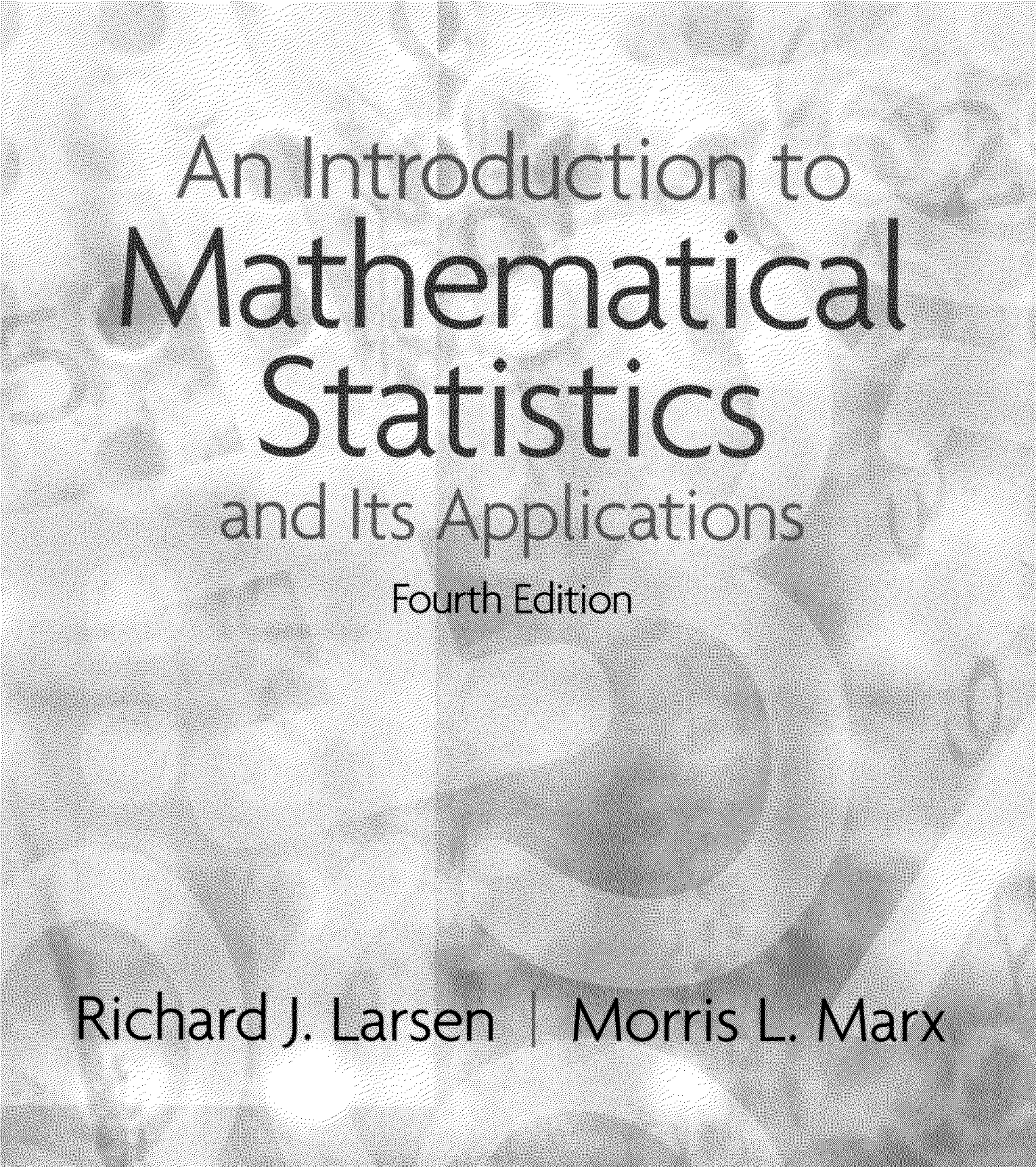


An Introduction to Mathematical Statistics and Its Applications

Fourth Edition

Richard J. Larsen | Morris L. Marx

An Introduction to Mathematical Statistics and Its Applications

Fourth Edition

Richard J. Larsen
Vanderbilt University

Morris L. Marx
University of West Florida

PEARSON
Prentice Hall

Upper Saddle River, New Jersey 07458

Library of Congress Cataloging-in-Publication Data

Larsen, Richard J.
An introduction to mathematical statistics and its applications / Richard J. Larsen, Morris L. Marx,—4th ed.
p. cm.
Includes bibliographical references and index.
ISBN 0-13-186793-8
1. Mathematical statistics. I. Marx, Morris L. II. Title.
CIP data available

Editor-in-Chief/Acquisitions Editor: *Sally Yagan*
Project Manager/Formatter: *Interactive Composition Corporation*
Assistant Managing Editor: *Bayani Mendoza de Leon*
Senior Managing Editor: *Linda Mihatov Behrens*
Executive Managing Editor: *Kathleen Schiaparelli*
Manufacturing Manager: *Alexis Heydt-Long*
Manufacturing Buyer: *Maura Zaldivar*
Marketing Manager: *Halee Dinsey*
Marketing Assistant: *Joon Won Moon*
Director of Creative Services: *Paul Belfanti*
Art Director: *Jayne Conte*
Cover Designer: *Bruce Kenselaar*
Editorial Assistant: *Jennifer Urban*
Cover Image: *Getty Images, Inc.*

Pearson Prentice Hall
Pearson Education, Inc.
Upper Saddle River, NJ 07458

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

ISBN 0-13-186793-8

Pearson Education Ltd., *London*
Pearson Education Australia PTY. Limited, *Sydney*
Pearson Education Singapore, Pte., Ltd
Pearson Education North Asia Ltd, *Hong Kong*
Pearson Education Canada, Ltd., *Toronto*
Pearson Education de Mexico, S.A. de C.V.
Pearson Education - Japan, *Tokyo*
Pearson Education Malaysia, Pte. Ltd

Table of Contents

Preface

We are pleased that our text has been sufficiently well received to justify this fourth edition. Students and instructors who use the text like the coupling of the rigorous and structured treatment of probability and statistics with real-world case studies and examples. The users of the book have been helpful in pointing out ways to improve our presentation. The changes found in this fourth edition reflect the many helpful suggestions we have received, as well as our own experience in teaching from the text.

Our first goal in writing this fourth edition was to continue strengthening the bridge between theory and practice. To that end, we have added sections at the end of each chapter called *Taking a Second Look at Statistics*. These sections discuss practical problems in applying the ideas in the chapter and also deal with common misunderstandings or faulty approaches. We also have included a new section on Bayesian estimation that integrates well into Chapter 5 on estimation and gives another view of how estimation can be applied. It introduces students to Bayesian ideas and also serves to reinforce the main concepts of estimation.

Some ideas that are useful and important lie beyond the mathematical scope of the text. To explore such topics within the mathematical context of the book, we have increased and enhanced the material on simulation and on the use of Monte Carlo studies. Since MINITAB is the main tool for simulations and demonstrating computer computations, the MINITAB sections have been rewritten to conform to Version 14, the latest release.

A barrier to efficient coverage of the book has been the length of time required to cover Chapters 2 and 3. One of the major changes in the fourth edition is a substantial revision of basic probability material. Chapters 2 and 3 have been reorganized and rewritten with the goal of a streamlined presentation. These chapters are now easier to teach and can be covered in less time, yet without loss of rigor.

In that same spirit, we have also improved and streamlined the development of the t, chi square and F distributions in Chapter 7, the heart of the book. The material there has been rewritten to simplify the development of the chi square distribution. In addition, we have made a much better division between the theoretical results and their applications.

Because of the efficiencies in the new edition, covering Chapters 1–7 plus other additional topics in one semester is now possible.

All in all, we feel that this new edition furthers our objective of writing a book that emphasizes the interrelation between probability theory, mathematical statistics, and data analysis. As in previous editions, real-world case studies and historical anecdotes provide valuable tools to effect the integration of these three areas. Our experience in the classroom has strengthened our belief in this approach. Students can better grasp the importance of each area when seen in the context of the other two.

SUPPLEMENTS

Instructor's Solutions Manual. This resource contains worked-out solutions to all text exercises.
Student Solutions Manual. Featuring complete solutions to selected exercises, this is a great tool for students as they study and work through the problem material.

ACKNOWLEDGMENTS

We would like to thank the following reviewers for their detailed and valuable comments, criticisms, and suggestions:

Ditlev Monrad, *University of Illinois at Urbana-Champaign*
Vidhu S. Prasad, *University of Massachusetts, Lowell*
Wen-Qing Xu, *California State University, Long Beach*
Katherine St. Clair, *Colby College*
Yimin Xiao, *Michigan State University*
Nicolas Christou, *University of California, Los Angeles.*
Daming Xu, *University of Oregon*
Maria Rizzo, *Ohio University*
Dimitris Politis, *University of California at San Diego*

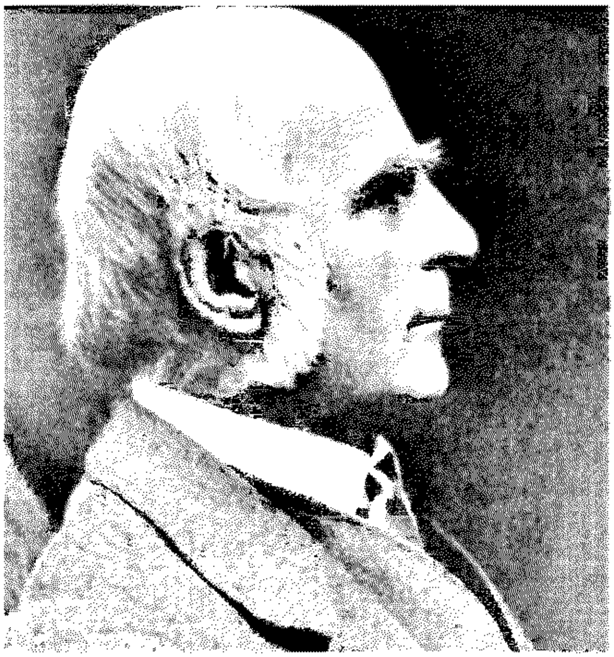
Finally, we convey our gratitude and appreciation to Prentice Hall's math editor-in-chief and acquisitions editor, Sally Yagan, assistant managing editor, Bayani Mendoza de Leon, and editorial assistant, Jennifer Urban, as well as to project manager, Jennifer Crotteau, of Interactive Composition Corporation, for their excellent teamwork in the production of this book.

Richard J. Larsen
Nashville, Tennessee
Morris L. Marx
Pensacola, Florida

CHAPTER 1

Introduction

Francis Galton



> *"Some people hate the very name of statistics, but I find them full of beauty and interest. Whenever they are not brutalized, but delicately handled by the higher methods, and are warily interpreted, their power of dealing with complicated phenomena is extraordinary. They are the only tools by which an opening can be cut through the formidable thicket of difficulties that bars the path of those who pursue the Science of man."*
>
> —Francis Galton

1.1 A BRIEF HISTORY

Statistics is the science of sampling. How one set of measurements differs from another and what the implications of those differences might be are its primary concerns. Conceptually, the subject is rooted in the mathematics of probability, but its applications are everywhere. Statisticians are as likely to be found in a research lab or a field station as they are in a government office, an advertising firm, or a college classroom.

Properly applied, statistical techniques can be enormously effective in clarifying and quantifying natural phenomena. Figure 1.1.1 illustrates a case in point. Pictured at the top is a facsimile of the kind of data routinely recorded by a seismograph—listed chronologically are the occurrence times and Richter magnitudes for a series of earthquakes. Viewed in that format, the numbers are largely meaningless: No patterns are evident, nor is there any obvious connection between the frequencies of tremors and their severities.

By way of contrast, the bottom of Figure 1.1.1 shows a statistical summary (using some of the regression techniques we will learn later) of a set of seismograph data recorded

Episode number	Date	Time	Severity (Richter scale)
⋮	⋮	⋮	⋮
217	6/19	4:53 P.M.	2.7
218	7/2	6:07 A.M.	3.1
219	7/4	8:19 A.M.	2.0
220	8/7	1:10 A.M.	4.1
221	8/7	10:46 P.M.	3.6
⋮	⋮	⋮	⋮

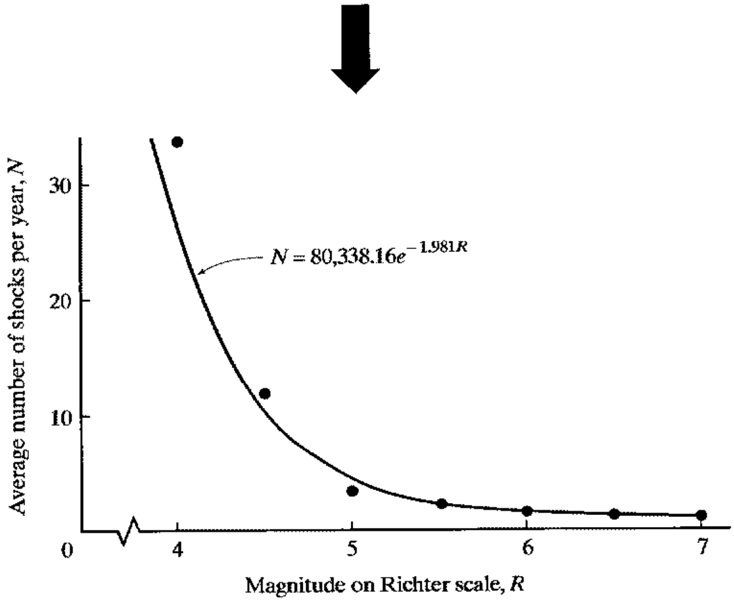


FIGURE 1.1.1

in southern California (66). Plotted above the Richter (R) value of 4.0, for example, is the average number (N) of earthquakes occurring per year in that region having magnitudes in the range 3.75 to 4.25. Similar points are included for R-values centered at 4.5, 5.0, 5.5, 6.0, 6.5, and 7.0. Now we can see that the two variables *are* related: Describing the (N, R)'s exceptionally well is the equation $N = 80{,}338.16e^{-1.981R}$.

In general, statistical techniques are employed either to (1) describe what *did* happen or (2) predict what *might* happen. The graph at the bottom of Figure 1.1.1 does both. Having "fit" the model $N = \beta_0 e^{-\beta_1 R}$ to the observed set of minor tremors (and finding that $\beta_0 = 80{,}338.16$ and $\beta_1 = -1.981$), we can then use that same equation to predict the likelihood of events *not* represented in the data set. If $R = 8.0$, for example, we would expect N to equal 0.01:

$$\begin{aligned} N &= 80{,}338.16e^{-1.981(8.0)} \\ &= 0.01 \end{aligned}$$

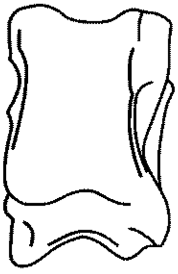
(which implies that Californians can expect catastrophic earthquakes registering on the order of 8.0 on the Richter scale to occur, on the average, once every 100 years).

It is unarguably true that the interplay between description and prediction—similar to what we see in Figure 1.1.1—is the single most important theme in statistics. Additional examples highlighting other aspects of that connection will be discussed in Section 1.2. To set the stage for the rest of the text, though, we will conclude Section 1.1 with brief histories of probability and statistics. Both are interesting stories, replete with large casts of unusual characters and plots that have more than a few unexpected twists and turns.

Probability: The Early Years

No one knows where or when the notion of chance first arose; it fades into our prehistory. Nevertheless, evidence linking early humans with devices for generating random events is plentiful: Archaeological digs, for example, throughout the ancient world consistently turn up a curious overabundance of *astragali,* the heel bones of sheep and other vertebrates. Why should the frequencies of these bones be so disproportionately high? One could hypothesize that our forbears were fanatical foot fetishists, but two other explanations seem more plausible: The bones were used for religious ceremonies *and for gambling.*

Astragali have six sides but are not symmetrical (see Figure 1.1.2). Those found in excavations typically have their sides numbered or engraved. For many ancient



Sheep astragalus

FIGURE 1.1.2

civilizations, astragali were the primary mechanism through which oracles solicited the opinions of their gods. In Asia Minor, for example, it was customary in divination rites to roll, or *cast,* five astragali. Each possible configuration was associated with the name of a god and carried with it the sought-after advice. An outcome of (1, 3, 3, 4, 4), for instance, was said to be the throw of the savior Zeus, and its appearance was taken as a sign of encouragement (36):

> One one, two threes, two fours
> The deed which thou meditatest, go do it boldly.
> Put thy hand to it. The gods have given thee
> favorable omens
> Shrink not from it in thy mind, for no evil
> shall befall thee.

A (4, 4, 4, 6, 6), on the other hand, the throw of the child-eating Cronos, would send everyone scurrying for cover:

> Three fours and two sixes. God speaks as follows.
> Abide in thy house, nor go elsewhere,
> Lest a ravening and destroying beast come nigh thee.
> For I see not that this business is safe. But bide
> thy time.

Gradually, over thousands of years, astragali were replaced by dice, and the latter became the most common means for generating random events. Pottery dice have been found in Egyptian tombs built before 2000 B.C.; by the time the Greek civilization was in full flower, dice were everywhere. (*Loaded* dice have also been found. Mastering the mathematics of probability would prove to be a formidable task for our ancestors, but they quickly learned how to cheat!)

The lack of historical records blurs the distinction initially drawn between divination ceremonies and recreational gaming. Among more recent societies, though, gambling emerged as a distinct entity, and its popularity was irrefutable. The Greeks and Romans were consummate gamblers, as were the early Christians (91).

Rules for many of the Greek and Roman games have been lost, but we can recognize the lineage of certain modern diversions in what was played during the Middle Ages. The most popular dice game of that period was called *hazard,* the name deriving from the Arabic *al zhar,* which means "a die." Hazard is thought to have been brought to Europe by soldiers returning from the Crusades; its rules are much like those of our modern-day craps. Cards were first introduced in the fourteenth century and immediately gave rise to a game known as *Primero,* an early form of poker. Board games, such as backgammon, were also popular during this period.

Given this rich tapestry of games and the obsession with gambling that characterized so much of the Western world, it may seem more than a little puzzling that a formal study of probability was not undertaken sooner than it was. As we will see shortly, the first instance of anyone *conceptualizing* probability, in terms of a mathematical model, occurred in the sixteenth century. That means that more than 2000 years of dice games, card games, and board games passed by before someone finally had the insight to write down even the simplest of probabilistic abstractions.

Historians generally agree that, as a subject, probability got off to a rocky start because of its incompatibility with two of the most dominant forces in the evolution of our Western culture, Greek philosophy and early Christian theology. The Greeks were comfortable with the notion of chance (something the Christians were not), but it went against their nature to suppose that random events could be quantified in any useful fashion. They believed that any attempt to reconcile mathematically what *did* happen with what *should have* happened was, in their phraseology, an improper juxtaposition of the "earthly plane" with the "heavenly plane."

Making matters worse was the antiempiricism that permeated Greek thinking. Knowledge, to them, was not something that should be derived by experimentation. It was better to reason out a question logically than to search for its explanation in a set of numerical observations. Together, these two attitudes had a deadening effect: The Greeks had no motivation to think about probability in any abstract sense, nor were they faced with the problems of interpreting data that might have pointed them in the direction of a probability calculus.

If the prospects for the study of probability were dim under the Greeks, they became even worse when Christianity broadened its sphere of influence. The Greeks and Romans at least accepted the *existence* of chance. They believed their gods to be either unable or unwilling to get involved in matters so mundane as the outcome of the roll of a die. Cicero writes:

> Nothing is so uncertain as a cast of dice, and yet there is no one who plays often who does not make a Venus-throw[1] and occasionally twice and thrice in succession. Then are we, like fools, to prefer to say that it happened by the direction of Venus rather than by chance?

For the early Christians, though, there was no such thing as chance: Every event that happened, no matter how trivial, was perceived to be a direct manifestation of God's deliberate intervention. In the words of St. Augustine:

> Nos eas causas quae dicuntur fortuitae ... non dicimus
> nullas, sed latentes; easque tribuimus vel veri Dei ...
> (We say that those causes that are said to be by chance
> are not non-existent but are hidden, and we attribute
> them to the will of the true God ...)

Taking Augustine's position makes the study of probability moot, and it makes a probabilist a heretic. Not surprisingly, nothing of significance was accomplished in the subject for the next fifteen hundred years.

It was in the sixteenth century that probability, like a mathematical Lazarus, arose from the dead. Orchestrating its resurrection was one of the most eccentric figures in the entire history of mathematics, Gerolamo Cardano. By his own admission, Cardano personified the best and the worst—the Jekyll and the Hyde—of the Renaissance man. He was born in 1501 in Pavia. Facts about his personal life are difficult to verify. He wrote an autobiography, but his penchant for lying raises doubts about much of what he says.

---

[1] When rolling four astragali, each of which is numbered on *four* sides, a Venus-throw was having each of the four numbers appear.

Whether true or not, though, his "one-sentence" self-assessment paints an interesting portrait (133):

> Nature has made me capable in all manual work, it has given me the spirit of a philosopher and ability in the sciences, taste and good manners, voluptuousness, gaiety, it has made me pious, faithful, fond of wisdom, meditative, inventive, courageous, fond of learning and teaching, eager to equal the best, to discover new things and make independent progress, of modest character, a student of medicine, interested in curiosities and discoveries, cunning, crafty, sarcastic, an initiate in the mysterious lore, industrious, diligent, ingenious, living only from day to day, impertinent, contemptuous of religion, grudging, envious, sad, treacherous, magician and sorcerer, miserable, hateful, lascivious, obscene, lying, obsequious, fond of the prattle of old men, changeable, irresolute, indecent, fond of women, quarrelsome, and because of the conflicts between my nature and soul I am not understood even by those with whom I associate most frequently.

Formally trained in medicine, Cardano's interest in probability derived from his addiction to gambling. His love of dice and cards was so all-consuming that he is said to have once sold all his wife's possessions just to get table stakes! Fortunately, something positive came out of Cardano's obsession. He began looking for a mathematical model that would describe, in some abstract way, the outcome of a random event. What he eventually formalized is now called the *classical definition of probability:* If the total number of possible outcomes, all equally likely, associated with some action is n, and if m of those n result in the occurrence of some given event, then the probability of that event is m/n. If a fair die is rolled, there are $n = 6$ possible outcomes. If the event "outcome is greater than or equal to 5" is the one in which we are interested, then $m = 2$ (the outcomes 5 and 6) and the probability of the event is $\frac{2}{6}$, or $\frac{1}{3}$ (see Figure 1.1.3).

Cardano had tapped into the most basic principle in probability. The model he discovered may seem trivial in retrospect, but it represented a giant step forward: His was the first recorded instance of anyone computing a *theoretical,* as opposed to an empirical, probability. Still, the actual impact of Cardano's work was minimal. He wrote a book in 1525, but its publication was delayed until 1663. By then, the focus of the Renaissance, as well as interest in probability, had shifted from Italy to France.

The date cited by many historians (those who are not Cardano supporters) as the "beginning" of probability is 1654. In Paris a well-to-do gambler, the Chevalier de Méré,

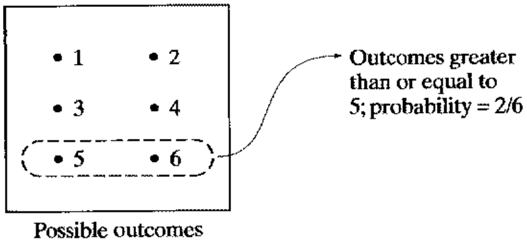


FIGURE 1.1.3

asked several prominent mathematicians, including Blaise Pascal, a series of questions, the best-known of which was the *problem of points:*

> Two people, A and B, agree to play a series of fair games until one person has won six games. They each have wagered the same amount of money, the intention being that the winner will be awarded the entire pot. But suppose, for whatever reason, the series is prematurely terminated, at which point A has won five games and B three. How should the stakes be divided?

[The correct answer is that A should receive seven-eighths of the total amount wagered. (Hint: Suppose the contest were resumed. What scenarios would lead to A's being the first person to win six games?)]

Pascal was intrigued by de Méré's questions and shared his thoughts with Pierre Fermat, a Toulouse civil servant and probably the most brilliant mathematician in Europe. Fermat graciously replied, and from the now famous Pascal-Fermat correspondence came not only the solution to the problem of points but the foundation for more general results. More significantly, news of what Pascal and Fermat were working on spread quickly. Others got involved, of whom the best known was the Dutch scientist and mathematician Christiaan Huygens. The delays and the indifference that plagued Cardano a century earlier were not going to happen again.

Best remembered for his work in optics and astronomy, Huygens, early in his career, was intrigued by the problem of points. In 1657 he published *De Ratiociniis in Aleae Ludo* (Calculations in Games of Chance), a very significant work, far more comprehensive than anything Pascal and Fermat had done. For almost fifty years it was the standard "textbook" in the theory of probability. Not surprisingly, Huygens has supporters who feel that *he* should be credited as the founder of probability.

Almost all the mathematics of probability was still waiting to be discovered. What Huygens wrote was only the humblest of beginnings, a set of fourteen propositions bearing little resemblance to the topics we teach today. But the foundation was there. The mathematics of probability was finally on firm footing.

Statistics: From Aristotle to Quetelet

Historians generally agree that the basic principles of statistical reasoning began to coalesce in the middle of the nineteenth century. What triggered this emergence was the union of three different "sciences," each of which had been developing along more or less independent lines (206).

The first of these sciences, what the Germans called *Staatenkunde,* involved the collection of comparative information on the history, resources, and military prowess of nations. Although efforts in this direction peaked in the seventeenth and eighteenth centuries, the concept was hardly new: Aristotle had done something similar in the fourth century B.C. Of the three movements, this one had the least influence on the development of modern statistics, but it did contribute some terminology: The word *statistics,* itself, first arose in connection with studies of this type.

The second movement, known as *political arithmetic,* was defined by one of its early proponents as "the art of reasoning by figures, upon things relating to government." Of more recent vintage than Staatenkunde, political arithmetic's roots were in seventeenth-century England. Making population estimates and constructing mortality tables were two of the problems it frequently dealt with. In spirit, political arithmetic was similar to what is now called *demography*.

The third component was the development of a *calculus of probability*. As we saw earlier, this was a movement that essentially started in seventeenth-century France in response to certain gambling questions, but it quickly became the "engine" for analyzing all kinds of data.

Staatenkunde: The Comparative Description of States

The need for gathering information on the customs and resources of nations has been obvious since antiquity. Aristotle is credited with the first major effort toward that objective: His *Politeiai,* written in the fourth century B.C., contained detailed descriptions of some 158 different city-states. Unfortunately, the thirst for knowledge that led to the *Politeiai* fell victim to the intellectual drought of the Dark Ages, and almost 2000 years elapsed before any similar projects of like magnitude were undertaken.

The subject resurfaced during the Renaissance, and the Germans showed the most interest. They not only gave it a name, *Staatenkunde,* meaning the comparative description of states, but they were also the first (in 1660) to incorporate the subject into a university curriculum. A leading figure in the German movement was Gottfried Achenwall, who taught at the University of Göttingen during the middle of the eighteenth century. Among Achenwall's claims to fame is that he was the first to use the word *statistics* in print. It appeared in the preface of his 1749 book *Abriss der Statswissenschaft der heutigen vornehmsten europaishen Reiche und Republiken*. (The word comes from the Italian root *stato,* meaning "state," implying that a statistician is someone concerned with government affairs.) As terminology, it seems to have been well-received: For almost one hundred years the word *statistics* continued to be associated with the comparative description of states. In the middle of the nineteenth century, though, the term was redefined, and statistics became the new name for what had previously been called political arithmetic.

How important was the work of Achenwall and his predecessors to the development of statistics? That would be difficult to say. To be sure, their contributions were more indirect than direct. They left no methodology and no general theory. But they did point out the need for collecting accurate data and, perhaps more importantly, reinforced the notion that something complex—even as complex as an entire nation—can be effectively studied by gathering information on its component parts. Thus, they were lending important support to the then growing belief that *induction,* rather than *deduction,* was a more sure-footed path to scientific truth.

Political Arithmetic

In the sixteenth century the English government began to compile records, called *bills of mortality,* on a parish-to-parish basis, showing numbers of deaths and their underlying causes. Their motivation largely stemmed from the plague epidemics that had periodically ravaged Europe in the not-too-distant past and were threatening to become a problem in England. Certain government officials, including the very influential Thomas Cromwell, The bill for the year—A General Bill for this present year, ending the 19 of December, 1665, according to the Report made to the King's most excellent Majesty, by the Co. of Parish Clerks of Lond., & c.—gives the following summary of the results; the details of the several parishes we omit, they being made as in 1625, except that the out-parishes were now 12:—

Buried in the 27 Parishes within the walls	15,207
Whereof of the plague	9,887
Buried in the 16 Parishes without the walls	41,351
Whereof of the plague	28,838
At the Pesthouse, total buried	159
Of the plague	156
Buried in the 12 out-Parishes in Middlesex and surrey	18,554
Whereof of the plague	21,420
Buried in the 5 Parishes in the City and Liberties of Westminster	12,194
Whereof the plague	8,403
The total of all the christenings	9,967
The total of all the burials this year	97,306
Whereof of the plague	68,596

Abortive and Stillborne	617
Aged	1,545
Ague & Feaver	5,257
Appolex and Suddenly	116
Bedrid	10
Blasted	5
Bleeding	16
Cold & Cough	68
Collick & Winde	134
Comsumption & Tissick	4,808
Convulsion & Mother	2,036
Distracted	5
Dropsie & Timpany	1,478
Drowned	50
Executed	21
Flox & Smallpox	655
Found Dead in streets, fields, &c.	20
French Pox	86
Frighted	23
Gout & Sciatica	27
Grief	46
Griping in the Guts	1,288
Hang'd & made away themselved	7
Headmould shot and mould fallen	14
Jaundice	110
Impostume	227
Kill by several accidents	46
King's Evill	86
Leprosie	2
Lethargy	14
Livergrown	20
Bloody Flux, Scowring & Flux	18
Burnt and Scalded	8
Calenture	3
Cancer, Cangrene & Fistula	56
Canker and Thrush	111
Childbed	625
Chrisomes and Infants	1,258
Meagrom and Headach	12
Measles	7
Murthered & Shot	9
Overlaid & Starved	45
Palsie	30
Plague	68,596
Plannet	6
Plurisie	15
Poysoned	1
Quinsie	35
Rickets	535
Rising of the Lights	397
Rupture	34
Scurry	105
Shingles & Swine Pox	2
Sores, Ulcers, Broken and Bruised Limbs	82
Spleen	14
Spotted Feaver & Purples	1,929
Stopping of the Stomach	332
Stone and Stranguary	98
Surfe	1,251
Teeth & Worms	2,614
Vomiting	51
Wenn	8

Christened-Males	5,114	Females	4,853	In all	9,967
Buried-Males	58,569	Females	48,737	In all	97,306

Of the Plague	68,596
Increase in the Burials in the 130 Parishes and the Pesthouse this year	79,009
Increase of the Plague in the 130 Parishes and the Pesthouse this year	68,590

FIGURE 1.1.4

felt that these bills would prove invaluable in helping to control the spread of an epidemic. At first, the bills were published only occasionally, but by the early seventeenth century they had become a weekly institution.[2]

Figure 1.1.4 (155) shows a portion of a bill that appeared in London in 1665. The gravity of the plague epidemic is strikingly apparent when we look at the numbers at the top: Out of 97,306 deaths, 68,596 (over 70%) were caused by the plague. The breakdown of certain other afflictions, though they caused fewer deaths, raises some interesting questions. What

---

[2] An interesting account of the bills of mortality is given in Daniel Defoe's *A Journal of the Plague Year,* which purportedly chronicles the London plague outbreak of 1665.

happened, for example, to the 23 people who were "frighted" or to the 397 who suffered from "rising of the lights"?

Among the faithful subscribers to the bills was John Graunt, a London merchant. Graunt not only read the bills, he studied them intently. He looked for patterns, computed death rates, devised ways of estimating population sizes, and even set up a primitive life table. His results were published in the 1662 treatise *Natural and Political Observations upon the Bills of Mortality*. This work was a landmark: Graunt had launched the twin sciences of vital statistics and demography, and, although the name came later, it also signaled the beginning of political arithmetic. (Graunt did not have to wait long for accolades; in the year his book was published, he was elected to the prestigious Royal Society of London.)

High on the list of innovations that made Graunt's work unique were his objectives. Not content simply to describe a situation, although he was adept at doing so, Graunt often sought to go beyond his data and make generalizations (or, in current statistical terminology, draw *inferences*). Having been blessed with this particular turn of mind, he almost certainly qualifies as the world's first statistician. All Graunt really lacked was the probability theory that would have enabled him to frame his inferences more mathematically. That theory, though, was just beginning to unfold several hundred miles away in France.

Other seventeenth-century writers were quick to follow through on Graunt's ideas. William Petty's *Political Arithmetick* was published in 1690, although it was probably written some fifteen years earlier. (It was Petty who gave the movement its name.) Perhaps even more significant were the contributions of Edmund Halley (of "Halley's comet" fame). Principally an astronomer, he also dabbled in political arithmetic, and in 1693 wrote *An Estimate of the Degrees of the Mortality of Mankind, drawn from Curious Tables of the Births and Funerals at the city of Breslaw; with an attempt to ascertain the Price of Annuities upon Lives*. (Book titles were longer then!) Halley shored up, mathematically, the efforts of Graunt and others to construct an accurate mortality table. In doing so, he laid the foundation for the important theory of annuities. Today, all life insurance companies base their premium schedules on methods similar to Halley's. (The first company to follow his lead was The Equitable, founded in 1765.)

For all its initial flurry of activity, political arithmetic did not fare particularly well in the eighteenth century, at least in terms of having its methodology fine-tuned. Still, the second half of the century did see some notable achievements for improving the quality of the databases: Several countries, including the United States in 1790, established a periodic census. To some extent, answers to the questions that interested Graunt and his followers had to be deferred until the theory of probability could develop just a little bit more.

Quetelet: The Catalyst

With political arithmetic furnishing the data and many of the questions, and the theory of probability holding out the promise of rigorous answers, the birth of statistics was at hand. All that was needed was a catalyst—someone to bring the two together. Several individuals served with distinction in that capacity. Karl Friedrich Gauss, the superb German mathematician and astronomer, was especially helpful in showing how statistical concepts could be useful in the physical sciences. Similar efforts in France were made by Laplace. But the man who perhaps best deserves the title of "matchmaker" was a Belgian, Adolphe Quetelet.

Quetelet was a mathematician, astronomer, physicist, sociologist, anthropologist, and poet. One of his passions was collecting data, and he was fascinated by the regularity of social phenomena. In commenting on the nature of criminal tendencies, he once wrote (69):

> Thus we pass from one year to another with the sad perspective of seeing the same crimes reproduced in the same order and calling down the same punishments in the same proportions. Sad condition of humanity! ... We might enumerate in advance how many individuals will stain their hands in the blood of their fellows, how many will be forgers, how many will be poisoners, almost we can enumerate in advance the births and deaths that should occur. There is a budget which we pay with a frightful regularity; it is that of prisons, chains and the scaffold.

Given such an orientation, it was not surprising that Quetelet would see in probability theory an elegant means for expressing human behavior. For much of the nineteenth century he vigorously championed the cause of statistics, and as a member of more than one hundred learned societies his influence was enormous. When he died in 1874, statistics had been brought to the brink of its modern era.

1.2 SOME EXAMPLES

Do stock markets rise and fall randomly? Is there a common element in the aesthetic standards of the ancient Greeks and the Shoshoni Indians? Can external forces, such as phases of the moon, affect admissions to mental hospitals? What kind of relationship exists between exposure to radiation and cancer mortality?

These questions are quite diverse in content, but they share some important similarities. They are all difficult or impossible to study in a laboratory, and none are likely to yield to deductive reasoning. Indeed, these are precisely the sorts of questions that are usually answered by collecting data, making assumptions about the conditions that generated the data, and then drawing inferences about those assumptions.

CASE STUDY 1.2.1

Each evening, radio and TV reporters offer a bewildering array of averages and indices that presumably indicate the state of the stock market. But do they? Are these numbers conveying any really useful information? Some financial analysts would say "No," arguing that speculative markets tend to rise and fall randomly, much as though some hidden roulette wheel were spinning out the figures. How might that "theory" be tested *statistically?*

We would begin by constructing a *model* that should describe the behavior of the market *if the (random) hypothesis were true.* To that end, the notion of "random movement" would be translated into two assumptions:

a. The chances of the market's rising or falling on a given day are unaffected by its actions on any previous days.

b. The market is equally likely to go up or down.

(Continued on next page)

(Case Study 1.2.1 continued)

Measuring the day-to-day randomness, or its absence, in the market's movements can be accomplished by looking at the lengths of *runs*. By definition, a *run of downturns of length k* is a sequence of days starting with a rise, followed by k consecutive declines, then followed by a rise. So, for example, a daily sequence of the form (rise, fall, fall, rise) is a run of length two.

If the actual behavior of the market's run lengths differs markedly from the predictions of assumptions (a) and (b), the random-movement hypothesis can be rejected. Fortunately, calculating the "expected" number of (randomly-generated) runs is straightforward.

Suppose a rise has occurred followed by a fall. For a run of length one, the market must next rise. By assumptions (a) and (b), this happens half the time, so a probability of $\frac{1}{2}$ would be assigned to a run of length one. The notation for this will be $P(1) = \frac{1}{2}$. The other half of the time the market falls, giving the sequence (rise, fall, fall). A run of length two occurs if there is now a rise. Again, this happens half the time, making its probability half of the half represented by the (rise, fall, fall) sequence. Thus, the probability of a run of length two is $P(2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. Continuing in this manner, it follows that a run of length k has probability $\left(\frac{1}{2}\right)^k$. Furthermore, if there are T total runs, it seems reasonable to expect $T \cdot \left(\frac{1}{2}\right)^k$ of them to be of length k.

Table 1.2.1 gives the distribution of 120 runs of downturns observed in daily closing prices of the Standard and Poor's 500 stock index between February 17, 1994 and February 9, 1996. The third column gives the corresponding expected numbers, as calculated from the expression $T \cdot \left(\frac{1}{2}\right)^k$, where $T = 120$.

Notice that the agreement between actual and predicted run frequencies seems good enough to lend at least some credence to assumptions (a) and (b). However, the expected numbers of longer runs (4, 5, and 6+) do not fit the distribution particularly well. The reason for that might be that the "equally likely" provision in assumption (b) is too restrictive and should be replaced by the probability p as given in assumption (c):

c. The likelihood of a fall in the market is some number p, where $0 \le p \le 1$.

TABLE 1.2.1: Runs in the Closing Prices for the S&P 500 Stock Index

Run Length, k	Observed	Expected
1	67	60.00
2	28	30.00
3	18	15.00
4	3	7.50
5	2	3.875
6+	2	3.75
	120	120.0

Invoking assumptions (a) and (c), then, allows for the run length probabilities to be recalculated. For example, following a (rise, fall) sequence, a rise would be expected

(Continued on next page)

$100(1 - p)\%$ of the time, so $P(1) = 1 - p$. Another fall, of course, would occur the remaining $p\%$ of the time. Since the chance of the next change being a rise is $1 - p$, the probability of the sequence (rise, fall, fall, rise)—that is, a run of length two—is $P(2) = p(1 - p)$. In general, $P(k) = p^{k-1}(1 - p)$.

Two questions now arise. Whichever of the two is more important for further study depends on the needs and interests of the model maker.

1. Is the initial assumption $p = \frac{1}{2}$ justified?
2. Given the observed data, what is the best choice (or *estimate*) for p?

To answer Question 1, we must decide whether the discrepancies between observed and expected run lengths are small enough to be attributed to chance or large enough to render the model invalid. One way to answer Question 2 is to seek the value of p that best "explains" the observations, in terms of maximizing their likelihood of occurring. For the data from which Table 1.2.1 was derived, this type of estimate turns out to be $p = 0.43$. The corresponding expected values, based on $P(k) = p^{k-1}(1 - p) = (0.43)^{k-1}(0.57)$, are given in column 3 of Table 1.2.2.

TABLE 1.2.2: Runs in the Closing Prices for the S&P 500 Stock Index

Run Length, k	Observed	Expected $[p = 0.43]$
1	67	68.4
2	28	29.4
3	18	12.6
4	3	5.4
5	2	2.3
6+	2	1.9
	120	120.0

Has assumption (c) provided a noticeably better fit? Yes. For five of the six run-length categories, the expected frequencies in Table 1.2.2 are closer to the corresponding observed frequencies than was true for their counterparts in Table 1.2.1. Moreover, both models—$p = \frac{1}{2}$ and $0 < p < 1$—are in substantial agreement with the hypothesis that up-and-down movements in the market look very much like a random sequence.

CASE STUDY 1.2.2

Not all rectangles are created equal. Since antiquity, societies have expressed aesthetic preferences for rectangles having certain width (w) to length (l) ratios. Plato, for example, wrote that rectangles whose sides were in a $1{:}\sqrt{3}$ ratio were especially pleasing. (These are the rectangles formed from the two halves of an equilateral triangle.)

(Continued on next page)

(Case Study 1.2.2 continued)

Another "standard" calls for the width-to-length ratio to be equal to the ratio of the length to the sum of the width and the length. That is,

$$\frac{w}{l} = \frac{l}{w + l} \tag{1.2.1}$$

Equation 1.2.1 implies that the width is $\frac{1}{2}(\sqrt{5} - 1)$, or approximately 0.618, times as long as the length. The Greeks called this the golden rectangle and used it often in their architecture (see Figure 1.2.1). Many other cultures were similarly inclined. The Egyptians, for example, built their pyramids out of stones whose faces were golden rectangles. Today, in our society, the golden rectangle remains an architectural and artistic standard, and even items such as drivers' licenses, business cards, and picture frames often have w/l ratios close to 0.618.

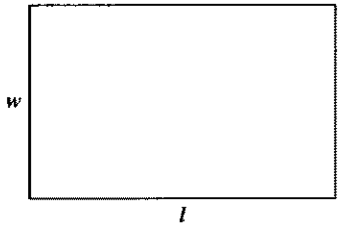


FIGURE 1.2.1: A golden rectangle $\left(\frac{w}{l} = \frac{l}{w + l}\right)$

The fact that many societies have embraced the golden rectangle as an aesthetic standard has two possible explanations. One, they "learned" to like it because of the profound influence that Greek writers, philosophers, and artists have had on cultures all over the world. Or two, there is something unique about human perception that predisposes a preference for the golden rectangle.

Researchers in the field of experimental aesthetics have tried to test the plausibility of those two hypotheses by seeing whether the golden rectangle is accorded any special status by societies that had no contact whatsoever with the Greeks or with their legacy. One such study (39) examined the w/l ratios of beaded rectangles sewn by the Shoshoni Indians as decorations on their blankets and clothes. Table 1.2.3 lists the ratios found for twenty such rectangles.

If, indeed, the Shoshonis also had a preference for golden rectangles, we would expect their ratios to be "close" to 0.618. The average value of the entries in Table 1.2.3, though, is *0.661*. What does that imply? Is 0.661 close enough to 0.618 to support the position that liking the golden rectangle is a human characteristic, or is 0.661 so far from 0.618 that the only prudent conclusion is that the Shoshonis did *not* agree with the aesthetics espoused by the Greeks?

(Continued on next page)

TABLE 1.2.3: Width-To-Length Ratios of Shoshoni Rectangles

0.693	0.749	0.654	0.670
0.662	0.672	0.615	0.606
0.690	0.628	0.668	0.611
0.606	0.609	0.601	0.553
0.570	0.844	0.576	0.933

Making that judgment is an example of *hypothesis testing,* one of the predominant formats used in statistical inference. Mathematically, hypothesis testing is based on a variety of probability results covered in Chapters 2 through 5. The Shoshonis and their rectangles, then, will have to be put on hold until Chapter 6, where we learn how to interpret the difference between a *sample mean* (= 0.661) and a *hypothesized mean* (= 0.618).

Comment. Like π and e, the ratio w/l for golden rectangles (more commonly referred to as either *phi* or the *golden ratio*), is a transcendental number with all sorts of fascinating properties and connections. Indeed, entire books have been written on phi—see, for example (106).

Algebraically, the solution of the equation

$$\frac{w}{l} = \frac{l}{w + l}$$

is the continued fraction

$$\frac{w}{l} = 1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cfrac{1}{1 + \cdots}}}}$$

Among the curiosities associated with phi is its relationship with the *Fibonacci series.* The latter, of course, is the famous sequence where each term is the sum of its two predecessors—that is,

$$1 \quad 1 \quad 2 \quad 3 \quad 5 \quad 8 \quad 13 \quad 21 \quad 34 \quad 55 \quad 89 \quad \ldots$$

Quotients of successive terms in the Fibonacci sequence alternate above and below phi *and they converge to phi:*

$$\begin{aligned}
1/1 &= 1.000000\\
2/1 &= 2.000000\\
3/2 &= 1.500000\\
5/3 &= 1.666666\\
8/5 &= 1.600000\\
13/8 &= 1.625000\\
21/13 &= 1.615385\\
34/21 &= 1.619048\\
55/34 &= 1.617647\\
89/55 &= 1.618182\\
&\vdots
\end{aligned}$$

But phi is not just about numbers—it has cosmological significance as well. Figure 1.2.2 shows a golden rectangle (of width w and length l), where a $w \times w$ square has been inscribed in its left-hand-side. What remains is a golden rectangle on the right, inscribed in which is an $l - w \times l - w$ square. Below that is another golden rectangle with a $w - (l - w) \times w - (l - w)$ square inscribed on its right-hand-side. Each such square leaves another golden rectangle, which can be inscribed with yet another square, and so on ad infinitum. Connecting the points where the squares touch the golden rectangles yields a *logarithmic spiral,* the beginning of which is pictured. These curves are quite common in nature and describe, for example, the shape of spiral galaxies, one of which being our own Milky Way (see Figure 1.2.3).

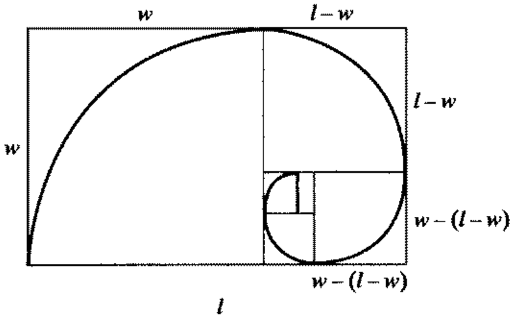


FIGURE 1.2.2

What does all this have to do with the Shoshonis? Absolutely nothing, but mathematical relationships like these are just too good to pass up! The famous astronomer Joannes Kepler once wrote (106):

> "Geometry has two great treasures; one is the Theorem of Pythagoras; the other [is the golden ratio]. The first we may compare to a measure of gold; the second we may name a precious jewel."

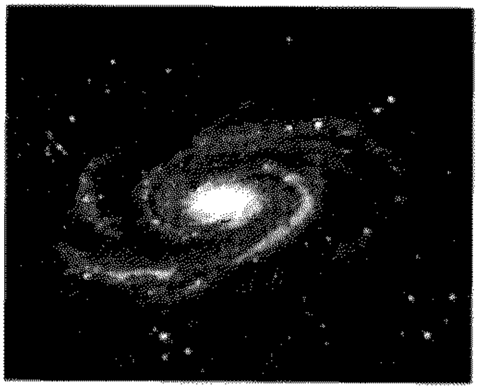


FIGURE 1.2.3

CASE STUDY 1.2.3

In folklore, the full moon is often portrayed as something sinister, a kind of evil force possessing the power to control our behavior. Over the centuries, many prominent writers and philosophers have shared this belief (132). Milton, in *Paradise Lost*, refers to

> Demoniac frenzy, moping melancholy
> And moon-struck madness.

And Othello, after the murder of Desdemona, laments:

> It is the very error of the moon,
> She comes more near the earth than she was wont
> And makes men mad.

On a more scholarly level, Sir William Blackstone, the renowned eighteenth-century English barrister, defined a "lunatic" as

> one who hath...lost the use of his reason and who hath lucid intervals, sometimes enjoying his senses and sometimes not, and that frequently depending upon changes of the moon.

The possibility of lunar phases influencing human affairs is a theory not without supporters among the scientific community. Studies by reputable medical researchers have attempted to link the "Transylvania effect," as it has come to be known, with higher suicide rates, pyromania, and even epilepsy.

The relationship between lunar cycles and mental breakdowns has also been studied. Table 1.2.4 shows the admission rates to the emergency room of a Virginia mental health clinic *before*, *during*, and *after* the twelve full moons from August 1971 to July 1972 (13).

(Continued on next page)

(Case Study 1.2.3 continued)

TABLE 1.2.4: Admission Rates (Patients/Day)

Month	Before Full Moon	During Full Moon	After Full Moon
Aug.	6.4	5.0	5.8
Sept.	7.1	13.0	9.2
Oct.	6.5	14.0	7.9
Nov.	8.6	12.0	7.7
Dec.	8.1	6.0	11.0
Jan.	10.4	9.0	12.9
Feb.	11.5	13.0	13.5
Mar.	13.8	16.0	13.1
Apr.	15.4	25.0	15.8
May	15.7	13.0	13.3
June	11.7	14.0	12.8
July	15.8	20.0	14.5
Averages	10.9	13.3	11.5

For these data, the average admission rate "during" the full moon *is* higher than the "before" and "after" admission rates: 13.3 versus 10.9 and 11.5. Does that imply that the Transylvania effect is real? Not necessarily. The question that needs to be addressed is whether sample means as different as 13.3, 10.9, and 11.5 could reasonably have occurred *by chance* if, in fact, the Transylvania effect does not exist. We will learn in Chapter 13 that the answer to that question appears to be "no."

CASE STUDY 1.2.4

The oil embargo of 1973 raised some very serious questions about energy policies in the United States. One of the most controversial is whether nuclear reactors should assume a more central role in the production of electric power. Those in favor point to their efficiency and to the availability of nuclear material; those against warn of nuclear "incidents" and emphasize the health hazards posed by low-level radiation. Illustrating the opponents' position was a serious safety lapse that occurred some years ago at a government facility located in Hanford, Washington. What happened there is what environmentalists fear will be a recurring problem if nuclear reactors are proliferated.

Until recently, Hanford was responsible for producing the plutonium used in nuclear weapons. One of the major safety problems encountered there was the storage of radioactive wastes. Over the years, significant quantities of strontium 90 and cesium 137 leaked from their open-pit storage areas into the nearby Columbia River, which

(Continued on next page)

flows along the Washington-Oregon border and eventually empties into the Pacific Ocean. The question raised by public health officials was whether exposure to that contamination contributed to any serious medical problems. And if so, to what extent?

As a starting point, an index of exposure was calculated for each of the nine Oregon counties having frontage on either the Columbia River or the Pacific Ocean. It was based on several factors, including the county's stream distance from Hanford and the average distance of its population from any water frontage. As a covariate, the cancer mortality rate was determined for each of the same counties (see Table 1.2.5) (42).

TABLE 1.2.5: Radioactive Contamination and Cancer Mortality in Oregon

County	Index of Exposure	Cancer Mortality per 100,000
Umatilla	2.49	147.1
Morrow	2.57	130.1
Gilliam	3.41	129.9
Sherman	1.25	113.5
Wasco	1.62	137.5
Hood River	3.83	162.3
Portland	11.64	207.5
Columbia	6.41	177.9
Clatsop	8.34	210.3

A graph of the data (see Figure 1.2.4) suggests that radiation exposure (x) and cancer mortality (y) *are* related and that the two vary *linearly*—that is $y = \beta_0 + \beta_1 x$. Finding the numerical values of β_0 and β_1 that orient the line in such a way that it "best" fits the data is a frequently-encountered problem in an area of statistics known as *regression analysis*. Here, the optimal line, based on methods described in Chapter 11, has the equation $y = 114.72 + 9.23x$.

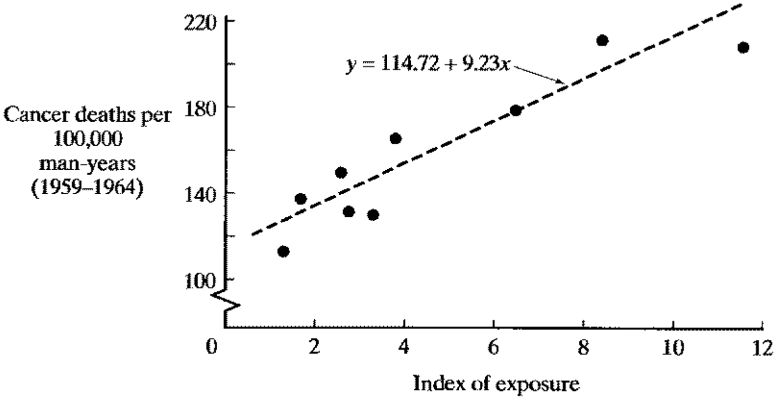


FIGURE 1.2.4

1.3 A CHAPTER SUMMARY

The concepts of probability lie at the very heart of all statistical problems, the case studies of Section 1.2 being typical examples. Acknowledging that fact, the next two chapters take a close look at some of those concepts. Chapter 2 states the axioms of probability and investigates their consequences. It also covers the basic skills for algebraically manipulating probabilities and gives an introduction to combinatorics, the mathematics of counting. Chapter 3 reformulates much of the material in Chapter 2 in terms of *random variables*, the latter being a concept of great convenience in applying probability to statistics. Over the years, particular measures of probability have emerged as being especially useful: The most prominent of these are profiled in Chapter 4.

Our study of statistics proper begins with Chapter 5, which is a first look at the theory of parameter estimation. Chapter 6 introduces the notion of hypothesis testing, a procedure that, in one form or another, commands a major share of the remainder of the book. From a conceptual standpoint, these are very important chapters: Most formal applications of statistical methodology will involve either parameter estimation or hypothesis testing, or both.

Among the probability functions featured in Chapter 4, the *normal distribution*—more familiarly known as the bell-shaped curve—is sufficiently important to merit even further scrutiny. Chapter 7 derives in some detail many of the properties and applications of the normal distribution as well as those of several related probability functions. Much of the theory that supports the methodology appearing in Chapters 9 through 13 comes from Chapter 7.

