)$ denotes the probability density function of a random variable X, then $f(x)$ represents a relative frequency curve, and probabilities, such as $P(a \leq X \leq b)$, are represented as areas under this curve. In symbolic terms,

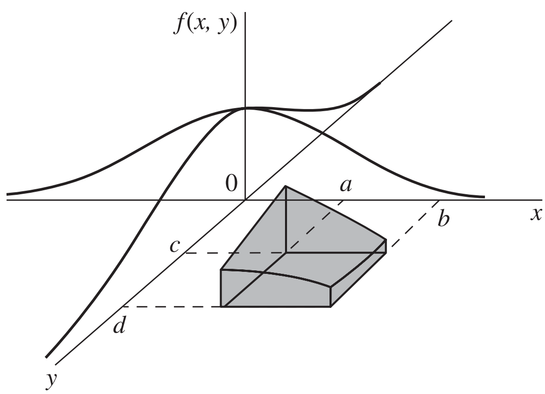
$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x.$$

Now suppose that we are interested in the joint behavior of two continuous random variables—say, X and Y—where X and Y might represent the amounts of two different hydrocarbons found in an air sample taken for a pollution study, for example. The relative frequency of these two random variables can be modeled by a bivariate function, $f(x, y)$, which forms a probability, or relative frequency, surface in three dimensions. Figure 6.1 shows such a surface. The probability that X lies in one interval and that Y lies in another interval is then represented as a volume under this surface. Thus,

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_c^d \int_a^b f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Notice that this integral simply gives the volume under the surface and over the shaded region in Figure 6.1. We trace the actual computations involved in such a bivariate problem with a simple example.

FIGURE **6.1**
A bivariate density function.

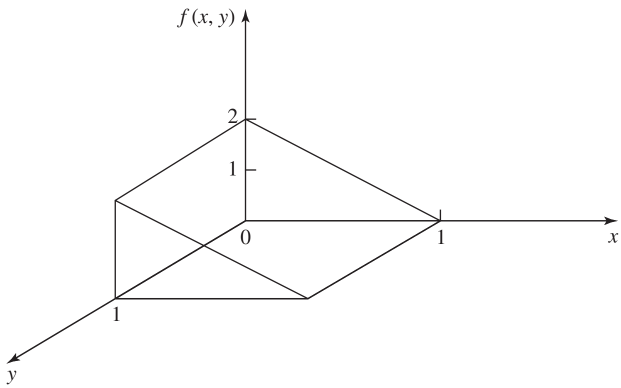


EXAMPLE 6.3 A certain process for producing an industrial chemical yields a product that contains two main types of impurities. For a certain volume of sample from this process, let X denote the proportion of total impurities in the sample, and let Y denote the proportion of type 1 impurity among all impurities found. Suppose that after investigation of many such samples, the joint distribution of X and Y can be adequately modeled by the following function:

$$f(x, y) = \begin{cases} 2(1-x), & \text{for } 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

This function graphs as the surface given in Figure 6.2. Calculate the probability that X is less than 0.5 and that Y is between 0.4 and 0.7.

FIGURE 6.2 Probability density function for Example 6.3.



Solution From the preceding discussion, we see that

$$\begin{aligned} P(0 \leq x \leq 0.5, 0.4 \leq y \leq 0.7) &= \int_{0.4}^{0.7} \int_{0}^{0.5} 2(1-x)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_{0.4}^{0.7} \left[-(1-x)^2\right]_0^{0.5} \mathrm{d}y \\ &= \int_{0.4}^{0.7} (0.75)\,\mathrm{d}y \\ &= 0.75y\Big]_{0.4}^{0.7} \\ &= (0.75)(0.3) \\ &= 0.225. \end{aligned}$$

Thus, the fraction of such samples having less than 50% impurities of which between 40% and 70% are type 1 impurities is 0.225. ▪

As with discrete bivariate distributions, the distribution function is

$$F(x, y) = P(X \le x, Y \le y), \qquad (x, y) \in \Re^2.$$

However, to find the probabilities, we integrate, instead of sum, over the appropriate regions. The four conditions of a distribution function are the same for both discrete and continuous random variables.

Just as the univariate (or marginal) probabilities were computed by summing over rows or columns in the discrete case, the univariate density function for X in the continuous case can be found by integrating ("summing") over values of Y. Thus, the marginal density function of X, $f_X(x)$, is given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y.$$

Similarly, the marginal density function of Y, $f_Y(y)$, is given by

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}x.$$

Note: We now have three different probability density functions corresponding to the density of X, the density of Y, and the joint density of X and Y. Subscripts are often used to clarify which distribution is being used. Here the joint density is given without subscripts. Subscripts for other densities, such as the marginal probability densities, will be used if needed to avoid possible confusion.

The definitions of **joint probability density, cumulative distribution,** and **marginal distribution function** are given in Definition 6.4.

DEFINITION 6.4

Let X and Y be continuous random variables. The **joint probability density function** of X and Y, if it exists, is given by a nonnegative function $f(x, y)$ such that

$$P(a \le X \le b, c \le Y \le d) = \int_{c}^{d}\int_{a}^{b} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The **cumulative distribution function** is

$$F_{(X,Y)}(a, b) = P(X \le a, Y \le b)$$
$$= \int_{-\infty}^{b}\int_{-\infty}^{a} f(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The **marginal probability functions** of X and Y, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx. \quad \blacksquare$$

EXAMPLE 6.4 For the case given in Example 6.3, find the marginal probability density functions for X and Y.

Solution Let us first try to visualize what the answers should look like before going through the integration. To find $f_X(x)$, we accumulate all the probabilities in the y direction. Look at Figure 6.2 and think of collapsing the wedge-shaped figure back onto the $[x, f(x, y)]$ plane. Then more probability mass will build up toward the 0 point of the x-axis than toward the unity point. In other words, the function $f_X(x)$ should be high at 0 and low at 1. Formally,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \int_0^1 2(1 - x)\, dy \\ &= 2(1 - x)y\Big]_0^1 \\ &= \begin{cases} 2(1 - x), & 0 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

The graph of $f_X(x)$ appears in Figure 6.3. Notice that our conjecture is correct. Thinking of how $f_Y(y)$ should look geometrically, imagine that the wedge of Figure 6.2 has been forced back onto the $[y, f(x, y)]$ plane. Then the probability mass should accumulate equally all along the (0, 1) interval on the y-axis. Mathematically,

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx$$

$$= \int_0^1 2(1 - x)\,dx$$

$$= \left[-(1 - x)^2\right]_0^1$$

$$= \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Thus, Y is a uniform random variable on the interval [0, 1]. So again our conjecture is verified, as shown in Figure 6.4.

FIGURE 6.3
Probability density function $f_X(x)$ for Example 6.3.

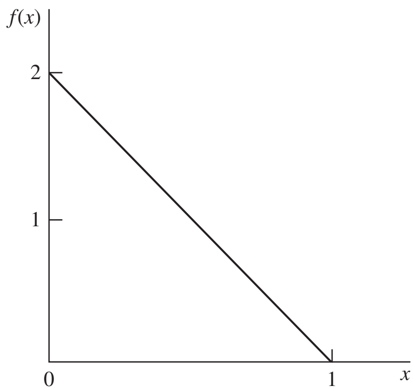
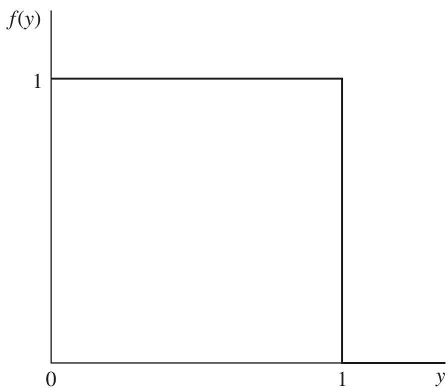


FIGURE 6.4
Probability density function $f_Y(y)$ for Example 6.3.



Following is another (and somewhat more complicated) example. ▪

EXAMPLE 6.5 A soft-drink machine has a random supply Y at the beginning of a given day and dispenses a random amount X during the day (with measurements in gallons). It is not resupplied during the day; hence, $X \leq Y$. It has been observed that X and Y have joint density

$$f(x, y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq y,\ 0 \leq y \leq 2 \\ 0, & \text{elsewhere.} \end{cases}$$

In other words, the points (x, y) are uniformly distributed over the triangle with the given boundaries.

1. Evaluate the probability that the supply is less than 1/2 gallon on a randomly selected day.
2. Evaluate the probability that the amount dispensed is less than 1/2 gallon on a randomly selected day.

Solution 1 Because the question refers to the supply in the soft-drink machine, the marginal behavior of Y is to be addressed. Thus, it is necessary to find the marginal density, $f_Y(y)$, by integrating the joint density of X and Y over X. Unlike the last example, the bounds on the region of positive probability for X variable depend on the value of Y. A graph depicting the region of positive probability is shown in Figure 6.5. From this graph we see that for a given value of $Y = Y$, the positive probability is associated with the values of X from 0 to Y. Thus, we have

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f(x, y)\, \mathrm{d}x \\ &= \begin{cases} \int_0^y \left(\frac{1}{2}\right) \mathrm{d}x = \left(\frac{1}{2}\right) y, & 0 \leq y \leq 2 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

The probability that the supply will be less than 1/2 gallon for a randomly selected day is

$$P(Y \leq y) = \int_0^{1/2} \frac{y}{2}\, \mathrm{d}y = \frac{y^2}{4}\bigg|_0^{1/2} = \frac{1}{16}.$$

Therefore, the probability of having a supply less than 1/2 gallon is 1/16, which is a fairly small probability.

2 Here the question refers to the amount dispensed from the soft-drink machine so the marginal behavior of X is to be addressed. To find the marginal density, $f_X(x)$, we integrate the joint density of X and Y over Y. In the first part of this example, we noted that the bounds on the region of positive probability for X depend on the value of Y. It is important to remember that if the bounds of positive probability for one random variable depend on the value of a second random variable, the bounds of the second random variable also depend on that of the first. As applied here, this

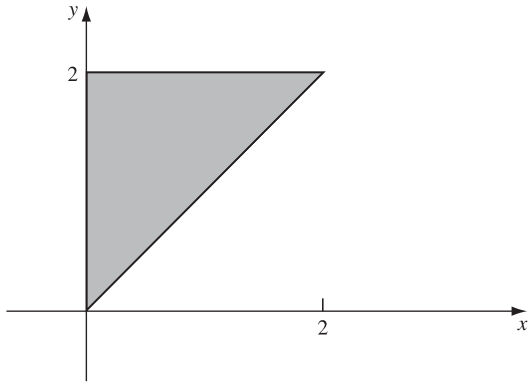


FIGURE 6.5
The region of positive probability for Example 6.5.

means that because the bounds on the region of positive probability for X depend on the value of Y, the region of positive probability for Y must also depend on the value of X. So whereas we might immediately think of integrating over Y from 0 to 2 simply because of the way the joint probability density function is presented, we know we must look deeper and note that Y must be greater than X; therefore, the graph depicting the region of positive probability shown in Figure 6.5 applies here as well. From this graph we see that for a given value of $X = x$, the positive probability is associated with the values of Y from x to 2. Thus, we have

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x, y)\,\mathrm{d}y \\ &= \int_{x}^{2} \left(\frac{1}{2}\right) \mathrm{d}y \\ &= \begin{cases} \frac{2-x}{2}, & 0 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Note: The marginal density of X (Y) does not involve Y (X), either in expressing the height of the density function or in designating the region of positive probability. This is true regardless of the joint behavior of X and Y.

The probability that less than $1/2$ a gallon of soft drink will be dispensed on a randomly selected day is

$$P(X \le x) = \int_{0}^{1/2} \frac{2-x}{2}\,\mathrm{d}x = \left(x - \frac{x^2}{4}\right)\Bigg|_{0}^{1/2} = \frac{1}{2} - \frac{1}{16} = \frac{7}{16}.$$

That is, the probability that less than $1/2$ gallon of soft drink will be dispensed from the machine on a randomly selected day is $1/4$. ▪

Exercises

6.1 Two construction contracts are to be randomly assigned to one or more of three firms. Numbering the firms I, II, and III, let X be the number of contracts assigned to firm I, and let Y be the number assigned to firm II. (A firm may receive more than one contract.)

a Find the joint probability distribution for X and Y.

b Find the marginal probability distribution for Y.

c Find $P(X = 1|Y = 1)$.

6.2 A mountain rescue service studied the behavior of lost hikers so that more effective search strategies could be devised. They decided to determine both the direction traveled and the experience level of hikers. From this study, it is known that the probabilities of a hiker being experienced or not and of going uphill or downhill or remaining in the same place are as shown in the table that follows. Let X be the random variable associated with experience and Y be the random variable associated with direction.

	Direction		
	Uphill	**Downhill**	**Remain in Same Place**
Novice	0.10	0.25	0.25
Experienced	0.05	0.10	0.25

a Define the random variables X and Y.

b Find the marginal distribution of the direction that a lost hiker travels.

6.3 The U.S. Census Bureau collects information on certain characteristics in the U.S. population. For 2004, it reported the following numbers of citizens by age and whether or not they were in poverty as shown in the following table (all counts are in 1000s of people).

Age (in years)	**Live in Poverty**	**Do Not Live in Poverty**	**Totals**
Under 18	347	12,680	13,027
18 to 64	517	19,997	20,514
65 and older	130	3,327	3,457
Totals	994	36,004	36,998

Suppose a person is selected at random from the citizens of the United States and his or her age and poverty status are observed. Let X, the random variable associated with age, and Y, the random variable representing poverty status, be defined as follows:

$$X = \begin{cases} 9, & \text{if age} < 18 \\ 41, & \text{if } 18 \leq \text{age} \leq 64 \\ 82, & \text{if age} \geq 65 \end{cases}$$

and

$$Y = \begin{cases} 0, & \text{if not living in poverty} \\ 1, & \text{if living in poverty.} \end{cases}$$

a Find the joint probability distribution of X and Y.

b Find the marginal probability distribution of age.

c Find the marginal probability distribution of poverty in the United States.

6.4 The National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC), among others, is interested in the causes of death and its relationship to gender, age, race, and so on. For children ages 1 to 4, the numbers of deaths for each sex and by cause during 2002 are presented in the table that follows. Let X be the random variable associated with cause of death and Y be the random variable representing sex.

Cause of Death	Female	Male	Total
Accidents (unintentional injuries)	617	1024	1641
Congenital Problems	248	282	530
Assault (homicide)	189	234	423
Other	998	1266	2264
Total	2052	2806	4858

Source: CDC/NCHS.

a Define the random variables X and Y.

b Find the joint distribution of sex and cause of death.

c Find the marginal distribution of sex.

d Find the marginal distribution of the cause of death.

6.5 A radioactive particle is randomly located in a square area with sides that are 1 unit in length. Let X and Y denote the coordinates of the particle. Because the particle is equally likely to fall in any subarea of fixed size, a reasonable model of (X, Y) is given by

$$f(x, y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Sketch the probability density surface.

b Find $P(X \le 0.2, Y \le 0.4)$.

c Find $P(0.1 \le X \le 0.3, Y > 0.4)$.

6.6 An environmental engineer measures the amount (by weight) of particulate pollution in air samples (of a certain volume) collected over the smokestack of a coal-fueled power plant. Let X denote the amount of pollutant per sample when a certain cleaning device on the stack is not operating, and let Y denote the amount of pollutants per sample when the cleaning device is operating under similar environmental conditions. It is observed that X is always greater than $2Y$, and the relative frequency of (X, Y) can be modeled by

$$f(x, y) = \begin{cases} k, & 0 \le x \le 2,\ 0 \le y \le 1,\ 2y \le x \\ 0, & \text{elsewhere.} \end{cases}$$

(In other words, X and Y are randomly distributed over the region inside the triangle bounded by $x = 2$, $y = 0$, and $2y = x$.)

a Find the value of k that makes this a probability density function.

b Find $P(X \ge 3Y)$. (In other words, find the probability that the cleaning device will reduce the amount of pollutant by $1/3$ or more.)

6.7 Refer to Exercise 6.5.

a Find the joint cumulative distribution function for X and Y.

b Use the joint cumulative distribution function to find the probability that $X < 0.5$ and $Y < 0.8$.

c Verify the probability in part (b) using the joint probability density function.

6.8 A child prepares a mixture of peanut butter, bananas, and jam for his school lunch daily. Let X and Y denote the proportions of peanut butter and bananas, respectively. The mixture proportions vary from day to day according to the joint probability density given by

$$f(x, y) = \begin{cases} 50, & 0.4 \le x \le 0.6,\ 0.10 \le y \le 0.30,\ 0 \le x + y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

For a randomly selected day, find the probability of the mixture having the following characteristics.

a The proportion of peanut butter is less than $1/2$ and the proportion of bananas is less than $1/5$.

b The proportion of peanut butter is less than $1/2$ and the proportion of bananas is greater than $1/5$.

c The proportion of peanut butter is less than $1/2$.

d The proportion of peanut butter is greater than $1/2$ given that the proportion of bananas is greater than $1/5$.

6.9 Let X and Y denote the proportions of two different chemicals found in a sample mixture of chemicals used as an insecticide. Suppose that X and Y have a joint probability density given by

$$f(x, y) = \begin{cases} 2, & 0 \le x \le 1,\ 0 \le y \le 1,\ 0 \le x + y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $P(X \le 1/4,\ Y \le 1/2)$.

b Find $P(X \le 3/4,\ Y \le 3/4)$.

c Find $P(X \le 1/4,\ | Y \le 1/2)$.

6.10 Consider the distribution of particulate pollution as in Exercise 6.6.

a Find the marginal density function of the amount of pollutants per sample when the cleaning device on the stack is not operating. Sketch the density function.

b Find the marginal density function of the amount of pollutants per sample when the cleaning device on the stack is operating and sketch the density function.

6.11 As in Exercise 6.5, the distribution of the location of a radioactive particle on a unit square is of interest.

a Find the marginal density function for X and sketch the density function.

b Find the marginal density function for Y.

6.12 Refer to the setting described in Exercise 6.8.

a Find the marginal density function of the proportion of peanut butter in the sandwich's mixture and sketch the density function.

b Find the marginal density of the proportion of bananas in the sandwich's mixture and sketch the density function.

6.13 Refer to the setting described in Exercise 6.9.

a Find the marginal density function for X and sketch the density function.

b Find the marginal density function for Y and sketch the density function.

6.14 Let X and Y denote the proportions of time out of one workweek that employees I and II, respectively, actually spend performing their assigned tasks. The joint relative frequency behavior of X and Y is modeled by the probability density function

$$f(x, y) = \begin{cases} x + y, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $P(X < 1/2,\ Y > 1/4)$.

b Find $P(X + Y \le 1)$.

c Find the marginal distributions of X and Y, the distributions of the time that each employee spends performing the assigned tasks.

d Find the probability that the first employee spends more than 75% of the workweek on his assigned task.

6.15 A bombing target is in the center of a circle with a radius of 1 mile. A bomb falls at a randomly selected point inside that circle. If the bomb destroys everything within $1/2$ mile of its landing point, what is the probability that it will destroy the target?

6.16 This problem is known as Buffon's needle problem, a problem first posed by the French naturalist Buffon in 1773 and which he subsequently solved in 1777. A floor is ruled with parallel lines, each a distance D apart. A needle of length L ($L \le D$) is randomly dropped onto the floor. What is the probability that the needle will intersect one of the lines? (The other possibility is that the needle will lie fully between two lines.)

6.2 Conditional Probability Distributions

Before considering conditional probability distributions, think back to the conditional probability discussed in Chapter 3. For two events A and B, the conditional probability of A given B may be computed as

$$P(A \mid B) = \frac{P(AB)}{P(B)}$$

where $P(B) > 0$. An analogous approach can be used with probability functions where the joint density of X and Y replaces $P(AB)$, and the marginal densities of X and Y replace $P(A)$ and $P(B)$. That is, an intuitive definition of the conditional density of X given Y is

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

for $f_Y(y) > 0$. We shall soon see that this intuition is correct.

Consider again the joint distribution of survival and passenger class of passengers on the *Titanic* given in Table 6.2. This distribution allows us to discuss the probabilities of passenger class and survival for randomly selected passengers. We can discuss the probability that a randomly selected passenger survived by working with the marginal distribution of survival. Similarly, by viewing the marginal distribution of class, we have the probability that a randomly selected passenger is in first class or some other class. Yet none of these address some of the interesting questions that have been raised since the *Titanic* sank. As an example, did survival depend on the class of the passenger? That is, was it more likely for a passenger in first class to survive than passengers in other classes? To answer these types of questions, we need to find the **conditional distribution** of survival given the class of passage. We might also be interested in whether a higher proportion of the surviving passengers had been in first class or were on the crew. The conditional distribution of class given survival would be needed to address that question.

For a bivariate discrete distribution, the conditional probability distribution X, given Y, fixes a value of Y (a row of Table 6.2) and looks at the relative frequencies for values of X in that row (see Definition 6.5). For example, conditioning on $Y = 1$, produces

$$\begin{aligned} P(X = 0 \mid Y = 1) &= \frac{P(X = 0, Y = 1)}{P(Y = 1)} \\ &= \frac{0.09}{0.15} = 0.60 \end{aligned}$$

and

$$P(X = 1 \mid Y = 1) = 0.40.$$

These numbers are summarized in Table 6.4.

The conditional probability distribution in Table 6.4 quantifies how survivorship behaved for passengers in first class. For third-class passengers, the conditional probability distribution is shown in the Table 6.5.

TABLE **6.4**
Conditional Distribution of Survival Given the Passenger Was in First Class.

X	0	1
$p(x \mid y = 1)$	0.60	0.40

TABLE **6.5**
Conditional Distribution of Survival Given the Passenger Was in Third Class.

X	0	1
$p(x \mid y = 3)$	0.25	0.75

What were the chances of survival for the second-class passengers and the crew? Notice that the probability of survival depends on passenger class; a randomly selected person in first class is much more likely to survive than one in third class.

Another set of conditional distributions here consists of those for Y (passenger class) conditioning on X (survival). In this case,

$$\begin{aligned}
P(Y = 1 \mid X = 0) &= \frac{P(X = 0, Y = 1)}{P(X = 0)} \\
&= \frac{0.09}{0.32} = 0.28 \\
P(Y = 2 \mid X = 0) &= \frac{0.05}{0.32} = 0.16 \\
P(Y = 3 \mid X = 0) &= \frac{0.08}{0.32} = 0.25
\end{aligned}$$

and

$$P(Y = 4 \mid X = 0) = \frac{0.10}{0.32} = 0.31.$$

The conditional probability of passenger class given survival is summarized in Table 6.6.

TABLE **6.6**
Conditional Distribution of Passenger Class Given the Passenger Survived.

Y	0	1	2	3
$p(y \mid x = 0)$	0.28	0.16	0.25	0.31

Notice that the probability that a randomly selected survivor is a crew member is 0.31. Does that mean that they had the highest rate of survival? Not necessarily. There were more crew members than any other passenger class. If the survival was the same for all classes, then we would expect more of the survivors to be crew members. Was the probability of a crew member surviving larger than that for other passenger classes?

We first worked with the data from the *Titanic* in Exercises 3.58 and 3.72. There we defined events and found the probabilities, odds, and odds ratios for particular outcomes. Here we have used random variables, but notice that the basic processes are the same. Keep in mind throughout this chapter that what we are doing here is not so much new as a different approach to the analysis.

DEFINITION 6.5

The **conditional distribution** of X given $Y = y$ is

$$P(X = x \mid Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}$$

which is defined for all real values of X. Similarly, for all real values of Y, the conditional distribution of Y given $X = x$ is

$$P(Y = y \mid X = x) = \frac{P(X = x, Y = y)}{P(X = x)}. \quad ▪$$

EXAMPLE 6.6

As in Example 6.2, suppose checkout counters are in operation at a local supermarket. Two customers arrive at the counters at different times when the counters are serving no other customers. It is assumed that the customers then choose a checkout station at random and independently of one another. Let X denote the number of times counter A is selected, and let Y denote the number of times counter B is selected by the two customers.

1 Find the probability that one of the customers visits counter B given that one of the customers is known to have visited counter A.

2 Find the conditional distribution of the number of customers visiting counter B given that one customer visited counter A.

Solution

1 The joint and marginal distributions of X and Y found in Example 6.2 are used here. We want to find the conditional probability of one customer visiting counter $B(Y = 1)$ given that one visited counter $A(X = 1)$.

$$\begin{aligned} P(Y = 1 \mid X = 1) &= \frac{P(X = 1, Y = 1)}{P(X = 1)} \\ &= \frac{2/9}{4/9} \\ &= \frac{1}{2}. \end{aligned}$$

Does this answer (1/2) agree with your intuition?

2 The conditional distribution of the number of customers visiting counter B given that one visited counter A was partially found in the first part. To express the conditional distribution, we need to find the probability associated with each possible outcome of Y given the particular value of X that was observed. Notice that if one customer visits counter A, it is not possible for two to visit counter B so the probability that $Y = 2$ is 0 and thus not included in the following table.

Conditional Distribution of Visiting Counter B Given That One Customer Visited Counter A

y	0	1
$p(y \mid x = 1)$	0.5	0.5

■

In the bivariate discrete case, the conditional probabilities of X for a given Y were found by fixing attention on the particular row in which $Y = y$ and then looking at the relative probabilities within that row; that is, the individual cell probabilities are divided by the marginal total for that row to obtain conditional probabilities.

In the bivariate continuous case, the form of the probability density function representing the conditional behavior of X for a given Y is found by slicing through the joint density in the x direction at the particular value of Y. The function then has to be weighted by the marginal density function for Y at that point. The manipulations to obtain conditional density functions in the continuous case are analogous to those used to obtain conditional probabilities in the discrete case, except that integrals are used instead of sums. A conditional density function is defined formally in Definition 6.6.

DEFINITION 6.6 Let X and Y be continuous random variables with joint probability density function $f(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$, respectively. Then, the **conditional probability density function** of X given $Y = y$ is defined by

$$f_{X|Y}(x \mid y) = \begin{cases} \frac{f(x,y)}{f_Y(y)}, & \text{for } f_Y(y) > 0 \\ 0, & \text{elsewhere} \end{cases}$$

and the conditional probability density function of Y given $X = X$, is defined by

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{f(x,y)}{f_X(x)}, & \text{for } f_X(x) > 0 \\ 0, & \text{elsewhere.} \end{cases} \quad ■$$

EXAMPLE 6.7 Consider again the soft-drink machine in Example 6.5. Evaluate the probability that less than $1/2$ gallon is sold given that the machine contains 1 gallon at the start of the day.

Solution We were given the joint probability density of X and Y and found the marginal density of Y in Example 6.5. By Definition 6.6,

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}$$
$$= \begin{cases} \frac{1/2}{(1/2)y} = \frac{1}{y}, & 0 \le x \le y \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Now, if we know that $Y = 1$, then the conditional density function becomes

$$f_{X|Y}(x \mid y = 1) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The probability of interest is

$$P(X < 1/2|Y = 1) = \int_{-\infty}^{\infty} f_{X|Y}(x \mid y = 1)\,\mathrm{d}x$$
$$= \int_{0}^{1/2} (1)\,\mathrm{d}x$$
$$= \frac{1}{2}.$$

Thus, the amount sold is highly dependent on the amount in supply. We chose to find the general form of the conditional distribution and then to substitute in the known value of Y. Instead we could have found

$$f_{X,Y}(x, 1) = \begin{cases} \frac{1}{2}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_Y(1) = \begin{cases} \left(\frac{1}{2}\right) y = \frac{1}{2}, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$$

and, finally,

$$f_{X|Y}(x \mid y = 1) = \frac{f_{X,Y}(x, 1)}{f_Y(1)} = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Then given that 1 gallon of soft drink is in the machine at the start of the day, the probability that less than 1/2 gallon will be sold on that day is

$$\begin{aligned} P(X < 1/2|Y = 1) &= \int_0^{1/2} f_{X|Y}(x \mid y = 1)\,\mathrm{d}x \\ &= \int_0^{1/2} 1\,\mathrm{d}x \\ &= \frac{1}{2}. \end{aligned}$$

The second approach may have been easier for this particular problem. However, if after responding we are asked what the probability is that less than 1/2 gallon is dispensed if 1.5 gallons are available at the beginning of the day, or some other similar question, we would have more work to do if we had initially used the second approach than if we had taken the more general approach.

Recall that the marginal density of X (or Y) does not involve Y (or X). However, note that a relationship between X and Y may again be evident in the conditional distribution of X given Y (or Y given X), both in expressing the height of the density function and in designating the region of positive probability. ▪

6.3 Independent Random Variables

Once again recall from Chapter 3 that two events A and B are **independent** if $P(AB) = P(A)P(B)$. Somewhat analogously, two discrete random variables are independent if and only if

$$P(X,\ Y) = P(X = x)P(Y = y)$$

for all real numbers x and y. Consider the data from the *Titanic* again. If X and Y are not independent, then we need only find one example for which $P(X = x, Y = y) \neq P(X = x)P(Y = y)$. From Table 6.2, we have

$$P(X = 0,\ Y = 1) = 0.09 \neq 0.048 = 0.32(0.15) = P(X = 0)P(Y = 1).$$

Thus, X and Y are not independent. To show independence, we must ensure that equality holds for all x and y, but to show that two variables are not independent, it is enough to find one case for which equality does not hold. Therefore, survival is not independent of passenger class.

An equivalent argument from a slightly different perspective is that because the probability of survival (X) depends on passenger class (Y), X (passenger class) and

Y (survival) are ***not*** independent. The foundation for this argument is that two random variables X and Y are independent if and only if the conditional probabilities of X given Y and of Y given X are equal to the marginal probabilities of X and Y, respectively. That is, if X and Y are independent, knowledge of the observed value of X (or Y) does not provide any information about the probability of observing specific values of Y (or X). That is not the case here. If we are given that a passenger is in first class ($Y = 1$), the probability of survival ($X = 0$) is 0.6, which is substantially greater than 0.32, which is the probability of survival of a randomly selected passenger. We can further verify this by comparing the conditional distributions of survival for first- and third-class passengers in Tables 6.3 and 6.4, respectively, to the marginal distribution of survival in Table 6.2. Because these distributions are not all equal, we know that survival was not independent of passenger class. In practice, we need not look at the question of independence from so many perspectives; it is enough to consider only one.

A similar idea carries over to the continuous case.

DEFINITION 6.7

Discrete random variables X and Y are said to be **independent** if

$$P(X = x,\ Y = y) = P(X = x)P(Y = y)$$

for all real numbers x and y.

Continuous random variables X and Y are said to be independent if

$$f(x,\ y) = f_X(x)f_Y(y)$$

for all real numbers x and y. ▪

The concepts of joint probability density functions and independence extend immediately to n random variables, where n is any finite positive integer. The n random variables $X_1,\ X_2,\ \ldots,\ X_n$ are said to be independent if their joint density function, $f(x_1,\ x_2,\ \ldots,\ x_n)$ is given by

$$f(x_1,\ x_2,\ \ldots,\ x_n) = f(x_1)f(x_2)\ \cdots\ f(x_n)$$

for all real numbers $x_1,\ x_2,\ \ldots,\ x_n$.

EXAMPLE 6.8 Show that the random variables in Example 6.3 are independent.

Solution Here,

$$f(x, y) = \begin{cases} 2(1-x), & 0 \le x \le 1;\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

We saw in Example 6.4 that

$$f_X(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere} \end{cases}$$

and

$$f_Y(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Thus, $f(x, y) = f_X(x) f_Y(y)$ for all real numbers x and y; and X and Y are independent random variables. ▪

Exercises

6.17 As described in Exercise 6.2, a mountain rescue service studied the behavior of lost hikers so that more effective search strategies could be devised. Use the definitions of X and Y developed there.

	Direction		
	Uphill	**Downhill**	**Remain in Same Place**
Novice	0.10	0.25	0.25
Experienced	0.05	0.10	0.25

a Find the conditional distribution of direction given that the hiker is experienced.

b Are experience and the direction a hiker walks independent? Justify your answer.

6.18 In Exercise 6.4, we were interested in the causes of death and its relationship to sex, age, race, and so on As we did there, let X be the random variable associated with cause of death and Y be the random variable representing sex.

Cause of Death	**Female**	**Male**	**Total**
Accidents (unintentional injuries)	617	1024	1641
Congenital Problems	248	282	530
Assault (homicide)	189	234	423
Other	998	1266	2264
Total	2052	2806	4858

Source: CDCN/NCHS.

a What is the probability that a randomly selected male in this age range who died had a congenital problem leading to his death?

b What is the probability that a child aged 1 to 4 who died as a consequence of a homicide was a female?

c Find the conditional distribution of cause of death given that the child was a male.

d Find the conditional distribution of sex given that the death was accidental.

e Are the cause of death and sex independent? Justify your answer.

6.19 In Exercise 6.3, we considered the distribution of age and poverty. As we did there, let X be the random variable associated with age and Y be the random variable representing poverty status.

Age (in years)	**Live in Poverty**	**Do Not Live in Poverty**	**Totals**
Under 18	347	12,680	13,027
18 to 64	517	19,997	20,514
65 and older	130	3,327	3,457
Totals	994	36,004	36,998

a Find the conditional distribution of age given that a person is in poverty.

b Find the conditional distribution of poverty given that a person is a child (under 18).

c Are age and whether or not a person lives in poverty in the United States independent? Justify your answer.

6.20 Refer to the situation described in Exercise 6.6.

a Find the conditional density function of the amount of pollutant per sample when a certain cleaning device on the stack is not operating given the amount per sample when it is working.

b Sketch the conditional density function of the amount of pollutant per sample when a certain cleaning device on the stack is not operating given that the amount per sample is 0.8 when it is working.

c Sketch the conditional density function of the amount of pollutant per sample when the amount per sample is 0.4 when it is working.

d Find the conditional density function of the amount of pollutant per sample when a certain cleaning device on the stack is operating.

e Are X and Y independent? Justify your answer.

f Find $P(Y \leq 1/4 \,|X = 1)$.

6.21 Refer to the situation described in Exercise 6.5.

a Find the conditional density function of the X position of the particle given its Y position.

b Sketch the conditional density function of the X position of the particle given that its Y position is 0.6.

c Are X and Y independent? Justify your answer.

d Find $P(X \geq 0.8|Y = 0.4)$.

6.22 As in Exercise 6.8, a child prepares a mixture of peanut butter, bananas, and jam for his school lunch daily. Let X and Y denote the proportions of peanut butter and bananas, respectively. The mixture proportions vary from day to day according to the joint probability density given by

$$f(x, y) = \begin{cases} 50, & 0.4 \leq x \leq 0.6,\ 0.10 \leq y \leq 0.30,\ 0 \leq x + y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability that the mixture is more than 50% peanut butter given that bananas comprise 25% of the mixture.

6.23 As in Exercises 6.9 and 6.13, let X and Y denote the proportions of chemicals I and II, respectively, found in a sample mixture of chemicals used as an insecticide. Suppose that X and Y have a joint probability density given by

$$f(x, y) = \begin{cases} 2, & 0 \leq x \leq 1,\ 0 \leq y \leq 1,\ 0 \leq x + y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that chemical I comprises more than half of the mixture if a fourth of the mixture is chemical II.

6.24 As in Exercise 6.14, let X and Y denote the proportions of time out of one workweek that employees I and II, respectively, actually spend performing their assigned tasks. The joint relative frequency behavior of X and Y is modeled by the probability density function

$$f(x, y) = \begin{cases} x + y, & 0 \leq x \leq 1,\ 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability that employee I spends more than 75% of the workweek on his assigned task, given that employee II spends exactly 25% of the workweek on her assigned task.

6.25 A bus arrives at a bus stop at a randomly selected time within a 1-hour period. A passenger arrives at the bus stop at a randomly selected time within the same hour. The passenger is willing to wait for the bus for up to $1/4$ of an hour. What is the probability that the passenger will catch the bus? (*Hint:* Let X denote the bus arrival time, and let Y denote the passenger arrival time.) If these arrivals are independent, then

$$f(x, y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Now find $P(Y \le X \le Y + 1/4)$.

6.26 A particular fast-food outlet is interested in the joint behavior of the random variable X, which is defined as the total time, in minutes, between a customer's arrival at the fast-food outlet and his or her leaving the service window, and Y, which is the time that the customer waits in line before reaching the service window. Because X includes the time that a customer waits in line, we must have $X \ge Y$. The relative frequency distribution of observed values of X and Y can be modeled by the probability density function

$$f(x, y) = \begin{cases} e^{-x}, & \text{for } 0 \le y \le x < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

a Find $P(X < 2, Y > 1)$.

b Find $P(X \ge 2Y)$.

c Find the marginal density function for the total waiting time X.

d Find the conditional density function of total waiting time given the time spent waiting in line before reaching the service window.

6.27 Two friends are to meet at a library. Each arrives at an independently selected time within a fixed 1-hour period. Each agrees to wait not more than 10 minutes for the other. Find the probability that they will meet.

6.28 Two quality control inspectors each interrupt a production line at randomly (and independently) selected times within a given day (of 8 hours). Find the probability that the two interruptions will be more than 4 hours apart.

6.4 Expected Values of Functions of Random Variables

When we encounter problems that involve more than one random variable, we often combine the variables into a single function. We may be interested in the life lengths of five different electronic components within the same system or the difference between two strength test measurements on the same section of cable. We now discuss how to find **expected values** of functions of more than one random variable. Definition 6.8 gives the basic result for finding expected values in the bivariate case. The definition, of course, can be generalized to more variables.

DEFINITION **6.8**

Suppose that the discrete random variables (X, Y) have a joint probability function given by $p(x, y)$. If $g(X, Y)$ is any real-valued function of (X, Y), then the **expected value** of $g(X, Y)$ is

$$E[g(X, Y)] = \sum_x \sum_y g(x, y) p(x, y).$$

The sum is over all values of (x, y) for which $p(x, y) > 0$. If (X, Y) are continuous random variables with probability density function $f(x, y)$, then

$$E[g(X, Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x, y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y. \quad ■$$

If X and Y are independent, the expected value of XY is equal to the product of their expectation as stated more formally in Theorem 6.1.

THEOREM 6.1

If X and Y are independent with means μ_X and μ_Y, respectively, then

$$E(XY) = E(X)E(Y).$$

Further, if X and Y are independent, g is a function of X alone, and h is a function of Y alone, then

$$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)].$$

The proof is left as an exercise. ■

A function of two variables that is commonly of interest in probabilistic and statistical problems is the **covariance.**

DEFINITION 6.9

The **covariance** between two random variables X and Y is given by

$$\mathrm{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$$

where

$$\mu_X = E(X) \quad \text{and} \quad \mu_Y = E(Y). \quad ■$$

The covariance helps us assess the relationship between two variables in the following sense. If Y tends to be large when X is large, and if Y tends to be small when X is small, then X and Y have a positive covariance. If, on the other hand, Y tends to be small when X is large and large when X is small, then the two variables have a negative covariance. As with the variance of a random variable, the covariance may be computed more easily than by using its definition as shown in Theorem 6.2.

THEOREM 6.2

If X has a mean μ_X and Y has a mean of μ_Y, then

$$\mathrm{cov}(X, Y) = E(XY) - \mu_X\mu_Y.$$

The proof is left as an exercise. ■

The covariance measures the direction of the association between two random variables. However, the covariance depends on the unit of measurement. As an example, suppose the covariance is found to be 0.2meter^2. If we decided to report the results in centimeters instead of meters, the covariance would be 200cm^2. We need a measure that allows us to judge the strength of the association regardless of the units. **Correlation** is such a measure.

DEFINITION **6.10**

Definition 6.10. The **correlation** between two random variables X and Y is given by

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}.$$

ρ is often called the **correlation coefficient**. ▪

The correlation is a unitless quantity that takes on values between -1 and $+1$. If $\rho = +1$ or $\rho = -1$, then Y must be a linear function of X. For $\rho = +1$, the slope of the line is positive, and the slope is negative when $\rho = -1$. As ρ decreases from $+1$ or increases from -1, the strength of the linear relationship becomes less and is smallest at $\rho = 0$. As we shall see, this does not mean that there is no relationship between X and Y when $\rho = 0$; it does mean that the relationship is not linear.

Suppose that X and Y are independent random variables. Then

$$\begin{aligned}\text{cov}(XY) &= E(XY) - E(X)E(Y)\\ &= E(X)E(Y) - E(X)E(Y)\\ &= 0.\end{aligned}$$

That is, if X and Y are independent, their covariance is 0. The converse is not necessarily true; that is, zero covariance does not imply that the variables are independent.

To see that a zero covariance between two variables does not necessarily mean that these variables are independent, look at the joint distribution in Table 6.7. Clearly, $E(X) = E(Y) = 0$, and $E(XY) = 0$ as well. Therefore, cov $(X, Y) = 0$. On the other hand,

$$P(X = -1,\ Y = -1) = \frac{1}{8} \neq \frac{9}{64} = P(X = -1)\, P(Y = -1)$$

TABLE **6.7**
Joint Distribution of Dependent Variables with a Covariance of Zero.

		X: -1	0	1	Totals
Y	-1	1/8	1/8	1/8	3/8
	0	1/8	0	1/8	2/8
	1	1/8	1/8	1/8	3/8
Totals		3/8	2/8	3/8	1

so the random variables are dependent. Because the covariance between X and Y must be zero if the two variables are independent, but a zero covariance does not ensure independence, we say that a zero covariance is necessary but not sufficient for independence.

We shall now calculate the covariance in an example.

EXAMPLE 6.9 A travel agency keeps track of the number of customers who call and the number of trips booked on any one day. Let X denote the number of calls, let Y denote the number of trips booked, and let $p(x, y)$ denote the joint probability function for (X, Y). Records show the following:

$p(0,0) = 0.04$	$p(2,1) = 0.20$	$p(3,1) = 0.16$
$p(1,0) = 0.08$	$p(2,2) = 0.12$	$p(3,2) = 0.10$
$p(1,1) = 0.06$	$p(3,0) = 0.10$	$p(3,3) = 0.02$
$p(2,0) = 0.12$		

Thus, for any given day, the probability of, say, two calls and one order is 0.20. Find $\text{cov}(X, Y)$ and the correlation between X and Y.

Solution Theorem 6.2 suggests that we first find $E(X, Y)$, which is

$$\begin{aligned}
E(XY) &= \sum_x \sum_y xyp(x, y) \\
&= (0 \times 0)p(0,0) + (1 \times 0)p(1,0) + (1 \times 1)p(1,1) + (2 \times 0)p(2,0) \\
&\quad + (2 \times 1)p(2,1) + (2 \times 2)p(2,2) + (3 \times 0)p(3,0) \\
&\quad + (3 \times 1)p(3,1) + (3 \times 2)p(3,2) + (3 \times 3)p(3,3) \\
&= 0(0.4) + 0(0.08) + 1(0.06) + 0(0.12) + 2(0.20) + 4(0.12) \\
&\quad + 0(0.10) + 3(0.16) + 6(0.10) + 9(0.02) \\
&= 2.20.
\end{aligned}$$

Now we must find $E(X) = \mu_X$ and $E(Y) = \mu_Y$. The marginal distributions of X and Y are given on the following charts:

x	$p(x)$
0	0.04
1	0.14
2	0.44
3	0.28

y	$p(y)$
0	0.34
1	0.42
2	0.22
3	0.02

It follows that

$$\mu_X = 0(0.04) + 1(0.14) + 2(0.44) + 3(0.28) = 1.86$$

and

$$\mu_Y = 0(0.34) + 1(0.42) + 2(0.22) + 3(0.02) = 0.92.$$

Thus,

$$\begin{aligned}\operatorname{cov}(X, Y) &= E(XY) - \mu_X \mu_Y \\ &= 2.20 - 1.86(0.92) \\ &= 0.4888.\end{aligned}$$

From the marginal distributions of X and Y, it follows that

$$\begin{aligned}V(X) &= E\left(X^2\right) - \mu_X^2 \\ &= 4.42 - (1.86)^2 \\ &= 0.9604\end{aligned}$$

and

$$\begin{aligned}V(Y) &= E\left(Y^2\right) - \mu_Y^2 \\ &= 1.48 - (0.92)^2 \\ &= 0.6336.\end{aligned}$$

Hence,

$$\begin{aligned}\rho &= \frac{\operatorname{cov}(X, Y)}{\sqrt{V(X)V(Y)}} \\ &= \frac{0.4888}{\sqrt{(0.9604)(0.6336)}} \\ &= 0.6266.\end{aligned}$$

A moderate positive association exists between the number of calls and the number of trips booked. Do the positive covariance and the strength of the correlation agree with your intuition? ▪

The covariance or correlation between two discrete random variables should only be considered if both are interval variables. That is, the difference between two values must be meaningful for each variable; otherwise, other measures of association are more appropriate. As an example, consider the *Titanic* example again. We could consider the passenger classes as being ordered: first class > second class > third class > crew. First and second class are one class apart, and second and third class are one class apart. However, it is not clear that these two differences, although each represents a change of one class, are the same. Certainly, the difference between any of the passenger classes and the crew is difficult to quantify. Thus, we would not compute the correlation between passenger class and survival. In contrast, the number of calls and the number of trips booked are both interval variables. Clearly, the values of the random variables and the differences in values of these variables are meaningful (booking three trips results in two more trips booked than if one trip had been booked).

When a problem involves n random variables $X_1, X_2, \ldots, X_n$, we are often interested in studying the linear combinations of those variables. For example, if the random variables measure the quarterly incomes for n plants in a corporation, we may want to look at their sum or average. If X represents the monthly cost of servicing defective plants before a new quality control system was installed, and Y denotes that cost after the system went into operation, then we may want to study $X - Y$. Theorem 6.3 gives a general result for the mean and variance of a linear combination of random variables.