Chapter 8 describes some of the basic principles of experimental "design." Its purpose is to provide a framework for comparing and contrasting the various statistical procedures profiled in Chapters 9 through 14.

Chapters 9, 12, and 13 continue the work of Chapter 7, but with the emphasis being on the comparison of several populations, similar to what was done in Case Study 1.2.3. Chapter 10 looks at the important problem of assessing the level of agreement between a set of data and the values predicted by the probability model from which those data presumably came (recall Case Study 1.2.1). Linear relationships, such as the one between radiation exposure and cancer mortality in Case Study 1.2.4, are examined in Chapter 11.

Chapter 14 is an introduction to nonparametric statistics. The objective there is to develop procedures for answering some of the same sorts of questions raised in Chapters 8, 9, 11, and 12, but with fewer initial assumptions.

As a general format, each chapter contains numerous examples and case studies, the latter being actual experimental data taken from a variety of sources, primarily newspapers, magazines, and technical journals. We hope that these applications will make it abundantly clear that, while the general orientation of this text is theoretical, the consequences of that theory are never too far from having direct relevance to the "real world."

CHAPTER 2

Probability

Pierre de Fermat

Blaise Pascal

One of the most influential of seventeenth-century mathematicians, Fermat earned his living as a lawyer and administrator in Toulouse. He shares credit with Descartes for the invention of analytic geometry, but his most important work may have been in number theory. Fermat did not write for publication, preferring instead to send letters and papers to friends. His correspondence with Pascal was the starting point for the development of a mathematical theory of probability. —Pierre de Fermat (1601–1665)

Pascal was the son of a nobleman. A prodigy of sorts, he had already published a treatise on conic sections by the age of sixteen. He also invented one of the early calculating machines to help his father with accounting work. Pascal's contributions to probability were stimulated by his correspondence, in 1654, with Fermat. Later that year he retired to a life of religious meditation. —Blaise Pascal (1623–1662)

2.1 INTRODUCTION

Experts have estimated that the likelihood of any given UFO sighting being genuine is on the order of one in one hundred thousand. Since the early 1950s, some ten thousand sightings have been reported to civil authorities. What is the probability that at least one of those objects was, in fact, an alien spacecraft? In 1978, Pete Rose of the Cincinnati Reds set a National League record by batting safely in forty-four consecutive games. How unlikely was that event, given that Rose was a lifetime .303 hitter? By definition, the *mean free path* is the average distance a molecule in a gas travels before colliding with another molecule. How likely is it that the distance a molecule travels between collisions will be at least twice its mean free path? Suppose a boy's mother and father both have genetic markers for sickle cell anemia, but neither parent exhibits any of the disease's symptoms. What are the chances that their son will also be asymptomatic? What are the odds that a poker player is dealt a full house or that a craps shooter makes his "point"? If a woman has lived to age seventy, how likely is it that she will die before her ninetieth birthday? In 1994, Tom Foley was Speaker of the House and running for re-election. The day after the election, his race had still not been "called" by any of the networks: he trailed his Republican challenger by 2174 votes, but 14,000 absentee ballots remained to be counted. Foley, however, conceded. Should he have waited for the absentee ballots to be counted, or was his defeat at that point a virtual certainty?

As the nature and variety of those questions would suggest, probability is a subject with an extraordinary range of real-world, everyday applications. What began as an exercise in understanding games of chance has proven to be useful everywhere. Maybe even more remarkable is the fact that the solutions to all of these diverse questions are rooted in just a handful of definitions and theorems. Those results, together with the problem-solving techniques they empower, are the sum and substance of Chapter 2. We begin, though, with a bit of history.

The Evolution of the Definition of Probability

Over the years, the definition of probability has undergone several revisions. There is nothing contradictory in the multiple definitions—the changes primarily reflected the need for greater generality and more mathematical rigor. The first formulation (often referred to as the *classical* definition of probability) is credited to Gerolamo Cardano (recall Section 1.1). It applies only to situations where (1) the number of possible outcomes is finite and (2) all outcomes are equally-likely. Under those conditions, the probability of an event comprised of m outcomes is the ratio m/n, where n is the total number of (equally-likely) outcomes. Tossing a fair, six-sided die, for example, gives $m/n = \frac{3}{6}$ as the probability of rolling an even number (that is, either 2, 4, or 6).

While Cardano's model was well-suited to gambling scenarios (for which it was intended), it was obviously inadequate for more general problems, where outcomes were not equally likely and/or the number of outcomes was not finite. Richard von Mises, a twentieth-century German mathematician, is often credited with avoiding the weaknesses in Cardano's model by defining "empirical" probabilities. In the von Mises approach, we

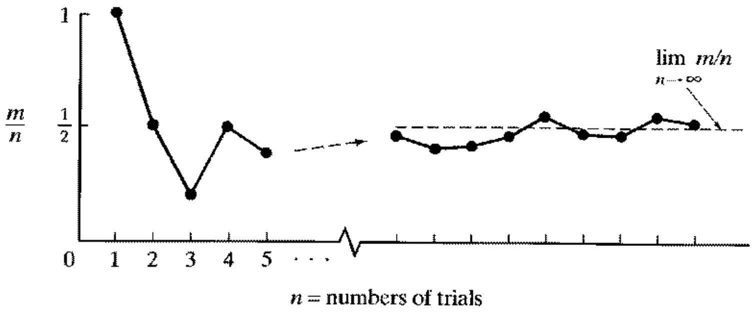


FIGURE 2.1.1

imagine an experiment being repeated over and over again *under presumably identical conditions*. Theoretically, a running tally could be kept of the number of times (m) the outcome belonged to a given event divided by n, the total number of times the experiment was performed. According to von Mises, the probability of the given event is the limit (as n goes to infinity) of the ratio m/n. Figure 2.1.1 illustrates the empirical probability of getting a head by tossing a fair coin: as the number of tosses continues to increase, the ratio m/n converges to $\frac{1}{2}$.

The von Mises approach definitely shores up some of the inadequacies seen in the Cardano model, but it is not without shortcomings of its own. There is some conceptual inconsistency, for example, in extolling the limit of m/n as a way of defining a probability *empirically*, when the very act of repeating an experiment under identical conditions an infinite number of times is physically impossible. And left unanswered is the question of how large n must be in order for m/n to be a good approximation for $\lim m/n$.

Andrei Kolmogorov, the great Russian probabilist, took a different approach. Aware that many twentieth-century mathematicians were having success developing subjects axiomatically, Kolmogorov wondered whether probability might similarly be defined operationally, rather than as a ratio (like the Cardano model) or as a limit (like the von Mises model). His efforts culminated in a masterpiece of mathematical elegance when he published *Grundbegriffe der Wahrscheinlichkeitsrechnung (Foundations of the Theory of Probability)* in 1933. In essence, Kolmogorov was able to show that a maximum of four simple axioms was necessary and sufficient to define the way any and all probabilities must behave. (These will be our starting point in Section 2.3.)

We begin Chapter 2 with some basic (and, presumably, familiar) definitions from set theory. These are important because probability will eventually be defined as a *set function*—that is, a mapping from a set to a number. Then, with the help of Kolmogorov's axioms in Section 2.3, we will learn how to calculate and manipulate probabilities. The chapter concludes with an introduction to *combinatorics*—the mathematics of systematic counting—and its application to probability.

2.2 SAMPLE SPACES AND THE ALGEBRA OF SETS

The starting point for studying probability is the definition of four key terms: *experiment, sample outcome, sample space,* and *event.* The latter three, all carryovers from classical set theory, give us a familiar mathematical framework within which to work; the former is what provides the conceptual mechanism for casting real-world phenomena into probabilistic terms.

By an *experiment* we will mean any procedure that (1) can be repeated, theoretically, an infinite number of times; and (2) has a well-defined set of possible outcomes. Thus, rolling a pair of dice qualifies as an experiment; so does measuring a hypertensive's blood pressure or doing a spectrographic analysis to determine the carbon content of moon rocks. Asking a would-be psychic to draw a picture of an image presumably transmitted by another would-be psychic does *not* qualify as an experiment, because the set of possible outcomes cannot be listed, characterized, or otherwise defined.

Each of the potential eventualities of an experiment is referred to as a *sample outcome,* s, and their totality is called the *sample space,* S. To signify the membership of s in S, we write $s \in S$. Any designated collection of sample outcomes, including individual outcomes, the entire sample space, and the null set, constitutes an *event.* The latter is said to *occur* if the outcome of the experiment is one of the members of the event.

EXAMPLE 2.2.1

Consider the experiment of flipping a coin three times. What is the sample space? Which sample outcomes make up the event A: Majority of coins show heads?

Think of each sample outcome here as an ordered triple, its components representing the outcomes of the first, second, and third tosses, respectively. Altogether, there are eight different triples, so those eight comprise the sample space:

$$S = \{\text{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}\}$$

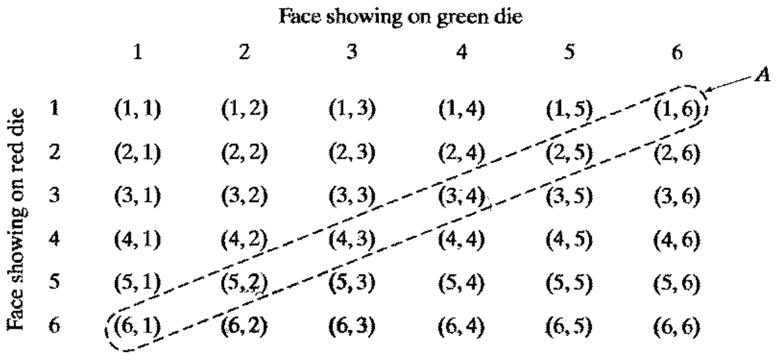
By inspection, we see that four of the sample outcomes in S constitute the event A:

$$A = \{\text{HHH, HHT, HTH, THH}\}$$

EXAMPLE 2.2.2

Imagine rolling two dice, the first one red, the second one green. Each sample outcome is an ordered pair (face showing on red die, face showing on green die), and the entire sample space can be represented as a 6×6 matrix (see Figure 2.2.1).

Gamblers are often interested in the event A that the sum of the faces showing is a 7. Notice in Figure 2.2.1 that the sample outcomes contained in A are the six diagonal entries, $(1, 6)$, $(2, 5)$, $(3, 4)$, $(4, 3)$, $(5, 2)$, and $(6, 1)$.



	1	2	3	4	5	6
1	(1, 1)	(1, 2)	(1, 3)	(1, 4)	(1, 5)	(1, 6)
2	(2, 1)	(2, 2)	(2, 3)	(2, 4)	(2, 5)	(2, 6)
3	(3, 1)	(3, 2)	(3, 3)	(3, 4)	(3, 5)	(3, 6)
4	(4, 1)	(4, 2)	(4, 3)	(4, 4)	(4, 5)	(4, 6)
5	(5, 1)	(5, 2)	(5, 3)	(5, 4)	(5, 5)	(5, 6)
6	(6, 1)	(6, 2)	(6, 3)	(6, 4)	(6, 5)	(6, 6)

FIGURE 2.2.1

EXAMPLE 2.2.3

A local TV station advertises two newscasting positions. If three women (W_1, W_2, W_3) and two men (M_1, M_2) apply, the "experiment" of hiring two coanchors generates a sample space of 10 outcomes:

$$S = \{(W_1, W_2), (W_1, W_3), (W_2, W_3), (W_1, M_1), (W_1, M_2), (W_2, M_1), (W_2, M_2), (W_3, M_1), (W_3, M_2), (M_1, M_2)\}$$

Does it matter here that the two positions being filled are equivalent? Yes. If the station were seeking to hire, say, a sports announcer and a weather forecaster, the number of possible outcomes would be 20: (W_2, M_1), for example, would represent a different staffing assignment than (M_1, W_2).

EXAMPLE 2.2.4

The number of sample outcomes associated with an experiment need not be finite. Suppose that a coin is tossed until the first tail appears. If the first toss is itself a tail, the outcome of the experiment is T; if the first tail occurs on the second toss, the outcome is HT; and so on. Theoretically, of course, the first tail may *never* occur, and the infinite nature of S is readily apparent:

$$S = \{\text{T, HT, HHT, HHHT}, \ldots\}$$

EXAMPLE 2.2.5

There are three ways to indicate an experiment's sample space. If the number of possible outcomes is small, we can simply list them, as we did in Examples 2.2.1 through 2.2.3. In some cases it may be possible to *characterize* a sample space by showing the structure its outcomes necessarily possess. This is what we did in Example 2.2.4. A third option is to state a mathematical formula that the sample outcomes must satisfy.

A computer programmer is running a subroutine that solves a general quadratic equation, $ax^2 + bx + c = 0$. Her "experiment" consists of choosing values for the three coefficients a, b, and c. Define (1) S and (2) the event A: Equation has two equal roots.

First, we must determine the sample space. Since presumably no combinations of finite a, b, and c are inadmissible, we can characterize S by writing a series of inequalities:

$$S = \{(a, b, c): -\infty < a < \infty, -\infty < b < \infty, -\infty < c < \infty\}$$

Defining A requires the well-known result from algebra that a quadratic equation has equal roots if and only if its discriminant, $b^2 - 4ac$, vanishes. Membership in A, then, is contingent on a, b, and c satisfying an equation:

$$A = \{(a, b, c): b^2 - 4ac = 0\}$$

QUESTIONS

2.2.1. A graduating engineer has signed up for three job interviews. She intends to categorize each one as being either a "success" or a "failure" depending on whether it leads to a plant trip. Write out the appropriate sample space. What outcomes are in the event A: Second success occurs on third interview? In B: First success never occurs? (*Hint:* Notice the similarity between this situation and the coin-tossing experiment described in Example 2.2.1.)

2.2.2. Three dice are tossed, one red, one blue, and one green. What outcomes make up the event A that the sum of the three faces showing equals five?

2.2.3. An urn contains six chips numbered 1 through 6. Three are drawn out. What outcomes are in the event "Second smallest chip is a 3"? Assume that the order of the chips is irrelevant.

2.2.4. Suppose that two cards are dealt from a standard 52-card poker deck. Let A be the event that the sum of the two cards is eight (assume that aces have a numerical value of one). How many outcomes are in A?

2.2.5. In the lingo of craps-shooters (where two dice are tossed and the underlying sample space is the matrix pictured in Figure 2.2.1) is the phrase "making a hard eight." What might that mean?

2.2.6. A poker deck consists of fifty-two cards, representing thirteen denominations (2 through Ace) and four suits (diamonds, hearts, clubs, and spades). A five-card hand is called a *flush* if all five cards are in the same suit, but not all five denominations are consecutive. Pictured next is a flush in hearts. Let N be the set of five cards in hearts that are *not* flushes. How many outcomes are in N? Note: In poker, the denominations (A, 2, 3, 4, 5) are considered to be consecutive (in addition to sequences such as (8, 9, 10, J, Q)).

		Denominations												
		2	3	4	5	6	7	8	9	10	J	Q	K	A
Suits	D													
	H	X	X				X				X	X		
	C													
	S													

2.2.7. Let P be the set of right triangles with a 5" hypotenuse and whose height and length are a and b, respectively. Characterize the outcomes in P.

2.2.8. Suppose a baseball player steps to the plate with the intention of trying to "coax" a base on balls by never swinging at a pitch. The umpire, of course, will necessarily call each pitch either a ball (B) or a strike (S). What outcomes make up the event A, that a batter walks on the sixth pitch? Note: A batter "walks" if the fourth ball is called before the third strike.

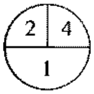
2.2.9. A telemarketer is planning to set up a phone bank to bilk widows with a Ponzi scheme. His past experience (prior to his most recent incarceration) suggests that each phone will be in use half the time. For a given phone at a given time, let *0* indicate that the phone is available and let *1* indicate that a caller is on the line. Suppose that the telemarketer's "bank" is comprised of four telephones.

(a) Write out the outcomes in the sample space.

(b) What outcomes would make up the event that exactly two phones were being used?

(c) Suppose the telemarketer had k phones. How many outcomes would allow for the possibility that at most one more call could be received?

2.2.10. Two darts are thrown at the following target:



(a) Let (u, v) denote the outcome that the first dart lands in region u and the second dart, in region v. List the sample space of (u, v)'s.

(b) List the outcomes in the sample space of *sums*, $u + v$.

2.2.11. A woman has her purse snatched by two teenagers. She is subsequently shown a police lineup consisting of five suspects, including the two perpetrators. What is the sample space associated with the experiment "Woman picks two suspects out of lineup"? Which outcomes are in the event A: She makes at least one incorrect identification?

2.2.12. Consider the experiment of choosing coefficients for the quadratic equation $ax^2 + bx + c = 0$. Characterize the values of a, b, and c associated with the event A: Equation has imaginary roots.

2.2.13. In the game of craps, the person rolling the dice (the *shooter*) wins outright if his first toss is a 7 or an 11. If his first toss is a 2, 3, or 12, he loses outright. If his first roll is something else, say, a 9, that number becomes his "point" and he keeps rolling the dice until he either rolls another 9, in which case he wins, or a 7, in which case he loses. Characterize the sample outcomes contained in the event "Shooter wins with a point of 9."

2.2.14. A probability-minded despot offers a convicted murderer a final chance to gain his release. The prisoner is given twenty chips, ten white and ten black. All twenty are to be placed into two urns, according to any allocation scheme the prisoner wishes, with the one proviso being that each urn contain at least one chip. The executioner will then pick one of the two urns at random and from that urn, one chip at random. If the chip selected is white, the prisoner will be set free; if it is black, he "buys the farm." Characterize the sample space describing the prisoner's possible allocation options. (Intuitively, which allocation affords the prisoner the greatest chance of survival?)

2.2.15. Suppose that ten chips, numbered 1 through 10, are put into an urn at one minute to midnight, and chip number 1 is quickly removed. At one-half minute to midnight, chips numbered 11 through 20 are added to the urn, and chip number 2 is quickly removed. Then at one-fourth minute to midnight, chips numbered 21 to 30 are added to the urn, and chip number 3 is quickly removed. If that procedure for adding chips to the urn continues, how many chips will be in the urn at midnight (152)?

Unions, Intersections, and Complements

Associated with events defined on a sample space are several operations collectively referred to as the *algebra of sets*. These are the rules that govern the ways in which one event can be combined with another. Consider, for example, the game of craps described in Question 2.2.13. The shooter wins on his initial roll if he throws either a 7 or an 11. In the language of the algebra of sets, the event "shooter rolls a 7 *or* an 11" is the *union* of two simpler events, "shooter rolls a 7" and "shooter rolls an 11." If E denotes the union and if A and B denote the two events making up the union, we write $E = A \cup B$. The next several definitions and examples illustrate those portions of the algebra of sets that we will find particularly useful in the chapters ahead.

Definition 2.2.1. Let A and B be any two events defined over the same sample space S. Then

a. The *intersection* of A and B, written $A \cap B$, is the event whose outcomes belong to both A and B.

b. The *union* of A and B, written $A \cup B$, is the event whose outcomes belong to either A or B or both.

EXAMPLE 2.2.6

A single card is drawn from a poker deck. Let A be the event that an ace is selected:

$$A = \{\text{ace of hearts, ace of diamonds, ace of clubs, ace of spades}\}$$

Let B be the event "Heart is drawn":

$$B = \{\text{2 of hearts, 3 of hearts}, \ldots, \text{ace of hearts}\}$$

Then

$$A \cap B = \{\text{ace of hearts}\}$$

and

$$A \cup B = \{\text{2 of hearts, 3 of hearts}, \ldots, \text{ace of hearts, ace of diamonds, ace of clubs, ace of spades}\}$$

(Let C be the event "club is drawn." Which cards are in $B \cup C$? In $B \cap C$?)

EXAMPLE 2.2.7

Let A be the set of x's for which $x^2 + 2x = 8$; let B be the set for which $x^2 + x = 6$. Find $A \cap B$ and $A \cup B$.

Since the first equation factors into $(x + 4)(x - 2) = 0$, its solution set is $A = \{-4, 2\}$. Similarly, the second equation can be written $(x + 3)(x - 2) = 0$, making $B = \{-3, 2\}$. Therefore,

$$A \cap B = \{2\}$$

and

$$A \cup B = \{-4, -3, 2\}$$

EXAMPLE 2.2.8

Consider the electrical circuit pictured in Figure 2.2.2. Let A_i denote the event that switch i fails to close, $i = 1, 2, 3, 4$. Let A be the event "Circuit is not completed." Express A in terms of the A_i's.

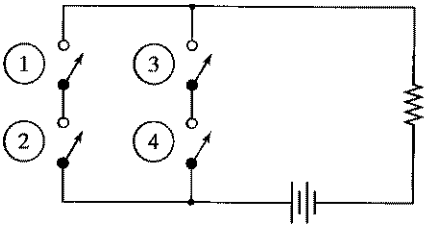


FIGURE 2.2.2

Call the ① and ② switches line a; call the ③ and ④ switches line b. By inspection, the circuit fails only if *both* line a and line b fail. But line a fails only if *either* ① *or* ② (or both) fail. That is, the event that line a fails is the union $A_1 \cup A_2$. Similarly, the failure of line b is the union $A_3 \cup A_4$. The event that the circuit fails, then, is an intersection:

$$A = (A_1 \cup A_2) \cap (A_3 \cup A_4)$$

Definition 2.2.2. Events A and B defined over the same sample space are said to be *mutually exclusive* if they have no outcomes in common—that is, if $A \cap B = \emptyset$, where $\emptyset$ is the null set.

EXAMPLE 2.2.9

Consider a single throw of two dice. Define A to be the event that the *sum* of the faces showing is odd. Let B be the event that the two faces themselves are odd. Then clearly the intersection is empty, the sum of two odd numbers necessarily being even. In symbols, $A \cap B = \emptyset$. (Recall the event $B \cap C$ asked for in Example 2.2.6.)

Definition 2.2.3. Let A be any event defined on a sample space S. The *complement* of A, written A^C, is the event consisting of all the outcomes in S other than those contained in A.

EXAMPLE 2.2.10

Let A be the set of (x, y)'s for which $x^2 + y^2 < 1$. Sketch the region in the xy-plane corresponding to A^C.

From analytic geometry, we recognize that $x^2 + y^2 < 1$ describes the interior of a circle of radius 1 centered at the origin. Figure 2.2.3 shows the complement—the points on the circumference of the circle and the points outside the circle.

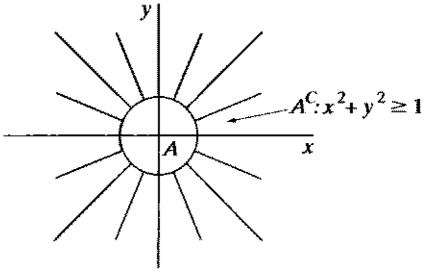


FIGURE 2.2.3

The notions of union and intersection can easily be extended to more than two events. For example, the expression $A_1 \cup A_2 \cup \cdots \cup A_k$ defines the set of outcomes belonging to *any* of the A_i's (or to any combination of the A_i's). Similarly, $A_1 \cap A_2 \cap \cdots \cap A_k$ is the set of outcomes belonging to *all* of the A_i's.

EXAMPLE 2.2.11

Suppose the events $A_1, A_2, \ldots, A_k$ are intervals of real numbers such that

$$A_i = \{x\colon\ 0 \leq x < 1/i\}, \quad i = 1, 2, \ldots, k$$

Describe the sets $A_1 \cup A_2 \cup \cdots \cup A_k = \cup_{i=1}^{k} A_i$ and $A_1 \cap A_2 \cap \cdots \cap A_k = \cap_{i=1}^{k} A_i$.

Notice that the A_i's are telescoping sets. That is, A_1 is the interval $0 \leq x < 1$, A_2 is the interval $0 \leq x < \frac{1}{2}$, and so on. It follows, then, that the *union* of the k A_i's is simply A_1 while the *intersection* of the A_i's (that is, their overlap) is A_k.

QUESTIONS

2.2.16. Sketch the regions in the xy-plane corresponding to $A \cup B$ and $A \cap B$ if

$$A = \{(x, y)\colon 0 < x < 3, 0 < y < 3\}$$

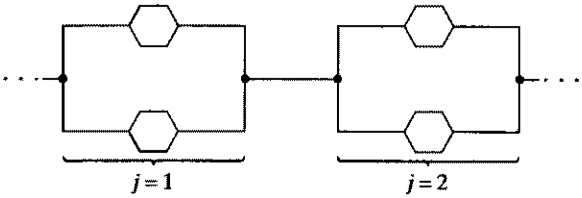
and

$$B = \{(x, y)\colon 2 < x < 4, 2 < y < 4\}$$

2.2.17. Referring to Example 2.2.7, find $A \cap B$ and $A \cup B$ if the two equations were replaced by inequalities: $x^2 + 2x \leq 8$ and $x^2 + x \leq 6$.

2.2.18. Find $A \cap B \cap C$ if $A = \{x\colon 0 \leq x \leq 4\}$, $B = \{x\colon 2 \leq x \leq 6\}$, and $C = \{x\colon x = 0, 1, 2, \ldots\}$.

2.2.19. An electronic system has four components divided into two pairs. The two components of each pair are wired in parallel; the two pairs are wired in series. Let A_{ij}, denote the event "ith component in jth pair fails," $i = 1, 2$; $j = 1, 2$. Let A be the event "System fails." Write A in terms of the A_{ij}'s.



2.2.20. Define $A = \{x\colon 0 \leq x \leq 1\}$, $B = \{x\colon 0 \leq x \leq 3\}$, and $C = \{x\colon -1 \leq x \leq 2\}$. Draw diagrams showing each of the following sets of points:
(a) $A^C \cap B \cap C$
(b) $A^C \cup (B \cap C)$
(c) $A \cap B \cap C^C$
(d) $((A \cup B) \cap C^C)^C$

2.2.21. Let A be the set of five-card hands dealt from a 52-card poker deck, where the denominations of the five cards are all consecutive—for example, (7 of Hearts, 8 of Spades, 9 of Spades, 10 of Hearts, Jack of Diamonds). Let B be the set of five-card hands where the suits of the five cards are all the same. How many outcomes are in the event $A \cap B$?

2.2.22. Suppose that each of the twelve letters in the word

$$T \quad E \quad S \quad S \quad E \quad L \quad L \quad A \quad T \quad I \quad O \quad N$$

is written on a chip. Define the events F, R, and C as follows:

F: letters in first half of alphabet
R: letters that are repeated
V: letters that are vowels

Which chips make up the following events:
(a) $F \cap R \cap V$
(b) $F^C \cap R \cap V^C$
(c) $F \cap R^C \cap V$

2.2.23. Let A, B, and C be any three events defined on a sample space S. Show that

(a) the outcomes in $A \cup (B \cap C)$ are the same as the outcomes in $(A \cup B) \cap (A \cup C)$

(b) the outcomes in $A \cap (B \cup C)$ are the same as the outcomes in $(A \cap B) \cup (B \cap C)$.

2.2.24. Let $A_1, A_2, \ldots, A_k$ be any set of events defined on a sample space S. What outcomes belong to the event

$$(A_1 \cup A_2 \cup \cdots \cup A_k) \cup \left(A_1^C \cap A_2^C \cap \cdots \cap A_k^C\right)$$

2.2.25. Let A, B, and C be any three events defined on a sample space S. Show that the operations of union and intersection are *associative* by proving that

(a) $A \cup (B \cup C) = (A \cup B) \cup C = A \cup B \cup C$

(b) $A \cap (B \cap C) = (A \cap B) \cap C = A \cap B \cap C$

2.2.26. Suppose that three events—A, B, and C—are defined on a sample space S. Use the union, intersection, and complement operations to represent each of the following events:

(a) none of the three events occurs

(b) all three of the events occur

(c) only event A occurs

(d) exactly one event occurs

(e) exactly two events occur

2.2.27. What must be true of events A and B if

(a) $A \cup B = B$

(b) $A \cap B = A$

2.2.28. Let events A and B and sample space S be defined as the following intervals:

$$S = \{x\colon\ 0 \le x \le 10\}$$
$$A = \{x\colon\ 0 < x < 5\}$$
$$B = \{x\colon\ 3 \le x \le 7\}$$

Characterize the following events:

(a) A^C

(b) $A \cap B$

(c) $A \cup B$

(d) $A \cap B^C$

(e) $A^C \cup B$

(f) $A^C \cap B^C$

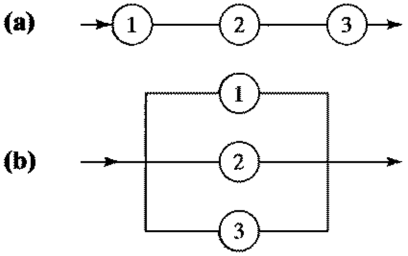
2.2.29. A coin is tossed four times and the resulting sequence of Heads and/or Tails is recorded. Define the events A, B, and C as follows:

A: Exactly two heads appear
B: Heads and tails alternate
C: First two tosses are heads

(a) Which events, if any, are mutually exclusive?

(b) Which events, if any, are subsets of other sets?

2.2.30. Pictured below are two organizational charts describing the way upper-management vets new proposals. For both models, three vice presidents—1, 2, and 3—each voice an opinion.



For (a), all three must concur if the proposal is to pass; if any one of the three favors the proposal in (b) it passes. Let A_i denote the event that vice-president i favors the proposal, $i = 1, 2, 3$, and let A denote the event that the proposal passes. Express A in terms of the A_i's for the two office protocols. Under what sorts of situations might one system be preferable to the other?

Expressing Events Graphically: Venn Diagrams

Relationships based on two or more events can sometimes be difficult to express using only equations or verbal descriptions. An alternative approach that can be highly effective is to represent the underlying events graphically in a format known as a *Venn diagram*. Figure 2.2.4 shows Venn diagrams for an intersection, a union, a complement, and for two events that are mutually exclusive. In each case, the shaded interior of a region corresponds to the desired event.

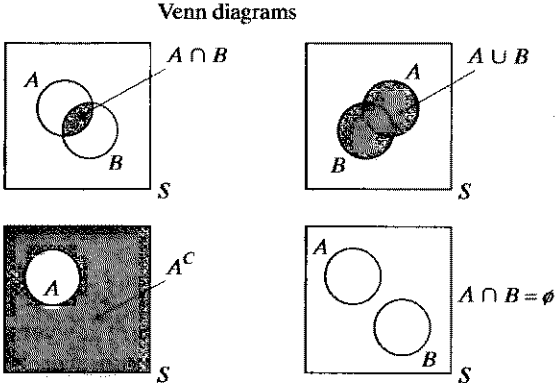


FIGURE 2.2.4

EXAMPLE 2.2.12

When two events A and B are defined on a sample space, we will frequently need to consider

a. the event that *exactly one* (of the two) occurs

b. the event that *at most one* (of the two) occurs

Getting expressions for each of these is easy if we visualize the corresponding Venn diagrams.

The shaded area in Figure 2.2.5 represents the event E that either A or B, but not both, occurs (that is, exactly one occurs).

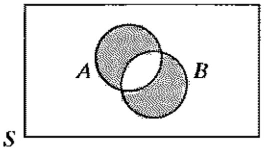


FIGURE 2.2.5

Just by looking at the diagram we can formulate an expression for E. The portion of A, for example, included in E is $A \cap B^C$. Similarly, the portion of B included in E is $B \cap A^C$. It follows that E can be written as a union:

$$E = (A \cap B^C) \cup (B \cap A^C)$$

(Convince yourself that an equivalent expression for E is $(A \cap B)^C \cap (A \cup B)$.)

Figure 2.2.6 shows the event F that *at most one* (of the two events) occurs. Since the latter includes every outcome except those belonging to *both* A and B, we can write

$$F = (A \cap B)^C$$

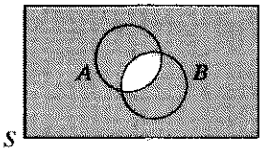


FIGURE 2.2.6

EXAMPLE 2.2.13

When Swampwater Tech's Class of '64 held its fortieth reunion, one hundred grads attended. Fifteen of those alumni were lawyers and rumor had it that thirty of the one hundred were psychopaths. If ten alumni were both lawyers and psychopaths, how many suffered from neither of those afflictions?

Let L be the set of lawyers and H, the set of psychopaths. If the symbol $N(Q)$ is defined to be the number of members in set Q, then

$$\begin{aligned} N(S) &= 100 \\ N(L) &= 15 \\ N(H) &= 30 \\ N(L \cap H) &= 10 \end{aligned}$$

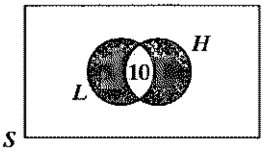


FIGURE 2.2.7

Summarizing all this information is the Venn diagram in Figure 2.2.7. Notice that

$$\begin{aligned} N(L \cup H) &= \text{number of alumni suffering from at least one affliction} \\ &= 5 + 10 + 20 \\ &= 35 \end{aligned}$$

which implies that 100 − 35, or *65* were neither lawyers nor psychopaths. In effect,

$$N(L \cup H) = N(L) + N(H) - N(L \cap H) \quad [= 15 + 30 - 10 = 35]$$

QUESTIONS

2.2.31. During orientation week, the latest Spiderman movie was shown twice at State University. Among the entering class of 6000 freshmen, 850 went to see it the first time, 690 the second time, while 4700 failed to see it either time. How many saw it twice?

2.2.32. Let A and B be any two events. Use Venn diagrams to show that

(a) the complement of their intersection is the union of their complements:

$$(A \cap B)^C = A^C \cup B^C$$

(b) the complement of their union is the intersection of their complements:

$$(A \cup B)^C = A^C \cap B^C$$

(These two results are known as *DeMorgan's laws*.)

2.2.33. Let A, B, and C be any three events. Use Venn diagrams to show that

(a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

(b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

2.2.34. Let A, B, and C be any three events. Use Venn diagrams to show that

(a) $A \cup (B \cup C) = (A \cup B) \cup C$

(b) $A \cap (B \cap C) = (A \cap B) \cap C$

2.2.35. Let A and B be any two events defined on a sample space S. Which of the following sets are necessarily subsets of which other sets?

$$A \qquad B \qquad A \cup B \qquad A \cap B \qquad A^C \cap B$$
$$A \cap B^C \qquad (A^C \cup B^C)^C$$

2.2.36. Use Venn diagrams to suggest an equivalent way of representing the following events:

(a) $(A \cap B^C)^C$

(b) $B \cup (A \cup B)^C$

(c) $A \cap (A \cap B)^C$

2.2.37. A total of twelve hundred graduates of State Tech have gotten into medical school in the past several years. Of that number, one thousand earned scores of twenty-seven or higher on the MCAT and four hundred had GPAs that were 3.5 or higher. Moreover, three hundred had MCATs that were twenty-seven or higher *and* GPAs that were 3.5 or higher. What proportion of those twelve hundred graduates got into medical school with an MCAT lower than twenty-seven and a GPA below 3.5?

2.2.38. Let A, B, and C be any three events defined on a sample space S. Let $N(A)$, $N(B)$, $N(C)$, $N(A \cap B)$, $N(A \cap C)$, $N(B \cap C)$, and $N(A \cap B \cap C)$ denote the numbers of outcomes in all the different intersections in which A, B, and C are involved. Use a Venn diagram to suggest a formula for $N(A \cup B \cup C)$. Hint: Start with the sum $N(A) + N(B) + N(C)$ and use the Venn diagram to identify the "adjustments" that need to be made to that sum before it can equal $N(A \cup B \cup C)$. As a precedent, recall from p. 35 that $N(A \cup B) = N(A) + N(B) - N(A \cap B)$. There, in the case of *two* events, subtracting $N(A \cap B)$ is the "adjustment."

2.2.39. A poll conducted by a potential presidential candidate asked two questions: (1) Do you support the candidate's position on taxes? and (2) Do you support the candidate's position on homeland security? A total of twelve hundred responses were received; six hundred said "yes" to the first question and four hundred said "yes" to the second. If three hundred respondents said "no" to the taxes question and "yes" to the homeland security question, how many said "yes" to the taxes question but "no" to the homeland security question?

2.2.40. For two events A and B defined on a sample space S, $N(A \cap B^C) = 15$, $N(A^C \cap B) = 50$, and $N(A \cap B) = 2$. Given that $N(S) = 120$, how many outcomes belong to neither A nor B?

2.3 THE PROBABILITY FUNCTION

Having introduced in Section 2.2 the twin concepts of "experiment" and "sample space," we are now ready to pursue in a formal way the all-important problem of assigning a *probability* to an experiment's outcome—and, more generally, to an event. Specifically, if A is any event defined on a sample space S, the symbol $P(A)$ will denote the *probability of* A, and we will refer to P as the *probability function*. It is, in effect, a mapping from a set (i.e., an event) to a number. The backdrop for our discussion will be the unions, intersections, and complements of set theory; the starting point will be the axioms referred to in Section 2.1 that were originally set forth by Kolmogorov.

If S has a finite number of members, Kolmogorov showed that as few as three axioms are necessary and sufficient for characterizing the probability function P:

Axiom 1. *Let A be any event defined over S. Then $P(A) \geq 0$.*

Axiom 2. $P(S) = 1$.

Axiom 3. *Let A and B be any two mutually exclusive events defined over S. Then*

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

When S has an infinite number of members, a fourth axiom is needed:

Axiom 4. *Let $A_1, A_2, \ldots,$ be events defined over S. If $A_i \cap A_j = \emptyset$ for each $i \neq j$, then*

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

From these simple statements come the general rules for manipulating the probability function that apply no matter what specific mathematical form it may take in a particular context.

Some Basic Properties of *P*

Some of the immediate consequences of Kolmogorov's axioms are the results given in Theorems 2.3.1 through 2.3.6. Despite their simplicity, several of these properties—as we will soon see—prove to be immensely useful in solving all sorts of problems.

Theorem 2.3.1. $P(A^C) = 1 - P(A)$.

Proof. By Axiom 2 and Definition 2.2.3,

$$P(S) = 1 = P(A \cup A^C)$$

But A and A^C are mutually exclusive, so

$$P(A \cup A^C) = P(A) + P(A^C)$$

and the result follows. ☐

Theorem 2.3.2. $P(\emptyset) = 0$.

Proof. Since $\emptyset = S^C$, $P(\emptyset) = P(S^C) = 1 - P(S) = 0$. ☐

Theorem 2.3.3. *If $A \subset B$, then $P(A) \leq P(B)$.*

Proof. Note that the event B may be written in the form

$$B = A \cup (B \cap A^C)$$

where A and $(B \cap A^C)$ are mutually exclusive. Therefore,

$$P(B) = P(A) + P(B \cap A^C)$$

which implies that $P(B) \geq P(A)$ since $P(B \cap A^C) \geq 0$. ☐

Theorem 2.3.4. *For any event A, $P(A) \leq 1$.*

Proof. The proof follows immediately from Theorem 2.3.3 because $A \subset S$ and $P(S) = 1$. ☐

Theorem 2.3.5. *Let $A_1, A_2, \ldots, A_n$ be events defined over S. If $A_i \cap A_j = \emptyset$ for $i \neq j$, then*

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

Proof. The proof is a straightforward induction argument with Axiom 3 being the starting point. □

Theorem 2.3.6. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

Proof. The Venn diagram for $A \cup B$ certainly suggests that the statement of the theorem is true (recall Figure 2.2.4). More formally, we have from Axiom 3 that

$$P(A) = P(A \cap B^C) + P(A \cap B)$$

and

$$P(B) = P(B \cap A^C) + P(A \cap B)$$

Adding these two equations gives

$$P(A) + P(B) = [P(A \cap B^C) + P(B \cap A^C) + P(A \cap B)] + P(A \cap B)$$

By Theorem 2.3.5, the sum in the brackets is $P(A \cup B)$. If we subtract $P(A \cap B)$ from both sides of the equation, the result follows. □

EXAMPLE 2.3.1

Let A and B be two events defined on a sample space S such that $P(A) = 0.3$, $P(B) = 0.5$, and $P(A \cup B) = 0.7$. Find (a) $P(A \cap B)$, (b) $P(A^C \cup B^C)$, and (c) $P(A^C \cap B)$.

a. Transposing the terms in Theorem 2.3.6 yields a general formula for the probability of an intersection:

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

Here

$$\begin{aligned} P(A \cap B) &= 0.3 + 0.5 - 0.7 \\ &= 0.1 \end{aligned}$$

b. The two cross-hatched regions in Figure 2.3.1 correspond to A^C and B^C. The union of A^C and B^C consists of those regions that have cross-hatching in either or both directions. By inspection, the only portion of S *not* included in $A^C \cup B^C$ is the intersection, $A \cap B$. By Theorem 2.3.1, then,

$$\begin{aligned} P(A^C \cup B^C) &= 1 - P(A \cap B) \\ &= 1 - 0.1 \\ &= 0.9 \end{aligned}$$

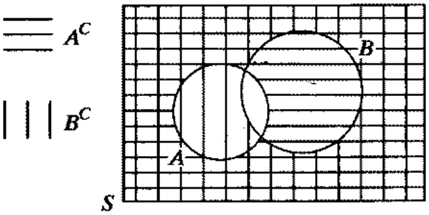


FIGURE 2.3.1

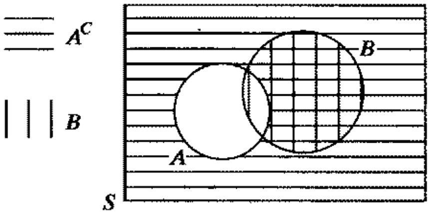


FIGURE 2.3.2

c. The event $A^C \cap B$ corresponds to the region in Figure 2.3.2 where the cross-hatching extends in *both* directions—that is, everywhere in B except the intersection with A. Therefore,

$$\begin{aligned} P(A^C \cap B) &= P(B) - P(A \cap B) \\ &= 0.5 - 0.1 \\ &= 0.4 \end{aligned}$$

EXAMPLE 2.3.2

Show that

$$P(A \cap B) \geq 1 - P(A^C) - P(B^C)$$

for any two events A and B defined on a sample space S.

From Example 2.3.1a and Theorem 2.3.1,

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

But $P(A \cup B) \leq 1$ from Theorem 2.3.4, so

$$P(A \cap B) \geq 1 - P(A^C) - P(B^C)$$

EXAMPLE 2.3.3

Two cards are drawn from a poker deck without replacement. What is the probability that the second is higher in rank than the first?

Let A_1, A_2, and A_3 be the events "First card is lower in rank," "First card is higher in rank," and "Both cards have same rank," respectively. Clearly, the three A_i's are mutually exclusive and they account for all possible outcomes, so from Theorem 2.3.5,

$$P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) = P(S) = 1$$

Once the first card is drawn, there are three choices for the second that would have the same rank—that is, $P(A_3) = \frac{3}{51}$. Moreover, symmetry demands that $P(A_1) = P(A_2)$, so

$$2P(A_2) + \frac{3}{51} = 1$$

implying that $P(A_2) = \frac{8}{17}$.

EXAMPLE 2.3.4

In a newly released martial arts film, the actress playing the lead role has a stunt double who handles all of the physically dangerous action scenes. According to the script, the actress appears in 40% of the film's scenes, her double appears in 30%, and the two of them are together 5% of the time. What is the probability that in a given scene (a) only the stunt double appears and (b) neither the lead actress nor the double appears?

a. If L is the event "Lead actress appears in scene" and D is the event "Double appears in scene," we are given that $P(L) = 0.40$, $P(D) = 0.30$, and $P(L \cap D) = 0.05$. It follows that

$$\begin{aligned} P(\text{Only double appears}) &= P(D) - P(L \cap D) \\ &= 0.30 - 0.05 \\ &= 0.25 \end{aligned}$$

(recall Example 2.3.1c).

b. The event "Neither appears" is the complement of the event "At least one appears." But $P(\text{At least one appears}) = P(L \cup D)$. From Theorems 2.3.1 and 2.3.6, then,

$$\begin{aligned} P(\text{Neither appears}) &= 1 - P(L \cup D) \\ &= 1 - [P(L) + P(D) - P(L \cap D)] \\ &= 1 - [0.40 + 0.30 - 0.05] \\ &= 0.35 \end{aligned}$$

EXAMPLE 2.3.5

Having endured (and survived) the mental trauma that comes from taking two years of chemistry, a year of physics, and a year of biology, Biff decides to test the medical school waters and sends his MCATs to two colleges, X and Y. Based on how his friends have fared, he estimates that his probability of being accepted at X is 0.7, and at Y is 0.4. He also suspects there is a 75% chance that at least one of his applications will be rejected. What is the probability that he gets at least one acceptance?

Let A be the event "School X accepts him" and B, the event "school Y accepts him." We are given that $P(A) = 0.7$, $P(B) = 0.4$, and $P(A^C \cup B^C) = 0.75$. The question is asking for $P(A \cup B)$.

From Theorem 2.3.6,

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

Recall from Question 2.2.32 that $A^C \cup B^C = (A \cap B)^C$, so

$$P(A \cap B) = 1 - P[(A \cap B)^C] = 1 - 0.75 = 0.25$$

It follows that Biff's prospects are not all that bleak—he has an 85% chance of getting in somewhere:

$$\begin{aligned} P(A \cup B) &= 0.7 + 0.4 - 0.25 \\ &= 0.85 \end{aligned}$$

Comment. Notice that $P(A \cup B)$ varies directly with $P(A^C \cup B^C)$:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - (1 - P(A^C \cup B^C)) \\ &= P(A) + P(B) - 1 + P(A^C \cup B^C) \end{aligned}$$

If $P(A)$ and $P(B)$, then, are fixed, we get the curious result that Biff's chances of getting at least one acceptance increase if his chances of at least one rejection increase.

QUESTIONS

2.3.1. According to a family-oriented lobbying group, there is too much crude language and violence on television. Forty-two percent of the programs they screened had language they found offensive, 27% were too violent, and 10% were considered excessive in both language and violence. What percentage of programs did comply with the group's standards?

2.3.2. Let A and B be any two events defined on S. Suppose that $P(A) = 0.4$, $P(B) = 0.5$, and $P(A \cap B) = 0.1$. What is the probability that A or B but not both occur?

2.3.3. Express the following probabilities in terms of $P(A)$, $P(B)$, and $P(A \cap B)$.

(a) $P(A^C \cup B^C)$

(b) $P(A^C \cap (A \cup B))$

2.3.4. Let A and B be two events defined on S. If the probability that at least one of them occurs is 0.3 and the probability that A occurs but B does not occur is 0.1, what is $P(B)$?

2.3.5. Suppose that three fair dice are tossed. Let A_i be the event that a 6 shows on the ith die, $i = 1, 2, 3$. Does $P(A_1 \cup A_2 \cup A_3) = \frac{1}{2}$? Explain.

2.3.6. Events A and B are defined on a sample space S such that $P((A \cup B)^C) = 0.6$ and $P(A \cap B) = 0.2$. What is the probability that either A or B but not both will occur?

2.3.7. Let $A_1, A_2, \ldots, A_n$ be a series of events for which $A_i \cap A_j = \emptyset$ if $i \neq j$ and $A_1 \cup A_2 \cup \cdots \cup A_n = S$. Let B be any event defined on S. Express B as a union of intersections.

2.3.8. Draw the Venn diagrams that would correspond to the equations (a) $P(A \cap B) = P(B)$ and (b) $P(A \cup B) = P(B)$.

2.3.9. In the game of "odd man out" each player tosses a fair coin. If all the coins turn up the same except for one, the player tossing the different coin is declared the odd man out and is eliminated from the contest. Suppose that three people are playing. What is the probability that someone will be eliminated on the first toss? (*Hint:* Use Theorem 2.3.1.)

2.3.10. An urn contains twenty-four chips, numbered 1 through 24. One is drawn at random. Let A be the event that the number is divisible by two and let B be the event that the number is divisible by three. Find $P(A \cup B)$.

2.3.11. If State's football team has a 10% chance of winning Saturday's game, a 30% chance of winning two weeks from now, and a 65% chance of losing both games, what are their chances of winning exactly once?

2.3.12. Events A_1 and A_2 are such that $A_1 \cup A_2 = S$ and $A_1 \cap A_2 = \emptyset$. Find p_2 if $P(A_1) = p_1$, $P(A_2) = p_2$, and $3p_1 - p_2 = \frac{1}{2}$.

2.3.13. Consolidated Industries has come under considerable pressure to eliminate its seemingly discriminatory hiring practices. Company officials have agreed that during the next five years, 60% of their new employees will be females and 30% will be minorities. One out of four new employees, though, will be white males. What percentage of their new hires will be minority females?

2.3.14. Three events—A, B, and C—are defined on a sample space, S. Given that $P(A) = 0.2$, $P(B) = 0.1$, and $P(C) = 0.3$, what is the smallest possible value for $P[(A \cup B \cup C)^C]$?

2.3.15. A coin is to be tossed four times. Define events X and Y such that

X: first and last coins have opposite faces

Y: exactly two heads appear

Assume that each of the sixteen Head/Tail sequences has the same probability. Evaluate

(a) $P(X^C \cap Y)$

(b) $P(X \cap Y^C)$

2.3.16. Two dice are tossed. Assume that each possible outcome has a $\frac{1}{36}$ probability. Let A be the event that the sum of the faces showing is 6, and let B be the event that the face showing on one die is twice the face showing on the other. Calculate $P(A \cap B^C)$.

2.3.17. Let, A, B, and C be three events defined on a sample space, S. Arrange the probabilities of the following events from smallest to largest:

(a) $A \cup B$

(b) $A \cap B$

(c) A

(d) S

(e) $(A \cap B) \cup (A \cap C)$

2.3.18. Lucy is currently running two dot-com scams out of a bogus chatroom. She estimates that the chances of the first one leading to her arrest are one in ten; the "risk" associated with the second is more on the order of one in thirty. She considers the likelihood that she gets busted for both to be 0.0025. What are Lucy's chances of avoiding incarceration?

2.4 CONDITIONAL PROBABILITY

In Section 2.3, we calculated probabilities of certain events by manipulating other probabilities whose values we were given. Knowing $P(A)$, $P(B)$, and $P(A \cap B)$, for example, allows us to calculate $P(A \cup B)$ (recall Theorem 2.3.6). For many real-world situations, though, the "given" in a probability problem goes beyond simply knowing a set of other probabilities. Sometimes, we know *for a fact* that certain events *have already occurred*, and those occurrences may have a bearing on the probability we are trying to find. In short, the probability of an event A may have to be "adjusted" if we know for certain that some related event B has already occurred. Any probability that is revised to take into account the (known) occurrence of other events is said to be a ***conditional probability.***

Consider a fair die being tossed, with A defined as the event "6 appears." Clearly, $P(A) = \frac{1}{6}$. But suppose that the die has already been tossed—by someone who refuses to tell us whether or not A occurred but does enlighten us to the point of confirming that B occurred, where B is the event "Even number appears." What are the chances of A now? Here, common sense can help us: There are three equally likely even numbers making up the event B—one of them satisfies the event A, so the "updated" probability is $\frac{1}{3}$.

Notice that the effect of additional information, such as the knowledge that B has occurred, is to revise—indeed, to *shrink*—the original sample space S to a new set of outcomes S'. In this example, the original S contained six outcomes, the conditional sample space, three (see Figure 2.4.1).

The symbol $P(A|B)$—read "the probability of A given B"—is used to denote a conditional probability. Specifically, $P(A|B)$ refers to the probability that A *will occur* given that B *has already occurred.*

It will be convenient to have a formula for $P(A|B)$ that can be evaluated in terms of the original S, rather than the revised S'. Suppose that S is a finite sample space with n outcomes, all equally likely. Assume that A and B are two events containing a and b outcomes, respectively, and let c denote the number of outcomes in the intersection of A and B (see Figure 2.4.2). Based on the argument suggested in Figure 2.4.1, the *conditional probability of A given B is the ratio of c to b.* But c/b can be written as the quotient of two other ratios,

$$\frac{c}{b} = \frac{c/n}{b/n}$$

so, for this particular case,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \tag{2.4.1}$$

The same underlying reasoning that leads to Equation 2.4.1, though, holds true even when the outcomes are not equally likely or when S is uncountably infinite.

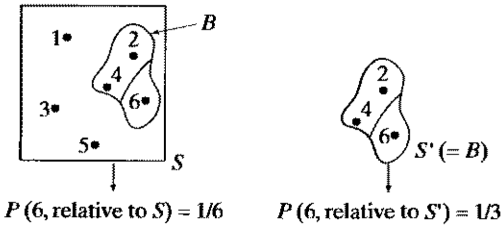


FIGURE 2.4.1

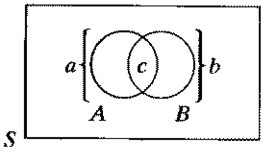


FIGURE 2.4.2

Definition 2.4.1. Let A and B be any two events defined on S such that $P(B) > 0$. The conditional probability of A, assuming that B has already occurred, is written $P(A|B)$

and is given by

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

Comment. Definition 2.4.1 can be cross-multiplied to give a frequently useful expression for the probability of an intersection. If $P(A|B) = P(A \cap B)/P(B)$, then

$$P(A \cap B) = P(A|B)P(B). \tag{2.4.2}$$

EXAMPLE 2.4.1

A card is drawn from a poker deck. What is the probability that the card is a club, given that the card is a king?

Intuitively, the answer is $\frac{1}{4}$: The king is equally likely to be a heart, diamond, club, or spade. More formally, let C be the event "Card is a club"; let K be the event "Card is a king." By Definition 2.4.1,

$$P(C|K) = \frac{P(C \cap K)}{P(K)}$$

But $P(K) = \frac{4}{52}$ and $P(C \cap K) = P(\text{card is a king of clubs}) = \frac{1}{52}$. Therefore, confirming our intuition,

$$P(C|K) = \frac{1/52}{4/52} = \frac{1}{4}$$

[Notice in this example that the conditional probability $P(C|K)$ is numerically the same as the unconditional probability $P(C)$—they both equal $\frac{1}{4}$. This means that our knowledge that K has occurred gives us no additional insight about the chances of C occurring. Two events having this property are said to be *independent*. We will examine the notion of independence and its consequences in detail in Section 2.5.]

EXAMPLE 2.4.2

Our intuitions can often be fooled by probability problems, even ones that appear to be simple and straightforward. The "two boys" problem described here is an often-cited case in point.

Consider the set of families having two children. Assume that the four possible birth sequences—(younger child is a boy, older child is a boy), (younger child is a boy, older child is a girl), and so on—are equally likely. What is the probability that both children are boys given that at least one is a boy?