THEOREM 6.3 Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be random variables, with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_i$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Then the following relationships hold:

a $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$

b $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2 \sum\sum_{i<j} a_i a_j \operatorname{cov}(Y_i, Y_j)$ where the double sum is over all pairs (i, j) with $i < j$

c $\operatorname{cov}(U_1, U_2) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_i b_j \operatorname{cov}(Y_i, X_j)$

Proof

a The proof for part (a) follows directly from the definition of expected value and from properties of sums or integrals.

b To prove this part, we resort to the definition of variance and write

$$\begin{aligned} V(U_1) &= E[U_1 - E(U_1)]^2 \\ &= E\left[\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right]^2 \\ &= E\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]^2 \end{aligned}$$

$$= E\left[\sum_{i=1}^{n} a_i^2 (Y_i - \mu_i)^2 + \sum_{i \neq j}\sum a_i a_j E\left[(Y_i - \mu_i)\left(Y_j - \mu_j\right)\right]\right]$$

$$= \sum_{i=1}^{n} a_i^2 E (Y_i - \mu_i)^2 + \sum_{i \neq j}\sum a_i a_j E\left[(Y_i - \mu_i)\left(Y_j - \mu_j\right)\right].$$

By the definition of variance and covariance, we then have

$$V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + \sum_{i \neq j}\sum a_i a_j \text{cov}\left(Y_i, Y_j\right).$$

Notice that $\text{cov}\left(Y_i, Y_j\right) = \text{cov}\left(Y_j, Y_i\right)$; hence, we can write

$$V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2\sum_{i < j}\sum a_i a_j \text{cov}\left(Y_i, Y_j\right).$$

c The proof of this part is obtained by similar steps. We have

$$\begin{aligned}
\text{cov}(U_1,\ U_2) &= E\{[U_1 - E(U_1)][U_2 - E(U_2)]\} \\
&= E\left[\left(\sum_{i=1}^{n} a_i Y_i - \sum_{i=1}^{n} a_i \mu_i\right)\left(\sum_{j=1}^{n} b_j X_j - \sum_{j=1}^{n} b_j \xi_j\right)\right] \\
&= E\left\{\left[\sum_{i=1}^{n} a_i (Y_i - \mu_i)\right]\left[\sum_{j=1}^{n} b_j (X_j - \xi_j)\right]\right\} \\
&= E\left[\sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j E[(Y_i - \mu_i)(X_j - \xi_j)]\right].
\end{aligned}$$

On observing that $\text{cov}(Y_i, Y_i) = V(Y_i)$, we see that part (b) is a special case of part (c). ▪

The proof of Theorem 6.3 is shown for discrete random variables. However, the same results hold for continuous random variables. The proofs for the continuous case are left as exercises. The next example illustrates the application of this theorem for two continuous random variables.

EXAMPLE 6.10 As in Examples 6.5 and 6.7, a soft-drink machine has a random supply Y at the beginning of a given day and dispenses a random amount X during the day (with measurements in gallons). The machine is not resupplied during the day. It has been observed that X and Y have joint density

$$f(x, y) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq y, \quad 0 \leq y \leq 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Then $Z = Y - X$ represents the amount of soft drink left over at the end of the day. Find the mean and the variance of Z.

Solution In Example 6.5, we found the marginal densities of X and Y to be

$$f_X(x) = \begin{cases} \frac{2-x}{2}, & 0 \leq x \leq 2 \\ 0, & \text{otherwise} \end{cases}$$

and

$$f_Y(y) = \begin{cases} \left(\frac{1}{2}\right) y, & 0 \leq y \leq 2 \\ 0, & \text{otherwise.} \end{cases}$$

We must first find the means and the variances of X and Y. Thus,

$$\begin{aligned} E(X) &= \int_0^2 x\left(\frac{2-x}{2}\right) \mathrm{d}x \\ &= \left[\frac{x^2}{2} - \frac{x^3}{6}\right]_0^2 \\ &= 2 - \frac{8}{6} \\ &= \frac{2}{3} \end{aligned}$$

and

$$\begin{aligned} E(Y) &= \int_0^2 (y)\frac{1}{2} y \,\mathrm{d}y \\ &= \frac{1}{2}\left[\frac{y^3}{3}\right]_0^2 \\ &= \frac{4}{3}. \end{aligned}$$

By similar arguments, it follows that

$$E(X^2) = \frac{2}{3}$$

$$
\begin{aligned}
V(X) &= \frac{2}{3} - \left(\frac{2}{3}\right)^2 \\
&= \frac{2}{9} \\
E(Y^2) &= 2
\end{aligned}
$$

and

$$
\begin{aligned}
V(Y) &= 2 - \left(\frac{4}{3}\right)^2 \\
&= \frac{2}{9}.
\end{aligned}
$$

The next step is to find cov(X, Y). Now,

$$
\begin{aligned}
E(XY) &= \int_0^2 \int_0^y (xy)\frac{1}{2}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^2 \left[\frac{x^2}{4}\right]_0^y y\,\mathrm{d}y \\
&= \int_0^2 \frac{y^3}{4}\,\mathrm{d}y \\
&= \left[\frac{y^4}{16}\right]_0^2 \\
&= 1
\end{aligned}
$$

and

$$
\begin{aligned}
\operatorname{cov}(X, Y) &= E(XY) - \mu_X \mu_Y \\
&= 1 - \left(\frac{2}{3}\right)\left(\frac{4}{3}\right) \\
&= \frac{1}{9}.
\end{aligned}
$$

From Theorem 6.2,

$$
\begin{aligned}
E(Z) &= E(Y) - E(X) \\
&= \frac{4}{3} - \frac{2}{3} \\
&= \frac{2}{3}
\end{aligned}
$$

and

$$\begin{aligned}
V(Z) &= V(Y - X) \\
&= V[(1)Y + (-1)X] \\
&= (1)^2 V(Y) + (-1)^2 V(X) + 2(1)(-1)\text{cov}(X, Y) \\
&= \frac{2}{9} + \frac{2}{9} - 2\left(\frac{1}{9}\right) \\
&= \frac{2}{9}.
\end{aligned}$$

The mean number of gallons left at the end of a randomly selected day is (2/3) gallons, and the variance of the number of gallons left is (10/9)gallons^2. ▪

When the random variables in use are independent, calculating the variance of a linear function simplifies because the covariances are zero. This simplification is illustrated in Example 6.11.

EXAMPLE 6.11 Let $X_1, X_2, \ldots, X_n$ be independent random variables, with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. (These variables may denote the outcomes of n independent trials of an experiment.) Defining

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

show that $E(\bar{X}) = \mu$ and $V(\bar{X}) = \frac{\sigma^2}{n}$.

Solution Notice that $\bar{X}$ is a linear function with all constants a_i equal to $1/n$; that is,

$$\bar{X} = \left(\frac{1}{n}\right)X_1 + \left(\frac{1}{n}\right)X_2 + \cdots + \left(\frac{1}{n}\right)X_n.$$

By Theorem 6.3 part (a),

$$\begin{aligned}
E(\bar{X}) &= \sum_{i=1}^{n} a_i\mu \\
&= \mu\sum_{i=1}^{n} a_i \\
&= \mu\sum_{i=1}^{n}\frac{1}{n} \\
&= \frac{n\mu}{n} \\
&= \mu.
\end{aligned}$$

By Theorem 6.3 part (b),

$$V\left(\bar{X}\right) = \sum_{i=1}^{n} a_i^2 V\left(X_i\right) + 2\sum_{i<j}\sum a_i a_j \operatorname{cov}\left(X_i, X_j\right).$$

However, the covariance terms are all zero, because the random variables are independent. Thus,

$$V\left(\bar{X}\right) = \sum_{i=1}^{n}\left(\frac{1}{n}\right)^2 \sigma^2 = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}. \quad \blacksquare$$

EXAMPLE 6.12 A firm purchases two types of industrial chemicals. The amount of type I chemical purchased per week, X, has $E(X) = 40$ gallons with $V(X) = 4$. The amount of type II chemical purchased, Y, has $E(Y) = 65$ gallons with $V(Y) = 8$. Type I costs \$3 per gallon, whereas type II costs \$5 per gallon. Find the mean and the standard deviation of the total weekly amount spent on these types of chemicals, assuming that X and Y are independent.

Solution The dollar amount spent each week is given by

$$Z = 3X + 5Y.$$

From Theorem 6.3,

$$\begin{aligned} E(Z) &= 3E(X) + 5E(Y) \\ &= 3(40) + 5(65) \\ &= 445 \end{aligned}$$

and

$$\begin{aligned} V(Z) &= (3)^2 V(X) + (5)^2 V(Y) \\ &= 9(4) + 25(8) \\ &= 236. \end{aligned}$$

The firm can expect to spend \$445 per week in chemicals and the standard deviation of cost is $\sqrt{V(Z)} = \sqrt{236} = \$15.36$. $\blacksquare$

EXAMPLE 6.13 Sampling problems involving finite populations can be modeled by selecting balls from urns. Suppose that an urn contains r white balls and $(N - r)$ black balls. A random sample of n balls is drawn without replacement, and Y, the number of white balls in the sample, is observed. From Chapter 4, we know that Y has a hypergeometric probability distribution. Find the mean and the variance of Y.

Solution We first observe some characteristics of sampling without replacement. Suppose that the sampling is done sequentially, and we observe outcomes from $X_1, X_2, \dots, X_n$, where

$$X_i = \begin{cases} 1, & \text{if the } i\text{th draw results in a white ball} \\ 0, & \text{otherwise.} \end{cases}$$

Unquestionably, $P(X_1) = r/N$. But, it is also true that $P(X_2) = r/N$, because

$$\begin{aligned} P(X_2 = 1) &= P(X_1 = 1, X_2 = 1) + P(X_1 = 0, X_2 = 1) \\ &= P(X_1 = 1)P(X_2 = 1 \mid X_1 = 1) + P(X_1 = 0)P(X_2 = 1 \mid X_1 = 0) \\ &= \frac{r}{N}\left(\frac{r-1}{N-1}\right) + \frac{N-r}{N}\left(\frac{r}{N-1}\right) \\ &= \frac{r(N-1)}{N(N-1)} \\ &= \frac{r}{N}. \end{aligned}$$

The same is true for X_k; that is,

$$P(X_k = 1) = \frac{r}{N}, \qquad k = 1, 2, \dots, n.$$

Thus, the probability of drawing a white ball on any draw, given no knowledge of the outcomes of previous draws, is r/N.

In a similar way, it can be shown that

$$P(X_j = 1, X_k = 1) = \frac{r(r-1)}{N(N-1)}, \qquad j \neq k.$$

Now observe that $Y = \sum_{i=1}^{n} X_i$; hence,

$$E(Y) = \sum_{i=1}^{n} E(X_i) = n\left(\frac{r}{N}\right).$$

To find $V(Y)$, we need $V(X_i)$ and $\text{cov}(X_i, X_j)$. Because X_i is 1 with probability r/N and 0 with probability $1 - r/N$, it follows that

$$V(X_i) = \frac{r}{N}\left(1 - \frac{r}{N}\right).$$

Also,

$$\begin{aligned}\operatorname{cov}(X_i,\ X_j) &= E(X_iX_j) - E(X_i)E(X_j)\\ &= \frac{r(r-1)}{N(N-1)} - \left(\frac{r}{N}\right)^2\\ &= -\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{1}{N-1}.\end{aligned}$$

Because $X_iX_j = 1$ if and only if $X_i = 1$ and $X_j = 1$. From Theorem 6.3, we know that

$$\begin{aligned}V(Y) &= \sum_{i=1}^{n} V(X_i) + 2\sum\sum_{i<j}\operatorname{cov}(X_i, X_j)\\ &= n\frac{r}{N}\left(1-\frac{r}{N}\right) + 2\sum\sum_{i<j}\left[-\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{1}{N-1}\right]\\ &= n\frac{r}{N}\left(1-\frac{r}{N}\right) - n(n-1)\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{1}{N-1}\end{aligned}$$

because there are $n(n-1)/2$ terms in the double summation. A little algebra yields

$$V(Y) = n\frac{r}{N}\left(1-\frac{r}{N}\right)\frac{N-n}{N-1}. \quad \blacksquare$$

The usefulness of Theorem 6.3 becomes especially clear when one tries to find the expected value and the variance for the hypergeometric random variable. Without it, by proceeding directly from the definition of an expectation, the necessary summations are exceedingly difficult to obtain.

Exercises

6.29 For each pair of random variables, indicate whether or not it is appropriate to compute the correlation.

- **a** Sex of a student (female, male) and major (business, education, mathematics, science)
- **b** Type of ice cream (brand A, brand B, brand C) and quality (poor, good)
- **c** View on death penalty (favor, oppose) and view on marijuana (favor, oppose)
- **d** Age (1 year, 2 years, 3 years, . . .) and height (18 inches, 19 inches, 20 inches, . . .)

6.30 For each pair of random variables, indicate whether or not it is appropriate to compute the correlation.

- **a** Time spent studying each week (1 hour, 2 hours, . . ., 10 hours) and grade (A, B, C, D, F)
- **b** Educational attainment (grade school, high school, some college, college and postgraduate training) and salary ($<$ \$20,000, \$20,000 to $<$ \$50,000, \$50,000 to $<$ \$100,000, at least \$100,000)

c Birth order (first child, second child, . . .) and level of extroversion (high, medium low)

d Weight and time to complete a 100-meter dash

6.31 Bollworms (*Heliothis* spp.) are a pest of cotton that cause economic damage if uncontrolled. Lady beetles (*Hippodamia convergens*) are natural pests of bollworms. If sufficient numbers of lady beetles are present in a particular cotton field, then chemical control may not be needed. Let X and Y denote the numbers of bollworms and lady beetles on a cotton plant, respectively. A researcher explored the relationship in these two variables in a cotton field and found that they had the following joint distribution:

$y \setminus x$	0	1	2	3
0	0.03	0.04	0.11	0.10
1	0.04	0.05	0.09	0.06
2	0.09	0.07	0.06	0.02
3	0.12	0.10	0.01	0.01

a Find the mean number of bollworms on the plants in this cotton field.

b Find the mean number of lady beetles on the plants in this cotton field.

c Find the covariance between the numbers of lady beetles and bollworms on a cotton plant.

d Find the correlation between the numbers of lady beetles and bollworms on a cotton plant.

6.32 In a particular state, a study was conducted to determine the relationship between X, the size of a family, and Y, the total number of pets in that family. The joint distribution of X and Y is displayed in the table that follows.

$y \setminus x$	2	3	4	5	6
0	0.12	0.12	0.10	0.02	0.01
1	0.03	0.05	0.08	0.10	0.04
2	0.02	0.03	0.06	0.07	0.05

a Find the mean number of pets in a family.

b Find the mean family size.

c Find the covariance between the number of pets in a family and the size of that family.

d Find the correlation between the number of pets in a family and the size of that family.

6.33 As in Exercises 6.5, 6.11, and 6.21, a radioactive particle is randomly located in a square area with sides that are 1 unit in length. Let X and Y denote the coordinates of the particle. Because the particle is equally likely to fall in any subarea of fixed size, a reasonable model of (X, Y) is given by

$$f(x, y) = \begin{cases} 1, & 0 \le x \le 1,\ 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the mean of X.

b Find the mean of Y.

c Find the covariance between X and Y.

d Find the correlation between X and Y.

6.34 As in Exercises 6.6, 6.10, and 6.20, in a study of the particulate pollution in air samples over a smokestack, X represents the amount of pollutants per sample when a cleaning device is not operating, and Y represents the amount per sample when the cleaning device is operating. Assume that (X, Y) has a joint probability density function

$$f(x, y) = \begin{cases} 1, & \text{for } 0 \le x \le 2; 0 \le y \le 1; 2y \le x \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the mean and the variance of pollutants per sample when the cleaning device is not in place.

b Find the mean and the variance of pollutants per sample when the cleaning device is in place.

6.35 Refer to Exercises 6.9, 6.13, and 6.23. The proportions X and Y of two chemicals found in samples in an insecticide have the joint probability density function

$$f(x, y) = \begin{cases} 2, & \text{for } 0 \le x \le 1; 0 \le y \le 1; 0 \le x + y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

The random variable $Z = X + Y$ denotes the proportion of the insecticides due to the combination of both chemicals.

a Find the mean and the variance of the proportion of the first chemical (X) used in the insecticide.

b Find the mean and the variance of the proportion of the insecticides due to the combination of both chemicals.

c Find an interval within which the values of Z should lie for at least 50% of the samples of insecticide.

6.36 Refer to Exercise 6.34. The random variable $Z = X - Y$ represents the amount by which the weight of emitted pollutant can be reduced by using the cleaning device.

a Find the mean and the variance of Z.

b Find an interval within which the values of Z should lie at least 75% of the time.

6.37 Refer to Exercise 6.35.

a Find the covariance between the proportions of the two chemicals (X and Y) used in the insecticide.

b Find the correlation between the proportions of the two chemicals (X and Y) used in the insecticides.

6.38 For a sheet-metal stamping machine in a certain factory, the time between failures, X, has a mean time between failure (MTBF) of 56 hours and a variance of 16 hours. The repair time, Y, has a mean time to repair (MTTR) of 5 hours and a variance of 4 hours. If X and Y are independent, find the expected value and the variance of $Z = X + Y$, which represents one operation-repair cycle.

6.39 Let X denote the amount of gasoline stocked in a bulk tank at the beginning of a week, and let Y denote the amount sold during the week; then $Z = X - Y$ represents the amount left over at the end of the week. Find the mean and the variance of Z, assuming that the joint density function of (X, Y) is given by

$$f(x, y) = \begin{cases} 3x, & \text{for } 0 \le y \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

6.40 Consider again the fast-food outlet discussed in Exercise 6.26. Again, let X be the total time in minutes between a customer's arrival at the fast-food outlet and his or her leaving the service window, and Y be the time that the customer waits in line before reaching the service window. The relative frequency distribution of observed values of X and Y can be modeled by the probability density function

$$f(x, y) = \begin{cases} e^{-x}, & \text{for } 0 \le y \le x < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

a Find the probability that the time spent at the service window $(X - Y)$ is more than a minute.

b Find the mean time spent at the service window.

c Find the standard deviation of the time spent at the service window.

d Is it highly likely that a customer will spend more than 2 minutes at the service window?

6.41 If X has a mean of μ_X and Y has a mean of μ_Y, prove that $\text{cov}(X, Y) = E(XY) - \mu_X \mu_Y$.

6.42 Refer again to Exercise 6.40.

a If a randomly observed customer's total waiting time for service is known to have been more than 2 minutes, find the probability that the customer waited less than 1 minute to be served.

b Suppose that a customer spends a length of time y_1 at the fast-food outlet. Find the probability that this customer spends less than half of that time at the service window.

6.43 Let X and Y be independent discrete random variables with means μ_X and μ_Y, respectively. Given that g is a function of X alone and h is a function of Y alone, prove that

$$E[g(X)h(Y)] = E[g(X)]\,E[h(Y)].$$

6.44 Let $Y_1, Y_2, \ldots, Y_n$ and $X_1, X_2, \ldots, X_m$ be continuous random variables, with $E(Y_i) = \mu_i$ and $E(X_i) = \xi_i$. Define

$$U_1 = \sum_{i=1}^{n} a_i Y_i \quad \text{and} \quad U_2 = \sum_{j=1}^{m} b_j X_j$$

for constants $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_m$. Prove that the following relationships hold.

a $E(U_1) = \sum_{i=1}^{n} a_i \mu_i$

b $V(U_1) = \sum_{i=1}^{n} a_i^2 V(Y_i) + 2\sum\sum_{i<j} a_i a_j \operatorname{cov}(Y_i, Y_j)$ where the double sum is over all pairs (i, j) with $i < j$

c $\operatorname{cov}(U_1, U_2) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_j \operatorname{cov}(Y_i, X_j)$

6.5 Conditional Expectations

Section 6.2 contains a discussion of conditional probability functions and conditional density functions, which we shall now relate to **conditional expectations.** Conditional expectations are defined in the same manner as univariate expectations, except that the conditional density function is used in place of the marginal density function.

DEFINITION **6.11**

If X and Y are any two random variables, the **conditional expectation** of X given that $Y = y$ is defined to be

$$E(X|Y = y) = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, dx$$

if X and Y are jointly continuous, and

$$E(X|Y = y) = \sum_{x} x p(x|y)$$

if X and Y are jointly discrete. ▪

EXAMPLE 6.14 Refer to Example 6.7, with X denoting the amount of soft drink sold, Y denoting the amount in supply, and

$$f(x, y) = \begin{cases} \frac{1}{2}, & 0 \le x \le y; 0 \le y \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the conditional expectation of amount of sales X given that $Y = 1$.

Solution In Example 6.7, we found that

$$f(x \mid y) = \begin{cases} \frac{1}{y}, & 0 \le x \le y \le 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Because we know that $Y = 1$, the conditional density is

$$f_{X|Y}(x \mid y) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Then, from Definition 6.11,

$$\begin{aligned} E(X|Y = 1) &= \int_{-\infty}^{\infty} x f_{X|Y}\left(x|y\right) \mathrm{d}x \\ &= \int_{0}^{1} x(1) \, \mathrm{d}x \\ &= \left. \frac{x^2}{2} \right|_0^1 \\ &= \frac{1}{2}. \end{aligned}$$

It follows that, if the soft-drink machine contains 1 gallon at the start of the day, the expected amount that will be sold that day is 1/2 gallon.

It is important to observe that we could solve this problem more generally. That is, given that the amount of supply for a given day is $Y = Y$, what is the expected amount that will be sold that day?

$$\begin{aligned} E\left(X|Y = y\right) &= \int_{-\infty}^{\infty} x f_{X|Y}\left(x|y\right) \mathrm{d}x \\ &= \int_{0}^{y} x \left(\frac{1}{y}\right) \mathrm{d}x \end{aligned}$$

$$= \left. \frac{x^2}{2y} \right|_0^y$$
$$= \frac{y}{2}, \qquad 0 \le y \le 2.$$

Notice that, if $Y = 1$, meaning that the soft-drink machine contains 1 gallon at the start of the day, the expected amount sold that day is 1/2 gallon, as found earlier. However, we can also easily determine that if the soft-drink machine contains the full 2 gallons at the start of the day, the expected amount sold is $2/2 = 1$ gallon. Similarly, we can determine the expected amount sold given any initial amount in the machine.

Suppose we kept records over many days of the gallons at the start of the day and the amount sold. We could plot the amount sold against the amount available at the start of the day. The points of the scatter plot should lie roughly along a line. If we fit a least squares regression line to these points, the intercept should be 0, and the slope should be about 1/2. In fact, we would be estimating $E\left(X|Y = y\right) = \frac{y}{2}$. Of course, here we know enough to obtain the exact relationship in X and Y. In real life, this is often not the case, and regression is used to obtain estimates of the very things we have derived exactly here. ▪

In Example 6.14, we found the conditional expectation of X, given $Y = Y$, obtaining the conditional expectation as being a function of the random variable Y. Hence, we can find the expected value of the conditional expectation. The results of this type of iterated expectation are given in Theorem 6.4.

THEOREM 6.4 Let X and Y denote random variables. Then

$$E(X) = E\left[E\left(X|Y\right)\right]$$

where on the right-hand side, the inside expectation is stated with respect to the conditional distribution of X given Y, and the outside expectation is stated with respect to the distribution of Y.

Proof Let X and Y have the joint density function $f(x, y)$ and marginal densities $f_X(x)$ and $f_Y(y)$, respectively. Then

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$$
$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x, y)\,\mathrm{d}x\,\mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{X|Y}(x|y) f_Y(y)\, \mathrm{d}x\, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} x f_{X|Y}(x|y)\, \mathrm{d}x \right] f_Y\left(y\right)\, \mathrm{d}y$$

$$= \int_{-\infty}^{\infty} E(X|Y = y) f_Y(y)\, \mathrm{d}y$$

$$= E\left[E\left(X|Y\right)\right].$$

The proof is similar for the discrete case. ▪

The proof of Theorem 6.4 assumes that X and Y are continuous random variables. However, the theorem is also valid for the discrete random variables X and Y. The proof of the discrete case is left as an exercise. Conditional expectations can also be used to find the variance of a random variable in the spirit of Theorem 6.3. The main result is given in Theorem 6.5.

THEOREM 6.5 For random variables X and Y,

$$V(X) = E\left[V\left(X|Y\right)\right] + V\left[E\left(X|Y\right)\right].$$

Proof We begin with $E\left[V\left(X|Y\right)\right]$ and write

$$\begin{aligned} E\left[V\left(X|Y\right)\right] &= E\left[E\left(X^2|Y\right) - \left[E\left(X|Y\right)\right]^2\right] \\ &= E\left[E\left(X^2|Y\right)\right] - E\left[E\left(X|Y\right)\right]^2 \\ &= E\left[X^2\right] - E\left[E\left(X|Y\right)\right]^2 \\ &= E\left[X^2\right] - \left[E\left(X\right)\right]^2 - E\left[E\left(X|Y\right)\right]^2 + \left[E\left(X\right)\right]^2 \\ &= V\left(X\right) - E\left[\left[E\left(X|Y\right)\right]^2\right] + \left[E\left[E\left(X|Y\right)\right]^2\right] \\ &= V\left(X\right) - V\left[E\left(X|Y\right)\right]. \end{aligned}$$

Rearranging terms, we have

$$V(X) = E\left[V\left(X|Y\right)\right] + V\left[E\left(X|Y\right)\right], \quad ▪$$

EXAMPLE **6.15** A quality control plan for an assembly line involves sampling n finished items per day and counting X, the number of defective items. If p denotes the probability of observing a defective item, then X has a binomial distribution when the number of items produced by the line is large. However, p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $1/4$.

1 Find the expected value of X for any given day.

2 Find the standard deviation of X.

3 Given $n = 10$ items are sampled, find the expected value and the standard deviation of the number of defectives for that day.

Solution **1** From Theorem 6.4, we know that

$$E(X) = E\left[E(X|p)\right].$$

For a given p, X has a binomial distribution; hence,

$$E(X|p) = np.$$

Thus,

$$\begin{aligned} E(X) &= E(np) \\ &= nE(p) \\ &= n\int_0^{1/4} 4p\,\mathrm{d}p \\ &= n\left(\frac{1}{8}\right). \end{aligned}$$

And for $n = 10$,

$$E(X) = \frac{10}{8} = \frac{5}{4}.$$

2 From Theorem 6.5,

$$\begin{aligned} V(X) &= E[V(X|p)] + V[E(X|p)] \\ &= E[np(1-p)] + V[np] \\ &= nE(p) - nE(p^2) + n^2V(p) \\ &= nE(p) - n\{[V(p)] + [E(p)]^2\} + n^2V(p) \\ &= nE(p)[1 - E(p)] + n(n-1)V(p). \end{aligned}$$

The standard deviation, of course, is simply

$$\sqrt{V(X)} = \sqrt{nE(p)[1 - E(p)] + n(n-1)V(p)}$$

Using the properties of the uniform distribution, we have that $E(p) = 1/8$ and $V(p) = \frac{1}{(4)^2(12)}$. Thus, for $n = 10$,

$$V(X) = 10\left(\frac{1}{8}\right)\left(\frac{9}{8}\right) + 10(9)\frac{1}{(4)^2(12)}$$
$$= 1.56$$

and

$$\sqrt{V(X)} = 1.25.$$

3 In the long run, if $n = 10$ items are inspected each day, this inspection policy should discover, on average, $5/4$ defective items per day with a standard deviation of 1.25 items. Note: The calculations could be checked by actually finding the unconditional distribution of X and computing $E(X)$ and $V(X)$ directly. ▪

Exercises

6.45 As in Exercises 6.6, 6.10, 6.20, and 6.34, an environmental engineer measures the amount (by weight) of particulate pollution in air samples (of a certain volume) collected over the smokestack of a coal-fueled power plant. Let X denote the amount of pollutant per sample when a certain cleaning device on the stack is not operating, and let Y denote the amount of pollutants per sample when the cleaning device is operating under similar environmental conditions. The relative frequency of (X, Y) can be modeled by

$$f(x, y) = \begin{cases} 1, & 0 \leq x \leq 2, 0 \leq y \leq 1, 2y \leq x \\ 0, & \text{elsewhere.} \end{cases}$$

a If the amount of pollutant is 0.5 when the cleaning device is operating, what is the mean of the pollutant emitted when the cleaning device is not operating?

b If the amount of pollutant is 0.5 when the cleaning device is operating, what is the variance of the pollutant emitted when the cleaning device is not operating?

6.46 As in Exercises 6.14 and 6.24, let X and Y denote the proportions of time out of one workweek that employees I and II, respectively, actually spend performing their assigned tasks. The joint relative frequency behavior of X and Y is modeled by the probability density function

$$f(x, y) = \begin{cases} x + y, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a If employee I works half of the time, what is the mean time that employee II works?

b If employee I works half of the time, what is the variance of the time that employee II works?

6.47 Refer to exercise 6.45.

a If the amount of pollutant is 0.5 when the cleaning device is not operating, what is the mean of the pollutant emitted when the cleaning device is operating?

b If the amount of pollutant is 0.5 when the cleaning device is not operating, what is the variance of the pollutant emitted when the cleaning device is operating?

6.48 As in Exercises 6.26, 6.40, and 6.42, for a particular fast-food outlet, let X be defined as the total time in minutes between a customer's arrival at the fast-food outlet and his or her leaving the service window, and Y be the time that the customer waits in line before reaching the service window. The relative frequency distribution of observed values of X and Y can be modeled by the probability density function

$$f(x, y) = \begin{cases} e^{-x}, & \text{for } 0 \le y \le x < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

a If a randomly observed customer's total time between arrival and leaving the service window is known to have been 2 minutes, find the probability that the customer waited less than 1 minute to be served.

b Suppose that a customer spends a length of time x in the fast-food outlet. Find the probability that this customer spends less than half of that time at the service window.

6.49 As in Exercises 6.9, 6.13, 6.23, and 6.35, the proportions X and Y of two chemicals found in samples in an insecticide have the joint probability density function

$$f(x, y) = \begin{cases} 2, & \text{for } 0 \le x \le 1;\ 0 \le y \le 1;\ 0 \le x + y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

a If 25% of the insecticide is chemical X, find the probability that less than 25% of the insecticide is chemical Y.

b Find the mean proportion of chemical Y in the insecticide if 25% of the insecticide is chemical X.

c Find the standard deviation of the proportion of chemical Y in the insecticide if 25% of the insecticide is chemical X.

6.50 Suppose that X is the random variable denoting the number of bacteria per cubic centimeter in water samples and that for a given location, X has a Poisson distribution with mean λ. But λ varies from location to location and has a gamma distribution with parameters α and β. Find the expressions for $E(X)$ and $V(X)$.

6.51 The number of eggs N found in nests of a certain species of turtles has a Poisson distribution with mean λ. Each egg has probability p of being viable, and this event is independent from egg to egg. Find the mean and the variance of the number of viable eggs per nest.

6.52 A particular road construction project has restricted passage on the road to one lane, causing motorists to sometimes wait for the traffic to move in the direction they are going. The time that a motorist spends waiting is exponentially distributed with a mean of λ, where λ is normally distributed with a mean of μ and a standard deviation of σ. Find the mean and the variance of the time a motorist spends waiting.

6.53 Let X and Y denote discrete random variables. Prove that

$$E(X) = E[E(X|Y)]$$

where on the right-hand side, the inside expectation is stated with respect to the conditional distribution of X given Y, and the outside expectation is stated with respect to the distribution of Y.

6.54 Let X and Y be independent, identically distributed binomial random variables with parameters n and p. Find the conditional expectation of X given $X + Y = m$.

6.55 A hiker is lost in a cave, and she simply wants to get out of the cave. Three passages are open to her. The first passage leads to the outside after 1 hour of travel. The second passage winds around and will return her to the same place after 3 hours of travel. The third and final passage also winds around and will return her to the same place after 3 hours of travel. If the hiker is at all times equally likely to choose any one of the passages, what is the expected length of time until she reaches the outside?

6.56 A quality control plan for an assembly line involves sampling n finished items per day and counting X, the number of defective items. If p denotes the probability of observing a defective item, then X has a binomial distribution when the number of items produced by the line is large. However, p varies from day to day and is assumed to have a uniform distribution on the interval from 0 to $1/4$.

- **a** Find the joint distribution of X and p.
- **b** Find the marginal distribution of X.
- **c** Find the mean of X using the marginal distribution found in part (b).
- **d** Find the variance of X using the marginal distribution found in part (b).

6.57 Suppose Joe is seeking a job. He decides to consider n possible jobs. However, after being offered a job, he must immediately decide whether to accept it or decline it and consider the next job that becomes available. Notice that when considering a new job offer, Joe only knows the relative rank of that job compared to the ones he has already declined. Assume that all $n!$ orderings are equally likely and that once a job is declined it is gone forever. Suppose Joe's strategy is to reject the first k jobs and then accept the first one that is better than all of those first k.

- **a** For a given value of k, what is the probability that Joe selects the best job?
- **b** What value of k would maximize the probability that Joe selects the best job under this strategy?

6.58 The game of craps was described in Exercise 3.40. Let X represent the number of rolls of the dice in a game of craps.

- **a** Find $E(X|\text{player wins})$.
- **b** Find $E(X|\text{player loses})$.
- **c** Find $E(X)$.

6.6 The Multinomial Distribution

Suppose that an experiment consists of n independent trials, much like the binomial case, except that each trial can result in any one of k possible outcomes. For example, a customer checking out of a grocery store may choose any one of k checkout counters. Now suppose that the probability that a particular trial results in outcome i is denoted by p_i, where $i = 1, 2, \ldots, k$ denote the number of the n trials that result in outcome i. In developing a formula for $P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k)$, we first observe that because of independence of trials, the probability of having x_1 outcomes of type 1 through x_k outcomes of type k in a particular order is

$$p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}.$$

The number of such orderings equals the number of ways to partition the n trials into x_1 type 1 outcomes, x_2 type 2 outcomes, and so on through x_k type k outcomes, or

$$\frac{n!}{x_1!\,x_2!\cdots x_k!}$$

where

$$\sum_{i=1}^{k} x_i = n.$$

Hence,

$$P\,(X_1 = x_1,\ X_2 = x_2,\ \ldots,\ X_k = x_k) = \frac{n!}{x_1!\,x_2!\ \cdots\ x_k!} p_1^{x_1} p_2^{x_2}\ \cdots\ p_k^{x_k},$$

This is called the multinomial probability distribution. Notice that for $k = 2$, we are back into the binomial case, because X_2 is then equal to $n - X_1$. The following example illustrates the computations involved.

EXAMPLE 6.16 Items under inspection are subject to two types of defects. About 70% of the items in a large lot are judged to be free of defects, another 20% have only a type A defect, and the final 10% have only a type B defect (none has both types of defect). If six of these items are randomly selected from the lot, find the probability that three have no defects, one has a type A defect, and two have type B defects.

Solution If we can legitimately assume that the outcomes are independent from trial to trial (that is, from item to item in our sample), which they usually would be in a large lot, then the multinomial distribution provides a useful model. Letting X_1, X_2, and X_3 denote the number of trials resulting in zero, type A, and type B defectives, respectively, we have $p_1 = 0.7$, $p_2 = 0.2$, and $p_3 = 0.1$. It follows that

$$\begin{aligned} P\,(X_1 = 3, X_2 = 1, X_3 = 2) &= \frac{6!}{3!\,1!\,2!}(0.7)^3(0.2)(0.1)^2 \\ &= 0.041. \quad \blacksquare \end{aligned}$$

EXAMPLE 6.17 Find $E(X_i)$ and $V(X_i)$ for the multinomial probability distribution.

Solution We are concerned with the marginal distribution of X_i, the number of trials that fall in cell i. Imagine that all the cells, excluding cell i, are combined into a single large cell. Hence, every trial will result either in cell i or not in cell i, with probabilities p_i and $1 - p_i$, respectively, and X_i possesses a binomial marginal probability distribution. Consequently,

$$E(X_i) = np_i$$

$$V(X_i) = np_iq_i$$

where

$$q_i = 1 - p_i.$$

The same results can be obtained by setting up the expectations and evaluating. For example,

$$E(X_i) = \sum_{x_1}\sum_{x_2}\cdots\sum_{x_k} x_1 \frac{n!}{x_1!\,x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2}\cdots p_k^{x_k}.$$

Because we have already derived the expected value and the variance of X_i, the tedious summation of this expectation is left to the interested reader. ▪

EXAMPLE 6.18 If $X_1, X_2, \ldots, X_k$ have the multinomial distribution given in Example 6.16, find $\text{cov}(X_s, X_t)$, with $s \neq t$.

Solution Thinking of the multinomial experiment as a sequence of n independent trials, we define

$$U_i = \begin{cases} 1, & \text{if trial } i \text{ results in class } s \\ 0, & \text{otherwise} \end{cases}$$

and

$$W_i = \begin{cases} 1, & \text{if trial } i \text{ results in class } t \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$X_s = \sum_{i=1}^{n} U_i$$

and

$$X_t = \sum_{j=1}^{n} W_j.$$

To evaluate $\text{cov}(X_s, X_t)$, we need the following results:

$$E(U_i) = p_s$$

$$E(W_j) = p_t$$

$$\text{cov}(U_i, W_j) = 0$$

if $i \neq j$, because the trials are independent; and

$$\begin{aligned} \text{cov}(U_i, W_j) &= E(U_i W_j) - E(U_i)E(W_j) \\ &= 0 - p_s p_t \end{aligned}$$

because $U_i W_i$ always equals zero. From Theorem 6.3, we then have

$$\text{cov}(X_s, X_t) = \sum_{i=1}^{n} \sum_{j=1}^{n} \text{cov}(U_i, W_j)$$

$$= \sum_{i=1}^{n} \text{cov}(U_i, W_i) + \sum_{i<j}\sum \text{cov}(U_i, W_j)$$
$$= \sum_{i=1}^{n} (-p_s p_t) + 0$$
$$= -np_s p_t.$$

The covariance is negative, which is to be expected because a large number of outcomes in cell s would force the number in cell t to be small. ▪

Multinomial Experment

Multinomial Experiment

1. The experiment consists of n identical trials.
2. The outcome of each trial falls into one of k classes or cells.
3. The probability that the outcome of a single trial will fall in a particular cell—say, cell i—is $p_i (i = 1, 2, \ldots, k)$ and remains the same from trial to trial. Notice that $p_1 + p_2 + \cdots p_k = 1$.
4. The trials are independent.
5. The random variables of interest are $X_1, X_2, \ldots, X_k$, where $X_i (i = 1, 2, \ldots, k)$ is equal to the number of trials in which the outcome falls in cell i. Notice that $X_1 + X_2 + \cdots + X_k = n$.

The Multinomial Distribution

$$P(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!\, x_2! \cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$$

where $\sum_{i=1}^{k} x_i = n$ and $\sum_{i=1}^{k} p_i = 1$

$$E(X_i) = np_i \quad V(X_i) = np_i\left(1 - p_i\right), \quad i = 1, 2, \ldots, k$$
$$\text{cov}\left(X_i, X_j\right) = -np_i p_j, \qquad i \neq j$$

Exercises

6.59 The National Fire Incident Reporting Service says that among residential fires, approximately 74% are in one- or two-family homes, 20% are in multifamily homes (such as apartments), and the other 6% are in other types of dwellings. If five fires are reported independently in 1 day, find the probability that two are in one- or two-family homes, two are in multifamily homes, and one is in another type of dwelling.

6.60 The typical cost of damages for a fire in a family home is \$50,000, whereas the typical cost of damage in an apartment fire is \$20,000 and in other dwellings only \$5000. Using the information given in Exercise 6.59, find the expected total damage cost for the five independently reported fires.

6.61 In inspections of aircraft, wing cracks are reported as nonexistent, detectable, or critical. The history of a certain fleet reveals that 70% of the planes inspected have no wing cracks, 25% have detectable wing cracks, and 5% have critical wing cracks. For the next four planes inspected, find the probabilities of the following events.

a One has a critical crack, one has a detectable crack, and two have no cracks.

b At least one critical crack is observed.

6.62 Of the 544,685 babies born in California in 2004, 24% were born in the winter (December, January, February), 24% were born in the spring (March, April, May), 26% were born in the summer (June, July, August), and 26% were born in the fall (September, October, November). If four babies are randomly selected, what is the approximate probability that one was born in each season? Does it matter whether the babies were born in 2005? In California?

6.63 The U.S. Bureau of Labor Statistics reported that as of 2005, 19% of the adult population under age 65 were between 16 and 24 years of age, 20% were between 25 and 34, 23% were between 35 and 44, and 38% were between 45 and 64. An automobile manufacturer wants to obtain opinions on a new design from five randomly chosen adults from those under age 65. Of the five people so selected, find the approximate probability that two will be between 16 and 24, two will be between 25 and 44, and one will be between 45 and 64.

6.64 Customers leaving a subway station can exit through any one of three gates. Assuming that any particular customer is equally likely to select any one of the three gates, find the probabilities of the following events among a sample of four customers.

a Two select gate A, one selects gate B, and one selects gate C.

b All four select the same gate.

c All three gates are used.

6.65 Among a large number of applicants for a certain position, 60% have only a high-school education, 30% have some college training, and 10% have completed a college degree. If five applicants are selected to be interviewed, find the probability that at least one will have completed a college degree. What assumptions must be true for your answer to be valid?

6.66 A roulette wheel has 38 numbers of which 18 are red, 18 are black, and 2 are green. When the roulette wheel is spun, the ball is equally likely to fall on any of the 38 numbers, and the outcome from one spin of the wheel is independent of the outcome on another spin of the wheel. Suppose the wheel is spun 10 times.

a Find the probability of obtaining five reds, three blacks, and two greens.

b Find the probability of obtaining five reds, three blacks, and two greens if the first four spins of the wheel result in all reds.

6.67 In a large lot of manufactured items, 10% contain exactly one defect, and 5% contain more than one defect. Ten items are randomly selected from this lot for sale, and the repair costs total

$$X_1 + 3X_2$$

where X_1 denotes the number among the 10 having exactly one defect, and X_2 denotes the number among the 10 having two or more defects. Find the expected value of the repair costs. Find the variance of the repair costs.

6.68 Vehicles arriving at an intersection can turn right or left or can continue straight ahead. In a study of traffic patterns at this intersection over a long period, engineers have noted that 40% of the vehicles turn left, 25% turn right, and the remainder continue straight ahead.

a For the next five cars entering this intersections, find the probability that one turns left, one turns right, and three continue straight ahead.

b For the next five cars entering the intersection, find the probability that at least one turns right.

c If 100 cars enter the intersection in a day, find the expected values and the variance of the number that turn left. What assumptions must be true for your answer to be valid?

6.7 More on the Moment-Generating Function

The use of moment-generating functions to identify distributions of random variables is particularly useful in work with sums of independent random variables. Suppose that X_1 and X_2 are independent random variables, and let $Y = X_1 + X_2$. Using the properties of expectations, we have

$$\begin{aligned} M_Y(t) &= E\left(e^{tY}\right) = E\left(e^{t(X_1+X_2)}\right) \\ &= E\left(e^{tX_1}e^{tX_2}\right) \\ &= E\left(e^{tX_1}\right)E\left(e^{tX_2}\right) \\ &= M_{X_1}(t)M_{X_2}(t). \end{aligned}$$

That is, the moment-generating function of the sum of two independent random variables is the product of the moment-generating functions of the two variables. This result can be extended to the sum of n independent random variables; that is, if $X_1, X_2, \ldots, X_n$ are n independent random variables, the moment-generating function of $Y = X_1 + X_2 + \cdots + X_n$ is the product of the n moment-generating functions:

$$M_Y(t) = M_{X_1}(t)M_{X_2}(t) \cdots M_{X_n}(t)$$

If we can then identify the form of the moment-generating function, we can determine the distribution of Y as illustrated in the next example.

EXAMPLE 6.19 Suppose that X_1 and X_2 are independent, exponential random variables, each with mean θ. Define $Y = X_1 + X_2$. Find the distribution of Y.

Solution The moment-generating functions of X_1 and X_2 are

$$M_{X_1}(t) = M_{X_2}(t) = (1 - \theta t)^{-1}.$$

Thus, the moment-generating function of Y is

$$\begin{aligned} M_Y(t) &= M_{X_1}(t)M_{X_2}(t) \\ &= (1-\theta t)^{-1}(1-\theta t)^{-1} \\ &= (1-\theta t)^{-2}. \end{aligned}$$

Notice that this is the moment-generating function of a gamma distribution. Because if they exist, moment-generating functions are unique, we conclude that Y has a gamma distribution with $\alpha = 2$ and $\beta = \theta$. This result can be generalized to the sum of independent gamma random variables with the common scale parameter β.

Note: A key to this method is identifying the moment-generating function of the derived random variable Y. ▪

EXAMPLE 6.20 Let X_1 denote the number of vehicles passing a particular point on the eastbound lane of a highway in 1 hour. Suppose that the Poisson distribution with mean λ_1 is a reasonable model for X_1. Now let X_2 denote the number of vehicles passing a point on the westbound lane of the same highway in 1 hour. Suppose that X_2 has a Poisson distribution with mean λ_2. Of interest is $Y = X_1 + X_2$, which is the total traffic count in both lanes in 1 hour. Find the probability distribution for Y if X_1 and X_2 are assumed to be independent.

Solution It is known from Chapter 4 that

$$M_{X_1}(t) = e^{\lambda_1(e^t-1)}$$

and

$$M_{X_2}(t) = e^{\lambda_2(e^t-1)}.$$

By the properties of moment-generating functions described earlier,

$$\begin{aligned} M_Y(t) &= M_{X_1}(t)M_{X_2}(t) \\ &= e^{\lambda_1(e^t-1)}e^{\lambda_2(e^t-1)} \\ &= e^{(\lambda_1+\lambda_2)(e^t-1)}. \end{aligned}$$

Now the moment-generating function for Y has the form of a Poisson moment-generating function with mean $\lambda_1 + \lambda_2$. Thus, by the uniqueness property, Y must have a Poisson distribution with mean $\lambda_1 + \lambda_2$. Being able to add independent Poisson random variables and still retain the Poisson properties is important in many applications. ▪

Exercises

6.69 A quality control inspector randomly selects n_1 items from the production line in the morning and determines the number of defectives, X_1. She randomly selects n_2 items from the same production line in the afternoon and determines the number of defectives, X_2.

a Find the distribution of the number of defectives observed on a randomly selected day. What assumptions did you have to make?

b If $n_1 = 5$ and $n_2 = 4$, what is the probability of not obtaining any defectives if the line is producing 1% defectives?