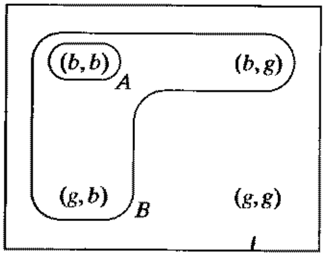
The answer is *not* $\frac{1}{2}$. The correct answer can be deduced from Definition 2.4.1. By assumption, the four possible birth sequences—(b, b), (b, g), (g, b), and (g, g)—each has a $\frac{1}{4}$ probability of occurring. Let A be the event that both children are boys, and let B be the event that at least one child is a boy. Then

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

since A is a subset of B (so the overlap between A and B is just A). But A has one outcome $\{(b, b)\}$ and B has three outcomes $\{(b, g), (g, b), (b, b)\}$. Applying Definition 2.4.1, then, gives

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

Another correct approach is to go back to the sample space and deduce the value of $P(A|B)$ from first principles. Figure 2.4.3 shows events A and B defined on the four family types that comprise the sample space S. Knowing that B has occurred redefines the sample space to include *three* outcomes, each now having a $\frac{1}{3}$ probability. Of those three possible outcomes, one—namely, (b, b)—satisfies the event A. It follows that $P(A|B) = \frac{1}{3}$.



S = sample space of two-child families
[outcomes written as (first born, second born)]

FIGURE 2.4.3

EXAMPLE 2.4.3

Two events A and B are defined such that (1) the probability that A occurs but B does not occur is 0.2, (2) the probability that B occurs but A does not occur is 0.1, and (3) the probability that neither occurs is 0.6. What is $P(A|B)$?

The three events whose probabilities are given are indicated on the Venn diagram shown in Figure 2.4.4. Since

$$P(\text{neither occurs}) = 0.6 = P((A \cup B)^C)$$

it follows that

$$P(A \cup B) = 1 - 0.6 = 0.4 = P(A \cap B^C) + P(A \cap B) + P(B \cap A^C)$$

so

$$\begin{aligned} P(A \cap B) &= 0.4 - 0.2 - 0.1 \\ &= 0.1 \end{aligned}$$

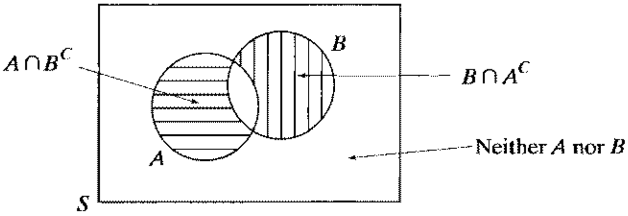


FIGURE 2.4.4

From Definition 2.4.1, then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A \cap B)}{P(A \cap B) + P(B \cap A^C)}$$
$$= \frac{0.1}{0.1 + 0.1}$$
$$= 0.5$$

EXAMPLE 2.4.4

The possibility of importing liquified natural gas (LNG) from Algeria has been suggested as one way of coping with a future energy crunch. Complicating matters, though, is the fact that LNG is highly volatile and poses an enormous safety hazard. Any major spill occurring near a U.S. port could result in a fire of catastrophic proportions. The question, therefore, of the *likelihood* of a spill becomes critical input for future policymakers who may have to decide whether or not to implement the proposal.

Two numbers need to be taken into account: (1) the probability that a tanker will have an accident near a port, and (2) the probability that a major spill will develop *given* that an accident has happened. Although no significant spills of LNG have yet occurred anywhere in the world, these probabilities can be approximated from records kept on similar tankers transporting less dangerous cargo. On the basis of such data, it has been estimated (44) that the probability is $\frac{8}{50{,}000}$ that an LNG tanker will have an accident on any one trip. Given that an accident *has* occurred, it is suspected that only 3 times in 15,000 will the damage be sufficiently severe that a major spill would develop. What are the chances that a given LNG shipment would precipitate a catastrophic disaster?

Let A denote the event "Spill develops" and let B denote the event "Accident occurs." Past experience is suggesting that $P(B) = \frac{8}{50{,}000}$ and $P(A|B) = \frac{3}{15{,}000}$. Of primary concern is the probability that an accident will occur *and* a spill will ensue—that is, $P(A \cap B)$. Using Equation 2.4.2, we find that the chances of a catastrophic accident are on the order of 3 in 100 million:

$$
\begin{aligned}
P(\text{accident occurs and spill develops}) &= P(A \cap B) \\
&= P(A|B)P(B) \\
&= \frac{3}{15{,}000} \cdot \frac{8}{50{,}000} \\
&= 0.000000032
\end{aligned}
$$

EXAMPLE 2.4.5

Max and Muffy are two myopic deer hunters who shoot simultaneously at a nearby sheepdog that they have mistaken for a 10-point buck. Based on years of well-documented ineptitude, it can be assumed that Max has a 20% chance of hitting a stationary target at close range, Muffy has a 30% chance, and the probability is 0.06 that they would both be on target. Suppose that the sheepdog is hit and killed by exactly one bullet. What is the probability that Muffy fired the fatal shot?

Let A be the event that Max hit the dog, and let B be the event that Muffy hit the dog. Then $P(A) = 0.2$, $P(B) = 0.3$, and $P(A \cap B) = 0.06$. We are trying to find

$$P(B|(A^C \cap B) \cup (A \cap B^C))$$

where the event $(A^C \cap B) \cup (A \cap B^C)$ is the union of A and B minus the intersection—that is, it represents the event that either A or B but not both occur (recall Figure 2.4.4).

Notice, also, from Figure 2.4.4 that the intersection of B and $(A^C \cap B) \cup (A \cap B^C)$ is the event $A^C \cap B$. Therefore, from Definition 2.4.1,

$$
\begin{aligned}
P(B|(A^C \cap B) \cup (A \cap B^C)) &= [P(A^C \cap B)]/[P\{(A^C \cap B) \cup (A \cap B^C)\}] \\
&= [P(B) - P(A \cap B)]/[P(A \cup B) - P(A \cap B)] \\
&= [0.3 - 0.06]/[0.2 + 0.3 - 0.06 - 0.06] \\
&= 0.63
\end{aligned}
$$

CASE STUDY 2.4.1 (Optional)

There once was a brainy baboon
Who always breathed down a bassoon
  For he said, "It appears
  That in billions of years
I shall certainly hit on a tune."
Eddington

(Continued on next page)

(Case Study 2.4.1 continued)

```
A GO THIS BABE AND JUDGEMENT OF TIMEDIOUS RETCH AND NOT LORD
WHAL IF THE EASELVES AND DO AND MAKE AND BASE GATHEM I AY
BEATELLOUS WE PLAY MEANS HOLY FOOL MOUR WORK FROM INMOST
BED BE CONFOULD HAVE MANY JUDGEMENT WAS IT YOU MASSURE'S TO
LADY WOULD HAT PRIME THAT'S OUR THROWN AND DID WIFE FATHER'ST
LIVENGTH SLEEP TITH I AMBITION TO THIN HIM AND FORCE AND LAW'S
MAY BUT SMELL SO AND SPURSELY SIGNOR GENT MUCH CHIEF MIXTURN
```

FIGURE 2.4.6

One can only wonder how "human" computer-generated text might be if conditional probabilities for, say, seven- or eight-letter sequences were available. Right now they are not, but given the rate that computer technology is developing, they soon will be. When that day comes, our monkey will probably still never come up with text as creative as Hamlet's soliloquy, but a fairly decent limerick might show up from time to time!

CASE STUDY 2.4.2 (Optional)

Several years ago, a television program (inadvertently) spawned a conditional probability problem that led to more than a few heated discussions, even in the national media. The show was *Let's Make a Deal*, and the question involved the strategy that contestants should take to maximize their chances of winning prizes.

On the program, a contestant would be presented with three doors: behind one of which was the prize. After the contestant had selected a door, the host, Monty Hall, would open one of the other two doors, showing that the prize was not there. Then he would give the contestant a choice—either stay with the door initially selected or switch to the "third" door that had not been opened.

For many viewers, common sense seemed to suggest that switching doors would make no difference. By assumption, the prize had a one-third chance of being behind each of the doors when the game began. Once a door was opened, it was argued that each of the remaining doors now had a one-half probability of hiding the prize, so contestants gained nothing by switching their bets.

Not so. An application of Definition 2.4.1, shows that it *does* make a difference—contestants, in fact, *double* their chances of winning by switching doors. To see why, consider a specific (but typical) case: the contestant has bet on Door #2 and Monty Hall has opened Door #3. Given that sequence of events, we need to calculate and compare the conditional probability of the prize being behind Door #1 and Door #2, respectively. If the former is larger (and we will prove that it is), the contestant should switch doors.

(Continued on next page)

TABLE 2.4.1

Character	Frequency	Probability	Random Number Range
Space	6934	0.1968	00001–06934
E	3277	0.0930	06935–10211
O	2578	0.0732	10212–12789
T	2557	0.0726	12790–15346
A	2043	0.0580	15347–17389
S	1856	0.0527	17390–19245
H	1773	0.0503	19246–21018
N	1741	0.0494	21019–22759
I	1736	0.0493	22760–24495
R	1593	0.0452	24496–26088
L	1238	0.0351	26089–27326
D	1099	0.0312	27327–28425
U	1014	0.0288	28426–29439
M	889	0.0252	29440–30328
Y	783	0.0222	30329–31111
W	716	0.0203	31112–31827
F	629	0.0178	31828–32456
C	584	0.0166	32457–33040
G	478	0.0136	33041–33518
P	433	0.0123	33519–33951
B	410	0.0116	33952–34361
V	309	0.0088	34362–34670
K	255	0.0072	34671–34925
'	203	0.0058	34926–35128
J	34	0.0010	35129–35162
Q	27	0.0008	35163–35189
X	21	0.0006	35190–35210
Z	14	0.0004	35211–35224

```
AOOAAORH ONNNDGELC TEFSISO VTALIDMA POESDHEMHIESWON
PJTOMJ FTL FIM TAOFERLMT O NORDEERH HMFIOMRETWOVRCA
OSRIE IEOBOTOGIM NUDSEEWU WHHS AWUA HIDNEVE NL SELTS
```

FIGURE 2.4.5

by a program knowing only single-letter frequencies (Table 2.4.1). Nowhere does even a single correctly spelled word appear. Contrast that with Figure 2.4.6, showing computer text generated by a program that had been given estimates for conditional probabilities corresponding to all 614,656 (= 28^4) *four*-letter sequences. What we get is still garble, but the improvement is astounding—more than 80% of the letter combinations are at least words.

(Continued on next page)

(Case Study 2.4.1 continued)

The image of a monkey sitting at a typewriter, pecking away at random until he gets lucky and types out a perfect copy of the complete works of William Shakespeare, has long been a favorite model of statisticians and philosophers to illustrate the distinction between something that is theoretically possible but for all practical purposes, impossible. But if that monkey and his typewriter are replaced by a high-technology computer and if we program in the right sorts of conditional probabilities, the prospects for generating *something* intelligible become a little less far-fetched—maybe even disturbingly less far-fetched (11).

Simulating nonnumerical English text requires that twenty-eight characters be dealt with: the twenty-six letters, the space, and the apostrophe. The simplest approach would be to assign each of those characters a number from 1 to 28. Then a random number in that range would be generated and the character corresponding to that number would be printed. A second random number would be generated, a corresponding second character would be printed, and so on.

Would that be a reasonable model? Of course not. Why should, say, X's have the same chance of being selected as E's when we know that the latter are much more common? At the very least, weights should be assigned to all the characters proportional to their relative probabilities. Table 2.4.1 shows the empirical distribution of the twenty-six letters, the space, and the apostrophe in the 35,224 characters making up Act III of *Hamlet*. Ranges of random numbers corresponding to each character's frequency are listed in the last column. If two random numbers were generated, say, 27351 and 11616, the computer would print the characters D and O. Doing that, of course, is equivalent to printing a D with probability $0.0312 = [(28425 - 27327 + 1)/35244 = 1099/35244]$ and an O with probability $0.0732 = [(12789 - 10212 + 1)/35244 = 2578/35244]$.

Extending this idea to *sequences* of letters requires an application of Definition 2.4.1. What is the probability, for example, that a T follows an E? By definition,

$$P(\text{T follows an E}) = P(\text{T}|\text{E}) = \frac{\text{number of ET's}}{\text{number of E's}}$$

The analog of Table 2.4.1, then, would be an array having twenty-eight rows and twenty-eight columns. The entry in the ith row and jth column would be $P(i|j)$, the probability that letter i follows letter j.

In a similar fashion, conditional probabilities for longer sequences could also be estimated. For example, the probability that an A follows the sequence QU would be the ratio of QUA's to QU's:

$$P(\text{A follows QU}) = P(\text{A}|\text{QU}) = \frac{\text{number of QUA's}}{\text{number of QU's}}$$

What does our monkey gain by having a typewriter programmed with probabilities of sequences? Quite a bit. Figure 2.4.5 shows three lines of computer text generated

(Continued on next page)

TABLE 2.4.2

(Prize Location, Door Opened)	Probability
(1, 3)	1/3
(2, 1)	1/6
(2, 3)	1/6
(3, 1)	1/3

Table 2.4.2 shows the sample space associated with the scenario just described. If the prize is actually behind Door #1, the host has no choice but to open Door #3; similarly, if the prize is behind Door #3, the host has no choice but to open Door #1. In the event that the prize is behind Door #2, though, the host would (theoretically) open Door #1 half the time and Door #3 half the time.

Notice that the four outcomes in S are not equally likely. There is necessarily a one-third probability that the prize is behind each of the three doors. However, the two choices that the host has when the prize is behind Door #2 necessitate that the two outcomes (2, 1) and (2, 3) share the one-third probability that represents the chances of the prize being behind Door #2. Each, then, has the one-sixth probability listed in Table 2.4.2.

Let A be the event that the prize is behind Door #2, and let B be the event that the host opened Door #3. Then

$$\begin{aligned} P(A|B) = P(\text{contestant wins by not switching}) &= [P(A \cap B)]/P(B) \\ &= \left[\tfrac{1}{6}\right] \Big/ \left[\tfrac{1}{3} + \tfrac{1}{6}\right] \\ &= \tfrac{1}{3} \end{aligned}$$

Now, let A^* be the event that the prize is behind Door #1, and let B (as before) be the event that the host opens Door #3. In this case,

$$\begin{aligned} P(A^*|B) = P(\text{contestant wins by switching}) &= [P(A^* \cap B)]/P(B) \\ &= \left[\tfrac{1}{3}\right] \Big/ \left[\tfrac{1}{3} + \tfrac{1}{6}\right] \\ &= \tfrac{2}{3} \end{aligned}$$

Common sense would have led us astray again! If given the choice, contestants should *always* switch doors. Doing so ups their chances of winning from one-third to two-thirds.

QUESTIONS

2.4.1. Suppose that two fair dice are tossed. What is the probability that the sum equals ten given that it exceeds eight?

2.4.2. Find $P(A \cap B)$ if $P(A) = 0.2$, $P(B) = 0.4$, and $P(A|B) + P(B|A) = 0.75$.

2.4.3. If $P(A|B) < P(A)$, show that $P(B|A) < P(B)$.

2.4.4. Let A and B be two events such that $P((A \cup B)^C) = 0.6$ and $P(A \cap B) = 0.1$. Let E be the event that either A or B but not both will occur. Find $P(E|A \cup B)$.

2.4.5. Suppose that in Example 2.4.2 we ignored the age of the children and distinguished only *three* family types; (boy, boy), (girl, boy), and (girl, girl). Would the conditional probability of both children being boys given that at least one is a boy be different from the answer found on p. 45? Explain.

2.4.6. Two events, A and B, are defined on a sample space S such that $P(A|B) = 0.6$, P(at least one of the events occurs) $= 0.8$, and P(exactly one of the events occurs) $= 0.6$. Find $P(A)$ and $P(B)$.

2.4.7. An urn contains one red chip and one white chip. One is drawn at random. If the chip selected is red, that chip together with two additional red chips are put back into the urn. If a white is drawn, the chip is returned to the urn. Then a second chip is drawn. What is the probability that both selections are red?

2.4.8. Given that $P(A) = a$ and $P(B) = b$, show that

$$P(A|B) \geq \frac{a + b - 1}{b}$$

2.4.9. An urn contains one white chip and a second chip that is equally likely to be white or black. A chip is drawn at random and returned to the urn. Then a second chip is drawn. What is the probability that a white appears on the second draw given that a white appeared on the first draw?

2.4.10. Suppose events A and B are such that $P(A \cap B) = 0.1$ and $P((A \cup B)^C) = 0.3$. If $P(A) = 0.2$, what does $P[(A \cap B)|(A \cup B)^C]$ equal? Hint: Draw the Venn diagram.

2.4.11. One hundred voters were asked their opinions of two candidates, A and B, running for mayor. Their responses to three questions are summarized below:

	Number Saying Yes
Do you like A?	65
Do you like B?	55
Do you like both?	25

(a) What is the probability that someone likes neither?
(b) What is the probability that someone likes exactly one?
(c) What is the probability that someone likes at least one?
(d) What is the probability that someone likes at most one?
(e) What is the probability that someone likes exactly one given that they like at least one?
(f) Of those who like at least one, what proportion like both?
(g) Of those who do not like A, what proportion like B?

2.4.12. A fair coin is tossed three times. What is the probability that at least two heads will occur given that at most two heads have occurred?

2.4.13. Two fair dice are rolled. What is the probability that the number on the first die was at least as large as 4 given that the sum of the two dice was eight?

2.4.14. Four cards are dealt from a standard 52-card poker deck. What is the probability that all four are aces given that at least three are aces? Note: There are 270, 725 different sets of four cards that can be dealt. Assume that the probability associated with each of those hands is $1/270, 725$.

2.4.15. Given that $P(A \cap B^C) = 0.3$, $P((A \cup B)^C) = 0.2$, and $P(A \cap B) = 0.1$, find $P(A|B)$.

2.4.16. Given that $P(A) + P(B) = 0.9$, $P(A|B) = 0.5$, and $P(B|A) = 0.4$, find $P(A)$.

2.4.17. Let A and B be two events defined on a sample space S such that $P(A \cap B^C) = 0.1$, $P(A^C \cap B) = 0.3$, and $P((A \cup B)^C) = 0.2$. Find the probability that at least one of the two events occurs given that at most one occurs.

2.4.18. Suppose two dice are rolled. Assume that each possible outcome has probability $1/36$. Let A be the event that the sum of the two dice is greater than or equal to eight, and let B be the event that at least one of the dice is a 5. Find $P(A|B)$.

2.4.19. According to your neighborhood bookie, there are five horses scheduled to run in the third race at the local track, and handicappers have assigned them the following probabilities of winning:

Horse	Probability of Winning
Scorpion	0.10
Starry Avenger	0.25
Australian Doll	0.15
Dusty Stake	0.30
Outandout	0.20

Suppose that Australian Doll and Dusty Stake are scratched from the race at the last minute. What are the chances that Outandout will prevail over the reduced field?

2.4.20. Andy, Bob, and Charley have all been serving time for grand theft auto. According to prison scuttlebutt, the warden plans to release two of the three next week. They all have identical records, so the two to be released will be chosen at random, meaning that each has a two-third probability of being included in the two to be set free. Andy, however, is friends with a guard who will know ahead of time which two will leave. He offers to tell Andy the name of a prisoner *other than himself* who will be released. Andy, however, declines the offer, believing that if he learns the name of a prisoner scheduled to be released, then *his* chances of being the other person set free will drop to one-half (since only two prisoners will be left at that point). Is his concern justified?

Applying Conditional Probability to Higher-Order Intersections

We have seen that conditional probabilities can be useful in evaluating intersection probabilities—that is, $P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$. A similar result holds for higher-order intersections. Consider $P(A \cap B \cap C)$. By thinking of $A \cap B$ as a single event—say, D—we can write

$$\begin{aligned} P(A \cap B \cap C) &= P(D \cap C) \\ &= P(C|D)P(D) \\ &= P(C|A \cap B)P(A \cap B) \\ &= P(C|A \cap B)P(B|A)P(A) \end{aligned}$$

Repeating this same argument for n events, $A_1, A_2, \ldots, A_n$, gives a formula for the general case:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_n | A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \cdot P(A_{n-1} | A_1 \cap A_2 \cap \cdots \cap A_{n-2}) \cdot \cdots \cdot P(A_2 | A_1) \cdot P(A_1) \tag{2.4.3}$$

EXAMPLE 2.4.6

An urn contains five white chips, four black chips, and three red chips. Four chips are drawn sequentially and without replacement. What is the probability of obtaining the sequence (white, red, white, black)?

Figure 2.4.7. shows the evolution of the urn's composition as the desired sequence is assembled. Define the following four events:

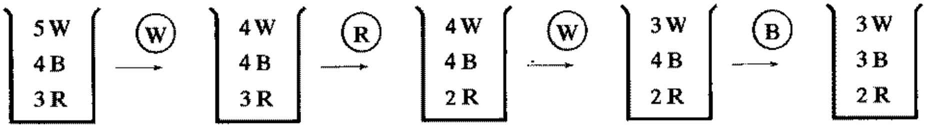


FIGURE 2.4.7

A: white chip is drawn on first selection
B: red chip is drawn on second selection
C: white chip is drawn on third selection
D: black chip is drawn on fourth selection

Our objective is to find $P(A \cap B \cap C \cap D)$.

From Equation 2.4.3,

$$P(A \cap B \cap C \cap D) = P(D|A \cap B \cap C) \cdot P(C|A \cap B) \cdot P(B|A) \cdot P(A)$$

Each of the probabilities on the right-hand side of the equation here can be gotten by just looking at the urns pictured in Figure 2.4.7: $P(D|A \cap B \cap C) = \frac{4}{9}$, $P(C|A \cap B) = \frac{4}{10}$, $P(B|A) = \frac{3}{11}$, and $P(A) = \frac{5}{12}$. Therefore, the probability of drawing a (white, red, white, black) sequence is 0.02:

$$\begin{aligned} P(A \cap B \cap C \cap D) &= \frac{4}{9} \cdot \frac{4}{10} \cdot \frac{3}{11} \cdot \frac{5}{12} \\ &= \frac{240}{11{,}880} \\ &= 0.02 \end{aligned}$$

CASE STUDY 2.4.3

Since the late 1940s, tens of thousands of eyewitness accounts of strange lights in the skies, unidentified flying objects, even alleged abductions by little green men, have made headlines. None of these incidents, though, has produced any hard evidence, any irrefutable *proof* that Earth has been visited by a race of extraterrestrials. Still, the haunting question remains—are we alone in the universe? Or are there other civilizations, more advanced than ours, making the occasional flyby?

Until, or unless, a flying saucer plops down on the White House lawn and a strange-looking creature emerges with the proverbial "Take me to your leader" demand, we may never know whether we have any cosmic neighbors. Equation 2.4.3, though, can help us speculate on the *probability* of our not being alone.

Recent discoveries suggest that planetary systems much like our own may be quite common. If so, there are likely to be many planets whose chemical makeups, temperatures, pressures, and so on, are suitable for life. Let those planets be the points in our sample space. Relative to them, we can define three events:

A: life arises

B: technical civilization arises (one capable of interstellar communication)

C: technical civilization is flourishing *now*

In terms of A, B, and C, the probability a habitable planet is presently supporting a technical civilization is the probability of an intersection—specifically, $P(A \cap B \cap C)$. Associating a number with $P(A \cap B \cap C)$ is highly problematic, but the task is simplified considerably if we work instead with the equivalent conditional formula, $P(C|B \cap A) \cdot P(B|A) \cdot P(A)$.

Scientists speculate (157) that life of some kind may arise on one-third of all planets having a suitable environment and that life on maybe 1% of all those planets will evolve into a technical civilization. In our notation, $P(A) = \frac{1}{3}$ and $P(B|A) = \frac{1}{100}$.

More difficult to estimate is $P(C|A \cap B)$. On Earth, we have had the capability of interstellar communication (that is, radio astronomy) for only a few decades, so $P(C|A \cap B)$, *empirically*, is on the order of 1×10^{-8}. But that may be an overly pessimistic estimate of a technical civilization's ability to endure. It may be true that if a civilization can avoid annihilating itself when it first develops nuclear weapons, its prospects for longevity are fairly good. If that were the case, $P(C|A \cap B)$ might be as large as 1×10^{-2}.

Putting these estimates into the computing formula for $P(A \cap B \cap C)$ yields a range for the probability of a habitable planet currently supporting a technical civilization. The chances may be as small as 3.3×10^{-11} or as "large" as 3.3×10^{-5}:

$$(1 \times 10^{-8})\left(\frac{1}{100}\right)\left(\frac{1}{3}\right) < P(A \cap B \cap C) < (1 \times 10^{-2})\left(\frac{1}{100}\right)\left(\frac{1}{3}\right)$$

or

$$0.000000000033 < P(A \cap B \cap C) < 0.000033$$

(Continued on next page)

(Case Study 2.4.3 continued)

A better way to put these figures in some kind of perspective is to think in terms of *numbers* rather than probabilities. Astronomers estimate there are 3×10^{11} habitable planets in our Milky Way galaxy. Multiplying that total by the two limits for $P(A \cap B \cap C)$ gives an indication of *how many* cosmic neighbors we are likely to have. Specifically, $3 \times 10^{11} \cdot 0.000000000033 \doteq 10$, while $3 \times 10^{11} \cdot 0.000033 \doteq 10{,}000{,}000$. So, on the one hand, we may be a galactic rarity. At the same time, the probabilities do not preclude the very real possibility that the heavens are abuzz with activity and that our neighbors number in the millions.

QUESTIONS

2.4.21. An urn contains six white chips, four black chips, and five red chips. Five chips are drawn out, one at a time and without replacement. What is the probability of getting the sequence (black, black, red, white, white)? Suppose that the chips are numbered 1 through 15. What is the probability of getting a specific sequence—say, (2, 6, 4, 9, 13)?

2.4.22. A man has n keys on a key ring, one of which opens the door to his apartment. Having celebrated a bit too much one evening, he returns home only to find himself unable to distinguish one key from another. Resourceful, he works out a fiendishly clever plan: He will choose a key at random and try it. If it fails to open the door, he will discard it and choose at random one of the remaining $n - 1$ keys, and so on. Clearly, the probability that he gains entrance with the first key he selects is $1/n$. Show that the probability the door opens with the *third* key he tries is also $1/n$. (*Hint:* What has to happen before he even gets to the third key?)

2.4.23. Suppose that four cards are drawn from a standard 52-card poker deck. What is the probability of drawing, in order, a 7 of diamonds, a jack of spades, a 10 of diamonds, and a 5 of hearts?

2.4.24. One chip is drawn at random from an urn that contains one white chip and one black chip. If the white chip is selected, we simply return it to the urn; if the black chip is drawn, that chip—together with another black—are returned to the urn. Then a second chip is drawn, with the same rules for returning it to the urn. Calculate the probability of drawing two whites followed by three blacks.

Calculating "Unconditional" Probabilities

We conclude this section with two very useful theorems that apply to *partitioned* sample spaces. By definition, a set of events $A_1, A_2, \ldots, A_n$ "partition" S if every outcome in

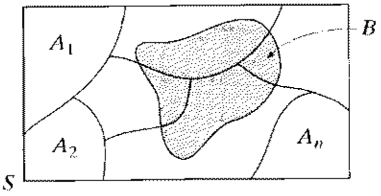


FIGURE 2.4.8

the sample space belongs to one and only one of the A_i's—that is, the A_i's are mutually exclusive and their union is S (see Figure 2.4.8).

Let B, as pictured, denote any event defined on S. The first result, Theorem 2.4.1, gives a formula for the "unconditional" probability of B (in terms of the A_i's). Then Theorem 2.4.2 calculates the set of conditional probabilities, $P(A_j|B)$, $j = 1, 2, \ldots, n$.

Theorem 2.4.1. *Let $\{A_i\}_{i=1}^n$ be a set of events defined over S such that $S = \bigcup_{i=1}^n A_i$, $A_i \cap A_j = \emptyset$ for $i \neq j$, and $P(A_i) > 0$ for $i = 1, 2, \ldots, n$. For any event B,*

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

Proof. By the conditions imposed on the A_i's,

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n)$$

and

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

But each $P(B \cap A_i)$ can be written as the product $P(B|A_i)P(A_i)$, and the result follows. □

EXAMPLE 2.4.7

Urn I contains two red chips and four white chips; urn II, three red and one white. A chip is drawn at random from urn I and transferred to urn II. Then a chip is drawn from urn II. What is the probability that the chip drawn from urn II is red?

Let B be the event "Chip drawn from urn II is red"; let A_1 and A_2 be the events "Chip transferred from urn I is red" and "Chip transferred from urn I is white," respectively. By inspection (see Figure 2.4.9), we can deduce all the probabilities appearing in the right-hand side of the formula in Theorem 2.4.1:

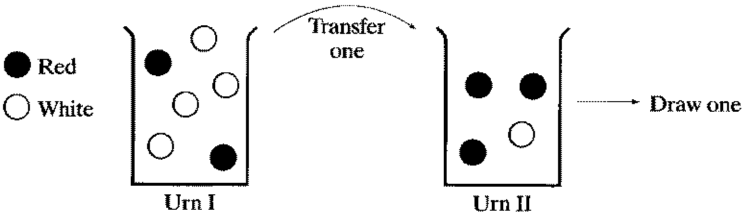


FIGURE 2.4.9

$$P(B|A_1) = \frac{4}{5} \quad P(B|A_2) = \frac{3}{5}$$

$$P(A_1) = \frac{2}{6} \quad P(A_2) = \frac{4}{6}$$

Putting all this information together, we see that the chances are two out of three that a red chip will be drawn from urn II:

$$\begin{aligned}P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2)\\ &= \frac{4}{5}\cdot\frac{2}{6}+\frac{3}{5}\cdot\frac{4}{6}\\ &= \frac{2}{3}\end{aligned}$$

EXAMPLE 2.4.8

A standard poker deck is shuffled and the card on top is removed. What is the probability that the *second* card is an ace?

Define the following events:

B: second card is an ace
A_1: top card was an ace
A_2: top card was not an ace

Then $P(B|A_1) = \frac{3}{51}$, $P(B|A_2) = \frac{4}{51}$, $P(A_1) = \frac{4}{52}$, and $P(A_2) = \frac{48}{52}$. Since the A_i's partition the sample space of two-card selections, Theorem 2.4.1 applies. Substituting into the expression for $P(B)$ shows that $\frac{4}{52}$ is the probability that the second card is an ace:

$$\begin{aligned}P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2)\\ &= \frac{3}{51}\cdot\frac{4}{52}+\frac{4}{51}\cdot\frac{48}{52}\\ &= \frac{4}{52}\end{aligned}$$

Comment. Notice that $P(B) = P(\text{2nd card is an ace})$ is numerically the same as $P(A_1) = P(\text{first card is an ace})$. The analysis in Example 2.4.8 illustrates a basic principle in probability that says, in effect, "what you don't know, doesn't matter." Here, removal of the top card is irrelevant to any subsequent probability calculations *if the identity of that card remains unknown.*

EXAMPLE 2.4.9

Ashley is hoping to land a summer internship with a public relations firm. If her interview goes well, she has a 70% chance of getting an offer. If the interview is a bust, though, her chances of getting the position drop to 20%. Unfortunately, Ashley tends to babble incoherently when she is under stress, so the likelihood of the interview going well is only 0.10. What is the probability that Ashley gets the internship?

Let B be the event "Ashley is offered internship," let A_1 be the event "Interview goes well," and let A_2 be the event "Interview does not go well." By assumption,

$$\begin{aligned}P(B|A_1) &= 0.70 &\quad P(B|A_2) &= 0.20\\ P(A_1) &= 0.10 &\quad P(A_2) &= 1 - P(A_1) = 1 - 0.10 = 0.90\end{aligned}$$

According to Theorem 2.4.1, Ashley has a 25% chance of landing the internship:

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) \\ &= (0.70)(0.10) + (0.20)(0.90) \\ &= 0.25 \end{aligned}$$

EXAMPLE 2.4.10

In an upstate congressional race, the incumbent Republican (R) is running against a field of three Democrats (D_1, D_2, and D_3) seeking the nomination. Political pundits estimate that the probabilities of D_1, D_2, and D_3 winning the primary are 0.35, 0.40, and 0.25, respectively. Furthermore, results from a variety of polls are suggesting that R would have a 40% chance of defeating D_1 in the general election, a 35% chance of defeating D_2, and a 60% chance of defeating D_3. Assuming all these estimates to be accurate, what are the chances that the Republican will retain his seat?

Let B denote the event that "R wins general election," and let A_i denote the event "D_i wins Democratic primary"; $i = 1, 2, 3$. Then

$$P(A_1) = 0.35 \qquad P(A_2) = 0.40 \qquad P(A_3) = 0.25$$

and

$$P(B|A_1) = 0.40 \qquad P(B|A_2) = 0.35 \qquad P(B|A_3) = 0.60$$

so

$$\begin{aligned} P(B) &= P(\text{Republican wins general election}) \\ &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ &= (0.40)(0.35) + (0.35)(0.40) + (0.60)(0.25) \\ &= 0.43 \end{aligned}$$

EXAMPLE 2.4.11

Three chips are placed in an urn. One is red on both sides, a second is blue on both sides, and the third is red on one side and blue on the other. One chip is selected at random and placed on a table. Suppose that the color showing on that chip is red. What is the probability that the color underneath is also red (see Figure 2.4.10)?

At first glance, it may seem that the answer is one-half: We know that the blue/blue chip has not been drawn, and only one of the remaining two—the red/red chip—satisfies the event that the color underneath is red. If this game were played over and over, though, and records were kept of the outcomes, it would be found that the proportion of times that a red top has a red bottom is two-thirds, not the one-half that our intuition might suggest. The correct answer follows from an application of Theorem 2.4.1.

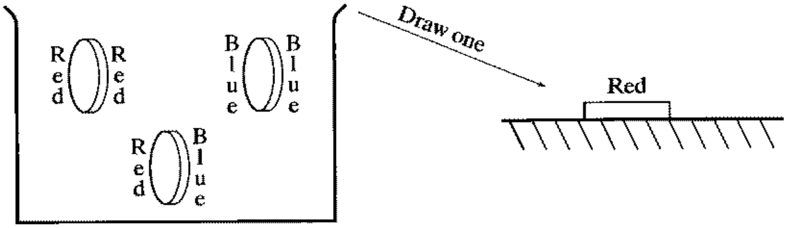


FIGURE 2.4.10

Define the following events:

A: bottom side of chip drawn is red
B: top side of chip drawn is red
A_1: red/red chip is drawn
A_2: blue/blue chip is drawn
A_3: red/blue chip is drawn

From the definition of conditional probability,

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

But $P(A \cap B) = P(\text{both sides are red}) = P(\text{red/red chip}) = \frac{1}{3}$. Theorem 2.4.1 can be used to find the denominator, $P(B)$:

$$\begin{aligned} P(B) &= P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3) \\ &= 1 \cdot \frac{1}{3} + 0 \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{2} \end{aligned}$$

Therefore,

$$P(A|B) = \frac{1/3}{1/2} = \frac{2}{3}$$

Comment. The question posed in Example 2.4.11 gives rise to a simple but effective con game. The trick is to convince a "mark" that the initial analysis given on page 59 is correct, meaning that the bottom has a fifty-fifty chance of being the same color as the top. Under that incorrect presumption that the game is "fair," both participants put up the same amount of money, but the gambler (knowing the correct analysis) always bets that the bottom is the same color as the top. In the long run, then, the con artist will be winning an even-money bet two-thirds of the time!

QUESTIONS

2.4.25. A toy manufacturer buys ball bearings from three different suppliers—50% of her total order comes from supplier 1, 30% from supplier 2, and the rest from supplier 3. Past experience has shown that the quality control standards of the three suppliers are not all the same. Two percent of the ball bearings produced by supplier 1 are defective, while suppliers 2 and 3 produce defective bearings 3% and 4% of the time, respectively. What proportion of the ball bearings in the toy manufacturer's inventory are defective?

2.4.26. A fair coin is tossed. If a head turns up, a fair die is tossed; if a tail turns up, two fair dice are tossed. What is the probability that the face (or the sum of the faces) showing on the die (or the dice) is equal to six?

2.4.27. Foreign policy experts estimate that the probability is 0.65 that war will break out next year between two Middle East countries if either side significantly escalates its terrorist activities: Otherwise, the likelihood of war is estimated to be 0.05. Based on what has happened this year, the chances of terrorism reaching a critical level in the next twelve months are thought to be three in ten. What is the probability that the two countries will go to war?

2.4.28. A telephone solicitor is responsible for canvassing three suburbs. In the past, 60% of the completed calls to Belle Meade have resulted in contributions, compared to 55% for Oak Hill and 35% for Antioch. Her list of telephone numbers includes one thousand households from Belle Meade, one thousand from Oak Hill, and two thousand from Antioch. Suppose that she picks a number at random from the list and places the call. What is the probability that she gets a donation?

2.4.29. If men constitute 47% of the population and tell the truth 78% of the time, while women tell the truth 63% of the time, what is the probability that a person selected at random will answer a question truthfully?

2.4.30. Urn I contains three red chips and one white chip. Urn II contains two red chips and two white chips. One chip is drawn from each urn and transferred to the other urn. Then a chip is drawn from the first urn. What is the probability that the chip ultimately drawn from urn I is red?

2.4.31. The crew of the Starship *Enterprise* is considering launching a surprise attack against the Borg in a neutral quadrant. Possible interference by the Klingons, though, is causing Captain Picard and Data to reassess their strategy. According to Data's calculations, the probability of the Klingons joining forces with the Borg is 0.2384. Captain Picard feels that the probability of the attack being successful is 0.8 if the *Enterprise* can catch the Borg alone, but only 0.3 if they have to engage both adversaries. Data claims that the mission would be a tactical misadventure if its probability of success were not at least 0.7306. Should the *Enterprise* attack?

2.4.32. Recall the "survival" lottery described in Question 2.2.14. What is the probability of release associated with the prisoner's optimal strategy?

2.4.33. State College is playing Backwater A&M for the conference football championship. If Backwater's first-string quarterback is healthy. A&M has a 75% chance of winning. If they have to start their backup quarterback, their chances of winning drop to 40%. The team physician says that there is a 70% chance that the first-string quarterback will play. What is the probability that Backwater wins the game?

2.4.34. An urn contains forty red chips and sixty white chips. Six chips are drawn out and discarded, and a seventh chip is drawn. What is the probability that the seventh chip is red?

2.4.35. A study has shown that seven out of ten people will say "heads" if asked to call a coin toss. Given that the coin is fair, though, a head occurs, on the average, only five times out of ten. Does it follow that you have the advantage if you let the other person call the toss? Explain.

2.4.36. Based on pretrial speculation, the probability that a jury returns a guilty verdict in a certain high-profile murder case is thought to be 15% if the defense can discredit the police department and 80% if they cannot. Veteran court observers believe that the skilled defense attorneys have a 70% chance of convincing the jury that the police either contaminated or planted some of the key evidence. What is the probability that the jury returns a guilty verdict?

2.4.37. As an incoming freshman, Marcus believes that he has a 25% chance of earning a GPA in the 3.5 to 4.0 range, a 35% chance of graduating with a 3.0 to 3.5 GPA, and a 40% chance of finishing with a GPA less than 3.0. From what the pre-med advisor has told him, he has an 8 in 10 chance of getting into medical school if his GPA is above 3.5, a 5 in 10 chance if his GPA is in the 3.0 to 3.5 range, and only a 1 in 10 chance if his GPA falls below 3.0. Based on those estimates, what is the probability that Marcus gets into medical school?

2.4.38. The governor of a certain state has decided to come out strongly for prison reform and is preparing a new early-release program. Its guidelines are simple: prisoners related to members of the governor's staff would have a 90% chance of being released early; the probability of early release for inmates not related to the governor's staff would be 0.01. Suppose that 40% of all inmates are related to someone on the governor's staff. What is the probability that a prisoner selected at random would be eligible for early release?

2.4.39. Following are the percentages of students of State College enrolled in each of the school's main divisions. Also listed are the proportions of students in each division who are women.

Division	%	% Women
Humanities	40	60
Natural Science	10	15
History	30	45
Social Science	20	75
	100	

Suppose the Registrar selects one person at random. What is the probability that the student selected will be a male?

Bayes Theorem

The second result in this section that is set against the backdrop of a partitioned sample space has a curious history. The first explicit statement of Theorem 2.4.2, coming in 1812, was due to Laplace, but it was named after the Reverend Thomas Bayes, whose 1763 paper (published posthumously) had already outlined the result. On one level, the theorem is a relatively minor extension of the definition of conditional probability. When viewed from a loftier perspective, though, it takes on some rather profound philosophical implications. The latter, in fact, have precipitated a schism among practicing statisticians: "Bayesians" analyze data one way; "non-Bayesians" often take a fundamentally different approach (see Section 5.8).

Our use of the result here will have nothing to do with its statistical interpretation. We will apply it simply as the Reverend Bayes originally intended, as a formula for evaluating a certain kind of "inverse" probability. If we know $P(B|A_i)$ for all i, the theorem enables us to compute conditional probabilities "in the other direction"—that is, we can deduce $P(A_j|B)$ from the $P(B|A_i)$'s.

Theorem 2.4.2. *(Bayes) Let* $\{A_i\}_{i=1}^n$ *be a set of* n *events, each with positive probability, that partition* S *in such a way that* $\cup_{i=1}^n A_i = S$ *and* $A_i \cap A_j = \emptyset$ *for* $i \neq j$. *For any event* B *(also defined on* S*), where* $P(B) > 0$,

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^{n} P(B|A_i)P(A_i)}$$

for any $1 \leq j \leq n$.

Proof. From Definition 2.4.1,

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

But Theorem 2.4.1 allows the denominator to be written as $\sum_{i=1}^{n} P(B|A_i)P(A_i)$, and the result follows. □

PROBLEM-SOLVING HINTS
(Working with Partitioned Sample Spaces)

Students sometimes have difficulty setting up problems that involve partitioned sample spaces—in particular, ones whose solution requires an application of either Theorem 2.4.1 or 2.4.2—because of the nature and amount of information that needs to be incorporated into the answers. The "trick" is learning to identify which part of the "given" corresponds to B and which parts correspond to the A_i's. The following hints may help.

1. As you read the question, pay particular attention to the last one or two sentences. Is the problem asking for an *unconditional probability* (in which case Theorem 2.4.1 applies) or a *conditional probability* (in which case Theorem 2.4.2 applies)?
2. If the question is asking for an unconditional probability, let B denote the event whose probability you are trying to find; if the question is asking for a conditional probability, let B denote the event that has *already happened.*
3. Once event B has been identified, reread the beginning of the question and assign the A_i's.

EXAMPLE 2.4.12

A biased coin, twice as likely to come up heads as tails, is tossed once. If it shows heads, a chip is drawn from urn I, which contains three white chips and four red chips; if it shows

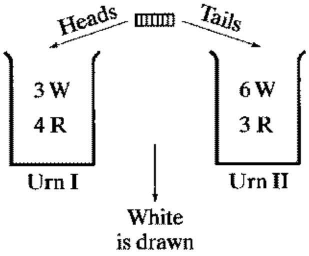


FIGURE 2.4.11

tails, a chip is drawn from urn II, which contains six white chips and three red chips. Given that a white chip was drawn, what is the probability that the coin came up tails (see Figure 2.4.11)?

Since $P(\text{Heads}) = 2P(\text{Tails})$, it must be true that $P(\text{Heads}) = \frac{2}{3}$ and $P(\text{Tails}) = \frac{1}{3}$. Define the events

B: white chip is drawn

A_1: coin came up heads (i.e., chip came from urn I)

A_2: coin came up tails (i.e., chip came from urn II)

Our objective is to find $P(A_2|B)$. From Figure 2.4.11,

$$P(B|A_1) = \frac{3}{7} \quad P(B|A_2) = \frac{6}{9}$$
$$P(A_1) = \frac{2}{3} \quad P(A_2) = \frac{1}{3}$$

so

$$\begin{aligned} P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{(6/9)(1/3)}{(3/7)(2/3) + (6/9)(1/3)} \\ &= \frac{7}{16} \end{aligned}$$

EXAMPLE 2.4.13

During a power blackout, one hundred persons are arrested on suspicion of looting. Each is given a polygraph test. From past experience it is known that the polygraph is 90% reliable when administered to a guilty suspect and 98% reliable when given to someone who is innocent. Suppose that of the one hundred persons taken into custody, only twelve were actually involved in any wrongdoing. What is the probability that a given suspect is innocent given that the polygraph says he is guilty?

Let B be the event "Polygraph says suspect is guilty," and let A_1 and A_2 be the events "Suspect is guilty" and "Suspect is not guilty," respectively. To say that the polygraph is "90% reliable when administered to a guilty suspect" means that $P(B|A_1) = 0.90$. Similarly, the 98% reliability for innocent suspects implies that $P(B^C|A_2) = 0.98$, or, equivalently, $P(B|A_2) = 0.02$.

We also know that $P(A_1) = \frac{12}{100}$ and $P(A_2) = \frac{88}{100}$. Substituting into Theorem 2.4.2, then, shows that the probability a suspect is innocent given that the polygraph says he is guilty is 0.14:

$$\begin{aligned} P(A_2|B) &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{(0.02)(88/100)}{(0.90)(12/100) + (0.02)(88/100)} \\ &= 0.14 \end{aligned}$$

EXAMPLE 2.4.14

As medical technology advances and adults become more health conscious, the demand for diagnostic screening tests inevitably increases. Looking for problems, though, when no symptoms are present can have undesirable consequences that may outweigh the intended benefits.

Suppose, for example, a woman has a medical procedure performed to see whether she has a certain type of cancer. Let B denote the event that the test *says* she has cancer, and let A_1 denote the event that she actually *does* (and A_2, the event that she *does not*). Furthermore, suppose the prevalence of the disease and the precision of the diagnostic test are such that

$$P(A_1) = 0.0001 \qquad [\text{and} \quad P(A_2) = 0.9999]$$

$$P(B|A_1) = 0.90 = P(\text{test } \textit{says} \text{ woman has cancer when, in fact, she does})$$

$$P(B|A_2) = P\left(B|A_1^C\right) = 0.001 = P(\text{false positive}) = P(\text{test } \textit{says} \text{ woman has cancer when, in fact, she does not})$$

What is the probability that she *does* have cancer, given that the diagnostic procedure says she does? That is, calculate $P(A_1|B)$.

Although the method of solution here is straightforward, the actual numerical answer is not what we would expect. From Theorem 2.4.2,

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P\left(B|A_1^C\right)P\left(A_1^C\right)} \\ &= \frac{(0.9)(0.0001)}{(0.9)(0.0001) + (0.001)(0.9999)} \\ &= 0.08 \end{aligned}$$

So, only 8% of those women identified as having cancer actually do! Table 2.4.3 shows the strong dependence of $P(A_1|B)$ on $P(A_1)$ and $P(B|A_1^C)$.

TABLE 2.4.3

$P(A_1)$	$P(B\|A_1^C)$	$P(A_1\|B)$
0.0001	0.001	0.08
	0.0001	0.47
0.001	0.001	0.47
	0.0001	0.90
0.01	0.001	0.90
	0.0001	0.99

In light of these probabilities, the practicality of screening programs directed at diseases having a low prevalence is open to question, especially when the diagnostic procedure, itself, poses a nontrivial health risk. (For precisely those two reasons, the use of chest X-rays to screen for tuberculosis is no longer advocated by the medical community.)

EXAMPLE 2.4.15

According to the manufacturer's specifications, your home burglar alarm has a 95% chance of going off if someone breaks into your house. During the two years you have lived there, the alarm went off on five different nights, each time for no apparent reason. Suppose the alarm goes off tomorrow night. What is the probability that someone is trying to break into your house? Note: Police statistics show that the chances of any particular house in your neighborhood being burglarized on any given night are two in ten thousand.

Let B be the event "Alarm goes off tomorrow night," and let A_1 and A_2 be the events "House is being burglarized" and "House is not being burglarized," respectively. Then

$$\begin{aligned} P(B|A_1) &= 0.95 \\ P(B|A_2) &= 5/730 \quad \text{(i.e., five nights in two years)} \\ P(A_1) &= 2/10{,}000 \\ P(A_2) &= 1 - P(A_1) = 9{,}998/10{,}000 \end{aligned}$$

The probability in question is $P(A_1|B)$.

Intuitively, it might seem that $P(A_1|B)$ should be close to one because the alarm's "performance" probabilities look good—$P(B|A_1)$ is close to one (as it should be) and $P(B|A_2)$ is close to zero (as it should be). Nevertheless, $P(A_1|B)$ turns out to be surprisingly small:

$$\begin{aligned} P(A_1|B) &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \\ &= \frac{(0.95)(2/10{,}000)}{(0.95)(2/10{,}000) + (5/730)(9998/10{,}000)} \\ &= 0.027 \end{aligned}$$

That is, if you hear the alarm going off, the probability is only 0.027 that the house is being burglarized.

Computationally, the reason $P(A_1|B)$ is so small is that $P(A_2)$ is so large. The latter makes the denominator of $P(A_1|B)$ large and, in effect, "washes out" the numerator. Even if $P(B|A_1)$ were substantially increased (by installing a more expensive alarm), $P(A_1|B)$ would remain largely unchanged (see Table 2.4.4).

TABLE 2.4.4

	$P(B\|A_1)$			
	0.95	0.97	0.99	0.999
$P(A_1\|B)$	0.027	0.028	0.028	0.028

EXAMPLE 2.4.16

Currently a college senior, Jeremy has had a secret crush on Emma ever since the third grade. Two weeks ago, fearing that his feelings would forever go unrequited, he broke his silence and sent Emma a letter through Campus Mail, acknowledging his twelve-year secret romance. Now, fourteen agonizing days later, he has yet to receive a response. Hoping against hope, Jeremy and his fragile psyche are clinging to the possibility that someone's letter was lost in the mail. Assuming that (1) Emma (who is actually secretly dating Jeremy's father) has a 70% chance of mailing a response if, in fact, she had received the letter and (2) the Campus Post Office has a one in fifty chance of losing any particular piece of mail, what is the probability that Emma never received Jeremy's confession of the heart?

Let B represent the event that Jeremy did not receive a response; let A_1 and A_2 denote the events that Emma did and did not, respectively, receive Jeremy's letter. The objective is to find $P(A_2|B)$.

From what we know about Emma's behavior and the incompetence of the Campus Post Office, $P(A_1) = \frac{49}{50}$, $P(A_2) = \frac{1}{50}$, and, of course, $P(B|A_2) = 1$. Also,

$$\begin{aligned} P(B|A_1) &= P(\text{Jeremy receives no response} \mid \text{Emma received Jeremy's letter}) \\ &= P[\text{Emma does not respond} \cup (\text{Emma responds} \cap \text{Post Office loses letter})] \\ &= P(\text{Emma does not respond}) + P(\text{letter is lost} \mid \text{Emma responds}) \\ &\quad \times P(\text{Emma responds}) \\ &= 0.30 + (1/50)(0.70) \\ &= 0.314 \end{aligned}$$

Therefore,

$$\begin{aligned} P(A_2|B) &= \frac{1(1/50)}{0.314(49/50) + 1(1/50)} \\ &= 0.061 \end{aligned}$$

Sadly, the magnitude of $P(A_2|B)$ is not good news for Jeremy. If $P(A_2|B) = 0.061$, it follows that Emma's probability of having received the letter but not caring enough to respond was almost 94%. "Faint heart ne'er won fair lady," but Jeremy would probably be well-advised to direct his romantic intentions elsewhere.

QUESTIONS

2.4.40. Urn I contains two white chips and one red chip; urn II has one white chip and two red chips. One chip is drawn at random from urn I and transferred to urn II. Then one chip is drawn from urn II. Suppose that a red chip is selected from urn II. What is the probability that the chip transferred was white?

2.4.41. Urn I contains three red chips and five white chips; Urn II contains four reds and four whites; Urn III contains five reds and three whites. One urn is chosen at random and one chip is drawn from that urn. Given that the chip drawn was red, what is the probability that III was the urn sampled?

2.4.42. A dashboard warning light is supposed to flash red if a car's oil pressure is too low. On a certain model, the probability of the light flashing when it should is 0.99; 2% of the time, though, it flashes for no apparent reason. If there is a 10% chance that the oil pressure really is low, what is the probability that a driver needs to be concerned if the warning light goes on?

2.4.43. Building permits were issued last year to three contractors starting up a new subdivision: Tara Construction built two houses; Westview, three houses; and Hearthstone, six houses. Tara's houses have a 60% probability of developing leaky basements; homes built by Westview and Hearthstone have that same problem 50% of the time and 40% of the time, respectively. Yesterday, the Better Business Bureau received a complaint from one of the new homeowners that his basement is leaking. Who is most likely to have been the contractor?

2.4.44. Two sections of a senior probability course are being taught. From what she has heard about the two instructors listed, Francesca estimates that her chances of passing the course are 0.85 if she gets professor X and 0.60 if she gets professor Y. The section into which she is put is determined by the registrar. Suppose that her chances of being assigned to professor X are four out of ten. Fifteen weeks later we learn that Francesca did, indeed, pass the course. What is the probability she was enrolled in professor X's section?

2.4.45. A liquor store owner is willing to cash personal checks for amounts up to $50, but she has become wary of customers who wear sunglasses. Fifty percent of checks written by persons wearing sunglasses bounce. In contrast, 98% of the checks written by persons not wearing sunglasses clear the bank. She estimates that 10% of her customers wear sunglasses. If the bank returns a check and marks it "insufficient funds," what is the probability it was written by someone wearing sunglasses?

2.4.46. Brett and Margo have each thought about murdering their rich Uncle Basil in hopes of claiming their inheritance a bit early. Hoping to take advantage of Basil's predilection for immoderate desserts, Brett has put rat poison in the cherries flambe; Margo, unaware of Brett's activities, has laced the chocolate mousse with cyanide. Given the amounts likely to be eaten, the probability of the rat poison being fatal is 0.60; the cyanide, 0.90. Based on other dinners where Basil was presented with the same dessert options, we can assume that he has a 50% chance of asking for the cherries flambe, a 40% chance of ordering the chocolate mousse, and a 10% chance of skipping dessert altogether. No sooner are the dishes cleared away when Basil drops dead. In the absence of any other evidence, who should be considered the prime suspect?