6.70 Let X_1 and X_2 denote independent, normally distributed random variables, not necessarily having the same mean and variance.

a Show that for any constants a and b, $Y = aX_1 + bX_2$ is normally distributed.

b Use the moment-generating function to find the mean and the variance of Y.

6.71 Resistors of a certain type have resistances that are normally distributed with a mean of 100 ohms and a standard deviation of 10 ohms. Two such resistors are connected in series, which causes the total resistance of the circuit to be the sum of the individual resistances. Find the probabilities of the following events.

a The total resistance exceeds 220 ohms.

b The total resistance is less than 190 ohms.

6.72 A certain type of elevator has a maximum weight capacity X, which is normally distributed with a mean and a standard deviation of 5000 and 300 pounds, respectively. For a certain building equipped with this type of elevator, the elevator loading Y is a normally distributed random variable with a mean and a standard deviation of 4000 and 400 pounds, respectively. For any given time that the elevator is in use, find the probability that it will be overloaded, assuming that X and Y are independent.

6.8 Compounding and Its Applications

The univariate probability distributions of Chapters 4 and 5 depend on one or more parameters; once the parameters are known, the distributions are specified completely. These parameters frequently are unknown, however, and (as in Example 6.16) sometimes may be regarded as random quantities. Assigning distributions to these parameters and then finding the marginal distribution of the original random variable is known as compounding. This process has theoretical as well as practical uses, as we see next.

EXAMPLE **6.21** Suppose that X denotes the number of bacteria per cubic centimeter in a certain liquid and that for a given location, X has a Poisson distribution with mean λ. Suppose, too, that λ varies from location to location and that for a location chosen at random, λ has a gamma distribution with parameters α and β, where α is a positive integer. Find the probability distribution for the bacteria count X at a randomly selected location.

Solution Because λ is random, the Poisson assumption applies to the conditional distribution of X for fixed λ. Thus,

$$p(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \qquad x = 0, 1, 2, \ldots$$

Also,

$$f(\lambda) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha}\lambda^{\alpha-1}e^{-\lambda\beta}, & \lambda > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Then the joint distribution of λ and X is given by

$$\begin{aligned} g(x, \lambda) &= p(x|\lambda)f(\lambda) \\ &= \frac{1}{x!\Gamma(\alpha)\beta^\alpha}\lambda^{x+\alpha-1}e^{-\lambda[1+(1/\beta)]}. \end{aligned}$$

The marginal distribution of X is found by integrating over λ; it yields

$$\begin{aligned} p(x) &= \frac{1}{x!\Gamma(\alpha)\beta^\alpha}\int_0^\infty \lambda^{x+\alpha-1}e^{-\lambda[1+(1/\beta)]}\,dx \\ &= \frac{1}{x!\Gamma(\alpha)\beta^\alpha}\Gamma(x+\alpha)\left(1+\frac{1}{\beta}\right)^{-(x+\alpha)}. \end{aligned}$$

Because α is an integer,

$$\begin{aligned} p(x) &= \frac{(x+\alpha-1)!}{(\alpha-1)!x!}\left(\frac{1}{\beta}\right)^\alpha\left(\frac{\beta}{1+\beta}\right)^{(x+\alpha)} \\ &= \binom{x+\alpha-1}{\alpha-1}\left(\frac{1}{1+\beta}\right)^\alpha\left(\frac{\beta}{1+\beta}\right)^x, \qquad x = 0, 1, 2, \ldots \end{aligned}$$

If we let $\alpha = r$ and $1/(1+\beta) = p$, then $p(x)$ has the form of a negative binomial distribution. Hence, the negative binomial distribution is a reasonable model for counts in which the mean count may vary. ■

Exercises

6.73 Suppose that a customer arrives at a checkout counter in a store just as the counter is opening. A random number of customers N will be ahead of him, because some customers may arrive early. Suppose that this number has the probability distribution

$$p(n) = P(N = n) = pq^n, \qquad n = 0, 1, 2, \ldots$$

where $0 < p < 1$ and $q = 1 - p$ (this is a form of the geometric distribution). Customer service times are assumed to be independent and identically distributed exponential random variables with a mean of θ.

a For a given n, find the waiting time W for the customer to complete his checkout.

b Find the distribution of waiting time for the customer to complete his checkout.

6.74 Refer to Exercise 6.73. Find the mean and the variance of the waiting time for the customer to complete his checkout.

6.9 Summary

Most phenomena that one studies, such as the effect of a drug on the body or the life length of a computer system, are the result of many variables acting jointly. In such studies, it is important to look not only at the *marginal distributions* of random variables individually, but also at the *joint distribution* of the variables acting together and at the *conditional distributions* of one variable for the fixed values of another. Expected values of joint distributions lead to *covariance* and *correlation*, which help assess the direction and strength of association between two random variables. *Moment-generating functions* help identify distributions as being of a certain type (more on this in Chapter 7) and help in the calculation of expected values.

As in the univariate case, multivariate probability distributions are models of reality and do not precisely fit real situations. Nevertheless, they serve as very useful approximations that help build understanding of the world around us.

Supplementary Exercises

6.75 Consider the joint probability density function given next:

$$f(x, y) = \begin{cases} cxy, & 0 \le y \le x, 0 \le x \le 2 \\ 0, & \text{otherwise.} \end{cases}$$

a Determine the constant c that makes f a legitimate joint probability density function.

b Find the marginal density of X.

c Find the marginal density of Y.

d Find the probability $P(X < 1/2,\ Y < 3/4)$.

e Find the probability $P(X \le 1/2 | Y < 3/4)$.

6.76 The joint density of X and Y is given by

$$f(x, y) = \begin{cases} 3x, & 0 \le y \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the marginal density functions of X and Y.

b Find $P(X \le 3/4,\ Y \le 1/2)$.

c Find $P(X < 3/4 | Y \le 1/2)$.

6.77 Refer to Exercise 6.75.

a Find the conditional density of Y given X.

b Find the expected value of Y given X.

c Are X and Y independent? Justify your answer.

6.78 Refer to the situation described in Exercise 6.76.

a Find the conditional density of X given $Y = y$.

b Find the conditional density of Y given $X = x$.

c Are X and Y independent? Justify your answer.

d Find the probability $P(X \leq 3/4 | Y = 1/2)$.

6.79 A committee of three persons is to be randomly selected from a group consisting of four Republicans, three Democrats, and two independents. Let X denote the number of Republicans on the committee, and let Y denote the number of Democrats on the committee.

a Find the joint probability distribution of X and Y.

b Find the marginal distributions of X and Y.

c Find the probability $P(X = 1 | Y \geq 1)$.

6.80 Two customers enter a bank at random times within a fixed 1-hour period. Assume that the arrivals are independent of each other.

a Find the probability that both arrive in the first half hour.

b Find the probability that the two customers are within 5 minutes of each other.

6.81 The joint density of X and Y is given by

$$f(x, y) = \begin{cases} 6x^2y, & 0 \leq x \leq 1, 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the mean and the variance of X.

b Find the mean and the variance of Y.

c Find the conditional distribution of X given Y.

6.82 Refer to Exercise 6.81.

a Find the covariance of X and Y.

b Find the correlation between X and Y.

6.83 Let X denote the amount of a certain bulk item stocked by a supplier at the beginning of a week, and suppose that X has a uniform distribution over the interval $(0, 1)$. Let Y denote the amount of this item sold by the supplier during the week, and suppose that Y has a uniform distribution over the interval $0 \leq y \leq x$, where x is a specific value of X.

a Find the joint density function of X and Y.

b If the supplier stocks an amount of the item of $1/2$, what is the probability that she sells an amount greater than $1/4$?

c If the supplier stocks an amount equal to $3/4$, what is the expected amount that she will sell during the week?

6.84 Let X and Y have the joint probability density function given by

$$f(x, y) = \begin{cases} 3y, & 0 \leq x \leq y;\ 0 \leq y \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the marginal density of X.

b Find the marginal density of Y.

c Find the conditional density of Y given $X = x$.

d Find the expected value of Y given $X = x$.

e Are X and Y independent? Justify your answer.

6.85 Let X and Y have a joint density given by

$$f(x, y) = \begin{cases} 2, & 0 \le x \le y,\ 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the covariance of X and Y.

b Find the correlation coefficient. What, if anything, does it tell you about the relationship of X and Y?

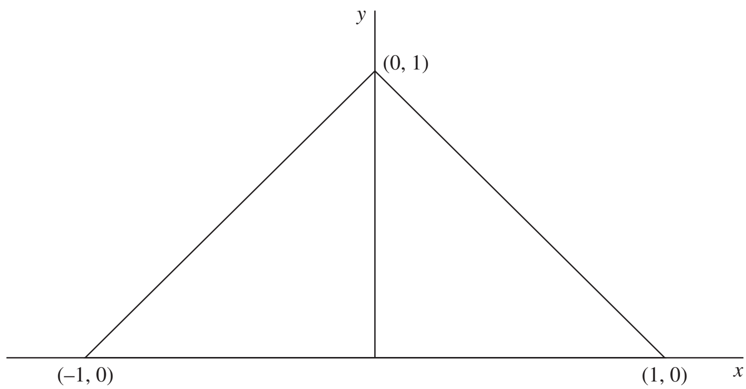
6.86 Let (X, Y) denote the coordinates of a point at random inside a unit circle with the center at the origin; that is, X and Y have a joint density function given by

$$f(x, y) = \begin{cases} 1/\pi, & x^2 + y^2 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the marginal density function of X.

b Find $P(X \le Y)$.

6.87 Let X and Y have a joint distribution that is uniform over the triangular region in the accompanying diagram.



a Find the marginal density of Y.

b Find the marginal density of X.

c Find the conditional density function of Y given X.

d Find $P[(X - Y) \ge 0]$.

6.88 Refer to the situation described in Exercise 6.87.

a Find $E(X)$.

b Find $V(X)$.

c Find $\text{cov}(X, Y)$.

6.89 Referring to Exercises 6.75 and 6.77, find $\text{cov}(X, Y)$.

6.90 Refer to the situation described in Exercise 6.79.

a Find $\text{cov}(X, Y)$.

b Find $E(X + Y)$ and $V(X + Y)$ by first finding the probability distribution of $X + Y$.

c Find $E(X + Y)$ and $V(X + Y)$ by using Theorem 6.2.

6.91 A quality control plan calls for randomly selecting three items from the daily production (assumed to be large) of a certain machine and observing the number of defectives. The proportion p of defectives produced by the machine varies from day to day and is assumed to have a uniform distribution on the interval (0, 1). For a randomly chosen day, find the unconditional probability that exactly two defectives are observed in the sample.

6.92 The number of defects per year, denoted by X, for a certain fabric is known to have a Poisson distribution with parameter λ. However, λ is not known and is assumed to be random with its probability density function given by

$$f(\lambda) = \begin{cases} e^{-\lambda}, & \lambda \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find the unconditional probability function for X.

6.93 The mean time that it takes any given machine of a particular type to fail is exponentially distributed with parameter λ. However, λ varies from machine to machine according to a gamma distribution with parameters r and θ.

a Find the mean time that it would take for a randomly selected machine to fail.

b Find the variance of the time that it would take for a randomly selected machine to break down.

6.94 The length of life X of a fuse has the probability density

$$f(\lambda) = \begin{cases} e^{-\lambda/\theta}/\theta, & x > 0,\ \theta > 0 \\ 0, & \text{otherwise.} \end{cases}$$

Three such fuses operate independently. Find the joint density of their lengths of life, X_1, X_2, and X_3.

6.95 A retail grocer figures that his daily gain X from sales is a normally distributed random variable with $\mu = 50$ and $\sigma^2 = 10$ (measurements in dollars). X could be negative if he is forced to dispose of perishable goods. He also figures daily overhead costs Y to have a gamma distribution with $\alpha = 4$ and $\beta = 2$. If X and Y are independent, find the expected value and the variance of his net daily gain. Would you expect his net gain for tomorrow to go above \$70?

6.96 A coin has probability p of coming up heads when tossed. In n independent tosses of the coin, let $Y_i = 1$ if the ith toss results in heads and $Y_i = 0$ if the ith toss results in tails. Then X, the number of heads in the n tosses, has a binomial distribution and can be represented as

$$X = \sum_{i=1}^{n} Y_i.$$

Find $E(X)$ and $V(X)$ using Theorem 6.2.

6.97 Refer to Exercises 6.75 and 6.77.

a Use Theorem 6.3 to find $E(Y)$.

b Find $E(Y)$ directly from the marginal density of Y.

6.98 Refer to Exercise 6.92.

a Find $E(X)$ by first finding the conditional expectation of X for given λ and then using Theorem 6.4.

b Find $E(X)$ directly from the probability distribution of X.

6.99 Among typical July days in Tampa, 30% have total radiation of at most 5 calories, 60% have total radiation of more than 5 calories but no more than 6 calories, and 10% have total radiation of more than 6 calories. A solar collector for a hot water system is to be run for 6 days. Find the probability that 3 days will produce no more than 5 calories each, 1 day will produce between 5 and 6 calories, and 2 days will produce at least 6 calories each. What assumption must be true for your answer to be correct?

6.100 Let X have a geometric distribution; that is, let

$$p(x) = p(1-p)^x, \qquad x = 0, 1, \ldots$$

The goal is to find $V(X)$ through a conditional argument.

a Let $Y = 1$ if the first trial is a success, and let $Y = 0$ if the first trial is a failure. Argue that

$$E(X^2|Y = 1) = 0$$

$$E(X^2|Y = 0) = E[(1 + X)^2].$$

b Write $E(X^2)$ in terms of $E(X^2|Y)$.

c Simplify the expression in part (b) to show that

$$E(X) = \frac{1-p}{p^2}.$$

d Show that

$$V(X) = \frac{1-p}{p^2}.$$

6.101 Let X be a continuous random variable with distribution function $F(x)$ and density function $f(x)$. We can then write for $x_1 \leq x_2$,

$$P(X \leq x_2|X \geq x_1) = \frac{F(x_2) - F(x_1)}{1 - F(x_1)}.$$

As a function of x_2 for fixed x_1, the right-hand side of this expression is called the *conditional distribution function* of X, given that $X \geq x_1$. On taking the derivative with respect to x_2, we see that the corresponding conditional density function is given by

$$\frac{f(x_2)}{1 - F(x_1)}, \qquad x_2 \geq x_1.$$

Suppose that a certain type of electronic component has life length X with the density function (life length measured in hours)

$$f(x) = \begin{cases} (1/400)e^{-x/400}, & x \geq 0 \\ 0, & \text{otherwise.} \end{cases}$$

Find the expected length of life for a component of this type that has already been in use for 100 hours.

6.102 Let X_1, X_2, and X_3 be random variables, either continuous or discrete. The joint moment-generating function of X_1, X_2, and X_3 is defined by

$$M(t_1, t_2, t_3) = E(e^{t_1X_1 + t_2X_2 + t_3X_3}).$$

a Show that $M(t_1, t_2, t_3)$ gives the moment-generating function for $X_1 + X_2 + X_3$.

b Show that $M(t_1, t_2, 0)$ gives the moment-generating function for $X_1 + X_2$.

c Show that

$$\left.\frac{\partial^{k_1+k_2+k_3} M(t_1, t_2, t_3)}{\partial t_1^{k_1} \partial t_2^{k_2} \partial t_3^{k_3}}\right|_{t_1=t_2=t_3=0} = E\left(X_1^{k_1} X_2^{k_2} X_3^{k_3}\right).$$

6.103 Let X_1, X_2, and X_3 have a multinomial distribution with probability function

$$p(x_1, x_2, x_3) = \frac{n!}{x_1!x_2!x_3!} p_1^{x_1} p_2^{x_2} p_3^{x_3}, \qquad \sum_{i=1}^{n} x_i = n.$$

Employ the results of Exercise 6.102 to perform the following tasks.

a Find the joint moment-generating function of X_1, X_2, and X_3.

b Use the joint moment-generating function to find $\text{cov}(X_1,\ X_2)$.

6.104 The negative binomial variable X is defined as the number of the failures prior to the rth success in a sequence of independent trials with constant probability p of success on each trial. Let Y_i denote a geometric random variable, which is defined as the number of failures prior to the trial on which the first success occurs. Then, we can write

$$X = \sum_{i=1}^{r} Y_i$$

for independent random variables $Y_1,\ Y_2,\ \ldots\ Y_r$. Use Theorem 6.2 to show that $E(X) = r/p$ and $V(X) = r(1-p)/p^2$.

6.105 A box contains four balls, numbered 1 through 4. One ball is selected at random from this box. Let

$$X_1 = 1 \text{ if ball 1 or ball 2 is drawn}$$
$$X_2 = 1 \text{ if ball 1 or ball 3 is drawn}$$
$$X_3 = 1 \text{ if ball 1 or ball 4 is drawn}$$

and let the values of X_i be zero otherwise. Show that any two of the random variables X_1, X_2, and X_3 are independent, but the three together are not.

6.106 Let X and Y be jointly distributed random variables with finite variances.

a Show that $[E(XY)]^2 \le E\left(X^2\right)E\left(Y^2\right)$. (*Hint*: Observe that for any real number t, $E\left[(tX - Y)^2\right] \ge 0$; or equivalently,

$$t^2E(X^2) - 2tE(XY) + E(Y^2) \ge 0.$$

This is a quadratic expression of the form $At^2 + Bt + C$, and because it is not negative, we must have $B^2 - 4AC \le 0$. (The preceding inequality follows directly.)

b Let ρ denote the correlation coefficient of X and Y; that is,

$$\rho = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}.$$

Using the inequality of part (a), show that $\rho^2 \le 1$.

6.107 A box contains N_1 white balls, N_2 black balls, and N_3 red balls ($N_1 + N_2 + N_3 = N$). A random sample of n balls is selected from the box without replacement. Let X_1, X_2, and X_3 denote the number of white, black, and red balls, respectively, observed in the sample. Find the correlation coefficient for X_1 and X_2. (Let $p_i = N_i/N$ for $i = 1,\ 2,\ 3$.)

6.108 Let $X_1,\ X_2,\ \ldots,\ X_n$ be independent random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$, for $i = 1,\ 2,\ \ldots,\ n$. Let

$$U_1 = \sum_{i=1}^{n} a_i X_i$$

and

$$U_2 = \sum_{i=1}^{n} b_i X_i$$

where $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_n$ are constants. U_1 and U_2 are said to be orthogonal if $\text{cov}(U_1, U_2) = 0$. Show that U_1 and U_2 are orthogonal if and only if $\sum_{i=1}^{n} a_i b_i = 0$.

6.109 The life length X of fuses of a certain type is modeled by the exponential distribution with

$$f(x) = \begin{cases} (1/3)e^{-x/3}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

The measurements are in hundreds of hours.

a If two such fuses have independent life lengths X and Y, find their joint probability density function.

b One fuse in part (a) is in a primary system, and the other is in a backup system that comes into use only if the primary system fails. The total effective life length of the two fuses is then $X + Y$. Find $P(X + Y \leq 1)$.

6.110 Referring to Exercise 6.109, suppose that three such fuses are operating independently in a system.

a Find the probability that exactly two of the three last longer than 500 hours.

b Find the probability that at least one of the three fails before 500 hours.

REFERENCES

Buffon, G. 1733. Editor's note concerning a lecture given 1733 by Mr. Le Clerc de Buffon to the Royal Academy of Sciences in Paris. *Histoire de l'Acad. Roy. des Sci.*, pp. 43–45.

Buffon, G. 1777. Essai d'arithmétique morale. *Histoire naturelle, générale er particulière, Supplément* 4:46–123.

Dawson, R. J. MacG. 1995. The "unusual episode" data revisited. *Journal of Statistics Education*, 3, no. 3.

7

Functions of Random Variables

7.1 Introduction

As we observed in Chapter 6, many situations we wish to study produce a set of random variables, $X_1, X_2, \ldots, X_n$, instead of a single random variable. Questions about the average life of components, the maximum price of a stock during a quarter, the time between two incoming telephone calls, or the total production costs across the plants in a manufacturing firm all involve the study of functions of random variables. This chapter considers the problem of finding the probability density function for a function of random variables with known probability distributions.

We have already seen some results along these lines in Chapter 6. Moment-generating functions were used to show that sums of exponential random variables have gamma distributions, and that linear functions of independent normal random variables are again normal. Those were special cases, however, and now more general methods for finding distributions of functions of random variables will be introduced.

If X and Y represent two different features of the phenomenon under study, it is sometimes convenient (and even necessary) to look at one random variable as a function of the other. For example, if the probability distribution for the velocity of a molecule in a uniform gas is known, the distribution of the kinetic energy can be found (see Exercise 7.78). And if the distribution of the radius of a sphere is known, the distribution of the volume can be found (see Exercise 7.77). Similarly, knowledge of the probability distribution for points in a plane allows us to find, in some cases, the distribution of the distance from a selected point to its neighbor. Knowledge of the distribution of life lengths of components in a complex system allows us to find the probability distribution for the life length of the system as a whole. Examples are endless, but those mentioned here should adequately illustrate the point that we are embarking on the study of a very important area of applied probability.

7.2 Functions of Discrete Random Variables

When working with functions of discrete random variables, the probabilities associated with each value x of X become associated with the value y corresponding to the function of X, say Y.

As an example, suppose a firm has a service contract on its copier. The firm pays a flat rate of \$150 each month plus \$50 for each service visit. Let X denote the number of repairs during a month. From experience the firm knows that the probability distribution of X is as follows (Table 7.1):

TABLE **7.1**
Number of Repairs during a Month.

x	**0**	**1**	**2**	**3**
$p(x)$	0.5	0.3	0.15	0.05

We now want to find the probability distribution of the monthly cost of maintenance for the copier. The cost C can be expressed as a function of X: $C = 150 + 50X$. Thus, for each $X = x$, there is a corresponding $C = c$. We can expand Table 7.1 as seen in Table 7.2.

TABLE **7.2**
Probabilities for X and the Transformed C.

x	**0**	**1**	**2**	**3**
c	150	200	250	300
$p(c)$	0.5	0.3	0.15	0.05

The important thing to note is that the probabilities of $C = c$ are equal to the corresponding probabilities of $X = x$. This is an intuitive, but correct, approach that works as long as the function is one to one.

Suppose now that we have the following probability distribution of X:

X	-1	0	1
$p(x)$	0.25	0.5	0.25

Define $Y = X^2$. Using the same approach as before we have the following:

X	-1	0	1
Y	1	0	1
$p(x)$	0.25	0.5	0.25

However, we have two cases in which $Y = 1$, each with a probability of 0.25. These probabilities need to be combined to form the total probability for $Y = 1$.

Y	0	1
$p(x)$	0.5	0.50

This technique can also be used when the form of the distribution of the random variable X is known, as illustrated in the next example.

EXAMPLE **7.1** A quality control manager samples from a large lot of items, testing each item until r ***defectives*** have been found. Find the distribution of Y, the number of ***items*** that are tested to obtain r defectives.

Solution Assuming that the probability p of obtaining a defective item is constant from trial to trial, the number of good items X sampled prior to the rth defective one is a negative binomial random variable. The density function of X is

$$P(X = x) = p(x) = \binom{x + r - 1}{r - 1} p^r q^x, \qquad x = 0,\ 1,\ \ldots$$

The number of trials, Y, required to get r defectives is equal to the sum of the number of good items before the rth defective and the number of defectives, r; that is, $Y = X + r$. Note that $X = Y - r$. Further, if the values of X with positive probability are $0, 1, 2, \ldots$, the values of Y with positive probability are $r,\ r + 1, \ldots$. Thus, making these substitutions in the density function, we have

$$P(Y = y) = p(y) = \binom{y - 1}{r - 1} p^r q^{y-r}, \qquad y = r,\ r + 1,\ \ldots$$

As mentioned in Chapter 4, some texts use Y in the definition of the negative binomial distribution instead of X as we have. ▪

This technique does not work for finding the probability density functions of continuous random variables because they have an infinite number of possible values for X. In this chapter, we will study four techniques for finding the probability density function for functions of continuous random variables.

Exercises

7.1 Among 10 applicants for an open position, 6 are women and 4 are men. Three applicants are randomly selected from the applicant pool for final interviews. Let X be the number of female applicants among the final three.

a Give the probability function for X.

b Define Y, the number of male applicants among the final three, as a function of X.

c Find the probability function for Y.

7.2 The median annual income for heads of households in a certain city is \$52,000. Five such heads of household are randomly selected for inclusion in an opinion poll. Let X be the number (out of the five) who have annual incomes below \$52,000.

a Give the probability distribution of X.

b Define Y, the number (out of the five) who have annual incomes of at least $52,000, as a function of X.

c Find the probability function of Y.

7.3 The employees of a firm that does asbestos cleanup are being tested for indications of asbestos in their lungs. The firm is asked to send four employees who have positive indications of asbestos to a medical center for further testing. Suppose 40% of the employees have positive indications of asbestos in their lungs. Let X be the number of employees who do not have asbestos in their lungs tested before finding the four who do have asbestos in their lungs.

a Give the probability function for X.

b Let Y be the total number of employees tested to find four with positive asbestos in their lungs. Define Y as a function of X.

c Find the probability function of Y.

7.4 Referring to Exercise 7.3, the cost to set up for testing is $300 and each test costs $40.

a Define C, the total cost of testing, in terms of Y.

b Find the probability function of C.

c Define C, the total cost of testing, in terms of X.

d Verify that the probability function of C, as defined in part (c), is the same as the probability function of C obtained in part (b).

7.3 Method of Distribution Functions

If X has a probability density function $f_X(x)$, and if Y is some function of X, then we can find $F_Y(y) = P(Y \leq y)$ by representing the event $Y \leq y$ in terms of X and then using the distribution function of X to determine that of Y. For example, suppose X is a continuous random variable with probability density f_X, and we want to find the distribution of $Y = X^2$. For $y \geq 0$,

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(X^2 \leq y) \\
&= P\left(-\sqrt{y} \leq X \leq \sqrt{y}\right) \\
&= F_X\left(\sqrt{y}\right) - F_X\left(-\sqrt{y}\right).
\end{aligned}$$

Notice that the critical step in the preceding method is moving from working with the distribution function of Y to working with the distribution function of X. We can find the probability density function of Y by differentiating the distribution function; that is,

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\sqrt{y}} f_X\left(\sqrt{y}\right) + \frac{1}{2\sqrt{y}} f_X\left(-\sqrt{y}\right) \\
&= \frac{1}{2\sqrt{y}} \left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right].
\end{aligned}$$

Now suppose that X is a normal random variable with mean 0 and standard deviation 1. What would be the probability density function of $Y = X^2$?

First, recall

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \qquad -\infty < x < \infty.$$

For $-\infty < x < \infty$, we have $y \geq 0$. Thus, for $y \geq 0$, we have

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}} \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\left(\sqrt{y}\right)^2/2\right) + \frac{1}{\sqrt{2\pi}} \exp\left(-\left(-\sqrt{y}\right)^2/2\right) \right] \\ &= \frac{1}{\sqrt{2\pi y}} \exp\left(-y/2\right) \\ &= \frac{1}{2^{1/2}\pi^{1/2}} y^{-1/2} e^{-y/2}. \end{aligned}$$

The density is 0 if $y < 0$. By looking at the exponents of y and e, we see that the density has the form of the gamma distribution with parameters $\alpha = 1/2$ and $\beta = 2$. We just need to confirm that the constants are the appropriate ones. The constants are

$$\frac{1}{\Gamma(\alpha)\beta^\alpha} = \frac{1}{\Gamma(1/2)2^{1/2}}.$$

Recalling that $\Gamma(1/2) = \sqrt{\pi}$, we see that the constants are the ones corresponding to a gamma distribution with parameters $\alpha = 1/2$ and $\beta = 2$.

Instead of representing the distribution function of Y in terms of the distribution function of X, we can find the $F_Y(y) = P(Y \leq y)$ directly by integrating $f_X(x)$ over the region for which $Y \leq y$. Then the probability density function for Y can be found by differentiating $F_Y(y)$.

EXAMPLE 7.2 The proportion of time X that a lathe is in use during a typical 40-hour workweek is a random variable whose probability density function is given by

$$f(x) = \begin{cases} 3x^2, & 0 \leq x \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

The actual number of *hours* out of a 40-hour week that the lathe is *not* in use then is

$$Y = 40(1 - X).$$

Find the probability density function for Y.

Solution Because the distribution function of Y looks at a region of the form $Y \leq y$, we must first find that region on the x scale. Now

$$\begin{aligned} Y \leq y &\Rightarrow 40(1 - X) \leq y \\ &\Rightarrow X > 1 - \frac{y}{40} \end{aligned}$$

so

$$
\begin{aligned}
F_Y(y) &= P(Y \le y) \\
&= P[40(1 - X) \le y] \\
&= P\left(X > 1 - \frac{y}{40}\right) \\
&= \int_{1-y/40}^{1} f(x)\,\mathrm{d}x \\
&= \int_{1-y/40}^{1} 3x^2\,\mathrm{d}x \\
&= \left[x^3\right]_{1-y/40}^{1}
\end{aligned}
$$

or

$$
F_Y(y) = \begin{cases} 0, & y < 0 \\ 1 - \left(1 - \frac{y}{40}\right)^3, & 0 \le y \le 40 \\ 1, & y > 40. \end{cases}
$$

The probability density function is found by differentiating the distribution function, so

$$
\begin{aligned}
f_Y(y) &= \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} \\
&= \begin{cases} \frac{3}{40}\left(1 - \frac{y}{40}\right)^2, & 0 \le y \le 40 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

We could now use $f_Y(y)$ to evaluate probabilities or to find the expected values related to the number of hours that the lathe is not in use. The reader should verify that $f_Y(y)$ has all the properties of a probability density function. ▪

The bivariate case is handled similarly, although it is often more difficult to transform the bivariate regions from statements about Y—a function of (X_1, X_2)—to statements about X_1 and X_2. Example 7.3 illustrates the point.

EXAMPLE **7.3** Two friends plan to meet at the library during a given 1-hour period. Their arrival times are independent and randomly distributed across the 1-hour period. Each agrees to wait for 15 minutes, or until the end of the hour. If the friend does not appear during that time, she will leave. What is the probability that the two friends will meet?

Solution If X_1 denotes one person's arrival time in (0, 1), the 1-hour period, and if X_2 denotes the second person's arrival time, then (X_1, X_2) can be modeled as having a two-dimensional uniform distribution over the unit square; that is,

$$f(x_1, x_2) = \begin{cases} 1, & 0 \le x_1 \le 1.\ 0 \le x_2 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

The event that the two friends will meet depends on the time Y between their arrivals where

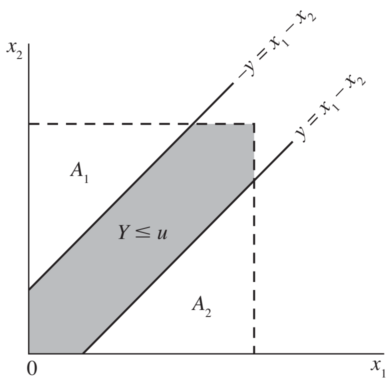
$$Y = |X_1 - X_2|.$$

We solve the specific problem by finding the probability density for Y. Now,

$$\begin{aligned} Y \le y &\Rightarrow |X_1 - X_2| \le y \\ &\Rightarrow -y \le X_1 - X_2 \le y. \end{aligned}$$

Figure 7.1 shows that the square region over which (X_1, X_2) has positive probability and the region defined by $Y \le y$. The probability of $Y \le y$ can be found by integrating the joint density function of (X_1, X_2) over the six-sided region shown in the center of Figure 7.1. This can be simplified by integrating over the triangles ($A_1$and A_2) and subtracting from 1, as we now see.

FIGURE **7.1**
Region $Y \le y$ for Example 7.3.



We have

$$\begin{aligned} F_Y(y) = P(Y \le y) &= \int\limits_{|x_1 - x_2| \le y} \int f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= 1 - \int\limits_{A_1}\int f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 - \int\limits_{A_2}\int f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= 1 - \int_y^1 \int_0^{x_2 - y} (1)\,\mathrm{d}x_1\,\mathrm{d}x_2 - \int_0^{1-y}\int_{x_2+y}^1 (1)\,\mathrm{d}x_1\,\mathrm{d}x_2 \end{aligned}$$

$$
\begin{aligned}
&= 1 - \int_{y}^{1} (x_2 - y)\,\mathrm{d}x_2 - \int_{0}^{1-y} (1 - y - x_2)\,\mathrm{d}x_2 \\
&= 1 - \left[\frac{1}{2}x_2^2 - x_2 y\right]_y^1 - \left[x_2 - x_2 y - \frac{1}{2}x_2^2\right]_0^{1-y} \\
&= 1 - \frac{1}{2}(1-y)^2 - \frac{1}{2}(1-y)^2 \\
&= 1 - (1-y)^2, \qquad 0 \le y \le 1.
\end{aligned}
$$

(Notice that the two double integrals really evaluate the volumes of prisms with triangular bases.) Because the probability density function is found by differentiating the distribution function, we have

$$
f_Y(y) = \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} = \begin{cases} 2(1-y), & 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}
$$

The problem asks for the probability of a meeting if each friend waits for up to 15 minutes. Because 15 minutes is $1/4$ hour, this can be found by evaluating

$$
\begin{aligned}
P\left(Y \le \frac{1}{4}\right) &= \int_{0}^{1/4} f_Y(y)\,\mathrm{d}y \\
&= F_Y\left(\frac{1}{4}\right) \\
&= 1 - \left(1 - \frac{1}{4}\right)^2 \\
&= 1 - \left(\frac{3}{4}\right)^2 \\
&= \frac{7}{16} \\
&= 0.4375.
\end{aligned}
$$

There is less than a 50-50 chance that the two friends will meet under this rule. ▪

Applications of probability often call for the use of sums of random variables. A study of downtimes of computer systems might require knowledge of the sum of the downtimes over a day or a week. The total cost of a building project can be studied as the sum of the costs for the major components of the project. The size of an animal population can be modeled as the sum of the sizes of the colonies within the population. The list of examples is endless. We now present an example that shows how to use the distribution functions to find the probability distribution of a sum of independent random variables. This complements the work in Chapter 6 on using moment-generating functions to find the distributions of sums.

EXAMPLE 7.4 Suppose that the length of time that a new, long-life bulb burns is exponentially distributed with a mean of 5 years. A homeowner has two such bulbs. She plans to place one bulb in the outside light of her home and to replace it with the second bulb when the first burns out. What is the probability that the bulbs will burn for at least 15 years before the second one burns out?

Solution Let X_1 and X_2 denote the length of time, respectively, that the first and second bulbs burn. The probability density function for either X_1 or X_2 has the form

$$f_X(x) = \begin{cases} \frac{1}{5}e^{-x/5}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}$$

and the distribution function of either X_1 or X_2 is given by

$$F_X(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-x/5}, & x \geq 0. \end{cases}$$

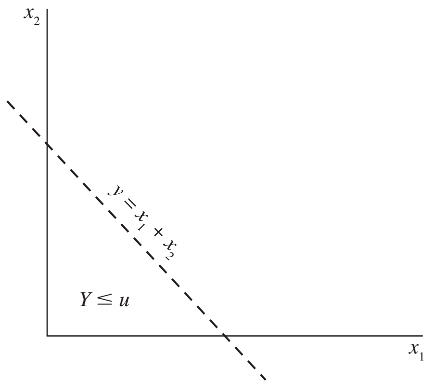
The length of time that one light bulb burns is independent of the time that the other one burns. Thus, the joint probability density function of X_1 and X_2 is

$$\begin{aligned} f_{12}(x_1, x_2) &= f_1(x_1)f_2(x_2) \\ &= \begin{cases} \frac{1}{25}e^{-(x_1+x_2)/5}, & x_1 > 0,\ x_2 > 0 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

The random variable $Y = X_1 + X_2$ denotes the length of time that both bulbs burn. To find $P(Y \leq y)$, we must integrate over the region shown in Figure 7.2. Thus,

$$\begin{aligned} F_Y(y) = P(Y \leq y) &= \int_0^y \int_0^{y-x_2} f_1(x_1)f_2(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 \\ &= \int_0^y F(y - x_2)f_2(x_2)\,\mathrm{d}x_2 \\ &= \int_0^y \left(1 - e^{-(y-x_2)/5}\right)\left(\frac{1}{5}\right)e^{-x_2/5}\,\mathrm{d}x_2 \\ &= \left[e^{-x_2/5}\right]_0^y - \left[\left(\frac{1}{5}\right)x_2 e^{-y}\right]_0^y \\ &= \begin{cases} 0, & y < 0 \\ 1 - e^{-y/5} - \left(\frac{1}{5}\right)ye^{-y/5}, & y \geq 0. \end{cases} \end{aligned}$$

FIGURE 7.2
Region $Y \leq y$ for Example 7.4.



Now we want to know the probability that both bulbs burn at least 15 years; that is,

$$\begin{aligned} P(Y \geq 15) &= 1 - P(Y < 15) \\ &= 1 - F(15) \\ &= 1 - \left(1 - e^{-3} - 3e^{-3}\right) \\ &= 4e^{-3} \\ &= 0.20. \end{aligned}$$

Further,

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} \\ &= \begin{cases} 0, & y < 0 \\ \frac{1}{25} y e^{-y/5}, & y \geq 0. \end{cases} \end{aligned}$$

Notice that $f_Y(y)$ is a gamma density function with $\alpha = 2$ and $\beta = 5$. This is consistent with results for the distribution of sums of independent exponential random variables found by using moment-generating functions. ▪

EXAMPLE 7.5 Let X have the probability density function given by

$$f_X(x) = \begin{cases} \frac{x+1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the density function for $Y = X^2$.

Solution Earlier in this section we found that

$$f_Y(y) = \frac{1}{2\sqrt{y}}\left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right].$$

On substituting into this equation, we obtain

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}}\left(\frac{\sqrt{y}+1}{2} + \frac{-\sqrt{y}+1}{2}\right) \\ &= \begin{cases} \frac{1}{2\sqrt{y}}, & 0 \le y \le 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Notice that because X has positive density over the interval $-1 \le x \le 1$, it follows that $Y = X^2$ has positive density over the interval $0 \le y \le 1$. ▪

Summary of the Distribution Function Method

Let Y be a function of the continuous random variables $X_1, X_2, \ldots, X_n$. Then,

1. Find the region $Y = y$ in the $(x_1, x_2, \ldots, x_n)$ space.
2. Find the region $Y \le y$.
3. Find $F_Y(y) = P(Y \le y)$ by integrating $f(x_1, x_2, \ldots, x_n)$ over the region $Y \le y$.
4. Find the density function $f_Y(y)$ by differentiating $F_Y(y)$. Thus, $f_Y(y) = \frac{dF_Y(y)}{dy}$.

Exercises

7.5 Let X be a random variable with a probability density function given by

$$f(x) = \begin{cases} 4(1-2x), & 0 \le x \le 0.5 \\ 0, & \text{otherwise.} \end{cases}$$

Find the density functions of the following variables.

a $Y_1 = 2X - 1$

b $Y_2 = 1 - 2X$

c $Y_3 = X^2$

7.6 Let X be a random variable with a probability density function given by

$$f(x) = \begin{cases} (3/2)x^2, & -1 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the density functions of the following variables.

a $Y_1 = 3X$

b $Y_2 = 3 - X$

c $Y_3 = X^2$

7.7 A supplier of kerosene has a weekly demand X possessing a probability density function given by

$$f(x) = \begin{cases} x, & 0 \le x \le 1 \\ 1, & 1 < x \le 1.5 \\ 0, & \text{elsewhere} \end{cases}$$

with measurements in hundreds of gallons. The supplier's profit is given by $Y = 10X - 4$.

a Find the probability density function for Y.

b Use the answer in part (a) to find $E(Y)$.

c Find $E(Y)$ by the methods in Chapter 5.

7.8 Let X be a random variable with cumulative distribution function given by

$$F(x) = \begin{cases} 1 - \frac{1}{x}, & x > 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability density function of $Y = 1/X$. Identify the distribution, including the parameters.

7.9 Let X be uniformly distributed over (0, 1). Find the probability density function of the random variable $Y = e^X$.

7.10 Consider the setting in Exercise 7.8. Find the probability density function of $Z = -\ln(Y)$. Identify the distribution, including the parameters.

7.11 Let X be exponentially distributed with a mean of θ. Find the probability density function of the random variable $Y = cX$, where c is some positive constant. Identify the distribution of Y, including the parameters.

7.12 The waiting time X for delivery of a new component for an industrial operation is uniformly distributed over the interval from 1 to 5 days. The cost of this delay is given by $Y = 2X^2 + 3$. Find the probability density function for Y.

7.13 Let the random variable X have the normal distribution with mean μ and variance σ^2. Find the probability density function of $Y = e^X$.

7.14 Suppose that a continuous random variable X has the distribution function $F(x)$. Show that $F(X)$ is uniformly distributed over the interval (0, 1).

7.15 The joint distribution of the amount of pollutant emitted from a smokestack without a cleaning device (X_1) and with a clean device (X_2) is given by

$$f(x) = \begin{cases} 1, & 0 \le x_1 \le 2;\ 0 \le x_2 \le 1;\ 2x_2 \le x_1 \\ 0, & \text{otherwise.} \end{cases}$$

The reduction in the amount of pollutant emitted due to the cleaning device is given by $Y = X_1 - X_2$.

a Find the probability density function for Y.

b Use the answer in part (a) to find $E(Y)$. (Compare this with the results of Exercises 6.34 and 6.36.)

7.16 The total time X_1 from arrival to completion of service at a fast-food outlet and the time spent waiting in line X_2 before arriving at the service window have a joint density function given by

$$f(x_1, x_2) = \begin{cases} e^{-x_1}, & 0 \le x_2 \le x_1 < \infty \\ 0, & \text{otherwise} \end{cases}$$

where $Y = X_1 - X_2$ represents the time spent at the service window.

a Find the probability density function for Y.

b Find $E(Y)$ and $V(Y)$, using the answer to part (a). (Compare this with the result of Exercise 6.40.)

7.4 Method of Transformations in One Dimension

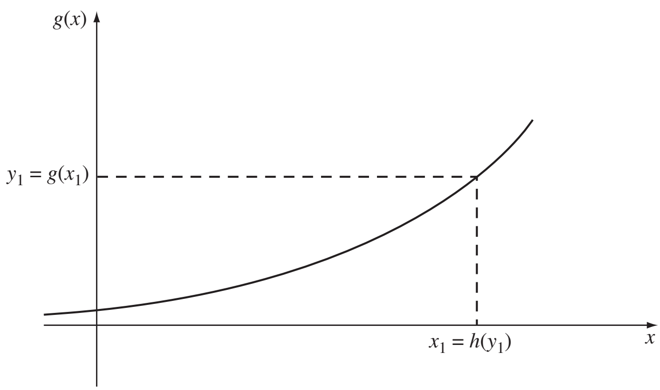
The transformation method for finding the probability distribution of a function of random variables is simply a generalization of the distribution function method (see Section 7.3). Through the distribution function approach, we can arrive at a single method of writing down the density function of $Y = g(X)$, provided that $g(x)$ is either decreasing or increasing. [By $g(x)$ increasing, we mean that, if $x_1 < x_2$, then $g(x_1) < g(x_2)$ for any real numbers x_1 and x_2.]

Suppose that $g(x)$ is an increasing function of x and that $Y = g(X)$, where X has density function $f_X(x)$. Let the inverse function be denoted by $h(y)$; that is, if $y = g(x)$, we can solve for x, obtaining $x = h(y)$. The graph of an increasing function $g(x)$ appears in Figure 7.3, where we see that the set of points x such that $g(x) \le y_1$ is precisely the same as the set of points x such that $x \le h(y_1)$. To find the density of $Y = h(X)$ by the distribution function method, we write

$$\begin{aligned} F_Y(y) &= P(Y \le y) \\ &= P[g(X) \le y] \\ &= P[X \le h(y)] \\ &= F_X[h(y)] \end{aligned}$$

where $F_X(x)$ is the distribution function of X.

FIGURE **7.3**
Increasing function.



To find the density function of Y—namely, $f_Y(y)$—we must differentiate $F_Y(y)$. Because $x = h(y)$,

$$F_Y(y) = F_X[h(y)].$$

Then,

$$\begin{aligned} f_Y(y) &= \frac{\mathrm{d}F_Y(y)}{\mathrm{d}y} \\ &= \frac{\mathrm{d}F_X(h(y))}{\mathrm{d}y} \\ &= f_X[h(y)]h'(y). \end{aligned}$$

Because h is an increasing function of x, $h'(y) > 0$. Suppose instead that g is a decreasing function of x so that $h'(y) < 0$. It is left as an exercise to show that $f_Y(y) = -f_X[h(y)]h'(y)$. Thus, in general, if g is a continuous, one-to-one function of x, then its inverse function h is a continuous one-to-one function of y and

$$f_Y(y) = f_X[h(y)] \left| h'(y) \right|.$$

This is summarized in the following theorem.

THEOREM **7.1**

Transformation of Random Variables: Let X be an absolutely continuous random variable with probability density function

$$f_X(x) = \begin{cases} > 0, & x \in A = (a, b) \\ 0, & x \notin A. \end{cases}$$

Let $Y = g(X)$ with inverse function $X = h(Y)$ such that h is a one-to-one, continuous function from $B = (\alpha, \beta)$ onto A. If $h'(y)$ exists and $h'(y) \neq 0$ for all $y \in B$, then $Y = g(X)$ determines a new random variable with density

$$f_Y(y) = \begin{cases} f_X[h(y)]|h'(y)|, & y \in B \\ 0, & y \notin B. \end{cases} \quad \blacksquare$$

The transformation process can be broken into steps as illustrated in the next example.