2.4.47. Josh takes a twenty-question multiple-choice exam where each question has five answers. Some of the answers he knows, while others he gets right just by making lucky guesses. Suppose that the conditional probability of his knowing the answer to a randomly selected question given that he got it right is 0.92. How many of the twenty questions was he prepared for?

2.4.48. Recently the U.S. Senate Committee on Labor and Public Welfare investigated the feasibility of setting up a national screening program to detect child abuse. A team of consultants estimated the following probabilities: (1) one child in ninety is abused, (2) a physician can detect an abused child 90% of the time, and (3) a screening program would incorrectly label 3% of all nonabused children as abused. What is the probability that a child is actually abused given that the screening program makes that diagnosis? How does the probability change if the incidence of abuse is one in one thousand? Or one in fifty?

2.4.49. At State University, 30% of the students are majoring in Humanities, 50% in History and Culture, and 20% in Science. Moreover, according to figures released by the Registrar, the percentages of women majoring in Humanities, History and Culture, and Science are 75%, 45%, and 30%, respectively. Suppose Justin meets Anna at a fraternity party. What is the probability that Anna is a History and Culture major?

2.4.50. An "eyes-only" diplomatic message is to be transmitted as a binary code of 0s and 1s. Past experience with the equipment being used suggests that if a 0 is sent, it will be (correctly) received as a 0 90% of the time (and mistakenly decoded as a 1 10% of the time). If a 1 is sent, it will be received as a 1 95% of the time (and as a 0 5% of the time). The text being sent is thought to be 70% 1s and 30% 0s. Suppose the next signal sent is received as a 1. What is the probability that it was sent as a 0?

2.4.51. When Zach wants to contact his girlfriend and he knows she is not at home, he is twice as likely to send her an e-mail as he is to leave a message on her answering machine. The probability that she responds to his e-mail within three hours is 80%; her chances of being similarly prompt in answering a phone message increase to 90%. Suppose she responded to the message he left this morning within two hours. What is the probability that Zach was communicating with her via e-mail?

2.4.52. A dot.com company ships products from three different warehouses (A, B, and C). Based on customer complaints, it appears that 3% of the shipments coming from A are somehow faulty, as are 5% of the shipments coming from B, and 2% coming from C. Suppose a customer is mailed an order and calls in a complaint the next day. What is the probability the item came from Warehouse C? Assume that Warehouses A, B, and C ship 30%, 20%, and 50% of the dot.com's sales, respectively.

2.4.53. A desk has three drawers. The first contains two gold coins, the second has two silver coins, and the third has one gold coin and one silver coin. A coin is drawn from a drawer selected at random. Suppose the coin selected was silver. What is the probability that the other coin in that drawer is gold?

INDEPENDENCE

Section 2.4 dealt with the problem of reevaluating the probability of a given event in light of the additional information that some other event has already occurred. It often is the case, though, that the probability of the given event remains unchanged, regardless of the outcome of the second event—that is, $P(A|B) = P(A) = P(A|B^C)$. Events sharing this property are said to be *independent*. Definition 2.5.1 gives a necessary and sufficient condition for two events to be independent.

Definition 2.5.1. Two events A and B are said to be *independent* if $P(A \cap B) = P(A) \cdot P(B)$.

Comment. The fact that the probability of the intersection of two independent events is equal to the product of their individual probabilities follows immediately from our first definition of independence, that $P(A|B) = P(A)$. Recall that the definition of conditional probability holds true for *any* two events A and B [provided that $P(B > 0)$]:

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

But $P(A|B)$ can equal $P(A)$ only if $P(A \cap B)$ factors into $P(A)$ times $P(B)$.

EXAMPLE 2.5.1

Let A be the event of drawing a king from a standard poker deck and B, the event of drawing a diamond. Then, by Definition 2.5.1, A and B are independent because the probability of their intersection—drawing a king of diamonds—is equal to $P(A) \cdot P(B)$:

$$P(A \cap B) = \frac{1}{52} = \frac{1}{4} \cdot \frac{1}{13} = P(A) \cdot P(B)$$

EXAMPLE 2.5.2

Suppose that A and B are independent events. Does it follow that A^C and B^C are also independent? That is, does $P(A \cap B) = P(A) \cdot P(B)$ guarantee that $P(A^C \cap B^C) = P(A^C) \cdot P(B^C)$?

Yes. The proof is accomplished by equating two different expressions for $P(A^C \cup B^C)$. First, by Theorem 2.3.6,

$$P(A^C \cup B^C) = P(A^C) + P(B^C) - P(A^C \cap B^C) \tag{2.5.1}$$

But the union of two complements is the complement of their intersection (recall Question 2.2.32). Therefore,

$$P(A^C \cup B^C) = 1 - P(A \cap B) \tag{2.5.2}$$

Combining Equations 2.5.1 and 2.5.2, we get

$$1 - P(A \cap B) = 1 - P(A) + 1 - P(B) - P(A^C \cap B^C)$$

Since A and B are independent, $P(A \cap B) = P(A) \cdot P(B)$, so

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

the latter factorization implying that A^C and B^C are, themselves, independent. (If A and B are independent, are A and B^C independent?)

EXAMPLE 2.5.3

Administrators-R-Us is responding to affirmative-action litigation by establishing hiring goals by race and sex for its office staff. So far they have agreed to employ the 120 people characterized in Table 2.5.1. How many black women do they need in order for the events A: Employee is female and B: Employee is black to be independent?

TABLE 2.5.1

	White	Black
Male	50	30
Female	40	

Let x denote the number of black women necessary for A and B to be independent. Then

$$P(A \cap B) = P(\text{Black female}) = x/(120 + x)$$

must equal

$$P(A)P(B) = P(\text{Female})P(\text{Black}) = [(40 + x)/(120 + x)] \cdot [(30 + x)/(120 + x)]$$

Setting $x/(120 + x) = [(40 + x)/(120 + x)] \cdot [(30 + x)/(120 + x)]$ implies that $x = 24$ black women need to be on the staff in order for A and B to be independent.

Comment. Having shown that "Employee is female" and "Employee is black" are independent, does it follow that, say, "Employee is male" and "Employee is white" are independent? Yes. By virtue of the derivation in Example 2.5.2, the independence of events A and B implies the independence of events A^C and B^C (as well as A and B^C and A^C and B). It follows, then, that the $x = 24$ black women not only make A and B independent, they also imply, more generally, that "race" and "sex" are independent.

EXAMPLE 2.5.4

Suppose that two events, A and B, each having nonzero probability, are mutually exclusive. Are they also independent?

No. If A and B are mutually exclusive, then $P(A \cap B) = 0$. But $P(A) \cdot P(B) > 0$ (by assumption), so the equality spelled out in Definition 2.5.1 that characterizes independence is not met.

Deducing Independence

Sometimes the physical circumstances surrounding two events make it obvious that the occurrence (or nonoccurrence) of one has absolutely no influence or effect on the occurrence (or nonoccurrence) of the other. If that should be the case, then the two events will necessarily be *independent* in the sense of Definition 2.5.1.

Suppose a coin is tossed twice. Clearly, whatever happens on the first toss has no physical connection or influence on the outcome of the second. If A and B, then, are events defined on the second and first tosses, respectively, it would have to be the case that $P(A|B) = P(a|B^C) = P(A)$. For example, let A be the event that the second toss of a fair coin is a head, and let B be the event that the first toss of that coin is a tail. Then

$$P(A|B) = P(\text{Head on second toss} \mid \text{Tail on first toss}) = P(\text{Head on second toss}) = \frac{1}{2}$$

Being able to infer that certain events are independent proves to be of enormous help in solving certain problems. The reason is that many events of interest are, in fact, intersections. If those events are independent, then the probability of that intersection reduces to a simple product (because of Definition 2.5.1)—that is, $P(A \cap B) = P(A) \cdot P(B)$. For the coin tosses just described,

$$\begin{aligned} P(A \cap B) &= P(\text{head on second toss} \cap \text{tail on first toss}) \\ &= P(A) \cdot P(B) \\ &= P(\text{head on second toss}) \cdot P(\text{tail on first toss}) \\ &= \frac{1}{2} \cdot \frac{1}{2} \\ &= \frac{1}{4} \end{aligned}$$

EXAMPLE 2.5.5

Myra and Carlos are summer interns working as proofreaders for a local newspaper. Based on aptitude tests, Myra has a 50% chance of spotting a hyphenation error, while Carlos picks up on that same kind of mistake 80% of the time. Suppose the copy they are editing contains a hyphenation error. What is the probability it goes undetected?

Let A and B be the events that Myra and Carlos, respectively, catch the mistake. By assumption, $P(A) = 0.50$ and $P(B) = 0.80$. What we are looking for is the probability of the complement of a union. That is,

$$\begin{aligned} P(\text{error goes undetected}) &= 1 - P(\text{error } \textit{is} \text{ detected}) \\ &= 1 - P(\text{Myra or Carlos or both see the mistake}) \\ &= 1 - P(A \cup B) \\ &= 1 - \{P(A) + P(B) - P(A \cap B)\} \quad \text{(from Theorem 2.3.6)} \end{aligned}$$

Since proofreaders invariably work by themselves, events A and B are necessarily independent, so $P(A \cap B)$ would reduce to the product, $P(A) \cdot P(B)$. It follows that such an error would go unnoticed *10%* of the time:

$$P(\text{error goes undetected}) = 1 - \{0.50 + 0.80 - (0.50)(0.80)\} = 1 - 0.90$$
$$= 0.10$$

EXAMPLE 2.5.6

Suppose that one of the genes associated with the control of carbohydrate metabolism exhibits two alleles—a dominant W and a recessive w. If the probabilities of the WW, Ww, and ww genotypes in the present generation are p, q, and r, respectively, for both males and females, what are the chances that an individual in the *next* generation will be a ww?

Let A denote the event that an offspring receives a w allele from its father; let B denote the event that it receives the recessive allele from its mother. What we are looking for is $P(A \cap B)$.

According to the information given,

$$\begin{aligned} p &= P(\text{parent has genotype WW}) = P(\text{WW}) \\ q &= P(\text{parent has genotype Ww}) = P(\text{Ww}) \\ r &= P(\text{parent has genotype ww}) = P(\text{ww}) \end{aligned}$$

If an offspring is equally likely to receive either of its parent's alleles, the probabilities of A and B can be computed using Theorem 2.4.1:

$$\begin{aligned} P(A) &= P(A \mid \text{WW})P(\text{WW}) + P(A \mid \text{Ww})P(\text{Ww}) + P(A \mid \text{ww})P(\text{ww}) \\ &= 0 \cdot p + \frac{1}{2} \cdot q + 1 \cdot r \\ &= r + \frac{q}{2} = P(B) \end{aligned}$$

Lacking any evidence to the contrary, there is every reason here to assume that A and B are independent events, in which case

$$\begin{aligned} P(A \cap B) &= P(\text{offspring has genotype ww}) \\ &= P(A) \cdot P(B) \\ &= \left(r + \frac{q}{2}\right)^2 \end{aligned}$$

(This particular model for allele segregation, together with the independence assumption, is called *random Mendelian mating.*)

EXAMPLE 2.5.7

Emma and Josh have just gotten engaged. What is the probability that they have different blood types? Assume that blood types for both men and women are distributed in the general population according to the following proportions:

Blood Type	Proportion
A	40%
B	10%
AB	5%
O	45%

First, note that the event "Emma and Josh have *different* blood types" includes more possibilities than does the event "Emma and Josh have the *same* blood type." That being the case, the complement will be easier to work with than the question originally posed. We can start, then, by writing

$$\begin{aligned} &P(\text{Emma and Josh have different blood types}) \\ &\quad = 1 - P(\text{Emma and Josh have the same blood type}) \end{aligned}$$

Now, if we let E_X and J_X represent the events that Emma and Josh, respectively, have blood type X, then the event "Emma and Josh have the same blood type" is a union of intersections, and we can write

$$\begin{aligned} P(\text{Emma and Josh have the same blood type}) = P\{&(E_A \cap J_A) \cup (E_B \cap J_B) \\ &\cup (E_{AB} \cap J_{AB}) \cup (E_O \cap J_O)\} \end{aligned}$$

Since the four intersections here are mutually exclusive, the probability of their union becomes the sum of their probabilities. Moreover, "blood type" is not a factor in the selection of a spouse, so E_X and J_X are independent events and $P(E_X \cap J_X) = P(E_X)P(J_X)$. It follows, then, that Emma and Josh have a *62.5%* chance of having different blood types:

$$\begin{aligned} P(\text{Emma and Josh have different blood types}) &= 1 - \{P(E_A)P(J_A) + P(E_B)P(J_B) \\ &\quad + P(E_{AB})P(J_{AB}) + P(E_O)P(J_O)\} \\ &= 1 - \{(0.40)(0.40) + (0.10)(0.10) \\ &\quad + (0.05)(0.05) + (0.45)(0.45)\} \\ &= 0.625 \end{aligned}$$

QUESTIONS

2.5.1. Suppose that $P(A \cap B) = 0.2$, $P(A) = 0.6$, and $P(B) = 0.5$.
- **(a)** Are A and B mutually exclusive?
- **(b)** Are A and B independent?
- **(c)** Find $P(A^C \cup B^C)$.

2.5.2. Spike is not a terribly bright student. His chances of passing chemistry are 0.35; mathematics, 0.40; and both, 0.12. Are the events "Spike passes chemistry" and "Spike passes mathematics" independent? What is the probability that he fails both subjects?

2.5.3. Two fair dice are rolled. What is the probability that the number showing on one will be twice the number appearing on the other?

2.5.4. Urn I has three red chips, two black chips, and five white chips; urn II has two red, four black, and three white. One chip is drawn at random from each urn. What is the probability that both chips are the same color?

2.5.5. Dana and Cathy are playing tennis. The probability that Dana wins at least one out of two games is 0.3. What is the probability that Dana wins at least one out of four?

2.5.6. Three points, X_1, X_2, and X_3, are chosen at random in the interval $(0, a)$. A second set of three points, Y_1, Y_2, and Y_3, are chosen at random in the interval $(0, b)$. Let A be the event that X_2 is between X_1 and X_3. Let B be the event that $Y_1 < Y_2 < Y_3$. Find $P(A \cap B)$.

2.5.7. Suppose that $P(A) = \frac{1}{4}$ and $P(B) = \frac{1}{8}$.

(a) What does $P(A \cup B)$ equal if

1. A and B are mutually exclusive?
2. A and B are independent?

(b) What does $P(A \mid B)$ equal if

1. A and B are mutually exclusive?
2. A and B are independent?

2.5.8. Suppose that events A, B, and C are independent.

(a) Use a Venn diagram to find an expression for $P(A \cup B \cup C)$ that does *not* make use of a complement.

(b) Find an expression for $P(A \cup B \cup C)$ that *does* make use of a complement.

2.5.9. A fair coin is tossed four times. What is the probability that the number of heads appearing on the first two tosses is equal to the number of heads appearing on the second two tosses?

2.5.10. Suppose that two cards are drawn from a standard 52-card poker deck. Let A be the event that both are either a jack, queen, king, or ace of hearts, and let B be the event that both are aces. Are A and B independent? Note: There are 1,326 equally-likely ways to draw two cards from a poker deck.

Defining the Independence of More Than Two Events

It is not immediately obvious how to extend Definition 2.5.1 to, say, *three* events. To call A, B, and C independent, should we require that the probability of the three-way intersection factors into the product of the three original probabilities,

$$P(A \cap B \cap C) = P(A) \cdot P(B) \cdot P(C) \tag{2.5.3}$$

or should we impose the definition we already have on the three *pairs* of events:

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B) \\ P(B \cap C) &= P(B) \cdot P(C) \\ P(A \cap C) &= P(A) \cdot P(C) \end{aligned} \tag{2.5.4}$$

Actually, neither condition by itself is sufficient. If three events satisfy Equations 2.5.3 *and* 2.5.4, we will call them independent (or *mutually independent*), but Equation 2.5.3 does not imply Equation 2.5.4, nor does Equation 2.5.4 imply Equation 2.5.3 (see Questions 2.5.11 and 2.5.12).

More generally, the independence of n events requires that the probabilities of all possible intersections equal the products of all the corresponding individual probabilities. Definition 2.5.2 states the result formally. Analogous to what was true in the case of *two* events, the practical applications of Definition 2.5.2 arise when n events are mutually independent, and we can calculate $P(A_1 \cap A_2 \cap \cdots \cap A_n)$ by computing the product $P(A_1) \cdot P(A_2) \cdots P(A_n)$.

Definition 2.5.2. Events $A_1, A_2, \ldots, A_n$ are said to be *independent* if for every set of indices $i_1, i_2, \ldots, i_k$ between 1 and n, inclusive,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k})$$

EXAMPLE 2.5.8

Audrey has registered for four courses in the upcoming fall term, one each in physics, English, economics, and sociology. Based on what has happened in the recent past, it would be reasonable to assume that she has a 20% chance of being bumped from the physics class, a 10% chance of being bumped from the English class, a 30% chance of being bumped from the economics class, and no chance of being bumped from the sociology class. What is the probability that she fails to get into at least one class?

For the events

A_1: Audrey is bumped from physics
A_2: Audrey is bumped from English
A_3: Audrey is bumped from economics
A_4: Audrey is bumped from sociology

$P(A_1) = 0.20$, $P(A_2) = 0.10$, $P(A_3) = 0.30$, and $P(A_4) = 0$. The chance that Audrey gets bumped from at least one class can be written as the probability of a union,

$$P(\text{Audrey is bumped from at least one class}) = P(A_1 \cup A_2 \cup A_3 \cup A_4) \qquad (2.5.5)$$

but evaluating Equation 2.5.5 is somewhat involved because the A_i's are not mutually exclusive. A much simpler solution is to express the complement of "bumped from at least one" as an intersection:

$$\begin{aligned} P\left(\begin{array}{c}\text{Audrey is bumped from}\\ \text{at least one class}\end{array}\right) &= 1 - P\left(\begin{array}{c}\text{Audrey is not bumped}\\ \text{from any classes}\end{array}\right) \\ &= 1 - P(A_1^C \cap A_2^C \cap A_3^C \cap A_4^C) \end{aligned}$$

Since different departments are involved, the A_i's are likely to be independent events, so the intersection "factors" and we can write

$$\begin{aligned} P(\text{Audrey is bumped from at least one class}) &= 1 - P(A_1^C)P(A_2^C)P(A_3^C)P(A_4^C) \\ &= 1 - (0.80)(0.90)(0.70)(1.00) \\ &= 0.496 \end{aligned}$$

EXAMPLE 2.5.9

The YouDie-WePay Insurance Company plans to assess its future liabilities by sampling the records of its current policyholders. A pilot study has turned up three clients—one living in Alaska, one in Missouri, and one in Vermont—whose estimated chances of surviving to the year 2010 are 0.7, 0.9, and 0.3, respectively. What is the probability that by the end of 2009 the company will have had to pay death benefits to exactly one of the three?

Let A_1 be the event "Alaska client survives through 2009." Define A_2 and A_3 analogously for the Missouri client and Vermont client, respectively. Then the event E: "Exactly one dies" can be written as the union of three intersections:

$$E = \left(A_1 \cap A_2 \cap A_3^C\right) \cup \left(A_1 \cap A_2^C \cap A_3\right) \cup \left(A_1^C \cap A_2 \cap A_3\right)$$

Since each of the intersections is mutually exclusive of the other two,

$$P(E) = P\left(A_1 \cap A_2 \cap A_3^C\right) + P\left(A_1 \cap A_2^C \cap A_3\right) + P\left(A_1^C \cap A_2 \cap A_3\right)$$

Furthermore, there is no reason to believe that for all practical purposes the fates of the three are not independent. That being the case, each of the intersection probabilities reduces to a product, and we can write

$$\begin{aligned} P(E) &= P(A_1) \cdot P(A_2) \cdot P\left(A_3^C\right) + P(A_1) \cdot P\left(A_2^C\right) \cdot P(A_3) + P\left(A_1^C\right) \cdot P(A_2) \cdot P(A_3) \\ &= (0.7)(0.9)(0.7) + (0.7)(0.1)(0.3) + (0.3)(0.9)(0.3) \\ &= 0.543 \end{aligned}$$

Comment. "Declaring" events independent for reasons other than those prescribed in Definition 2.5.2 is a necessarily subjective endeavor. Here we might feel fairly certain that a "random" person dying in Alaska will not affect the survival chances of a "random" person residing in Missouri (or Vermont). But there may be special circumstances that invalidate that sort of argument. For example, what if the three individuals in question were mercenaries fighting in an African border war and were all crew members assigned to the same helicopter? In practice, all we can do is look at each situation on an individual basis and try to make a reasonable judgment as to whether the occurrence of one event is likely to influence the outcome of another.

EXAMPLE 2.5.10

Protocol for making financial decisions in a certain corporation follows the "circuit" pictured in Figure 2.5.1. Any budget is first screened by 1. If he approves it, the plan is forwarded to 2, 3, and 5. If either 2 or 3 concurs, it goes to 4. If either 4 or 5 say "yes," it moves on to 6 for a final reading. Only if 6 is also in agreement does the proposal pass. Suppose that 1, 5, and 6 each has a 50% chance of saying "yes," whereas 2, 3, and 4 will each concur with a probability of 0.70. If everyone comes to a decision independently, what is the probability that a budget will pass?

Probabilities of this sort are calculated by reducing the circuit to its component unions and intersections. Moreover, if all decisions are made independently, which is the case here, then every intersection becomes a product.

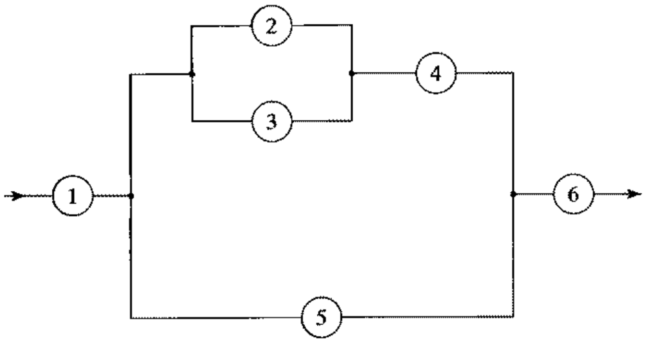


FIGURE 2.5.1

Let A_i be the event that person i approves the budget, $i = 1, 2, \ldots, 6$. Looking at Figure 2.5.1, we see that

$$\begin{aligned} P(\text{budget passes}) &= P(A_1 \cap \{[(A_2 \cup A_3) \cap A_4] \cup A_5\} \cap A_6) \\ &= P(A_1)P\{[(A_2 \cup A_3) \cap A_4] \cup A_5\}P(A_6) \end{aligned}$$

By assumption, $P(A_1) = 0.5$, $P(A_2) = 0.7$, $P(A_3) = 0.7$, $P(A_4) = 0.7$, $P(A_5) = 0.5$, and $P(A_6) = 0.5$, so

$$\begin{aligned} P\{[(A_2 \cup A_3) \cap A_4]\} &= [P(A_2) + P(A_3) - P(A_2)P(A_3)]P(A_4) \\ &= [0.7 + 0.7 - (0.7)(0.7)](0.7) \\ &= 0.637 \end{aligned}$$

Therefore,

$$\begin{aligned} P(\text{budget passes}) &= (0.5)\{0.637 + 0.5 - (0.637)(0.5)\}(0.5) \\ &= 0.205 \end{aligned}$$

Repeated Independent Events

We have already seen several examples where the event of interest was actually an intersection of independent simpler events (in which case the probability of the intersection reduced to a product). There is a special case of that basic scenario that deserves special mention because it applies to numerous real-world situations. If the events making up the intersection all arise from the same physical circumstances and assumptions (i.e., they represent repetitions of the same experiment), they are referred to as *repeated independent trials*. The number of such trials may be finite or infinite.

EXAMPLE 2.5.11

Suppose the string of Christmas tree lights you just bought has twenty-four bulbs wired in series. If each bulb has a 99.9% chance of "working" the first time current is applied, what is the probability that the string, itself, will *not* work.

Let A_i be the event that the ith bulb fails, $i = 1, 2, \ldots, 24$. Then

$$\begin{aligned}
P(\text{string fails}) &= P(\text{at least one bulb fails})\\
&= P(A_1 \cup A_2 \cup \cdots \cup A_{24})\\
&= 1 - P(\text{string works})\\
&= 1 - P(\text{all twenty-four bulbs work})\\
&= 1 - P(A_1^C \cap A_2^C \cap \cdots \cap A_{24}^C)
\end{aligned}$$

If we assume that bulb failures are independent events,

$$P(\text{string fails}) = 1 - P(A_1^C)P(A_2^C)\cdots P(A_{24}^C)$$

Moreover, since all the bulbs are presumably manufactured the same way, $P(A_i^C)$ is the same for all i, so

$$\begin{aligned}
P(\text{string fails}) &= 1 - \left\{P(A_i^C)\right\}^{24}\\
&= 1 - (0.999)^{24}\\
&= 1 - 0.98\\
&= 0.02
\end{aligned}$$

The chances are one in fifty, in other words, that the string would not work the first time you take it out of the box.

EXAMPLE 2.5.12

A box contains one two-headed coin and eight fair coins. One is drawn at random and tossed seven times. Suppose that all seven tosses come up heads. What is the probability that the coin is fair?

This is basically a Bayes' problem, but the conditional probabilities on the right-hand side of Theorem 2.4.2 appeal to the notion of independence as well. Define the events

B: seven heads occurred in seven tosses
A_1: coin tossed has two heads
A_2: coin tossed was fair

The question is asking for $P(A_2 \mid B)$.

By virtue of the composition of the box, $P(A_1) = \frac{1}{9}$ and $P(A_2) = \frac{8}{9}$. Also,

$$\begin{aligned}
P(B \mid A_1) &= P(\text{head on first toss} \cap \cdots \cap \text{head on seventh toss} \mid \text{coin has two heads})\\
&= 1^7 = 1
\end{aligned}$$

Similarly, $P(B \mid A_2) = \left(\frac{1}{2}\right)^7$. Substituting into Bayes's formula shows that the probability is 0.06 that the coin is fair:

$$P(A_2 \mid B) = \frac{P(B \mid A_2)P(A_2)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2)}$$

$$= \frac{\left(\frac{1}{2}\right)^7\left(\frac{8}{9}\right)}{1\left(\frac{1}{9}\right) + \left(\frac{1}{2}\right)^7\left(\frac{8}{9}\right)}$$

$$= 0.06$$

Comment. Let B_n denote the event that the coin chosen at random is tossed n times with the result being that n heads appear. As our intuition would suggest, $P(A_2 \mid B_n) \rightarrow 0$ as $n \rightarrow \infty$:

$$\lim_{n\to\infty} P(A_2 \mid B_n) = \lim_{n\to\infty} \frac{8\left(\frac{1}{2}\right)^n}{1 + 8\left(\frac{1}{2}\right)^n} = 0$$

EXAMPLE 2.5.13

During the 1978 baseball season, Pete Rose of the Cincinnati Reds set a National League record by hitting safely in 44 consecutive games. Assume that Rose is a .300 hitter and that he comes to bat four times each game. If each at-bat is assumed to be an independent event, what probability might reasonably be associated with a hitting streak of that length?

For this problem we need to invoke the repeated independent trials model *twice*—once for the four at-bats making up a game and a second time for the forty-four games making up the streak. Let A_i denote the event "Rose hits safely in ith game," $i = 1, 2, \ldots, 44$. Then

$$\begin{aligned} P(\text{Rose hits safely in forty-four consecutive games}) &= P(A_1 \cap A_2 \cap \cdots \cap A_{44}) \\ &= P(A_1) \cdot P(A_2) \cdot \cdots \cdot P(A_{44}) \end{aligned} \tag{2.5.6}$$

Since all the $P(A_i)$'s are equal, we can further simplify Equation 2.5.6 by writing

$$P(\text{Rose hits safely in 44 consecutive games}) = [P(A_1)]^{44}$$

To calculate $P(A_1)$ we should focus on the *complement* of A_1. Specifically,

$$\begin{aligned} P(A_1) &= 1 - P\left(A_1^C\right) \\ &= 1 - P(\text{Rose does } \textit{not} \text{ hit safely in Game 1}) \\ &= 1 - P(\text{Rose makes four outs}) \\ &= 1 - (0.700)^4 \quad \text{(why?)} \\ &= 0.76 \end{aligned}$$

Therefore, the probability of a .300 hitter putting together a forty-four-game streak (during a given set of forty-four games) is 0.0000057:

$$\begin{aligned} P(\text{Rose hits safely in forty-four consecutive games}) &= (0.76)^{44} \\ &= 0.0000057 \end{aligned}$$

Comment. The analysis described here has the basic "structure" of a repeated independent trials problem, but the assumptions that the latter makes are not entirely satisfied by the data. Each at-bat, for example, is not really a repetition of the same experiment, nor is $P(A_i)$ the same for all i. Rose would obviously have different probabilities of getting a hit against different pitchers. Moreover, "four" was probably the typical number of official at-bats that he had during a game, but there would certainly have been many instances where he had either fewer or more. Modest deviations from game to game, though, would not have a major effect on the probability associated with Rose's forty-four-game streak.

EXAMPLE 2.5.14

In a certain third world nation, statistics show that only eight out of ten children born in the early 1980s reached the age of twenty-one. If the same mortality rate is operative over the next generation, how many children does a woman need to bear if she wants to have at least a 75% probability that at least one of her offspring survives to adulthood?

Restated, the question is asking for the smallest integer n such that

$$P(\text{at least one of } n \text{ children survives to adulthood}) \geq 0.75$$

Assuming that the fates of the n children are independent events,

$$\begin{aligned} P(\text{at least one (of } n) \text{ survives to age twenty-one}) &= 1 - P(\text{all } n \text{ die before adulthood}) \\ &= 1 - (0.80)^n \end{aligned}$$

Table 2.5.2 shows the value of $1 - (0.80)^n$ as a function of n.

TABLE 2.5.2

n	$1 - (0.80)^n$
5	0.67
6	0.74
7	0.79

By inspection, we see that the smallest number of children for which the probability is at least 0.75 that at least one of them survives to adulthood is *seven*.

EXAMPLE 2.5.15 (Optional)

In the game of craps, one of the ways a player can win is by rolling (with two dice) one of the sums four, five, six, eight, nine, or ten, and then rolling that sum again before rolling a sum of seven. For example, the sequence of sums six, five, eight, eight, six would result in the player winning on his fifth roll. In gambling parlance, "six" is the player's "point," and he "made his point." On the other hand, the sequence of sums eight, four, ten, seven would result in the player losing on his fourth roll: his point was an eight, but he rolled a sum of seven before he rolled a second eight. What is the probability that a player wins with a point of ten?

TABLE 2.5.3

Sequence of Rolls	Probability
(10, 10)	(3/36)(3/36)
(10, no 10 or 7, 10)	(3/36)(27/36)(3/36)
(10, no 10 or 7, no 10 or 7, 10)	(3/36)(27/36)(27/36)(3/36)
⋮	⋮

Table 2.5.3 shows some of the ways a player can make a point of ten. Each sequence, of course, is an intersection of independent events, so its probability becomes a product. The event "Player wins with a point of ten" is then the union of all the sequences that could have been listed in the first column. Since all those sequences are mutually exclusive, the probability of winning with a point of ten reduces to the sum of an infinite number of products:

$$\begin{aligned} P(\text{Player wins with a point of 10}) &= \frac{3}{36}\cdot\frac{3}{36} + \frac{3}{36}\cdot\frac{27}{36}\cdot\frac{3}{36} \\ &\quad + \frac{3}{36}\cdot\frac{27}{36}\cdot\frac{27}{36}\cdot\frac{3}{36} + \cdots \\ &= \frac{3}{36}\cdot\frac{3}{36}\sum_{k=0}^{\infty}\left(\frac{27}{36}\right)^k \end{aligned} \tag{2.5.7}$$

Recall from algebra that if $0 < r < 1$,

$$\sum_{k=0}^{\infty} r^k = 1/(1 - r)$$

Applying the formula for the sum of a geometric series to Equation 2.5.7 shows that the probability of winning at craps with a point of ten is $\frac{1}{36}$:

$$\begin{aligned} P(\text{Player wins with a point of ten}) &= \frac{3}{36}\cdot\frac{3}{36}\cdot\frac{1}{(1-\frac{27}{36})} \\ &= \frac{1}{36} \end{aligned}$$

TABLE 2.5.4

Point	P (makes point)
4	1/36
5	16/360
6	25/396
8	25/396
9	16/360
10	1/36

Comment. Table 2.5.4 shows the probabilities of a person "making" each of the possible six points—4, 5, 6, 8, 9, and 10. According to the rules of craps, a player wins by either (1) getting a sum of seven or eleven on the first roll or (2) getting a 4, 5, 6, 8, 9, or 10 on the first roll and making the point. But $P(\text{sum} = 7) = 6/36$ and $P(\text{sum} = 11) = 2/36$, so

$$\begin{aligned} P(\text{player wins}) &= \frac{6}{36} + \frac{2}{36} + \frac{1}{36} + \frac{16}{360} + \frac{25}{396} + \frac{25}{396} + \frac{16}{360} + \frac{1}{36} \\ &= 0.493 \end{aligned}$$

As even-money games go, craps is relatively fair—the probability of the shooter winning is not much less than 0.500.

QUESTIONS

2.5.11. Suppose that two fair dice (one red and one green) are rolled. Define the events

A: a 1 or a 2 shows on the red die
B: a 3, 4, or 5 shows on the green die
C: the dice total is four, eleven, or twelve

Show that these events satisfy Equation 2.5.3 but not Equation 2.5.4.

2.5.12. A roulette wheel has thirty-six numbers colored red or black according to the pattern indicated below:

Roulette wheel pattern

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
R	R	R	R	R	B	B	B	B	R	R	R	R	B	B	B	B	B
36	35	34	33	32	31	30	29	28	27	26	25	24	23	22	21	20	19

Define the events

A: Red number appears
B: Even number appears
C: Number is less than or equal to eighteen

Show that these events satisfy Equation 2.5.4 but not Equation 2.5.3.

2.5.13. How many probability equations need to be verified to establish the mutual independence of *four* events?

2.5.14. In a roll of a pair of fair dice (one red and one green), let A be the event the red die shows a 3, 4, or 5; let B be the event the green die shows a 1 or a 2; and let C be the event the dice total is seven. Show that A, B, and C are independent.

2.5.15. In a roll of a pair of fair dice (one red and one green), let A be the event of an odd number on the red die, let B be the event of an odd number on the green die, and let C be the event that the sum is odd. Show that any pair of these events are independent but that A, B, and C are not mutually independent.

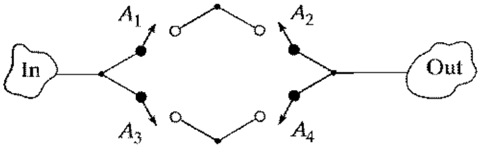
2.5.16. On her way to work, a commuter encounters four traffic signals. Assume that the distance between each of the four is sufficiently great that her probability of getting a green light at any intersection is independent of what happened at any previous intersection. The first two lights are green for forty seconds of each minute; the last two, for thirty seconds of each minute. What is the probability that the commuter has to stop at least three times?

2.5.17. School board officials are debating whether to require all high school seniors to take a proficiency exam before graduating. A student passing all three parts (mathematics, language skills, and general knowledge) would be awarded a diploma; otherwise, he would receive only a certificate of attendance. A practice test given to this year's ninety-five hundred seniors resulted in the following numbers of failures:

Subject Area	Number of Students Failing
Mathematics	3325
Language skills	1900
General knowledge	1425

If "Student fails mathematics," "Student fails language skills," and "Student fails general knowledge" are independent events, what proportion of next year's seniors can be expected to fail to qualify for a diploma? Does independence seem a reasonable assumption in this situation?

2.5.18. Consider the following four-switch circuit:



If all switches operate independently and P(switch closes) $= p$, what is the probability the circuit is completed?

2.5.19. A fast-food chain is running a new promotion. For each purchase, a customer is given a game card that may win \$10. The company claims that the probability of a person winning at least once in five tries is 0.32. What is the probability that a customer wins \$10 on his or her first purchase?

2.5.20. Players A, B, and C toss a fair coin in order. The first to throw a head wins. What are their respective chances of winning?

2.5.21. Andy, Bob, and Charley have gotten into a disagreement over a female acquaintance and decide to settle their dispute with a three-cornered pistol duel. Of the three, Andy is the worst shot, hitting his target only 30% of the time. Charley, a little better, is on-target 50% of the time, while Bob never misses. The rules they agree to are simple: They are to fire at the targets of their choice in succession, and cyclically, in the order Andy, Bob, Charley, Andy, Bob, Charley, and so on until only one of them is left standing. (On each "turn," they get only one shot. If a combatant is hit, he no longer participates, either as a shooter or as a target.) Show that Andy's optimal strategy, assuming he wants to maximize his chances of staying alive, is to fire his first shot into the ground.

2.5.22. According to an advertising study, 15% of television viewers who have seen a certain automobile commercial can correctly identify the actor who does the voice-over. Suppose that 10 such people are watching TV and the commercial comes on. What is the probability that at least one of them can name the actor? What is the probability that exactly one can name the actor?

2.5.23. A fair die is rolled and then n fair coins are tossed, where n is the number showing on the die. What is the probability that no heads appear?

2.5.24. Each of m urns contains three red chips and four white chips. A total of r samples with replacement are taken from each urn. What is the probability that at least one red chip is drawn from at least one urn?

2.5.25. If two fair dice are tossed, what is the smallest number of throws, n, for which the probability of getting at least one double six exceeds 0.5? (*Note:* This was one of the first problems that de Méré communicated to Pascal in 1654.)

2.5.26. A pair of fair dice are rolled until the first sum of eight appears. What is the probability that a sum of seven does not precede that first sum of eight?

2.5.27. An urn contains w white chips, b black chips, and r red chips. The chips are drawn out at random, one at a time, with replacement. What is the probability that a white appears before a red?

2.5.28. A Coast Guard dispatcher receives an SOS from a ship that has run aground off the shore of a small island. Before the captain can relay her exact position, though, her radio goes dead. The dispatcher has n helicopter crews he can send out to conduct a search. He suspects the ship is somewhere either south in area I (with probability p) or north in area II (with probability $1 - p$). Each of the n rescue parties is equally competent and has probability r of locating the ship given it has run aground in the sector being searched. How should the dispatcher deploy the helicopter crews to maximize the probability that one of them will find the missing ship? *Hint:* Assume that m search crews are sent to area I and $n - m$ are sent to area II. Let B denote the event that the ship is found, let A_1 be the event that the ship is in area I, and let A_2 be the event that the ship is in area II. Use Theorem 2.4.1 to get an expression for $P(B)$; then differentiate with respect to m.

2.5.29. A computer is instructed to generate a random sequence using the digits 0 through 9; repetitions are permissible. What is the shortest length the sequence can be and still have at least a 70% probability of containing at least one 4?

COMBINATORICS

Combinatorics is a time-honored branch of mathematics concerned with counting, arranging, and ordering. While blessed with a wealth of early contributors (there are references to combinatorial problems in the Old Testament), its emergence as a separate discipline is often credited to the German mathematician and philosopher Gottfried Wilhelm Leibniz (1646–1716), whose 1666 treatise, *Dissertatio de arte combinatoria,* was perhaps the first monograph written on the subject (111).

Applications of combinatorics are rich in both diversity and number. Users range from the molecular biologist trying to determine how many ways genes can be positioned along a chromosome, to a computer scientist studying queuing priorities, to a psychologist modeling the way we learn, to a weekend poker player wondering whether he should draw to a straight, or a flush, or a full house. Surprisingly enough, solutions to all of these questions are rooted in the same set of four basic theorems and rules, despite the considerable differences that seem to distinguish one question from another.

Counting Ordered Sequences: The Multiplication Rule

More often than not, the relevant "outcomes" in a combinatorial problem are ordered sequences. If two dice are rolled, for example, the outcome (4, 5)—that is, the first die comes up 4 and the second die comes up 5—is an ordered sequence of length two. The number of such sequences is calculated by using the most fundamental result in combinatorics, the *multiplication rule.*

Multiplication Rule. *If operation A can be performed in m different ways and operation B in n different ways, the sequence (operation A, operation B) can be performed in $m \cdot n$ different ways.*

Proof. At the risk of belaboring the obvious, we can verify the multiplication rule by considering a *tree* diagram (see Figure 2.6.1). Since each version of A can be followed by any of n versions of B, and there are m of the former, the total number of "A, B" sequences that can be pieced together is obviously the product $m \cdot n$. ☐

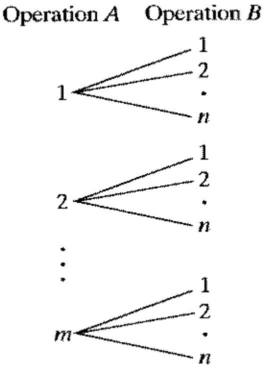


FIGURE 2.6.1

Corollary. *If operation A_i, $i = 1, 2, \ldots, k$, can be performed in n_i ways, $i = 1, 2, \ldots, k$, respectively, then the ordered sequence (operation A_1, operation $A_2, \ldots,$ operation A_k) can be performed in $n_1 \cdot n_2 \cdot \cdots \cdot n_k$ ways.*

EXAMPLE 2.6.1

The combination lock on a briefcase has two dials, each marked off with 16 notches (see Figure 2.6.2). To open the case, a person first turns the left dial in a certain direction for two revolutions and then stops on a particular mark. The right dial is set in a similar fashion, after having been turned in a certain direction for two revolutions. How many different settings are possible?

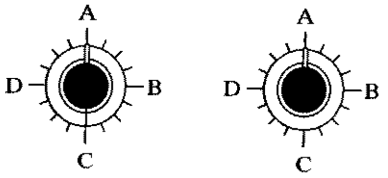


FIGURE 2.6.2

In the terminology of the multiplication rule, opening the briefcase corresponds to the four-step sequence (A_1, A_2, A_3, A_4) detailed in Table 2.6.1. Applying the previous corollary, we see that 1,024 different settings are possible:

$$\begin{aligned}\text{Number of different settings} &= n_1 \cdot n_2 \cdot n_3 \cdot n_4\\ &= 2 \cdot 16 \cdot 2 \cdot 16\\ &= 1{,}024\end{aligned}$$

TABLE 2.6.1

Operation	Purpose	Number of Options
A_1	Rotating the left dial in a particular direction	2
A_2	Choosing an endpoint for the left dial	16
A_3	Rotating the right dial in a particular direction	2
A_4	Choosing an endpoint for the right dial	16

Comment. Designers of locks should be aware that the number of dials, as opposed to the number of notches on each dial, is the critical factor in determining how many different settings are possible. A two-dial lock, for example, where each dial has twenty notches, gives rise to only $2 \cdot 20 \cdot 2 \cdot 20 = 1600$ settings. If those forty notches, though, are distributed among *four* dials (10 to each dial), the number of different settings increases a hundredfold to *160,000* $(= 2 \cdot 10 \cdot 2 \cdot 10 \cdot 2 \cdot 10 \cdot 2 \cdot 10)$.

EXAMPLE 2.6.2

Alphonse Bertillon, a nineteenth-century French criminologist, developed an identification system based on eleven anatomical variables (height, head width, ear length, etc.) that presumably remained essentially unchanged during an individual's adult life. The range of each variable was divided into three subintervals: small, medium, and large. A person's *Bertillon configuration* was an ordered sequence of eleven letters, say

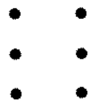
$$s, s, m, m, l, s, l, s, s, m, s$$

where a letter indicated the individual's "size" relative to a particular variable. How populated does a city have to be before it can be guaranteed that at least two citizens will have the same Bertillon configuration?

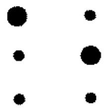
Viewed as an ordered sequence, a Bertillon configuration is an eleven-step classification system, where three options are available at each step. By the multiplication rule, a total of 3^{11}, or 177,147, distinct sequences are possible. Therefore, any city with at least 177,148 adults would necessarily have at least two residents with the same pattern. (The limited number of possibilities generated by Bertillon's variables proved to be one of its major weaknesses. Still, it was widely used in Europe for criminal identification before the development of fingerprinting.)

EXAMPLE 2.6.3

In 1824 Louis Braille invented what would eventually become the standard alphabet for the blind. Based on an earlier form of "night writing" used by the French army for reading battlefield communiqués in the dark, Braille's system replaced each written character with a six-dot matrix:



where certain dots were raised, the choice depending on the character being transcribed. The letter *e*, for example, has two raised dots and is written



Punctuation marks, common words, suffixes, and so on also have specified dot patterns. In all, how many different characters can be enciphered in Braille?

Think of the dots as six distinct operations, numbered 1 to 6 (see Figure 2.6.3). In forming a Braille letter, we have two options for each dot: We can raise it or *not* raise it. The letter *e*, for example, corresponds to the six-step sequence (raise, do not raise, do

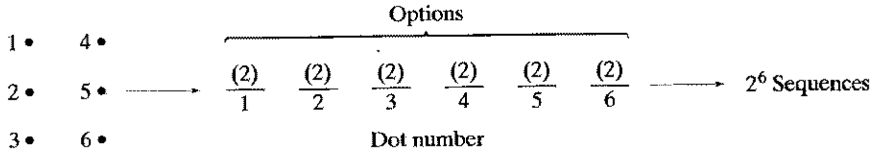


FIGURE 2.6.3

not raise, do not raise, raise, do not raise). The number of such sequences, with $k = 6$ and $n_1 = n_2 = \cdots = n_6 = 2$, is 2^6, or *64*. One of those sixty-four configurations, though, has *no* raised dots, making it of no use to a blind person. Figure 2.6.4 shows the entire sixty-three-character Braille alphabet.

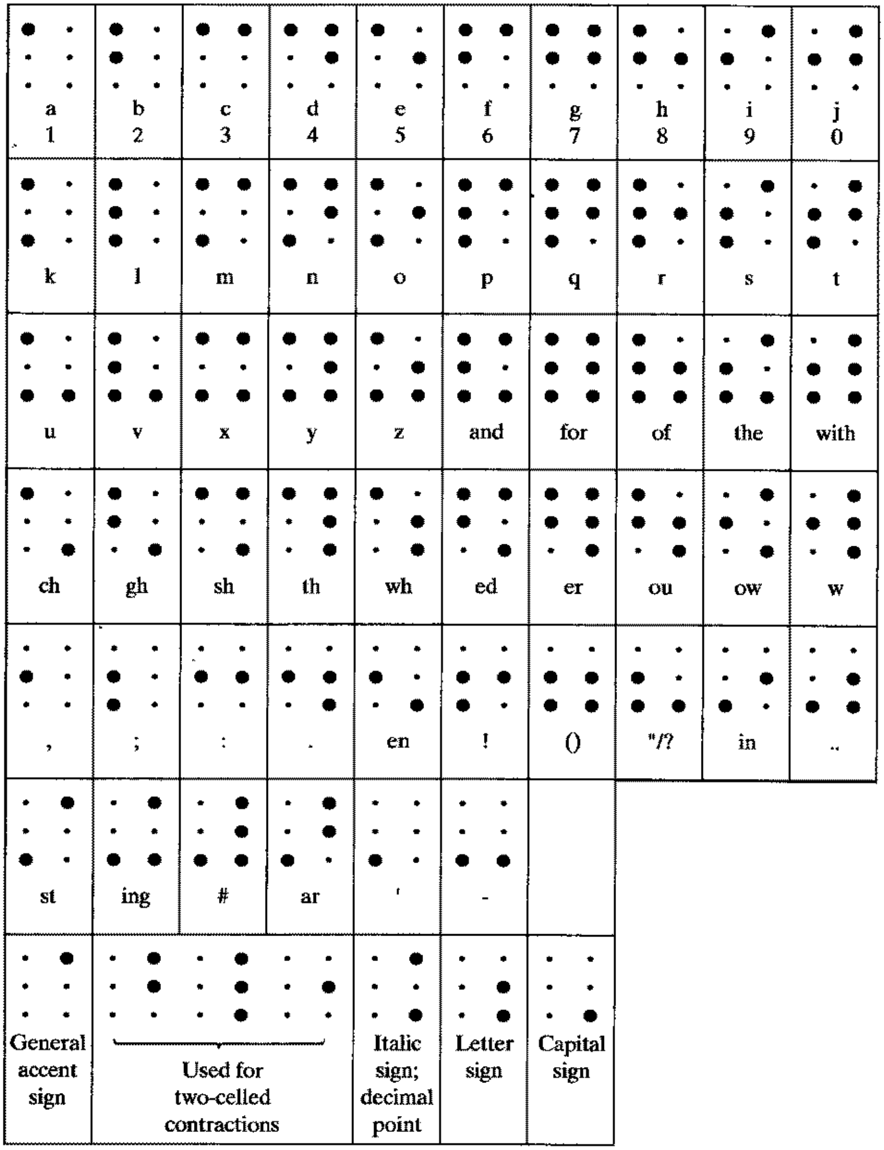


FIGURE 2.6.4

EXAMPLE 2.6.4

The annual NCAA ("March Madness") basketball tournament starts with a field of sixty-four teams. After six rounds of play, the squad that remains unbeaten is declared the national champion. How many different configurations of winners and losers are possible, starting with the first round? Assume that the initial pairing of the sixty-four invited teams into thirty-two first-round matches has already been done.

Counting the number of ways a tournament of this sort can play out is an exercise in applying the multiplication rule twice. Notice, first, that the thirty-two first-round games can be decided in 2^{32} ways. Similarly, the resulting sixteen second-round games can generate 2^{16} different winners, and so on. Overall, the tournament can be pictured as a six-step sequence, where the number of possible outcomes at the six steps are $2^{32}, 2^{16}, 2^8, 2^4, 2^2$, and 2^1, respectively. It follows that the number of possible tournaments (not all of which, of course, would be equally likely!) is the product $2^{32} \cdot 2^{16} \cdot 2^8 \cdot 2^4 \cdot 2^2 \cdot 2^1$, or 2^{63}.

EXAMPLE 2.6.5

An octave contains twelve distinct notes (on a piano, five black keys and seven white keys). How many different eight-note melodies within a single octave can be written if the black keys and white keys need to alternate?

Choices:	5	7	5	7	5	7	5	7	**or**	7	5	7	5	7	5	7	5
	B	W	B	W	B	W	B	W		W	B	W	B	W	B	W	B
	1	2	3	4	5	6	7	8		1	2	3	4	5	6	7	8
	Keys (a)									Keys (b)							

FIGURE 2.6.5

There are two fundamentally different ways in which the black and white keys can alternate—the black keys could produce notes 1, 3, 5, and 7 in the melody, or they could produce notes 2, 4, 6, and 8. Figure 2.6.5 diagrams the two cases. Consider the first, where the black keys produce the odd-numbered notes in the melody. In Multiplication Rule terminology, notes 1, 3, 5, and 7 correspond to Operations A_1, A_3, A_5, and A_7 for which the numbers of available options are $n_1 = 5$, $n_3 = 5$, $n_5 = 5$, and $n_7 = 5$. The white keys (that is, Operations A_2, A_4, A_6, and A_8) all have $n_i = 7$, $i = 2, 4, 6, 8$, so the number of different "alternating" melodies—where a black note comes first—is the product $5^4\ 7^4$, or *1,500,625*.

By the same argument, the second case (where the black keys produce the even-numbered notes in a melody) also generates $7^4\ 5^4 = 1{,}500{,}625$ melodies. Altogether, then, the number of different melodies with alternating black and white notes is the sum $1{,}500{,}625 + 1{,}500{,}625$, or *3,001,250*.

PROBLEM-SOLVING HINTS
(Doing combinatorial problems)

Combinatorial questions sometimes call for problem-solving techniques that are not routinely used in other areas of mathematics. The three listed below are especially helpful.

1. Draw a diagram that shows the structure of the outcomes that are being counted. Be sure to include (or indicate) all relevant variations. A case in point is Figure 2.6.5. Recognizing at the outset that there are two mutually exclusive ways for the black keys and white keys to alternate (i.e., the black keys can be either the odd-numbered notes or the even-numbered notes) is a critical first step in solving the problem. Almost invariably, diagrams such as these will suggest the formula, or combination of formulas, that should be applied.
2. Use enumerations to "test" the appropriateness of a formula. Typically, the answer to a combinatorial problem—that is, the number of ways to do something—will be so large that listing all possible outcomes is not feasible. It often *is* feasible, though, to construct a simple, but analogous, problem for which the entire set of outcomes can be identified (and counted). If the proposed formula does not agree with the simple-case enumeration, we know that our analysis of the original question is incorrect.
3. If the outcomes to be counted fall into structurally different categories, the total number of outcomes will be the *sum* (not the product) of the number of outcomes in each category. Recall Example 2.6.5. Alternating melodies fall into two structurally-different categories: black keys can be the odd-numbered notes or they can be the even-numbered notes (there is no third possibility). Associated with each category is a different set of outcomes, implying that the total number of alternating melodies is the *sum* of the numbers of outcomes associated with the two categories.

QUESTIONS

2.6.1. A chemical engineer wishes to observe the effects of temperature, pressure, and catalyst concentration on the yield resulting from a certain reaction. If she intends to include two different temperatures, three pressures, and two levels of catalyst, how many different runs must she make in order to observe each temperature-pressure-catalyst combination exactly twice?

2.6.2. A coded message from a CIA operative to his Russian KGB counterpart is to be sent in the form Q4ET, where the first and last entries must be consonants; the second, an integer 1 through 9; and the third, one of the six vowels. How many different ciphers can be transmitted?

2.6.3. How many terms will be included in the expansion of

$$(a + b + c)(d + e + f)(x + y + u + v + w)$$

Which of the following will be included in that number: *aeu, cdx, bef, xvw*?

2.6.4. Suppose that the format for license plates in a certain state is two letters followed by four numbers.

- **(a)** How many different plates can be made?
- **(b)** How many different plates are there if the letters can be repeated but no two numbers can be the same?
- **(c)** How many different plates can be made if repetitions of numbers and letters is allowed except that no plate can have four zeros?

2.6.5. How many integers between 100 and 999 have distinct digits, and how many of those are odd numbers?

2.6.6. A fast-food restaurant offers customers a choice of eight toppings that can be added to a hamburger. How many different hamburgers can be ordered?