EXAMPLE **7.6** Let X have the probability density function given by

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

Find the density function of $Y = -2X + 5$.

Solution

1. The probability density function is given in the problem.
2. Solving $Y = g(X) = -2X + 5$ for X, we obtain the inverse function

$$X = h(Y) = \frac{5 - Y}{2}$$

where h is a continuous, one-to-one function from $B = (3, 5)$ onto $A = (0, 1)$.

3. $h'(y) = -\frac{1}{2} \neq 0$ for any $y \in B$
4. Then

$$\begin{aligned} f_Y(y) &= f_X[h(y)]\left|h'(y)\right| \\ &= 2\left(\frac{5-y}{2}\right)\left|-\frac{1}{2}\right| \\ &= \begin{cases} \frac{5-y}{2}, & 3 < y < 5 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$ ■

Notice that we used a series of steps to perform the transformation. These are summarized next.

Summary of the Univariate Transformation Method

Let Y be a function of the continuous random variables X; that is, $Y = g(X)$. Then,

1. Write the probability density function of X.
2. Find the inverse function h such that $X = h(Y)$. Verify that h is a continuous one-to-one function from $B = (\alpha, \beta)$ onto $A = (a, b)$ where for $x \in A, f(x) > 0$.
3. Verify $\frac{d[h(y)]}{dy} = h'(y)$ exists and is not zero for any $y \in B$.
4. Find $f_Y(y)$ by calculating

$$f_Y(y) = f_X[h(y)]\ \left|h'(y)\right|.$$

Exercises

7.17 Let X be a random variable with a probability density function given by

$$f(x) = \begin{cases} 4(1 - 2x), & 0 \leq x \leq 0.5 \\ 0, & \text{otherwise.} \end{cases}$$

Find the density functions of the following variables using the transformation technique.

a $Y_1 = 2X - 1$ **b** $Y_2 = 1 - 2X$

c $Y_3 = X^2$

7.18 Let X be a random variable with a probability density function given by

$$f(x) = \begin{cases} (3/2)x^2, & -1 \le x \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the density functions of the following variables using the transformation technique.

a $Y_1 = 3X$ **b** $Y_2 = 3 - X$

c $Y_3 = X^3$

7.19 Let X be a random variable with a cumulative distribution function given by

$$F(x) = \begin{cases} 1 - \frac{1}{x}, & x > 1 \\ 0, & \text{otherwise.} \end{cases}$$

a Find the probability density function of $Y = 1/X$ using the transformation technique. Identify the distribution, including the parameters.

b Find the probability density function of $Z = -\ln(Y)$ using the transformation technique. Identify the distribution, including the parameters.

7.20 Let X be uniformly distributed over (0, 1). Find the probability density function of the random variable $Y = e^X$ using the transformation technique.

7.21 The waiting time X for delivery of a new component for an industrial operation is uniformly distributed over the interval from 1 to 5 days. The cost of this delay is given by $Y = 2X^2 + 3$. Find the probability density function for Y using the transformation technique.

7.22 Let the random variable X have the normal distribution with mean μ and variance σ^2. Using the transformation technique, find the probability density function of $Y = e^X$.

7.23 A density function sometimes used to model lengths of life of electronic components is the Rayleigh density, which is given by

$$f(x) = \begin{cases} \left(\frac{2x}{\theta}\right) e^{-x^2/\theta}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

a If X has a Rayleigh density, find the probability density function for $Y = X^2$.

b Use the result in part (a) to find $E(X)$ and $V(X)$.

7.24 The Weibull density function is given by

$$f(x) = \begin{cases} \frac{1}{\alpha} m x^{m-1} e^{-x^m/\alpha}, & x > 0 \\ 0, & \text{otherwise} \end{cases}$$

where α and m are positive constants. Suppose X has the Weibull density.

a Find the density function of $Y = X^m$.

b Find $E(X^k)$ for any positive integer k.

7.5 Method of Conditioning

Conditional density functions frequently provide a convenient path for finding distributions of random variables. Suppose that X_1 and X_2 have a joint density function, and that we want to find the density function for $Y = h(X_1, X_2)$. We observe that the density of Y—namely, $F_Y(y)$—can be written as

$$f_Y(y) = \int_{-\infty}^{\infty} f(y, x_2)\,\mathrm{d}x_2.$$

Because the conditional density of Y given X_2 is equal to

$$f_Y(y \mid x_2) = \frac{f(y, x_2)}{f_2(x_2)}$$

it follows that

$$f_Y(y) = \int_{-\infty}^{\infty} f(y \mid x_2) f_2(x_2)\,\mathrm{d}x_2.$$

EXAMPLE **7.7** Let X_1 and X_2 be independent random variables, each with the density function

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

Find the density function for $Y = X_1/X_2$.

Solution We want to find the conditional density of Y given $X_2 = x_2$. When X_2 is held fixed at the constant value x_2, Y is simply X_1/x_2, where X_1 has the exponential density function and x_2 is treated as a known constant. Thus, we can use the transformation method of Section 7.4. We know the distribution of X_1 and want to find the distribution of the new function, $Y = g(x_1) = X_1/x_2$. We will follow the steps as suggested in Section 7.4.

1 The probability density function of X_1 is given above.

2 The inverse function $X_1 = h(Y) = Yx_2$ is a continuous, one-to-one function from $B = (0, \infty)$ onto $A = (0, \infty)$.

3 $h'(y) = x_2 \neq 0$ for all $y \in B$.

4 The conditions are satisfied, so by the method of Section 7.4,

$$\begin{aligned} f(y \mid x_2) &= f_{X_1}(yx_2)\ \left|h'(y)\right| \\ &= \begin{cases} x_2 e^{-x_2 y}, & y > 0 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

Now,

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f(y \mid x_2) f_2(x_2)\, dx_2 \\
&= \int_0^{\infty} e^{-yx_2} x_2 \left(e^{-x_2}\right) dx_2 \\
&= \int_0^{\infty} x_2 e^{-x_2(y+1)}\, dx_2 \\
&= \begin{cases} (y+1)^{-2}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}$$

Evaluating this integral is made easier by observing that the integrand is a gamma function. ▪

Summary of the Conditioning Method

Let Y be a function of the random variables X_1 and X_2.

1 Find the conditional density of Y given $X_2 = x_2$. (This is usually found by means of transformations.)

2 Find $f_Y(y)$ from the relation

$$f_Y(y) = \int_{-\infty}^{\infty} f(y \mid x_2) f(x_2)\, dx_2.$$

Exercises

7.25 The joint distribution of the amount of pollutant emitted from a smokestack without a cleaning device (X_1) and with a cleaning device (X_2) is given by

$$f(x) = \begin{cases} 1, & 0 \le x_1 \le 2.\ 0 \le x_2 \le 1;\ 2x_2 \le x_1 \\ 0, & \text{otherwise.} \end{cases}$$

The reduction in the amount of pollutant emitted due to the cleaning device is given by $Y = X_1 - X_2$. Find the probability density function for Y using the method of conditioning.

7.26 The total time X_1 from arrival to completion of service at a fast-food outlet and the time spent waiting in line X_2 before arriving at the service window have a joint density function given by

$$f(x_1, x_2) = \begin{cases} e^{-x_1}, & 0 \le x_2 \le x_1 < \infty \\ 0, & \text{otherwise} \end{cases}$$

where $Y = X_1 - X_2$ represents the time spent at the service window. Using the method of conditioning, find the probability density function for Y.

7.27 Suppose that a unit of mineral ore contains a proportion X_1 of metal A and a proportion X_2 of metal B. Experience has shown that the joint probability density function of (X_1, X_2) is uniform over the region $0 \le x_1 \le 1,\ 0 \le x_2 \le 1,\ 0 \le x_1 + x_2 \le 1$. Let $Y = X_1 + X_2$, the proportion of metals A and B per unit. Using the method of conditioning, find the probability density function for Y.

7.28 In a process that involves sintering two types of copper powder, the density function for X_1, the volume proportion of solid copper in a sample, is given by

$$f_1(x_1) = \begin{cases} 6x_1(1 - x_1), & 0 \le x_1 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The density function for X_2, the proportion of type A crystals among the solid copper, is given by

$$f_2(x_2) = \begin{cases} 3x_2^2, & 0 \le x_2 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The variable $Y = X_1 X_2$ represents the proportion of the sample volume due to type A crystals. Find the probability density function for Y, assuming that X_1 and X_2 are independent.

7.6 Method of Moment-Generating Functions

The definitions and uses of moment-generating functions have been introduced in Chapters 4, 5, and 6. Now, we will pull those ideas together and show how they form a powerful technique for finding distributions of functions of random variables in some common situations.

To review, the moment-generating function of a random variable Y is given by

$$M(t) = E\left(e^{tX}\right)$$

when such an expected value exists for values of t in a neighborhood of zero.

Key to our use of moment-generating functions is the *uniqueness* theorem.

THEOREM **7.2**

Uniqueness of the Moment-Generating Function: Let X and Y be two random variables with distribution functions $F_X(x)$and $F_Y(y)$, respectively. Suppose that the respective moment-generating functions, $M_X(t)$ and $M_Y(t)$, exist and are equal for $|t| < h,\ h > 0$. Then the two functions $F_X(x)$ and $F_Y(y)$ are equal.

The proof requires theory beyond the level of this book. ▪

Theorem 7.2 tells us that if we can identify the moment-generating function of Y as belonging to some particular distribution, then Y must have that particular distribution.

Note: Being able to recognize that the moment-generating function of a new random variable is the moment-generating function of a particular probability distribution is critical to successfully using this transformation method.

Before looking at special cases, let us review two other basic properties of moment-generating functions. If

$$Y = aX + b$$

for constants a and b, then

$$M_Y(t) = e^{bt}M_X(at).$$

If X and Y are independent random variables, then

$$M_{X+Y}(t) = M_X(t)M_Y(t).$$

7.6.1 Gamma Case

The moment-generating function is quite useful in dealing with certain functions of random variables that come from the gamma family in which the probability density function is given by

$$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1}e^{-x/\beta}, & x > 0 \\ 0, & \text{otherwise.} \end{cases}$$

In Exercise 5.140, the moment-generating function of a gamma random variable is found to be

$$M_X(t) = (1 - \beta t)^{-\alpha}.$$

As we have seen, sums of random variables occur frequently in applications. If $X_1, X_2, \ldots, X_n$ are independent gamma random variables, each having the preceding density function, then

$$Y = \sum_{i=1}^{n} X_i$$

has the moment-generating function

$$\begin{aligned} M_Y(t) &= M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t) \\ &= \left[(1-\beta t)^{-\alpha}\right]^n \\ &= (1-\beta t)^{-\alpha n}. \end{aligned}$$

Thus, Y has a gamma distribution with parameters $n\alpha$ and β. What happens to Y if the α-parameter is allowed to change from X_i to X_j?

7.6.2 Normal Case

Moment-generating functions are extremely useful in finding distributions of functions of normal random variables. If X has a normal distribution with mean μ and variance σ^2, then we know from Chapters 5 and 6 that

$$M_{X_i}(t) = \exp\left[\mu t + \frac{t^2\sigma^2}{2}\right].$$

Suppose that $X_1, X_2, \ldots, X_n$ are independent normal random variables with

$$E(X_i) = \mu_i \quad \text{and} \quad V(X_i) = \sigma_i^2.$$

Let

$$Y = \sum_{i=1}^{n} a_i X_i$$

for constants $a_1, a_2, \ldots, a_n$. The moment-generating function for Y is then

$$\begin{aligned} M_Y(t) &= M_{a_1X_1}(t)M_{a_{21}X_2}(t)\cdots M_{a_nX_n}(t) \\ &= \exp\left[a_1\mu_1 + \frac{t^2a_1^2\sigma_1^2}{2}\right]\exp\left[a_2\mu_2 + \frac{t^2a_2^2\sigma_2^2}{2}\right]\cdots\exp\left[a_n\mu_n + \frac{t^2a_n^2\sigma_n^2}{2}\right] \\ &= \exp\left[t\sum_{i=1}^{n} a_i\mu_i + \frac{t^2}{2}\sum_{i=1}^{n} a_i\sigma_i^2\right]. \end{aligned}$$

Thus, Y has a normal distribution with mean $\sum_{i=1}^{n} a_i\mu_i$ and $\sum_{i=1}^{n} a_i\sigma_i^2$. This distribution-preserving property of linear functions of normal random variables has numerous theoretical and applied uses.

7.6.3 Normal and Gamma Relationships

The two cases considered so far—gamma and normal—have connections to each other. Understanding these connections begins with looking at squares of normally distributed random variables.

Suppose that Z has a standard normal distribution. What can we say about the distribution of Z^2? Moment-generating functions might help. We have

$$\begin{aligned} M_{Z^2}(t) &= \int_{-\infty}^{\infty} e^{tz^2}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}z^2(1-2t)\right]dz \end{aligned}$$

$$
\begin{aligned}
&= (1-2t)^{-t/2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}(1-2t)^{-1/2}} \exp\left[-\frac{1}{2}z^2(1-2t)\right] \mathrm{d}z \\
&= (1-2t)^{-1/2}(1) \\
&= (1-2t)^{-1/2}.
\end{aligned}
$$

Now we begin to see the connection. Z^2 has a gamma distribution with parameters $1/2$ and 2.

What about the sum of squares of independent standard normal random variables? Suppose that $Z_1, Z_2, \ldots, Z_n$ are independent standard normal random variables; and let

$$Y = \sum_{i=1}^{n} Z_i^2.$$

Then,

$$
\begin{aligned}
M_Y(t) &= M_{Z_1^2}(t) M_{Z_2^2}(t) \cdots M_{Z_n^2}(t) \\
&= \left[(1-2t)^{-1/2}\right]^n \\
&= (1-2t)^{-n/2}
\end{aligned}
$$

and Y has a gamma distribution with parameters $n/2$ and 2. This particular type of gamma function comes up so often that it is given a specific name—the *chi-squared distribution*. In summary, a gamma distribution with parameters $n/2$ and 2 is called a *chi-squared distribution with n degrees of freedom*.

In order to pursue more of the connections between normal and chi-squared distributions, we must review some basic properties of joint moment-generating functions. If X_1 and X_2 have the joint probability density function $f(x_1, x_2)$, then their joint moment-generating function is given by

$$
\begin{aligned}
M_{X_1,X_2}(t_1, t_2) &= E\left[e^{t_1X_1+t_2X_2}\right] \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{t_1X_1+t_2X_2} f(x_1, x_2)\, \mathrm{d}x_1\, \mathrm{d}x_2.
\end{aligned}
$$

If X_1 and X_2 are independent,

$$M_{X_1,X_2}(t_1, t_2) = M_{X_1}(t_1)\, M_{X_2}(t_2).$$

We make use of these results in the ensuing discussion.

If $X_1, X_2, \ldots, X_n$ represent a set of independent random variables from a common normal (μ, σ^2) distribution, then the sample mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

is taken to be the estimator of an unknown μ. The sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

is taken to be the best estimator of σ^2. From our earlier work on linear functions, it is clear that $\bar{X}$ has another normal distribution. But what about the distribution of S^2?

The first task is to show that $\bar{X}$ and S^2 are independent random variables. It is possible to write S^2 as

$$S^2 = \frac{1}{2n(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i - X_j)^2$$

so that we only need show that differences $(X_1 - X_2)$ are independent of $X_1 + X_2 + \cdots + X_n$, the random part of $\bar{X}$. Let us see how this can be shown for the case of $n = 2$; the general case follows similarly.

Consider for two independent normal (μ, σ^2) random variables X_1 and X_2,

$$T = X_1 + X_2 \quad \text{and} \quad D = X_2 - X_1.$$

Now the moment-generating function of T is

$$M_Y(t) = \exp[t(2\mu) + t^2\sigma^2]$$

and the moment-generating function of D is

$$M_D(t) = \exp[t^2\sigma^2].$$

Also,

$$\begin{aligned}
M_{T,D}(t_1, t_2) &= E\{\exp[t_1(X_1 + X_2) + t_2(X_2 - X_1)]\} \\
&= E\{\exp[(t_1 - t_2)X_1 + (t_1 + t_2)X_2]\} \\
&= E\{\exp[(t_1 - t_2)X_1]\}E\{\exp[(t_1 + t_2)X_2]\} \\
&= \exp\left[(t_1 - t_2)\mu + \frac{(t_1 - t_2)^2}{2}\sigma^2 + (t_1 + t_2)\mu + \frac{(t_1 + t_2)^2}{2}\sigma^2\right] \\
&= \exp\left[t_1(2\mu) + t_1^2\sigma^2 + t_2^2\sigma^2\right] \\
&= M_T(t_1)\, M_D(t_2).
\end{aligned}$$

Thus, T and D are independent (and incidentally, each has a normal distribution).

The second task is to show how S^2 relates to quantities whose distribution we know. We begin with $\sum_{i=1}^{n}(X_i - \mu)^2$, which can be written

$$\sum_{i=1}^{n}(X_i - \mu)^2 = \sum_{i=1}^{n}[(X - \bar{X})^2 + (\bar{X} - \mu)^2]$$
$$= \sum_{i=1}^{n}(X_i - \bar{X})^2 + n(\bar{X} - \mu)^2.$$

(Why does the cross-product term vanish?) Now,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{n}{\sigma^2}(\bar{X} - \mu)^2$$

or

$$\sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$

or

$$\sum_{i=1}^{n} Z_i^2 = \frac{(n-1)S^2}{\sigma^2} + Z^2$$

where Z and all Z_i have a standard normal distribution. Relating these functions to known distributions, the quantity on the left has a chi-squared distribution with n degrees of freedom, and the one on the right has a chi-squared distribution with 1 degree of freedom. Relating these distributions to moment-generating functions (and using the independence of S^2 and $\bar{X}$), we have

$$(1 - 2t)^{-n/2} = M_{(n-1)S^2/\sigma^2}(t)(1 - 2t)^{-1/2}.$$

Thus,

$$M_{(n-1)S^2/\sigma^2}(t) = (1 - 2t)^{\frac{-n}{2}+\frac{1}{2}}$$
$$= (1 - 2t)^{-\frac{(n-1)}{2}}$$

and therefore, $(n - 1)S^2/\sigma^2$ has a chi-squared distribution with $(n - 1)$ degrees of freedom. This is a fundamental result in classical statistical inference. Its derivation shows something of the power of the moment-generating function approach.

Exercises

7.29 Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables, each having a binomial distribution with parameters n and p.

a If n and p are the same for all distributions, find the distribution of $\sum_{i=1}^{k} X_i$.

b What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the n's change from X_i to X_j, but the probability of success p is the same for each X?

c What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the p's change from X_i to X_j, but the number of trials n is the same for each X?

7.30 Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables, each having a negative binomial distribution with parameters r and p.

a If r and p are the same for all distributions, find the distribution of $\sum_{i=1}^{k} X_i$.

b What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the r's change from X_i to X_j, but the probability of success p is the same for each X?

c What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the p's change from X_i to X_j, but the parameter r is the same for each X.

7.31 Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables, each having a Poisson distribution, each with mean λ.

a Find the distribution of $\sum_{i=1}^{k} X_i$.

b What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the λ's change from X_i to X_j?

c For constants $a_1, a_2, \ldots, a_k$, does $\sum_{i=1}^{k} a_i X_i$ have a Poisson distribution? Justify your answer.

7.32 Suppose that $X_1, X_2, \ldots, X_k$ are independent random variables, each having a gamma distribution with parameters α and β.

a What is the distribution of $\sum_{i=1}^{k} X_i$?

b What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the α's change from X_i to X_j, but the parameter β is the same for each X?

c What happens to the distribution of $\sum_{i=1}^{k} X_i$ if the β's change from X_i to X_j, but the parameter α is the same for each X?

7.33 Suppose that $X_1, X_2, \ldots, X_k$ are independent observations from a normal (μ, σ^2) distribution. Find the distribution of $\bar{X}$.

7.34 Show that

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

can be written as

$$S^2 = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} (X_i - X_j)^2.$$

[*Hint*: Add and subtract $\bar{X}$ inside $(X_i - X_j)^2$.]

7.35 Suppose that $X_1, X_2, \ldots, X_n$ is a set of independent observations from a normal distribution with mean μ and variance σ^2.

a Find $E(S^2)$.

b Find $V(S^2)$.

c Find $E(S)$.

7.36 Projectiles aimed at a target land at coordinates (X_1, X_2), where X_1 and X_2 are independent standard normal random variables. Suppose that two such projectiles land independently at (X_1, X_2) and (X_3, X_4).

a Find the probability distribution for D^2, the square of the distance between the two points.

b Find the probability distribution for D.

7.7 Method of Transformation—Two Dimensions

In Section 7.4, we used the method of transformation to find the probability density function of a new random variable $Y = g(X)$, where the probability density function of X is known. However, sometimes we know the joint distribution of X_1 and X_2 and want to find the distribution of Y_1 and Y_2, which are functions of X_1 and X_2. We can extend the transformation method to two or more dimensions.

Let X_1 and X_2 be jointly distributed random variables with probability density function

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &> 0, \quad (x_1, x_2) \in A \\ &= 0, \quad (x_1, x_2) \notin A. \end{aligned}$$

Now suppose we want to find the distribution of Y_1 and Y_2, which are functions of X_1 and X_2. Specifically, suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2 from $(x_1, x_2) \in A$ onto $(y_1, y_2) \in B$. Assume that the functions g_1 and g_2 satisfy the following conditions:

1 The equations $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.

2 The functions h_1 and h_2 have continuous partial derivatives at all points $(x_1, x_2) \in A$ and are such that the following 2×2 determinant, called the Jacobian of the transformation,

$$J(y_1, y_2) = \begin{vmatrix} \dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} \\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} \end{vmatrix} = \frac{\partial h_1}{\partial y_1}\frac{\partial h_2}{\partial y_2} - \frac{\partial h_1}{\partial y_2}\frac{\partial h_2}{\partial y_1} \neq 0, \qquad (y_1, y_2) \in B.$$

Under these two conditions, the random variables Y_1 and Y_2 are jointly continuous with joint probability density function given by

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_{X_1,X_2}\left(h_1(y_1, y_2), h_2(y_1, y_2)\right) \left|J(y_1, y_2)\right|, \quad &(y_1, y_2) \in B \\ &= 0, &(y_1, y_2) \notin B. \end{aligned}$$

As with the univariate transformation technique, it is helpful to break down the method into steps as we illustrate in Example 7.8.

EXAMPLE 7.8 Let X_1 be a gamma random variable with parameters α_1 and β, and let X_2 be a gamma random variable with parameters α_2 and β. Further, suppose that X_1 and X_2 are independent. Find the joint and marginal distributions of $Y_1 = X_1 + X_2$ and $Y_2 = X_1/X_2$.

Solution 1 Find the joint probability density function of X_1 and X_2. Because X_1 and X_2 are independent, their joint probability density function is the product of the marginal probability density functions:

$$\begin{aligned} f_{X_1,X_2}(x_1, x_2) &= f_{X_1}(x_1) f_{X_2}(x_2), \qquad x_1 > 0,\ x_2 > 0 \\ &= \frac{1}{\Gamma(\alpha_1)\beta^{\alpha_1}} x_1^{\alpha_1 - 1} e^{-x_1/\beta} \frac{1}{\Gamma(\alpha_2)\beta^{\alpha_2}} x_2^{\alpha_2 - 1} e^{-x_2/\beta} \\ &= \begin{cases} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} x_1^{\alpha_1-1} x_2^{\alpha_2-1} e^{-(x_1+x_2)/\beta}, & x_1 > 0,\ x_2 > 0 \\ 0, & \text{otherwise} \end{cases} \end{aligned}$$

2 Find the inverse functions and check the conditions. Because the joint probability density function of X_1 and X_2 is positive for $x_1 > 0$ and $x_2 > 0$, $A = \{(x_1, x_2) \mid x_1 > 0,\ x_2 > 0\}$. Then, $B = \{(y_1, y_2) \mid y_1 > 0, y_2 > 0\}$. Now, solving $y_1 = x_1 + x_2$ and $y_2 = x_1/x_2$ for x_1 and x_2, we have

$$x_1 = h_1(y_1, y_2) = \frac{y_1 y_2}{1 + y_2}$$

and

$$x_2 = h_2(y_1, y_2) = \frac{y_1}{1 + y_2}.$$

Note that h_1 and h_2 are continuous, one to one, and differentiable.

3 Now we find the Jacobian of the transformation and verify that it is not equal to zero for any $(y_1, y_2) \in B$.

$$\begin{aligned} J(y_1, y_2) &= \begin{vmatrix} \dfrac{y_2}{1+y_2} & \dfrac{y_1}{(1+y_2)^2} \\ \dfrac{1}{1+y_2} & -\dfrac{y_1}{(1+y_2)^2} \end{vmatrix} = -\frac{y_1 y_2}{(1+y_2)^3} - \frac{y_1}{(1+y_2)^3} \\ &= -\frac{y_1}{(1+y_2)^2} \neq 0, \quad (y_1, y_2) \in B \end{aligned}$$

The conditions are satisfied for the transformation technique.

4 Find the joint density of Y_1 and Y_2. Thus,

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}\left(h_1(y_1, y_2),\ h_2(y_1, y_2)\right) |J(y_1, y_2)|, \qquad (y_1, y_2) \in B \\ &= \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} \left(\frac{y_1 y_2}{1+y_2}\right)^{\alpha_1-1} \left(\frac{y_1}{1+y_2}\right)^{\alpha_2-1} e^{-\left(\frac{y_1 y_2}{1+y_2} + \frac{y_1}{1+y_2}\right)/\beta} \left| -\frac{y_1}{(1+y_2)^2} \right| \\ &= \begin{cases} \frac{1}{\Gamma(\alpha_1)\Gamma(\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta} \frac{y_2^{\alpha_1-1}}{(1+y_2)^{\alpha_1+\alpha_2}}, & y_1 > 0, y_2 > 0 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Notice that $f_{Y_1,Y_2}(y_1, y_2)$ can be written as follows:

$$f_{Y_1,Y_2}(y_1, y_2)$$
$$= \begin{cases} \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{y_2^{\alpha_1-1}}{(1+y_2)^{\alpha_1+\alpha_2}}, & y_1 > 0, y_2 > 0 \\ 0, & \text{otherwise} \end{cases}$$

The marginal density of Y_1 is found by integrating over Y_2:

$$f_{Y_1}(y_1)$$
$$= \frac{1}{\Gamma(\alpha_1 + \alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta} \int_0^\infty \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{y_2^{\alpha_1-1}}{(1 + y_2)^{\alpha_1+\alpha_2}} dy_2$$
$$= \begin{cases} \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta}, & y_1 > 0 \\ 0, & \text{otherwise} \end{cases}$$

Thus, the marginal distribution of Y_1 is gamma with parameters $\alpha_1 + \alpha_2$ and β.

Note: Recognizing that the integral is equal to 1 was key to finding the marginal density of Y_1. It follows from the fact that

$$\int_0^\infty \frac{y^{\alpha_1-1}}{(1 + y)^{\alpha_1+\alpha_2}} dy = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)}$$

(see Exercise 5.180).

Similarly, the marginal density of Y_2 is determined by integrating over Y_1:

$$f_{Y_2}(y_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{y_2^{\alpha_1+\alpha_2-1}}{(1 + y_2)^{\alpha_1+\alpha_2}} \int_0^\infty \frac{1}{\Gamma(\alpha_1 + \alpha_2)\beta^{\alpha_1+\alpha_2}} y_1^{\alpha_1+\alpha_2-1} e^{-y_1/\beta} dy_1$$
$$= \begin{cases} \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \frac{y_2^{\alpha_1+\alpha_2-1}}{(1+y_2)^{\alpha_1+\alpha_2}}, & y_2 > 0 \\ 0, & \text{otherwise} \end{cases}$$

■

Note: When using this technique, it is important to remember that the transformation is from two dimensions to two dimensions or, more generally, from n dimensions to n dimensions. If we are only interested in $Y_1 = g(X_1, X_2)$, we must define a second variable, say Y_2. (We choose a Y_2 that will make the transformation as simple as possible.) Using the joint probability density of X_1 and X_2 and the transformation technique, we find the joint probability density of Y_1 and Y_2. Then the probability density function of Y_1 is found by integrating over Y_2.

Summary of the Transformation Method

Let Y_1 and Y_2 be functions of the continuous random variables X_1 and X_2. Then,

1. Write the joint probability density function of X_1 and X_2.
2. Suppose $Y_1 = g_1(X_1, X_2)$ and $Y_2 = g_2(X_1, X_2)$ for some functions g_1 and g_2 from $(x_1, x_2) \in A$ onto $(y_1, y_2) \in B$. Verify that $y_1 = g_1(x_1, x_2)$ and $y_2 = g_2(x_1, x_2)$ can be uniquely solved for x_1 and x_2 in terms of y_1 and y_2 with solutions given by, say $x_1 = h_1(y_1, y_2)$ and $x_2 = h_2(y_1, y_2)$.
3. Further suppose that the functions h_1 and h_2 have continuous partial derivatives at all points $(x_1, x_2) \in A$ and are such that the 2×2 determinant (the Jacobian of the transformation)

$$J(y_1, y_2) = \begin{vmatrix} \dfrac{\partial h_1}{\partial y_1} & \dfrac{\partial h_1}{\partial y_2} \\ \dfrac{\partial h_2}{\partial y_1} & \dfrac{\partial h_2}{\partial y_2} \end{vmatrix} = \frac{\partial h_1}{\partial y_1}\frac{\partial h_2}{\partial y_2} - \frac{\partial h_1}{\partial y_2}\frac{\partial h_2}{\partial y_1} \neq 0, \qquad (y_1, y_2) \in B.$$

4. Assuming the above holds, the random variables Y_1 and Y_2 are jointly continuous with joint probability density function given by

$$\begin{aligned} f_{Y_1,Y_2}(y_1, y_2) &= f_{X_1,X_2}\left(h_1(y_1, y_2), h_2(y_1, y_2)\right)\left|J(y_1, y_2)\right|, & (y_1, y_2) \in B \\ &= 0, & (y_1, y_2) \notin B. \end{aligned}$$

Exercises

7.37 Let X_1 and X_2 be independent, uniform random variables on the interval (0, 1). Let $Y_1 = 2X_1$ and $Y_2 = 3X_2$.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.
c Find the marginal distribution of Y_2.
d Are Y_1 and Y_2 independent? Justify your answer.

7.38 Let X_1 and X_2 be independent, normal random variables, each with a mean of 0 and a standard deviation of 1. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.
c Find the marginal distribution of Y_2.
d Are Y_1 and Y_2 independent? Justify your answer.

7.39 Let X_1 and X_2 be independent, exponential random variables, each with parameter θ. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_2$.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.

7.40 For an expensive item, two inspectors must approve each item before it is shipped. Let X_1, the time required for the first inspector to review an item, and X_2, the time required for the second inspector to review an item, be independent, exponential random variables with parameter θ. Let Y_1 be the proportion of the total inspection time taken by the first inspector and Y_2 be the total time taken by the inspectors.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.
c Find the marginal distribution of Y_2.
d Are Y_1 and Y_2 independent? Justify your answer.

7.41 Let X_1 and X_2 be independent, uniform random variables on the interval (0, 1). Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.
c Find the marginal distribution of Y_2.
d Are Y_1 and Y_2 independent? Justify your answer.

7.42 Let X_1 and X_2 be independent, exponential random variables, each with parameter θ. Let $Y_1 = X_1 + X_2$ and $Y_2 = X_1 - X_2$.

a Find the joint probability distribution of Y_1 and Y_2.
b Find the marginal distribution of Y_1.
c Find the marginal distribution of Y_2.
d Are Y_1 and Y_2 independent? Justify your answer.

7.43 The joint distribution of amount of pollutant emitted from a smokestack without a cleaning device (X_1) and with a clean device (X_2) is given by

$$f(x) = \begin{cases} 1, & 0 \le x_1 \le 2.\ 0 \le x_2 \le 1;\ 2x_2 \le x_1 \\ 0, & \text{otherwise.} \end{cases}$$

The reduction in amount of pollutant emitted due to the cleaning device is given by $Y = X_1 - X_2$. Find the probability density function for Y using the transformation method.

7.44 The total time X_1 from arrival to completion of service at a fast-food outlet and the time spent waiting in line X_2 before arriving at the service window have a joint density function given by

$$f(x_1, x_2) = \begin{cases} e^{-x_1}, & 0 \le x_2 \le x_1 < \infty \\ 0, & \text{otherwise.} \end{cases}$$

Let $Y = X_1 - X_2$ represent the time spent at the service window. Using the transformation method, find the probability density function for Y.

7.45 Suppose that a unit of mineral ore contains a proportion X_1 of metal A and a proportion X_2 of metal B. Experience has shown that the joint probability density function of (X_1, X_2) is uniform over the region $0 \le x_1 \le 1$, $0 \le x_2 \le 1$, $0 \le x_1 + x_2 \le 1$. Let $Y = X_1 + X_2$, the proportion of metals A and B per unit. Using the transformation method, find the probability density function for Y.

7.46 In a process that involves sintering two types of copper powder, the density function for X_1, the volume proportion of solid copper in a sample, is given by

$$f_1(x_1) = \begin{cases} 6x_1(1 - x_1), & 0 \le x_1 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The density function for X_2, the proportion of type A crystals among the solid copper, is given by

$$f_2(x_2) = \begin{cases} 3x_2^2, & 0 \le x_2 \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

The variable $Y = X_1X_2$ represents the proportion of the sample volume due to type A crystals. Using the transformation method, find the probability density function for Y, assuming that X_1 and X_2 are independent.

7.8 Order Statistics

Many functions of random variables of interest in practice depend on the relative magnitudes of the observed variables. For instance, we may be interested in the fastest time in an automobile race or the heaviest mouse among those fed a certain diet. Thus, we often order observed random variables according to their magnitudes. The resulting ordered variables are called *order statistics*.

Formally, let $X_1, X_2, \ldots, X_n$ denote independent continuous random variables with distribution function $F(X)$ and density function $f(X)$. We shall denote the ordered random variables X_i by $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, where $X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(n)}$. (Because the random variables are continuous, the equality signs can be ignored.) Thus,

$$X_{(1)} = \min(X_1, X_2, \ldots, X_n),$$

the minimum value of X_i, and

$$X_{(n)} = \max(X_1, X_2, \ldots, X_n),$$

the maximum value of X_i.

The probability density functions for $X_{(1)}$ and $X_{(n)}$ are found easily. Looking at $X_{(n)}$ first, we see that

$$P[X_{(n)} \leq x] = P(X_1 \leq x, X_2 \leq x, \ldots, X_n \leq x)$$

because $[X_{(n)} \leq x]$ implies that all the values of X_i must be less than or equal to x, and vice versa. However, $X_1, X_2, \ldots, X_n$ are independent; hence,

$$P[X_{(n)} \leq x] = [F(x)]^n.$$

Letting $g_n(x)$ denote the density function of $X_{(n)}$, we see, on taking derivatives on both sides, that

$$g_n(x) = n[F(x)]^{n-1}f(x).$$

The density function of $X_{(1)}$, denoted by $g_1(x)$, can be found by a similar device. We have

$$\begin{aligned} P[X_{(1)} \leq x] &= 1 - P[X_{(1)} > x] \\ &= 1 - P(X_1 > x, X_2 > x, \ldots, X_n > x) \\ &= 1 - [1 - F(x)]^n. \end{aligned}$$

Hence,

$$g_1(x) = n[1 - F(x)]^{n-1}f(x).$$

Similar reasoning can be used to derive the probability density function of $X_{(j)}$ for any j from 1 to n, but the details are a bit tedious. Nevertheless, we present one way of doing this, beginning with constructing the distribution function for $X_{(j)}$. Now,

$$\begin{aligned}
P[X_{(j)} \le x] &= P\text{ [at least } j \text{ of the values of } X_j \text{ are less than or equal to } x] \\
&= \sum_{k=j}^{n} P\text{ (Exactly } k \text{ of the values of } X_j \text{ are less than or equal to } x) \\
&= \sum_{k=j}^{n} \binom{n}{k} [F(x)]^k [1 - F(x)]^{n-k}.
\end{aligned}$$

The last expression comes about because the event "exactly k of the values of X_j are less than or equal to x" has a binomial construction with $p = F(x) = P(X_j \le x)$.

The probability density function of $X_{(j)}$ is the derivative of the distribution function for $X_{(j)}$, as given in the preceding equation. This derivative, with respect to x, is

$$\sum_{k=j}^{n} \binom{n}{k} \left\{ k[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x) - (n-k)[F(x)]^k [1 - F(x)]^{n-k-1} f(x) \right\}.$$

We now separate the term for $k = j$ from the first half of the expression, obtaining

$$\begin{aligned}
&\binom{n}{j} j[F(x)]^{j-1}[1 - F(x)]^{n-j} f(x) \\
&\quad + \sum_{k=j+1}^{n} \binom{n}{k} k[F(x)]^{k-1}[1 - F(x)]^{n-k} f(x) \\
&\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F(x)]^k [1 - F(x)]^{n-k-1} f(x)
\end{aligned}$$

The second summation can end at $(n - 1)$, because the $k = n$ term is zero. To make these two summations extend over the same range of values, we change the index of summation in the first one from k to $k + 1$ (that is, we replace k by $k + 1$ in the first summation). Then, the sums are

$$\begin{aligned}
&\sum_{k=j}^{n-1} \binom{n}{k+1} (k+1)[F(x)]^k [1 - F(x)]^{n-k-1} f(x) \\
&\quad - \sum_{k=j}^{n-1} \binom{n}{k} (n-k)[F(x)]^k [1 - F(x)]^{n-k-1} f(x).
\end{aligned}$$

By writing out the full factorial version, we can see that

$$\binom{n}{k+1}(k+1) = \binom{n}{k}(n-k)$$

and, therefore, the two summations are identical, canceling each other out. Thus, we are left with the probability density function for $X_{(j)}$.

$$g_j(x) = \binom{n}{j} j[F(x)]^{j-1}[1 - F(x)]^{n-j} f(x)$$

$$= \frac{n!}{(j-1)!\,(n-j)!}[F(x)]^{j-1}[1 - F(x)]^{n-1} f(x)$$

Careful inspection of this form will reveal a pattern that can be generalized to joint density functions of two order statistics. Think of $f(x)$ as being proportional to the probability that one of the values of X_i falls into a small interval about the value of x. The density function of $X_{(j)}$ requires $(j - 1)$ of the values of X_j to fall below x, one of them to fall on (or at least close to) x and $(n - j)$ of the n to exceed x. Thus, the probability that $X_{(j)}$ falls close to x is proportional to a multinomial probability with three cells, as follows:

$$\begin{array}{c|c|c} (j-1)X_i's & \begin{array}{c}\text{one}\\ X_i\end{array} & (n-j)X_i's \\ \hline & x & \end{array}$$

The probability that a value of X_i will fall in the first cell is $F(x)$; the probability for the second cell is proportional to $f(x)$; and the probability for the third cell is $[1 - F(x)]$. Thus,

$$g_j(x) = \frac{n!}{(j-1)!\,1!\,(n-j)!}[F(x)]^{j-1} f(x)[1 - F(x)]^{n-j}.$$

If we generalize to the joint probability density function for $X_{(i)}$ and $X_{(j)}$, where $i < j$, the diagram is

$$\begin{array}{c|c|c|c|c} (i-1)X_i's & \begin{array}{c}\text{one}\\ X_i\end{array} & (j-i-1)X_i's & \begin{array}{c}\text{one}\\ X_i\end{array} & (n-j)X_i's \\ \hline & x_i & & x_j & \end{array}$$

and the joint density (as proportional to a multinomial probability) becomes

$$g_{ij}(x_i, x_j) = \frac{n!}{(i-1)!\,(j-i-1)!\,(n-j)!}$$
$$\times [F(x_i)]^{i-1}[F(x_j) - F(x_i)]^{j-i-1}[1 - F(x_j)]^{n-j} f(x_i) f(x_j).$$

In particular, the joint density function of $X_{(1)}$ and $X_{(n)}$ becomes

$$g_{1n}(x_1, x_n) = \frac{n!}{(n-2)!}[F(x_n) - F(x_1)]^{n-2} f(x_1) f(x_n).$$

The same method can be used to find the joint density of $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$, which turns out to be

$$g_{12\cdots n}(x_1, x_2, \ldots, x_n) = \begin{cases} n!\,f(x_1)f(x_2)\cdots f(x_n), & x_1 \le x_2 \le \cdots \le x_n \\ 0, & \text{elsewhere.} \end{cases}$$

The marginal density function for any of the order statistics can be found from this joint density function, but we shall not pursue the matter in this text.

EXAMPLE 7.9 Electronic components of a certain type have a life length X with a probability density given by

$$f(x) = \begin{cases} \frac{1}{100}e^{-x/100}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

(Life length is measured in hours.) Suppose that two such components operate independently and in series in a certain system; that is, the system fails when either component fails. Find the density function for Y, the life length of the system.

Solution Because the system fails at the failure of the first component, $Y = \min(X_1, X_2)$, where X_1 and X_2 are independent random variables with the given density. Then, because $F(x) = 1 - e^{-x/100}$ for $x \geq 0$,

$$\begin{aligned} f_Y(y) &= g_1(x) \\ &= n[1 - F(x)]^{n-1} f(x) \\ &= 2e^{-x/100}\left(\frac{1}{100}\right)e^{-x/100} \\ &= \begin{cases} \left(\frac{1}{50}\right)e^{-x/50}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

Thus, we see that the minimum of two exponentially distributed random variables has an exponential distribution, but the mean is reduced by a half. ▪

EXAMPLE 7.10 Suppose that, in the situation described in Example 7.9, the components operate in parallel; that is, the system does not fail until both components fail. Find the density functions for Y, the life length of the system.

Solution Now, $Y = \max(X_1, X_2)$ and

$$\begin{aligned} f_Y(y) &= g_2(x) \\ &= n[F(x)]^{n-1} f(x) \\ &= 2(1 - e^{-x/100})\left(\frac{1}{100}\right)e^{-x/100} \\ &= \begin{cases} \left(\frac{1}{50}\right)(e^{-x/100} - e^{-x/50}), & x > 0 \\ 0, & \text{elsewhere.} \end{cases} \end{aligned}$$

We see that the maximum of two exponential random variables is not an exponential random variable. ▪

EXAMPLE 7.11 Suppose that $X_1, X_2, \ldots, X_n$ are independent random variables, each with a uniform distribution on the interval (0, 1).

1 Find the probability density function of the range $R = X_{(n)} - X_{(1)}$.

2 Find the mean and the variance of R.

Solution **1** For the uniform case,

$$f(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

and

$$F(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x < 1 \\ 1, & x \ge 1. \end{cases}$$

Thus, the joint density function of $X_{(1)}$ and $X_{(n)}$ is

$$f_{1n}(x_1, x_n) = \begin{cases} n(n-1)[x_n - x_1]^{n-2}, & 0 < x_1 < x_n < 1 \\ 0, & \text{elsewhere.} \end{cases}$$

To find the density function for $R = X_{(n)} - X_{(1)}$, fix the value of x_1 and let $R = X_n - x_1$. Thus $R = g(X_n) = X_n - x_1$, and the inverse function is $X_n = h(R) = R + x_1$. From the transformation methods,

$$f_{1r}(x_1, r) = n(n-1)(r)^{n-2}(1).$$

Now this is still a function of x_1 and r, and x_1 cannot be larger than $1 - r$ (why?). By integrating out x_1, we obtain

$$\begin{aligned} f_R(r) &= \int_0^{1-r} n(n-1)r^{n-2}\, dx_1 \\ &= \begin{cases} n(n-1)r^{n-2}(1-r), & 0 < r < 1 \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

2 Notice that this can be written as

$$f_R(r) = \int_0^{1-r} n(n-1)r^{n-2}\,dx_1$$

$$= \begin{cases} n(n-1)r^{(n-1)-1}(1-r)^{2-1}, & 0 < r < 1 \\ 0, & \text{otherwise} \end{cases}$$

which is a beta density function with parameters $\alpha = n - 1$ and $\beta = 2$. From the properties of the beta distribution, we have

$$E(R) = \frac{\alpha}{\alpha + \beta} = \frac{n-1}{n+1}$$

and

$$V(R) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

$$= \frac{2(n-1)}{(n+1)^2(n+2)}. \quad ■$$

Exercises

7.47 Let $X_1, X_2, \ldots, X_n$ represent a random sample from the uniform distribution over the interval (a,b).

a Find the distribution function of the last order statistic, $X_{(n)}$.

b Find the probability density function of the last order statistic, $X_{(n)}$.

c Find the expected value of the last order statistic, $X_{(n)}$.

7.48 The opening prices per share of two similar stocks, X_1 and X_2, are independent random variables, each with density function

$$f(x) = \begin{cases} \frac{1}{2}e^{-(x-4)/2}, & x > 4 \\ 0, & \text{elsewhere.} \end{cases}$$

On a given morning, Mr. A is going to buy shares of whichever stock is less expensive. Find the probability density function for the price per share that Mr. A will have to pay.

7.49 Suppose that the length of time it takes a worker to complete a certain task X has the probability density function

$$f(x) = \begin{cases} e^{-(x-\theta)}, & x > \theta \\ 0, & \text{elsewhere} \end{cases}$$

where θ is a positive constant that represents the minimum time from start to completion of the task. Let $X_1, X_2, \ldots, X_n$ denote independent random variables from this distribution.

a Find the probability density function for $X_{(1)} = \min(X_1, X_2, \ldots, X_n)$.

b Find $E(X_{(1)})$.