2.6.7. In baseball there are twenty-four different "base-out" configurations (runner on first—two outs, bases loaded—none out, and so on). Suppose that a new game, sleazeball, is played where there are seven bases (excluding home plate) and each team gets five outs an inning. How many base-out configurations would be possible in sleazeball?

2.6.8. When they were first introduced, postal zip codes were five-digit numbers, theoretically ranging from 00000 to 99999. (In reality, the lowest zip code was 00601 for San Juan, Puerto Rico; the highest was 99950 for Ketchikan, Alaska.) An additional four digits have recently been added, so each zip code is now a nine-digit number. How many zip codes are at least as large as 60000–0000, are even numbers, and have a seven as their third digit?

2.6.9. A restaurant offers a choice of four appetizers, fourteen entrees, six desserts, and five beverages. How many different meals are possible if a diner intends to order only three courses? (Consider the beverage to be a "course.")

2.6.10. Proteins are chains of molecules chosen (with repetition) from some 20 different amino acids. In a living cell, proteins are synthesized through the *genetic code*, a mechanism whereby ordered sequences of nucleotides in the messenger RNA dictate the formation of a particular amino acid. The four key nucleotides are adenine, guanine, cytosine, and uracil (A, G, C, and U). Assuming A, G, C, or U can appear any number of times in a nucleotide chain and that all sequences are physically possible, what is the minimum length the chains must attain to have the capability of encoding the entire set of amino acids? *Note:* Each sequence in the genetic code must have the same number of nucleotides.

2.6.11. Residents of a condominium have an automatic garage door opener that has a row of eight buttons. Each garage door has been programmed to respond to a particular set of buttons being pushed. If the condominium houses 250 families, can residents be assured that no two garage doors will open on the same signal? If so, how many additional families can be added before the eight-button code becomes inadequate? *Note:* The order in which the buttons are pushed is irrelevant.

2.6.12. In international Morse code, each letter in the alphabet is symbolized by a series of dots and dashes: the letter "a," for example, is encoded as ".–". What is the maximum number of dots and/or dashes needed to represent any letter in the English alphabet?

2.6.13. The decimal number corresponding to a sequence of n binary digits $a_0, a_1 \ldots, a_{n-1}$, where each a_i is either 0 or 1, is defined to be

$$a_0 2^0 + a_1 2^1 + \cdots + a_{n-1} 2^{n-1}$$

For example, the sequence 0 1 1 0 is equal to 6 $(= 0 \cdot 2^0 + 1 \cdot 2^1 + 1 \cdot 2^2 + 0 \cdot 2^3)$. Suppose a fair coin is tossed nine times. Replace the resulting sequence of H's and T's with a binary sequence of 1's and 0's (1 for H, 0 for T). For how many sequences of tosses will the decimal corresponding to the observed set of heads and tails exceed 256?

2.6.14. Given the letters in the word

ZOMBIES

in how many ways can two of the letters be arranged such that one is a vowel and one is a consonant?

2.6.15. Suppose that two cards are drawn—in order—from a standard 52-card poker deck. In how many ways can one of the cards be a club and one of the cards be an ace?

2.6.16. Monica's vacation plans require that she fly from Nashville to Chicago to Seattle to Anchorage. According to her travel agent, there are three available flights from Nashville to Chicago, five from Chicago to Seattle, and two from Seattle to Anchorage. Assume that the numbers of options she has for return flights are the same. How many round-trip itineraries can she schedule?

Counting Permutations (when the objects are all distinct)

Ordered sequences arise in two fundamentally different ways. The first is the scenario addressed by the multiplication rule—a process is comprised of k operations, each allowing n_i options, $i = 1, 2, \ldots, k$; choosing one version of each operation leads to $n_1 n_2 \ldots n_k$ possibilities.

The second occurs when an ordered arrangement of some specified length k is formed from a finite collection of objects. Any such arrangement is referred to as a *permutation of length k*. For example, given the three objects A, B, and C, there are six different permutations of length two that can be formed if the objects cannot be repeated: AB, AC, BC, BA, CA, and CB.

Theorem 2.6.1. *The number of permutations of length k that can be formed from a set of n distinct elements, repetitions not allowed, is denoted by the symbol* ${}_nP_k$, *where*

$$ {}_nP_k = n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!} $$

Proof. Any of the n objects may occupy the first position in the arrangement, any of $n-1$ the second, and so on—the number of choices available for filling the kth position will be $n-k+1$ (see Figure 2.6.6). The theorem follows, then, from the multiplication rule: There will be $n(n-1)\cdots(n-k+1)$ ordered arrangements. ☐

Corollary. *The number of ways to permute an entire set of n distinct objects is* ${}_nP_n = n(n-1)(n-2)\cdots 1 = n!$.

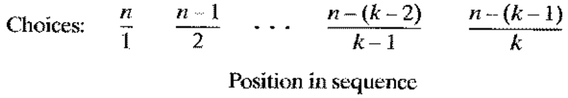


FIGURE 2.6.6

EXAMPLE 2.6.6

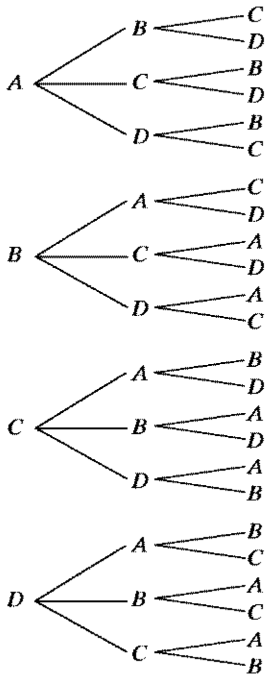
How many permutations of length $k = 3$ can be formed from the set of $n = 4$ distinct elements, A, B, C, and D?

According to Theorem 2.6.1, the number should be 24:

$$\frac{n!}{(n-k)!} = \frac{4!}{(4-3)!} = \frac{4 \cdot 3 \cdot 2 \cdot 1}{1} = 24$$

Confirming that figure, Table 2.6.2 lists the entire set of 24 permutations and illustrates the argument used in the proof of the theorem.

TABLE 2.6.2



1. (ABC)
2. (ABD)
3. (ACB)
4. (ACD)
5. (ADB)
6. (ADC)
7. (BAC)
8. (BAD)
9. (BCA)
10. (BCD)
11. (BDA)
12. (BDC)
13. (CAB)
14. (CAD)
15. (CBA)
16. (CBD)
17. (CDA)
18. (CDB)
19. (DAB)
20. (DAC)
21. (DBA)
22. (DBC)
23. (DCA)
24. (DCB)

EXAMPLE 2.6.7

In her sonnet with the famous first line, "How do I love thee? Let me count the ways," Elizabeth Barrett Browning listed eight. Suppose Ms. Browning had decided that writing greeting cards afforded her a better format for expressing her feelings. For how many years could she have corresponded with her favorite beau on a daily basis and never sent the same card twice? Assume that each card contains exactly four of the eight "ways" and that order matters.

In selecting the verse for a card, Ms. Browning would be creating a permutation of length $k = 4$ from a set of $n = 8$ distinct objects. According to Theorem 2.6.1,

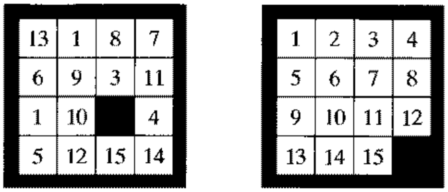
$$\begin{aligned}\text{Number of different cards} = {}_8P_4 = \frac{8!}{(8-4)!} &= 8 \cdot 7 \cdot 6 \cdot 5 \\ &= 1680\end{aligned}$$

At the rate of a card a day, she could keep the correspondence going for more than four and one-half years.

EXAMPLE 2.6.8

Years ago—long before Rubik cubes and electronic games had become epidemic—puzzles were much simpler. One of the more popular combinatorial-related diversions was a four by four grid consisting of fifteen movable squares and one empty space. The object was to maneuver as quickly as possible an arbitrary configuration (Figure 2.6.7a) into a specific pattern (Figure 2.6.7b). How many different ways could the puzzle be arranged?

Take the empty space to be square number 16 and imagine the four rows of the grid laid end to end to make a sixteen-digit sequence. Each permutation of that sequence corresponds to a different pattern for the grid. By the corollary to Theorem 2.6.1, the number of ways to position the tiles is 16!, or more than twenty trillion (20,922,789,888,000, to be exact). *That total is more than fifty times the number of stars in the entire Milky Way galaxy.* (*Note:* Not all of the 16! permutations can be generated without physically removing some of the tiles. Think of the two by two version of Figure 2.6.7 with tiles numbered 1 through 3. How many of the 4! theoretical configurations can actually be formed?)



13	1	8	7
6	9	3	11
1	10		4
5	12	15	14

(a)

1	2	3	4
5	6	7	8
9	10	11	12
13	14	15	

(b)

FIGURE 2.6.7

EXAMPLE 2.6.9

A deck of 52 cards is shuffled and dealt face up in a row. For how many arrangements will the four aces be adjacent?

This is a good example for illustrating the problem-solving benefits that come from drawing diagrams, as mentioned earlier. Figure 2.6.8 shows the basic structure that needs

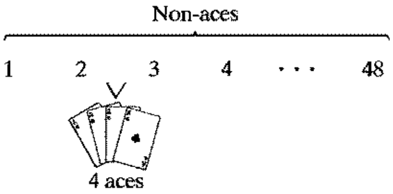


FIGURE 2.6.8

to be considered: The four aces are positioned as a "clump" somewhere between or around the forty-eight non-aces.

Clearly, there are forty-nine "spaces" that could be occupied by the four aces (in front of the first non-ace, between the first and second non-aces, and so on). Furthermore, by the corollary to Theorem 2.6.1, once the four aces are assigned to one of those forty-nine positions, they can still be permuted in ${}_4P_4 = 4!$ ways. Similarly, the forty-eight non-aces can be arranged in ${}_{48}P_{48} = 48!$ ways. It follows from the multiplication rule, then, that the number of arrangements having consecutive aces is the product, $49 \cdot 4! \cdot 48!$, or, approximately, 1.46×10^{64}.

Comment. Computing $n!$ can be quite cumbersome, even for n's that are fairly small: We saw in Example 2.6.8, for instance, that 16! is already in the trillions. Fortunately, an easy-to-use approximation is available. According to *Stirling's formula,*

$$n! \doteq \sqrt{2\pi}\, n^{n+1/2} e^{-n}$$

In practice, we apply Stirling's formula by writing

$$\log_{10}(n!) \doteq \log_{10}\left(\sqrt{2\pi}\right) + \left(n + \frac{1}{2}\right)\log_{10}(n) - n\log_{10}(e)$$

and then exponentiating the right-hand side.

Recall Example 2.6.9, where the number of arrangements was calculated to be $49 \cdot 4! \cdot 48!$, or $24 \cdot 49!$. Substituting into Stirling's formula, we can write

$$\begin{aligned}\log_{10}(49!) &\doteq \log_{10}\left(\sqrt{2\pi}\right) + \left(49 + \frac{1}{2}\right)\log_{10}(49) - 49\log_{10}(e)\\ &\approx 62.783366\end{aligned}$$

Therefore,

$$\begin{aligned}24 \cdot 49! &\doteq 24 \cdot 10^{62.78337}\\ &= 1.46 \times 10^{64}\end{aligned}$$

EXAMPLE 2.6.10

In chess a rook can move vertically and horizontally (see Figure 2.6.9). It can capture any unobstructed piece located anywhere in its own row or column. In how many ways can eight distinct rooks be placed on a chessboard (having eight rows and eight columns) so that no two can capture one another?

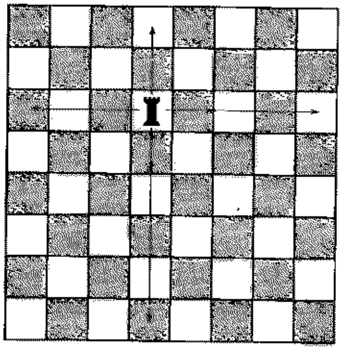


FIGURE 2.6.9

To start with a simpler problem, suppose that the eight rooks are all identical. Since no two rooks can be in the same row or same column (why?), it follows that each row must contain exactly one. The rook in the first row, however, can be in any of eight columns; the rook in the second row is then limited to being in one of seven columns, and so on. By the multiplication rule, then, the number of noncapturing configurations for eight identical rooks is $_8P_8$, or 8! (see Figure 2.6.10).

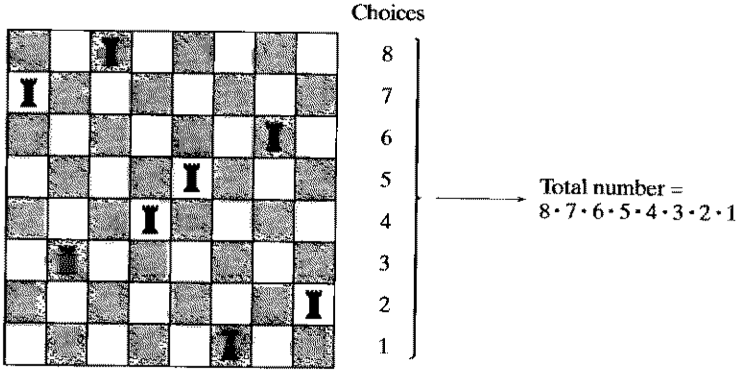


FIGURE 2.6.10

Now imagine the eight rooks to be distinct—they might be numbered, for example, 1 through 8. The rook in the first row could be marked with any of eight numbers; the rook in the second row with any of the remaining seven numbers; and so on. Altogether, there would be 8! numbering patterns *for each configuration*. The total number of ways to position eight distinct, noncapturing rooks, then, is 8! · 8!, or 1,625,702,400.

EXAMPLE 2.6.11

A new horror movie, *Friday the 13th, Part X*, stars Jason's great-grandson as a psychotic trying to dismember, decapitate, or do whatever else it takes to dispatch eight camp counselors, four men and four women. (a) How many scenarios (i.e., victim orders) can the screenwriters devise, assuming they want Jason to do away with all the men before going after any of the women? (b) How many scripts are possible if the only restriction imposed on Jason is that he save Muffy for last?

a. Suppose the male counselors are denoted A, B, C, and D, and the female counselors, W, X, Y, and Z. Among the admissible plots would be the sequence pictured in Figure 2.6.11, where B is done in first, then D, and so on. The men, if they are to be restricted to the first four positions, can still be permuted in ${}_4P_4 = 4!$ ways. The same number of arrangements can be found for the women. Furthermore, the plot in its entirety can be thought of as a two-step sequence: first the men are eliminated, then the women. Since 4! ways are available to do the former and 4! the latter, the total number of different scripts, by the multiplication rule, is 4! 4!, or *576*.

Men				Women			
B	D	A	C	Y	Z	W	X
1	2	3	4	5	6	7	8

Order of killing

FIGURE 2.6.11

b. If the only condition to be met is that Muffy be dealt with last, the number of admissible scripts is simply ${}_7P_7 = 7!$, that being the number of ways to permute the other seven counselors (see Figure 2.6.12).

B	W	Z	C	Y	A	D	Muffy
1	2	3	4	5	6	7	8

Order of killing

FIGURE 2.6.12

EXAMPLE 2.6.12

Consider the set of nine-digit numbers that can be formed by rearranging without repetition the integers 1 through 9. For how many of those permutations will the 1 and the 2 precede the 3 and the 4? That is, we want to count sequences like 7 *2* 5 *1 3* 6 9 *4* 8 but not like 6 8 *1* 5 *4 2* 7 *3* 9.

At first glance, this seems to be a problem well beyond the scope of Theorem 2.6.1. With the help of a symmetry argument, though, its solution is surprisingly simple.

Think of just the digits 1 through 4. By the Corollary on page 93, those four numbers give rise to 4!(= 24) permutations. Of those 24, only 4—(1, 2, 3, 4), (2, 1, 3, 4), (1, 2, 4, 3), and (2, 1, 4, 3)—have the property that the 1 and the 2 come before the 3 and the 4. It follows that $\frac{4}{24}$ of the total number of nine-digit permutations should satisfy the condition being imposed on 1, 2, 3, and 4. Therefore,

$$\begin{aligned}\text{Number of permutations where 1 and 2 precede 3 and 4} &= \frac{4}{24} \cdot 9! \\ &= 60{,}480\end{aligned}$$

QUESTIONS

2.6.17. The board of a large corporation has six members willing to be nominated for office. How many different "president/vice president/treasurer" slates could be submitted to the stockholders?

2.6.18. How many ways can a set of four tires be put on a car if all the tires are interchangeable? How many ways are possible if two of the four are snow tires?

2.6.19. Use Stirling's formula to approximate 30!.
(*Note:* The exact answer is 265,252,859,812,268,935,315, 188,480,000,000.)

2.6.20. The nine members of the music faculty baseball team, the Mahler Maulers, are all incompetent, and each can play any position equally poorly. In how many different ways can the Maulers take the field?

2.6.21. A three-digit number is to be formed from the digits 1 through 7, with no digit being used more than once. How many such numbers would be less than 289?

2.6.22. Four men and four women are to be seated in a row of chairs numbered 1 through 8.
- **(a)** How many total arrangements are possible?
- **(b)** How many arrangements are possible if the men are required to sit in alternate chairs?

2.6.23. An engineer needs to take three technical electives sometime during his final four semesters. The three are to be selected from a list of ten. In how many ways can he schedule those classes, assuming that he never wants to take more than one technical elective in any given term?

2.6.24. How many ways can a twelve-member cheerleading squad (six men and six women) pair up to form six male-female teams? How many ways can six male-female teams be positioned along a sideline? What might the number $6!6!2^6$ represent? What might the number $6!6!2^6 2^{12}$ represent?

2.6.25. Suppose that a seemingly interminable German opera is recorded on all six sides of a three-record album. In how many ways can the six sides be played so that at least one is out of order?

2.6.26. A group of n families, each with m members, are to be lined up for a photograph. In how many ways can the nm people be arranged if members of a family must stay together?

2.6.27. Suppose that ten people, including you and a friend, line up for a group picture. How many ways can the photographer rearrange the line if she wants to keep exactly three people between you and your friend?

2.6.28. Theorem 2.6.1 was the first mathematical result known to have been proved by induction, that feat being accomplished in 1321 by Levi ben Gerson. Assume that we do not know the multiplication rule. Prove the theorem the way Levi ben Gerson did.

2.6.29. In how many ways can a pack of fifty-two cards be dealt to thirteen players, four to each, so that every player has one card of each suit?

2.6.30. If the definition of $n!$ is to hold for all nonnegative integers n, show that it follows that $0!$ must equal one.

2.6.31. The crew of Apollo 17 consisted of a pilot, a copilot, and a geologist. Suppose that NASA had actually trained nine aviators and four geologists as candidates for the flight. How many different crews could they have assembled?

2.6.32. Uncle Harry and Aunt Minnie will both be attending your next family reunion. Unfortunately, they hate each other. Unless they are seated with at least two people between them, they are likely to get into a shouting match. The side of the table at which they will be seated has seven chairs. How many seating arrangements are available for those seven people if a safe distance is to be maintained between your aunt and your uncle?

2.6.33. In how many ways can the digits 1 through 9 be arranged such that

- **(a)** all the even digits precede all the odd digits
- **(b)** all the even digits are adjacent to each other
- **(c)** two even digits begin the sequence and two even digits end the sequence
- **(d)** the even digits appear in either ascending or descending order?

Counting Permutations (when the objects are not all distinct)

The corollary to Theorem 2.6.1 gives a formula for the number of ways an entire set of n objects can be permuted *if the objects are all distinct.* Fewer than $n!$ permutations are possible, though, if some of the objects are identical. For example, there are $3! = 6$ ways to permute the three distinct objects A, B, and C:

$$\begin{gathered} ABC \\ ACB \\ BAC \\ BCA \\ CAB \\ CBA \end{gathered}$$

If the three objects to permute, though, are A, A, and B—that is, if two of the three are identical—the number of permutations decreases to three:

$$\begin{gathered} AAB \\ ABA \\ BAA \end{gathered}$$

As we will see, there are many real-world applications where the n objects to be permuted belong to r different categories, each category containing one or more identical objects.

Theorem 2.6.2. *The number of ways to arrange n objects, n_1 being of one kind, n_2 of a second kind, …, and n_r of an rth kind, is*

$$\frac{n!}{n_1!n_2!\cdots n_r!}$$

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

Proof. Let N denote the total number of such arrangements. For any one of those N, the similar objects (if they were actually different) could be arranged in $n_1!n_2!\cdots n_r!$ ways. (Why?) It follows that $N \cdot n_1!n_2!\cdots n_r!$ is the total number of ways to arrange n (distinct) objects. But $n!$ equals that same number. Setting $N \cdot n_1!n_2!\cdots n_r!$ equal to $n!$ gives the result. □

Comment. Ratios like $n!/(n_1!n_2!\cdots n_r!)$ are called *multinomial coefficients* because the general term in the expansion of

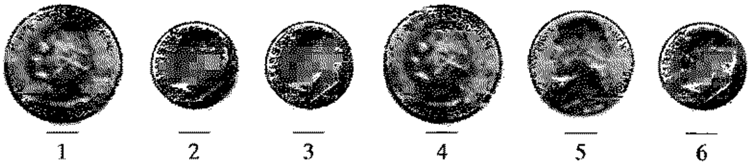
$$(x_1 + x_2 + \cdots + x_r)^n$$

is

$$\frac{n!}{n_1!n_2!\cdots n_r!}x_1^{n_1}x_2^{n_2}\cdots x_r^{n_r}$$

EXAMPLE 2.6.13

A pastry in a vending machine costs 85¢. In how many ways can a customer put in two quarters, three dimes, and one nickel?



Order in which coins are deposited

FIGURE 2.6.13

If all coins of a given value are considered identical, then a typical deposit sequence, say *QDDQND* (see Figure 2.6.13), can be thought of as a permutation of $n = 6$ objects belonging to $r = 3$ categories, where

$$n_1 = \text{number of nickels} = 1$$
$$n_2 = \text{number of dimes} = 3$$
$$n_3 = \text{number of quarters} = 2$$

By Theorem 2.6.2, there are sixty such sequences:

$$\frac{n!}{n_1!n_2!n_3!} = \frac{6!}{1!3!2!} = 60$$

Of course, had we assumed the coins were distinct (having been minted at different places and different times), the number of distinct permutations would be 6!, or 720.

EXAMPLE 2.6.14

Prior to the seventeenth century there were no scientific journals, a state of affairs that made it difficult for researchers to document discoveries. If a scientist sent a copy of his work to a colleague, there was always a risk that the colleague might claim it as his own. The obvious alternative—wait to get enough material to publish a book—invariably resulted in lengthy delays. So, as a sort of interim documentation, scientists would sometimes send each other anagrams—letter puzzles that, when properly unscrambled, summarized in a sentence or two what had been discovered.

When Christiaan Huygens (1629–1695) looked through his telescope and saw the ring around Saturn, he composed the following anagram (203):

$$\begin{gathered} aaaaaaa,\ ccccc,\ d,\ eeeee,\ g,\ h,\ iiiiiii,\ llll,\ mm, \\ nnnnnnnnn,\ oooo,\ pp,\ q,\ rr,\ s,\ ttttt,\ uuuuu \end{gathered}$$

How many ways can the sixty-two letters in Huygens's anagram be arranged?

Let $n_1 (= 7)$ denote the number of a's, $n_2 (= 5)$ the number of c's, and so on. Substituting into the appropriate multinomial coefficient, we find

$$N = \frac{62!}{7!5!1!5!1!1!7!4!2!9!4!2!1!2!1!5!5!}$$

as the total number of arrangements. To get a feeling for the magnitude of N, we need to apply Stirling's formula to the numerator. Since

$$62! \doteq \sqrt{2\pi}\, e^{-62} 62^{62.5}$$

then

$$\begin{aligned} \log(62!) &\doteq \log\left(\sqrt{2\pi}\right) - 62 \cdot \log(e) + 62.5 \cdot \log(62) \\ &\doteq 85.49731 \end{aligned}$$

The antilog of 85.49731 is 3.143×10^{85}, so

$$N \doteq \frac{3.143 \times 10^{85}}{7!5!1!5!1!1!7!4!2!9!4!2!1!2!1!5!5!}$$

is a number on the order of 3.6×10^{60}. Huygens was clearly taking no chances! (*Note* When appropriately rearranged, the anagram becomes "Annulo cingitur tenui, plano nusquam cohaerente, ad eclipticam inclinato," which translates to "Surrounded by a thir ring, flat, suspended nowhere, inclined to the ecliptic.")

EXAMPLE 2.6.15

What is the coefficient of x^{23} in the expansion of $(1 + x^5 + x^9)^{100}$?

To understand how this question relates to permutations, consider the simpler problem of expanding $(a + b)^2$:

$$\begin{aligned}(a + b)^2 &= (a + b)(a + b) \\ &= a \cdot a + a \cdot b + b \cdot a + b \cdot b \\ &= a^2 + 2ab + b^2\end{aligned}$$

Notice that each term in the first $(a + b)$ is multiplied by each term in the second $(a + b)$. Moreover, the coefficient that appears in front of each term in the expansion corresponds to the number of ways that that term can be formed. For example, the 2 in the term $2ab$ reflects the fact that the product ab can result from two different multiplications:

$$\underbrace{(a + b)(a + b)}_{ab} \qquad \text{or} \qquad (a + \underbrace{b)(a}_{ab} + b)$$

By analogy, the coefficient of x^{23} in the expansion of $(1 + x^5 + x^9)^{100}$ will be the number of ways that one term from each of the one hundred factors $(1 + x^5 + x^9)$ can be multiplied together to form x^{23}. The only factors that will produce x^{23}, though, is the set of two x^9's, one x^5, and ninety-seven 1's:

$$x^{23} = x^9 \cdot x^9 \cdot x^5 \cdot 1 \cdot 1 \cdots 1$$

It follows that the *coefficient* of x^{23} is the number of ways to permute two x^9's, one x^5, and ninety-seven 1's. So, from Theorem 2.6.2,

$$\begin{aligned}\text{coefficient of } x^{23} &= \frac{100!}{2!1!97!} \\ &= 485{,}100\end{aligned}$$

EXAMPLE 2.6.16

A palindrome is a phrase whose letters are in the same order whether they are read backward or forward, such as Napoleon's lament

Able was I ere I saw Elba

or the often cited

Madam, I'm Adam.

Words themselves can become the units in a palindrome, as in the sentence

Girl, bathing on Bikini, eyeing boy,
finds boy eyeing bikini on bathing girl.

Suppose the members of a set consisting of four objects of one type, six of a second type, and two of a third type are to be lined up in a row. How many of those permutations are palindromes?

Think of the twelve objects to arrange as being four A's, six B's, and two C's. If the arrangement is to be a palindrome, then half of the A's, half of the B's, and half of the C's must occupy the first six positions in the permutation. Moreover, the final six members of the sequence must be in the reverse order of the first six. For example, if the objects comprising the first half of the permutation were

$$C \quad A \quad B \quad A \quad B \quad B$$

then the last six would need to be in the order

$$B \quad B \quad A \quad B \quad A \quad C$$

It follows that the number of palindromes is the number of ways to permute the first six objects in the sequence, because once the first six are positioned, there is only one arrangement of the last six that will complete the palindrome. By Theorem 2.6.2, then,

$$\text{number of palindromes} = 6!/(2!3!1!) = 60$$

EXAMPLE 2.6.17

A deliveryman is currently at Point X and needs to stop at Point 0 before driving through to Point Y (see Figure 2.6.14). How many different routes can he take without ever going out of his way?

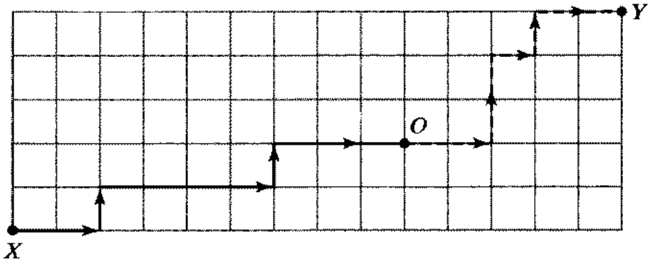


FIGURE 2.6.14

Notice that any admissible path from, say, X to 0 is an ordered sequence of 11 "moves"—nine East and two North. Pictured in Figure 2.6.14, for example, is the particular X to 0 route

$$E \quad E \quad N \quad E \quad E \quad E \quad E \quad N \quad E \quad E \quad E$$

Similarly, any acceptable path from 0 to Y will necessarily consist of five moves East and three moves North (the one indicated is $E\ E\ N\ N\ E\ N\ E\ E$).

Since each path from X to 0 corresponds to a unique permutation of nine E's and two N's, the *number* of such paths (from Theorem 2.6.2) is the quotient

$$11!/(9!2!) = 55$$

For the same reasons, the number of different paths from 0 to Y is

$$8!/(5!3!) = 56$$

By the Multiplication Rule, then, the total number of admissible routes from X to Y that pass through 0 is the product of 55 and 56, or *3080*.

QUESTIONS

2.6.34. Which state name can generate more permutations, TENNESSEE or FLORIDA?

2.6.35. How many numbers greater than 4,000,000 can be formed from the digits 2, 3, 4, 4, 5, 5, 5?

2.6.36. An interior decorator is trying to arrange a shelf containing eight books, three with red covers, three with blue covers, and two with brown covers.

- **(a)** Assuming the titles and the sizes of the books are irrelevant, in how many ways can she arrange the eight books?
- **(b)** In how many ways could the books be arranged if they were all considered distinct?
- **(c)** In how many ways could the books be arranged if the red books were considered indistinguishable, but the other five were considered distinct?

2.6.37. Four Nigerians (A, B, C, D), three Chinese (#, *, &), and three Greeks (α, β, γ) are lined up at the box office, waiting to buy tickets for the World's Fair.

- **(a)** How many ways can they position themselves if the Nigerians are to hold the first four places in line; the Chinese, the next three; and the Greeks, the last three?
- **(b)** How many arrangements are possible if members of the same nationality must stay together?
- **(c)** How many different queues can be formed?
- **(d)** Suppose a vacationing Martian strolls by and wants to photograph the ten for her scrapbook. A bit myopic, the Martian is quite capable of discerning the more obvious differences in human anatomy but is unable to distinguish one Nigerian (N) from another, one Chinese (C) from another, or one Greek (G) from another. Instead of perceiving a line to be $B * \beta AD\#\&C\alpha\gamma$, for example, she would see $NCGNNCCNGG$. From the Martian's perspective, in how many different ways can the ten funny-looking Earthlings line themselves up?

2.6.38. How many ways can the letters in the word

$$SLUMGULLION$$

be arranged so that the three L's precede all the other consonants?

2.6.39. A tennis tournament has a field of $2n$ entrants, all of whom need to be scheduled to play in the first round. How many different pairings are possible?

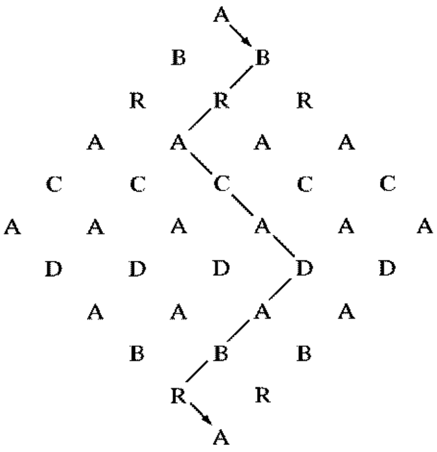
2.6.40. What is the coefficient of x^{12} in the expansion of $(1 + x^3 + x^6)^{18}$?

2.6.41. In how many ways can the letters of the word

ELEEMOSYNARY

be arranged so that the S is always immediately followed by a Y?

2.6.42. In how many ways can the word ABRACADABRA be formed in the array pictured below? Assume that the word must begin with the top A and progress diagonally downward to the bottom A.



2.6.43. Suppose a pitcher faces a batter who never swings. For how many different ball/strike sequences will the batter be called out on the fifth pitch?

2.6.44. What is the coefficient of $w^2x^3yz^3$ in the expansion of $(w + x + y + z)^9$?

2.6.45. Imagine six points in a plane, no three of which lie on a straight line. In how many ways can the six points be used as vertices to form two triangles? (*Hint:* Number the points 1 through 6. Call one of the triangles A and the other B. What does the permutation

$$\begin{array}{cccccc} A & A & B & B & A & B \\ 1 & 2 & 3 & 4 & 5 & 6 \end{array}$$

represent?)

2.6.46. Show that $(k!)!$ is divisible by $k!^{(k-1)!}$. (*Hint:* Think of a related permutation problem whose solution would require Theorem 2.6.2.)

2.6.47. In how many ways can the letters of the word

BROBDINGNAGIAN

be arranged without changing the order of the vowels?

2.6.48. Make an anagram out of the familiar expression STATISTICS IS FUN. In how many ways can the letters in the anagram be permuted?

2.6.49. Linda is taking a five-course load her first semester: English, math, French, psychology, and history. In how many different ways can she earn three A's and two B's? Enumerate the entire set of possibilities. Use Theorem 2.6.2 to verify your answer.

Counting Combinations

Order is not always a meaningful characteristic of a collection of elements. Consider a poker player being dealt a five-card hand. Whether he receives a 2 of hearts, 4 of clubs, 9 of clubs, jack of hearts, and ace of diamonds *in that order*, or in any one of the other $5! - 1$ permutations of those particular five cards is irrelevant—the hand is still the same. As the last set of examples in this section bear out, there are many such situations—problems where our only legitimate concern is with the composition of a set of elements, not with any particular arrangement.

We call a collection of k *unordered* elements a *combination of size* k. For example, given a set of $n = 4$ distinct elements—A, B, C, and D—there are *six* ways to form combinations of size 2:

A and B
A and C
A and D
B and C
B and D
C and D

A general formula for counting combinations can be derived quite easily from what we already know about counting permutations.

Theorem 2.6.3. *The number of ways to form combinations of size k from a set of n distinct objects, repetitions not allowed, is denoted by the symbols* $\binom{n}{k}$ *or* ${}_nC_k$, *where*

$$\binom{n}{k} = {}_nC_k = \frac{n!}{k!(n-k)!}$$

Proof. Let the symbol $\binom{n}{k}$ denote the number of combinations satisfying the conditions of the theorem. Since each of those combinations can be ordered in $k!$ ways, the product $k!\binom{n}{k}$ must equal the number of *permutations* of length k that can be formed from n distinct elements. But n distinct elements can be formed into permutations of length k in $n(n-1)\cdots(n-k+1) = n!/(n-k)!$ ways. Therefore,

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

Solving for $\binom{n}{k}$ gives the result. □

Comment. It often helps to think of combinations in the context of drawing objects out of an urn. If an urn contains n chips labeled 1 through n, the number of ways we can reach in and draw out different samples of size k is $\binom{n}{k}$. In deference to this sampling interpretation for the formation of combinations, $\binom{n}{k}$ is usually read "n things taken k at a time" or "n choose k."

Comment. The symbol $\binom{n}{k}$ appears in the statement of a familiar theorem from algebra,

$$(x + y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Since the expression being raised to a power involves two terms, x and y, the constants $\binom{n}{k}$, $k = 0, 1, \ldots, n$, are commonly referred to as *binomial coefficients*.

EXAMPLE 2.6.18

Eight politicians meet at a fund-raising dinner. How many greetings can be exchanged if each politician shakes hands with every other politician exactly once?

Imagine the politicians to be eight chips—1 through 8—in an urn. A handshake corresponds to an unordered sample of size 2 chosen from that urn. Since repetitions are not allowed (even the most obsequious and overzealous of campaigners would not shake hands with himself!), Theorem 2.6.3 applies, and the total number of handshakes is

$$\binom{8}{2} = \frac{8!}{2!6!}$$

or 28.

EXAMPLE 2.6.19

The basketball recruiter for Swampwater Tech has scouted sixteen former NBA starters that he thinks he can pass off as Junior College transfers—six are guards, seven are forwards, and three are centers. Unfortunately, his slush fund of illegal alumni donations is at an all-time low and he can afford to buy new Corvettes for only nine of the players. If he wants to keep three guards, four forwards, and two centers, how many ways can he parcel out the cars?

This is a combination problem that also requires an application of the multiplication rule. First, note there are $\binom{6}{3}$ *sets* of three guards that could be chosen to receive Corvettes (think of drawing a set of three names out of an urn containing six names). Similarly, the forwards and centers can be bribed in $\binom{7}{4}$ and $\binom{3}{2}$ ways, respectively. It follows from the multiplication rule, then, that the total number of ways to divvy up the cars is the product

$$\binom{6}{3} \cdot \binom{7}{4} \cdot \binom{3}{2}$$

or 2100 $(= 20 \cdot 35 \cdot 3)$.

EXAMPLE 2.6.20

Your statistics teacher announces a twenty-page reading assignment on Monday that is to be finished by Thursday morning. You intend to read the first x_1 pages Monday, the next x_2 pages Tuesday, and the final x_3 pages Wednesday, where $x_1 + x_2 + x_3 = 20$ and each $x_i \geq 1$. In how many ways can you complete the assignment? That is, how many different sets of values can be chosen for x_1, x_2, and x_3?

Imagine the nineteen spaces *between* the twenty pages (see Figure 2.6.15). Choosing any two of those spaces automatically partitions the twenty pages into three nonempty sets. Spaces 3 and 7, for example, would correspond to reading three pages on Monday, four pages on Tuesday, and thirteen pages on Wednesday. The number of different values for the set (x_1, x_2, x_3), then, must equal the number of ways to select two "markers"—namely, $\binom{19}{2}$, or *171*.

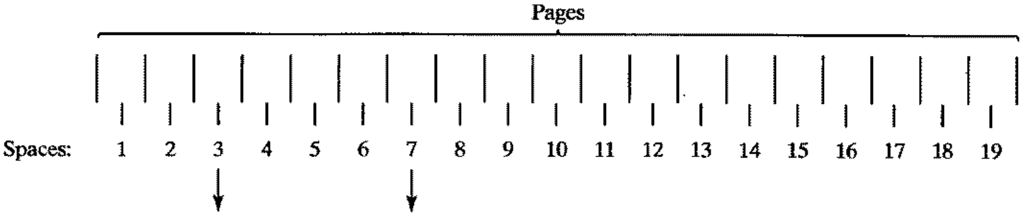


FIGURE 2.6.15

EXAMPLE 2.6.21

Mitch is trying to put a little zing into his cabaret act by telling four jokes at the beginning of each show. His current engagement is booked to run four months. If he gives one performance a night and never wants to repeat the same set of jokes on any two nights, what is the minimum number of jokes he needs in his repertoire?

Four months of performances create a demand for roughly 120 different sets of jokes. Let n denote the number of jokes that Mitch can tell. The question is asking for the smallest n for which $\binom{n}{4} \geq 120$. Trial-and-error calculations summarized in Table 2.6.3 show that the optimal n is surprisingly small: A set of only *nine* jokes is sufficient to keep Mitch from having to repeat his opening monologue.

TABLE 2.6.3

n	$\binom{n}{4}$	$\geq$120?
7	35	No
8	70	No
→ 9	126	Yes

EXAMPLE 2.6.22

Binomial coefficients have many interesting properties. Perhaps the most familiar is Pascal's triangle,[1] a numerical array where each entry is equal to the sum of the two numbers appearing diagonally above it (see Figure 2.6.16). Notice that each entry in Pascal's triangle can be expressed as a binomial coefficient, and the relationship just described appears to reduce to a simple equation involving those coefficients:

$$\binom{n+1}{k} = \binom{n}{k} + \binom{n}{k-1} \tag{2.6.1}$$

Prove that Equation 2.6.1 holds for all positive integers n and k.

	Row	
1	0	$\binom{0}{0}$
1 1	1	$\binom{1}{0}$ $\binom{1}{1}$
1 2 1	2	$\binom{2}{0}$ $\binom{2}{1}$ $\binom{2}{2}$
1 3 3 1	3	$\binom{3}{0}$ $\binom{3}{1}$ $\binom{3}{2}$ $\binom{3}{3}$
1 4 6 4 1	4	$\binom{4}{0}$ $\binom{4}{1}$ $\binom{4}{2}$ $\binom{4}{3}$ $\binom{4}{4}$
⋮ ⋮	⋮	⋮ ⋮

FIGURE 2.6.16

Consider a set of $n+1$ distinct objects $A_1, A_2, \ldots, A_{n+1}$. We can obviously draw samples of size k from that set in $\binom{n+1}{k}$ different ways. Now, consider any particular object—for example, A_1. Relative to A_1, each of those $\binom{n+1}{k}$ samples belongs to one of two categories: those containing A_1 and those not containing A_1. To form samples containing A_1, we need to select $k-1$ additional objects from the remaining n. This can be done in $\binom{n}{k-1}$ ways. Similarly, there are $\binom{n}{k}$ ways to form samples not containing A_1. Therefore, $\binom{n+1}{k}$ must equal $\binom{n}{k} + \binom{n}{k-1}$.

EXAMPLE 2.6.23

The answers to combinatorial questions can sometimes be obtained using quite different approaches. What invariably distinguishes one solution from another is the way in which outcomes are characterized.

For example, suppose you have just ordered a roast beef sub at a sandwich shop, and now you need to decide which, if any, of the available toppings (lettuce, tomato, onions,

---

[1] Despite its name, Pascal's triangle was not discovered by Pascal. Its basic structure was known hundreds of years before the French mathematician was born. It was Pascal, though, who first made extensive use of its properties.

Add?	Y	Y	Y	N	N	N	N	N
	Lettuce	Tomato	Onion	Mustard	Relish	Mayo	Pickles	Peppers

FIGURE 2.6.17

etc.) to add. If the store has eight "extras" to choose from, how many different subs can you order?

One way to answer this question is to think of each sub as an ordered sequence of length eight, where each position in the sequence corresponds to one of the toppings. At each of those positions, you have two choices—"add" or "do not add" that particular topping. Pictured in Figure 2.6.17 is the sequence corresponding to the sub that has lettuce, tomato, and onion but no other toppings. Since two choices ("add" or "do not add") are available for each of the eight toppings, the multiplication rule tells us that the number of different roast beef subs that would be requested is 2^8, or *256*.

An ordered sequence of length eight, though, is not the only model capable of characterizing a roast beef sandwich. We can also distinguish one roast beef sub from another by the particular *combination* of toppings that each one has. For example, there are $\binom{8}{4} = 70$ different subs having exactly four toppings. It follows that the total number of different sandwiches is the total number of different combinations of size k, where k ranges from 0 to 8. Reassuringly, that sum agrees with the ordered sequence answer:

$$\begin{aligned}\text{total number of different roast beef subs} &= \binom{8}{0} + \binom{8}{1} + \binom{8}{2} + \cdots + \binom{8}{8} \\ &= 1 + 8 + 28 + \cdots + 1 \\ &= 256\end{aligned}$$

What we have just illustrated here is another property of binomial coefficients—namely, that

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n \tag{2.6.2}$$

The proof of Equation 2.6.2 is a direct consequence of Newton's binomial expansion (see the second comment following Theorem 2.6.3).

QUESTIONS

2.6.50. How many straight lines can be drawn between five points (A, B, C, D, and E), no three of which are collinear?

2.6.51. The Alpha Beta Zeta sorority is trying to fill a pledge class of nine new members during fall rush. Among the twenty-five available candidates, fifteen have been judged marginally acceptable and ten highly desirable. How many ways can the pledge class be chosen to give a two-to-one ratio of highly desirable to marginally acceptable candidates?

2.6.52. A boat has a crew of eight: Two of those eight can row only on the stroke side, while three can row only on the bow side. In how many ways can the two sides of the boat be manned?

2.6.53. Nine students, five men and four women, interview for four summer internships sponsored by a city newspaper.

(a) In how many ways can the newspaper choose a set of four interns?

(b) In how many ways can the newspaper choose a set of four interns if it must include two men and two women in each set?

(c) How many sets of four can be picked such that not everyone in a set is of the same sex?

2.6.54. The final exam in History 101 consists of five essay questions that the professor chooses from a pool of seven that are given to the students a week in advance. For how many possible sets of questions does a student need to be prepared? In this situation does order matter?

2.6.55. Ten basketball players meet in the school gym for a pickup game. How many ways can they form two teams of five each?

2.6.56. A chemist is trying to synthesize part of a straight-chain aliphatic hydrocarbon polymer that consists of twenty-one radicals—ten ethyls (E), six methyls (M), and five propyls (P). Assuming all arrangements of radicals are physically possible, how many different polymers can be formed if no two of the methyl radicals are to be adjacent?

2.6.57. In how many ways can the letters in

$$MISSISSIPPI$$

be arranged so that no two I's are adjacent?

2.6.58. Prove that $\sum_{k=0}^{n}\binom{n}{k} = 2^n$. *Hint:* Use the binomial expansion mentioned on page 108.

2.6.59. Prove that

$$\binom{n}{0}^2 + \binom{n}{1}^2 + \cdots + \binom{n}{n}^2 = \binom{2n}{n}$$

(*Hint:* Rewrite the left-hand side as

$$\binom{n}{0}\binom{n}{n} + \binom{n}{1}\binom{n}{n-1} + \binom{n}{2}\binom{n}{n-2} + \cdots$$

and consider the problem of selecting a sample of n objects from an original set of $2n$ objects.)

2.6.60. Show that

$$\binom{n}{1} + \binom{n}{3} + \cdots = \binom{n}{0} + \binom{n}{2} + \cdots$$

(*Hint:* Consider the expansion of $(x - y)^n$.)

2.6.61. Prove that successive terms in the sequence $\binom{n}{0}, \binom{n}{1}, \ldots, \binom{n}{n}$ first increase and then decrease. (*Hint:* Examine the ratio of two successive terms, $\binom{n}{j+k}\Big/\binom{n}{j}$.)

2.6.62. Imagine n molecules of a gas confined to a rigid container divided into two chambers by a semipermeable membrane. If i molecules are in the left chamber, the *entropy* of the system is defined by the equation

$$\text{Entropy} = \log \binom{n}{i}$$

If n is even, for what configuration of molecules will the entropy be maximized? (Entropy is a concept physicists find useful in characterizing heat exchanges, particularly those involving gases. In general terms, the entropy of a system is a measure of its disorder: As the "randomness" of the position and velocity vectors of a system of particles increases, so does its entropy.) (*Hint:* See Question 2.6.61.)

2.6.63. Compare the coefficients of t^k in $(1 + t)^d(1 + t)^e = (1 + t)^{d+e}$ to prove that

$$\sum_{j=0}^{k} \binom{d}{j}\binom{e}{k-j} = \binom{d+e}{k}$$

COMBINATORIAL PROBABILITY

In Section 2.6 our concern focused on counting the number of ways a given operation, or sequence of operations, could be performed. In Section 2.7 we want to couple those enumeration results with the notion of probability. Putting the two together makes a lot of sense—there are many combinatorial problems where an enumeration, by itself, is not particularly relevant. A poker player, for example, is not interested in knowing the total *number* of ways he can draw to a straight; he *is* interested, though, in his *probability* of drawing to a straight.

In a combinatorial setting, making the transition from an enumeration to a probability is easy. If there are n ways to perform a certain operation and a total of m of those satisfy some stated condition—call it A—then $P(A)$ is defined to be the ratio, m/n. This assumes, of course, that all possible outcomes are equally likely.

Historically, the "m over n" idea is what motivated the early work of Pascal, Fermat, and Huygens (recall Section 1.1). Today we recognize that not all probabilities are so easily characterized. Nevertheless, the m/n model—the so-called *classical* definition of probability—is entirely appropriate for describing a wide variety of phenomena.

EXAMPLE 2.7.1

An urn contains eight chips, numbered 1 through 8. A sample of three is drawn without replacement. What is the probability that the largest chip in the sample is a 5?

Let A be the event "Largest chip in sample is a 5." Figure 2.7.1 shows what must happen in order for A to occur: (1) the 5 chip must be selected, and (2) two chips must be drawn from the subpopulation of chips numbered 1 through 4. By the multiplication rule, the number of samples satisfying event A is the product $\binom{1}{1} \cdot \binom{4}{2}$.

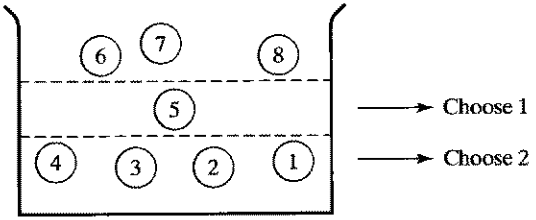


FIGURE 2.7.1

The sample space S for the experiment of drawing three chips from the urn contains $\binom{8}{3}$ outcomes, all equally likely. In this situation, then, $m = \binom{1}{1} \cdot \binom{4}{2}$, $n = \binom{8}{3}$, and

$$P(A) = \frac{\binom{1}{1} \cdot \binom{4}{2}}{\binom{8}{3}}$$
$$= 0.11$$

EXAMPLE 2.7.2

An urn contains n red chips numbered 1 through n, n white chips numbered 1 through n, and n blue chips numbered 1 through n (see Figure 2.7.2). Two chips are drawn at random and without replacement. What is the probability that the two drawn are either the same color or the same number?

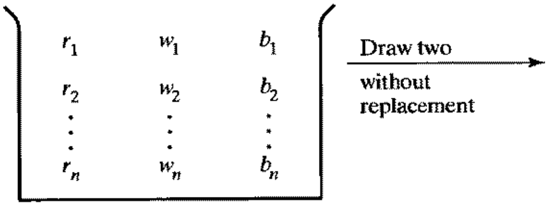


FIGURE 2.7.2

Let A be the event that the two chips drawn are the same color; let B be the event that they have the same number. We are looking for $P(A \cup B)$.

Since A and B here are mutually exclusive,

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

With $3n$ chips in the urn, the total number of ways to draw an unordered sample of size two is $\binom{3n}{2}$. Moreover,

$$\begin{aligned} P(A) &= P(2\,\text{reds} \cup 2\,\text{whites} \cup 2\,\text{blues}) \\ &= P(2\,\text{reds}) + P(2\,\text{whites}) + P(2\,\text{blues}) \\ &= 3\binom{n}{2} \Big/ \binom{3n}{2} \end{aligned}$$

and

$$P(B) = P(\text{two } 1\text{'s} \cup \text{ two } 2\text{'s} \cup \cdots \cup \text{ two } n\text{'s})$$
$$= n\binom{3}{2} \Big/ \binom{3n}{2}$$

Therefore,

$$P(A \cup B) = \frac{3\binom{n}{2} + n\binom{3}{2}}{\binom{3n}{2}}$$
$$= \frac{n+1}{3n-1}$$

EXAMPLE 2.7.3

Twelve fair dice are rolled. What is the probability that

a. the first six dice all show one face and the last six dice all show a second face?

b. not all the faces are the same?

c. each face appears exactly twice?

a. The sample space that corresponds to the "experiment" of rolling twelve dice is the set of ordered sequences of length twelve, where the outcome at every position in the sequence is one of the integers 1 through 6. If the dice are fair, all 6^{12} such sequences are equally likely.

Let A be the set of rolls where the first six dice show one face and the second six show another face. Figure 2.7.3 shows one of the sequences in the event A. Clearly, the face that appears for the first half of the sequence could be any of the six integers from 1 through 6.

Faces

2	2	2	2	2	2	4	4	4	4	4	4
1	2	3	4	5	6	7	8	9	10	11	12

Position in sequence

FIGURE 2.7.3

Five choices would be available for the last half of the sequence (since the two faces cannot be the same). The number of sequences in the event A, then, is ${}_6P_2 = 6 \cdot 5 = 30$. Applying the "*m/n*" rule gives

$$P(A) = 30/6^{12} = 1.4 \times 10^{-8}$$

b. Let B be the event that not all the faces are the same. Then

$$P(B) = 1 - P(B^C)$$
$$= 1 - 6/12^6$$

since there are six sequences—(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,), . . . , (6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6,)—where the twelve faces *are* all the same.

c. Let C be the event that each face appears exactly twice. From Theorem 2.6.2, the *number* of ways each face can appear exactly twice is $12!/(2! \cdot 2! \cdot 2! \cdot 2! \cdot 2! \cdot 2!)$. Therefore,

$$\begin{aligned} P(C) &= \frac{12!/(2! \cdot 2! \cdot 2! \cdot 2! \cdot 2! \cdot 2!)}{6^{12}} \\ &= 0.0034 \end{aligned}$$

EXAMPLE 2.7.4

A fair die is tossed n times. What is the probability that the sum of the faces showing is $n + 2$?

The sample space associated with rolling a die n times has 6^n outcomes, all of which in this case are equally likely because the die is presumed fair. There are two "types" of outcomes that will produce a sum of $n + 2$—(a) $n - 1$ 1s and one 3 and (b) $n - 2$ 1s and two 2s (see Figure 2.7.4). By Theorem 2.6.2 the number of sequences having $n - 1$ 1's and one 3 is $\frac{n!}{1!(n-1)!} = n$; likewise, there are $\frac{n!}{2!(n-2)!} = \binom{n}{2}$ outcomes having $n - 2$ 1s and two 2s. Therefore,

$$P(\text{sum} = n + 2) = \frac{n + \binom{n}{2}}{6^n}$$

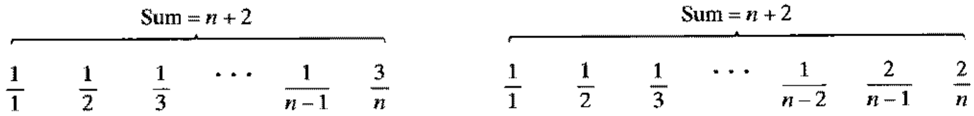


FIGURE 2.7.4

EXAMPLE 2.7.5

To keep the monkey entertained, Tarzan gives Cheetah the following letters from a Scrabble set to play with:

$$AAA \quad EE \quad I \quad J \quad K \quad L \quad NN \quad R \quad T \quad Z$$

What is the probability that Cheetah (who can't spell) rearranges the letters at random and forms the following sequence:

TARZAN LIKE JANE

(Ignore the spaces between the words).

If similar letters are considered indistinguishable, Theorem 2.6.2 applies, and the total number of ways to arrange the fourteen letters is 14!/(3!2!1!1!1!1!2!1!1!1!), or 3,632,428,800. Only one of those sequences is the desired arrangement, so

$$P(\text{“TARZAN LIKE JANE”}) = \frac{1}{3{,}632{,}428{,}800}$$

Notice that the same answer is obtained if the fourteen tiles are considered distinct. Under that scenario, the total number of permutations is 14!, but the number of ways to spell TARZAN LIKE JANE increases to 3!2!2!, because all the A's, E's, and N's can be permuted. Therefore,

$$P(\text{“TARZAN LIKE JANE”}) = \frac{3!2!2!}{14!} = \frac{1}{3{,}632{,}428{,}800}$$

EXAMPLE 2.7.6

Suppose that k people are selected at random from the general population. What are the chances that at least two of those k were born on the same day? Known as the *birthday problem,* this is a particularly intriguing example of combinatorial probability because its statement is so simple, its analysis is straightforward, yet its solution, as we will see, goes strongly contrary to our intuition.

Picture the k individuals lined up in a row to form an ordered sequence. If leap year is omitted, each person might have any of 365 birthdays. By the multiplication rule, the group as a whole generates a sample space of 365^k birthday sequences (see Figure 2.7.5).

Define A to be the event "at least two people have the same birthday." If each person is assumed to have the same chance of being born on any given day, the 365^k sequences in Figure 2.7.5 are equally likely, and

$$P(A) = \frac{\text{Number of sequences in } A}{365^k}$$

Counting the number of sequences in the numerator here is prohibitively difficult because of the complexity of the event A; fortunately, counting the number of sequences in A^c is quite easy. Notice that each birthday sequence in the sample space belongs to exactly one of two categories (see Figure 2.7.6):

1. At least two people have the same birthday.
2. All k people have different birthdays.

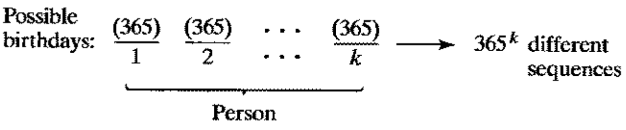


FIGURE 2.7.5

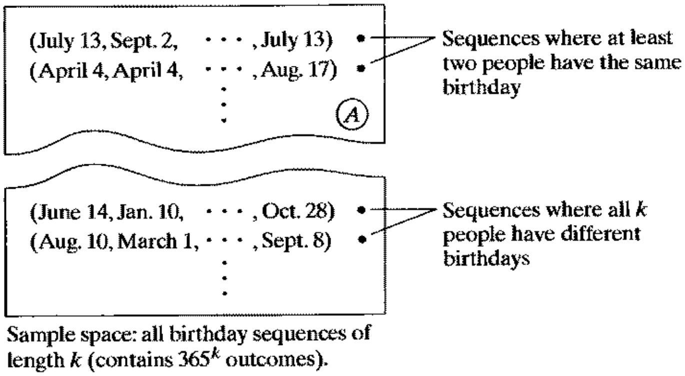


FIGURE 2.7.6

It follows that

$$\text{Number of sequences in } A = 365^k - \text{number of sequences where all } k \text{ people have different birthdays}$$

The number of ways to form birthday sequences for k people subject to the restriction that all k birthdays must be different is simply the number of ways to form permutations of length k from a set of 365 distinct objects:

$$_{365}P_k = 365(364)\cdots(365 - k + 1)$$

Therefore,

$$P(A) = P(\text{at least two people have the same birthday})$$
$$= \frac{365^k - 365(364)\cdots(365 - k + 1)}{365^k}$$

Table 2.7.1 shows $P(A)$ for k values of 15, 22, 23, 40, 50, and 70. Notice how the $P(A)$'s greatly exceed what our intuition would suggest.

Comment. Presidential biographies offer one opportunity to "confirm" the unexpectedly large values that Table 2.7.1 gives for $P(A)$. Among our first $k = 40$ presidents, two did have the same birthday: Harding and Polk were both born on November 2. More

TABLE 2.7.1

k	$P(A) = P$ (at least two have same birthday)
15	0.253
22	0.476
23	0.507
40	0.891
50	0.970
70	0.999

surprising, though, are the death dates of the presidents: Adams, Jefferson, and Monroe all died on July 4, and Fillmore and Taft both died on March 8.

Comment. The values for $P(A)$ in Table 2.7.1 are actually slight *underestimates* for the true probabilities that at least two of k people will be born on the same day. The assumption made earlier that all 365^k birthday sequences are equally likely is not entirely true: Births are somewhat more common during the summer than they are during the winter. It has been proven, though, that any sort of deviation from the equally-likely model will only serve to *increase* the chances that two or more people will share the same birthday (120). So, if $k = 40$, for example, the probability is slightly greater than 0.891 that at least two were born on the same day.

EXAMPLE 2.7.7

One of the more instructive—and to some, one of the more useful—applications of combinatorics is the calculation of probabilities associated with various poker hands. It will be assumed in what follows that five cards are dealt from a poker deck and that no other cards are showing, although some may already have been dealt. The sample space is the set of $\binom{52}{5} = 2{,}598{,}960$ different hands, each having probability $1/2{,}598{,}960$. What are the chances of being dealt (a) a *full house,* (b) *one pair,* and (c) a *straight?* [Probabilities for the various other kinds of poker hands (two pairs, three-of-a-kind, flush, and so on) are gotten in much the same way.]

a. *Full house.* A full house consists of three cards of one denomination and two of another. Figure 2.7.7 shows a full house consisting of three 7s and two Queens. Denominations for the three-of-a-kind can be chosen in $\binom{13}{1}$ ways. Then, given that a denomination has been decided on, the three requisite suits can be selected in $\binom{4}{3}$ ways. Applying the same reasoning to the pair gives $\binom{12}{1}$ available denominations, each having $\binom{4}{2}$ possible choices of suits. Thus, by the multiplication rule,

$$P(\text{full house}) = \frac{\binom{13}{1}\binom{4}{3}\binom{12}{1}\binom{4}{2}}{\binom{52}{5}} = 0.00144$$

	2	3	4	5	6	7	8	9	10	J	Q	K	A
D													
H						×					×		
C						×							
S						×					×		

FIGURE 2.7.7

	2	3	4	5	6	7	8	9	10	J	Q	K	A
D			×										×
H					×		×						
C					×								
S													

FIGURE 2.7.8

b. *One pair.* To qualify as a one-pair hand, the five cards must include two of the same denomination and three "single" cards—cards whose denominations match neither the pair nor each other. Figure 2.7.8 shows a pair of 6's. For the pair, there are $\binom{13}{1}$ possible denominations and, once selected, $\binom{4}{2}$ possible suits. Denominations for the three single cards can be chosen $\binom{12}{3}$ ways (see Question 2.7.16), and each card can have any of $\binom{4}{1}$ suits. Multiplying these factors together and dividing by $\binom{52}{2}$ gives a probability of 0.42:

$$P(\text{one pair}) = \frac{\binom{13}{1}\binom{4}{2}\binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1}}{\binom{52}{5}} = 0.42$$

c. *Straight.* A straight is five cards having consecutive denominations *but not all in the same suit*—for example, a 4 of diamonds, 5 of hearts, 6 of hearts, 7 of clubs, and 8 of diamonds (see Figure 2.7.9). An ace may be counted "high" or "low," which means that (10, jack, queen, king, ace) is a straight and so is (ace, 2, 3, 4, 5). (If five consecutive cards are all in the same suit, the hand is called a *straight flush*. The latter is considered a fundamentally different type of hand in the sense that a straight flush "beats" a straight.) To get the numerator for $P(\text{straight})$, we will first ignore the condition that all five cards not be in the same suit and simply count the number of hands having consecutive denominations. Note there are ten sets of consecutive denominations of length five: (ace, 2, 3, 4, 5), (2, 3, 4, 5, 6), ..., (10, jack, queen, king, ace). With no restrictions on the suits, each card can be either a diamond, heart, club, or spade. It follows, then, that the number of five-card hands having consecutive denominations is $10 \cdot \binom{4}{1}^5$. But forty $(= 10 \cdot 4)$ of those hands are straight flushes. Therefore,

$$P(\text{straight}) = \frac{10 \cdot \binom{4}{1}^5 - 40}{\binom{52}{5}} = 0.00392$$

Table 2.7.2 shows the probabilities associated with all the different poker hands. Hand i beats hand j if $P(\text{hand } i) < P(\text{hand } j)$.

	2	3	4	5	6	7	8	9	10	J	Q	K	A
D			×				×						
H				×	×								
C						×							
S													

FIGURE 2.7.9

TABLE 2.7.2

Hand	Probability
One pair	0.42
Two pairs	0.048
Three-of-a-kind	0.021
Straight	0.0039
Flush	0.0020
Full house	0.0014
Four-of-a-kind	0.00024
Straight flush	0.000014
Royal flush	0.0000015

PROBLEM-SOLVING HINTS
(Doing combinatorial probability problems)

Listed on p. 91 are several hints that can be helpful in counting the number of ways to do something. Those same hints apply to the solution of combinatorial *probability* problems, but a few others should be kept in mind as well.

1. The solution to a combinatorial probability problem should be set up as a quotient of numerator and denominator *enumerations*. Avoid the temptation to multiply probabilities associated with each position in the sequence. The latter approach will always "sound" reasonable, but it will frequently oversimplify the problem and give the wrong answer.
2. Keep the numerator and denominator consistent with respect to *order*—if permutations are being counted in the numerator, be sure that permutations are being counted in the denominator; likewise, if the outcomes in the numerator are combinations, the outcomes in the denominator should also be combinations.
3. The number of outcomes associated with any problem involving the rolling of n six-sided dice is 6^n; similarly, the number of outcomes associated with tossing a coin n times is 2^n. The number of outcomes associated with dealing a hand of n cards from a standard 52-card poker deck is ${}_{52}C_n$.

QUESTIONS

2.7.1. Ten equally-qualified marketing assistants are candidates for promotion to associate buyer; seven are men and three are women. If the company intends to promote four of the ten at random, what is the probability that exactly two of the four are women?

2.7.2. An urn contains six chips, numbered 1 through 6. Two are chosen at random and their numbers are added together. What is the probability that the resulting sum is equal to five?

2.7.3. An urn contains twenty chips, numbered 1 through 20. Two are drawn simultaneously. What is the probability that the numbers on the two chips will differ by more than two?

2.7.4. A bridge hand (thirteen cards) is dealt from a standard 52-card deck. Let A be the event that the hand contains four aces; let B be the event that the hand contains four kings. Find $P(A \cup B)$.

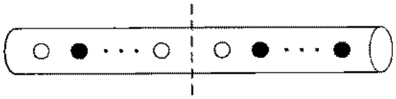
2.7.5. Consider a set of ten urns, nine of which contain three white chips and three red chips each. The tenth contains five white chips and one red chip. An urn is picked at random. Then a sample of size three is drawn without replacement from that urn. If all three chips drawn are white, what is the probability the urn being sampled is the one with five white chips?

2.7.6. A committee of fifty politicians is to be chosen from among our one hundred U.S. Senators. If the selection is done at random, what is the probability that each state will be represented?

2.7.7. Suppose that n fair dice are rolled. What are the chances that all n faces will be the same?

2.7.8. Five fair dice are rolled. What is the probability that the faces showing constitute a "full house"—that is, three faces show one number and two faces show a second number?

2.7.9. Imagine that the test tube pictured contains $2n$ grains of sand, n white and n black. Suppose the tube is vigorously shaken. What is the probability that the two colors of sand will completely separate; that is, all of one color fall to the bottom, and all of the other color lie on top? (*Hint:* Consider the $2n$ grains to be aligned in a row. In how many ways can the n white and the n black grains be permuted?)



2.7.10. Does a monkey have a better chance of rearranging

$$ACCLLUUS \quad \text{to spell} \quad CALCULUS$$

or

$$AABEGLR \quad \text{to spell} \quad ALGEBRA?$$

2.7.11. An apartment building has eight floors. If seven people get on the elevator on the first floor, what is the probability they all want to get off on different floors? On the same floor? What assumption are you making? Does it seem reasonable? Explain.

2.7.12. If the letters in the phrase

$$A\ ROLLING\ STONE\ GATHERS\ NO\ MOSS$$

are arranged at random, what are the chances that not all the S's will be adjacent?

2.7.13. Suppose each of ten sticks is broken into a long part and a short part. The twenty parts are arranged into ten pairs and glued back together, so that again there are ten sticks. What is the probability that each long part will be paired with a short part? (*Note:* This problem is a model for the effects of radiation on a living cell. Each chromosome, as a result of being struck by ionizing radiation, breaks into two parts, one part containing the centromere. The cell will die unless the fragment containing the centromere recombines with one not containing a centromere.)

2.7.14. Six dice are rolled one time. What is the probability that each of the six faces appears?

2.7.15. Suppose that a randomly selected group of k people are brought together. What is the probability that exactly one pair has the same birthday?

2.7.16. For one-pair poker hands, why is the number of denominations for the three single cards $\binom{12}{3}$ rather than $\binom{12}{1}\binom{11}{1}\binom{10}{1}$?

2.7.17. Dana is not the world's best poker player. Dealt a 2 of diamonds, an 8 of diamonds, an ace of hearts, an ace of clubs, and an ace of spades, she discards the three aces. What are her chances of drawing to a flush?

2.7.18. A poker player is dealt a 7 of diamonds, a queen of diamonds, a queen of hearts, a queen of clubs, and an ace of hearts. He discards the 7. What is his probability of drawing to either a full house or four-of-a-kind?

2.7.19. Tim is dealt a 4 of clubs, a 6 of hearts, an 8 of hearts, a 9 of hearts, and a king of diamonds. He discards the 4 and the king. What are his chances of drawing to a straight flush? to a flush?

2.7.20. Five cards are dealt from a standard 52-card deck. What is the probability that the sum of the faces on the five cards is 48 or more?

2.7.21. Nine cards are dealt from a 52-card deck. Write a formula for the probability that three of the five even numerical denominations are represented twice, one of the three face cards appears twice, and a second face card appears once. *Note:* Face cards are the jacks, queens, and kings; 2, 4, 6, 8, and 10 are the even numerical denominations.

2.7.22. A coke hand in bridge is one where none of the thirteen cards is an ace or is higher than a 9. What is the probability of being dealt such a hand?

2.7.23. A pinochle deck has forty-eight cards, two of each of six denominations (9, J, Q, K, 10, A) and the usual four suits. Among the many hands that count for meld is a *roundhouse*, which occurs when a player has a king and queen of each suit. In a hand of twelve cards, what is the probability of getting a "bare" roundhouse (a king and queen of each suit and no other kings or queens)?

2.7.24. A somewhat inebriated conventioneer finds himself in the embarrassing predicament of being unable to predetermine whether his next step will be forward or backward. What is the probability that after hazarding n such maneuvers he will have stumbled forward a distance of r steps? (*Hint:* Let x denote the number of steps he takes forward and y, the number backward. Then $x + y = n$ and $x - y = r$.)

AKING A SECOND LOOK AT STATISTICS (ENUMERATION AND MONTE CARLO TECHNIQUES)

It is a characteristic of probability and combinatorial problems that proposed solutions can sound so right and yet be so wrong. Intuition can easily be fooled, and verbal arguments are often inadequate to deal with questions having even a modicum of complexity. There are some problem-solving strategies available, though, that can be very helpful. In general, approaches that go back to basics are especially useful.

Making a List and Checking It Twice

Ask a realtor to list the three most important features that a house for sale can have and the answer is likely to be "location, location, location." Ask a probabilist to name the three most helpful techniques for solving difficult combinatorial problems and the answe might very well be "enumerate, enumerate, enumerate." Making a partial list of the se of outcomes comprising an event can often show that a proposed solution is incorrect an what the right answer should be. Sometimes, though, the magnitudes of the numbers i a problem are so large that making even a partial list of outcomes is not a viable optio *for that particular problem*. In those cases, the trick is to enumerate a much smaller-scal problem, one that has all the essential features of the original.

For example, suppose a student government council is to be comprised of thre freshmen, three sophomores, three juniors, three seniors, and one at-large representative who could be a member of any of the four classes. Moreover, suppose ten candidate from each of the four classes have been nominated. How many different thirteen-membe councils can be formed?

One approach that may seem reasonable is to choose the council members in a way tha mimics the statement of the question. That is, three representatives from each class ca be chosen in $\binom{10}{3}$ ways; then the at-large member would be selected from the remainin 28(= 40 − 12) nominees. Applying the multiplication rule gives

$$\text{number of different councils} = \binom{10}{3}\binom{10}{3}\binom{10}{3}\binom{10}{3}\binom{28}{1}$$
$$= 5{,}806{,}080{,}000$$

Another approach, which also may seem reasonable, is to realize that one of the classe will necessarily have four representatives, while the other three will each have three. An of the four classes, of course, could be the one with four representatives. Electing fo freshmen, for example, and three from each of the other three classes can be done i $\binom{10}{4}\binom{10}{3}\binom{10}{3}\binom{10}{3}$ ways. Allowing for the fact that the four representatives could in any class, it follows that the total number of thirteen-member councils is *1,451,520,00*

$$\text{number of different councils} = \binom{10}{4}\binom{10}{3}\binom{10}{3}\binom{10}{3} + \binom{10}{3}\binom{10}{4}\binom{10}{3}\binom{10}{3}$$
$$+ \binom{10}{3}\binom{10}{3}\binom{10}{4}\binom{10}{3} + \binom{10}{3}\binom{10}{3}\binom{10}{3}\binom{10}{4}$$
$$= 1{,}451{,}520{,}000$$

Is the first approach overcounting the number of different councils or is the seco approach undercounting? The two proposed solutions differ by a factor of four. Enume ating (by hand) even a portion of the possible outcomes is not feasible here because the sheer magnitude of the combinatorial factors. A very simple analogous question c be posed, though, that *is* easily enumerated. Suppose there were only *two* classes—sa

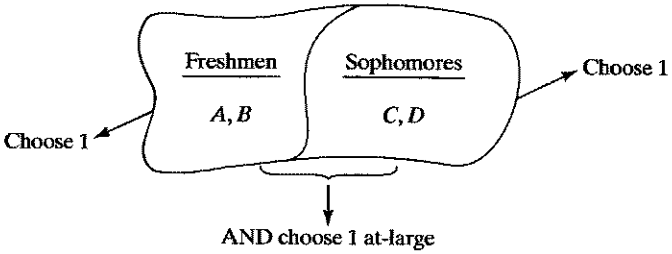


FIGURE 2.8.1

TABLE 2.8.1

	First Approach			Second Approach				
	Fresh.	Soph.	At-large		Fresh.		Soph.	
8	A	C	B	4	A	B	C	
	A	C	D		A	B	D	
	A	D	B		A		C	D
	A	D	C		B		C	D
	B	C	A					
	B	C	D					
	B	D	A					
	B	D	C					
			Duplicates					

freshmen and sophomores—and only two nominees from each class. Furthermore, suppose a three-member council is to be formed, consisting of one freshman, one sophomore, and one representative at large (see Figure 2.8.1).

Applied to Figure 2.8.1, the first approach would claim that the number of different councils is $\binom{2}{1}\binom{2}{1}\binom{4-2}{1}$, or *8*. The second approach would imply that the number of different councils is *4* $\left(=\binom{2}{2}\binom{2}{1}+\binom{2}{1}\binom{2}{2}\right)$. Table 2.8.1 is a listing of the outcomes generated by the two strategies. By inspection, it is now clear that the first approach is incorrect—every possible outcome is double-counted. The outcome ACB, for example, where B is the at-large representative, reappears as BCA, where A is the at-large representative. The second approach, on the other hand, prevents any such overlapping from occurring (but does include all possible councils).

Play It Again, Sam

Recall the von Mises definition of probability given on p. 23: If an experiment is repeated n times under identical conditions, and if the event E occurs on m of those repetitions, then

$$P(E) = \lim_{n\to\infty} \frac{m}{n} \qquad (2.8.1)$$

To be sure, Equation 2.8.1 is an asymptotic result, but it suggests an obvious (and *very* useful) approximation—if n is finite,

$$P(E) \doteq \frac{m}{n}$$

In general, efforts to estimate probabilities by simulating repetitions of an experiment (usually with a computer) are referred to as *Monte Carlo* studies. Usually the technique is used in situations where an exact probability is difficult to calculate. It can also be used, though, as an empirical justification for choosing one proposed solution over another.

For example, consider the game described in Example 2.4.11. An urn contains a red chip, a blue chip, and a two-color chip (red on one side, blue on the other). One chip is drawn at random and placed on a table. The question is, if *blue* is showing, what is the probability that the color underneath is also *blue?*

Pictured in Figure 2.8.2 are two ways of conceptualizing the question just posed. The outcomes in (a) are assuming that a *chip* was drawn. Starting with that premise, the answer to the question is $\frac{1}{2}$—the red chip is obviously eliminated and only one of the two remaining chips is blue on both sides.

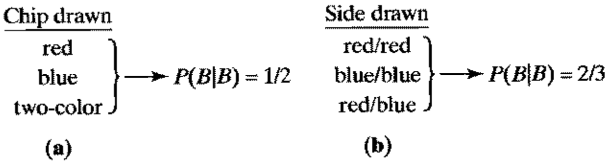


FIGURE 2.8.2

By way of contrast, the outcomes in (b) are assuming that the *side* of a chip was drawn If so, the blue color showing could be any of three blue sides, two of which are blue underneath. According to model (b), then, the probability of both sides being blue is $\frac{2}{3}$.

The formal analysis on pp. 60, of course, resolves the debate—the correct answer is $\frac{2}{3}$. But suppose that such a derivation was unavailable. How might we assess the relative plausibilities of $\frac{1}{2}$ and $\frac{2}{3}$? The answer is simple—just play the game a number of times and see what proportion of outcomes that show blue on top have blue underneath.

To that end, Table 2.8.2 summarizes the results of one hundred random drawings. Fo a total of fifty-two, blue was showing (S) when the chip was placed on a table; for thirty-six of the trials (those marked with an asterisk), the color underneath (U) was also blue Using the approximation suggested by Equation 2.8.1,

$$P(\text{blue is underneath} \mid \text{blue is on top}) = P(B \mid B) \doteq \frac{36}{52} = 0.69$$

a figure much more consistent with $\frac{2}{3}$ than with $\frac{1}{2}$.

TABLE 2.8.2

Trial #	S	U	Trial #	S	U	Trial #	S	U	Trial #	S	U
1	R	B	26	B	R	51	B	R	76	B	B*
2	B	B*	27	R	R	52	R	B	77	B	B*
3	B	R	28	R	B	53	B	B*	78	R	R
4	R	R	29	R	B	54	R	B	79	B	B*
5	R	B	30	R	R	55	R	R	80	R	R
6	R	B	31	R	B	56	R	B	81	R	B
7	R	R	32	B	B*	57	R	R	82	R	B
8	R	R	33	R	B	58	B	B*	83	R	R
9	B	B*	34	B	B*	59	B	R	84	B	R
10	B	R	35	B	B*	60	B	B*	85	B	R
11	R	R	36	R	R	61	B	R	86	R	R
12	B	B*	37	B	R	62	R	B	87	B	B*
13	R	R	38	B	B*	63	R	R	88	R	B
14	B	R	39	R	R	64	R	R	89	B	R
15	B	B*	40	B	B*	65	B	B*	90	R	R
16	B	B*	41	B	B*	66	B	R	91	R	B
17	R	B	42	B	R	67	R	R	92	R	R
18	B	R	43	B	B*	68	B	B*	93	R	R
19	B	B*	44	B	B*	69	B	B*	94	R	B
20	B	B*	45	B	B*	70	R	R	95	B	B*
21	R	R	46	R	R	71	R	R	96	B	B*
22	R	R	47	B	B*	72	B	B*	97	B	R
23	B	B*	48	B	B*	73	R	B	98	R	R
24	B	R	49	R	R	74	R	R	99	B	B*
25	B	B*	50	R	R	75	B	B*	100	B	B*

The point of these examples is not to downgrade the importance of rigorous derivations and exact answers. Far from it. The application of Theorem 2.4.1 to solve the problem posed in Example 2.4.11 is obviously superior to the Monte Carlo approximation illustrated in Table 2.8.2. Still, enumerations of outcomes and replications of experiments can often provide valuable insights and call attention to nuances that might otherwise go unnoticed. As problem-solving techniques in probability and combinatorics, they are extremely, extremely important.

CHAPTER 3

Random Variables

Jakob (Jacques) Bernoulli

One of a Swiss family producing eight distinguished scientists, Jakob was forced by his father to pursue theological studies, but his love of mathematics eventually led him to a university career. He and his brother, Johann, were the most prominent champions of Leibniz's calculus on continental Europe, the two using the new theory to solve numerous problems in physics and mathematics. Bernoulli's main work in probability, Ars Conjectandi, was published after his death by his nephew, Nikolaus, in 1713.

—Jakob (Jacques) Bernoulli (1654–1705)

INTRODUCTION

Throughout Chapter 2, probabilities were assigned to *events*—that is, to sets of sample outcomes. The events we dealt with were composed of either a finite or a countably infinite number of sample outcomes, in which case the event's probability was simply the sum of the probabilities assigned to its outcomes. One particular probability function that came up over and over again in Chapter 2 was the assignment of $\frac{1}{n}$ as the probability associated with each of the n points in a finite sample space. This is the model that typically describes games of chance (and all of our combinatorial probability problems in Chapter 2).

The first objective of this chapter is to look at several other useful ways for assigning probabilities to sample outcomes. In so doing, we confront the desirability of "redefining" sample spaces using functions known as *random variables*. How and why these are used—and what their mathematical properties are—become the focus of virtually everything covered in Chapter 3.

As a case in point, suppose a medical researcher is testing eight elderly adults for their allergic reaction (yes or no) to a new drug for controlling blood pressure. One of the $2^8 = 256$ possible sample points would be the sequence (yes, no, no, yes, no, no, yes, no), signifying that the first subject had an allergic reaction, the second did not, the third did not, and so on. Typically, in studies of this sort, the particular subjects experiencing reactions is of little interest: what does matter is the *number* who show a reaction. If that were true here, the outcome's relevant information (i.e., the number of allergic reactions) could be summarized by the number *3*.[1]

Suppose X denotes the number of allergic reactions among a set of eight adults. Then X is said to be a *random variable* and the number *3* is the *value* of the random variable for the outcome (yes, no, no, yes, no, no, yes, no).

In general, random variables are functions that associate numbers with some attribute of a sample outcome that is deemed to be especially important. If X denotes the random variable and s denotes a sample outcome, then $X(s) = t$, where t is a real number. For the allergy example, $s =$ (yes, no, no, yes, no, no, yes, no) and $t = 3$.

Random variables can often create a dramatically simpler sample space. That certainly is the case here—the original sample space has *256* $(= 2^8)$ outcomes, each being an ordered sequence of length eight. The random variable X, on the other hand, has only *nine* possible values, the integers from 0 to 8, inclusive.

In terms of their fundamental structure, all random variables fall into one of two broad categories, the distinction resting on the number of possible values the random variable can equal. If the latter is finite or countably infinite (which would be the case with the allergic reaction example), the random variable is said to be *discrete;* if the outcomes can be any real number in a given interval, the number of possibilities is uncountably infinite, and the random variable is said to be *continuous*. The difference between the two is critically important, as we will learn in the next several sections.

---

[1] By Theorem 2.6.2, of course, there would be a total of *fifty-six* (= 8!/3!5!) outcomes having exactly three yeses. All fifty-six would be equivalent in terms of what they imply about the drug's likelihood of causing allergic reactions.

The purpose of Chapter 3 is to introduce the important definitions, concepts, and computational techniques associated with random variables, both discrete and continuous. Taken together, these ideas form the bedrock of modern probability and statistics.

3.2 BINOMIAL AND HYPERGEOMETRIC PROBABILITIES

This section looks at two specific probability scenarios that are especially important, both for their theoretical implications as well as for their ability to describe real-world problems. What we learn in developing these two models will help us understand random variables in general, the formal discussion of which begins in Section 3.3.

The Binomial Probability Distribution

Binomial probabilities apply to situations involving a series of independent and identical trials, where each trial can have only one of two possible outcomes. Imagine three distinguishable coins being tossed, each having a probability p of coming up heads. The set of possible outcomes are the eight listed in Table 3.2.1. If the probability of any of the coins coming up heads is p, then the probability of the *sequence* (H, H, H) is p^3, since the coin tosses qualify as independent trials. Similarly, the probability of (T, H, H) is $(1 - p)p^2$. The fourth column of Table 3.2.1 shows the probabilities associated with each of the three-coin sequences.

Suppose our main interest in the coin tosses is the *number* of heads that occur. Whether the actual sequence is, say, (H, H, T) or (H, T, H) is immaterial, since each outcome contains exactly two heads. The last column of Table 3.2.1 shows the number of heads in each of the eight possible outcomes. Notice that there are *three* outcomes with exactly two heads, each having an individual probability of $p^2(1 - p)$. The probability, then, of the event "two heads" is the sum of those three individual probabilities—that is, $3p^2(1 - p)$. Table 3.2.2 lists the probabilities of tossing k heads, where $k = 0, 1, 2,$ or 3.

TABLE 3.2.1

1st Coin	2nd Coin	3rd Coin	Probability	Number of Heads
H	H	H	p^3	3
H	H	T	$p^2(1 - p)$	2
H	T	H	$p^2(1 - p)$	2
T	H	H	$p^2(1 - p)$	2
H	T	T	$p(1 - p)^2$	1
T	H	T	$p(1 - p)^2$	1
T	T	H	$p(1 - p)^2$	1
T	T	T	$(1 - p)^3$	0

TABLE 3.2.2

Number of Heads	Probability
0	$(1 - p)^3$
1	$3p(1 - p)^2$
2	$3p^2(1 - p)$
3	p^3

Now, more generally, suppose that n coins are tossed, in which case the number of heads can equal any integer from 0 through n. By analogy,

$$P(k \text{ heads}) = \begin{pmatrix} \text{number of ways} \\ \text{to arrange } k \\ \text{heads and } n - k \text{ tails} \end{pmatrix} \cdot \begin{pmatrix} \text{probability of any} \\ \text{particular sequence} \\ \text{having } k \text{ heads} \\ \text{and } n - k \text{ tails} \end{pmatrix}$$

$$= \begin{pmatrix} \text{number of ways} \\ \text{to arrange } k \\ \text{heads and } n - k \text{ tails} \end{pmatrix} \cdot p^k(1 - p)^{n-k}$$

The number of ways to arrange k Hs and $n - k$ Ts, though, is $\dfrac{n!}{k!(n - k)!}$, or $\dbinom{n}{k}$ (recall Theorem 2.6.2).

Theorem 3.2.1. *Consider a series of n independent trials, each resulting in one of two possible outcomes, "success" or "failure." Let $p = P$ (success occurs at any given trial) and assume that p remains constant from trial to trial. Then*

$$P(k \text{ successes}) = \binom{n}{k} p^k(1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Comment. The probability assignment given by the Equation in Theorem 3.2.1 is known as the *binomial distribution.*

EXAMPLE 3.2.1

As the lawyer for a client accused of murder, you are looking for ways to establish "reasonable doubt" in the minds of the jurors. Central to the prosecutor's case is testimony from a forensics expert who claims that a blood sample taken from the scene of the crime matches the DNA of your client. One-tenth of 1% of the time, though, such tests are in error.

Suppose your client is actually guilty. If six other laboratories in the country are capable of doing this kind of DNA analysis (and you hire them all), what are the chances that at least one will make a mistake and conclude that your client is innocent?

Each of the six analyses constitutes an independent trial, where $p = P$ (lab makes a mistake) $= 0.001$. Substituting into Theorem 3.2.1 shows that the lawyer's strategy is not likely to work:

$$\begin{aligned} P(\text{at least one lab says client is innocent}) &= 1 - P(0 \text{ labs make a mistake}) \\ &= 1 - \binom{6}{0}(0.001)^0(0.999)^6 \\ &= 0.006 \end{aligned}$$

For the defendant, the calculated 0.006 is hardly reassuring. Given such small values for n and p, though, getting contradictory forensic results would be a longshot at best. But then again, as the erstwhile TV detective, Baretta, was fond of saying. "If you can't do the time, don't do the crime."

EXAMPLE 3.2.2

Kingwest Pharmaceuticals is experimenting with a new affordable AIDS medication, PM-17, that may have the ability to strengthen a victim's immune system. Thirty monkeys infected with the HIV complex have been given the drug. Researchers intend to wait six weeks and then count the number of animals whose immunological responses show a marked improvement. Any inexpensive drug capable of being effective 60% of the time would be considered a major breakthrough; medications whose chances of success are 50% or less are not likely to have any commercial potential.

Yet to be finalized are guidelines for interpreting results. Kingwest hopes to avoid making either of two errors: (1) rejecting a drug that would ultimately prove to be marketable and (2) spending additional development dollars on a drug whose effectiveness, in the long run, would be 50% or less. As a tentative "decision rule," the project manager suggests that unless *16 or more* of the monkeys show improvement, research on PM-17 should be discontinued.

a. What are the chances that the "sixteen or more" rule will cause the company to reject PM-17, *even if the drug is 60% effective?*

b. How often will the "sixteen or more" rule allow a 50%-effective drug to be perceived as a major breakthrough?

(a) Each of the monkeys is one of $n = 30$ independent trials, where the outcome is either a "success" (monkey's immune system is strengthened) or a "failure" (monkey's immune system is not strengthened). By assumption, the probability that PM-17 produces an immunological improvement in any given monkey is $p = P$ (success) $= 0.60$.

By Theorem 3.2.1, the probability that exactly k monkeys (out of thirty) will show improvement after six weeks is $\binom{30}{k}(0.60)^k(0.40)^{30-k}$. The probability, then, that the "sixteen or over" rule will cause a 60%-effective drug to be discarded is the sum of "binomial" probabilities for k values ranging from 0 to 15:

$$\begin{aligned} P(60\%\text{-effective drug fails "sixteen or more" rule}) &= \sum_{k=0}^{15} \binom{30}{k}(0.60)^k(0.40)^{30-k} \\ &= 0.1754 \end{aligned}$$

Roughly 18% of the time, in other words, a "breakthrough" drug such as PM-17 will produce test results so mediocre (as measured by the "sixteen or more" rule) that the company will be misled into thinking it has no potential.

(b) The other error Kingwest can make is to conclude that PM-17 warrants further study when, in fact, its value for p is below a marketable level. The chance that particular incorrect inference will be drawn here is the probability that the number of that successes will be greater than or equal to sixteen when $p = 0.5$. That is,

$$\begin{aligned} &P(50\%\text{-effective PM-17 appears to be marketable}) \\ &\quad = P(\text{sixteen or more successes occur}) \\ &\quad = \sum_{k=16}^{30} \binom{30}{k}(0.5)^k(0.5)^{30-k} \\ &\quad = 0.43 \end{aligned}$$

Thus, even if PM-17's success rate is an unacceptably low 50%, it has a 43% chance of performing sufficiently well in thirty trials to satisfy the "sixteen or more" criterion.

Comment. Evaluating binomial summations can be tedious, even with a calculator. Statistical software packages offer a convenient alternative. Appendix 3.A.1 describes how one such program, MINITAB, can be used to answer the sorts of questions posed in Example 3.2.2.

EXAMPLE 3.2.3

The Stanley Cup playoff in professional hockey is a seven-game series, where the first team to win four games is declared the champion. The series, then, can last anywhere from four to seven games (just like the World Series in baseball). Calculate the likelihoods that the series will last four, five, six, and seven games. Assume that (1) each game is an independent event and (2) the two teams are evenly matched.

Consider the case where Team A wins the series in *six* games. For that to happen, they must win exactly three of the first five games *and* they must win the sixth game. Because of the independence assumption, we can write

$$\begin{aligned} P(\text{Team A wins in six games}) &= P(\text{Team A wins three of first five}) \cdot P(\text{Team A wins sixth} \\ &= \left[\binom{5}{3}(0.5)^3(0.5)^2\right] \cdot (0.5) = 0.15625 \end{aligned}$$

Since the probability that Team B wins the series in six games is the same (why?),

$$\begin{aligned} P(\text{series ends in six games}) &= P(\text{Team A wins in six games} \cup \text{Team B wins in six games}) \\ &= P(\text{A wins in six}) + P(\text{B wins in six}) \qquad (\text{why?}) \\ &= 0.15625 + 0.15625 \\ &= 0.3125 \end{aligned}$$

A similar argument allows us to calculate the probabilties of four-, five-, and seven- game series:

$$\begin{aligned} P(\text{four game series}) &= 2(0.5)^4 = 0.125 \\ P(\text{five game series}) &= 2\left[\binom{4}{3}(0.5)^3(0.5)\right](0.5) = 0.25 \\ P(\text{seven game series}) &= 2\left[\binom{6}{3}(0.5)^3(0.5)^3\right](0.5) = 0.3125 \end{aligned}$$

Having calculated the "theoretical" probabilities associated with the possible lengths of a Stanley Cup playoff raises an obvious question: How do those likelihoods compare with the actual distribution of playoff lengths? For a recent fifty-nine year period, Column 2 in Table 3.2.3 shows the proportion of playoffs that lasted 4, 5, 6, and 7 games, respectively.

Clearly, the agreement between the entries in Columns two and three is not very good: Particularly noticeable is the excess of short playoffs (four games) and the deficit of long playoffs (seven games). What this "lack of fit" suggests is that one or more of the binomial distribution assumptions is not satisfied. Consider, for example, the parameter p, which we assumed to equal $\frac{1}{2}$. In reality, its value might be something quite different—just

TABLE 3.2.3

Series Length	Observed Proportion	Theoretical Probability
4	19/59 = 0.322	0.125
5	15/59 = 0.254	0.250
6	15/59 = 0.254	0.3125
7	10/59 = 0.169	0.3125

because the teams playing for the championship won their respective divisions, it does not necessarily follow that the two are equally good. Indeed, if the two contending teams were frequently mismatched, the consequence would be an increase in the number of short playoffs and a decrease in the number of long playoffs. It may also be the case that momentum is a factor in a team's chances of winning a given game. If so, the independence assumption implicit in the binomial model is rendered invalid.

EXAMPLE 3.2.4

Doomsday Airlines ("Come Take the Flight of Your Life") has two aircraft—a dilapidated two-engine prop plane and an equally outdated and under-maintained four-engine prop plane. Each plane will land safely only if at least half its engines are working properly. Given that you wish to remain among the living, under what conditions would you opt to fly on the two-engine plane? Assume that each engine on each plane has the same probability p of failing and that any such failures are independent events.

For the two-engine plane,

$$\begin{aligned} P(\text{flight lands safely}) &= P(\text{one or more engines work properly}) \\ &= \sum_{k=1}^{2} \binom{2}{k}(1-p)^k p^{2-k} \qquad (3.2.1) \end{aligned}$$

For the four-engine plane,

$$\begin{aligned} P(\text{flight lands safely}) &= P(\text{two or more engines work properly}) \\ &= \sum_{k=2}^{4} \binom{4}{k}(1-p)^k p^{4-k} \qquad (3.2.2) \end{aligned}$$

When to opt for the two-engine plane, then, reduces to an algebra problem: We look for the values of p for which

$$\sum_{k=1}^{2} \binom{2}{k}(1-p)^k p^{2-k} > \sum_{k=2}^{4} \binom{4}{k}(1-p)^k p^{4-k}$$

or, equivalently,

$$\sum_{k=0}^{1} \binom{4}{k}(1-p)^k p^{4-k} > \binom{2}{0}(1-p)^0 p^2$$

Simplifying the inequality

$$\binom{4}{0}(1-p)^0 p^4 + \binom{4}{1}(1-p)^1 p^3 > \binom{2}{0}(1-p)^0 p^2$$

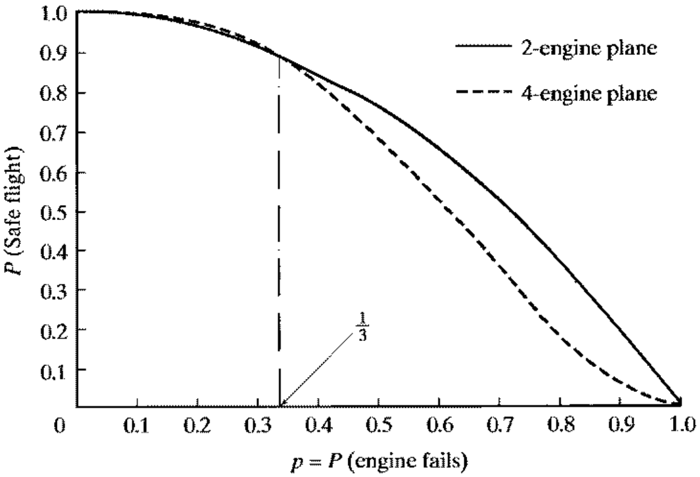


FIGURE 3.2.1

gives

$$(3p - 1)(p - 1) < 0 \tag{3.2.3}$$

But $(p - 1)$ is never positive, so Inequality 3.2.3 will be true only when $(3p - 1) > 0$, which gives $p > \frac{1}{3}$ as the desired solution set. Figure 3.2.1 compares the two "safe return" probabilities as a function of p.

QUESTIONS

3.2.1. An investment analyst has tracked a certain blue-chip stock for the past six months and found that on any given day it either goes up a point or down a point. Furthermore, it went up on 25% of the days and down on 75%. What is the probability that at the close of trading four days from now the price of the stock will be the same as it is today? Assume that the daily fluctuations are independent events.

3.2.2. In a nuclear reactor, the fission process is controlled by inserting special rods into the radioactive core to absorb neutrons and slow down the nuclear chain reaction. When functioning properly, these rods serve as a first-line defense against a core meltdown. Suppose a reactor has 10 control rods, each operating independently and each having a 0.80 probability of being properly inserted in the event of an "incident". Furthermore, suppose that a meltdown will be prevented if at least half the rods perform satisfactorily. What is the probability that, upon demand, the system will fail?

3.2.3. Suppose that since the early 1950s some 10,000 independent UFO sightings have been reported to civil authorities. If the probability that any sighting is genuine is on the order of 1 in 100,000, what is the probability that at least 1 of the 10,000 was genuine?

3.2.4. The probability that a circuit board coming off an assembly line needs rework is 0.15. Suppose that 12 boards are tested.

(a) What is the probability that exactly 4 will need rework?

(b) What is the probability that at least one needs rework?

3.2.5. A manufacturer has 10 machines that die cut cardboard boxes. The probability that, on a given day, any one of the machines will be out of service for repair or maintenance is 0.05. If the day's production requires the availability of at least seven of the machines, what is the probability the work will get done?

3.2.6. Two lighting systems are being proposed for an employee work area. One requires 50 bulbs, each having a probability of 0.05 of burning out within a month's time. The second has 100 bulbs, each with a 0.02 burnout probability. Whichever system is installed will be inspected once a month for the purpose of replacing burned-out bulbs. Which system is likely to require less maintenance? Answer the question by comparing the probabilities that each will require at least one bulb to be replaced at the end of 30 days.

3.2.7. The great English diarist Samuel Pepys asked his friend Sir Isaac Newton the following question: Is it more likely to get at least one 6 when 6 dice are rolled, at least two 6's when 12 dice are rolled, or at least three 6's when 18 dice are rolled? After considerable correspondence (see (162)). Newton convinced the skeptical Pepys that the first event is the most likely. Compute the three probabilities.

3.2.8. The gunner on a small assault boat fires six missiles at an attacking plane. Each has a 20% chance of being on target. It two or more of the shells find their mark, the plane will crash. At the same time, the pilot of the plane fires 10 air-to-surface rockets, each of which has a 0.05 chance of critically disabling the boat. What you rather be on the plane or the boat?

3.2.9. If a family has four children, is it more likely they will have two boys and two girls or three of one sex and one of the other? Assume that the probability of a child being a boy is $\frac{1}{2}$ and that the births are independent events.

3.2.10. Experience has shown that only $\frac{1}{3}$ of all patients having a certain disease will recover if given the standard treatment. A new drug is to be tested on a group of 12 volunteers. If the FDA requires that at least seven of these patients recover before it will license the new drug, what is the probability that the treatment will be discredited even if it has the potential to increase an individual's recovery rate to $\frac{1}{2}$?

3.2.11. Transportation to school for a rural county's 76 children is provided by a fleet of four buses. Drivers are chosen on a day-to-day basis and come from a pool of local farmers who have agreed to be "on call". What is the smallest number of drivers that need to be in the pool if the county wants to have at least a 95% probability on any given day that all the buses will run? Assume that each driver has an 80% chance of being available if contacted.

3.2.12. The captain of a Navy gunboat orders a volley of 25 missiles to be fired at random along a 500-foot stretch of shoreline that he hopes to establish as a beachhead. Dug into the beach is a 30-foot-long bunker serving as the enemy's first line of defense. The captain has reason to believe that the bunker will be destroyed if at least three of the missiles are on target. What is the probability of that happening?

3.2.13. A computer has generated seven random numbers over the interval 0 to 1. Is it more likely that (1) exactly three will be in the interval $\frac{1}{2}$ to 1 or (2) fewer than three will be greater than $\frac{3}{4}$?

3.2.14. Listed in the following table is the length distribution of World Series competition for the 52 years from 1950 to 2002.

World Series Lengths

Number of Games, X	Number of Years
4	9
5	8
6	11
7	24
	52

Assuming that each World Series game is an independent event and that the probability of either team's winning any particular contest is 0.5, find the probability of each series length. How well does the model fit the data? (Compute the "expected" frequencies, that is, multiply the probability of a given length series times 52).

3.2.15. Use the expansion of $(x + y)^n$ (recall the comment in Section 2.6 on page 108) to verify that the binomial probabilities sum to 1; that is, $\sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} = 1$

3.2.16. Suppose a series of n independent trials can end in one of *three* possible outcomes. Let k_1 and k_2 denote the number of trials that result in outcomes 1 and 2, respectively. Let p_1 and p_2 denote the probabilities associated with outcomes 1 and 2. Generalize Theorem 3.2.1. to deduce a formula for the probability of getting k_1 and k_2 occurrences of outcomes 1 and 2, respectively.

3.2.17. Repair calls for central air conditioners fall into three general categories: coolant leakage, compressor failure, and electrical malfunction. Experience has shown that the probabilities associated with the three are 0.5, 0.3, and 0.2, respectively. Suppose that a dispatcher has logged in 10 service requests for tomorrow morning. Use the answer to Question 3.2.16 to calculate the probability that 3 of those 10 will involve coolant leakage and 5 will be compressor failures.

The Hypergeometric Distribution

The second "special" distribution that we want to look at formalizes the urn problems that frequented Chapter 2. Our solutions to those earlier problems tended to be enumerations. We listed the entire set of possible samples, and then counted the ones that satisfied the event in question. The inefficiency and redundancy of that approach should be painfully obvious. What we are seeking here is a general formula that can be applied to any and all such problems, much like the expression in Theorem 3.2.1 can handle the full range of questions arising from the binomial model.

Suppose an urn contains r red chips and w white chips, where $r + w = N$. Imagine drawing n chips from the urn one-at-a-time without replacing any of the chips selected. At each drawing we record the color of the chip removed. The question is, what is the probability that exactly k red chips are included among the n that are removed?

Notice that the experiment just described is similar in some respects to the binomial model, but the method of sampling creates a critical distinction. *If* each chip drawn was replaced prior to making another selection, then each drawing would be an independent trial, the chances of drawing a red at any given try would be a constant r/N, and the probability that exactly k red chips would ultimately be included in the n selections would be a direct application of Theorem 3.2.1:

$$P(k \text{ reds drawn}) = \binom{n}{k}(r/N)^k(1 - r/N)^{n-k}, \quad k = 0, 1, 2, \ldots, n$$

However, if the chips drawn are *not* replaced, then the probability of drawing a red on any given attempt is not necessarily r/N: Its value would depend on the colors of the chips selected earlier. Since $p = P(\text{red is drawn}) = P(\text{success})$ does not remain constant from drawing to drawing, the binomial model of Theorem 3.2.1 does not apply. Instead, probabilities that arise from the "no replacement" scenario just described are said to follow the *hypergeometric distribution.*

Theorem 3.2.2. *Suppose an urn contains r red chips and w white chips, where $r + w = N$. If n chips are drawn out at random, without replacement, and if k denotes the number of red chips selected, then*

$$P(k \text{ red chips are chosen}) = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}} \tag{3.2.4}$$

where k varies over all the integers for which $\binom{r}{k}$ and $\binom{w}{n-k}$ are defined. The probabilities appearing on the right-hand side of Equation 3.2.4 are known as the hypergeometric distribution.

Proof. Assume the chips are distinguishable. We need to count the number of elements making up the event of getting k red chips and $n - k$ white chips. The number of ways to select the red chips, regardless of the order in which they are chosen, is ${}_rP_k$. Similarly, the number of ways to select the $n - k$ white chips is ${}_wP_{n-k}$. However, the order in which the chips are selected does matter. Each outcome is an n-long ordered sequence of red and white. There are $\binom{n}{k}$ ways to choose where in the sequence the red chips go. Thus, the number of elements in the event of interest is $\binom{n}{k}{}_rP_k\,{}_wP_{n-k}$. Now, the total number of ways to choose n elements from N, in order, without replacement is ${}_NP_n$, so

$$P(k \text{ red chips are chosen}) = \frac{\binom{n}{k}{}_rP_k\,{}_wP_{n-k}}{{}_NP_n}$$

This quantity, while correct, is not in the form of the statement of the theorem. To make that conversion, we have to change all of the terms in the expression to factorials:

$$P(k \text{ red chips are chosen}) = \frac{\binom{n}{k} {}_rP_k \, {}_wP_{n-k}}{{}_NP_n}$$

$$= \frac{\dfrac{n!}{k!(n-k)!} \dfrac{r!}{(r-k)!} \dfrac{w!}{(w-n+k)!}}{\dfrac{N}{(N-n)!}}$$

$$= \frac{\dfrac{r!}{k!(r-k)!} \dfrac{w!}{(n-k)!(w-n+k)!}}{\dfrac{N!}{n!(N-n)!}} = \frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}} \qquad \square$$

Comment. The appearance of binomial coefficients suggests a model of selecting unordered subsets. Indeed, one can consider the model of selecting a subset of size n simultaneously, where order doesn't matter. In that case, the question remains: what is the probability of getting k red chips and $n - k$ white chips. A moment's reflection will show that the hypergeometric probabilities given in the statement of the theorem also answer that question. So, if our interest is simply counting the number of red and white chips in the sample, the probabilities are the same whether the drawing of the sample is simultaneous, or the chips are drawn in order without repetition.

Comment. The name *hypergeometric* derives from a series introduced by the Swiss mathematician and physicist, Leonhard Euler, in 1769:

$$1 + \frac{ab}{c}x + \frac{a(a+1)b(b+1)}{2!c(c+1)}x^2 + \frac{a(a+1)(a+2)b(b+1)(b+2)}{3!c(c+1)(c+2)}x^3 + \cdots$$

This is an expansion of considerable flexibility: Given appropriate values for a, b, and c, it reduces to many of the standard infinite series used in analysis. In particular, if a is set equal to 1, and b and c are set equal to each other, it reduces to the familiar *geometric* series,

$$1 + x + x^2 + x^3 + \cdots$$

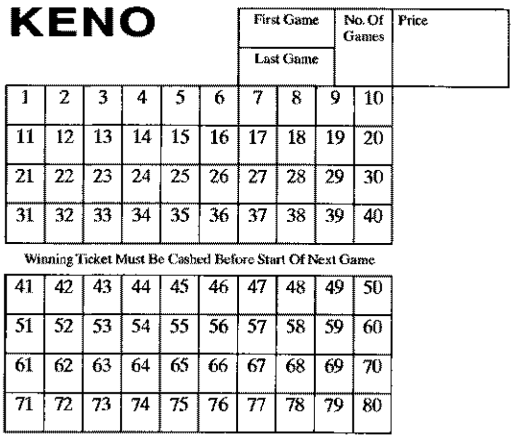
hence the name *hypergeometric*. The relationship of the probability function in Theorem 3.2.2 to Euler's series becomes apparent if we set $a = -n$, $b = -r$, $c = w - n + 1$, and multiply the series by $\binom{w}{n} \Big/ \binom{N}{n}$. Then the coefficient of x^k will be

$$\frac{\binom{r}{k}\binom{w}{n-k}}{\binom{N}{n}}$$

the value the theorem gives for $P(k$ red chips are chosen).

EXAMPLE 3.2.5

Keno is among the most popular games played in Las Vegas even though it ranks as one of the least "fair" in the sense that the odds are overwhelmingly in favor of the house. (Betting on keno is only a little less foolish than playing a slot machine!) A keno card has eighty numbers, 1 through 80, from which the player selects a sample of size k, where k can be anything from 1 to 15 (see Figure 3.2.2). The "caller" then announces twenty winning numbers, chosen at random from the eighty. If—and how much—the player wins depends on how many of his numbers match the twenty identified by the caller. Suppose that a player bets on a ten-spot ticket. What is his probability of "catching" five numbers?



KENO

First Game | No. Of Games | Price

Last Game

1	2	3	4	5	6	7	8	9	10
11	12	13	14	15	16	17	18	19	20
21	22	23	24	25	26	27	28	29	30
31	32	33	34	35	36	37	38	39	40

Winning Ticket Must Be Cashed Before Start Of Next Game

41	42	43	44	45	46	47	48	49	50
51	52	53	54	55	56	57	58	59	60
61	62	63	64	65	66	67	68	69	70
71	72	73	74	75	76	77	78	79	80

FIGURE 3.2.2

Consider an urn containing eighty numbers, twenty of which are winners and sixty of which are losers (see Figure 3.2.3). By betting on a ten-spot ticket, the player, in effect, is drawing a sample of size ten from that urn. The probability of "catching" five numbers is the probability that five of the numbers the player has bet on are contained in the set of twenty winning numbers.