7.50 Suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample from a uniform distribution over the interval $(0, \theta)$ for some constant $\theta > 0$.

a Find the probability density function for $X_{(n)}$.

b Find $E(X_{(n)})$.

c Find $V(X_{(n)})$.

d Find a constant b such that $E(bX_{(n)}) = 0$.

e Find the probability density function for the range $R = X_{(n)} - X_{(1)}$.

f Find $E(R)$.

7.51 Suppose that $X_1, X_2, \ldots, X_n$ represent a random sample from the uniform distribution over the interval (θ_1, θ_2).

a Find $E(X_{(1)})$.

b Find $E(X_{(n)})$.

c Find a function of $X_{(1)}$ and $X_{(n)}$ such that the expected value of that function is $(\theta_2 - \theta_1)$.

7.52 Suppose that X_1 and X_2 are independent standard normal random variables. Find the probability density function for $R = X_{(2)} - X_{(1)}$.

7.53 Let $X_1, X_2, \ldots, X_n$ constitute a random sample from an exponential distribution with mean θ.

a Find the distribution of the first order statistic.

b Find the mean of the first order statistic.

c Find the variance of the first order statistic.

7.54 The time until a light bulb burns out is exponentially distributed with a mean of 1000 hours. Suppose that 100 such bulbs are used in a large factory. Answer the following, using the results of Exercise 7.53.

a Find the probability that the time until the first bulb burns out is greater than 20 hours.

b When a bulb burns out, it is immediately replaced. Find the probability that the time until the next bulb burns out is greater than 20 hours.

7.9 Probability-Generating Functions: Applications to Random Sums of Random Variables

Probability-generating functions (see Section 4.10) also are useful tools for determining the probability distribution of a function of random variables. We examine their applicability by developing some properties of random sums of random variables.

Let $X_1, X_2, \ldots, X_n$ be a sequence of independent and identically distributed random variables, each with the probability-generating function $R(x)$. Let N be an integer-valued random variable with the probability-generating function $Q(s)$. Define S_N as

$$S_N = X_1 + X_2 + \cdots + X_N.$$

The following are illustrative of the many applications of such random variables. N may represent the number of customers per day at a checkout counter, and X_i may represent the number of items purchased by the ith customer; then, S_N represents the total number of items sold per day. Or N may represent the number of seeds produced by a plant, and X_i may be defined by

$$X_i = \begin{cases} 1, & \text{if the } i\text{th seed germinates} \\ 0, & \text{otherwise.} \end{cases}$$

Then S_N is the total number of germinating seeds produced by the plant.

We now assume that S_N has the probability-generating function $P(s)$, and we investigate its construction. By definition,

$$P(s) = E\left(s^{S_N}\right)$$

which can be written as

$$\begin{aligned} P(s) &= E\left[E\left(s^{S_n} | N = n\right)\right] \\ &= \sum_{n=0}^{\infty} E\left(s^{S_n} | N = n\right) P(N = n). \end{aligned}$$

Notice that

$$\begin{aligned} E\left(s^{S_n}\right) &= E\left(s^{X_1 + X_2 + \cdots + X_n}\right) \\ &= E\left(s^{X_1} s^{X_2} \cdots s^{X_n}\right) \\ &= E\left(s^{X_1}\right) E\left(s^{X_2}\right) \cdots E\left(s^{X_n}\right) \\ &= [R(s)]^n \end{aligned}$$

because $R(s)$ is the probability-generating function of S. Also,

$$Q(s) = E\left(s^{S_N}\right) = \sum_{n=0}^{\infty} s^n P(N = n).$$

If N is independent of the values of X_i, then

$$\begin{aligned} P(s) &= \sum_{n=0}^{\infty} E\left(s^{S_n}\right) P(N = n) \\ &= \sum_{n=0}^{\infty} [R(s)]^n P(N = n) \end{aligned}$$

or

$$P(s) = Q\left[R(s)\right].$$

The expression $Q\left[R(s)\right]$ is the probability-generating function $Q(s)$ evaluated at $R(s)$ instead of at s.

Suppose that N has a Poisson distribution with mean λ. Then,

$$Q(s) = e^{\lambda(s-1)} = e^{-\lambda+\lambda s}.$$

It follows that

$$\begin{aligned} P(s) &= Q\,[R(s)] \\ &= e^{-\lambda+\lambda R(s)}. \end{aligned}$$

This is often referred to as the *probability-generating function* of the compound Poisson distribution.

EXAMPLE 7.12 Suppose that N, the number of animals caught in a trap per day, has a Poisson distribution with mean λ. The probability of any one animal's being male is p. Find the probability-generating function and the expected value of the number of males caught per day.

Solution Let $X_i = 1$ if the ith animal caught is male, and let $X_i = 0$ otherwise. Then $S_N = X_1 + X_2 + \cdots + X_N$ denotes the total number of males caught per day. Now,

$$R(s) = E(s^X) = q + ps$$

where $q = 1 - p$; and $P(s)$, the probability-generating function of S_N, is given by

$$\begin{aligned} P(s) &= e^{-\lambda+\lambda R(s)} \\ &= e^{-\lambda+\lambda(q+ps)} \\ &= e^{-\lambda(1-q)+\lambda ps} \\ &= e^{-\lambda p+\lambda ps}. \end{aligned}$$

Thus, S_N has a Poisson distribution with parameter λp. ■

Exercises

7.55 In the compound Poisson distribution of Section 6.8, let

$$P(X_i = n) = \frac{1}{n}\alpha q^n, \qquad n = 1, 2, \ldots$$

for some constants α and q, where $0 < q < 1$. (X_i is said to have a logarithmic series distribution, which is frequently used in ecological models.) Show that the resulting compound Poisson distribution is related to the negative binomial distribution.

7.56 Show that the random variable S_N, defined in Section 7.9, has a mean $E(N)E(X)$ and a variance $E(N)V(X) + V(N)E^2(X)$. Use probability-generating functions.

7.57 Let $X_1, X_2, \ldots$, denote a sequence of independent and identically distributed random variables with $P(X_i = 1) = p$ and $P(X_i = -1) = q = 1 - p$. If a value of 1 signifies success, then

$$S_n = X_1 + X_2 + \cdots + X_n$$

is the excess of successes over failures for n trials. A success could represent, for example, the completion of a sale or a win at a gambling device. Let $p_n^{(y)}$ equal the probability that S_n reaches the value y for the first time on the nth trial. We outline the steps for obtaining the probability-generating function $P(s)$ for $p_n^{(1)}$:

$$P(s) = \sum_{n=1}^{\infty} s_n p_n^{(1)}$$

a Let $P^{(2)}(s)$ denote the probability-generating function for $p_n^{(2)}$; that is,

$$P^{(2)}(s) = \sum_{n=1}^{\infty} s_n p_n^{(2)}.$$

Because a first accumulation of S_n to 2 represents two successive first accumulations to 1, reason that

$$P^{(2)}(s) = P^2(s).$$

b For $n > 1$, if $S_n = 1$ for the first time on trial n, this implies that S_n must have been at 2 on trial $n - 1$, and that a failure must have occurred on trial n. Because $p_n^{(1)} = qp_{n-1}^{(2)}$, show that

$$P(s) = ps + qsP^2(s).$$

c Solve the quadratic equation in part (b) to show that

$$P(s) = \frac{1}{2qs}[1 - (1 - 4pqs^2)^{1/2}].$$

d Show that

$$P(1) = \frac{1}{2q}(1 - |p - q|)$$

so the probability that S_n ever becomes positive equals 1 or p/q, whichever is smaller.

7.58 Refer to Exercise 7.57. Show that for the case $p = q - 1/2$, the expected number of trials until the first passage of S_n through 1 is infinite.

7.10 Summary

Many interesting applications of probability involve one or more functions of random variables, such as the sum, the average, the maximum value, and the range. Five methods for finding distributions of functions of random variables are outlined in this chapter. The *method of distribution functions* works well when the cumulative probability of the transformed variable can be calculated in closed form. If the transformed variable is a monotonic function of a univariate random variable, the *method of transformations* is appropriate and easy to apply. If the transformed variable is a function of more than one random variable, the *method of conditioning* or the *method of transformation* for more than one dimension might be the method of choice. The *method of moment-generating functions* is particularly useful for finding the distributions of sums of independent random variables.

Order statistics for independent random variables are of great practical utility, because it is often important to consider extreme values, medians, and ranges, among other statistical entities. A general form for the joint or marginal distribution of any set of order statistics is developed and used as a model in practical applications.

In finding the distribution of a function of random variables, one should bear in mind that there is no single best method for solving any one type of problem. In light of the preceding guidelines, one might try a number of approaches to see which one seems to be the most efficient.

Supplementary Exercises

7.59 Let X be a binomial random variable with parameters n and p.

a Find the distribution of the random variable $Y = n - X$.

b If X is the number of successes in n independent Bernoulli trials, what does Y represent?

7.60 Let X be a negative binomial random variable with parameters r and p.

a Find the probability function of $Y = X + r$.

b If X is the number of failures prior to the rth success in a sequence of independent Bernoulli trials, what does Y represent?

7.61 Let X_1 and X_2 be independent and uniformly distributed over the interval (0, 1). Find the probability density functions of the following variables.

a $Y_1 = X_1^2$ **b** $Y_2 = X_1/X_2$

c $Y_3 = -\ln(X_1X_2)$ **d** $Y_4 = X_1X_2$

7.62 Let X_1 and X_2 be independent Poisson random variables with mean λ_1 and λ_2, respectively.

a Find the probability function of $X_1 + X_2$.

b Find the conditional probability function of X_1, given that $X_1 + X_2 = m$.

7.63 Refer to Exercise 7.6.

a Find the expected values of Y_1, Y_2, and Y_3 directly (by using the density function of X, and not the density function of Y_1, Y_2, and Y_3).

b Find the expected values of Y_1, Y_2, and Y_3 by using the derived density functions for these random variables.

7.64 A parachutist wants to land at a target T but finds that she is equally likely to land at any point on a straight line (A, B) of which T is the midpoint. Find the probability density function of the distance between her landing point and the target. (*Hint*: Denote A by -1, B by $+1$, and T by 0. Then the parachutist's landing point has a coordinate X that is uniformly distributed between -1 and $+1$. The distance between X and T is $|X|$.)

7.65 Let X_1 denote the amount of a bulk item stocked by a supplier at the beginning of a day, and let X_2 denote the amount of that item sold during the day. Suppose that X_1 and X_2 have the joint density function

$$f(x) = \begin{cases} 2, & 0 \le x_2 \le x_1 \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Of interest to this supplier is the random variable $Y = X_1 - X_2$, which denotes the amount of the item left at the end of the day.

a Find the probability density function for Y.

b Find $E(Y)$.

c Find $V(Y)$.

7.66 An efficiency expert takes two independent measurements, X_1 and X_2, on the length of time required for workers to complete a certain task. Each measurement is assumed to have the density function given by

$$f(x) = \begin{cases} \frac{1}{4}xe^{-x/2}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for

$$Y = \frac{X_1 + X_2}{2},$$

the average of the times recorded by the efficiency expert.

7.67 The length of time that a certain machine operates without failure is denoted by X_1, and the length of time needed to repair a failure is denoted by X_2. After repair the machine is assumed to operate as if it were a new machine. X_1 and X_2 are independent, and each has the density function

$$f(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability density function for

$$Y = \frac{X_1}{X_1 + X_2}$$

which represents the proportion of time that the machine is in operation during any one operation-repair cycle.

7.68 Two sentries are sent to patrol a road that is 1 mile long. The sentries are sent to points chosen independently and at random along the road. Find the probability that the sentries will be less than $1/2$ mile apart when they reach their assigned posts.

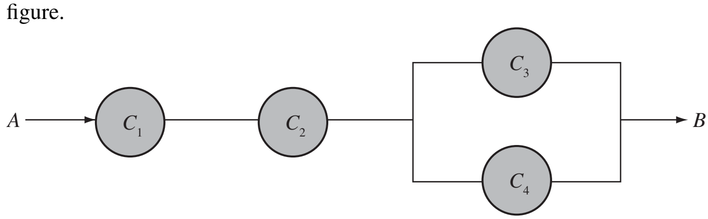
7.69 Let X_1 and X_2 be independent standard normal random variables. Find the probability density function of $Y_1 = X_1/X_2$, using:

a the method of conditioning.

b the transformation technique (*Hint*: Let $Y_2 = X_2$).

7.70 Let X be uniformly distributed over the interval $(-1, 3)$. Find the probability density function of $Y = X^2$.

7.71 If X denotes the life length of a component, and if $F(x)$ is the distribution function of X, then $P(X > x) = 1 - F(x)$ is called the reliability of the component. Suppose that a system consists of four components with identical reliability functions, $1 - F(x)$, operating as indicated in the following figure.



The system operates correctly if an unbroken chain of components is in operation between A and B. Assuming that the four components operate independently, find the reliability of the system in terms of $F(x)$.

7.72 Let $X_1, X_2, \ldots, X_n$ denote a random sample from the uniform distribution on (0, 1); that is, $f(x) = 1$, where $0 \le x \le 1$. Find the probability density function for the range $R = X_{(n)} - X_{(1)}$.

7.73 Suppose that U and V are independent random variables, with U having a standard normal distribution and V having a chi-squared distribution with n degrees of freedom. Define T by

$$T = \frac{U}{V/\sqrt{n}}.$$

Then T has a t (or Student's t) distribution, and the density function can be obtained as follows.

a If V is fixed at v, then T is given by U/c, where $c = \sqrt{v/n}$. Use this idea to find the conditional density of T for fixed $V = v$.

b Find the joint density of T and V—namely, $f(t, v)$—by using

$$f(t, v) = f(t|v)f(v).$$

c Integrate over v to show that

$$f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n}\Gamma(n/2)}(1 + t^2/n)^{-(n+1)/2}, \qquad -\infty < t < \infty.$$

7.74 Suppose that V and W are independent chi-squared random variables with n_1 and n_2 degrees of freedom, respectively. Then F, defined by

$$F = \frac{V/n_1}{W/n_2}$$

is said to have an F distribution with n_1 and n_2 degrees of freedom, and the density function can be obtained as follows.

a If W is fixed at w, then $F = V/c$, where $c = wn_1/n_2$. Find the conditional density of F for fixed $W = w$.

b Find the joint density of F and W.

c Integrate over w to show that the probability density function of F—namely $g(f)$—is given by

$$g(f) = \frac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n1/2}}{\Gamma(n_1/2)\Gamma(n_2/2)}(f)^{(n_1/2)-1}\left(1 + \frac{n_1 f}{n_2}\right)^{-(n_1+n_2)/2}, \qquad 0 < f < \infty.$$

7.75 An object is to be dropped at point 0 in the plane but lands at point (X, Y) instead, where X and Y denote horizontal and vertical distances, respectively, on a coordinate system centered at 0. If X and Y are independent, normally distributed random variables, each with a mean of 0 and a variance of σ^2, find the distribution of the distances between the landing point (X, Y) and 0. (The resulting distribution is called the *Rayleigh distribution*.)

7.76 Suppose that n electronic components, each having an exponentially distributed life length with a mean of θ, are put into operation at the same time. The components operate independently and are observed until r of them have failed $(r \le n)$. Let W_j denote the length of time until the jth failure $(W_1 \le W_2 \le \cdots \le W_n)$. Let $T_j = W_j - W_{j-1}$ for $j \ge 2$ and $T_1 = W_1$.

a Show that T_j, for $j = 1, 2, \ldots, r$, has an exponential distribution with mean $\theta/(n - j + 1)$.

b Show that $U_r = \sum_{j=1}^{r} W_j + (n-1)W_r = \sum_{j=1}^{r}(n - j + 1)T_j$; and, hence, that $E(U_r) = r\theta$. [This suggests that $(1/r)U_r$ can be used as an estimator of θ.]

7.77 A machine produces spherical containers whose radii vary according to the probability density function

$$f(r) = \begin{cases} 2r, & 0 \le r \le 1 \\ 0, & \text{otherwise.} \end{cases}$$

Find the probability density function for the volume of the containers.

7.78 Let V denote the velocity of a molecule having mass m in a uniform gas at equilibrium. The probability density function of V is known to be

$$f(v) = \begin{cases} \frac{4}{\sqrt{\pi}} b^{3/2} v^2 e^{-bv^2}, & v > 0 \\ 0, & \text{otherwise} \end{cases}$$

where $b = m/2KT$; in the latter equation, T denotes the absolute temperature of the gas, and K represents Boltzmann's constant. Find the probability density function for the kinetic energy E, given by $E = (1/2)mV^2$.

7.79 Suppose that members of a certain animal species are randomly distributed over a planar area, so that the number found in a randomly selected quadrant of unit area has a Poisson distribution with mean λ. Over a specified period, each animal has a probability θ of dying, and deaths are assumed to be independent from animal to animal. Ignoring births and other population changes, what is the distribution of animals at the end of the time period?

8

Some Approximations to Probability Distributions: Limit Theorems

8.1 Introduction

As noted in Chapter 7, we frequently are interested in functions of random variables, such as their average or their sum. Unfortunately, the application of the methods in Chapter 7 for finding probability distributions for functions of random variables may lead to intractable, mathematical problems. Hence, we need some simple methods for approximating the probability distributions of functions of random variables.

In this chapter, we discuss properties of functions of random variables when the number of variables n gets large (approaches infinity). We shall see, for example, that the distribution for certain functions of random variables can easily be approximated for large n, even though the exact distribution for fixed n may be difficult to obtain. Even more importantly, the approximations are sometimes good for samples of modest size and, in some instances, for samples as small as $n = 5$ or 6. In Section 8.2, we present some extremely useful theorems, called limit theorems, that give properties of random variables as n tends toward infinity.

8.2 Convergence in Probability

Suppose that a coin has probability p, with $0 \leq p \leq 1$, of coming up heads on a single flip, and suppose that we flip the coin n times. What can be said about the fraction of heads observed in the n flips? Intuition tells us that the sampled fraction of heads provides an estimate of p, and we would expect the estimate to fall closer to p for larger sample sizes—that is, as the quantity of information in the sample increases. Although our supposition and intuition may be correct for many problems of estimation, it is not always true that larger sample sizes lead to better estimates. Hence, this example gives rise to a question that occurs in all estimation problems: What can be said about the random distance between an estimate and its target parameter?

Notationally, let X denote the number of heads observed in the n tosses. Then $E(X) = np$ and $V(X) = np(1-p)$. One way to measure the closeness of X/n to p is to ascertain the probability that the distance $|(X/n)-p|$ will be less than a preassigned real number ε. This probability

$$P\left(\left|\frac{X}{n} - p\right| < \varepsilon\right)$$

should be close to unity for larger n, if our intuition is correct. Definition 8.1 formalizes this convergence concept.

DEFINITION 8.1

The sequence of random variables $X_1, X_2, \ldots, X_n$, is said to **converge in probability** to the constant c if, for every positive number ε,

$$\lim_{n\to\infty} P\left(\left|X_n - c\right| < \varepsilon\right) = 1. \quad \blacksquare$$

Theorem 8.1 often provides a mechanism for proving **convergence in probability**. We now apply Theorem 8.1 to our coin-tossing example.

EXAMPLE 8.1 Let X be a binomial random variable with probability of success p and number of trials n. Show that X/n converges in probability toward p.

Solution We have seen that we can write X as $\sum_{i=1}^{n} Y_i$, where $Y_i = 1$ if the ith trial results in success, and $Y_i = 0$ otherwise. Then

$$\frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i.$$

Also, $E(Y_i) = p$ and $V(Y_i) = p(1-p)$. The conditions of Theorem 8.1 are then fulfilled with $\mu = p$ and $o^2 = p(1-p)$, and we conclude that, for any possible ε,

$$\lim_{n\to\infty} P\left(\left|\frac{X}{n} - p\right| \geq \varepsilon\right) = 0. \quad \blacksquare$$

Theorem 8.1, sometimes called the (*weak*) *law of large numbers*, is the theoretical justification for the averaging process employed by many experimenters to obtain precision in measurements. For example, an experimenter may take the average of five measurements of the weight of an animal to obtain a more precise estimate of the animal's weight. The experimenter's feeling—a feeling borne out by Theorem 8.1—is that the average of a number of independently selected weights has a high probability of being quite close to the true weight.

Like the law of large numbers, the theory of convergence in probability has many applications. Theorem 8.2, which is presented without proof, points out some properties of the concept of convergence in probability.

THEOREM **8.1**

Weak Law of Large Numbers: Let $X_1, X_2, \dots, X_n$ be independent and identically distributed random variables, with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$. Let $\bar{X}_n = (1/n)\sum_{i=1}^{n} X_i$. Then, for any positive real number ε,

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

or

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1.$$

Thus, $\bar{X}_n$ converges in probability toward μ.

Proof

Notice that $E(\bar{X}_n) = \mu$ and $V(\bar{X}_n) = \sigma^2/n$. To prove the theorem, we appeal to Tchebysheff's theorem (see Section 4.2 or Section 5.2), which states that

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

where $E(X) = \mu$ and $V(X) = \sigma^2$. In the context of our theorem, X is to be replaced by $\bar{X}_n$, and σ^2 is to be replaced by σ^2/n. It then follows that

$$P\left(|\bar{X}_n - \mu| \geq k\frac{\sigma}{\sqrt{n}}\right) \leq \frac{1}{k^2}.$$

Notice that k can be any real number, so we shall choose

$$k = \frac{\varepsilon}{\sigma}\sqrt{n}.$$

Then,

$$P\left(|\bar{X}_n - \mu| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\left(\frac{\sigma}{\sqrt{n}}\right)\right) \leq \frac{\sigma^2}{\varepsilon^2 n}$$

or

$$P\left(|\bar{X}_n - \mu| \geq \varepsilon\right) \leq \frac{\sigma^2}{\varepsilon^2 n}.$$

Now let us take the limit of this expression as n tends toward infinity. Recall that σ^2 is finite and that ε is a positive real number. On taking the limit as n tends toward infinity, we have

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0.$$

The conclusion that $\lim_{n\to\infty} P(|\bar{X}_n - \mu| < \varepsilon) = 1$ follows directly because

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = 1 - P(|\bar{X}_n - \mu| < \varepsilon). \quad \blacksquare$$

THEOREM 8.2 Suppose that X_n converges in probability toward μ_1 and Y_n converges in probability toward μ_2. Then the following statements are also true.

1. $X_n + Y_n$ converges in probability toward $\mu_1 + \mu_2$.
2. $X_n Y_n$ converges in probability toward $\mu_1 \mu_2$.
3. X_n / Y_n converges in probability toward μ_1/μ_2, provided that $\mu_2 \neq 0$.
4. $\sqrt{X_n}$ converges in probability toward $\sqrt{\mu_1}$, provided that $P(X_n \geq 0) = 1$. $\blacksquare$

EXAMPLE 8.2 Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with $E(X_i) = \mu$, $E(X_i^2) = \mu_2'$, $E(X_i^3) = \mu_3'$, and $E(X_i^4) = \mu_4'$, all assumed finite. Let S'^2 denote the sample variance given by

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Show that S'^2 converges in probability to $V(X_i)$.

Solution First, notice that

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

where

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

To show that S'^2 converges in probability to $V(X_i)$, we apply both Theorems 8.1 and 8.2. Look at the terms in S'^2. The quantity $(1/n)\sum_{i=1}^{n} X_i^2$ is the average of n independent and identically distributed variables of the form X_i^2 with $E(X_i^2) = \mu_2'$

and $V(X_i^2) = \mu_4' - (\mu_2')^2$. Because $V(X_i^2)$ is assumed to be finite, Theorem 8.1 tells us that $(1/n)\sum_{i=1}^n X_i^2$ converges in probability toward μ_2'. Now consider the limit of $\bar{X}^2$ as n approaches infinity. Theorem 8.1 tells us that $\bar{X}$ converges in probability toward μ; and it follows from Theorem 8.2, part (b), that $\bar{X}^2$ converges in probability toward μ^2. Having shown that $(1/n)\sum_{i=1}^n X_i^2$ and $\bar{X}^2$ converge in probability toward μ_2' and μ^2, respectively, we can conclude from Theorem 8.2 that

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

converges in probability toward $\mu_2' - \mu^2 = V(X_i)$.

This example shows that, for large samples, the sample variance has a high probability of being close to the population variance. ▪

8.3 Convergence in Distribution

In Section 8.2, we dealt only with the convergence of certain random variables toward constants and said nothing about the form of the probability distributions. In this section, we look at what happens to the probability distributions of certain types of random variables as n tends toward infinity. We need the following definition before proceeding.

DEFINITION **8.2**

Let X_n be a random variable with distribution function $F_n(x)$. Let X be a random variable with distribution function $F(x)$. If

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

at every point x for which $F(x)$ is continuous, then X_n is said to **converge in distribution** toward X, $F(x)$ is called the *limiting distribution function* of X_n. ▪

EXAMPLE **8.3**

Let $X_1, X_2, \ldots, X_n$ be independent uniform random variables over the interval $(\theta, 0)$ for a negative constant θ. In addition, let $Y_n = \min(X_1, X_2, \ldots, X_n)$. Find the limiting distribution of Y_n.

Solution The distribution function for the uniform random variable X_i is

$$F(X_i) = P(X_i \le x) = \begin{cases} 0, & x < \theta \\ \frac{x-\theta}{-\theta}, & \theta \le x \le 0 \\ 1, & x > 0. \end{cases}$$

In Section 7.8, we found that the distribution function for Y_n is

$$\begin{aligned} G(y) = P(Y_n \le y) &= 1 - \left[1 - F_X(y)\right]^n \\ &= \begin{cases} 0, & y < \theta \\ 1 - \left(1 - \frac{y-\theta}{-\theta}\right)^n = 1 - \left(\frac{y}{\theta}\right)^n, & \theta \le y \le 0 \\ 1, & y > 0. \end{cases} \end{aligned}$$

where $F_X(y)$ is the distribution function for X_i evaluated at y. Then

$$\begin{aligned} \lim_{n\to\infty} G(y) &= \begin{cases} 0, & y < \theta \\ \lim_{n\to\infty} \left[1 - \left(\frac{y}{\theta}\right)^n\right], & \theta \le y \le 0 \\ 1, & y > 0 \end{cases} \\ &= \begin{cases} 0, & y < \theta \\ 1, & y \ge \theta. \end{cases} \end{aligned}$$

Thus, Y_n **converges in distribution** toward a random variable that has a probability of 1 at the point θ and a probability of 0 elsewhere. ▪

It is often easier to find limiting distributions by working with moment-generating functions. Theorem 8.3 gives the relationship between convergence of distribution functions and convergence of moment-generating functions.

THEOREM **8.3**

Let X_n and X be random variables with moment-generating functions $M_n(t)$ and $M(t)$, respectively. If

$$\lim_{n\to\infty} M_n(t) = M(t)$$

for all real t, then X_n converges in distribution toward X.

The proof of Theorem 8.3 is beyond the scope of this text. ▪

EXAMPLE **8.4**

Let X_n be a binomial random variable with n trials and probability p of success on each trial. If n tends toward infinity and p tends toward zero with np remaining fixed, show that X_n converges in distribution toward a Poisson random variable.

Solution This problem was solved in Chapter 4, when we derived the Poisson probability distribution. We now solve it by using moment-generating functions and Theorem 8.3.

We know that the moment-generating function of X_n—namely, $M_n(t)$—is given by

$$M_n(t) = (q + pe^t)^n$$

where $q = 1 - p$. This can be rewritten as

$$M_n(t) = [1 + p(e^t - 1)]^n.$$

Letting $np = \lambda$ and substituting into $M_n(t)$, we obtain

$$M_n(t) = \left[1 + \frac{\lambda}{n}(e^t - 1)\right]^n.$$

Now, let us take the limit of this expression as n approaches infinity.

From calculus, you may recall that

$$\lim_{n\to\infty}\left(1 + \frac{k}{n}\right)^n = e^k.$$

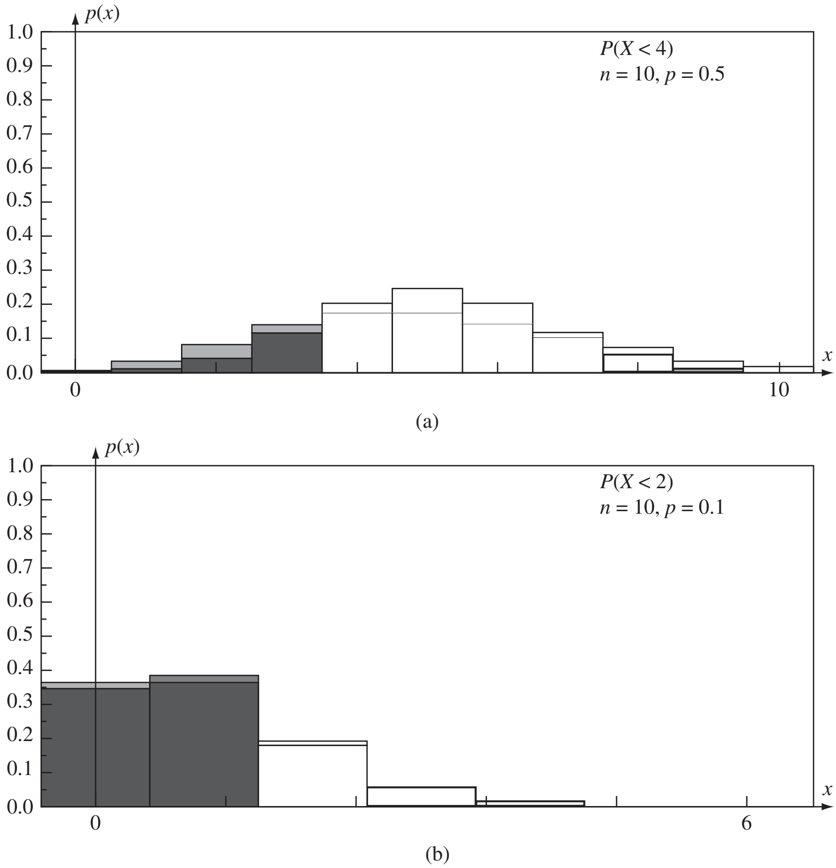


FIGURE **8.1** The Poisson approximation to the binomial distribution.

Letting $k = \lambda(e^t - 1)$, we have

$$\lim_{n\to\infty} M_n(t) = \exp[\lambda(e^t - 1)].$$

We recognize the right-hand expression as the moment-generating function for the Poisson random variable. Hence, it follows from Theorem 8.3 that X_n converges in distribution toward a Poisson random variable.

From the preceding discussion as well as the derivation of the Poisson distribution in Chapter 4, it is no surprise that the Poisson distribution is often used to approximate binomial probabilities, especially when n is large and p is small. The *Approximations to Distributions* applet allows one to explore how well the Poisson distribution approximates binomial probabilities over a range of possible parameter values. As an example, if X is a binomial random variable with $n = 10$ and $p = 0.5$, the probability that a randomly observed value of X is less than 4 is 0.17187, and the Poisson approximation of this probability is 0.26502 (Figure 8.1a). If instead X is a binomial random variable with $n = 10$ and $p = 0.1$, the Poisson approximation of the probability that a randomly observed value of X is less than two is 0.73575, which is close to the true probability of 0.73609 (Figure 8.1b). ▪

EXAMPLE 8.5 In monitoring for pollution, an experimenter collects a small volume of water and counts the number of bacteria in the sample. Unlike earlier problems, we have only one observation. For purposes of approximating the probability distribution of counts, we can think of the volume as the quantity that is getting large.

Let X denote the bacteria count per cubic centimeter of water, and assume that X has a Poisson probability distribution with mean λ. We want to approximate the probability distribution of X for large values of λ, which we do by showing that

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}$$

converges in distribution toward a standard normal random variable as λ tends toward infinity.

Specifically, if the allowable pollution in a water supply is a count of 110 bacteria per cubic centimeter, approximate the probability that X will be at most 110, assuming that $\lambda = 100$.

Solution We proceed by taking the limit of the moment-generating function of Y as $\lambda \to \infty$ and then using Theorem 8.3. The moment-generating function of X—namely, $M_X(t)$—is given by

$$M_X(t) = e^{\lambda(e^t - 1)}$$

and hence the moment-generating function for Y—namely, $M_Y(t)$—is

$$\begin{aligned} M_Y(t) &= e^{-t\sqrt{\lambda}} M_X\left(\frac{t}{\sqrt{\lambda}}\right) \\ &= e^{-t\sqrt{\lambda}} \exp\left[\lambda(e^{t/\sqrt{\lambda}} - 1)\right]. \end{aligned}$$

The term $e^{t/\sqrt{\lambda}} - 1$ can be written as

$$e^{t/\sqrt{\lambda}} - 1 = \frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{6\lambda^{3/2}} + \cdots$$

Thus, on adding exponents, we have

$$\begin{aligned} M_Y(t) &= \exp\left[-t\sqrt{\lambda} + \lambda\left(\frac{t}{\sqrt{\lambda}} + \frac{t^2}{2\lambda} + \frac{t^3}{6\lambda^{3/2}} + \cdots\right)\right] \\ &= \exp\left(\frac{t^2}{2} + \frac{t^3}{6\sqrt{\lambda}} + \cdots\right). \end{aligned}$$

In the exponent of $M_Y(t)$, the first term $(t^2/2)$ is free of λ, and the remaining terms all have a λ to some positive power in the denominator. Therefore, as $\lambda \to \infty$, all the terms after the first will tend toward zero sufficiently quickly to allow

$$\lim_{n\to\infty} M_Y(t) = e^{t^2/2}$$

and the right-hand expression is the moment-generating function for a standard normal random variable. We now want to approximate $P(X \le 110)$. Notice that

$$P(X \le 110) = P\left(\frac{X - \lambda}{\sqrt{\lambda}} \le \frac{110 - \lambda}{\sqrt{\lambda}}\right).$$

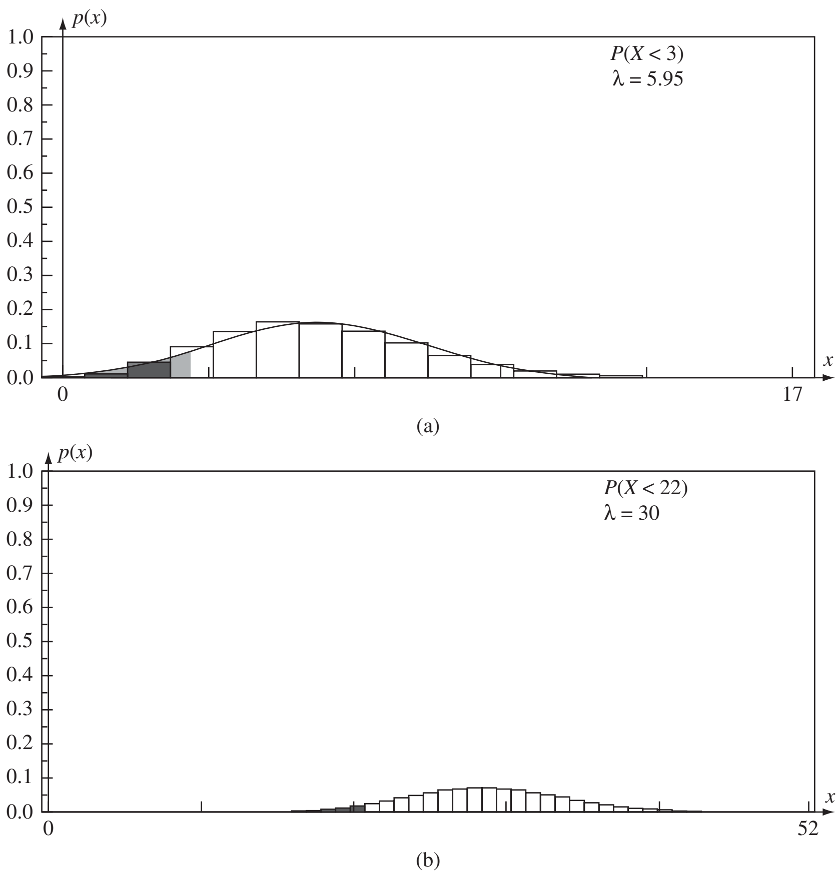
We have shown that $Y = (X - \lambda)/\sqrt{\lambda}$ is approximately a standard normal random variable for large λ. Hence, for $\lambda = 100$, we have

$$\begin{aligned} P\left(Y \le \frac{110 - 100}{10}\right) &= P(Y \le 1) \\ &= 0.8413 \end{aligned}$$

from either Table 4 in the Appendix or the *Continuous Distributions* applet. ▪

The normal approximation to Poisson probabilities improves as λ increases, and it works reasonably well for $\lambda \ge 25$. To see this, use the *Approximations of Distributions* applet to compare the exact Poisson probabilities to the normal approximations of those probabilities. As an example, if X has a Poisson distribution with mean $\lambda = 5.95$, the probability that a random observation from this distribution is less than 3 is 0.06423. The normal approximation to this probability is 0.11325 (Figure 8.2a), an error of more than 75%. In contrast, if X has a Poisson distribution with mean $\lambda = 30$, the probability that a random observation from this distribution is less than 22 is 0.05444; the normal approximation of the probability is 0.07206, giving an error of about 32% (Figure 8.2b). The approximation continues to improve as λ increases. Estimating the tail probabilities precisely requires a larger lambda than estimation in the center of the distribution. To illustrate, for $\lambda = 5.95$, the exact probability of obtaining a randomly observed value greater than 5 is 0.54625, and the normal approximation to this error is 0.65153, an error of about 19%, which is much smaller than the 75% error observed for the tail probability. For the larger mean of $\lambda = 30$, the normal approximation of $P(X > 28) = 0.58633$ is 0.63254, an error of less than 8%.

FIGURE 8.2 Normal approximation to the Poisson distribution.



Exercises

8.1 Let $X_1, X_2, \ldots, X_n$ be independent random variables, each with the probability density function

$$f(x) = \begin{cases} 2(1-x), & 0 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges in probability toward a constant as $n \to \infty$, and find the constant.

8.2 Let $X_1, X_2, \ldots, X_n$ be independent random variables, each with the probability density function

$$f(x) = \begin{cases} \frac{3}{2}x^2, & -1 \le x \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

converges in probability toward a constant as $n \to \infty$, and find the constant.

8.3 Let $X_1, X_2, \ldots, X_m$ be independent binomial random variables, each of whose density function is given by

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, 2, \ldots, n \\ 0, & \text{elsewhere.} \end{cases}$$

Show that the mean $\bar{X}$ converges in probability toward a constant as $m \to \infty$, and find the constant.

8.4 Let $X_1, X_2, \ldots, X_n$ be independent Poisson random variables, each possessing the density function given by

$$f(x) = \begin{cases} \frac{\lambda^x}{x!} e^{-\lambda}, & x = 0, 1, 2, \ldots \\ 0, & \text{elsewhere.} \end{cases}$$

Show that the mean $\bar{X}$ converges in probability toward a constant as $n \to \infty$, and find the constant.

8.5 Let $X_1, X_2, \ldots, X_n$ be independent gamma random variables, each possessing the density function given by

$$f(x) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, & x \ge 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that the mean $\bar{X}$ converges in probability toward a constant as $n \to \infty$, and find the constant.

8.6 Let $X_1, X_2, \ldots, X_n$ be independent random variables, each uniformly distributed over the interval $(0, \theta)$.

a Show that the mean $\bar{X}$ converges in probability toward a constant as $n \to \infty$, and find the constant.

b Show that $\max(X_1, X_2, \ldots, X_n)$ converges in probability toward θ as $n \to \infty$.

8.7 Let $X_1, X_2, \ldots, X_n$ be independent uniform random variables over the interval $(0, \theta)$ for a positive constant θ. In addition, let $Y_n = \max(X_1, X_2, \ldots, X_n)$. Find the limiting distribution of Y_n.

8.8 Let $X_1, X_2, \ldots, X_n$ be independent random variables, each possessing the density function given by

$$f(x) = \begin{cases} \frac{2}{x^2}, & x \geq 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Does the law of large numbers apply to $\bar{X}$ in this case? If so, find the limit in the probability of $\bar{X}$.

8.9 Suppose that the probability that a person will suffer a bad reaction from an injection of a certain serum is 0.001. We want to determine the probability that two or more will suffer a bad reaction if 1000 persons receive the serum.

a Justify using the Poisson approximation to the binomial probability for this application.

b Approximate the probability of interest.

8.10 The number of accidents per year X at a given intersection is assumed to have a Poisson distribution. Over the past few years, an average number of 32 accidents per year has occurred at this intersection. If the number of accidents per year is at least 40, an intersection can qualify to be rebuilt under an emergency program set up by the state. Approximate the probability that the intersection in question will qualify under the emergency program at the end of next year.

8.4 The Central Limit Theorem

Example 8.5 gives a random variable that converges in distribution toward the standard normal random variable. That this phenomenon is shared by a large class of random variables is shown in Theorem 8.4.

Another way to say that a random variable converges in distribution toward a standard normal is to say that it is asymptotically normal. It is noteworthy that Theorem 8.4 is not the most general form of the Central Limit Theorem. Similar theorems exist for certain cases in which the values of X_i are not identically distributed and in which they are dependent.

The probability distribution that arises from looking at many independent values of $\bar{X}$, for a fixed sample size, selected from the same population is called the *sampling distribution of* $\bar{X}$. The practical importance of the Central Limit Theorem is that for large n, the sampling distribution of $\bar{X}$ can be closely approximated by a normal distribution. More precisely,

$$\begin{aligned} P(\bar{X} \leq b) &= P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{b - \mu}{\sigma/\sqrt{n}}\right) \\ &= P\left(Z \leq \frac{b - \mu}{\sigma/\sqrt{n}}\right) \end{aligned}$$

where Z is a standard normal random variable.

THEOREM **8.4**

The Central Limit Theorem. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$. Define Y_n as

$$Y_n = \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right)$$

where

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Then Y_n converges in distribution toward a standard normal random variable.

Proof

We sketch a proof for the case in which the moment-generating function for X_i exists. (This is not the most general proof, because moment-generating functions do not always exist.)

Define a random variable Z_i by

$$Z_i = \frac{X_i - \mu}{\sigma}.$$

Notice that $E(Z_i) = 0$ and $V(Z_i) = 1$. The moment-generating function of Z_i—namely, $M_Z(t)$—can be written as

$$M_Z(t) = 1 + \frac{t^2}{2!} + \frac{t^3}{3!}E(Z_i^3) + \cdots$$

Now,

$$\begin{aligned}
Y_n &= \sqrt{n}\left(\frac{\bar{X}_n - \mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{n}}\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma}\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_i
\end{aligned}$$

and the moment-generating function of Y_n—namely $M_n(t)$—can be written as

$$M_n(t) = \left[M_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n.$$

Recall that the moment-generating function of the sum of independent random variables is the product of their individual moment-generating functions. Hence,

$$\begin{aligned}
M_n(t) &= \left[M_Z\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \left(1 + \frac{t^2}{2n} + \frac{t^3}{3!\,n^{3/2}}k + \cdots\right)^n
\end{aligned}$$

where $k = E(Z_i^3)$.

Now take the limit of $M_n(t)$ as $n \to \infty$. One way to evaluate the limit is to consider $\ln M_n(t)$ where

$$\ln M_n(t) = n \ln\left[1 + \left(\frac{t^2}{2n} + \frac{t^3 k}{6n^{3/2}} + \cdots\right)\right].$$

A standard series expansion for $\ln(1 + x)$ is

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

Let

$$x = \left(\frac{t^2}{2n} + \frac{t^3 k}{6n^{3/2}} + \cdots\right).$$

Then ,

$$\begin{aligned}
\ln M_n(t) &= n \ln(1 + x) \\
&= n\left(x - \frac{x^2}{2} + \cdots\right) \\
&= n\left[\left(\frac{t^2}{2n} + \frac{t^3 k}{6n^{3/2}} + \cdots\right) - \frac{1}{2}\left(\frac{t^2}{2n} + \frac{t^3 k}{6n^{3/2}} + \cdots\right)^n + \cdots\right]
\end{aligned}$$

where the succeeding terms in the expansion involve x^2, x^4, and so on. Multiplying through by n, we see that the first term, $t^2/2$, does not involve n, whereas all other terms have n to a positive power in the denominator. Thus,

$$\lim_{n\to\infty} \ln M_n(t) = \frac{t^2}{2}$$

or

$$\lim_{n\to\infty} M_n(t) = e^{t^2/2}$$

which is the moment-generating function for a standard normal random variable. Applying Theorem 8.3, we conclude that Y_n converges in distribution toward a standard normal random variable. ▪

We can observe an approximate sampling distribution of $\bar{X}$ by looking at the following results from a computer simulation. Samples of size n were drawn from a population having the probability density function

$$f(x) = \begin{cases} \frac{1}{10}e^{-x/10}, & x > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

The sample mean was then computed for each sample. The relative frequency histogram of these mean values for 1000 samples of size $n = 5$ is shown in Figure 8.3. Figures 8.4 and 8.5 show similar results for 1000 samples of size $n = 25$ and $n = 100$, respectively. Although all the relative frequency histograms have roughly a bell shape, the tendency toward a symmetric normal curve becomes stronger as n increases. A smooth curve drawn through the bar graph of Figure 8.5 would be nearly identical to a normal density function with a mean of 10 and a variance of $(10)^2/100 = 1$.

FIGURE **8.3**
Relative frequency histogram for $\bar{x}$ from 1000 samples of size $n = 5$.