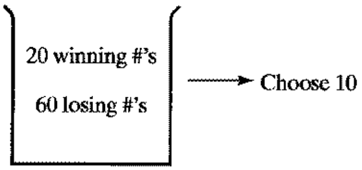


FIGURE 3.2.3

By Theorem 3.2.2 (with $r = 20$, $w = 60$, $n = 10$, $N = 80$, and $k = 5$), the player has approximately a 5% chance of guessing exactly five winning numbers:

$$P(\text{five winning numbers are selected}) = \frac{\binom{20}{5}\binom{60}{5}}{\binom{80}{10}} = 0.05$$

EXAMPLE 3.2.6

A hung jury is one that is unable to reach a unanimous decision. Suppose that a pool of twenty-five potential jurors is assigned to a murder case where the evidence is so overwhelming against the defendant that twenty-three of the twenty-five would return a guilty verdict. The other two potential jurors would vote to acquit regardless of the facts. What is the probability that a twelve-member panel chosen at random from the pool of twenty-five will be unable to reach a unanimous decision?

Think of the jury pool as an urn containing twenty-five chips, twenty-three of which correspond to jurors who would vote "guilty" and two of which correspond to jurors who would vote "not guilty." If either or both of the jurors who would vote "not guilty" are included in the panel of twelve, the result would be a hung jury. Applying Theorem 3.2.2 (twice) gives *0.74* as the probability that the jury impanelled would not reach a unanimous decision:

$$\begin{aligned}P(\text{hung jury}) &= P(\text{decision is not unanimous})\\ &= \binom{2}{1}\binom{23}{11}\Big/\binom{25}{12} + \binom{2}{2}\binom{23}{10}\Big/\binom{25}{12}\\ &= 0.74\end{aligned}$$

EXAMPLE 3.2.7

When a bullet is fired it becomes scored with minute striations produced by imperfections in the gun barrel. Appearing as a series of parallel lines, these striations have long been recognized as a basis for matching a bullet with a gun, since repeated firings of the same weapon will produce bullets having substantially the same configuration of markings. Until recently, deciding how close two patterns had to be before it could be concluded the bullets came from the same weapon was largely subjective. A ballistics expert would simply look at the two bullets under a microscope and make an informed judgment based on past experience. Today, criminologists are beginning to address the problem more quantitatively, partly with the help of the hypergeometric distribution.

Suppose a bullet is recovered from the scene of a crime, along with the suspect's gun. Under a microscope, a grid of m cells, numbered 1 to m, is superimposed over the bullet. If m is chosen large enough so the width of the cells is sufficiently small, each of that

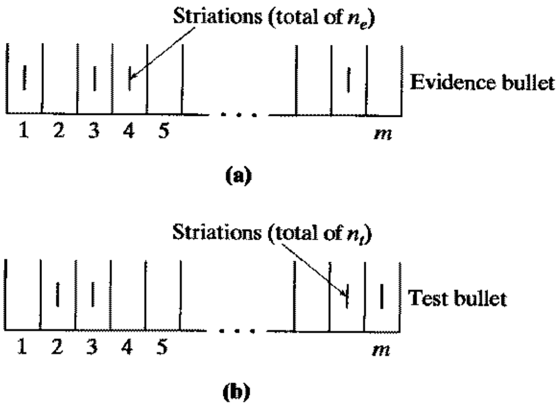


FIGURE 3.2.4

evidence bullet's n_e striations will fall into a different cell (see Figure 3.2.4(a)). Then the suspect's gun is fired, yielding a test bullet, which will have a total of n_t striations located in a possibly different set of cells (see Figure 3.2.4(b)). How might we assess the similarities in cell locations for the two striation patterns?

As a model for the striation pattern on the evidence bullet, imagine an urn containing m chips, with n_e corresponding to the striation locations. Now, think of the striation pattern on the *test* bullet as representing a sample of size n_t from the evidence urn. By Theorem 3.2.2, the probability that k of the cell locations will be shared by the two striation patterns is

$$\frac{\binom{n_e}{k}\binom{m-n_e}{n_t-k}}{\binom{m}{n_t}}$$

Suppose the bullet found at a murder scene is superimposed with a grid having $m = 25$ cells, n_e of which contain striations. The suspect's gun is fired and the bullet is found to have $n_t = 3$ striations, one of which matches the location of one of the striations on the evidence bullet. What do you think a ballistics expert would conclude?

Intuitively, the similarity between the two bullets would be reflected in the probability that *one or more* striations in the suspect's bullet matched the evidence bullet. The smaller that probability is, the stronger would be our belief that the two bullets were fired by the same gun. Based on the values given for m, n_e, and n_t,

$$P(\text{one or more matches}) = \frac{\binom{4}{1}\binom{21}{2}}{\binom{25}{3}} + \frac{\binom{4}{2}\binom{21}{1}}{\binom{25}{3}} + \frac{\binom{4}{3}\binom{21}{0}}{\binom{25}{3}}$$
$$= 0.42$$

If P(one or more matches) had been a very small number—say, 0.001—the inference would have been clear-cut: The same gun fired both bullets. But, here with the probability of one or more matches being so large, we cannot rule out the possibility that the bullets were fired by two different guns (and, presumably, by two different people).

EXAMPLE 3.2.8

Wipe Your Feet, a carpet cleaning company, is trying to establish name recognition in a community consisting of sixty thousand households. The company's management team estimates that five thousand of those families would do business with the firm if they were contacted and informed of the services available. With that in mind, the company has hired a staff of telemarketers to place one thousand calls. Write a formula for the probability that at least one hundred new customers will be identified.

Conceptually, this is an urn problem not unlike the previous three examples, except for the fact that the numbers of "chips" are powers of ten larger than what we have encountered up to this point. In the terminology of Theorem 3.2.2, $N = 60{,}000$, $r = 5{,}000$, $w = 55{,}000$, $n = 1000$, and

$$\begin{aligned}&P(\text{telemarketers identify } k \text{ new customers})\\&\quad= \frac{\binom{5000}{k}\binom{55{,}000}{1000-k}}{\binom{60{,}000}{1000}}, \quad k = 0, 1, \ldots, 1000\end{aligned}$$

It follows that

$$\begin{aligned}&P(\text{one hundred or more new customers are identified})\\&\quad= 1 - P(\text{ninety-nine or fewer new customers are identified})\\&\quad= 1 - \sum_{k=0}^{99} \frac{\binom{5000}{k}\binom{55{,}000}{1000-k}}{\binom{60{,}000}{1000}} \qquad (3.2.5)\end{aligned}$$

Needless to say, evaluating Equation 3.2.5 directly is very difficult because of the number of terms involved and the large factorials implicit in both the numerator and denominator. In Chapter 4 we will learn a series of approximations that virtually trivialize the evaluation.

CASE STUDY 3.2.1

Biting into a plump, juicy apple is one of the innocent pleasures of autumn. Critical to that enjoyment is the *firmness* of the apple, a property that growers and shippers monitor closely. The apple industry goes so far as to set a lowest acceptable limit for firmness, which is measured (in lbs) by inserting a probe into the apple. For the Red Delicious variety, for example, firmness is supposed to be at least 12 lbs; in the state of Washington, wholesalers are not allowed to sell apples if more than 10% of their shipment falls below that 12 lb limit.

All of this raises an obvious question: How can shippers demonstrate that their apples meet the 10% standard? Testing each one is not an option—the probe that measures firmness renders an apple unfit for sale. That leaves *sampling* as the only viable strategy.

Suppose, for example, a shipper has a supply of 144 apples. She decides to select 15 at random and measure each one's firmness, with the intention of selling the remaining apples if 2 or fewer in the sample are substandard. What are the consequences of her plan? More specifically, does it have a good chance of "accepting" a shipment that meets the 10% rule and a good chance of "rejecting" one that does not? (If either or both of those objectives are not met, the plan is inappropriate.)

For example, suppose there are actually *10* defective apples among the original 144. Since $\frac{10}{144} \times 100 = 6.9\%$, that shipment would be suitable for sale because fewer than 10% failed to meet the firmness standard. The question is, how likely is it that a sample of 15 chosen at random from that shipment will pass inspection?

Notice, here, that the number of substandard apples in the sample has a hypergeometric distribution with $r = 10$, $w = 134$, $n = 15$, and $N = 144$. Therefore,

$$\begin{aligned} P(\text{sample passes inspection}) &= P(2 \text{ or fewer substandard apples are found}) \\ &= \frac{\binom{10}{0}\binom{134}{15}}{\binom{144}{15}} + \frac{\binom{10}{1}\binom{134}{14}}{\binom{144}{15}} + \frac{\binom{10}{2}\binom{134}{13}}{\binom{144}{15}} \\ &= 0.320 + 0.401 + 0.208 = 0.929 \end{aligned}$$

So, the probability is reassuringly high that a supply of apples this good would, in fact, be judged acceptable to ship. Of course, it also follows from this calculation that roughly 7% of the time, the number of substandard apples found will be *greater* than 2, in which case the apples would be (incorrectly) assumed to be unsuitable for sale (earning them an undeserved one-way ticket to the applesauce factory...)

How good is the proposed sampling plan at recognizing apples that would, in fact, be inappropriate to ship? Suppose, for example, that *30*, or *21*%, of the 144 apples

(Continued on next page)

(Case Study 3.2.1 continued)

would fall below the 12 lb limit. Ideally, the probability here that a sample passes inspection should be small. The number of substandard apples found in this case would be hypergeometric with $r = 30$, $w = 114$, $n = 15$, and $N = 144$, so

$$P(\text{sample passes inspection}) = \frac{\binom{30}{0}\binom{114}{15}}{\binom{144}{15}} + \frac{\binom{30}{1}\binom{114}{14}}{\binom{144}{15}} + \frac{\binom{30}{2}\binom{114}{13}}{\binom{144}{15}}$$

$$= 0.024 + 0.110 + 0.221 = 0.355$$

Here the bad news is that the sampling plan will allow a 21% defective supply to be shipped *36%* of the time. The good news is that *64%* of the time, the number of substandard apples in the sample will exceed 2, meaning that the correct decision "not to ship" will be made.

Figure 3.2.5 shows P(sample passes) plotted against the percentage of defectives in the entire supply. Graphs of this sort are called *operating characteristic* (or *OC*) *curves:* They summarize how a sampling plan will respond to all possible levels of quality.

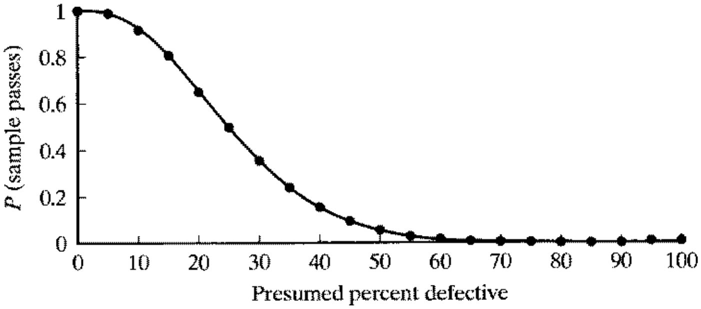


FIGURE 3.2.5

Comment. Every sampling plan invariably allows for two kinds of errors—rejecting shipments that should be accepted and accepting shipments that should be rejected. In practice, the probabilities of committing these errors can be manipulated by redefining the decision rule and/or changing the sample size. Some of these options will be explored later in Chapter 6.

QUESTIONS

3.2.18. A corporate board contains 12 members. The board decides to create a five person Committee to Hide Corporation Debt. Suppose four members of the board are accountants. What is the probability that the Committee will contain two accountants and three non-accountants?

3.2.19. One of the popular tourist attractions in Alaska is watching black bears catch salmon, swimming upstream to spawn. Not all "black" bears are black, though—some are tan-colored. Suppose that six black bears and three tan-colored bears are working the rapids of a salmon stream. Over the course of an hour, six different bears are sighted. What is the probability that those six will include at least twice as many black bears as tan-colored bears?

3.2.20. A city has 4050 children under the age of 10, including 514 who have not been vaccinated for measles. Sixty-five of the city's children are enrolled in the ABC Day Care Center. Suppose the municipal health department sends a doctor and a nurse to ABC to immunize any child who has not already been vaccinated. Find a formula for the probability that exactly k of the children at ABC have not been vaccinated.

3.2.21. Country A inadvertently launches 10 guided missiles—6 armed with nuclear warheads—at Country B. In response, Country B fires 7 antiballistic missiles, each of which will destroy exactly one of the incoming rockets. The antiballistic missiles have no way of detecting, though, which of the 10 rockets are carrying nuclear warheads. What are the chances that Country B will be hit by at least one nuclear missile?

3.2.22. Anne is studying for a history exam covering the French Revolution that will consist of five essay questions selected at random from a list of 10 the professor has handed out to the class in advance. Not exactly a Napoleon buff, Anne would like to avoid researching all 10 questions but still be reasonably assured of getting a fairly good grade. Specifically, she wants to have at least an 85% chance of getting at least four of the five questions right. Will it be sufficient if she studies eight of the 10 questions?

3.2.23. Each year a college awards five merit-based scholarships to members of the entering freshmen class who have exceptional high school records. The initial pool of applicants for the upcoming academic year has been reduced to a "short list" of eight men and ten women, all of whom seem equally deserving. If the awards are made at random from among the 18 finalists, what are the chances that both men and women will be represented?

3.2.24. A local lottery is conducted weekly by choosing five chips at random and without replacement from a population of 40 chips, numbered 1 through 40; order does not matter. The winning numbers are announced on five successive commercials during the Monday night broadcast of a televised movie. Suppose the first three winning numbers match three of yours. What are your chances at that point of winning the lottery?

3.2.25. A display case contains 35 gems, of which 10 are real diamonds and 25 are fake diamonds. A burglar removes four gems at random, one at a time and without replacement. What is the probability that the last gem she steals is the second real diamond in the set of four?

3.2.26. A bleary-eyed student awakens one morning, late for an 8:00 class, and pulls two socks out of a drawer that contains two black, six brown, and two blue socks, all randomly arranged. What is the probability that the two he draws are a matched pair?

3.2.27. Show directly that the set of probabilities associated with the hypergeometric distribution sum to 1. *Hint:* Expand the identity

$$(1 + \mu)^N = (1 + \mu)^r (1 + \mu)^{N-r}$$

and equate coefficients.

3.2.28. Urn I contains five red chips and four white chips; Urn II contains four red and five white chips. Two chips are drawn simultaneously from Urn I and placed in Urn II.

Then a single chip is drawn from Urn II. What is the probability that the chip drawn from Urn II is white? *Hint:* Use Theorem 2.4.1.

3.2.29. As the owner of a chain of sporting goods stores, you have just been offered a "deal" on a shipment of 100 robot table tennis machines. The price is right, but the prospect of picking up the merchandise at midnight from an unmarked van parked on the side of the New Jersey Turnpike is a bit disconcerting. Being of low repute yourself, you do not consider the legality of the transaction to be an issue, but you do have concerns about being cheated. If too many of the machines are in poor working order, the offer ceases to be a bargain. Suppose you decide to close the deal only if a sample of 10 machines contains no more than one defective. Construct the corresponding operating characteristic curve. For approximately what incoming quality will you accept a shipment 50% of the time?

3.2.30. Suppose that r of N chips are red. Divide the chips into three groups of sizes n_1, n_2, and n_3, where $n_1 + n_2 + n_3 = N$. Generalize the hypergeometric distribution to find the probability that the first group contains r_1 red chips, the second group r_2 red chips, and the third group r_3 red chips, where $r_1 + r_2 + r_3 = r$.

3.2.31. Some nomadic tribes, when faced with a life-threatening contagious disease, will try to improve their chances of survival by dispersing into smaller groups. Suppose a tribe of 21 people, of whom four are carriers of the disease, split into three groups of 7 each. What is the probability that at least one group is free of the disease? *Hint:* Find the probability of the complement.

3.2.32. Suppose a population contains n_1 objects of one kind, n_2 objects of a second kind, ..., and n_t objects of a tth kind, where $n_1 + n_2 + \cdots + n_t = N$. A sample of size n is drawn at random and without replacement. Deduce an expression for the probability of drawing k_1 objects of the first kind, k_2 objects of the second kind, ..., and k_t objects of the tth kind by generalizing Theorem 3.2.2.

3.2.33. Sixteen students—five freshmen, four sophomores, four juniors, and three seniors—have applied for membership in their school's Communications Board, a group that oversees the college's newspaper, literary magazine, and radio show. Eight positions are open. If the selection is done at random, what is the probability that each class gets two representatives? (*Hint:* Use the generalized hypergeometric model asked for in Question 3.2.32.)

3.3 DISCRETE RANDOM VARIABLES

The binomial and hypergeometric distributions described in Section 3.2 are special cases of some important general concepts that we want to explore more fully in this section. Previously in Chapter 2, we studied in depth the situation where every point in a sample space is equally likely to occur (recall Section 2.6). The sample space of independent trials that ultimately led to the binomial distribution presented a quite different scenario: specifically, individual points in S had different probabilities. For example, if $n = 4$ and $p = \frac{1}{3}$, the probabilities assigned to the sample points (s, f, s, f) and (f, f, f, f) are $(1/3)^2(2/3)^2 = \frac{4}{81}$ and $(2/3)^4 = \frac{16}{81}$, respectively. Allowing for the possibility that different outcomes may have different probabilities will obviously broaden enormously the range of real-world problems that probability models can address.

How to assign probabilities to outcomes that are not binomial or hypergeometric is one of the major questions investigated in this chapter. A second critical issue is the nature of the sample space itself and whether it makes sense to redefine the outcomes and create, in effect, an alternative sample space. Why we would want to do that has already come up in our discussion of independent trials. The "original" sample space in such cases is a set of ordered sequences, where the ith member of a sequence is either an "s" or an "f," depending on whether the ith trial ended in either success or failure, respectively. However, knowing which particular trials ended in success is typically less important than knowing the *number* that did (recall the clinical trial discussion on p. 129). That being the case, it often makes sense to replace each ordered sequence with the number of successes that sequence contains. Doing so collapses the original set of 2^n ordered sequences (i.e., outcomes) in S to the set of $n + 1$ integers ranging from 0 to n. The probabilities assigned to those integers, of course, are given by the binomial formula in Theorem 3.2.1.

In general, a function that assigns numbers to outcomes is called a *random variable.* The purpose of such functions in practice is to define a new sample space whose outcomes speak more directly to the objectives of the experiment. That was the rationale that ultimately motivated both the binomial and hypergeometric distributions.

The purpose of this section is to (1) outline the general conditions under which probabilities can be assigned to sample spaces and (2) explore the ways and means of redefining sample spaces through the use of random variables. The notation introduced in this section is especially important and will be used throughout the remainder of the book.

Assigning Probabilities: The Discrete Case

We begin with the general problem of assigning probabilities to sample outcomes, the simplest version of which occurs when the number of points in S is either finite or countably infinite. The probability functions, $p(s)$, that we are looking for in those cases satisfy the conditions in Definition 3.3.1.

Definition 3.3.1. Suppose that S is a finite or countably infinite sample space. Let p be a real-valued function defined for each element of S such that

a. $0 \leq p(s)$ for each $s \in S$

b. $\sum_{\text{all } s \in S} p(s) = 1$

Then p is said to be a *discrete probability function.*

Comment. Once $p(s)$ is defined for all s, it follows that the probability of any event A—that is, $P(A)$—is the sum of the probabilities of the outcomes comprising A:

$$P(A) = \sum_{\text{all } s \in A} p(s) \tag{3.3.1}$$

Defined in this way, the function $P(A)$ satisfies the probability axioms given in Section 2.3. The next several examples illustrate some of the specific forms that $p(s)$ can have and how $P(A)$ is calculated.

EXAMPLE 3.3.1

Ace-six flats are a type of crooked dice where the cube is foreshortened in the one-six direction, the effect being that 1s and 6s are more likely to occur than any of the other four faces. Let $p(s)$ denote the probability that the face showing is s. For many ace-six flats, the "cube" is asymmetric to the extent that $p(1) = p(6) = \frac{1}{4}$, while $p(2) = p(3) = p(4) = p(5) = \frac{1}{8}$. Notice that $p(s)$ here qualifies as a discrete probability function because each $p(s)$ is greater than or equal to 0 and the sum of $p(s)$, over all s, is $1(= 2(\frac{1}{4}) + 4(\frac{1}{8}))$.

Suppose A is the event that an even number occurs. It follows from Equation 3.3.1 that $P(A) = P(2) + P(4) + P(6) = \frac{1}{8} + \frac{1}{8} + \frac{1}{4} = \frac{1}{2}$.

Comment. If two ace-six flats are rolled, the probability of getting a sum equal to seven is equal to $2p(1)p(6) + 2p(2)p(5) + 2p(3)p(4) = 2(\frac{1}{4})^2 + 4(\frac{1}{8})^2 = \frac{3}{16}$. If two *fair* dice are rolled, the probability of getting a sum equal to seven is $2p(1)p(6) + 2p(2)p(5) + 2p(3)p(4) = 6(\frac{1}{6})^2 = \frac{1}{6}$, which is less than $\frac{3}{16}$. Gamblers cheat with ace-six flats by switching back and forth between fair dice and ace-six flats, depending on whether or not they want a sum of seven to be rolled.

EXAMPLE 3.3.2

Suppose a fair coin is tossed until a head comes up for the first time. What are the chances of that happening on an odd-numbered toss?

Note that the sample space here is countably infinite and so is the set of outcomes making up the event whose probability we are trying to find. The $P(A)$ that we are looking for, then, will be the sum of an infinite number of terms.

Let $p(s)$ be the probability that the first head appears on the sth toss. Since the coin is presumed to be fair, $p(1) = \frac{1}{2}$. Furthermore, we would expect half the time, when a tail appears, the next toss would be a head, so $p(2) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$. In general, $p(s) = (\frac{1}{2})^s$, $s = 1, 2, \ldots$

Does $p(s)$ satisfy the conditions stated in Definition 3.3.1? Yes. Clearly, $p(s) \geq 0$ for all s. To see that the sum of the probabilities is 1, recall the formula for the sum of a geometric series: If $0 < r < 1$,

$$\sum_{s=0}^{\infty} r^s = \frac{1}{1-r} \tag{3.3.2}$$

Applying Equation 3.3.2 to the sample space here confirms that $P(S) = 1$:

$$P(S) = \sum_{s=1}^{\infty} p(s) = \sum_{s=1}^{\infty} \left(\frac{1}{2}\right)^s = \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^s - \left(\frac{1}{2}\right)^0 = 1 \Big/ \left(1 - \frac{1}{2}\right) - 1 = 1$$

Now, let A be the event that the first head appears on an odd-numbered toss. Then $P(A) = p(1) + p(3) + p(5) + \cdots$ But

$$p(1) + p(3) + p(5) + \cdots = \sum_{s=0}^{\infty} p(2s + 1) = \sum_{s=0}^{\infty} \left(\frac{1}{2}\right)^{2s+1} = \left(\frac{1}{2}\right) \sum_{s=0}^{\infty} \left(\frac{1}{4}\right)^{s}$$

$$= \left(\frac{1}{2}\right)\left(1 \Big/ \left(1 - \frac{1}{4}\right)\right) = \frac{2}{3}$$

CASE STUDY 3.3.1

For good pedagogical reasons, the principles of probability are always introduced by considering events defined on familiar sample spaces generated by simple experiments. To that end, we toss coins, deal cards, roll dice, and draw chips from urns. It would be a serious error, though, to infer that the importance of probability extends no further than the nearest casino. In its infancy, gambling and probability were, indeed, intimately related: Questions arising from games of chance were often the catalyst that motivated mathematicians to study random phenomena in earnest. But more than 340 years have passed since Huygens published *De Ratiociniis*. Today, the application of probability to gambling is relatively insignificant (the NCAA March basketball tournament notwithstanding) compared to the depth and breadth of uses the subject finds in business, medicine, engineering, and science.

Probability functions—properly chosen—can "model" complex real-world phenomena every bit as well as $P(\text{heads}) = \frac{1}{2}$ describes the behavior of a fair coin. The following set of actuarial data is a case in point. Over a period of three years (= 1096 days) in London, records showed that a total of 903 deaths occurred among males eighty-five years of age and older (188). Columns one and two of Table 3.3.1 give the breakdown of those 903 deaths according to the number occurring on a given day. Column three gives the *proportion* of days for which exactly s elderly men died.

TABLE 3.3.1

(1) Number of Deaths, s	(2) Number of Days	(3) Proportion [= Col.(2)/1096]	(4) $p(s)$
0	484	0.442	0.440
1	391	0.357	0.361
2	164	0.150	0.148
3	45	0.041	0.040
4	11	0.010	0.008
5	1	0.001	0.003
6+	0	0.000	0.000
	1096	1	1

For reasons that will be gone into at length in Chapter 4, the probability function that describes the behavior of this particular phenomenon is

$$p(s) = P(s \text{ elderly men die on a given day})$$
$$= \frac{e^{-0.82}(0.82)^s}{s!}, \qquad s = 0, 1, 2, \ldots \tag{3.3.3}$$

How do we know that the $p(s)$ in Equation 3.3.3 is an appropriate way to assign probabilities to the "experiment" of elderly men dying? Because it accurately predicts what happened. Column four of Table 3.3.1 shows $p(s)$ evaluated for $s = 0, 1, 2, \ldots$ To two decimal places, the agreement between the entries in Column three and Column four is perfect.

EXAMPLE 3.3.3

Consider the following experiment: Every day for the next month you copy down each number that appears in the stories on the front pages of your hometown newspaper. Those numbers would necessarily be extremely diverse: One might be the age of a celebrity who just died, another might report the interest rate currently paid on government Treasury bills, and still another might give the number of square feet of retail space recently added to a local shopping mall.

Suppose you then calculated the proportion of those numbers whose leading digit was a 1, the proportion whose leading digit was a 2, and so on. What relationship would you expect those proportions to have? Would numbers starting with a 2, for example, occur as often as numbers starting with a 6?

Let $p(s)$ denote the probability that the first significant digit of a "newspaper number" is s, $s = 1, 2, \ldots, 9$. Our intuition is likely to tell us that the nine first digits should be equally probable—that is, $p(1) = p(2) = \cdots = p(9) = \frac{1}{9}$. Given the diversity and the randomness of the numbers, there is no obvious reason why one digit should be more common than another. Our intuition, though, would be wrong—first digits are *not* equally likely. Indeed, they are not even close to being equally likely!

Credit for making this remarkable discovery goes to Simon Newcomb, a mathematician who observed more than a hundred years ago that some portions of logarithm tables are used more than others (77). Specifically, pages at the beginning of such tables are more dog-eared than pages at the end, suggesting that users had more occasion to look up logs of numbers starting with small digits than they did numbers starting with large digits.

Almost fifty years later, a physicist, Frank Benford, reexamined Newcomb's claim in more detail and looked for a mathematical explanation. What is now known as *Benford's law* asserts that the first digits of many different types of measurements, or combinations of measurements, often follow the discrete probability model:

$$p(s) = P(\text{1st significant digit is } s) = \log\left(1 + \frac{1}{s}\right), \qquad s = 1, 2, \ldots, 9$$

Table 3.3.2 compares Benford's law to the uniform assumption that $p(s) = \frac{1}{9}$, for all s. The differences are striking. According to Benford's law, for example, 1s are the most frequently occurring first digit, appearing 6.5 times (= 0.301/0.046) as often as 9s.

TABLE 3.3.2

s	"Uniform" Law	Benford's Law
1	0.111	0.301
2	0.111	0.176
3	0.111	0.125
4	0.111	0.097
5	0.111	0.079
6	0.111	0.067
7	0.111	0.058
8	0.111	0.051
9	0.111	0.046

Comment. A key to *why* Benford's law is true are the differences in proportional changes associated with each leading digit. To go from one thousand to two thousand, for example, represents a 100% increase; to go from eight thousand to nine thousand, on the other hand, is only a 12.5% increase. That would suggest that evolutionary phenomena such as stock prices would be more likely to start with 1s and 2s than with 8s and 9s—and they are. Still, the precise conditions under which $p(s) = \log\left(1 + \frac{1}{s}\right), s = 1, 2, \ldots, 9$ are not fully understood and remain a topic of research.

EXAMPLE 3.3.4

Is

$$p(s) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s, \quad s = 0, 1, 2, \ldots; \quad \lambda > 0$$

a discrete probability function? Why or why not?

To qualify as a discrete probability function, a given $p(s)$ needs to satisfy Parts (a) and (b) of Definition 3.3.1. A simple inspection shows that Part (a) is satisfied. Since $\lambda > 0$, $p(s)$ is, in fact, greater than or equal to 0 for all $s = 0, 1, 2, \ldots$ Part (b) is satisfied if the sum of all the probabilities defined on the outcomes in S is 1. But

$$\begin{aligned}\sum_{\text{all } s \in S} p(s) &= \sum_{s=0}^{\infty} \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s \\ &= \frac{1}{1+\lambda}\left(\frac{1}{1-\dfrac{\lambda}{1+\lambda}}\right) \qquad \text{why?} \\ &= \frac{1}{1+\lambda} \cdot \frac{1+\lambda}{1} \\ &= 1\end{aligned}$$

The answer, then, is "yes"—$p(s) = \frac{1}{1+\lambda}\left(\frac{\lambda}{1+\lambda}\right)^s$, $s = 0, 1, 2, \ldots; \lambda > 0$ *does* qualify as a discrete probability function. Of course, whether it has any practical value depends on whether the set of values for $p(s)$ actually do describe the behavior of real-world phenomena.

Defining "New" Sample Spaces

We have seen how the function $p(s)$ associates a probability with each outcome, s, in a sample space. Related is the key idea that outcomes can often be grouped or reconfigured in ways that may facilitate problem-solving. Recall the sample space associated with a series of n independent trials, where each s is an ordered sequence of successes and failures. The most relevant information in such outcomes is often the *number* of successes that occur, not a detailed listing of which trials ended in success and which ended in failure. That being the case, it makes sense to define a "new" sample space by grouping the original outcomes according to the number of successes they contained. The outcome $(f, f, \ldots, f)$, for example, had *0* successes. On the other hand, there were n outcomes that yielded *1* success—$(s, f, f, \ldots, f)$, $(f, s, f, \ldots, f)$, $\ldots$, and $(f, f, \ldots, s)$. As we saw earlier in this chapter, that particular regrouping of outcomes ultimately led to the binomial distribution.

The function that replaces the outcome $(s, f, f, \ldots, f)$ with the numerical value *1* is called a *random variable*. We conclude this section with a discussion of some of the concepts, terminology, and applications associated with random variables.

Definition 3.3.2. A function whose domain is a sample space S and whose values form a finite or countably infinite set of real numbers is called a *discrete random variable*. We denote random variables by upper case letters, often X or Y.

EXAMPLE 3.3.5

Consider tossing two dice, an experiment for which the sample space is a set of ordered pairs, $S = \{(i, j) \mid i = 1, 2, \ldots, 6;\ j = 1, 2, \ldots, 6\}$. For a variety of games ranging from Monopoly to craps, the *sum* of the numbers showing is what matters on a given turn. That being the case, the original sample space S of thirty-six ordered pairs would not provide a particularly convenient backdrop for discussing the rules of those games. It would be better to work directly with the sums. Of course, the eleven possible sums (from two to twelve) are simply the different values of the random variable X, where $X(i, j) = i + j$.

Comment. In the above example, suppose we define a random variable X_1 that gives the result on the first die and X_2 that gives the result on the second die. Then $X = X_1 + X_2$. Note how easily we could extend this idea to the toss of *three* dice, or *ten* dice. The ability to conveniently express complex events in terms of simpler ones is an advantage of the random variable concept that we will see playing out over and over again.

The Probability Density Function

We began this section discussing the function $p(s)$, which assigns a probability to each outcome s in S. Now, having introduced the notion of a random variable X as a real-valued function defined on S—that is, $X(s) = k$—we need to find a mapping analogous to $p(s)$ that assigns probabilities to the different values of k.

Definition 3.3.3. Associated with every discrete random variable X is a *probability density function* (or *pdf*), denoted $p_X(k)$, where

$$p_X(k) = P(\{s \in S \mid X(s) = k\})$$

Note that $p_X(k) = 0$ for any k not in the range of X. For notational simplicity, we will usually delete all references to s and S and write $p_X(k) = P(X = k)$.

Comment. We have already discussed at length two examples of the function $p_X(k)$. Recall the binomial distribution derived in Section 3.2. If we let the random variable X denote the number of successes in n independent trials, then Theorem 3.2.1 states that

$$P(X = k) = p_X(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \dots, n$$

A similar result was given in that same section in connection with the hypergeometric distribution. If a sample of size n is drawn without replacement from an urn containing r red chips and w white chips, and if we let the random variable X denote the number of red chips included in the sample, then (according to Theorem 3.2.2),

$$P(X = k) = p_X(k) = \binom{r}{k}\binom{w}{n - k} \Bigg/ \binom{r + w}{n}$$

EXAMPLE 3.3.6

Consider again the rolling of two dice as described in Example 3.3.5. Let i and j denote the faces showing on the first and second die, respectively, and define the random variable X to be the sum of the two faces: $X(i, j) = i + j$. Find $p_X(k)$.

According to Definition 3.3.3, each value of $p_X(k)$ is the sum of the probabilities of the outcomes that get mapped by X onto the value k. For example,

$$\begin{aligned} P(X = 5) &= p_X(5) = P(\{s \in S \mid X(s) = 5\}) \\ &= P((1, 4) \cup (4, 1) \cup (2, 3) \cup (3, 2)) \\ &= P(1, 4) + P(4, 1) + P(2, 3) + P(3, 2) \\ &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\ &= \frac{4}{36} \end{aligned}$$

TABLE 3.3.3

k	$p_X(k)$	k	$p_X(k)$
2	1/36	8	5/36
3	2/36	9	4/36
4	3/36	10	3/36
5	4/36	11	2/36
6	5/36	12	1/36
7	6/36		

assuming the dice are fair. Values of $p_X(k)$ for other k are calculated similarly. Table 3.3.3 shows the random variable's entire pdf.

EXAMPLE 3.3.7

Acme Industries typically produces three electric power generators a day; some pass the company's quality control inspection on their first try and are ready to be shipped; others need to be retooled. The probability of a generator needing further work is 0.05. If a generator is ready to ship, the firm earns a profit of \$10,000. If it needs to be retooled, it ultimately costs the firm \$2000. Let X be the random variable quantifying the company's daily profit. Find $p_X(k)$.

The underlying sample space here is a set of $n = 3$ independent trials, where $p = P(\text{generator passes inspection}) = 0.95$. If the random variable X is to measure the company's daily profit, then

$$\begin{aligned} X = \$10{,}000 &\times (\text{no. of generators passing inspection}) \\ &- \$2{,}000 \times (\text{no. of generators needing retooling}) \end{aligned}$$

For instance, $X(s, f, s) = 2(\$10{,}000) - 1(\$2{,}000) = \$18{,}000$. Moreover, the random variable X equals \$18,000 whenever the day's output consists of two successes and one failure. That is, $X(s, f, s) = X(s, s, f) = X(f, s, s)$. It follows that

$$P(X = \$18{,}000) = p_X(18{,}000) = \binom{3}{2}(0.95)^2(0.05)^1 = 0.135375$$

TABLE 3.3.4

No. Defectives	k = Profit	$p_X(k)$
0	\$30,000	0.857375
1	\$18,000	0.135375
2	\$6,000	0.007125
3	−\$6,000	0.000125

Table 3.3.4 shows $p_X(k)$ for the four possible values of k (\$30,000, \$18,000, \$6,000, and −\$6,000).

EXAMPLE 3.3.8

As part of her warm-up drill, each player on State's basketball team is required to shoot free throws until two baskets are made. If Rhonda has a 65% success rate at the foul line, what is the pdf of the random variable X that describes the number of throws it takes her to complete the drill? Assume that individual throws constitute independent events.

Figure 3.3.1 illustrates what must occur if the drill is to end on the kth toss, $k = 2, 3, 4, \ldots$: First, Rhonda needs to make exactly one basket sometime during the first $k - 1$ attempts, and, second, she needs to make a basket on the kth toss. Written formally,

$$\begin{aligned} p_X(k) &= P(X = k) = P(\text{drill ends on } k\text{th throw}) \\ &= P((1 \text{ basket and } k\text{-}2 \text{ misses in first } k\text{-}1 \text{ throws}) \cap (\text{basket on } k\text{th throw})) \\ &= P(1 \text{ basket and } k\text{-}2 \text{ misses}) \cdot P(\text{basket}) \end{aligned}$$

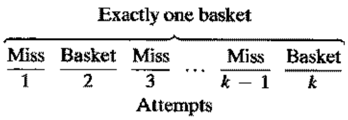


FIGURE 3.3.1

Notice that k-1 different sequences have the property that exactly one of the first k-1 throws results in a basket:

$$k\text{-}1 \text{ sequences} \left\{ \begin{array}{cccccc} \frac{B}{1} & \frac{M}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{M}{k-1} \\ \frac{M}{1} & \frac{B}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{M}{k-1} \\ & & \vdots & & & \\ \frac{M}{1} & \frac{M}{2} & \frac{M}{3} & \frac{M}{4} & \cdots & \frac{B}{k-1} \end{array} \right.$$

Since each sequence has probability $(0.35)^{k-2}(0.65)$,

$$P(1 \text{ basket and } k\text{-}2 \text{ misses}) = (k\text{-}1)(0.35)^{k-2}(0.65)$$

Therefore,

$$\begin{aligned} p_X(k) &= (k\text{-}1)(0.35)^{k-2}(0.65) \cdot (0.65) \\ &= (k\text{-}1)(0.35)^{k-2}(0.65)^2, \quad k = 2, 3, 4, \ldots \end{aligned} \tag{3.3.4}$$

Table 3.3.5 shows the pdf evaluated for specific values of k. Although the range of k is infinite, the bulk of the probability associated with X is concentrated in the values two through seven: It is highly unlikely, for example, that Rhonda would need more than seven shots to complete the drill.

TABLE 3.3.5

k	$p_X(k)$
2	0.4225
3	0.2958
4	0.1553
5	0.0725
6	0.0317
7	0.0133
8+	0.0089

Transformations

Transforming a variable from one scale to another is a problem that is comfortably familiar. If a thermometer says the temperature outside is 83°F, we know that the temperature *in degrees Centigrade* is 28:

$$°\mathrm{C} = \left(\frac{5}{9}\right)(°\mathrm{F} - 32) = \left(\frac{5}{9}\right)(83 - 32) = 28$$

An analogous question arises in connection with random variables. Suppose that X is a discrete random variable with pdf $p_X(k)$. If a second random variable, Y, is defined to be $aX + b$, where a and b are constants, what can be said about the pdf for Y?

Theorem 3.3.1. *Suppose X is a discrete random variable. Let $Y = aX + b$, where a and b are constants. Then $p_Y(y) = p_X\left(\frac{y-b}{a}\right)$.*

Proof. $p_Y(y) = P(Y = y) = P(aX + b = y) = P\left(X = \frac{y-b}{a}\right) = p_X\left(\frac{y-b}{a}\right)$ ☐

EXAMPLE 3.3.9

Let X be a random variable for which $p_X(k) = \frac{1}{10}$, for $k = 1, 2, \ldots, 10$. What is the probability distribution associated with the random variable Y, where $Y = 4X - 1$? That is, find $p_Y(y)$.

From Theorem 3.3.1, $P(Y = y) = P(4X - 1 = y) = P(X = (y + 1)/4) = p_X\left(\frac{y+1}{4}\right)$, which implies that $p_Y(y) = \frac{1}{10}$ for the ten values of $(y + 1)/4$ that equal $1, 2, \ldots, 10$. But $(y + 1)/4 = 1$ when $y = 3$, $(y + 1)/4 = 2$ when $y = 7, \ldots$, $(y + 1)/4 = 10$ when $y = 39$. Therefore, $p_Y(y) = \frac{1}{10}$, for $y = 3, 7, \ldots, 39$.

The Cumulative Distribution Function

In working with random variables, we frequently need to calculate the probability that the value of a random variable is somewhere between two numbers. For example, suppose we have an integer-valued random variable. We might want to calculate an expression like $P(s \le X \le t)$. If we know the pdf for X, then

$$P(s \le X \le t) = \sum_{k=s}^{t} p_X(k).$$

but depending on the nature of $p_X(k)$ and the number of terms that need to be added, calculating the sum of $p_X(k)$ from $k = s$ to $k = t$ may be quite difficult. An alternate strategy is to use the fact that

$$P(s \le X \le t) = P(X \le t) - P(X \le s - 1)$$

where the two probabilities on the right represent *cumulative* probabilities of the random variable X. If the latter were available (and they often are), then evaluating $P(s \le X \le t)$ by one simple subtraction would clearly be easier than doing all the calculations implicit in $\sum_{k=s}^{t} p_X(k)$.

Definition 3.3.4. Let X be a discrete random variable. For any real number t, the probability that X takes on a value $\le t$ is the *cumulative distribution function* (*cdf*) of X (written $F_X(t)$). In formal notation, $F_X(t) = P(\{s \in S \mid X(s) \le t\})$. As was the case with pdfs, references to s and S are typically deleted, and the cdf is written $F_X(t) = P(X \le t)$.

EXAMPLE 3.3.10

Suppose we wish to compute $P(21 \le X \le 40)$ for a binomial random variable X with $n = 50$ and $p = 0.6$. From Theorem 3.2.1, we know the formula for $p_X(k)$, so $P(21 \le X \le 40)$ can be written as a simple, although computationally cumbersome, sum:

$$P(21 \le X \le 40) = \sum_{k=21}^{40} \binom{50}{k} (0.6)^k (0.4)^{50-k}$$

Equivalently, the probability we are looking for can be expressed as the difference between two cdfs:

$$P(21 \le X \le 40) = P(X \le 40) - P(X \le 20) = F_X(40) - F_X(20)$$

As it turns out, values of the cdf for a binomial random variable are widely available, both in books and in computer software. Here, for example, $F_X(40) = 0.9992$ and $F_X(20) = 0.0034$, so

$$\begin{aligned} P(21 \le X \le 40) &= 0.9992 - 0.0034 \\ &= 0.9958 \end{aligned}$$

EXAMPLE 3.3.11

Suppose that two fair dice are rolled. Let the random variable X denote the larger of the two faces showing: (a) Find $F_X(t)$ for $t = 1, 2, \ldots, 6$ and (b) Find $F_X(2.5)$.

a. The sample space associated with the experiment of rolling two fair dice is the set of ordered pairs, $s = (i, j)$, where the face showing on the first die is i and the face showing on the second die is j. By assumption, all 36 possible outcomes are equally likely. Now, suppose t is some integer from 1 to 6, inclusive. Then

$$\begin{aligned} F_X(t) &= P(X \leq t) \\ &= P(\text{Max}\,(i, j) \leq t) \\ &= P(i \leq t \quad \text{and} \quad j \leq t) && \text{(why?)} \\ &= P(i \leq t) \cdot P(j \leq t) && \text{(why?)} \\ &= \frac{t}{6} \cdot \frac{t}{6} \\ &= \frac{t^2}{36}, \quad t = 1, 2, 3, 4, 5, 6 \end{aligned}$$

b. Even though the random variable X has non-zero probability only for the integers 1 through 6, the cdf is defined for *any* real number from $-\infty$ to $+\infty$. By definition, $F_X(2.5) = P(X \leq 2.5)$. But

$$\begin{aligned} P(X \leq 2.5) &= P(X \leq 2) + P(2 < X \leq 2.5) \\ &= F_X(2) + 0 \end{aligned}$$

so

$$F_X(2.5) = F_X(2) = \frac{2^2}{36} = \frac{1}{9}$$

What would the graph of $F_X(t)$ as a function of t look like?

QUESTIONS

3.3.1. An urn contains five balls numbered 1 to 5. Two balls are drawn simultaneously.
  (a) Let X be the larger of the two numbers drawn. Find $p_X(k)$.
  (b) Let V be the sum of the two numbers drawn. Find $p_V(k)$.

3.3.2. Repeat Question 3.3.1. for the case where the two balls are drawn *with replacement*.

3.3.3. Suppose a fair die is tossed three times. Let X be the largest of the three faces that appear. Find $p_x(k)$.

3.3.4. Suppose a fair die is tossed three times. Let X be the number of different faces that appear (so $X = 1, 2,$ or 3). Find $p_x(k)$.

3.3.5. A fair coin is tossed three times. Let X be the number of heads in the tosses minus the number of tails. Find $p_x(k)$.

3.3.6. Suppose die one has spots $1, 2, 2, 3, 3, 4$ and die two has spots $1, 3, 4, 5, 6, 8$. If both dice are rolled, what is the sample space? Let X = total spots showing. Show that the pdf for X is the same as for normal dice.

3.3.7. Suppose a particle moves along the x-axis beginning at 0. It moves one integer step to the left or right with equal probability. What is the pdf of its position after 4 steps?

3.3.8. How would the pdf asked for in Question 3.3.7. be affected if the particle was twice as likely to move to the right as to the left?

3.3.9. Suppose that five people, including you and a friend, line up at random. Let the random variable X denote the number of people standing between you and your friend. What is $p_x(k)$?

3.3.10. Urn I and Urn II each have two red chips and two white chips. Two chips are drawn simultaneously from each urn. Let X_1 be the number of red chips in the first sample and X_2 the number of red chips in the second sample. Find the pdf of $X_1 + X_2$.

3.3.11. Suppose X is a binomial random variable with $n = 4$ and $p = \frac{1}{3}$. What is the pdf of $2X + 1$?

3.3.12. Find the cdf for the random variable X in Question 3.3.3.

3.3.13. A fair die is rolled four times. Let the random variable X denote the number of 6's that appear. Find and graph the cdf for X.

3.3.14. At the points $x = 0, 1, \ldots, 6$, the cdf for the discrete random variable X has the value $F_X(x) = x(x + 1)/42$. Find the pdf for X.

3.3.15. Find the pdf for the discrete random variable X whose cdf at the points $x = 0, 1, \ldots, 6$ is given by $F_X(x) = x^3/216$.

CONTINUOUS RANDOM VARIABLES

The statement was made in Chapter 2 that all sample spaces belong to one of two generic types—*discrete* sample spaces are ones that contain a finite or a countably infinite number of outcomes and *continuous* sample spaces are those that contain an uncountably infinite number of outcomes. Rolling a pair of dice and recording the faces that appear is an experiment with a discrete sample space; choosing a number at random from the interval [0, 1] would have a continuous sample space.

How we assign probabilities to these two types of sample spaces is different. Section 3.3 focussed on discrete sample spaces. Each outcome s is assigned a probability by the discrete probability function $p(s)$. If a random variable X is defined on the sample space, the probabilities associated with its outcomes are assigned by the probability density function $p_X(k)$. Applying those same definitions, though, to the outcomes in a continuous sample space will not work. The fact that a continuous sample space has an uncountably infinite number of outcomes eliminates the option of assigning a probability to each point as we did in the discrete case with the function $p(s)$. We begin this section with a particular pdf defined on a discrete sample space that suggests how we might define probabilities, in general, on a continuous sample space.

Suppose an electronic surveillance monitor is turned on briefly at the beginning of every hour and has a 0.905 probability of working properly, regardless of how long it has remained in service. If we let the random variable X denote the hour at which the monitor first fails, then $p_X(k)$ is the product of k individual probabilities:

$$\begin{aligned} p_X(k) &= P(X = k) = P(\text{monitor fails for the first time at the } k\text{th hour}) \\ &= P(\text{monitor functions properly for first } k-1 \text{ hours} \cap \text{monitor fails at the } k\text{th hour}) \\ &= (0.905)^{k-1}(0.095), \quad k = 1, 2, 3, \ldots \end{aligned}$$

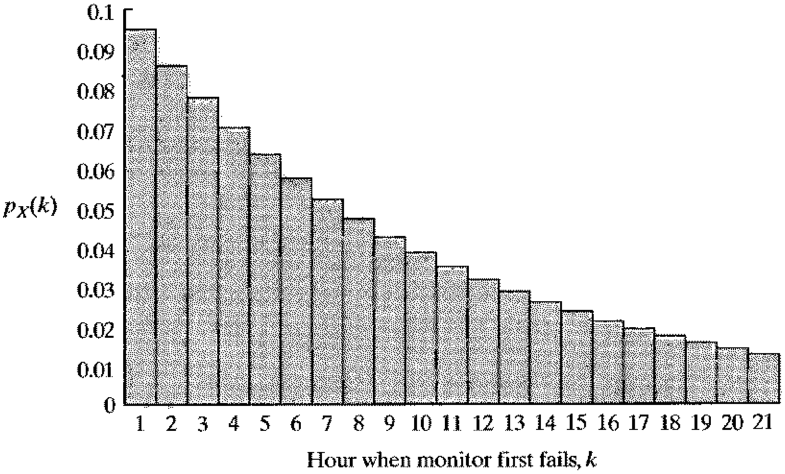


FIGURE 3.4.1

Figure 3.4.1 shows a probability histogram of $p_X(k)$ for k values ranging from 1 to 21. Here the height of the kth bar is $p_X(k)$, and since the width of each bar is 1, the *area* of the kth bar is also $p_X(k)$.

Now, look at Figure 3.4.2, where the exponential curve $y = 0.1e^{-0.1x}$ is superimposed on the graph of $p_X(k)$. Notice how closely the area under the curve approximates the area of the bars. It follows that the probability that X lies in some given interval will be numerically similar to the integral of the exponential curve above that same interval.

For example, the probability that the monitor fails sometime during the first four hours would be the sum

$$\begin{aligned} P(0 \le X \le 4) &= \sum_{k=0}^{4} p_X(k) \\ &= \sum_{k=0}^{4} (0.905)^{k-1}(0.095) \\ &= 0.3297 \end{aligned}$$

To four decimal places, the corresponding area under the exponential curve is the same:

$$\int_0^4 0.1e^{-0.1x}\, dx = 0.3297$$

Implicit in the similarity here between $p_X(k)$ and the exponential curve $y = 0.1e^{-0.1x}$ is our sought-after alternative to $p(s)$ for continuous sample spaces. Instead of defining probabilities for individual points, we will define probabilities for *intervals* of points, and those probabilities will be areas under the graph of some function (such as $y = 0.1e^{-0.1x}$), where the shape of the function will reflect the desired probability "measure" to be associated with the sample space.

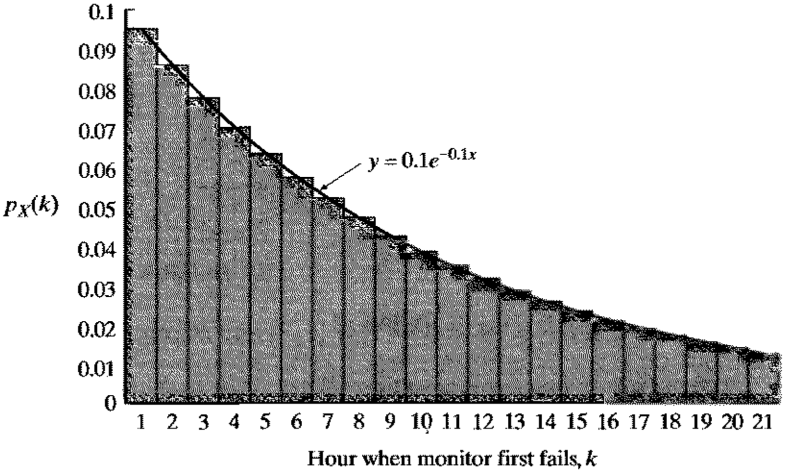


FIGURE 3.4.2

Definition 3.4.1. A probability function P on a set of real numbers S is called *continuous* if there exists a function $f(t)$ such that for any closed interval $[a, b] \subset S$, $P([a, b]) = \int_a^b f(t)\,dt$.

Comment. If a probability function P satisfies Definition 3.4.1, then $P(A) = \int_A f(t)\,dt$ for any set A where the integral is defined.

Conversely, suppose a function $f(t)$ has the two properties

1. $f(t) \geq 0$ for all t
2. $\int_{-\infty}^{\infty} f(t)\,dt = 1$.

If $P(A) = \int_A f(t)\,dt$ for all A, then P will satisfy the probability axioms given in Section 2.3.

Choosing the Function $f(t)$

We have seen that the probability structure of any sample space with a finite or countably infinite number of outcomes is defined by the function $p(s) = P(\text{outcome is } s)$. For sample spaces having an uncountably infinite number of possible outcomes, the function $f(t)$ serves an analogous purpose. Specifically, $f(t)$ defines the probability structure of S in the sense that the probability of any *interval* in the sample space is the *integral* of $f(t)$. The next set of examples illustrate several different choices for $f(t)$.

EXAMPLE 3.4.1

The continuous equivalent of the equiprobable probability model on a discrete sample space is the function $f(t)$ defined by $f(t) = 1/(b - a)$ for all t in the interval $[a, b]$ (and $f(t) = 0$, otherwise). This particular $f(t)$ places equal probability weighting on every closed interval of the same length contained in the interval $[a, b]$. For example, suppose

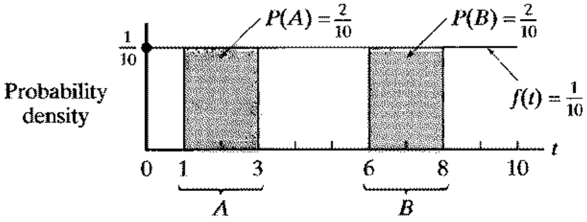


FIGURE 3.4.3

$a = 0$ and $b = 10$, and let $A = [1, 3]$ and $B = [6, 8]$. Then $f(t) = \frac{1}{10}$, and

$$P(A) = \int_1^3 \left(\frac{1}{10}\right) dt = \frac{2}{10} = P(B) = \int_6^8 \left(\frac{1}{10}\right) dt$$

(see Figure 3.4.3).

EXAMPLE 3.4.2

Could $f(t) = 3t^2$, $0 \le t \le 1$ be used to define the probability function for a continuous sample space whose outcomes consisted of all the real numbers in the interval [0, 1]? Yes, because (1) $f(t) \ge 0$ for all t, and (2) $\int_0^1 f(t)\,dt = \int_0^1 3t^2\,dt = t^3\big|_0^1 = 1$.

Notice that the shape of $f(t)$ (see Figure 3.4.4) implies that outcomes close to 1 are more likely to occur than are outcomes close to 0. For example, $P([0, \frac{1}{3}]) = \int_0^{1/3} 3t^2\,dt = t^3\big|_0^{1/3} = \frac{1}{27}$, while $P([\frac{2}{3}, 1]) = \int_{2/3}^1 3t^2\,dt = t^3 \int_{2/3}^1 = 1 - \frac{8}{27} = \frac{19}{27}$.