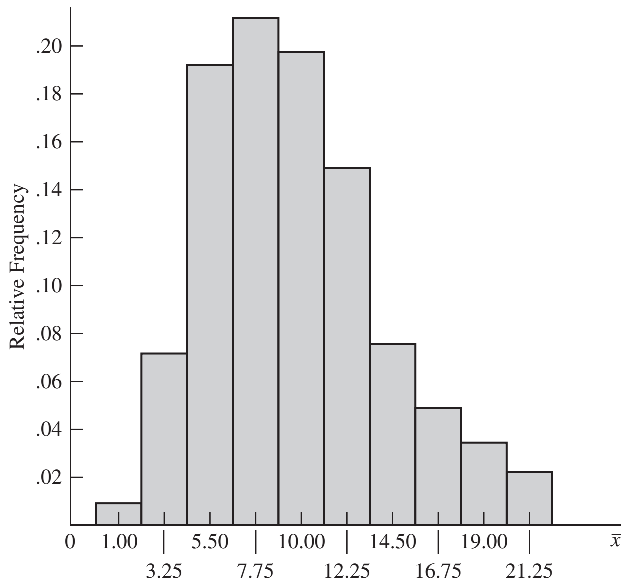
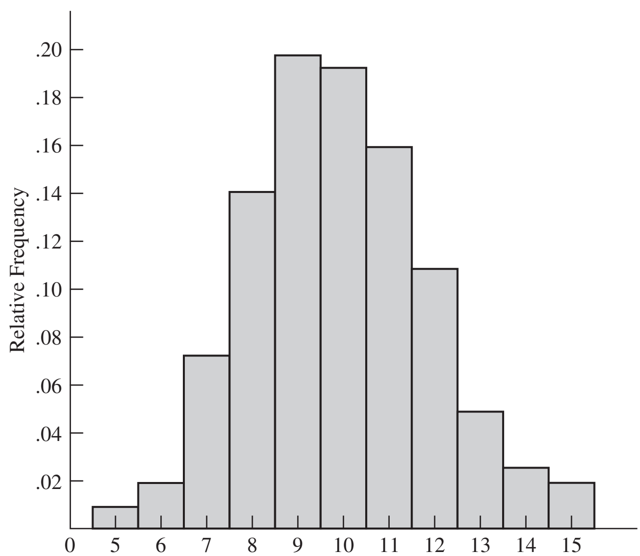


FIGURE **8.4**
Relative frequency histogram for $\bar{x}$ from 1000 samples of size $n = 25$.



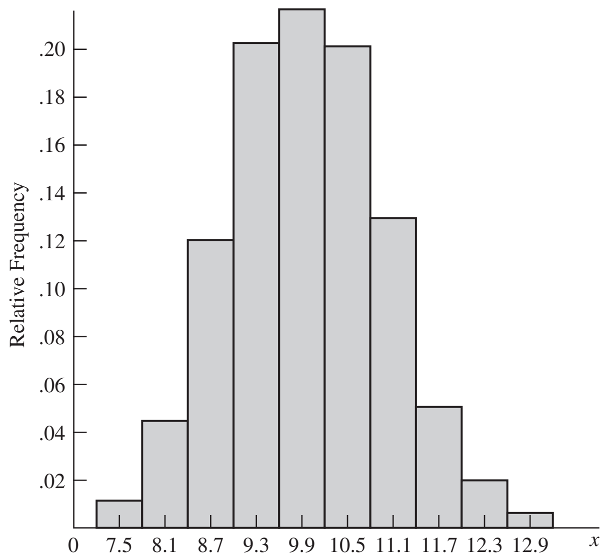


FIGURE 8.5
Relative frequency histogram for $\bar{x}$ from 1000 samples of size $n = 1000$.

The Central Limit Theorem provides a very useful result for statistical inference, because it enables us to know not only that $\bar{X}$ has mean μ and variance σ^2/n if the population has mean μ and variance σ^2, but also that the probability distribution for $\bar{X}$ is approximately normal. For example, suppose that we wish to find an interval, (a, b), such that

$$P(a \le \bar{X} \le b) = 0.95.$$

This probability is equivalent to

$$P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \le \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \le \frac{b-\mu}{\sigma/\sqrt{n}}\right) = 0.95$$

for constants μ and σ. Because $(\bar{X}-\mu)/(\sigma/\sqrt{n})$ has approximately a standard normal distribution, the equality can be approximated by

$$P\left(\frac{a-\mu}{\sigma/\sqrt{n}} \le Z \le \frac{b-\mu}{\sigma/\sqrt{n}}\right) = 0.95$$

where Z has a standard normal distribution. From Table 4 in the Appendix or the *Continuous Distributions* applet, we know that

$$P(-1.96 \le Z \le 1.96) = 0.95$$

and, hence,

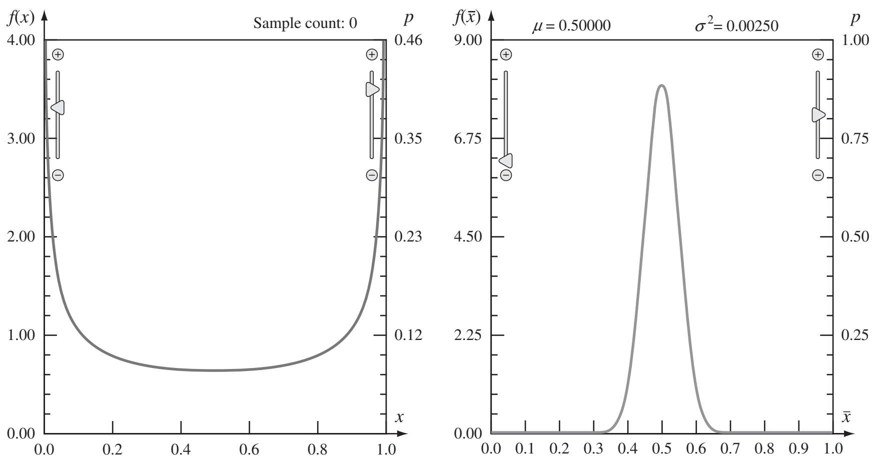
$$\frac{a-\mu}{\sigma/\sqrt{n}} = -1.96 \qquad \frac{b-\mu}{\sigma/\sqrt{n}} = 1.96$$

or

$$a = \mu - \frac{1.96\sigma}{\sqrt{n}} \qquad b = \mu + \frac{1.96\sigma}{\sqrt{n}}.$$

The *Central Limit Theorem* applet can be used to explore the Central Limit Theorem for a variety of distributions. As an example, samples of size 50 were drawn from a beta distribution with parameters $\alpha = \beta = 0.5$. The probability density function of the beta distribution is shown in Figure 8.6 (left), and the sampling distribution of the sample means is displayed in Figure 8.6 (right).

FIGURE **8.6**
The beta probability density function with $\alpha = \beta = 0.5$ (left) and the sampling distribution of sample means based on samples of size 50 drawn from that distribution (right).



EXAMPLE **8.6** From 1976 to 2002, a mechanical golfer, Iron Byron, whose swing was modeled after that of Byron Nelson (a leading golfer in the 1940s), was used to determine whether golf balls met the Overall Distance Standard. Specifically, Iron Byron would be used to hit the golf balls. If the average distance of 24 golf balls tested exceeded 296.8 yards, then that brand would be considered nonconforming. Under these rules, suppose a manufacturer produces a new golf ball that travels an average distance of 297.5 yards with a standard deviation of 10 yards.

1 What is the probability that the ball will be determined to be nonconforming when tested?

2 Find an interval that includes the average overall distance of 24 golf balls with a probability of 0.95.

Solution **1** We assume that $n = 24$ is large enough for the sample mean to have an approximate normal distribution. The average overall distance $\bar{X}$ has a mean of $\mu = 297.5$ and a standard deviation of

$$\frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{24}} = 2.04.$$

Thus,

$$P(\bar{X} > 296.8) = P\left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} > \frac{296.8 - \mu}{\sigma/\sqrt{n}}\right)$$

is approximately equal to

$$P\left(Z > \frac{296.8 - 297.5}{10/\sqrt{24}}\right) = P\left(Z > \frac{-0.7}{2.04}\right)$$
$$= P(Z > -0.34)$$
$$= 0.1331 + 0.5$$
$$= 0.6331$$

from the normal probability table, Table 4, or from the *Continuous Distributions* applet. For this manufacturer, although there is a significant probability that the new golf ball will meet the standard, the probability of having it declared nonconforming is a little less than $2/3$.

2 We have seen that

$$P\left[\mu - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) \leq \bar{X} \leq \mu + 1.96\left(\frac{\sigma}{\sqrt{n}}\right)\right] = 0.95$$

for a normally distributed $\bar{X}$. In this problem,

$$\mu - 1.96\left(\frac{\sigma}{\sqrt{n}}\right) = 297.5 - 1.96\left(\frac{10}{\sqrt{24}}\right) = 293.5$$

and

$$\mu + 1.96\left(\frac{\sigma}{\sqrt{n}}\right) = 297.5 + 1.96\left(\frac{10}{\sqrt{24}}\right) = 301.5.$$

Approximately 95% of sample mean overall distances, for samples of size 24, should be between 293.5 and 301.5 yards. ▪

EXAMPLE **8.7** A certain machine that is used to fill bottles with liquid has been observed over a long period, and the variance in the amounts of fill has been found to be approximately $\sigma^2 = 1$ ounce. The mean ounces of fill μ, however, depends on an adjustment that may change from day to day or from operator to operator. If $n = 36$ observations on ounces of fill dispensed are to be taken on a given day (all with the same machine setting), find the probability that the sample mean will be within 0.3 ounce of the true population mean for that setting.

Solution We shall assume that $n = 36$ is large enough for the sample mean $\bar{X}$ to have approximately a normal distribution. Then

$$\begin{aligned}
P(|\bar{X} - \mu| \leq 0.3) &= P[-0.3 \leq (\bar{X} - \mu) \leq 0.3] \\
&= P\left[-\frac{0.3}{\sigma/\sqrt{n}} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \frac{0.3}{\sigma/\sqrt{n}}\right] \\
&= P\left[-0.3\sqrt{36} \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 0.3\sqrt{36}\right] \\
&= P\left[-1.8 \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq 1.8\right].
\end{aligned}$$

Because $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ has approximately a standard normal distribution, the preceding probability is approximately

$$P(| - 1.8 \leq Z \leq 1.8) = 2(0.4641) = 0.9282. \quad ■$$

We have seen in Chapter 4 that a binomially distributed random variable X can be written as a sum of independent Bernoulli random variables Y_i. Symbolically,

$$X = \sum_{i=1}^{n} Y_i$$

where $Y_i = 1$ with probability p and $Y_i = 0$ with probability $1-p$, for $i = 1, 2, \ldots, n$. X can represent the number of successes in a sample of n trials or measurements, such as the number of thermistors conforming to standards in a sample of n thermistors.

Now the fraction of successes in n trials is

$$\frac{X}{n} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$$

so X/n is a sample mean. In particular, for large n, we can affirm that X/n has approximately a normal distribution with a mean of

$$E\left(\frac{Y}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = p$$

and a variance of

$$V\left(\frac{Y}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} V(X_i)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} p(1-p)$$
$$= \frac{p(1-p)}{n}.$$

The normality follows from the Central Limit Theorem. Because $X = n\bar{Y}$, we know that X has approximately a normal distribution with a mean of np and a variance of $np(1-p)$. Because binomial probabilities are cumbersome to calculate for a large n, we make extensive use of this normal approximation to the binomial distribution.

Figure 8.7 shows the histogram of a binomial distribution for $n = 20$ and $p = 0.6$. The heights of the bars represent the respective binomial probabilities. For this distribution, the mean is $np = 20(0.6) = 12$, and the variance is $np(1-p) = 20(0.6)(0.4) = 4.8$. Superimposed on the binomial distribution is a normal distribution with mean $\mu = 2$ and variance $\sigma^2 = 4.8$. Notice how closely the normal curve approximates the binomial distribution. Use the *Approximations to Distributions* applet to explore how well the normal approximates the binomial distribution for a range of n and p values.

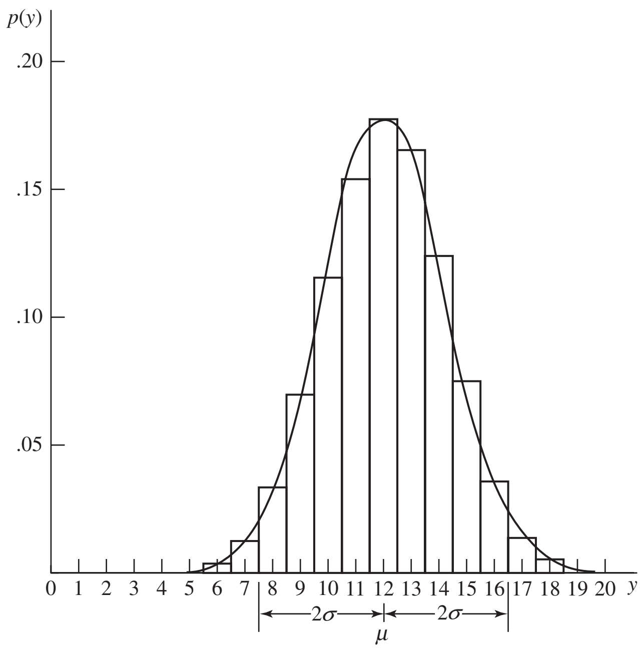


FIGURE **8.7** A binomial distribution $n = 20$, $p = 0.6$, and a normal distribution $\mu = 12$, $\sigma^2 = 4.8$.

For the situation displayed in Figure 8.7, suppose that we wish to find $P(X \leq 10)$. By the exact binomial probabilities found using the *Discrete Distributions* applet, the *Approximations to Distributions* applet, or Table 2 of the Appendix,

$$P(X \leq 10) = 0.245.$$

This value is the sum of the heights of the bars from $x = 0$ up to and including $x = 10$.

Looking at the normal curve in Figure 8.4, we can see that the areas in the bars at $x = 10$ and below are best approximated by the area under the curve to the left of 10.5. The extra 0.5 is added so that the total bar at $x = 10$ is included in the area under consideration. Thus, if W represents a normally distributed random variable with $\mu = 12$ and $\sigma^2 = 4.8$ ($\sigma = 2.2$), then

$$\begin{aligned} P(X \leq 10) &\approx P(W \leq 10.5) \\ &= P\left(\frac{W - \mu}{\sigma} \leq \frac{10.5 - 12}{2.2}\right) \\ &= P(Z \leq -0.68) \\ &= 0.5 - 0.2517 \\ &= 0.2483 \end{aligned}$$

from the normal probability table, Table 4 in the Appendix, or either the *Continuous Distributions* or *Approximations to Distributions* applet. We see that the normal approximation of 0.248 is close to the exact binomial probability of 0.245. The approximation would even be better if n were larger.

The normal approximation to the binomial distribution works well even for moderately large n, as long as p is not close to 0 or to 1. A useful rule of thumb is to make sure that n is large enough to guarantee that $p \pm 2\sqrt{p(1-p)/n}$ will lie within the interval (0, 1), before using the normal approximation. Otherwise, the binomial distribution may be so asymmetric that the symmetric normal distribution cannot provide a good approximation.

EXAMPLE 8.8 Six percent of the apples in a large shipment are damaged. Before accepting each shipment, the quality control manager of a large store randomly selects 100 apples. If four or more are damaged, the shipment is rejected. What is the probability that this shipment is rejected?

Solution The number of damaged apples X in the sample is well modeled by a binomial distribution if the shipment indeed is large. Before using the normal approximation, we should check to confirm that

$$\begin{aligned} p \pm 2\sqrt{\frac{p(1-p)}{n}} &= 0.06 \pm 2\sqrt{\frac{(0.06)(0.94)}{100}} \\ &= 0.06 \pm 0.024 \end{aligned}$$

is entirely within the interval (0, 1), which it is. Thus, the normal approximation should work well.

Therefore, the probability of rejecting the lot is

$$P(X \geq 4) \approx P(W \geq 3.5)$$

where W is a normally distributed random variable with $\mu = np = 6$ and $\sigma = \sqrt{np(1-p)} = 2.4$. It follows that

$$\begin{aligned} P(W \geq 3.5) &= P\left(\frac{W-\mu}{\sigma} \geq \frac{3.5-6}{2.4}\right) \\ &= P(Z \geq -1.05) \\ &= 0.3531 + 0.5 \\ &= 0.8531. \end{aligned}$$

There is a large probability of rejecting a shipment with 6% damaged apples. ▪

EXAMPLE 8.9 Candidate A believes that she can win a city election if she receives at least 55% of the votes from precinct I. Unknown to the candidate, 50% of the registered voters in the precinct favor her. If $n = 100$ voters show up to vote at precinct I, what is the probability that candidate A will receive at least 55% of that precinct's votes.

Solution Let X denote the number of voters in precinct I who vote for candidate A. We must approximate $P(X/n \geq 0.55)$, when p, the probability that a randomly selected voter favors candidate A, is 0.5. If we think of the $n = 100$ voters at precinct I as a random sample from all potential voters in that precinct, then X has a binomial distribution with $p = 0.5$.

Applying Theorem 8.4, we find that

$$\begin{aligned} P(X/n \geq 0.55) &= P\left[Z \geq \frac{0.545 - 0.5}{\sqrt{0.5(0.5)/100}}\right] \\ &= P(Z \geq 0.9) \\ &= 0.5 - 0.3159 = 0.1841 \end{aligned}$$

from the *Continuous Distributions* applet or Table 4 in the Appendix. Candidate A has a 18.4% chance of getting at least 55% of the precinct's votes. ▪

Exercises

8.11 The fracture strength of a certain type of glass has been found to have a standard deviation of approximately 0.4 thousands of pounds per square inch. If the fracture strength of 100 pieces of glass is to be tested, find the approximate probability that the sample mean is within 0.2 thousands of pounds per square inch of the true population mean.

8.12 If the fracture strength measurements of glass have a standard deviation of 0.4 thousands of pounds per square inch, how many glass fracture strength measurements should be taken if the sample mean is to be within 0.2 thousands of pounds per square inch with a probability of approximately 0.95?

8.13 Soil acidity is measured by a quantity called the pH, which may range from 0 to 14 for soils ranging from extreme alkalinity to extreme acidity. Many soils have an average pH in the more-or-less-neutral 5 to 8 range. A scientist wants to estimate the average pH for a large field from n randomly selected core samples, measuring the pH of each sample. From experience he expects the standard deviation of the measurements to be 0.5. If the scientist selects $n = 40$ samples, find the approximate probability that the sample mean of the 40 pH measurements will be within 0.2 of the true average pH for the field.

8.14 Suppose that the scientist in Exercise 8.13 would like the sample mean to be within 0.1 of the true mean with a probability of 0.90. How many core samples should he take?

8.15 Resistors of a certain type have resistances that average 200 ohms with a standard deviation of 10 ohms. Suppose that 25 of these resistors are to be used in a circuit.

a Find the probability that the average resistance of the 25 resistors is between 199 and 202 ohms.

b Find the probability that the *total* resistance of the 25 resistors does not exceed 5100 ohms. [*Hint:* Notice that

$$P\left(\sum_{i=1}^{n} X_i > a\right) = P(n\bar{X} > a) = P(\bar{X} > a/n)$$

for the situation described.]

c What assumptions are necessary for the answers in parts (a) and (b) to be good approximations?

8.16 One-hour carbon monoxide concentrations in air samples from a large city average 12 ppm with a standard deviation of 9 ppm. Find the probability that the average concentration in 100 samples selected randomly will exceed 14 ppm.

8.17 In the manufacture of compact discs (CDs), the block error rate (BLER) is a measure of quality. The BLER is the raw digital error rate before any error correction. According to the industry standard, a CD is allowed a BLER of up to 220 before it is considered a "bad" disc. A CD manufacturer produces CDs with an average of 218 BLERS per disc and a standard deviation of 12. Find the probability that 25 randomly selected CDs from this manufacturer have an average BLER of more than 220.

8.18 The average time that Mark spends commuting to and from work each day is 1.2 hours, and the standard deviation is 0.2 hour.

a Find the probability that the average daily commute time for a period of 36 working days is between 1.15 and 1.25 hours.

b Find the probability that the total commute time for the 36 days is less than 40 hours.

c What assumptions must be true for the answers in parts (a) and (b) to be valid approximations?

8.19 The strength of a thread is a random variable with a mean of 0.5 lb and a standard deviation of 0.2 lb. Assume that the strength of a rope is the sum of the strengths of the threads in the rope.

a Find the probability that a rope consisting of 100 threads will hold 45 lb.

b How many threads are needed for a rope to provide 99% assurance that it will hold 45 lb?

8.20 Many bulk products, such as iron ore, coal, and raw sugar, are sampled for quality by a method that requires many small samples to be taken periodically as the material moves along a conveyor belt. The small samples are then aggregated and mixed to form one composite sample. Let X_i denote the volume of the ith small sample from a particular lot, and suppose that $X_1, X_2, \ldots, X_n$ constitute a random sample, with each X_i having a mean of μ and a variance of σ^2. The average volume μ of the samples can be set by adjusting the size of the sampling device. Suppose that the variance σ^2 of sampling volumes is known to be 4 for a particular situation (measurements are in cubic inches). The total volume of the composite sample is required to exceed 200 cubic inches with a probability of approximately 0.95 when $n = 50$ small samples are selected. Find a setting for μ that will satisfy the sampling requirements.

8.21 The service time for customers coming through a checkout counter in a grocery store are independent random variables with a mean of 2.5 minutes and a variance of 1 minute. Approximate the probability that 100 customers can be serviced in less than 4 hours of total service time.

8.22 Referring to Exercise 8.21, find the number of customers n such that the probability of servicing all n customers in less than 4 hours is approximately 0.1.

8.23 Suppose that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ constitute random samples from populations with mean μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Then the Central Limit Theorem can be extended to show that $\bar{X} - \bar{Y}$ is approximately normally distributed for large n_1 and n_2 with mean $\mu_1 - \mu_2$ and variance $\left(\sigma_1^2/n_1 + \sigma_2^2/n_2\right)$. Water flow through soil depends on, among other things, the porosity (volume proportion due to voids) of the soil. To compare two types of sandy soil, $n_1 = 50$ measurements are to be taken on the porosity of soil A, and $n_2 = 100$ measurements are to be taken on the porosity of soil B. Assume that $\sigma_1^2 = 0.01$ and $\sigma_2^2 = 0.02$. Find the approximate probability that the difference between the sample means will be within 0.05 unit of the true difference between population means $\mu_1 - \mu_2$.

8.24 The length of the same species of fish can differ with location. To compare the lengths of largemouth bass in two rivers in Wisconsin, $n_1 = 30$ fish from the first river were collected at random and measured, and $n_2 = 36$ fish from the second river were collected at random and measured. Assume that $\sigma_1^2 = 2.2$ inches^2 and $\sigma_2^2 = 2.4$ inches^2. Find the approximate probability that the difference between the sample means will be within 0.5 inch of the true difference between population means $\mu_1 - \mu_2$.

8.25 Referring to Exercise 8.24, suppose that samples are to be selected with $n_1 = n_2 = n$. Find the value of n that will allow the difference between the sample means to be within 0.4 inch of $\mu_1 - \mu_2$ with a probability of approximately 0.90.

8.26 An experiment is designed to test whether operator A or operator B gets the job of operating a new machine. Each operator is timed on 50 independent trials involving the performance of a certain task on the machine. If the sample means for the 50 trials differ by more than 1 second, the operator with the smaller mean will get the job. Otherwise, the outcome of the experiment will be considered a tie. If the standard deviations of times for both operators are assumed to be 2 seconds, what is the probability that operator A will get the job on the basis of the experiment, even though both operators have equal ability?

8.27 The median age of residents of the United States is 36.4 years. If a survey of 100 randomly selected United States residents is taken, find the approximate probability that at least 60 of them will be under 31 years of age.

8.28 A lot acceptance sampling plan for large lots calls for sampling 50 items and accepting the lot if the number of nonconformances is no more than 5. Find the approximate probability of acceptance if the true proportion of nonconformances in the lot is as follows.

a 10%
b 20%
c 30%

8.29 Of the customers who enter a store for stereo speakers, only 24% make purchases. If 38 customers enter the showroom tomorrow, find the approximate probability that at least 10 will make purchases.

8.30 The block error rate (BLER), which is the raw digital error rate, is a measure of the quality of compact discs (CDs). According to industry standards, a CD-ROM is considered defective if the BLER exceeds 220. For a certain brand of CDs, 6% are generally found to be defective because the BLER exceeds 220. If 125 CD-ROMs are inspected, find the approximate probability that 4 or fewer are defective as measured by the BLER.

8.31 Sixty-eight percent of the U.S. population prefer toilet paper to be hung so that the paper dangles over the top; 32% prefer the paper to dangle at the back. In a poll of 500 people, the respondents are each asked whether they prefer toilet paper to be hung over the top or to dangle down the back.

a What is the probability that at least 70% respond "over the top"?

b What is the probability that at least half respond "down the back"?

c With a probability of 0.95, how close should the sample proportion of those responding "over the top" be to the population proportion with that response?

8.32 The capacitances of a certain type of capacitor are normally distributed with a mean of $53\mu f$ and a standard deviation of $2\mu f$. If 64 such capacitors are to be used in an electronic system, approximate the probability that at least 12 of them will have capacitances below $50\mu f$.

8.33 The daily water demands for a pumping station exceed 500,000 gallons with a probability of only 0.15. Over a 30-day period, find the approximate probability that the demand for more than 500,000 gallons per day occurs no more than twice.

8.34 Waiting times at a service counter in a pharmacy are exponentially distributed with a mean of 10 minutes. If 100 customers come to the service counter in a day, approximate the probability that at least half of them must wait for more than 10 minutes.

8.35 A large construction firm has won 70% of the jobs for which it has bid. Suppose this firm bids on 25 jobs next month.

a Approximate the probability that it will win at least 20 of these jobs.

b Find the exact binomial probability that it will win at least 20 of these jobs. Compare the result to your answer in part (a).

c What assumptions must be true for your answers in parts (a) and (b) to be valid?

8.36 An auditor samples 100 of a firm's travel vouchers to determine how many of these vouchers are improperly documented. Find the approximate probability that more than 30% of the sampled vouchers will show up as being improperly documented if, in fact, only 20% of the firm's vouchers are improperly documented.

8.5 Combination of Convergence in Probability and Convergence in Distribution

We may often be interested in the limiting behavior of the product or quotient of several functions of a set of random variables. The following theorem, which combines convergence in probability with convergence in distribution, applies to the quotient of two functions, X_n and Y_n.

THEOREM 8.5

Suppose that X_n **converges in distribution** toward a random variable X, and that Y_n converges in probability toward unity. Then X_n/Y_n converges in distribution toward X.

The proof of Theorem 8.5 is beyond the scope of this text, but we can observe its usefulness in the following example. ▪

EXAMPLE **8.10** Suppose that $X_1, X_2, \ldots, X_n$ are independent and identically distributed random variables with $E(X_i) = \mu$ and $V(X_i) = \sigma^2$. Define S'^2 as

$$S'^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Show that

$$\sqrt{n}\frac{\bar{X} - \mu}{S'}$$

converges in distribution toward a standard normal random variable.

Solution In Example 8.2, we showed that S'^2 converges in probability toward σ^2. Hence, it follows from Theorem 8.2, parts (c) and (d), that S'^2/σ^2 (and hence S'/σ) converges in probability toward 1. We also know from Theorem 8.4 that

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right)$$

converges in distribution toward a standard normal random variable. Therefore,

$$\sqrt{n}\left(\frac{\bar{X} - \mu}{S'}\right) = \sqrt{n}\left(\frac{\bar{X} - \mu}{\sigma}\right)\Big/\left(\frac{S'}{\sigma}\right)$$

converges in distribution toward a standard normal random variable by Theorem 8.5. ▪

8.6 Summary

Distributions of functions of random variables are important in both theoretical and practical work. Often, however, the distribution of a certain function cannot be found—at least not without a great deal of effort. Thus, approximations to the distributions of functions of random variables play a key role in probability theory. To begin with, probability itself can be thought of as a ratio of the number of "successes" (a random variable) to the total number of trials in a random experiment. That this ratio converges, under certain conditions, toward a constant is a fundamental result in probability on which much theory and many applications are built. This *law of large numbers* is the reason, for example, that opinion polls work, if the sample underlying the poll is truly random.

The average, or mean, of random variables is one of the most commonly used functions. The *Central Limit Theorem* provides a very useful normal approximation to the distribution of such averages. This approximation works well under very general conditions and is one of the most widely used results in all of probability theory.

Supplementary Exercises

8.37 An anthropologist wishes to estimate the average height of men for a certain race of people. If the population standard deviation is assumed to be 2.5 inches and if she randomly samples 100 men, find the probability that the difference between the sample mean and the true population mean will not exceed 0.5 inch.

8.38 Suppose that the anthropologist of Exercise 8.37 wants the difference between the sample mean and the population mean to be less than 0.4 inch with a probability of 0.95. How many men should she sample to achieve this objective?

8.39 A large industry has an average hourly wage of \$9 per hour with a standard deviation of \$0.75. A certain ethnic group consisting of 81 workers has an average wage of \$8.90. Is it reasonable to assume that the ethnic group is a random sample of workers from the industry? (Calculate the probability of randomly obtaining a sample mean that is less than or equal to \$8.90 per hour.)

8.40 A machine is shut down for repairs if a random sample of 100 items selected from the daily output of the machine reveals at least 10% defectives. (Assume that the daily output is a large number of items.) If the machine, in fact, is producing only 8% defective items, find the probability that it will shut down on a given day.

8.41 A pollster believes that 60% of the voters in a certain area favor a particular candidate. If 64 others are randomly sampled from the large number of voters in this area, approximate the probability that the sampled fraction of voters favoring the bond issue will not differ from the true fraction by more than 0.06.

8.42 Twenty-five heat lamps are connected in a greenhouse so that when one lamp fails another takes over immediately. (Only one lamp is turned on at any time.) The lamps operate independently, each with a mean life of 50 hours and a standard deviation of 4 hours. If the greenhouse is not checked for 1300 hours after the lamp system is turned on, what is the probability that a lamp will be burning at the end of the 1300-hour period?

8.43 Suppose that $X_1, X_2, \dots, X_n$ are independent random variables, each with a mean of μ_1 and a variance of σ_1^2. Suppose, too, that $Y_1, Y_2, \dots, Y_n$ are independent random variables, each with a mean of μ_2 and a variance of σ_2^2. Show that, as $n \to \infty$, the random variable

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{(\sigma_1^2 + \sigma_2^2)/n}}$$

converges in distribution toward a standard normal random variable.

8.44 Let X have a chi-squared distribution with n degrees of freedom; that is, Y has the density function

$$f(x) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}, & x \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that the random variable

$$\frac{X - n}{\sqrt{2n}}$$

is asymptotically standard normal in distribution, as $n \to \infty$.

8.45 A machine in a heavy-equipment factory produces steel rods of length X, where X is a normal random variable with a mean μ of 6 inches and a variance σ^2 of 0.2. The cost C of repairing a rod that is not exactly 6 inches in length is proportional to the square of the error and is given (in dollars) by

$$C = 4(X - \mu)^2.$$

If 50 rods with independent lengths are produced in a given day, approximate the probability that the total cost for repairs for that day will exceed \$48.

9

Extensions of Probability Theory

9.1 The Poisson Process

In this section, we apply probability theory to modeling random phenomena that change over time—the so-called *stochastic processes*. Thus, a time parameter is introduced, and the random variable $Y(t)$ is regarded as being a function of time. (Later we shall see that the concept of time can be generalized to include space.) Examples are endless: $Y(t)$ could represent the size of a biological population at time t; the cost of operating a complex industrial system for t time units; the distance displaced by a particle in time t; or the number of customers waiting to be served at a checkout counter t time units after its opening.

Although we shall not discuss stochastic processes exhaustively, a few elementary concepts are in order. A stochastic process $Y(t)$ is said to have *independent* increments if, for any set of time points $t_0 < t_1 < \cdots < t_n$, the random variables $[Y(t_i) - Y(t_{i-1})]$ and $[Y(t_j) - Y(t_{j-1})]$ are independent for $i \neq j$. (See Section 6.3 for a discussion of independent random variables.) The process is said to have *stationary* independent increments if, in addition, the random variable $[Y(t_2 + h) - Y(t_1 + h)]$ and $[Y(t_2) - Y(t_1)]$ has identical distributions for any $h > 0$. We begin our discussion of specific processes by considering an elementary counting process with wide applicability: the Poisson process.

Let $Y_{(t)}$ denote the number of occurrences of some event in the time interval $(0, t)$, where $t > 0$. This could be the number of accidents at a particular intersection, the number of times a computer breaks down, or any similar count. We derived the Poisson distribution in Section 4.7 by looking at a limiting form of the binomial distribution, but now we derive the Poisson process by working directly from a set of axioms.

The Poisson process is defined as satisfying the following four axioms:

Axiom 1: $P[Y(0) = 0] = 1$.

Axiom 2: $Y(t)$ has stationary independent increments. (The numbers of occurrences in two nonoverlapping time intervals of the same length are independent and have the same probability distribution.)

Axiom 3: $P\{[Y(t+h) - Y(t)] = 1\} = \lambda h + o(h)$, where $h > 0$ and $o(h)$ is a generic notation for a term such that $o(h)/h \to 0$ as $h \to 0$, and where λ is a constant.

Axiom 4: $P\{[Y(t+h) - Y(t)] > 1\} = o(h)$

Axiom 3 says that the chance of one occurrence in a small interval of time h is roughly proportional to h, and Axiom 4 states that the chance of more than one such occurrence tends to reach zero faster than h itself does. These two axioms imply that

$$P\{[Y(t+h) - Y(t)] = 0\} = 1 - \lambda h - o(h).$$

The probability distribution can now be derived from the axioms. Consider the interval $(0, t+h)$ partitioned into the two disjoint pieces $(0, t)$ and $(t, t+h)$. Then we can write

$$P[Y(t+h) = k] = \sum_{j=0}^{k} P\{[Y(t) = j], [Y(t+h) - Y(t)] = k - j\}$$

because the events inside the right-hand probability are mutually exclusive. Because independent increments are assumed, we have

$$P[Y(t+h) = k] = \sum_{j=0}^{k} P[Y(t) = j] P\{[Y(t+h) - Y(t)] = k - j\}.$$

By stationarity and Axiom 1, this becomes

$$P[Y(t+h) = k] = \sum_{j=0}^{k} P[Y(t) = j] P[Y(h) = k - j].$$

To simplify the notation, let $P[Y(t) = j] = P_j(t)$. Then

$$P_k(t+h) = \sum_{j=0}^{k} P_j(t) P_{k-j}(h)$$

and Axiom 4 reduces this to

$$P_k(t+h) = P_{k-1}(t) P_1(h) + P_k(t) P_0(h) + o(h).$$

Axiom 3 further reduces the equation to

$$\begin{aligned}P_k(t+h) &= P_{k-1}(t)[\lambda h + o(h)] + P_k(t)[1 - \lambda h - o(h)] \\ &= P_{k-1}(t)(\lambda h) + P_k(t)(1 - \lambda h) + o(h).\end{aligned}$$

We can write this equation as

$$\frac{1}{h}[P_k(t+h) - P_k(t)] = \lambda P_{k-1}(t) - \lambda P_k(t) + \frac{o(h)}{h}$$

and on taking limits as $h \to 0$, we have

$$\frac{\mathrm{d}P_k(t)}{\mathrm{d}t} = \lim_{h\to 0}\frac{1}{h}[P_k(t+h) - P_k(t)] = \lambda P_{k-1}(t) - \lambda P_k(t).$$

To solve this differential equation, we take $k = 0$ to obtain

$$\frac{\mathrm{d}P_k(t)}{\mathrm{d}t} = P_0'(t) = -\lambda P_0(t)$$

because $P_{k-1}(t)$ will be replaced by zero for $k = 0$. Thus, we have

$$\frac{P_0'(t)}{P_0(t)} = -\lambda$$

or (on integrating both sides)

$$\ln P_0(t) = -\lambda t + c$$

The constant is evaluated by the boundary condition $P_0(0) = P[Y(0) = 0] = 1$, by Axiom 1, which gives $c = 0$. Hence,

$$\ln P_0(t) = -\lambda t$$

or

$$P_0(t) = e^{-\lambda t}.$$

By letting $k = 1$ in the original differential equation, we can obtain

$$P_1(t) = \lambda t e^{-\lambda t}$$

and recursively,

$$P_k(t) = \frac{1}{k!}(\lambda t)^k e^{-\lambda t}, \quad k = 0, 1, 2, \ldots$$

Thus, $Y(t)$ has a Poisson distribution with mean λt.

The notion of a Poisson process is easily extended to points randomly dispersed in a plane, where the interval (0, t) is replaced by a planar region (quadrat) of area A. Then, under analogous axioms, $Y(A)$—the number of points in a quadrat of area A—has a Poisson distribution with mean λA. Similarly, it can be applied to points randomly dispersed in space when the interval (0, t) is replaced by a cube, a sphere, or some other three-dimensional figure in volume V.

EXAMPLE 9.1 Suppose that in a certain area, plants are randomly dispersed with a mean density of 20 per square yard. If a biologist randomly locates 100 sampling quadrats of 2 square yards each in the area, how many of them can be expected to contain no plants?

Solution Assuming that the plant counts per unit area have a Poisson distribution with a mean of 20 per square yard, the probability of no plants in a 2-square-yard area is

$$\begin{aligned} P[Y(2) = 0] &= e^{-2\lambda} \\ &= e^{-2(20)} \\ &= e^{-40}. \end{aligned}$$

If the 100 quadrats are randomly located, the expected number of quadrats containing zero plants is

$$100P[Y(2) = 0] = 100e^{-40}. \quad ■$$

9.2 Birth and Death Processes: Biological Applications

In Section 9.1, we were interested in counting the number of occurrences of a single type of event, such as accidents, defects, or plants. Let us extend this idea to counts of two types of events, which we call *births* and *deaths*. For illustrative purposes, a birth may denote a literal birth in a population of organisms, and similarly for a death. More specifically, we shall think of a birth as a cell division that produces two new cells, and we shall think of a death as the removal of a cell from the system.

Birth and death processes are of use in modeling the dynamics of population growth. In human populations, such models may be used to influence decisions on housing programs, food management, natural resource management, and a host of related economic matters. In animal populations, birth and death models are used to predict the size of insect populations and to measure the effectiveness of nutrition or eradication programs.

Let λ denote the birth rate, and let θ denote the death rate of each cell in the population; assume that the probability of birth or death for an individual cell is independent of the size and age of the population. If $Y(t)$ denotes the size of the population at time t, we write

$$P[Y(t) = n] = P_n(t).$$

More specifically, assume that the probability of a birth in a small interval of time h, given that the population size at the start of the interval is n, has the form $n\lambda h + o(h)$. Similarly, the probability of a death is given by $n\theta h + o(h)$. The probability of more than one birth or death in h is an order $o(h)$. For an individual cell (the case $n = 1$), this says that the probability of division in a small interval of time is $\lambda h + o(h)$. A differential equation for $P_n(t)$ is developed as follows. If a population is of size n at time $(t + h)$, it must have been of size $(n - 1)$, n, or $(n + 1)$ at time t. Thus

$$P_n(t+h) = \lambda(n-1)hP_{n-1}(t) + [1 - n\lambda h - n\theta h]P_n(t) + \theta(n+1)hP_{n+1}(t) + o(h)$$

or

$$\frac{1}{h}[P_n(t+h) - P_n(t)] = \lambda(n-1)P_{n-1}(t) - n(\lambda+\theta)P_n(t) + \theta(n+1)P_{n+1}(t) + \frac{o(h)}{h}.$$

On taking limits as $h \to 0$, we have

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = \lambda(n-1)P_{n-1}(t) - n(\lambda+\theta)P_n(t) + \theta(n+1)P_{n+1}(t).$$

A general solution to this equation can be obtained with some difficulty, but let us instead look at some special cases. If $\theta = 0$, which implies that no deaths are taking place in the time of interest, we have a pure birth process exemplified by the differential equation

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = \lambda(n-1)P_{n-1}(t) - n\lambda P_n(t).$$

A solution to this equation is

$$P_n(t) = \binom{n-1}{i-1} e^{-\lambda it}(1 - e^{-\lambda t})^{n-i}$$

where i is the size of the population at time $t = 0$; that is, $P_i(0) = 1$. Notice that this solution is a negative binomial probability distribution (alternative parameterization). The pure birth process may be a reasonable model of a real population if the time interval is short or if deaths are neglected.

EXAMPLE 9.2 Find the expected size of a population, governed by a pure birth process, at time t if the size was i at time 0.

Solution The mean of a negative binomial distribution written as

$$\binom{n-1}{r-1} p^r(1-p)^{n-r}$$

is given by r/p. Thus, the mean of the pure birth process is $i/e^{-\lambda t}$ or $ie^{\lambda t}$. If the birth rate is known (at least approximately), this provides a formula for estimating what the expected population size will be t time units in the future. ▪

Returning to the case where $\theta > 0$, we can show that

$$P_0(t) = \left[\frac{\theta e^{(\lambda-\theta)t} - \theta}{\lambda e^{(\lambda-\theta)t} - \theta}\right]^t$$

(see Bailey 1964).

On taking the limit of $P_0(t)$ as $t \to \infty$, we get the probability of ultimate extinction of the population, which turns out to be

$$\lim_{t\to\infty} P_0(t) = \begin{cases} (\theta/\lambda)^t, & \lambda > \theta \\ 1, & \lambda < \theta \\ 1, & \lambda = \theta. \end{cases}$$

Thus, the population has a chance of persisting indefinitely only if the birth rate λ is larger than the death rate θ; otherwise, it is certain to become extinct. We can find $E[Y(t)]$ for the birth and death processes without first finding the probability distribution $P_n(t)$. The method is outlined in Exercise 9.2, and the solution is

$$E[Y(t)] = ie^{(\lambda-\theta)t}.$$

9.3 Queues: Engineering Applications

Queuing theory is concerned with probabilistic models that predict the behavior of customers arriving at a certain station and demanding some kind of service. The word "customer" may refer to actual customers at a service counter, or more generally, it may refer to automobiles entering a highway or a service station, telephone calls coming into a switchboard, breakdowns of machines in a factory, or comparable phenomena. Queues are classified according to an input distribution (the distribution of customer arrivals); a service distribution (the distribution of service time per customer); and a queue discipline, which specifies the number of servers and the manner of dispensing service (such as "first come, first served").

Queuing theory forms useful probabilistic models for a wide range of practical problems, including the design of stores and public buildings, the optimal arrangement of workers in a factory, and the planning of highway systems. Consider a system that involves one station dispensing service to customers on a first-come, first-served basis. This could be a store with a single checkout counter or a service station with a single gasoline pump. Suppose that customer arrivals constitute a Poisson process with intensity λ per hour, and that departures of customers from the station constitute an independent Poisson process with intensity θ per hour. (This implies that the service time is an exponential random variable with mean $1/\theta$.) Hence the probability of a customer arrival in a small interval of time h is $\lambda h + o(h)$, and the probability of a departure is $\theta h + o(h)$, with a probability of more than one arrival for departure of $o(h)$. Notice that these probabilities are identical to the ones for births and deaths (discussed in the preceding section), but with the dependence on n suppressed. Thus, if $Y(t)$ denotes the number of customers in the system (being served and waiting to be served) at time t, and if $P_n(t) = P[Y(t) = n]$, then, as in Section 9.2,

$$\frac{\mathrm{d}P_n(t)}{\mathrm{d}t} = \lambda P_{n-1}(t) - (\lambda + \theta)P_n(t) + \theta P_{n+1}(t).$$

It can be shown that, for a large value of t, this equation has a solution P_n that does not depend on t. Such a solution is called an *equilibrium distribution*. If the solution is free of t, it must satisfy

$$0 = \lambda P_{n-1} - (\lambda + \theta)P_n + \theta P_{n+1}.$$

A solution to this equation is given by

$$P_n = \left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^n, \quad n = 0, 1, 2, \ldots$$

provided that $\lambda < \theta$, P_n represents the probability of these being n customers in the system at any time t that is far removed from the start of the system.

EXAMPLE 9.3 In a system operating as has just been indicated, find the expected number of customers (including the one being served) at the station at some time far removed from the start.

Solution The equilibrium distribution is the geometric distribution given in Chapter 4. If X has the distribution given by P_n, then from Chapter 4 we know that

$$E(X) = \frac{(\lambda/\theta)}{\left(1 - \frac{\lambda}{\theta}\right)} = \frac{\lambda}{\theta - \lambda}$$

again assuming that $\lambda < \theta$. ■

9.4 Arrival Times for the Poisson Process

Let $Y(t)$ denote a Poisson process (see Section 9.1), with mean λ per unit time. It is sometimes of interest to study properties of the actual times at which events have occurred, given that $Y(t)$ is fixed. Suppose the $Y(t) = n$, and let $0 < U_{(1)} < U_{(2)} < \cdots < U_{(n)} < t$ be the actual times that events occur in the interval from 0 to t. Let the joint density function of $U_{(1)}, \ldots, U_{(n)}$ be denoted by $g(u_1, \ldots, u_n)$. If we think of $\mathrm{d}u_i$ as being a very small interval, then $g(u_1, \ldots, u_n)\mathrm{d}u_1 \ldots \mathrm{d}u_n$ is equal to the probability that one event will occur in each of the intervals $(u_i,\ u_i + \mathrm{d}u_i)$ and none will occur elsewhere, given that n events occur in $(0, t)$. Thus,

$$g(u_1, \ldots, u_n)\,\mathrm{d}u_1 \cdots \mathrm{d}u_n = \frac{1}{\left(\frac{(\lambda t)^n e^{-\lambda t}}{n!}\right)}[\lambda \mathrm{d}u_1 e^{-\lambda \mathrm{d}u_1} \cdots \lambda \mathrm{d}u_n e^{-\lambda \mathrm{d}u_n} e^{-\lambda(t - \mathrm{d}u_1 - \cdots - \mathrm{d}u_n)}]$$

$$= \frac{n!}{t^n}\mathrm{d}u_1 \cdots \mathrm{d}u_n.$$

It follows that

$$g(u_1, \ldots, u_n) = \frac{n!}{t^n}, \quad u_1 < u_2 < \cdots < u_n$$

or in other words, $U_{(1)}, \ldots, U_{(n)}$ behave as an ordered set of n independent observations from the uniform distribution on $(0,\ t)$. This implies that the unordered occurrence times $U_1, \ldots, U_n$ are independent uniform random variables.

EXAMPLE 9.4 Suppose that the number of telephone calls coming into a switchboard follows a Poisson process with a mean of 10 calls per minute. A particularly slow period of 2 minutes' duration yielded only four calls. Find the probability that all four calls came in the first minute.

Solution Here we are given that $Y(2) = 4$. We can then evaluate the density function of $U_{(4)}$ to be

$$g_4(u) = \left(\frac{4}{2^4}\right)u^3, \quad 0 \le u \le 2.$$

This comes from the fact that the $U_{(4)}$ is the largest order statistic in a sample of $n = 4$ observations from a uniform distribution. Thus, the probability in question is

$$P([U_{(4)} \le 1] = \int_0^1 g_4(u)\,\mathrm{d}u$$

$$= \frac{4}{2^4}\int_0^1 u^3 \mathrm{d}u$$

$$= \frac{1}{2^4}$$

$$= \frac{1}{16}. \quad \blacksquare$$

9.5 Infinite Server Queue

We now consider a queue with an infinite number of servers. This queue could model a telephone system with a very large number of channels or the claims department of an insurance company in which all claims are processed immediately. Customers arrive at random times and keep a server busy for a random length of time. The number of servers is large enough, however, that a customer never has to wait for service to begin. A random variable of much interest in this system is $X(t)$, the number of servers busy at time t.

Let the customer arrivals be a Poisson process with a mean of λ per unit time. The customer arriving at time U_n, measured from the start of the system at $t = 0$, keeps a server busy for a random service time Y_n. The service times $Y_1, Y_2, \ldots$, are assumed to be independent and identically distributed with distribution function $F(y)$. $X(t)$, the number of servers busy at time t, can then be written as

$$X(t) = \sum_{n=1}^{N(t)} w(t, U_n, Y_n)$$

where $N(t)$, the number of arrivals in $(0, t)$, has a Poisson distribution and

$$w(t, U_n, Y_n) = \begin{cases} 1, & 0 < U_n \leq t \leq U_n + Y_n \\ 0, & \text{otherwise} \end{cases}$$

The condition $U_n \leq t \leq U_n + Y_n$ is precisely the condition that a customer arriving at time U_n is still being served in time t. The function $w(t, U_n, Y_n)$ simply keeps track of those customers still receiving services at t. Operations similar to those used in Section 7.8 for the compound Poisson distribution will reveal that $X(t)$ has a Poisson distribution with a mean given by

$$E[X(t)] = E[N(t)]E[w(t, U_n, Y_n)].$$

Because $E[N(t)] = \lambda t$, it remains for us to evaluate

$$E[w(t, U_n, Y_n)] = P(0 \leq U_n \leq t \leq U_n + Y_n) = p(Y_n \geq t - U_n).$$

Recall that U_n is uniformly distributed on $(0, t)$; and thus,

$$\begin{aligned} P(Y_n \geq t - U_n) &= \int_0^1 P(Y_n \geq t - u | U_n = u) \frac{\mathrm{d}u}{t} \\ &= \frac{1}{t} \int_0^t P(Y_n \geq t - u)\, \mathrm{d}u \\ &= \frac{1}{t} \int_0^1 [1 - F(t - u)]\, \mathrm{d}u. \end{aligned}$$

Making the change of variable $t - u = s$, we have

$$P(Y_n \geq t - U_n) = \frac{1}{t} \int_0^t [1 - F(s)]\, \mathrm{d}s$$

or

$$E[X(t)] = \lambda \int_0^1 [1 - F(s)]\, \mathrm{d}s.$$

EXAMPLE 9.5 Suppose that the telephone switchboard of Example 9.4 (which averaged $\lambda = 10$ incoming call per minute) has a large number of channels. The calls are of random and independent length, but they average 4 minutes each. Approximate the probability that after the switchboard has been in service for a long time, there will be no busy channels at some specified time t_0.

Solution Notice that $\int_0^\infty [1 - F(s)]\, \mathrm{d}s = E(Y)$. Thus, for a large value of t_0,

$$\begin{aligned} E[X(t_0)] &= \lambda \int_0^{t_0} [1 - F(s)]\, \mathrm{d}s \\ &= \lambda E(Y) \\ &= 10(4) \\ &= 40. \end{aligned}$$

Because $X(t)$ has a Poisson distribution

$$\begin{aligned} P[X(t_0) = 0] &= \exp\left\{ -\lambda \int_0^{t_0} [1 - F(s)]\, \mathrm{d}s \right\} \\ &= \exp(-40). \quad \blacksquare \end{aligned}$$

9.6 Renewal Theory: Reliability Applications

Reliability theory deals with probabilistic models that predict the behavior of systems of components, enabling one to formulate optimum maintenance policies, including replacement and repair of components, and to measure operating characteristics of the system, such as the proportion of downtime, the expected life, and the probability of survival beyond a specified age. The "system" could be a complex electronic system (such as a computer), an automobile, a factory, or even a human body. In the remainder of this section, this system will be simplified to a single component that has a random life length but may be replaced before it fails—such as a light bulb in a socket.

For a Poisson process, as described in Section 9.1, the time between any two successive occurrences of the event that is being observed has an exponential distribution, with mean $1/\lambda$, where λ is the expected number of occurrences per unit time (see Section 5.4). A more general stochastic process can be obtained by treating interarrival times as nonnegative random variables that are identically distributed but not necessarily exponential.

Suppose that $X_1, X_2, \ldots, X_n$ represent a sequence of independent, identically distributed random variables, each with distribution function $F(x)$ and probability density function $f(x)$. Assume, too, that $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$. The values of X_i could represent lifetimes of identically constructed electronic components, for example. If a system operates by inserting a new component, letting it burn continuously until it fails, and then repeating the process with an identical new component that is assumed to operate independently, then the occurrences of interest are in-service failures of components; and X_i is the interarrival time between failure number $(i - 1)$ and failure number i. A random variable of interest in such problems is N_i, the number of in-service failures in the time interval $(0, t)$. Notice that

$$N_t = \text{Maximum integer } k, \text{ such that } \sum_{i=1}^{k} X_i \le t$$

with $N_t = 0$ if $X_i > t$. Let us now consider the probability distribution of N_t. If X_i is an exponential random variable, then N_t has a Poisson distribution. If X_i is not exponential, however, the distribution of N_t may be difficult or impossible to obtain. Thus we shall investigate the asymptotic behavior of N_t as t tends toward infinity.

Notice that with $S_r = \sum_{i=1}^{r} X_i$,

$$P(N_t \ge r) = P(S_r \le t).$$

The limiting probabilities will be unity unless $r \to \infty$, and some restrictions must be placed on the relationship between r and t. Thus, suppose that $r \to \infty$ and $t \to \infty$ in such a way that

$$\frac{t - r\mu}{\sqrt{r}\sigma} \to c$$

for some constant c. This implies that $(\mu r/t) \to 1$ as $r \to \infty$ and $t \to \infty$. Now the preceding probability equality can be written as

$$P\left[\frac{N_t - (t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} \ge \frac{r - (t/\mu)}{t^{1/2}\mu^{-3/2}\sigma}\right] = P\left(\frac{S_r - r\mu}{r^{1/2}\sigma} \le \frac{t - r\mu}{r^{1/2}\sigma}\right).$$

We have

$$\frac{r-(t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} = \frac{\mu r - t}{r^{1/2}\sigma}\left(\frac{\mu r}{t}\right)^{1/2}$$

and this quantity tends toward $-c$ as $t \to \infty$ and $r \to \infty$. It follows that

$$P\left[\frac{N_t-(t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} \geq -c\right] = P\left(\frac{S_r - r\mu}{r^{1/2}\sigma} \leq c\right).$$

However,

$$\frac{S_r - r\mu}{r^{1/2}\sigma} = \sqrt{r}\left(\frac{\bar{X}-\mu}{\sigma}\right)$$

which has approximately a standard normal distribution for large values of r. Letting $\Phi(x)$ denote the standard normal distribution function, we can conclude that

$$P\left[\frac{N_t-(t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} \geq -c\right] \to \Phi(c).$$

as $t \to \infty$ and $r \to \infty$, or

$$P\left[\frac{N_t-(t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} \leq -c\right] = [1-\Phi(c)] = \Phi(-c).$$

Hence, for large t, the values of N_t can be regarded as approximately normally distributed with a mean of t/μ and a variance of $t\sigma^2\mu^3$.

EXAMPLE 9.6 A fuse in an electronic system is replaced with an identical fuse upon failure. It is known that these fuses have a mean life of 10 hours with a variance of 2.5 hours. Beginning with a new fuse, the system is to operate for 400 hours.

1 If the cost of a fuse is \$5, find the expected amount to be paid for fuses over the 400-hour period.

2 If 42 replacement fuses are in stock, find the probability that they will all be used in the 400-hour period.

Solution **1** No distribution of lifetimes for the fuses is given, so asymptotic results must be employed. (Notice that $t = 400$ is large in comparison to $\mu = 10$.) Letting N_t denote the number of replacements, we have

$$E(N_t) = \frac{1}{\mu} = \frac{400}{10} = 40.$$

We would expect to use 40 replacements, plus the one initially placed into the system. Thus, the expected cost of fuses is

$$(5)(41) = \$205.$$

2 N_t is approximately normal, with mean $t/\mu = 40$ and variance

$$\frac{t\sigma^2}{\mu^3} = \frac{400(2.5)}{1000} = 1.$$

Thus,

$$\begin{aligned} P(N_t \geq 42) &= P\left(\frac{N_t - (t/\mu)}{t^{1/2}\mu^{-3/2}\sigma} \geq \frac{42-4}{1}\right) \\ &= P(Z \geq 2) \\ &= 0.0228 \end{aligned}$$

where Z denotes a standard normal random variable. ▪

To minimize the number of in-service failures, components often are replaced at age T or at failure, whichever comes first. Here, age refers to length of time in service and T is a constant. Under this age replacement scheme, the interarrival times are given by

$$Y_t = \min(X_i, T).$$

Consider an age replacement policy in which c_1 represents the cost of replacing a component that has failed in service, and c_2 represents the cost of replacing a component (which has not failed) at age T. (Usually, $c_1 > c_2$.) To find the total replacement costs up to time t, we must count the total number of replacements N_t and the number of in-service failures. Let

$$W_i = \begin{cases} 1, & X_i < T \\ 0, & X \geq T. \end{cases}$$

Then $\sum_{i=1}^{N_t} W_i$ denotes the total number of in-service failures. Thus, the total replacement cost up to time t is

$$C_t = c_1 \sum_{i=1}^{N_t} W_i + c_2 \left(N_t - \sum_{i=1}^{N_t} W_i\right).$$

The expected replacement cost per unit time is frequently of interest, and this can be approximated after observing that, under the conditions stated in this section,

$$E\left(\sum_{i=1}^{N_t} W_i\right) = E\left(N_t\right) E\left(W_i\right).$$

Thus,

$$E(C_t) = c_1 E(N_t)E(W_i) + c_2 E(N_t)[1 - E(W_i)]$$

which, for large t, becomes

$$\frac{t}{E(Y_i)}\left[c_1 P(X_i < T) + c_2 P(X_i \geq T)\right].$$

It follows that

$$\frac{1}{t}E(C_t) = \frac{1}{E(Y_i)}\left\{c_1 F(T) + c_2[1 - F(T)]\right\}$$

where $F(x)$ is still the distribution function of X_i.

An optimum replacement policy can be found by choosing T so as to minimize the expected cost per unit time. This concept is continued in Exercise 9.11.

9.7 Summary

The study of the behavior of random events that occur over time is called *stochastic processes*. A few examples of both the theory and the application of stochastic processes are presented here in order to round out the study of important areas of probability.

Perhaps the most common model for random events occurring over time is the Poisson process—an extension of the Poisson distribution introduced in Chapter 4. The Poisson process is generally considered the standard model for events that occur in a purely random fashion across time.

If two types of items—say, births and deaths—are being counted across time, a reasonable probabilistic model is a little more difficult to develop. Many applications of models for birth and death processes have been developed in the biological sciences. One interesting result is that the pure birth process leads to the negative binomial model for population size.

The number of people in a waiting line, or queue, and the way that number changes with time is another interesting stochastic process. It can be shown that, under equilibrium conditions, the number of people in a queue follows a geometric distribution. In an infinite-server queue, there is no waiting time for customers and the problem of interest switches to the number of servers who are busy at any given moment.

Other details related to stochastic processes include the actual arrival times of the events under study and their interarrival times. For the Poisson process, the arrival times within a fixed time interval are shown to be related to the order statistics from a uniform distribution. Renewal theory provides a useful normal approximation to the distribution of the number of events that occur during a fixed interval of time.

Exercises

9.1 During the workday, telephone calls come into an office at the rate of one call every 3 minutes.

a Find the probability that no more than one call will come into this office during the next 5 minutes. (Are the assumptions for the Poisson process reasonable here?)

b An observer in the office hears two calls come in during the first 2 minutes of a 5-minute visit. Find the probability that no more calls will arrive during the visit.

9.2 For the birth and death processes of Section 9.2, $E[Y(t)] = ie^{(\lambda-\theta)t}$, where i is the population size at $t = 0$. Show this by observing that

$$m(t) = E[Y(t)] = \sum_{n=1}^{\infty} nP_n(t)$$

and

$$m'(t) = \frac{dm(t)}{dt} = \sum_{n=0}^{\infty} nP'_n(t).$$

Now use the expression for $P'_n(t)$ given in Section 9.2 and evaluate the sum to obtain a differential equation relating $m'(t)$ to $m(t)$. The solution follows.

9.3 Referring to Section 9.3, show that the geometric equilibrium distribution P_n is a solution to the differential equation defining the equilibrium state.

9.4 In a single-server queue, as defined in Section 9.3, let W denote the total waiting time, including her own service time, of a customer entering the queue a long time after the start of operations. (Assume that $\lambda < \theta$.) Show that W has an exponential distribution by writing

$$\begin{aligned} P(W \le w) &= \sum_{n=0}^{\infty} P[W \le w | n \text{ customers in queue}]P_n \\ &= \sum_{n=0}^{\infty} P[W \le w | n \text{ customers in queue}] \left(\frac{\lambda}{\theta}\right)^n \left(1 - \frac{\lambda}{\theta}\right) \end{aligned}$$

and observing that the conditional distribution of W for fixed n is a gamma distribution with $\alpha = n+1$ and $\beta = 1/\theta$. (The interchange of Σ and ∞ is permitted here.)

9.5 If $Y(t)$ and $X(t)$ are independent Poisson processes, show that $Y(t) + X(t)$ is also a Poisson process.

9.6 Claims arrive at an insurance company according to a Poisson process with a mean of 20 per day. The claims are placed in the processing system immediately on arrival, but the service time per claim is an exponential random variable with a mean of 10 days. On a certain day, the company was completely caught up on its claims service (no claims were left to be processed). Find the probability that 5 days later the company was engaged in servicing at least two claims.

9.7 Assume that telephone calls arrive at a switchboard according to a Poisson process with an average rate of λ per minute. One time period t minutes long is known to have included exactly two calls. Find the probability that the length of time between these calls exceeds d minutes for some constant $d \le t$.

9.8 Let $Y(A)$ denote a Poisson process in a plane, as given in Section 9.1, and let R denote the distance from a random point to the nearest point of realization of the process.

a Show that $U = R^2$ is an exponential random variable with mean $(1/\lambda\pi)$.

b Suppose that n independent values of R can be obtained by sampling n points. Let W denote the sum of the squares of these values. Show that W has a gamma distribution.

c Show that $(n-1)/\pi W$ is an unbiased estimator of λ; that is, $E[(n-1)/\pi W] = \lambda$.

9.9 Let V denote the volume of a three-dimensional figure, such as a sphere. The axioms of the Poisson process as applied to three-dimensional point processes—such as stars in the heavens or bacteria in water—yield the fact that $Y(V)$, the number of points in a figure of volume V, has a Poisson distribution with mean λV. Moreover, $Y(V_1)$ and $Y(V_2)$ are independent if V_1 and V_2 are not overlapping volumes.

a If a point is chosen at random in three-dimensional space, show that the distance R to the nearest point in the realization of a Poisson process has the density function

$$f(r) = \begin{cases} 4\lambda\pi r^2 e^{-(4/3)\lambda\pi r^3}, & r \geq 0 \\ 0, & \text{elsewhere.} \end{cases}$$

b If R is as in part (a), show that $U = R^3$ has an exponential distribution.

9.10 Let Y denote a random variable with distribution function $F(y)$, such that $F(0) = 0$. If $X = min(Y, T)$ for a constant T, show that

$$E(X) = \int_0^T [1 - F(x)]\,\mathrm{d}x.$$

9.11 Referring to Section 9.6, the optimum age replacement interval T_0 is defined as the one that minimizes the expected cost per unit time, as given in Section 9.6.

a Find T_0, assuming that the components have exponential life lengths with mean θ.

b Show that a finite T_0 always exists if $r(t)$ is strictly increasing toward infinity. (Notice that this will be the case for Weibull life lengths if $m > 1$.)

REFERENCE

Bailey, N. T. J. 1964. *The Elements of Stochastic Processes with Applications to the Natural Sciences*. New York: John Wiley & Sons.

Appendix Tables

TABLE 1 Random numbers

Row	Column 1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	10480	15011	01536	02011	81647	91646	69179	14194	62590	36207	20969	99570	91291	90700
2	22368	46573	25595	85393	30995	89198	27982	53402	93965	34095	52666	19174	39615	99505
3	24130	48360	22527	97265	76393	64809	15179	24830	49340	32081	30680	19655	63348	58629
4	42167	93093	06243	61680	07856	16376	39440	53537	71341	57004	00849	74917	97758	16379
5	37570	39975	81837	16656	06121	91782	60468	81305	49684	60672	14110	06927	01263	54613
6	77921	06907	11008	42751	27756	53498	18602	70659	90655	15053	21916	81825	44394	42880
7	99562	72905	56420	69994	98872	31016	71194	18738	44013	48840	63213	21069	10634	12952
8	96301	91977	05463	07972	18876	20922	94595	56869	69014	60045	18425	84903	42508	32307
9	89579	14342	63661	10281	17453	18103	57740	84378	25331	12566	58678	44947	05585	56941
10	85475	36857	53342	53988	53060	59533	38867	62300	08158	17983	16439	11458	18593	64952
11	28918	69578	88231	33276	70997	79936	56865	05859	90106	31595	01547	85590	91610	78188
12	63553	40961	48235	03427	49626	69445	18663	72695	52180	20847	12234	90511	33703	90322
13	09429	93969	52636	92737	88974	33488	36320	17617	30015	08272	84115	27156	30613	74952
14	10365	61129	87529	85689	48237	52267	67689	93394	01511	26358	85104	20285	29975	89868
15	07119	97336	71048	08178	77233	13916	47564	81056	97735	85977	29372	74461	28551	90707
16	51085	12765	51821	51259	77452	16308	60756	92144	49442	53900	70960	63990	75601	40719
17	02368	21382	52404	60268	89368	19885	55322	44819	01188	65255	64835	44919	05944	55157
18	01011	54092	33362	94904	31273	04146	18594	29852	71585	85030	51132	01915	92747	64951
19	52162	53916	46369	58586	23216	14513	83149	98736	23495	64350	94738	17752	35156	35749
20	07056	97628	33787	09998	42698	06691	76988	13602	51851	46104	88916	19509	25625	58104
21	48663	91245	85828	14346	09172	30168	90229	04734	59193	22178	30421	61666	99904	32812
22	54164	58492	22421	74103	47070	25306	76468	26384	58151	06646	21524	15227	96909	44592
23	32639	32363	05597	24200	13363	38005	94342	28728	35806	06912	17012	64161	18296	22851
24	29334	27001	87637	87308	58731	00256	45834	15398	46557	41135	10367	07684	36188	18510
25	02488	33062	28834	07351	19731	92420	60952	61280	50001	67658	32586	86679	50720	94953
26	81525	72295	04839	96423	24878	82651	66566	14778	76797	14780	13300	87074	79666	95725
27	29676	20591	68086	26432	46901	20849	89768	81536	86645	12659	92259	57102	80428	25280
28	00742	57392	39064	66432	84673	40027	32832	61362	98947	96067	64760	64584	96096	98253
29	05366	04213	25669	26422	44407	44048	37937	63904	45766	66134	75470	66520	34693	90449
30	91921	26418	64117	94305	26766	25940	39972	22209	71500	64568	91402	42416	07884	69618
31	00582	04711	87917	77341	42206	35126	74087	99547	81817	42607	43808	76655	62028	76630
32	00725	69884	62797	56170	86324	88072	76222	36086	84637	93161	76038	65855	77919	88006
33	69011	65795	95876	55293	18988	27354	26575	08625	40801	59920	29841	80150	12777	48501

TABLE 1 (continued)

Row	Column 1	2	3	4	5	6	7	8	9	10	11	12	13	14
34	25976	57948	29888	88604	67917	48708	18912	82271	65424	69774	33611	54262	85963	03547
35	09763	83473	73577	12908	30883	18317	28290	35797	05998	41688	34952	37888	38917	88050
36	91576	42595	27958	30134	04024	86385	29880	99730	55536	84855	29080	09250	79656	73211
37	17955	56349	90999	49127	20044	59931	06115	20542	18059	02008	73708	83517	36103	42791
38	46503	18584	18845	49618	02304	51038	20655	58727	28168	15475	56942	53389	20562	87338
39	92157	89634	94824	78171	84610	82834	09922	25417	44137	48413	25555	21246	35509	20468
40	14577	62765	35605	81263	39667	47358	56873	56307	61607	49518	89656	20103	77490	18062
41	98427	07523	33362	64270	01638	92477	66969	98420	04880	45585	46565	04102	46880	45709
42	34914	63976	88720	82765	34476	17032	87589	40836	32427	70002	70663	88863	77775	69348
43	70060	28277	39475	46473	23219	53416	94970	25832	69975	94884	19661	72828	00102	66794
44	53976	54914	06990	67245	68350	82948	11398	42878	80287	88267	47363	46634	06541	97809
45	76072	29515	40980	07391	58745	25774	22987	80059	39911	96189	41151	14222	60697	59583
46	90725	52210	83974	29992	65831	38857	50490	83765	55657	14361	31720	57375	56228	41546
47	64364	67412	33339	31926	14883	24413	59744	92351	97473	89286	35931	04110	23726	51900
48	08962	00358	31662	25388	61642	34072	81249	35648	56891	69352	48373	45578	78547	81788
49	95012	68379	93526	70765	10592	04542	76463	54328	02349	17247	28865	14777	62730	92277
50	15664	10493	20492	38391	91132	21999	59516	81652	27195	48223	46751	22923	32261	85653
51	16408	81899	04153	53381	79401	21438	83035	92350	36693	31238	59649	91754	72772	02338
52	18629	81953	05520	91962	04739	13092	97662	24822	94730	06496	35090	04822	86774	98289
53	73115	35101	47498	87637	99016	71060	88824	71013	18735	20286	23153	72924	35165	43040
54	57491	16703	23167	49323	45021	33132	12544	41035	80780	45393	44812	12515	98931	91202
55	30405	83946	23792	14422	15059	45799	22716	19792	09983	74353	68668	30429	70735	25499
56	16631	35006	85900	98275	32388	52390	16815	69298	82732	38480	73817	32523	41961	44437
57	96773	20206	42559	78985	05300	22164	24369	54224	35083	19687	11052	91491	60383	19746
58	38935	64202	14349	82674	66523	44133	00697	35552	35970	19124	63318	29686	03387	59846
59	31624	76384	17403	53363	44167	64486	64758	75366	76554	31601	12614	33072	60332	92325
60	78919	19474	23632	27889	47914	02584	37680	20801	72152	39339	34806	08930	85001	87820
61	03931	33309	57047	74211	63445	17361	62825	39908	05607	91284	68833	25570	38818	46920
62	74426	33278	43972	10119	89917	15665	52872	73823	73144	88662	88970	74492	51805	99378
63	09066	00903	20795	95452	92648	45454	09552	88815	16553	51125	79375	97596	16296	66092
64	42238	12426	87025	14267	20979	04508	64535	31355	86064	29472	47689	05974	52468	16834
65	16153	08002	26504	41744	81959	65642	74240	56302	00033	67107	77510	70625	28725	34191
66	21457	40742	29820	96783	29400	21840	15035	34527	33310	06116	95240	15957	16572	06004
67	21581	57802	02050	89728	17937	37621	47075	42080	97403	48626	68995	43805	33386	21597
68	55612	78095	83197	33732	05810	24813	86902	60397	16489	03264	88525	42786	05269	92532
69	44657	66999	99324	51281	84463	60563	79312	93454	68876	25471	93911	25650	12682	73572
70	91340	84979	46949	81973	37949	61023	43997	15263	80644	43942	89203	71795	99533	50501
71	91227	21199	31935	27022	84067	05462	35216	14486	29891	68607	41867	14951	91696	85065
72	50001	38140	66321	19924	72163	09538	12151	06878	91903	18749	34405	56087	82790	70925
73	65390	05224	72958	28609	81406	39147	25549	48542	42627	45233	57202	94617	23772	07896
74	27504	96131	83944	41575	10573	08619	64482	73923	36152	05184	94142	25299	84387	34925
75	37169	94851	39117	89632	00959	16487	65536	49071	39782	17095	02330	74301	00275	48280
76	11508	70225	51111	38351	19444	66499	71945	05422	13442	78675	84081	66938	93654	59894
77	37449	30362	06694	54690	04052	53115	62757	95348	78662	11163	81651	50245	34971	52924
78	46515	70331	85922	38329	57015	15765	97161	17869	45349	61796	66345	81073	49106	79860

Source: Reprinted with permission from W. H. Beyer, ed., *CRC Standard Mathematical Tables,* 24th ed. (abridged). Copyright CRC Press, Inc., Boca Raton, FL, 1976.

TABLE 2 Binomial probabilities

Tabulated values are $\sum_{x=0}^{k} p(x)$. (Computations are rounded at the third decimal place.)

(a) $n = 5$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.951	.774	.590	.328	.168	.078	.031	.010	.002	.000	.000	.000	.000
1	.999	.977	.919	.737	.528	.337	.188	.087	.031	.007	.000	.000	.000
2	1.000	.999	.991	.942	.837	.683	.500	.317	.163	.058	.009	.001	.000
3	1.000	1.000	1.000	.993	.969	.913	.812	.663	.472	.263	.081	.023	.001
4	1.000	1.000	1.000	1.000	.998	.990	.969	.922	.832	.672	.410	.226	.049

(b) $n = 10$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.904	.599	.349	.107	.028	.006	.001	.000	.000	.000	.000	.000	.000
1	.996	.914	.736	.376	.149	.046	.011	.002	.000	.000	.000	.000	.000
2	1.000	.988	.930	.678	.383	.167	.055	.012	.002	.000	.000	.000	.000
3	1.000	.999	.987	.879	.650	.382	.172	.055	.011	.001	.000	.000	.000
4	1.000	1.000	.998	.967	.850	.633	.377	.166	.047	.006	.000	.000	.000
5	1.000	1.000	1.000	.994	.953	.834	.623	.367	.150	.033	.002	.000	.000
6	1.000	1.000	1.000	.999	.989	.945	.828	.618	.350	.121	.013	.001	.000
7	1.000	1.000	1.000	1.000	.998	.988	.945	.833	.617	.322	.070	.012	.000
8	1.000	1.000	1.000	1.000	1.000	.998	.989	.954	.851	.624	.264	.086	.004
9	1.000	1.000	1.000	1.000	1.000	1.000	.999	.994	.972	.893	.651	.401	.096

TABLE 2 (continued)

(c) $n = 15$

k	p .01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.860	.463	.206	.035	.005	.000	.000	.000	.000	.000	.000	.000	.000
1	.990	.829	.549	.167	.035	.005	.000	.000	.000	.000	.000	.000	.000
2	1.000	.964	.816	.398	.127	.027	.004	.000	.000	.000	.000	.000	.000
3	1.000	.995	.944	.648	.297	.091	.018	.002	.000	.000	.000	.000	.000
4	1.000	.999	.987	.836	.515	.217	.059	.009	.001	.000	.000	.000	.000
5	1.000	1.000	.998	.939	.722	.403	.151	.034	.004	.000	.000	.000	.000
6	1.000	1.000	1.000	.982	.869	.610	.304	.095	.015	.001	.000	.000	.000
7	1.000	1.000	1.000	.996	.950	.787	.500	.213	.050	.004	.000	.000	.000
8	1.000	1.000	1.000	.999	.985	.905	.696	.390	.131	.018	.000	.000	.000
9	1.000	1.000	1.000	1.000	.996	.966	.849	.597	.278	.061	.002	.000	.000
10	1.000	1.000	1.000	1.000	.999	.991	.941	.783	.485	.164	.013	.001	.000
11	1.000	1.000	1.000	1.000	1.000	.998	.982	.909	.703	.352	.056	.005	.000
12	1.000	1.000	1.000	1.000	1.000	1.000	.996	.973	.873	.602	.184	.036	.000
13	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.833	.451	.171	.010
14	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.995	.965	.794	.537	.140

(d) $n = 20$

k	p .01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.818	.358	.122	.002	.001	.000	.000	.000	.000	.000	.000	.000	.000
1	.983	.736	.392	.069	.008	.001	.000	.000	.000	.000	.000	.000	.000
2	.999	.925	.677	.206	.035	.004	.000	.000	.000	.000	.000	.000	.000
3	1.000	.984	.867	.411	.107	.016	.001	.000	.000	.000	.000	.000	.000
4	1.000	.997	.957	.630	.238	.051	.006	.000	.000	.000	.000	.000	.000
5	1.000	1.000	.989	.804	.416	.126	.021	.002	.000	.000	.000	.000	.000
6	1.000	1.000	.998	.913	.608	.250	.058	.006	.000	.000	.000	.000	.000
7	1.000	1.000	1.000	.968	.772	.416	.132	.021	.001	.000	.000	.000	.000
8	1.000	1.000	1.000	.990	.887	.596	.252	.057	.005	.000	.000	.000	.000
9	1.000	1.000	1.000	.997	.952	.755	.412	.128	.017	.001	.000	.000	.000
10	1.000	1.000	1.000	.999	.983	.872	.588	.245	.048	.003	.000	.000	.000
11	1.000	1.000	1.000	1.000	.995	.943	.748	.404	.113	.010	.000	.000	.000
12	1.000	1.000	1.000	1.000	.999	.979	.868	.584	.228	.032	.000	.000	.000
13	1.000	1.000	1.000	1.000	1.000	.994	.942	.750	.392	.087	.002	.000	.000
14	1.000	1.000	1.000	1.000	1.000	.998	.979	.874	.584	.196	.011	.000	.000
15	1.000	1.000	1.000	1.000	1.000	1.000	.994	.949	.762	.370	.043	.003	.000
16	1.000	1.000	1.000	1.000	1.000	1.000	.999	.984	.893	.589	.133	.016	.000
17	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.965	.794	.323	.075	.001
18	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.992	.931	.608	.264	.017
19	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.999	.988	.878	.642	.182

TABLE 2 (continued)

(e) $n = 25$

k \ p	.01	.05	.10	.20	.30	.40	.50	.60	.70	.80	.90	.95	.99
0	.778	.277	.072	.004	.000	.000	.000	.000	.000	.000	.000	.000	.000
1	.974	.642	.271	.027	.002	.000	.000	.000	.000	.000	.000	.000	.000
2	.998	.873	.537	.098	.009	.000	.000	.000	.000	.000	.000	.000	.000
3	1.000	.966	.764	.234	.033	.002	.000	.000	.000	.000	.000	.000	.000
4	1.000	.993	.902	.421	.090	.009	.000	.000	.000	.000	.000	.000	.000
5	1.000	.999	.967	.617	.193	.029	.002	.000	.000	.000	.000	.000	.000
6	1.000	1.000	.991	.780	.341	.074	.007	.000	.000	.000	.000	.000	.000
7	1.000	1.000	.998	.891	.512	.154	.022	.001	.000	.000	.000	.000	.000
8	1.000	1.000	1.000	.953	.677	.274	.054	.004	.000	.000	.000	.000	.000
9	1.000	1.000	1.000	.983	.811	.425	.115	.013	.000	.000	.000	.000	.000
10	1.000	1.000	1.000	.994	.902	.586	.212	.034	.002	.000	.000	.000	.000
11	1.000	1.000	1.000	.998	.956	.732	.345	.078	.006	.000	.000	.000	.000
12	1.000	1.000	1.000	1.000	.983	.846	.500	.154	.017	.000	.000	.000	.000
13	1.000	1.000	1.000	1.000	.994	.922	.655	.268	.044	.002	.000	.000	.000
14	1.000	1.000	1.000	1.000	.998	.966	.788	.414	.098	.006	.000	.000	.000
15	1.000	1.000	1.000	1.000	1.000	.987	.885	.575	.189	.017	.000	.000	.000
16	1.000	1.000	1.000	1.000	1.000	.996	.946	.726	.323	.047	.000	.000	.000
17	1.000	1.000	1.000	1.000	1.000	.999	.978	.846	.488	.109	.002	.000	.000
18	1.000	1.000	1.000	1.000	1.000	1.000	.993	.926	.659	.220	.009	.000	.000
19	1.000	1.000	1.000	1.000	1.000	1.000	.998	.971	.807	.383	.033	.001	.000
20	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.910	.579	.098	.007	.000
21	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.967	.766	.236	.034	.000
22	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.991	.902	.463	.127	.002
23	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.998	.973	.729	.358	.026
24	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	.996	.928	.723	.222

TABLE 3
Poisson distribution function

$$F(x, \lambda) = \sum_{k=0}^{x} e^{-\lambda} \frac{\lambda^k}{k!}$$

λ	x: 0	1	2	3	4	5	6	7	8	9
.02	.980	1.000								
.04	.961	.999	1.000							
.06	.942	.998	1.000							
.08	.923	.997	1.000							
.10	.905	.995	1.000							
.15	.861	.990	.999	1.000						
.20	.819	.982	.999	1.000						
.25	.779	.974	.998	1.000						
.30	.741	.963	.996	1.000						
.35	.705	.951	.994	1.000						
.40	.670	.938	.992	.999	1.000					
.45	.638	.925	.989	.999	1.000					
.50	.607	.910	.986	.998	1.000					
.55	.577	.894	.982	.998	1.000					
.60	.549	.878	.977	.997	1.000					
.65	.522	.861	.972	.996	.999	1.000				
.70	.497	.844	.966	.994	.999	1.000				
.75	.472	.827	.959	.993	.999	1.000				
.80	.449	.809	.953	.991	.999	1.000				
.85	.427	.791	.945	.989	.998	1.000				
.90	.407	.772	.937	.987	.998	1.000				
.95	.387	.754	.929	.981	.997	1.000				
1.00	.368	.736	.920	.981	.996	.999	1.000			
1.1	.333	.699	.900	.974	.995	.999	1.000			
1.2	.301	.663	.879	.966	.992	.998	1.000			
1.3	.273	.627	.857	.957	.989	.998	1.000			
1.4	.247	.592	.833	.946	.986	.997	.999	1.000		
1.5	.223	.558	.809	.934	.981	.996	.999	1.000		
1.6	.202	.525	.783	.921	.976	.994	.999	1.000		
1.7	.183	.493	.757	.907	.970	.992	.998	1.000		
1.8	.165	.463	.731	.891	.964	.990	.997	.999	1.000	
1.9	.150	.434	.704	.875	.956	.987	.997	.999	1.000	
2.0	.135	.406	.677	.857	.947	.983	.995	.999	1.000	
2.2	.111	.355	.623	.819	.928	.975	.993	.998	1.000	
2.4	.091	.308	.570	.779	.904	.964	.988	.997	.999	1.000
2.6	.074	.267	.518	.736	.877	.951	.983	.995	.999	1.000
2.8	.061	.231	.469	.692	.848	.935	.976	.992	.998	.999
3.0	.050	.199	.423	.647	.815	.916	.966	.988	.996	.999

TABLE 3 (continued)

λ	x = 0	1	2	3	4	5	6	7	8	9
3.2	.041	.171	.380	.603	.781	.895	.955	.983	.994	.998
3.4	.033	.147	.340	.558	.744	.871	.942	.977	.992	.997
3.6	.027	.126	.303	.515	.706	.844	.927	.969	.988	.996
3.8	.022	.107	.269	.473	.668	.816	.909	.960	.984	.994
4.0	.018	.092	.238	.433	.629	.785	.889	.949	.979	.992
4.2	.015	.078	.210	.395	.590	.753	.867	.936	.972	.989
4.4	.012	.066	.185	.359	.551	.720	.844	.921	.964	.985
4.6	.010	.056	.163	.326	.513	.686	.818	.905	.955	.980
4.8	.008	.048	.143	.294	.476	.651	.791	.887	.944	.975
5.0	.007	.040	.125	.265	.440	.616	.762	.867	.932	.968
5.2	.006	.034	.109	.238	.406	.581	.732	.845	.918	.960
5.4	.005	.029	.095	.213	.373	.546	.702	.822	.903	.951
5.6	.004	.024	.082	.191	.342	.512	.670	.797	.886	.941
5.8	.003	.021	.072	.170	.313	.478	.638	.771	.867	.929
6.0	.002	.017	.062	.151	.285	.446	.606	.744	.847	.916
6.2	.002	.015	.054	.134	.259	.414	.574	.716	.826	.902
6.4	.002	.012	.046	.119	.235	.384	.542	.687	.803	.886
6.6	.001	.010	.040	.105	.213	.355	.511	.658	.780	.869
6.8	.001	.009	.034	.093	.192	.327	.480	.628	.755	.850
7.0	.001	.007	.030	.082	.173	.301	.450	.599	.729	.830
7.2	.001	.006	.025	.072	.156	.276	.420	.569	.703	.810
7.4	.001	.005	.022	.063	.140	.253	.392	.539	.676	.788
7.6	.001	.004	.019	.055	.125	.231	.365	.510	.648	.765
7.8	.000	.004	.016	.048	.112	.210	.338	.481	.620	.741
8.0	.000	.003	.014	.042	.100	.191	.313	.453	.593	.717
8.5	.000	.002	.009	.030	.074	.150	.256	.386	.523	.653
9.0	.000	.001	.006	.021	.055	.116	.207	.324	.456	.587
9.5	.000	.001	.004	.015	.040	.089	.165	.269	.392	.522
10.0	.000	.000	.003	.010	.029	.067	.130	.220	.333	.458
10.5	.000	.000	.002	.007	.021	.050	.102	.179	.279	.397
11.0	.000	.000	.001	.005	.015	.038	.079	.143	.232	.341
11.5	.000	.000	.001	.003	.011	.028	.060	.114	.191	.289
12.0	.000	.000	.001	.002	.008	.020	.046	.090	.155	.242
12.5	.000	.000	.000	.002	.005	.015	.035	.070	.125	.201
13.0	.000	.000	.000	.001	.004	.011	.026	.054	.100	.166
13.5	.000	.000	.000	.001	.003	.008	.019	.041	.079	.135
14.0	.000	.000	.000	.000	.002	.006	.014	.032	.062	.109
14.5	.000	.000	.000	.000	.001	.004	.010	.024	.048	.088
15.0	.000	.000	.000	.000	.001	.003	.008	.018	.037	.070

TABLE 3
(continued)

	x									
λ	10	11	12	13	14	15	16	17	18	19
2.8	1.000									
3.0	1.000									
3.2	1.000									
3.4	.999	1.000								
3.6	.999	1.000								
3.8	.998	.999	1.000							
4.0	.997	.999	1.000							
4.2	.996	.999	1.000							
4.4	.994	.998	.999	1.000						
4.6	.992	.997	.999	1.000						
4.8	.990	.996	.999	1.000						
5.0	.986	.995	.998	.999	1.000					
5.2	.982	.993	.997	.999	1.000					
5.4	.977	.990	.996	.999	1.000					
5.6	.972	.988	.995	.998	.999	1.000				
5.8	.965	.984	.993	.997	.999	1.000				
6.0	.957	.980	.991	.996	.999	.999	1.000			
6.2	.949	.975	.989	.995	.998	.999	1.000			
6.4	.939	.969	.986	.994	.997	.999	1.000			
6.6	.927	.963	.982	.992	.997	.999	.999	1.000		
6.8	.915	.955	.978	.990	.996	.998	.999	1.000		
7.0	.901	.947	.973	.987	.994	.998	.999	1.000		
7.2	.887	.937	.967	.984	.993	.997	.999	.999	1.000	
7.4	.871	.926	.961	.980	.991	.996	.998	.999	1.000	
7.6	.854	.915	.954	.976	.989	.995	.998	.999	1.000	
7.8	.835	.902	.945	.971	.986	.993	.997	.999	1.000	
8.0	.816	.888	.936	.966	.983	.992	.996	.998	.999	1.000
8.5	.763	.849	.909	.949	.973	.986	.993	.997	.999	.999
9.0	.706	.803	.876	.926	.959	.978	.989	.995	.998	.999
9.5	.645	.752	.836	.898	.940	.967	.982	.991	.996	.998
10.0	.583	.697	.792	.864	.917	.951	.973	.986	.993	.997
10.5	.521	.639	.742	.825	.888	.932	.960	.978	.988	.994
11.0	.460	.579	.689	.781	.854	.907	.944	.968	.982	.991
11.5	.402	.520	.633	.733	.815	.878	.924	.954	.974	.986
12.0	.347	.462	.576	.682	.772	.844	.899	.937	.963	.979
12.5	.297	.406	.519	.628	.725	.806	.869	.916	.948	.969
13.0	.252	.353	.463	.573	.675	.764	.835	.890	.930	.957
13.5	.211	.304	.409	.518	.623	.718	.798	.861	.908	.942
14.0	.176	.260	.358	.464	.570	.669	.756	.827	.883	.923
14.5	.145	.220	.311	.413	.518	.619	.711	.790	.853	.901
15.0	.118	.185	.268	.363	.466	.568	.664	.749	.819	.875

TABLE 3 (continued)

λ	x 20	21	22	23	24	25	26	27	28	29
8.5	1.000									
9.0	1.000									
9.5	.999	1.000								
10.0	.998	.999	1.000							
10.5	.997	.999	.999	1.000						
11.0	.995	.998	.999	1.000						
11.5	.992	.996	.998	.999	1.000					
12.0	.988	.994	.997	.999	.999	1.000				
12.5	.983	.991	.995	.998	.999	.999	1.000			
13.0	.975	.986	.992	.996	.998	.999	1.000			
13.5	.965	.980	.989	.994	.997	.998	.999	1.000		
14.0	.952	.971	.983	.991	.995	.997	.999	.999	1.000	
14.5	.936	.960	.976	.986	.992	.996	.998	.999	.999	1.000
15.0	.917	.947	.967	.981	.989	.994	.997	.998	.999	1.000

λ	x 4	5	6	7	8	9	10	11	12	13
16	.000	.001	.004	.010	.022	.043	.077	.127	.193	.275
17	.000	.001	.002	.005	.013	.026	.049	.085	.135	.201
18	.000	.000	.001	.003	.007	.015	.030	.055	.092	.143
19	.000	.000	.001	.002	.004	.009	.018	.035	.061	.098
20	.000	.000	.000	.001	.002	.005	.011	.021	.039	.066
21	.000	.000	.000	.000	.001	.003	.006	.013	.025	.043
22	.000	.000	.000	.000	.001	.002	.004	.008	.015	.028
23	.000	.000	.000	.000	.000	.001	.002	.004	.009	.017
24	.000	.000	.000	.000	.000	.000	.001	.003	.005	.011
25	.000	.000	.000	.000	.000	.000	.001	.001	.003	.006

TABLE 3
(continued)

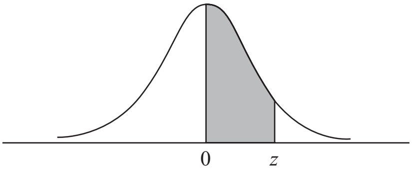
λ	*x* 14	15	16	17	18	19	20	21	22	23
16	.368	.467	.566	.659	.742	.812	.868	.911	.942	.963
17	.281	.371	.468	.564	.655	.736	.805	.861	.905	.937
18	.208	.287	.375	.469	.562	.651	.731	.799	.855	.899
19	.150	.215	.292	.378	.469	.561	.647	.725	.793	.849
20	.105	.157	.221	.297	.381	.470	.559	.644	.721	.787
21	.072	.111	.163	.227	.302	.384	.471	.558	.640	.716
22	.048	.077	.117	.169	.232	.306	.387	.472	.556	.637
23	.031	.052	.082	.123	.175	.238	.310	.389	.472	.555
24	.020	.034	.056	.087	.128	.180	.243	.314	.392	.473
25	.012	.022	.038	.060	.092	.134	.185	.274	.318	.394

λ	*x* 24	25	26	27	28	29	30	31	32	33
16	.978	.987	.993	.996	.998	.999	.999	1.000		
17	.959	.975	.985	.991	.995	.997	.999	.999	1.000	
18	.932	.955	.972	.983	.990	.994	.997	.998	.999	1.000
19	.893	.927	.951	.969	.980	.988	.993	.996	.998	.999
20	.843	.888	.922	.948	.966	.978	.987	.992	.995	.997
21	.782	.838	.883	.917	.944	.963	.976	.985	.991	.994
22	.712	.777	.832	.877	.913	.940	.959	.973	.983	.989
23	.635	.708	.772	.827	.873	.908	.936	.956	.971	.981
24	.554	.632	.704	.768	.823	.868	.904	.932	.953	.969
25	.473	.553	.629	.700	.763	.818	.863	.900	.929	.950

λ	*x* 34	35	36	37	38	39	40	41	42	43
19	.999	1.000								
20	.999	.999	1.000							
21	.997	.998	.999	.999	1.000					
22	.994	.996	.998	.999	.999	1.000				
23	.988	.993	.996	.997	.999	.999	1.000			
24	.979	.987	.992	.995	.997	.998	.999	.999	1.000	
25	.966	.978	.985	.991	.995	.997	.998	.999	.999	1.000

Source: Reprinted by permission from E. C. Molina, *Poisson's Exponential Binomial Limit* (Princeton, N.J.: D. Van Nostrand Company, 1947).

TABLE 4
Normal curve areas



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0174	.0753
.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: Abridged from Table I of A. Hald. *Statistical Tables and Formulas* (New York: John Wiley & Sons, 1952). Reproduced by permission of A. Hald and the publisher.

Answers to Selected Problems

Chapter 2

2.5 Let F and M be the events that a randomly selected baby is femle or male, respectively

a $S = \{F, M\}$ **b** $\emptyset, \{F\}, \{M\}, S$ **c** $\{F\}, S$

2.7 **a** WB; $n(WB) = 4$ **b** $W \cup B$; $n(W \cup B) = 16$

c $\bar{W}B$; $n(\bar{W}B) = 4$ **d** $\bar{W}B \cup W\bar{B}$; $n(\bar{W}B \cup W\bar{B}) = 12$

2.9 Let B, G, GF, and N be the outcomes that the selected fish is a black molly, guppy, goldfish, and neon, respectively

a $S = \{(B, G), (B, GF), (B, N), (G, GF), (G, N), (GF, N)\}$

b $A = \{(B, G), (B, GF), (B, N)\}$; $n(A) = 3$

c Let E be the event of selections containing a guppy but no goldfish or neon
$E = \{(G, B)\}$; $n(E) = 1$

d $\bar{A} = \{(G, GF), (G, N), (GF, N)\}$
$AB \cup CD = \{(B, G), (GF, N)\}$
$\overline{AB \cup CD} = \{(B, GF), (B, N), (G, GF), (G, N)\}$
$\overline{(A \cup C)(B \cup D)} = \{(B, GF), (G, N)\}$

2.11 **a** T; $n(T) = 47$ **b** $\bar{T} \cap \bar{E}$; $n(\bar{T}\bar{E}) = 6$

c $E\bar{T}$; $n(E\bar{T}) = 7$ **d** $E\bar{T} \cup \bar{E}T$; $n(E\bar{T} \cup \bar{E}T) = 44$

2.13 Let S, L, and N be the events that the student reads the school, local, and national papers respectively

a $P(\overline{S \cup L \cup N}) = 0.18$ **b** $P(L) = 0.30$

c $P(SL\bar{N} \cup S\bar{L}N \cup \bar{S}LN) = 0.15$

d $P(SL\bar{N} \cup S\bar{L}N \cup \bar{S}LN \cup SLN) = 0.16$

2.17 **a** 1/2 **b** 1/6 **c** 5/6

2.19 **a** 0.78 **b** 0.12 **c** 0.73

2.21 **a** 0.5 **b** 0.59 **c** 0.3 **d** 0.14

e 0.25

2.23 **a** 0.06 **b** 0.16 **c** 0.14 **d** 0.84

2.25 **a** 734 **b** 159

2.35 **a** $S = \{(I, I), (I, II), (II, I), (II, II)\}$ **b** 0.25 **c** 0.5

2.37 40

2.39 24

2.41 210

2.43 **a** $15! = 1.31 \times 10^{12}$ **b** $8!\,7! = 203,212,800$ **c** $(2)8!\,7! = 406,425,600$

2.45 2520

2.47 0.40

2.49 **a** 24 **b** 0.5

2.51 **a** 1/1024 **b** 0.38

2.53 0.0022

2.55 $1/4^{35}$

2.57 **a** 1680 **b** 1/12

2.59 0.4914

2.61 **a** 0.09375 **b** 0.14 **c** 0.58 **d** 0.56

2.63 **a** $\frac{1}{r!}$ **b** $\frac{1}{2!} - \frac{1}{3!} + \frac{1}{4!} + \cdots + \frac{(-1)^{N+1}}{N!}$

2.67 78

2.69 **a** 0.000729

2.71 **a** 0.729 **b** 0.729

2.73 Let L, R, and S represent the outcomes that the car moves left, right, and straight, respectively, at the intersection

a $S = \{L, R, S\}$ **b** $\emptyset, \{L\}, \{R\}, \{S\}, \{L, R\}, \{L, S\}, \{R, S\}, S$

c $\{R\}, \{L, R\}, \{R, S\}, S$ **d** 0.667

2.75 Let A, W, and D represent the outcomes that a randomly selected student was admitted, wait listed, or denied, respectively, to medical school

a $S = \{A, W, D\}$ **b** $\emptyset, \{A\}, \{W\}, \{D\}, \{A, W\}, \{A, D\}, \{W, D\}, S$

c $\{A\}, \{A, W\}, \{A, D\}, S$ **d** $P(A) = 0.2,\ P(W) = 0.1,\ P(D) = 0.6$

2.77 **a** 36 **b** 0.167

2.79 **a** 0.28 **b** 0.28 **c** 0.20

2.81 120

2.83 19,380

2.85 **a** 0.00024 **b** 0.23

2.87 **a** 177,100 **b** 0.31 **c** 0.9988

2.89 0.00001

2.91 Close to 0

Chapter 3

3.1 **a** 0.583 **b** 0.468 **c** 0.281 **d** 0.719

3.3 **a** 0.019 **b** 0.46 **c** 0.52 **d** 0.813

3.5 0.5

3.7 0.1092

3.9 0.09

3.11 **a** 0.632 **b** 0.368 **c** 0.652

3.13 0.1333

3.15 **a** 1 **b** 0.988 **c** 0.972

3.17 0.75

3.19 **a** 0.92 **b** 0.87

3.23 **a** Percentage of U.S. Men in Each Age Group

Age Group	16 to 24	25 to 44	45 to 64	65 and Older
Probability	0.36	0.30	0.24	0.10

b Conditional Distribution of Age Group for Women

Age Group	16 to 24	25 to 44	45 to 64	65 and Older
Probability	0.34	0.28	0.24	0.14

3.25 **a** Conditional distribution of speed limit given a fatal crash in a rural area

Speed Limit	≤ 30	35 or 40	45 or 50	55	60 or higher	No statuatory limit
P(Speed Limit | Rural)	0.043	0.090	0.162	0.446	0.254	0.004

b Conditional distribution of speed limit given a fatal crash in an urban area

Speed Limit	≤ 30	35 or 40	45 or 50	55	60 or higher	No statuatory limit
P(Speed Limit | Rural)	0.190	0.289	0.230	0.137	0.152	0.002

3.27 **a** 0.6 **b** 0.4

3.31 $P(AB) = P(A) = P(B) = P(O) = 1/4$

3.33 0.82

3.35 0.9912

3.39 **a** W/M **b** k/n

3.41 **a** $2(1-p)p^3 + 2p(1-p)^3$ **b** $\dfrac{p^2}{1-2p(1-p)}$

3.43 0.5

3.45 0.92

3.47 0.000945

3.49 **a** 0.183 **b** 0.034

3.51 $\dfrac{np}{np+1-p}$

3.53 0.111

3.55 **b** Level 159:

Group	MI Yes	MI No
Aspirin	2	380
Placebo	9	397

OR = 0.232

Level 160–209:

Group	MI Yes	MI No
Aspirin	12	1575
Placebo	37	1474

OR = 0.306

Level 210–259:

Group	MI Yes	MI No
Aspirin	26	1409
Placebo	43	1401

OR = 0.601

Level Greater than or equal to 260:

Group	MI Yes	MI No
Aspirin	14	568
Placebo	23	547

OR = 0.586

3.57 **a**

Treatment	Infection +	Infection −
Vaccine	0	768
Placebo	41	720

b 0.057 **c** 0 **d** 0

3.59 **a** 0.712 **b** 0.25 **c** 0.412 **d** 0.0442

3.61 **a** 0.253 **b** 0.12 **c** 0.28 **d** 0.23

3.63 Let F and M be the events that a randomly selected person is female and male, respectively. Let S be the event that a randomly selected person has smoked a cigarette during the past month. Let A be the age of a randomly selected person.

a

Age	$P(A\|FS)$
12–17	0.054
18–25	0.198
26+	0.748

b

Age	$P(A\|MS)$
12–17	0.045
18–25	0.218
26+	0.737

c

Age	$P(A\|F\bar{S})$
12–17	0.113
18–25	0.108
26+	0.778

d

Age	$P(A\|M\bar{S})$
12–17	0.136
18–25	0.108
26+	0.756

3.65 **a** 0.3125 **b** 0.234

3.67 0.986

3.69 **a** 0.000061 **b** 0.141

3.71 **a** 5.6 **b** 58958.7 **c** 0.000137

3.77 0.5

Chapter 4

4.1 **a**

x	0	1	2
$p(x)$	$\frac{2}{15}$	$\frac{8}{15}$	$\frac{1}{3}$

c

$$F(x) = \begin{cases} 0, & x < 0 \\ \dfrac{2}{15}, & 0 \le x < 1 \\ \dfrac{2}{3}, & 1 \le x < 2 \\ 1, & x \ge 1 \end{cases}$$

4.3 **a**

x	0	1	2	3	4
$p(x)$	0.0625	0.25	0.375	0.25	0.0625

c

x	$x < 0$	$0 \le x < 1$	$1 \le x < 2$	$2 \le x < 3$	$3 \le x < 4$	$4 \le x$
$F(x)$	0	0.0625	0.3125	0.6875	0.9375	1

4.5 Assume each person is equally likely to choose each entrance.

x	0	1	2	3
$p(x)$	0.422	0.422	0.141	0.016

4.7 **a**

x	0	1	2	3	4	5	6
$p(x)$	0.25	0.3	0.24	0.14	0.0525	0.015	0.0025

b 0.45

4.9 **a**

x	0	1	2	3
$p(x)$	0.30	0.44	0.22	0.04

y	0	1	2	3
$p(y)$	0.857	0.135	0.007	0.000125

b

$x+y$	0	1	2	3
$p(x+y)$	0.235	0.437	0.2718	0.056

4.11 **b**

x	1	2	3	4	5	6
$p(x)$	0.05	0.10	0.20	0.30	0.20	0.15

4.13 $P(X = x) = \frac{1}{n}$, $x = 1, 2, \ldots, n$

4.15 Let X = the score from one arrow

a

x	1	2	3	4	5	6	7	8	9	10
$p(x)$	0.19	0.17	0.15	0.13	0.11	0.09	0.07	0.05	0.03	0.01

b

x	1	2	3	4	5	6	7	8	9	10
$p(x)$	0.19	0.17	0.15	0.13	0.11	0.09	0.07	0.05	0.03	0.01

4.17 **a** $\mathrm{E}(G_1) = 0,\ \mathrm{V}(G_1) = 0.667$ **b** $\mathrm{E}(G_2) = 0,\ \mathrm{V}(G_2) = 2.4$

4.19 **a**

x (midpoint of age interval)	7	17	22	27	32	37
$p(x)$	0.0026	0.0079	0.0431	0.0861	0.1153	0.1934
x (midpoint of age interval)	42	47	52	57	62	75*
$p(x)$	0.2107	0.1504	0.0947	0.0501	0.0240	0.0217

b $\mu = 41.08,\ \sigma = 10.86$

4.21 **a**

x (midpoint of age interval)	7	17	22	27	32	37
$p(x)$	0.051	0.197	0.202	0.136	0.120	0.108
x (midpoint of age interval)	42	47	52	57	62	75*
$p(x)$	0.087	0.051	0.025	0.012	0.006	0.005

b $\mu = 28.44,\ \sigma = 1.92$

4.23 $E(X) = 0.2,\ V(X) = 0.18$

4.25 $E(X) = 1,\ V(X) = 0.5$

4.27 **a** (3, 7) breakdowns.

4.29 **a** (109.2, 140.8) minutes.

4.31 No, the mean net revenue is -7¢.

4.33 2 items

4.35 **a** 0.015 **b** $E(X) = 3.9,\ V(X) = 0.269$

4.37

x	0	1	2	3	4
$p(x)$	0.42	0.5	0.083	0	0.042

4.39 **a** $\mu = 3.85,\ \sigma = 2.35$ **b** $\mu = 19.25,\ \sigma = 11.75$

4.41 $E(X) = 0.25$
$V(X) = 0.5625$

4.43 **a** 0.1323 **b** 0.96927 **c** 0.16308 **d** 1.5
e 1.05

4.45 **a** 0.5405 **b** 0.1117

4.47 **a** 0.6229 **b** 0.2235 **c** 9 donors

4.49 **b** 0.7379

4.51 **a** 0.25 **b** 0.00098

4.53 **a** 14 people

4.55 \$3.96 **4.57** 7260 **4.59** $A(p + 1.05)$ **4.61** 0.617

4.63 **a** 0.429 **b** $2.982\,P_1 + 0.018\,P_2 - 2.262\,C$

4.65 **a** 0.81 **b** 0.81

4.67 0.009

4.69 **a** 0.00012 **b** 0.00074

4.71 **a** 0.80 **b** 0.032 **c** 0.20

4.73 $E(C) = \$400,\ V(C) = \$24,000$

4.75 **a** 0.141 **b** 0.0672

4.77 **a** 0.02 **b** $E(X + 4) = 4.21,\ V(X + 4) = 0.22$

4.79 **a** 0.9728 **b** 15 customers

4.81 **a** $p = 0.60,\ r = 36$ **b** 0.0000000103 **c** > 0.999

4.83 **a** 0.216, 0.26, 0.21 **b** 0.648

4.85 $\binom{2N-k}{N}(0.5)^{(2N-k)}$

4.89 **a** 0.156 **b** 0.629 **c** 0.371 **d** 0.371

4.91 **a** 0.363 **b** 0.999

4.93 **a** 0.60 **b** 0.296 **c** 0.993

4.95 **a** $\mu = 70$ minutes, $\sigma^2 = 700$ minutes^2

4.97 **a** 0.0002 **b** 0.607 **c** 0.393

4.99 **a** 0.042 **b** 0.082

4.101 **a** $\mu = 2.427$ goals/game

b

x	0	1	2	3	4
Expected	61.5	149.2	181.0	146.4	88.8

x	5	6	7	8	≥ 9
Expected	43.1	17.4	6.0	1.8	0.65

4.103 **a** $P(Y = y) = \dfrac{\dfrac{\lambda^y}{y!}e^{-\lambda}}{1 - e^{-\lambda}}, \ y = 1, 2, \ldots$ **b** $E(Y) = \dfrac{\lambda}{1 - e^{-\lambda}}$

4.109 **a** 0.14 **b** 0.889 **c** 0.312

4.111 **a** 0.576 **b** 0.9697 **c** 0.998

4.113 0.0333

4.115 **a** 0.2 **b** 0.8

4.117 0.476

4.119 **a**

x	0	1	2
$p(x)$	0.47	0.47	0.07

b

x	0	1	2	3
$p(x)$	0.17	0.50	0.30	0.033

4.121 **a** \$160 **b** (\$158.37, \$161.63)

4.123 **a** 1 **b** 0.75 **c** 0.95 **d** 0.86

4.125 $pe^t + q, t \in \Re$

4.133 $P(t) = e^{-\lambda(1-t)}$

4.135 $E(X) = \lambda$
$E[(X)(X-1)] = \lambda^2$

4.145 (i) $p = 0.2$

x	0	1	2	3	4	5	6	7
$p(x)$	0.2097	0.367	0.2753	0.1147	0.0287	0.0043	0.0004	0.00001

(ii) $p = 0.05$

x	0	1	2	3	4	5	6	7
$p(x)$	0.0078	0.0547	0.1641	0.2734	0.2734	0.1641	0.0547	0.0078

(iii) $p = 0.05$

x	0	1	2	3	4	5	6	7
$p(x)$	0.0001	0.0044	0.0043	0.0287	0.1147	0.2753	0.367	0.2097

4.147 **a** 0.41 **b** 0.9997

4.149 **a** 1 **b** 0.17 **c** 0 **d** 0.59
e 0.031

4.151 **a** 0.018

4.153 **a** 0.93 **b** $\mu = 15, \sigma = 3.87$ **c** (7.26, 22.74)

4.155 **a** 5 **b** 0.00674 **c** 0.125

4.157 **a** 0.25 **b** 0.016 **c** 0.016

4.159 0.8

4.161 0.04096

4.163 **a** 0.463

b

x	0	1	2	3	4	≥ 5
Expected	406.3	188.9	44.0	7.1	0.7	0.1

c

x	0	1	2	3	4	≥ 5
Expected	441.6	140.2	44.5	14.1	4.5	2.1

4.165 **a** 0.837

4.167 **a** 0.081 **b** 0.81 **c** 0.5905 **d** 0.0009

4.169 **a** 0.019 **b** 0.1745

4.171 **a**

x	0	1	2	3	4
$p(x)$	0.3585	0.3774	0.1887	0.0596	0.0133

b

x	0	1	2	3	4
$p(x)$	0.379	0.3679	0.1839	0.0613	0.0153

Chapter 5

5.1 **a** X = distance a randomly selected insect moves after pupation; Continuous
b X = number of stem cells that differentiate into brain cells in a randomly selected baby
Discrete
c X = proportion of toxins in a randomly selected volume of water
Continuous
d X = length (in cm) of randomly selected fish
Continuous

5.3 **a** 0.222 **b** 0.778 **c** 0.457

d $$F(x) = \begin{cases} 0, & x < -1 \\ \frac{x^3}{9} + \frac{1}{9}, & -1 \leq x < 2 \\ 1, & x \geq 2 \end{cases}$$

5.5 **a** 0.10 **b** $F(x) = \begin{cases} 1 - e^{-\frac{x}{10}} & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$ **c** 0.2231
d 0.1653

5.7 **b** 0.2031 **c** $f(x) = \begin{cases} \frac{3}{2}x^2 & 0 \le x < 1 \\ \frac{1}{2} & 1 \le x \le 2 \\ 0 & \text{elsewhere} \end{cases}$

5.9 **b** $F(x) = \begin{cases} \frac{x^3}{16} - \frac{3x^4}{256} & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$ **c** 0.6875 **d** 0.2617

5.11 **a** 0.75 **b** 0.80 **c** 1

5.13 $E(X) = 4,\ \sigma_x = 0.89$

5.15 **a** $E(X) = 0.5,\ V(X) = 0.05$ **b** (0.05, 0.95) **c** (0.22, 0.78)

5.17 **a** $E(X) = 2.4,\ V(X) = 0.64,\ \sigma_X = 0.8$ **b** (0.8, 4)

5.19 **a** $E(X) = 0.667,\ V(X) = 0.06,\ \sigma_X = 0.24$

b $E(Y) = \$186.80,\ V(Y) = 8817.86$ dollars^2

c $(-\$1, \$374.60)$

5.21 3 months

5.23 \$9.22

5.25 $B = -20,\ A = 15.33.$

5.27 **a** $\frac{b-c}{b-a}$ **b** $\frac{b-d}{b-c}$

5.29 **a** 0.25 **b** 0.0833 **c** 0.1667

5.31 0.2

5.33 **a** 0.375 **b** 21.6 hours

5.35 **a** 0.25 **b** 0.50

5.37 $\mu = 0.000006\pi$ cm, $\sigma^2 = 9.64 \times 10^{-10}\pi^2\text{cm}^2$

5.39 **a** 0.5 **b** 0.25 **c** 0.375

5.41 **a** 0.647 **b** 0.189 **c** 0.148

5.43 **a** $\sigma^2 = 2.4^2,\ \sigma = 2.4$ **b** 5.53

5.45 **a** $E(X) = 0.5,\ V(X) = 0.25$ **b** 0.998

5.47 **a** $E(X) = 0.667,\ V(X) = 0.444$ **b** 0.223 **c** 0.26

d 0.367

5.49 **a** 0.51 **b** 1936

5.51 **a** 0.61 **b** 0.78

5.53 **a** 0.082 **b** 0.0273

5.55 **a** 0.257 **b** 0.263 **c** 0.310

5.57 **a** 0.667 **b** 0 **c** 0.133

5.59 **a** 0.5553 **b** 0.5553

5.61 **a** 0.2425 **b** 0.2457 **c** $k = 82.89$ minutes

5.63 $m = 0.6931\vartheta$

5.65 **a** $E(X) = 8,\ V(X) = 160,\ \sigma_X = 12.6$ **b** (0 mm, 33.2 mm)

5.67 **a** $E(L) = 228.4,\ V(L) = 26,223.75$ **b** (\$0, \$716.60)

5.69 **a** $E(Y) = 20,\ V(Y) = 200$

b Let A be the average time to complete both tasks, $E(A) = 10,\ V(A) = 50$

5.71 **a** 0.082 **b** 0.393 **c** 0.318

d Gamma $(\alpha = 2,\ \beta = 20)$, $f(x) = \begin{cases} \frac{1}{400}xe^{-x/20}, & x \ge 0 \\ 0, & \text{otherwise} \end{cases}$

5.73 **a** $E(Y) = 140,\ V(Y) = 280$

$$f(y) = \begin{cases} \frac{1}{\Gamma(70)2^{70}} y^{69} e^{\frac{-y}{2}}, & y > 0 \\ 0 & y \le 0 \end{cases}$$

b 206.93

5.75 $\frac{\beta^k \Gamma(\alpha+k)}{\Gamma(\alpha)}$

5.77 **a** 0.7881 **b** 0.84 **c** 1.08 **d** −0.50 **e** 1.645 **f** 0.13

5.79 **a** 0 **b** 0.3085 **c** 0.3399 **d** 0.5595 **e** 0.0455

5.81 **a** 0.3413 **b** 0.3413 **c** 0.2857 **d** 0.3830 **e** 0.8664

5.83 **a** 12 **b** 14.985 **c** 14.22 **d** 11.32

5.85 0.0062

5.87 **a** 0.0730 **b** $\mu = 1.00$

5.89 **a** 0.95 **b** 0.83

5.91 0.5859

5.93 **a** 0.1498 **b** 0.0224

5.95 **a** 0.0548 **b** 203.95 hours

5.97 **a** 11.67 ounces **b** 0.871

5.101 **b** $E(X) = 64.30,\ V(X) = 6.50,\ \sigma_x = 2.55$

5.105 **a** 105 **b** $E(X) = 0.625$ $V(X) = 0.026$

5.107 **a** 0.6 **b** 0.0272

5.109 **a** $E(C) = 17.33,\ V(C) = 29.96$ **b** (6.38, 28.28)

5.111 $E(X) = 0.667$, Angle $= 240°$

5.113 **a** $E(X) = 0.003,\ V(X) = 0.000008353$ **b** 0.9685

5.115 **a** $E(X) = 0.5,\ V(X) = 0.0357$ **b** $E(X) = 0.5,\ V(X) = 0.05$ **c** $E(X) = 0.5,\ V(X) = 0.0833$ **d** Case (a)

5.119 **a** 0.632 **b** $\sqrt{\pi}$

5.121 0.6576 **5.123** 0.031 **5.125** 0.0657

5.127 $\gamma = 6.22,\ \theta = 43.52$ **5.129** $R(t) = e^{\frac{-t_2}{\theta}},\ t_2 > 0$

5.131 **a** $r(t) = \frac{\gamma}{\theta} t^{\gamma-1} \quad t > 0$

5.133 **a** $$f(x) = \begin{cases} \frac{e^{-\frac{(x-\mu)}{\beta}}}{\beta\left(1+e^{-\frac{(x-\mu)}{\beta}}\right)^2} & -\infty < x < \infty \end{cases}$$ **b** $r(t) = \frac{1}{\beta}$

5.137 $M(t) = \frac{e^{bt}-e^{at}}{t(b-a)}$

5.139 $\frac{2}{ct^2}\left(e^{t\sqrt{c}}\left(t\sqrt{c}-1\right)+1\right)$

5.143 $M_{Z^2}(t) = (1-2t)^{-\frac{1}{2}}$

Gamma $\left(\alpha = \frac{1}{2},\ \beta = 2\right)$

5.147 0.4 lbs

5.149 **a** Bernoulli ($p = 0.8$) **b** Exponential ($\lambda = 1$)

c $F(x) = \begin{cases} 0, & x < 0 \\ 0.2, & x = 0 \\ 1 - 0.8\, e^{-x}, & x > 0 \end{cases}$

5.151 **a** 0.5 **b** 0.25

5.153 **a** -0.375 **b** $F(x) = \begin{cases} 0 & x < 0 \\ \frac{x^2}{2} - \frac{x^3}{8} & 0 \le x \le 2 \\ 1 & x > 2 \end{cases}$

d 0, 0.1094, 0.375 **e** 0.625

f $E(X) = 1.1667,\ V(X) = 0.2389$

5.155 **b** $f(x) = \begin{cases} \frac{2}{9}x(3 - x), & 0 < x \le 3 \\ 0, & \text{otherwise} \end{cases}$ **d** 0.4259

e $E(X) = 1.5,\ V(X) = 5.85$

5.157 **a** 0.1151 **b** 0.1587 **c** 0.0015

5.159 **a** 4 **b** $E(X) = 1,\ V(X) = 0.50$ **c** $M(t) = \frac{4}{(2-t)^2}$

5.161 **a** 0.9179 **b** $f(x) = \begin{cases} \frac{1}{\Gamma(4)24^4} y^{4-1} e^{-\frac{y}{24}} & y > 0 \\ 0 & y \le 0 \end{cases}$

5.163 **a** $k = 105$ **b** $E(X) = 0.375,\ V(X) = 0.026$

5.165 **c** 0.4424 **d** 0.1587

5.167 0.9817 **5.169** \$53.58

5.171 **a** 1.39 **b** 5.99 **c** 0.3678

5.173 **a** 0.0446 **b** 0.8849

5.175 **a** $E(X) = 90.017,\ V(X) = 13{,}923.38$ **b** (0, 326)

5.177 $M(t) = \frac{1}{1-t^2},\ E(X) = 0$

Chapter 6

6.1 **a** Let X and Y be the number of construction contracts obtained by Firms I and II, respectively.

		x = 0	1	2
y	0	1/9	2/9	1/9
	1	2/9	2/9	0
	2	1/9	0	0

b

x	0	1	2
$p(x)$	4/9	4/9	1/9

c 0.50

6.3 **a**

		y = 0	y = 1
x	9	0.343	0.009
	41	0.540	0.014
	82	0.090	0.004

b The marginal distribution of age X:

x	1	2	3
$P(x)$	0.352	0.554	0.093

c The marginal distribution of poverty Y:

y	0	1
$P(y)$	0.973	0.027

6.5 **b** 0.80 **c** 0.12

6.7 **a** $$F(x,y)=\begin{cases}0, & x<0,\ y<0\\ xy, & 0\le x\le 1,\ 0\le y\le 1\\ y, & x\ge 1,\ 0\le y\le 1\\ x, & 0\le x\le 1,\ y\ge 1\\ 1, & x>1,\ y>1\end{cases}$$

b 0.40 **c** 0.40

6.9 **a** 0.25 **b** 0.875 **c** 1 **d** 0.667

6.11 **a** $f(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{otherwise}\end{cases}$ **b** $f(y)=\begin{cases}1, & 0\le y\le 1\\ 0, & \text{otherwise}\end{cases}$

6.13 **a** $f(x)=\begin{cases}2(1-x) & 0<x<1\\ 0 & \text{otherwise}\end{cases}$ **b** $f(y)=\begin{cases}2(1-y) & 0<y<1\\ 0 & \text{otherwise}\end{cases}$

6.15 0.25

6.17 **a**

y	1	0	−1
$P(y)$	0.125	0.25	0.625

6.19 **a** Given that a person is in poverty the conditional distribution of age is below:

y	9	41	82
$p(y)$	0.35	0.52	0.13

b

$x \mid y$	0	1
$p(x \mid y=9)$	0.973	0.027

6.21 **a** $f(x)=\begin{cases}1, & 0\le x\le 1\\ 0, & \text{otherwise}\end{cases}$ **d** 0.20

6.23 0.333 **6.25** 0.156 **6.27** 0.306

6.29 **a** no **b** no **c** no **d** yes

6.31 **a** 1.37 **b** 1.44 **c** −0.6128 **d** −0.4999

6.33 **a** 0.5 **b** 0.5 **c** 0.25 **d** 0

6.35 **a** $E(X)=0.333,\ \ V(x)=0.0556$

b $E(X+Y)=0.667,\ \ V(X+Y)=0.0556$ **c** (0.333, 1)

6.37 **a** −0.028 **b** −0.50

6.39 $E(X)=0.75,\ \ E(Y)=0.375,\ \ E(Z)=0.375,\ \ V(Z)=0.0593$

6.45 **a** 1.5 **b** 0.0833

6.47 **a** 0.125 **b** 0.012

6.49 **a** 0.3333 **b** 0.375 **c** 0.2165

6.51 $E(X) = p\lambda$
$V(X) = p\lambda$

6.55 7 hours

6.57 **a** $p(k) = -\frac{k}{n}\log\left(\frac{k}{n}\right)$ **b** $\frac{1}{e}$

6.59 0.0394

6.61 **a** 0.3675 **b** 0.2262

6.63 0.076

6.65 0.4095

6.67 $E(X_1 + 3X_2) = 2.5$
$V(X_1 + 3X_2) = 4.875$

6.69 **a** Binomial $(n_1 + n_2,\ p)$ **b** $P(X = 0) = 0.914$

6.71 **a** 0.1587 **b** 0.3085

6.73 **a** $\text{Gamma}(\alpha = n+1,\ \beta = \theta)$ **b** $W \sim \text{Exp}\left(\frac{\theta}{p}\right)$

6.75 **a** $c = 0.5$ **b** $f(x) = \begin{cases} \frac{1}{4}x^3, & 0 \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$

c $f(y) = \begin{cases} y, & 0 \le y \le 2 \\ 0, & \text{otherwise} \end{cases}$ **d** 0.0039

e 0.0149

6.77 **a** $f(y\,|\,x) = \begin{cases} 2\frac{y}{x^2}, & 0 \le y \le x \le 2 \\ 0, & \text{otherwise} \end{cases}$ **b** $\frac{2x}{3}$

6.79 **a** and **b**

y \ x	0	1	2	3	$p(x)$
0	0	0.036	0.071	0.012	0.12
1	0.048	0.286	0.142	0	0.476
2	0.143	0.214	0	0	0.357
3	0.048	0	0	0	0.0476
$p(y)$	0.24	0.536	0.214	0.012	

c 0.5625

6.81 **a** $E(X) = 0.75,\quad V(X) = 0.0375$ **b** $E(Y) = 0.667,\quad V(Y) = 0.055$

c $f(x\,|\,y) = \begin{cases} 3x^2 & 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

6.83 **a** $f(x, y) = \begin{cases} \frac{1}{x} & 0 \le y \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$

b 0.5 **c** 0.375

6.85 **a** 0.028 **b** 0.506

6.87 **a** $f(y) = \begin{cases} 2(1-y) & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$ **b** $f(y) = \begin{cases} 1 - |x| & |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$

c $f(y\,|\,x) = \begin{cases} \frac{1}{1-|x|} & |x| \le 1 \\ 0 & \text{otherwise} \end{cases}$ **d** 0.25

6.89 -0.462

6.91 0.25

6.93 **a** $r\theta$ **b** $r\theta(1 + 2\theta)$

6.95
$$\begin{aligned} E(X - Y) &= \$42 \\ V(X - Y) &= \$26 \\ P(G > 70) &\le 0.03 \end{aligned}$$

6.97 **a** 1.067 **b** 1.067

6.99 0.00486

6.103 **a** $\left(p_1 e^{t_1} + p_2 e^{t_2} + p_3 e^{t_3}\right)^n$ **b** $-np_1 p_2$

6.107 $-\sqrt{\frac{p_1 p_2}{(1-p_1)(1-p_2)}}$

6.109 **a** $f(x, y) = \begin{cases} \frac{1}{9} e^{\frac{-(x+y)}{3}} & x > 0,\ y > 0 \\ 0 & \text{otherwise} \end{cases}$ **b** 0.0446

Chapter 7

7.1 **a** $f(x) = \begin{cases} \dfrac{\binom{6}{x}\binom{4}{3-x}}{\binom{10}{3}}, & x = 0, 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$ **b** $X = 3 - Y$

c $f(y) = \begin{cases} \dfrac{\binom{6}{3-y}\binom{4}{y}}{\binom{10}{3}}, & y = 0, 1, 2, 3 \\ 0, & \text{otherwise} \end{cases}$

7.3 **a** $f(x) = \begin{cases} \binom{x+3}{3} (0.4)^4 (0.6)^x, & x = 0, 1, 2, \ldots \\ 0, & \text{otherwise} \end{cases}$ **b** $X = Y - 4$

c $f(y) = \begin{cases} \binom{y-1}{3} (0.4)^4 (0.6)^{y-4}, & y = 4, 5, 6, \ldots \\ 0, & \text{otherwise} \end{cases}$

7.5 **a** $F_{Y_1}(y) = \begin{cases} -2y & -1 \le y \le 0 \\ 0 & \text{otherwise} \end{cases}$ **b** $F_{Y_2}(y) = \begin{cases} 2y & 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$

c $f_{Y_3}(y) = \begin{cases} \frac{2}{\sqrt{y}} - 4 & 0 \le y \le 0.25 \\ 0 & \text{otherwise} \end{cases}$

7.7 **a** $F_Y(y) = \begin{cases} \frac{y+4}{100} & -4 \le y \le 6 \\ \frac{1}{10} & 6 \le y \le 11 \\ 0 & \text{otherwise} \end{cases}$ **b** 5.583

c $E(X) = 0.9583$
$E(Y) = 5.583$

7.9 $f(y) = \begin{cases} \frac{1}{y}, & 1 \le y \le e \\ 0, & \text{otherwise} \end{cases}$

7.11 $Y \sim$ Exponential $(c\theta)$

7.13 $f_Y(y) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - \mu)}{2\sigma^2}}, \quad y > 0$

7.15 **a** $F_Y(y) = P(Y \le y) = P(X_1 - X_2 \le y) = P(X_1 \le y + X_2)$

Case $1 : 0 \le y \le 1$

$f_Y(y) = \frac{y^2}{2}$

Case $2 : 1 < y \le 2$

$$f_y(y) = \begin{cases} y, & 0 \le y \le 1 \\ 2 - y, & 1 < y \le 2 \\ 0, & \text{otherwise} \end{cases}$$

b $E(Y) = 1$

7.17 **a** $f_{Y_1}(y) = -2y, \quad \text{where } -1 \le y \le 0$ **b** $f_{Y_2}(y) = 2y, \quad \text{where } 0 \le y \le 1$

c $f_{Y_3}(y) = \frac{2}{\sqrt{y}} - 4, \quad \text{where } 0 \le y \le 0.25$

7.19 **a** $f_Y(y) = \begin{cases} 1, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

$U \sim \text{Uniform}(0, 1)$

b $f_Z(z) = \begin{cases} e^{-z}, & z \ge 0 \\ 0, & \text{otherwise} \end{cases}$

$Z \sim \text{Exponential}(1)$

7.21 $f_Y(y) = \begin{cases} \frac{1}{8\sqrt{2y-6}}, & 5 \le y \le 53 \\ 0, & \text{otherwise} \end{cases}$

7.23 **a** $f_Y(y) = \begin{cases} \frac{1}{\theta} e^{\frac{-y}{\theta}}, & y \ge 0 \\ 0, & \text{otherwise} \end{cases}$ **b** $E(Y) = \theta, \;\; V(Y) = \theta^2$

7.25 $f_y(y) = \begin{cases} y, & 0 \le y \le 1 \\ 2 - y, & 1 < y \le 2 \\ 0, & \text{otherwise} \end{cases}$

7.27 $f_Y(y) = \begin{cases} 2(1 - y), & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

7.29 **a** $M_Y(t) = \left(pe^t + q\right)^{kn} \to Y \sim \text{Binomial } (kn, p)$

b $M_Z(t) = \left(pe^t + q\right)^{\sum_{i=1}^{k} n_i} \to Z \sim \text{Binomial } \left(\sum_{i=1}^{k} n_i, p\right)$

c $M_W(t) = \prod_{i=1}^{k} \left(p_i e^t + q_i\right)^n \to$ We cannot identify this moment-generating function. It is not that of a binomial random variable

7.31 **a** $M_Y(t) = e^{k\lambda(e^t - 1)} \;\; Y \sim \text{poisson } (k\lambda)$

7.33 $M_{\overline{X}}(t) = e^{xt + \frac{\sigma^2 t^2}{2n}}$

$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

7.35 **a** $E(S^2) = \sigma^2$

b $V(S^2) = E\left(S^2\right)^2 - \left(\sigma^2\right)^2$

$E\left(S^2\right)^2 = \sigma^4 + \frac{2\sigma^4}{n-1}$

$V(S^2) = \frac{2\sigma^4}{n-1}$

c $E(S) = \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \sqrt{\frac{2}{n-1}} \sigma$

7.37 **a** $f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} \frac{1}{6}, & 0 \le y_1 \le 2, \; 0 \le y_2 \le 3 \\ 0, & \text{otherwise} \end{cases}$

b $f_{Y_1}(y_1) = \begin{cases} \frac{1}{2}, & 0 \le y_1 \le 2 \\ 0, & \text{otherwise} \end{cases}$

$Y_1 \sim \text{Uniform}(0, 2)$

c $f_{Y_2}(y_2) = \begin{cases} \frac{1}{3}, & 0 \le y_2 \le 3 \\ 0, & \text{otherwise} \end{cases}$

$Y_2 \sim \text{Uniform}(0, 3)$

d Yes, Y_1 and Y_2 are independent because $f_{Y_1, Y_2}(y_1, y_2) = f_{Y_1}(y_1) f_{Y_2}(y_2)$.

7.39 **a** $f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{1}{\theta^2} e^{-\frac{y_1}{\theta}}, & y_1 > y_2 > 0 \\ 0, & \text{otherwise} \end{cases}$

b $f_{Y_1}(y_1) = \begin{cases} \frac{y_1}{\theta^2} e^{-\frac{y_1}{\theta}}, & y_1 > 0 \\ 0, & \text{otherwise} \end{cases}$

$Y_1 \sim \text{Gamma}(2, \theta)$

7.41 **a** $f_{Y_1,Y_2}(y_1,y_2) = \begin{cases} \frac{1}{2}, & (y_1,y_2) \in A \\ 0, & \text{otherwise} \end{cases}$

b $f(y_1) = \begin{cases} y_1, & 0 < y_1 < 1 \\ 2 - y_1, & 1 \le y_1 < 2 \\ 0, & \text{otherwise} \end{cases}$

c $f(y_2) = \begin{cases} (y_2 + 1), & -1 < y_2 < 0 \\ (1 - y_2), & 0 \le y_2 \le 1 \\ 0, & \text{otherwise} \end{cases}$

d No, for $y_1 = \frac{1}{2}$, $y_2 = \frac{1}{2}$

$$\begin{aligned} f(y_1)f(y_2) &= 4y_1(4)(1 - y_2) \\ &= 16y_1(1 - y_2) \\ &\ne 2 = f(y_1, y_2) \end{aligned}$$

7.43 $f_y(y) = \begin{cases} y, & 0 \le y \le 1 \\ 2 - y, & 1 < y \le 2 \\ 0, & \text{otherwise} \end{cases}$

7.45 $f_Y(y) = \begin{cases} 2y, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

7.47 **a** $F_{X_{(n)}}(x) = \begin{cases} 0, & x < a \\ \left(\frac{x-a}{b-a}\right)^n, & a \le x \le b \\ 1, & x > b \end{cases}$

b $f_{X_{(n)}}(x) = \begin{cases} n\frac{(x-a)^{n-1}}{(b-a)^n}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$

c $E(X_{(n)}) = \frac{nb+a}{n+1}$

7.49 **a** $f_{X_{(1)}}(x) = \begin{cases} ne^{-n(x-\theta)}, & x > \theta \\ 0, & \text{elsewhere} \end{cases}$

b $\frac{e^{n\theta}}{n}$

7.51 **a** $\theta_1 + \frac{1}{n+1}(\theta_2 - \theta_1)$

b $\theta_2 - \frac{1}{n+1}(\theta_2 - \theta_1)$

c $\frac{n+1}{n-1}(x_n - x_1)$

7.53 **a** $X_{(1)} \sim \text{Exp}\left(\frac{\theta}{n}\right)$

b $\frac{\theta}{n}$

c $\frac{\theta^2}{n^2}$

7.59 **a** $f(y) = \binom{n}{n-y} p^{n-y}(1-p)^y, \qquad y = 0, 1, \ldots, n$

b y is the number of failures in n independent Bernoulli trials.

7.61 **a** $f_{Y_1}(y) = \frac{1}{2\sqrt{y}}$ where $0 \le y \le 1$

b $f_{Y_2}(y) = \frac{dF_{Y_2}(y)}{dy} = \begin{cases} \frac{1}{2}, & 0 \le y \le 1 \\ \frac{1}{2y^2}, & 1 < y \\ 0, & \text{otherwise} \end{cases}$

c $f_{Y_3}(y) = \frac{dF_{Y_3}(y)}{dy} = \begin{cases} ye^{-y}, & y > 0 \\ 0, & \text{otherwise} \end{cases}$

d $f_{Y_4}(y) = \frac{dF_{Y_4}(y)}{dy} = \begin{cases} 1 - y\left(\frac{1}{y}\right) - \ln y = \ln y, & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

7.63 **a** $E(Y_1) = 0$
$E(Y_2) = 3$
$E(Y_3) = 0.6$

b $E(Y_1) = 0$
$E(Y_2) = 3$
$E(Y_3) = 0.6$

7.65 **a** $f_Y(y) = \begin{cases} 2(1 - y), & 0 \le y \le 1 \\ 0, & \text{otherwise} \end{cases}$

b $E(Y) = 0.3333$

c $E(Y^2) = 0.1667$
$V(Y) = 0.0556$

7.67 $Y \sim \text{Uniform}(0, 1)$

7.69 $f_Y(y) = \frac{1}{\pi(1+y^2)}$, where $-\infty < y < \infty$

7.71 $[1 - F(x)]^3 [1 + F(x)]$

7.73 **a** $f_{T|v}(t|v) = \sqrt{\frac{v}{2n\pi}}e^{-\frac{t^2u}{2n}}$ **b** $f(t,v) = \dfrac{\frac{1}{2}^{\frac{n+1}{2}} v^{\frac{n+1}{2}-1} e^{-\frac{v}{2}\left(\frac{t^2}{n}+1\right)}}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}$

c $f(t) = \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n}\Gamma\left(\frac{n}{2}\right)\left(\frac{t^2}{n}+1\right)^{\frac{n+1}{2}}}, \quad -\infty < t < \infty$

7.75 $f_u(u) = \begin{cases} ue^{-\frac{u^2}{2}}, & u > 0 \\ 0, & \text{otherwise} \end{cases}$ **7.77** $f_U(u) = \begin{cases} \frac{1}{2\pi}\left(\frac{3u}{4\pi}\right)^{-\frac{1}{3}}, & 0 \le u \le \frac{4}{3}\pi \\ 0, & \text{otherwise} \end{cases}$

7.79 Poisson(θ, λ)

Chapter 8

8.1 $\frac{1}{3}$ **8.3** np **8.5** $\alpha\beta$ **8.7** $F_{X_{(n)}}(x) = \begin{cases} 0, & x < \theta \\ 1, & x \ge \theta \end{cases}$

$F_{X_{(n)}}(x) = \begin{cases} 1, & x = 0 \\ 0, & \text{otherwise} \end{cases}$

8.9 0.264 **8.11** 0.032 **8.13** 0.9090

8.15 **a** 0.5328 **b** 0.9772

8.17 0.7976

8.19 **a** 0.9938 **b** $n = 116$

8.21 0.1587 **8.23** 0.9876 **8.25** 80 **8.27** 0.0287

8.29 0.44

8.31 **a** 0 **b** 0 **c** $d = 0.04$ **8.33** 0.1539

8.35 Normal Approximation: 0.1922
Binomial Exact: 0.193 **8.37** 0.9544

8.39 0.1151 **8.41** 0.327

8.45 0.1587

Chapter 9

9.1 0.5037

9.3 Let P_n be equal to the geometric equilibrium distribution: $P_n = \left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^n$. The geometric equilibrium distribution is a solution to the differential equation defining the equilibrium state: $\lambda P_{n-1} - (\lambda + \theta)P_n + \theta P_{n+1} = 0$. Therefore:

$\lambda P_{n-1} - (\lambda + \theta)p_n + \theta P_{n+1} = \lambda\left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^{n-1} - (\lambda + \theta)\left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^n + \theta\left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^{n+1}$

$= \left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^{n-1}\left\{\lambda - (\lambda + \theta)\left(\frac{\lambda}{\theta}\right) + \theta\left(\frac{\lambda}{\theta}\right)\right\} = \left(1 - \frac{\lambda}{\theta}\right)\left(\frac{\lambda}{\theta}\right)^{n-1}\left\{\lambda - \frac{\lambda^2}{\theta} - \lambda + \frac{\lambda^2}{\theta}\right\} = 0$

9.5 $(X(t) + Y(t)) \sim Poisson\left(\lambda_1 + \lambda_2\right)t$

9.7 $P(T > d) = \left(1 - \frac{d}{t}\right)^2$

9.9 **a** $f_R(r) = \frac{\partial F_R(r)}{\partial r} = \begin{cases} 4\lambda\pi r^2 e^{-\frac{4\pi r^3\lambda}{3}}, & r < 0 \\ 0, & \text{otherwise} \end{cases}$ **b** $U \sim Exp\left(\frac{3}{4\pi\lambda}\right)$

9.11 **a** $T_0 \to \infty$, minimizes $\frac{E(C_t)}{t}$

b $r(T)\int_0^T \overline{F}(u)du - F(T) = \frac{c_2}{c_1 - c_2} > 0.$

$T = 0 \Rightarrow$ the left side of the equation is 0, the left side will eventually cross $\frac{c_2}{c_1 - c_2}$, for a finite T.

This page intentionally left blank

INDEX

'b' indicates boxed material; 'f' indicates a figure; 't' indicates a table