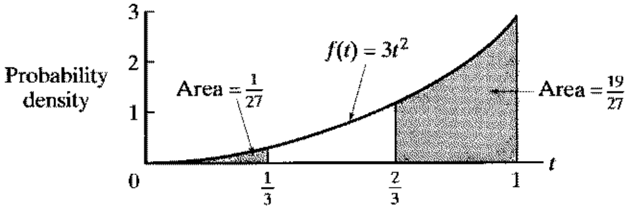


FIGURE 3.4.4

EXAMPLE 3.4.3

By far the most important of all continuous probability functions is the "bell-shaped" curve, known more formally as the *normal* (or *Gaussian*) *distribution*. The sample space for the normal distribution is the entire real line; its probability function is given by

$$f(t) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right], \quad -\infty < t < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$$

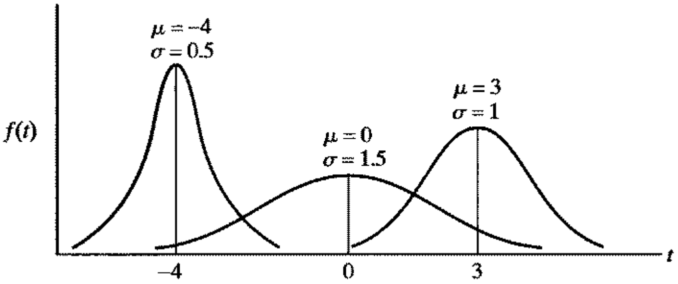


FIGURE 3.4.5

Depending on the values assigned to the parameters μ and σ, $f(t)$ can take on a variety of shapes and locations; three are illustrated in Figure 3.4.5.

Fitting $f(t)$ to Data: The Density-Scaled Histogram

The notion of using a continuous probability function to approximate an integer-valued discrete probability model has already been discussed (recall Figure 3.4.2). The "trick" there was to replace the spikes that define $p_X(k)$ with rectangles whose heights are $p_X(k)$ and whose widths are one. Doing that makes the sum of the areas of the rectangles corresponding to $p_X(k)$ equal to one, which is the same as the total area under the approximating continuous probability function. Because of the equality of those two areas, it makes sense to superimpose (and compare) the "histogram" of $p_X(k)$ and the continuous probability function on the same set of axes.

Now, consider the related, but slightly more general, problem of using a continuous probability function to model the distribution of a set of n measurements, $y_1, y_2, \ldots, y_n$. Following the approach taken in Figure 3.4.2, we would start by making a histogram of the n observations. The problem is, the sum of the areas of the bars comprising that histogram would not necessarily equal one.

As a case in point, Table 3.4.1 shows a set of forty observations. Grouping those y_i's into five classes, each of width ten, produces the distribution and histogram pictured in Figure 3.4.6. Furthermore, suppose we have reason to believe that these forty y_i's may be a random sample from a uniform probability function defined over the interval [20, 70]—that is,

$$f(t) = \frac{1}{70 - 20} = \frac{1}{50}, \quad 20 \leq t \leq 70$$

TABLE 3.4.1

33.8	62.6	42.3	62.9	32.9	58.9	60.8	49.1	42.6	59.8
41.6	54.5	40.5	30.3	22.4	25.0	59.2	67.5	64.1	59.3
24.9	22.3	69.7	41.2	64.5	33.4	39.0	53.1	21.6	46.0
28.1	68.7	27.6	57.6	54.8	48.9	68.4	38.4	69.0	46.6

Class	Frequency
$20 \le y < 30$	7
$30 \le y < 40$	6
$40 \le y < 50$	9
$50 \le y < 60$	8
$60 \le y < 70$	10
	40

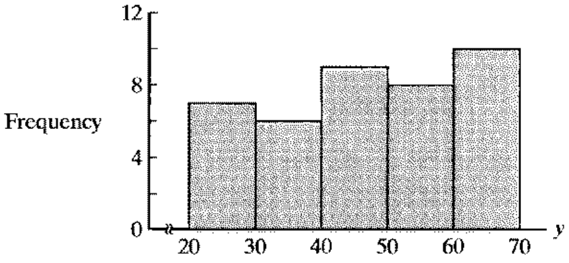


FIGURE 3.4.6

(recall Example 3.4.1). How can we appropriately draw the distribution of the y_i's and the uniform probability model on the same graph?

Note, first, that $f(t)$ and the histogram are not compatible in the sense that the area under $f(t)$ is (necessarily) *one* ($= 50 \times \frac{1}{50}$), but the sum of the areas of the bars making up the histogram is *four hundred:*

$$\begin{aligned}\text{histogram area} &= 10(7) + 10(6) + 10(9) + 10(8) + 10(10)\\ &= 400\end{aligned}$$

Nevertheless, we can "force" the total area of the five bars to match the area under $f(t)$ by redefining the scale of the vertical axis on the histogram. Specifically, *frequency* needs to be replaced with the analog of *probability density,* which would be the scale used on the vertical axis of any graph of $f(t)$. Intuitively, the density associated with, say, the interval [20, 30) would be defined as the quotient

$$\frac{7}{40 \times 10}$$

because integrating that constant over the interval [20, 30) would give $\frac{7}{40}$, and the latter does represent the estimated probability that an observation belongs to the interval [20, 30).

Figure 3.4.7 shows a histogram of the data in Table 3.4.1 where the height of each bar has been converted to a *density,* according to the formula

$$\text{density (of a class)} = \frac{\text{class frequency}}{\text{total no. of observations} \times \text{class width}}$$

Superimposed is the uniform probability model, $f(t) = \frac{1}{50}$, $20 \le t \le 70$. Scaled in this fashion, areas under both $f(t)$ and the histogram are one.

In practice, density-scaled histograms offer a simple, but effective, format for examining the "fit" between a set of data and a presumed continuous model. We will use it often in the chapters ahead. Applied statisticians have especially embraced this particular graphical technique. Indeed, computer software packages that include *Histograms* on their menus routinely give users the choice of putting either *frequency* or *density* on the vertical axis.

Class	Density
$20 \le y < 30$	$7/[40(10)] = 0.0175$
$30 \le y < 40$	$6/[40(10)] = 0.0150$
$40 \le y < 50$	$9/[40(10)] = 0.0225$
$50 \le y < 60$	$8/[40(10)] = 0.0200$
$60 \le y < 70$	$10/[40(10)] = 0.0250$

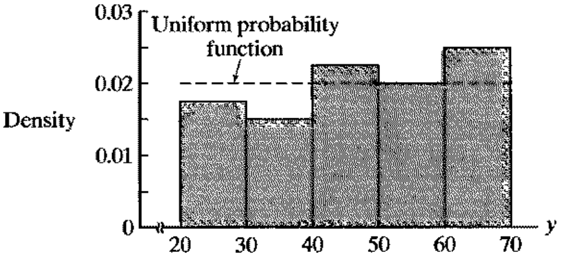


FIGURE 3.4.7

CASE STUDY 3.4.1

Years ago, the V805 transmitter tube was standard equipment on many aircraft radar systems. Table 3.4.2 summarizes part of a reliability study done on the V805; listed are the lifetimes (in hrs) recorded for 903 tubes (37). Grouped into intervals of width eighty, the densities for the nine classes are shown in the last column.

TABLE 3.4.2

Lifetime (hrs)	Number of Tubes	Density
0–80	317	0.0044
80–160	230	0.0032
160–240	118	0.0016
240–320	93	0.0013
320–400	49	0.0007
400–480	33	0.0005
480–560	17	0.0002
560–700	26	0.0002
700+	20	0.0002
	903	

Experience has shown that lifetimes of electrical equipment can often be nicely modeled by the exponential probability function,

$$f(t) = \lambda e^{-\lambda t}, \quad t > 0$$

where the value of λ (for reasons explained in Chapter 5) is set equal to the reciprocal of the average lifetime of the tubes in the sample. Can the distribution of these data also be described by the exponential model?

One way to answer such a question is to superimpose the proposed model on a graph of the density-scaled histogram. The extent to which the two graphs are similar then becomes an obvious measure of the appropriateness of the model.

(Continued on next page)

(Case Study 3.4.1 continued)

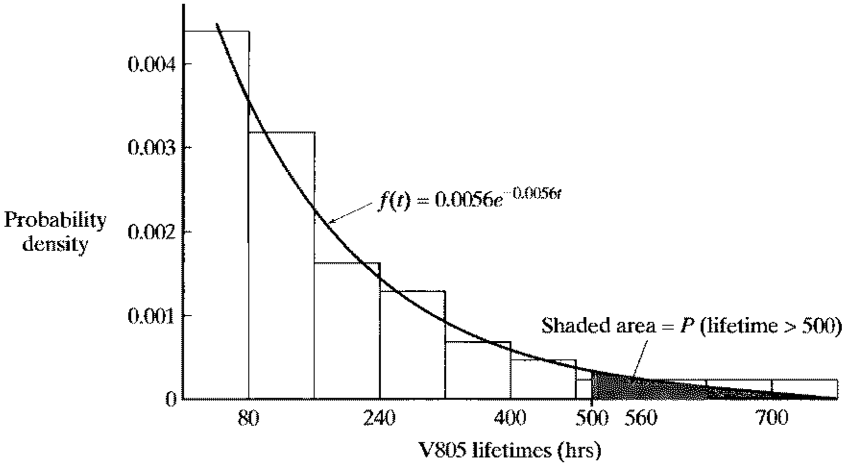


FIGURE 3.4.8

For these data, λ would be 0.0056. Figure 3.4.8 shows the function

$$f(t) = 0.0056e^{-0.0056t}$$

plotted on the same axes as the density-scaled histogram. Clearly, the agreement is excellent, and we would have no reservations about using areas under $f(t)$ to estimate lifetime probabilities. How likely is it, for example, that a V805 tube will last longer than five hundred hrs? Based on the exponential model, that probability would be *0.0608:*

$$\begin{aligned} P(\text{V805 lifetime exceeds 500 hrs}) &= \int_{500}^{\infty} 0.0056e^{-0.0056y}\,dy \\ &= -e^{-0.0056y}\Big|_{500}^{\infty} = e^{-0.0056(500)} = e^{-2.8} = 0.0608 \end{aligned}$$

Continuous Probability Density Functions

We saw in Section 3.3 how the introduction of discrete random variables facilitated the solution of certain problems. The same sort of function can also be defined on sample spaces with an uncountably infinite number of outcomes. In practice, *continuous random variables* are often simply an identity mapping, so they do not radically redefine the sample space in the way that a binomial random variable does. Nevertheless, it helps to have the same notation for both kinds of sample spaces.

Definition 3.4.2. A function Y that maps a subset of the real numbers into the real numbers is called a *continuous random variable*. The *pdf* of Y is the function $f_Y(y)$

having the property that for any numbers a and b,

$$P(a \leq Y \leq b) = \int_a^b f_Y(y)\,dy$$

EXAMPLE 3.4.4

We saw in Case Study 3.4.1 that lifetimes of V805 radar tubes can be nicely modeled by the exponential probability function,

$$f(t) = 0.0056e^{-0.0056t}, \quad t > 0$$

To couch that statement in random variable notation would simply require that we define Y to be the life of a V805 radar tube. Then Y would be the identity mapping and the pdf for the random variable Y would be the same as the probability function, $f(t)$. That is, we would write

$$f_Y(y) = 0.0056e^{-0.0056y}, \quad y \geq 0$$

Similarly, when we work with the bell-shaped normal distribution in later chapters we will write the model in random variable notation as

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

EXAMPLE 3.4.5

Suppose we would like a continuous random variable Y to "select" a number between 0 and 1 in such a way that intervals near the middle of the range would be more likely to be represented than intervals near either 0 or 1. One pdf having that property is the function $f_Y(y) = 6y(1 - y), 0 \leq y \leq 1$ (see Figure 3.4.9). Do we know for certain that the function pictured in Figure 3.4.9 is a "legitimate" pdf? Yes, because $f_Y(y) \geq 0$ for all y, and $\int_0^1 6y(1 - y)\,dy = 6[y^2/2 - y^3/3]|_0^1 = 1$.

Comment. To simplify the way pdfs are written, it will be assumed that $f_Y(y) = 0$ for all y outside the range actually specified in the funtion's definition. In Example 3.4.5,

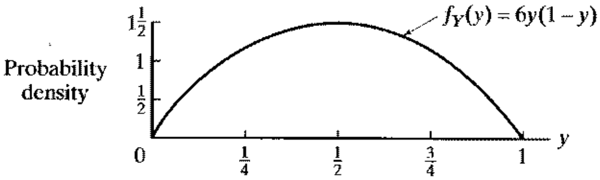


FIGURE 3.4.9

for instance, the statement $f_Y(y) = 6y(1 - y)$, $0 \leq y \leq 1$ is to be interpreted as an abbreviation for

$$f_Y(y) = \begin{cases} 0, & y < 0 \\ 6y(1 - y), & 0 \leq y \leq 1 \\ 0, & y > 1 \end{cases}$$

Continuous Cumulative Distribution Functions

Associated with every random variable, discrete or continuous, is a cumulative distribution function. For discrete random variables (recall Definition 3.3.4), the cdf is a nondecreasing step function, where the "jumps" occur at the values of t for which the pdf has positive probability. For continuous random variables, the cdf is a monotonically nondecreasing. continuous function. In both cases, the cdf can be helpful in calculating the probability that a random variable takes on a value in a given interval. As we will see in later chapters. there are also several important relationships that hold for continuous cdfs and pdfs. One such relationship is cited in Theorem 3.4.1.

Definition 3.4.3. The cdf for a continuous random variable Y is an indefinite integral of its pdf:

$$F_Y(y) = \int_{-\infty}^{y} f_Y(r)\, dr = P(\{s \in S \mid Y(s) \leq y\}) = P(Y \leq y)$$

Theorem 3.4.1. *Let $F_Y(y)$ be the cdf of a continuous random variable Y. Then*

$$\frac{d}{dy} F_Y(y) = f_Y(y)$$

Proof. The statement of Theorem 3.4.1 follows immediately from the Fundamental Theorem of Calculus. □

Theorem 3.4.2. *Let Y be a continuous random variable with cdf $F_Y(y)$. Then*

a. $P(Y > s) = 1 - F_Y(s)$

b. $P(r < Y \leq s) = F_Y(s) - F_Y(r)$

c. $\lim\limits_{y \to \infty} F_Y(y) = 1$

d. $\lim\limits_{y \to -\infty} F_Y(y) = 0$

Proof.

a. $P(Y > s) = 1 - P(Y \leq s)$ since $(Y > s)$ and $(Y \leq s)$ are complementary events. But $P(Y \leq s) = F_Y(s)$, and the conclusion follows.

b. Since the set $(r < Y \leq s) = (Y \leq s) - (Y \leq r)$, $P(r < Y \leq s) = P(Y \leq s) - P(Y \leq r) = F_Y(s) - F_Y(r)$.

c. Let $\{y_n\}$ be a set of values of Y, $n = 1, 2, 3, \ldots$, where $y_n < y_{n+1}$ for all n, and $\lim_{n\to\infty} y_n = \infty$. If $\lim_{n\to\infty} F_Y(y_n) = 1$ for every such sequence $\{y_n\}$, then $\lim_{y\to\infty} F_Y(y) = 1$. To that end, set $A_1 = (Y \le y_1)$, and let $A_n = (y_{n-1} < Y \le y_n)$ for $n = 2, 3, \ldots$ Then $F_Y(y_n) = P(\cup_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$, since the A_k are disjoint. Also, the sample space $S = \cup_{k=1}^{\infty} A_k$, and by Axiom 4, $1 = P(S) = P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$. Then putting these equalities together gives $1 = \sum_{k=0}^{\infty} P(A_k) = \lim_{n\to\infty} \sum_{k=0}^{n} P(A_k) = \lim_{n\to\infty} F_Y(y_n)$

d. $$\begin{aligned}\lim_{y\to-\infty} F_Y(y) &= \lim_{y\to-\infty} P(Y \le y) = \lim_{y\to-\infty} P(-Y \ge -y) = \lim_{y\to-\infty} [1 - P(-Y \le -y)] \\ &= 1 - \lim_{y\to-\infty} P(-Y \le -y) = 1 - \lim_{y\to\infty} P(-Y \le y) \\ &= 1 - \lim_{y\to\infty} F_{-Y}(y) = 0\end{aligned}$$ $\square$

Transformations

If X and Y are two discrete random variables and a and b are constants such that $Y = aX + b$, the pdf for Y can be expressed in terms of the pdf for X. Theorem 3.3.1 provided the details. Here we give the analogous result for a linear transformation involving two *continuous* random variables.

Theorem 3.4.3. *Suppose X is a continuous random variable. Let $Y = aX + b$, where $a \ne 0$ and b are constants. Then*

$$f_Y(y) = \frac{1}{|a|} f_x\left(\frac{y-b}{a}\right)$$

Proof. We begin by writing an expression for the cdf of Y:

$$F_Y(y) = P(Y \le y) = P(aX + b \le y) = P(aX \le y - b)$$

At this point we need to consider two cases, the distinction being the sign of a. Suppose, first, that $a > 0$. Then

$$F_Y(y) = P(aX \le y - b) = P\left(X \le \frac{y-b}{a}\right)$$

and differentiating $F_Y(y)$ yields $f_Y(y)$:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_x\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right)$$

If $a < 0$,

$$F_Y(y) = P(aX \le y - b) = P\left(X > \frac{y-b}{a}\right) = 1 - P\left(X \le \frac{y-b}{a}\right)$$

Differentiation in this case gives

$$f_Y(y) = \frac{d}{dy}F_Y(y) = \frac{d}{dy}\left[1 - F_X\left(\frac{y-b}{a}\right)\right] = -\frac{1}{a}f_X\left(\frac{y-b}{a}\right) = \frac{1}{|a|}f_X\left(\frac{y-b}{a}\right)$$

and the theorem is proved. ☐

QUESTIONS

3.4.1. Suppose $f_Y(y) = 4y^3, 0 \le y \le 1$. Find $P(0 \le Y \le \frac{1}{2})$.

3.4.2. For the random variable Y with pdf $f_Y(y) = \frac{2}{3} + \frac{2}{3}y, 0 \le y \le 1$, find $P(\frac{3}{4} \le Y \le 1)$

3.4.3. Let $f_Y(y) = \frac{3}{2}y^2, -1 \le y \le 1$. Find $P(|Y - \frac{1}{2}| < \frac{1}{4})$. Draw a graph of $f_Y(y)$ and show the area representing the desired probability.

3.4.4. For persons infected with a certain form of malaria, the length of time spent in remission is described by the continuous pdf $f_Y(y) = \frac{1}{9}y^2, 0 \le y \le 3$, where Y is measured in years. What is the probability that a malaria patient's remission lasts longer than one year?

3.4.5. The length of time, Y, that a customer spends in line at a bank teller's window before being served is described by the exponential pdf $f_Y(y) = 0.2e^{-0.2y}, y \ge 0$.

- **(a)** What is the probability that a customer will wait more than 10 minutes?
- **(b)** Suppose the customer will leave if the wait is more than 10 minutes. Assume that the customer goes to the bank twice next month. Let the random variable X be the number of times the customer leaves without being served. Calculate $p_X(1)$.

3.4.6. Let n be a positive integer. Show that $f_Y(y) = (n+2)(n+1)y^n(1-y), 0 \le y \le 1$, is a pdf.

3.4.7. Find the cdf for the random variable Y given in Question 3.4.1. Calculate $P(0 \le Y \le \frac{1}{2})$ using $F_Y(y)$.

3.4.8. If Y is an exponential random variable, $f_Y(y) = \lambda e^{-\lambda y}, y \ge 0$, find $F_Y(y)$.

3.4.9. If the pdf for Y is

$$f_Y(y) = \begin{cases} 0, & |y| > 1 \\ 1 - |y|, & |y| \le 1 \end{cases}$$

find and graph $F_Y(y)$.

3.4.10. A continuous random variable Y has a cdf given by

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ y^2 & 0 \le y < 1 \\ 1 & y \ge 1 \end{cases}$$

Find $P(\frac{1}{2} < Y \le \frac{3}{4})$ two ways—first, by using the cdf and second, by using the pdf.

3.4.11. A random variable Y has cdf

$$F_Y(y) = \begin{cases} 0 & y < 1 \\ \ln y & 1 \le y \le e \\ 1 & e < y \end{cases}$$

Find

(a) $P(Y < 2)$

(b) $P\left(2 < Y \le 2\frac{1}{2}\right)$

(c) $P\left(2 < Y < 2\frac{1}{2}\right)$

(d) $f_Y(y)$

3.4.12. The cdf for a random variable Y is defined by $F_Y(y) = 0$ for $y < 0$; $F_Y(y) = 4y^3 - 3y^4$ for $0 \le y \le 1$; and $F_Y(y) = 1$ for $y > 1$. Find $P\left(\frac{1}{4} \le Y \le \frac{3}{4}\right)$ by integrating $f_Y(y)$.

3.4.13. Suppose $F_Y(y) = \frac{1}{12}(y^2 + y^3), 0 \le y \le 2$. Find $f_Y(y)$.

3.4.14. In a certain country, the distribution of a family's disposable income, Y, is described by the pdf $f_Y(y) = ye^{-y}, y \ge 0$. Find the median of the income distribution—that is, find the value m such that $F_Y(m) = 0.5$.

3.4.15. Let Y be the random variable described in Question 3.4.3. Define $W = 3Y + 2$. Find $f_W(w)$. For which values of w is $f_W(w) \ne 0$?

3.4.16. Suppose that $f_Y(y)$ is a continuous and symmetric pdf, where *symmetry* is the property that $f_Y(y) = f_Y(-y)$ for all y. Show that $P(-a \le Y \le a) = 2F_Y(y) - 1$.

3.4.17. Let Y be a random variable denoting the age at which a piece of equipment fails. In reliability theory, the probability that an item fails at time y given that it has survived until time y is called the *hazard rate*, $h(y)$. In terms of the pdf and cdf,

$$h(y) = \frac{f_Y(y)}{1 - F_Y(y)}$$

Find $h(y)$ if Y has an exponential pdf (see Question 3.4.8).

EXPECTED VALUES

Probability density functions, as we have already seen, provide a global overview of a random variable's behavior. If X is discrete, $p_X(k)$ gives $P(X = k)$ for all k; if Y is continuous, and A is any interval, or countable union of intervals, $P(Y \in A) = \int_A f_Y(y)\,dy$. Detail that explicit, though, is not always necessary—or even helpful. There are times when a more prudent strategy is to focus the information contained in a pdf by summarizing certain of its features with single numbers.

The first such feature that we will examine is *central tendency*, a term referring to the "average" value of a random variable. Consider the pdf's $p_X(k)$ and $f_Y(y)$ pictured in Figure 3.5.1. Although we obviously cannot predict with certainty what values any future X's and Y's will take on, it seems clear that X values will tend to lie somewhere near, μ_X, and Y values, somewhere near μ_Y. In some sense, then, we can characterize $p_X(k)$ by μ_X, and $f_Y(y)$ by μ_Y.

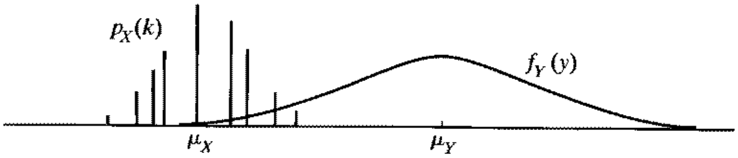


FIGURE 3.5.1

The most frequently used measure for describing central tendency—that is, for quantifying μ_X and μ_Y—is the *expected value.* Discussed at some length in this section and in Section 3.9, the expected value of a random variable is a slightly more abstract formulation of what we are already familiar with in simple discrete settings as the arithmetic average. Here, though, the values included in the average are "weighted" by the pdf.

Gambling affords a familiar illustration of the notion of an expected value. Consider the game of roulette. After bets are placed, the croupier spins the wheel and declares one of thirty-eight numbers, 00, 0, 1, 2, ..., 36, to be the winner. Disregarding what seems to be a perverse tendency of many roulette wheels to land on numbers for which no money has been wagered, we will assume that each of these thirty-eight numbers is equally likely (although only the eighteen numbers 1, 3, 5, ..., 35 are considered to be odd and only the eighteen numbers 2, 3, 4, ..., 36 are considered to be even). Suppose that our particular bet (at "even money") is $1 on odds. If the random variable X denotes our winnings, then X takes on the value *1* if an odd number occurs, and −1, otherwise. Therefore,

$$p_X(1) = P(X = 1) = \frac{18}{38} = \frac{9}{19}$$

and

$$p_X(-1) = P(X = -1) = \frac{20}{38} = \frac{10}{19}$$

Then $\frac{9}{19}$ of the time we will win one dollar and $\frac{10}{19}$ of the time we will lose one dollar. Intuitively, then, if we persist in this foolishness, we stand to *lose,* on the average, a little more than five cents each time we play the game:

$$\begin{aligned}\text{"expected" winnings} &= \$1 \cdot \frac{9}{19} + (-\$1) \cdot \frac{10}{19} \\ &= -\$0.053 \doteq -5¢\end{aligned}$$

The number −0.053 is called the *expected value of* X.

Physically, an expected value can be thought of as a center of gravity. Here, for example, imagine two bars of height $\frac{10}{19}$ and $\frac{9}{19}$ positioned along a weightless X-axis at the points −1 and +1, respectively (see Figure 3.5.2). If a fulcrum were placed at the point −0.053, the system would be in balance, implying that we can think of that point as marking off the center of the random variable's distribution.

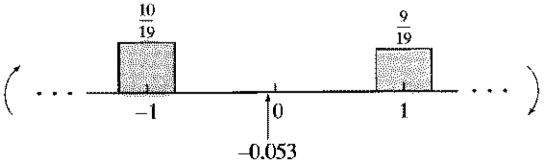


FIGURE 3.5.2

If X is a discrete random variable taking on each of its values with the same probability, the expected value of X is simply the everyday notion of an arithmetic average or mean:

$$\text{expected value of } X = \sum_{\text{all } k} k \cdot \frac{1}{n} = \frac{1}{n} \sum_{\text{all } k} k$$

Extending this idea to a discrete X described by an arbitrary pdf, $p_X(k)$, gives

$$\text{expected value of } X = \sum_{\text{all } k} k \cdot p_X(k) \tag{3.5.1}$$

For a continuous random variable, Y, the summation in Equation 3.5.1 is replaced by an integration and $k \cdot p_X(k)$ becomes $y \cdot f_Y(y)$.

Definition 3.5.1. Let X be a discrete random variable with probability function $p_X(k)$. The *expected value of* X is denoted $E(X)$ (or sometimes μ or μ_X) and is given by

$$E(X) = \mu = \mu_X = \sum_{\text{all } k} k \cdot p_X(k)$$

Similarly, if Y is a continuous random variable with pdf $f_Y(y)$,

$$E(Y) = \mu = \mu_Y = \int_{-\infty}^{\infty} y \cdot f_Y(y)\,dy$$

Comment. We assume that both the sum and the integral in Definition 3.5.1 converge absolutely:

$$\sum_{\text{all } k} |k| p_X(k) < \infty \qquad \int_{-\infty}^{\infty} |y| f_Y(y)\,dy < \infty$$

If not, we say that the random variable has no finite expected value. One immediate reason for requiring *absolute* convergence is that a convergent sum that is not absolutely convergent depends on the order in which the terms are added, and order should obviously not be a consideration when defining an average.

EXAMPLE 3.5.1

Suppose X is a binomial random variable with $p = \frac{5}{9}$ and $n = 3$. Then $p_X(k) = P(X = k) = \binom{3}{k}\left(\frac{5}{9}\right)^k\left(\frac{4}{9}\right)^{3-k}$, $k = 0, 1, 2, 3$. What is the expected value of X?

Applying Definition 3.5.1 gives

$$\begin{aligned} E(X) &= \sum_{k=0}^{3} k \cdot \binom{3}{k}\left(\frac{5}{9}\right)^k\left(\frac{4}{9}\right)^{3-k} \\ &= (0)\left(\frac{64}{729}\right) + (1)\left(\frac{240}{729}\right) + (2)\left(\frac{300}{729}\right) + (3)\left(\frac{125}{729}\right) = \frac{1215}{729} = \frac{5}{3} = 3\left(\frac{5}{9}\right) \end{aligned}$$

Comment. Notice that the expected value here reduces to five-thirds, which can be written as three times five-ninths, the latter two factors being n and p, respectively. As the next theorem proves, that relationship is not a coincidence.

Theorem 3.5.1. *Suppose X is a binomial random variable with parameters n and p. Then $E(X) = np$.*

Proof. According to Definition 3.5.1, $E(X)$ for a binomial random variable is the sum

$$\begin{aligned} E(X) = \sum_{k=0}^{n} k \cdot p_X(k) &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1 - p)^{n-k} \\ &= \sum_{k=0}^{n} \frac{k \cdot n!}{k!(n - k)!} p^k (1 - p)^{n-k} \\ &= \sum_{k=1}^{n} \frac{n!}{(k - 1)!(n - k)!} p^k (1 - p)^{n-k} \end{aligned} \tag{3.5.2}$$

At this point, a trick is called for. If $E(X) = \sum_{\text{all } k} g(k)$ can be factored in such a way that $E(X) = h \sum_{\text{all } k} p_{X^*}(k)$, where $p_{X^*}(k)$ is the pdf for some random variable X^*, then $E(X) = h$, since the sum of a pdf over its entire range is one. Here, suppose that np is factored out of Equation 3.5.2. Then

$$\begin{aligned} E(X) &= np \sum_{k=1}^{n} \frac{(n - 1)!}{(k - 1)!(n - k)!} p^{k-1} (1 - p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n - 1}{k - 1} p^{k-1} (1 - p)^{n-k} \end{aligned}$$

Now, let $j = k - 1$. It follows that

$$E(X) = np \sum_{j=0}^{n-1} \binom{n - 1}{j} p^j (1 - p)^{n-j-1}$$

Finally, letting $m = n - 1$ gives

$$E(X) = np \sum_{j=0}^{m} \binom{m}{j} p^j (1 - p)^{m-j}$$

and, since the value of the sum is 1 (why?),

$$E(X) = np \tag{3.5.3}$$

□

Comment. The statement of Theorem 3.5.1 should come as no surprise. If a multiple-choice test, for example, has one hundred questions, each with five possible answers, we would "expect" to get twenty correct, just by guessing. But if the random variable X denotes the number of correct answers (out of one hundred), $20 = E(X) = 100(\frac{1}{5}) = np$.

EXAMPLE 3.5.2

An urn contains nine chips, five red and four white. Three are drawn out at random without replacement. Let X denote the number of red chips in the sample. Find $E(X)$.

From Section 3.2, we recognize X to be a hypergeometric random variable, where

$$P(X = k) = p_X(k) = \frac{\binom{5}{k}\binom{4}{3-k}}{\binom{9}{3}}, \quad k = 0, 1, 2, 3$$

Therefore,

$$\begin{aligned}E(X) &= \sum_{k=0}^{3} k \cdot \frac{\binom{5}{k}\binom{4}{3-k}}{\binom{9}{3}} \\ &= (0)\left(\frac{4}{84}\right) + (1)\left(\frac{30}{84}\right) + (2)\left(\frac{40}{84}\right) + (3)\left(\frac{10}{84}\right) \\ &= \frac{5}{3}\end{aligned}$$

Comment. As was true in Example 3.5.1, the value found here for $E(X)$ suggests a general formula—in this case, for the expected value of a hypergeometric random variable.

Theorem 3.5.2. *Suppose X is a hypergeometric random variable with parameters r, w, and n. That is, suppose an urn contains r red balls and w white balls. A sample of size n is drawn simultaneously from the urn. Let X be the number of red balls in the sample. Then* $E(X) = \dfrac{rn}{r + w}$.

Proof. See Question 3.5.25. □

Comment. Let p represent the proportion of red balls in an urn—that is, $p = \dfrac{r}{r + w}$. The formula, then for the expected value of a hypergeometric random variable has the same structure as the formula for the expected value of a binomial random variable:

$$E(X) = \frac{rn}{r + w} = n\frac{r}{r + w} = np$$

EXAMPLE 3.5.3

Among the more common versions of the "numbers" racket is a game called D.J., its name deriving from the fact that the winning ticket is determined from Dow Jones averages. Three sets of stocks are used: Industrials, Transportations, and Utilities. Traditionally, the three are quoted at two different times, 11 A.M. and noon. The last digits of the earlier quotation are arranged to form a three-digit number; the noon quotation generates a second three-digit number, formed the same way. Those two numbers are then added together and the last three digits of that sum become the winning pick. Figure 3.5.3 shows a set of quotations for which *906* would be declared the winner.

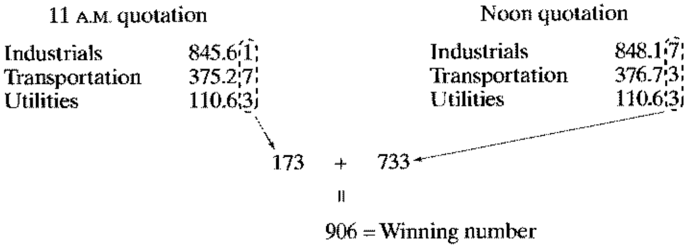


FIGURE 3.5.3

The payoff in D.J. is 700 to 1. Suppose that we bet \$5. How much do we stand to win, or lose, *on the average?*

Let p denote the probability of our number being the winner and let X denote our earnings. Then

$$X = \begin{cases} \$3500 & \text{with probability } p \\ -\$5 & \text{with probability } 1 - p \end{cases}$$

and

$$E(X) = \$3500 \cdot p - \$5 \cdot (1 - p)$$

Our intuition would suggest (and this time it would be correct!) that each of the possible winning numbers, 000 through 999, is equally likely. That being the case, $p = 1/1000$ and

$$E(X) = \$3500 \cdot \left(\frac{1}{1000}\right) - \$5 \cdot \left(\frac{999}{1000}\right) = -\$1.50$$

On the average, then, we lose \$1.50 on a \$5.00 bet.

EXAMPLE 3.5.4

Suppose that fifty people are to be given a blood test to see who has a certain disease. The obvious laboratory procedure is to examine each person's blood individually, meaning that fifty tests would eventually be run. An alternative strategy is to divide each person's blood sample into two parts—say, A and B. All of the A's would then be mixed together and treated as one sample. If that "pooled" sample proved to be negative for the disease, all fifty individuals must necessarily be free of the infection, and no further testing would need to be done. If the pooled sample gave a positive reading, of course, all fifty B samples would have to be analyzed separately. Under what conditions would it make sense for a laboratory to consider pooling the fifty samples?

In principle, the pooling strategy is preferable (i.e., more economical) if it can substantially reduce the number of tests that need to be performed. Whether or not it can depends ultimately on the probability p that a person is infected with the disease.

Let the random variable X denote the number of tests that will have to be performed if the samples are pooled. Clearly,

$$X = \begin{cases} 1 & \text{if none of the fifty is infected} \\ 51 & \text{if at least one of the fifty is infected} \end{cases}$$

But

$$\begin{aligned} P(X = 1) = p_X(1) &= P(\text{none of the fifty is infected}) \\ &= (1 - p)^{50} \end{aligned}$$

(assuming independence), and

$$P(X = 51) = p_X(51) = 1 - P(X = 1) = 1 - (1 - p)^{50}$$

Therefore,

$$E(X) = 1 \cdot (1 - p)^{50} + 51 \cdot [1 - (1 - p)^{50}]$$

Table 3.5.1 shows $E(X)$ as a function of p. As our intuition would suggest, the pooling strategy becomes increasingly feasible as the prevalence of the disease diminishes. If the chance of a person being infected is 1 in 1000, for example, the pooling strategy requires an average of only 3.4 tests, a dramatic improvement over the 50 tests that would be needed if the samples were tested one by one. On the other hand, if 1 in 10 individuals is infected, pooling would be clearly inappropriate, requiring *more* than 50 tests [$E(X) = 50.7$].

TABLE 3.5.1

p	$E(X)$
0.5	51.0
0.1	50.7
0.01	20.8
0.001	3.4
0.0001	1.2

EXAMPLE 3.5.5

Consider the following game. A fair coin is flipped until the first tail appears; we win \$2 if it appears on the first toss, \$4 if it appears on the second toss, and, in general, $\$2^k$ if it first occurs on the kth toss. Let the random variable X denote our winnings. How much should we have to pay in order for this to be a fair game? [*Note:* A fair game is one where the difference between the ante and $E(X)$ is 0.]

Known as the St. Petersburg paradox, this problem has a rather unusual answer. First, note that

$$p_X(2^k) = P(X = 2^k) = \frac{1}{2^k}, \quad k = 1, 2, \ldots$$

Therefore,

$$E(X) = \sum_{\text{all } k} 2^k p_X(2^k) = \sum_{k=1}^{\infty} 2^k \cdot \frac{1}{2^k} = 1 + 1 + 1 + \cdots$$

which is a divergent sum. That is, X does not have a finite expected value, so in order for this game to be fair, our ante would have to be an infinite amount of money!

Comment. Mathematicians have been trying to "explain" the St. Petersburg paradox for almost two hundred years (56). The answer seems clearly absurd—no gambler would consider paying even \$25 to play such a game, much less an infinite amount—yet the computations involved in showing that X has no finite expected value are unassailably correct. Where the difficulty lies, according to one common theory, is with our inability to put in perspective the very small probabilities of winning very large payoffs. Furthermore, the problem assumes that our opponent has infinite capital, which is an impossible state of affairs. We get a much more reasonable answer for $E(X)$ if the stipulation is added that our winnings can be at most, say, \$1000 (see Question 3.5.19) or if the payoffs are assigned according to some formula other than 2^k (see Question 3.5.20).

Comment. There are two important lessons to be learned from the St. Petersburg paradox. First is the realization that $E(X)$ is not necessarily a meaningful characterization of the "location" of a distribution. Question 3.5.24 shows another situation where the formal computation of $E(X)$ gives a similarly inappropriate answer. Second, we need to be aware that the notion of expected value is not necessarily synonymous with the concept of *worth*. Just because a game, for example, has a positive expected value—even a very *large* positive expected value—does not imply that someone would want to play it. Suppose, for example, that you had the opportunity to spend your last \$10,000 on a sweepstakes ticket where the prize was a billion dollars but the probability of winning was only 1 in 10,000. The expected value of such a bet would be over \$90,000,

$$\begin{aligned} E(X) &= \$1{,}000{,}000{,}000\left(\frac{1}{10{,}000}\right) + (-\$10{,}000)\left(\frac{9{,}999}{10{,}000}\right) \\ &= \$90{,}001 \end{aligned}$$

but it is doubtful that many people would rush out to buy a ticket. (Economists have long recognized the distinction between a payoff's numerical value and its perceived desirability. They refer to the latter as *utility*.)

EXAMPLE 3.5.6

The distance, Y, that a molecule in a gas travels before colliding with another molecule can be modeled by the exponential pdf

$$f_Y(y) = \frac{1}{\mu}e^{-y/\mu}, \quad y \geq 0$$

where μ is a positive constant known as the *mean free path*. Find $E(Y)$.

Since the random variable here is continuous, its expected value is an integral:

$$E(Y) = \int_0^\infty y\frac{1}{\mu}e^{-y/\mu}\,dy$$

Let $w = y/\mu$, so that $dw = 1/\mu\, dy$. Then $E(Y) = \mu\int_0^\infty we^{-w}dw$. Setting $u = w$ and $dv = e^{-w}dw$ and integrating by parts gives

$$E(Y) = \mu[-we^{-w} - e^{-w}]\Big|_0^\infty = \mu \tag{3.5.4}$$

Equation 3.5.4 shows that μ is aptly named—it does, in fact, represent the average distance a molecule travels, free of any collisions. Nitrogen (N_2), for example, at room temperature and standard atmospheric pressure has $\mu = 0.00005$ cm. An N_2 molecule, then, travels that far before colliding with another N_2 molecule, *on the average*.

EXAMPLE 3.5.7

One continuous pdf that has a number of interesting applications in physics is the *Rayleigh distribution*, where the pdf is given by

$$f_Y(y) = \frac{y}{a^2}e^{-y^2/2a^2}, \quad a > 0; \quad 0 \leq y < \infty \tag{3.5.5}$$

Calculate the expected value for a random variable having a Rayleigh distribution.

From Definition 3.5.1,

$$E(Y) = \int_0^\infty y \cdot \frac{y}{a^2}e^{-y^2/2a^2}dy$$

Let $v = y/(\sqrt{2}a)$. Then

$$E(Y) = 2\sqrt{2}a\int_0^\infty v^2e^{-v^2}dv$$

The integrand here is a special case of the general form $v^{2k}e^{-v^2}$. For $k = 1$,

$$\int_0^\infty v^{2k}e^{-v^2}dv = \int_0^\infty v^2e^{-v^2}dv = \frac{1}{4}\sqrt{\pi}$$

Therefore,

$$E(Y) = 2\sqrt{2}a \cdot \frac{1}{4}\sqrt{\pi}$$
$$= a\sqrt{\pi/2}$$

Comment. The pdf here is named for John William Strutt, Baron Rayleigh, the nineteenth- and twentieth-century British physicist who showed that Equation 3.5.5 is the solution to a problem arising in the study of wave motion. If two waves are superimposed, it is well known that the height of the resultant at any time t is simply the algebraic sum of the corresponding heights of the waves being added (see Figure 3.5.4). Seeking to extend that notion, Rayleigh posed the following question: If n waves, each having the same amplitude h and the same wavelength, are superimposed randomly with respect to phase, what can we say about the amplitude R of the resultant? Clearly, R is a random variable, its value depending on the particular collection of phase angles represented by the sample. What Rayleigh was able to show in his 1880 paper (173) is that when n is large, the probabilistic behavior of R is described by the pdf

$$f_R(r) = \frac{2r}{nh^2} \cdot e^{-r^2/nh^2}, \quad r > 0$$

which is just a special case of Equation 3.5.5 with $a = \sqrt{2/nh^2}$.

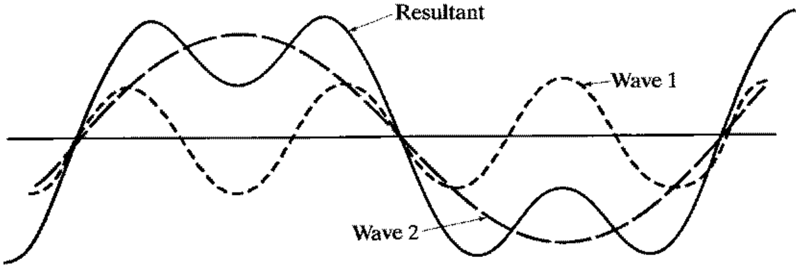


FIGURE 3.5.4

A Second Measure of Central Tendency: The Median

While the expected value is the most frequently used measure of a random variable's central tendency, it does have a weakness that sometimes makes it misleading and inappropriate. Specifically, if one or several possible values of a random variable are either much smaller or much larger than all the others, the value of μ can be distorted in the sense that it no longer reflects the center of the distribution in any meaningful way. For example, suppose a small community consists of a homogeneous group of middle-range salary earners, and then Bill Gates moves to town. Obviously, the town's average salary before and after the multibillionaire arrives will be quite different, even though he represents only one new value of the "salary" random variable.

It would be helpful to have a measure of central tendency that was not so sensitive to "outliers" or to probability distributions that are markedly skewed. One such measure is the *median,* which, in effect, divides the area under a pdf into two equal areas.

Definition 3.5.2. If X is a discrete random variable, the median, m, is that point for which $P(X < m) = P(X > m)$. In the event that $P(X \leq m) = 0.5$ and $P(X \geq m') = 0.5$, the median is defined to be the arithmetic average, $(m + m')/2$.

If Y is a continuous random variable, its median is the solution to the integral equation, $\int_{-\infty}^{m} f_Y(y)\,dy = 0.5$.

EXAMPLE 3.5.8

If a random variable's pdf is symmetric, both μ and m will be equal. Should $p_X(k)$ or $f_Y(y)$ not be symmetric, though, the difference between the expected value and the median can be considerable, especially if the asymmetry takes the form of extreme skewness. The situation described here is a case in point.

Soft-glow makes a 60-watt light bulb that is advertised to have an average life of one thousand hours. Assuming that that performance claim is valid, is it reasonable for consumers to conclude that the Soft-glow bulbs they buy will last for approximately one-thousand hours?

No! If the average life of a bulb is one thousand hours, the (continuous) pdf, $f_Y(y)$, modeling the length of time, Y, that it remains lit before burning out is likely to have the form

$$f_Y(y) = 0.001e^{-0.001y}, \quad y > 0 \qquad (3.5.6)$$

(for reasons explained in Chapter 4). But Equation 3.5.6 is a very skewed pdf, having a shape much like the curve drawn in Figure 3.4.8. The median for such a distribution will lie considerably to the left of the mean.

More specifically, the median lifetime for these bulbs—according to Definition 3.5.2—is the value m for which

$$\int_0^m 0.001e^{-0.001y}dy = 0.5$$

But $\int_0^m 0.001e^{-0.001y}dy = 1 - e^{-0.001m}$. Setting the latter equal to 0.5 implies that

$$m = (1/-0.001)\ln(0.5) = 693$$

So, even though the *average* life of a bulb is 1000 hours, there is a 50% chance that the one you buy will last less than 693 hours.

QUESTIONS

3.5.1. Recall the game of Keno described in Example 3.2.5. The following are all the payoffs on a \$1 wager where the player has bet on 10 numbers. Calculate $E(X)$, where the random variable X denotes the amount of money won.

Number of Correct Guesses	Payoff	Probability
< 5	$-\$1$	.935
5	2	.0514
6	18	.0115
7	180	.0016
8	1,300	1.35×10^{-4}
9	2,600	6.12×10^{-6}
10	10,000	1.12×10^{-7}

3.5.2. Cracker Jack first appeared in 1893 at the Chicago World's Fair. Enormously popular ever since (250 million boxes are sold each year), the snack owes more than a little of its success, especially with children, to the toy included in each box. When a new Nutty Deluxe flavor was introduced in the mid-1990s, that familiar marketing gimmick was raised to a new level. Placed in one box was a certificate redeemable for a \$10,000 ring; in 50 other boxes were certificates for a *Breakfast at Tiffany's* video (a movie in which the leading character, Holly Golightly, finds her engagement ring in a Cracker Jack box); the usual toys and puzzles were put in all the other boxes (183). Calculate the expected value of the prize in a box of Nutty Deluxe Cracker Jack. Assume that 5 million boxes were distributed during that first year. Also, assume that each video was worth \$30 and each other prize 1.2¢.

3.5.3. The pdf describing the daily profit, X, earned by Acme Industries was derived in Example 3.3.7. Find the company's *average* daily profit.

3.5.4. In the game of redball, two drawings are made without replacement from a bowl that has four white ping-pong balls and two red ping-pong balls. The amount won is determined by how many of the red balls are selected. For a \$5 bet, a player can opt to be paid under either Rule A or Rule B, as shown. If you were playing the game, which would you choose? Why?

A		B	
No. of Red Balls Drawn	Payoff	No. of Red Balls Drawn	Payoff
0	0	0	0
1	\$2	1	\$1
2	\$10	2	\$20

3.5.5. Recall the telemarketing campaign launched by the Wipe Your Feet carpet cleaning company described in Example 3.2.8. On the average, how many new customers would that effort identify? How many calls would they have to make in order to find an average of 100 new customers?

3.5.6. A manufacturer has 100 memory chips in stock, 4% of which are likely to be defective (based on past experience). A random sample of 20 chips is selected and shipped to a factory that assembles laptops. Let X denote the number of computers that receive faulty memory chips. Find $E(X)$.

3.5.7. Records show that 642 new students have just entered a certain Florida school district. Of those 642, a total of 125 are not adequately vaccinated. The district's physician has scheduled a day for students to receive whatever shots they might need. On any given day, though, 12% of the district's students are likely to be absent. How many new students, then, can be expected to remain inadequately vaccinated?

3.5.8. Calculate $E(Y)$ for the following pdf's:

(a) $f_Y(y) = 3(1 - y)^2, 0 \leq y \leq 1$

(b) $f_Y(y) = 4ye^{-2y}, y \geq 0$

(c) $f_Y(y) = \begin{cases} \frac{3}{4}, & 0 \leq y \leq 1 \\ \frac{1}{4}, & 2 \leq y \leq 3 \\ 0, & \text{elsewhere} \end{cases}$

(d) $f_Y(y) = \sin y, \quad 0 \leq y \leq \frac{\pi}{2}$

3.5.9. Recall Question 3.4.4, where the length of time Y (in years) that a malaria patient spends in remission has pdf $f_Y(y) = \frac{1}{9}y^2, 0 \leq y \leq 3$. What is the average length of time that such a patient spends in remission?

3.5.10. Let the random variable Y have the uniform distribution over $[a, b]$; that is $f_Y(y) = \frac{1}{b-a}$ for $a \leq y \leq b$. Find $E(Y)$ using Definition 3.5.1. Also, deduce the value of $E(Y)$, knowing that the expected value is the center of gravity of $f_Y(y)$.

3.5.11. Show that the expected value associated with the exponential distribution, $f_Y(y) = \lambda e^{-\lambda y}, y > 0$, is $1/\lambda$, where λ is a positive constant.

3.5.12. Show that

$$f_Y(y) = \frac{1}{y^2}, \quad y \geq 1$$

is a valid pdf but that Y does not have a finite expected value.

3.5.13. Based on recent experience, 10-year-old passenger cars going through a motor vehicle inspection station have an 80% chance of passing the emissions test. Suppose that 200 such cars will be checked out next week. Write two formulas that show the number of cars that are expected to pass.

3.5.14. Suppose that 15 observations are chosen at random from the pdf $f_Y(y) = 3y^2, 0 \leq y \leq 1$. Let X denote the number that lie in the interval $\left(\frac{1}{2}, 1\right)$. Find $E(X)$.

3.5.15. A city has 74,806 registered automobiles. Each is required to display a bumper decal showing that the owner paid an annual wheel tax of $50. By law, new decals need to be purchased during the month of the owner's birthday. How much wheel tax revenue can the city expect to receive in November?

3.5.16. Regulators have found that 23 of the 68 investment companies that filed for bankruptcy in the past five years failed because of fraud, not for reasons related to the economy. Suppose that nine additional firms will be added to the bankruptcy rolls during the next quarter. How many of those failures are likely to be attributed to fraud?

3.5.17. An urn contains four chips numbered 1 through 4. Two are drawn without replacement. Let the random variable X denote the larger of the two. Find $E(X)$.

3.5.18. A fair coin is tossed three times. Let the random variable X denote the total number of heads that appear times the number of heads that appear on the first and third tosses. Find $E(X)$.

3.5.19. How much would you have to ante to make the St. Petersburg game "fair" (recall Example 3.5.5) if the most you could win was \$1000? That is, the payoffs are $\$2^k$ for $1 \le k \le 9$, and \$1000 for $k \ge 10$.

3.5.20. For the St. Petersburg problem (Example 3.5.5) find the expected payoff if

- **(a)** the amounts won are c^k instead of 2^k, where $0 < c < 2$.
- **(b)** the amounts won are $\log 2^k$. [This was a modification suggested by D. Bernoulli (a nephew of James Bernoulli) to take into account the decreasing marginal utility of money—the more you have, the less useful a bit more is.]

3.5.21. A fair die is rolled 3 times. Let X denote the number of different faces showing, $X = 1, 2, 3$. Find $E(X)$.

3.5.22. Two distinct integers are chosen at random from the first five positive integers. Compute the expected value of the absolute value of the difference of the two numbers.

3.5.23. Suppose that two evenly matched teams are playing in the World Series. On the average, how many games will be played? (The winner is the first team to get four victories.) Assume that each game is an independent event.

3.5.24. An urn contains one white chip and one black chip. A chip is drawn at random. If it is white, the "game" is over; if it is black, that chip and another black one are put into the urn. Then another chip is drawn at random from the "new" urn and the same rules for ending or continuing the game are followed (if the chip is white, the game is over; if the chip is black, it is replaced in the urn, together with another chip of the same color). The drawings continue until a white chip is selected. Show that the expected number of drawings necessary to get a white chip is not finite.

3.5.25. A random sample of size n is drawn without replacement from an urn containing r red chips and w white chips. Define the random variable X to be the number of red chips in the sample. Use the summation technique described in Theorem 3.5.1 to prove that $E(X) = rn/(r + w)$.

3.5.26. Given that X is a nonnegative, integer-valued random variable, show that

$$E(X) = \sum_{k=1}^{\infty} P(X \ge k)$$

The Expected Value of a Function of a Random Variable

There are many situations that call for finding the expected value of a *function* of a random variable—say, $Y = g(X)$. One common example would be change of scale problems, where $g(X) = aX + b$ for constants a and b. Sometimes the pdf of the new random variable Y can be easily determined, in which case $E(Y)$ can be calculated by simply applying Definition 3.5.1. Often, though, $f_Y(y)$ can be difficult to derive, depending on the complexity of $g(X)$. Fortunately, Theorem 3.5.3 allows us to calculate the expected value of Y without knowing the pdf for Y.

Theorem 3.5.3. *Suppose X is a discrete random variable with pdf $p_X(k)$. Let $g(X)$ be a function of X. Then the expected value of the random variable $g(X)$ is given by*

$$E[g(X)] = \sum_{\text{all } k} g(k) \cdot p_X(k)$$

provided that $\sum_{\text{all } k} |g(k)| p_X(k) < \infty$.

If Y is a continuous random variable with pdf $f_Y(y)$, and if $g(Y)$ is a continuous function, then the expected value of the random variable $g(Y)$ is

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y) \cdot f_Y(y)\, dy$$

provided that $\int_{-\infty}^{\infty} |g(y)| f_Y(y)\, dy < \infty$.

Proof. We will prove the result for the discrete case. See (150) for details showing how the argument is modified when the pdf is continuous. Let $W = g(X)$. The set of all possible k-values, $k_1, k_2, \ldots$, will give rise to a set of w-values, $w_1, w_2, \ldots$, where, in general, more than one k may be associated with a given w. Let S_j be the set of k's for which $g(k) = w_j$ [so $\cup_j S_j$ is the entire set of k-values for which $p_X(k)$ is defined]. We obviously have that $P(W = w_j) = P(X \in S_j)$, and we can write

$$\begin{aligned}
E(W) = \sum_j w_j \cdot P(W = w_j) &= \sum_j w_j \cdot P(X \in S_j) \\
&= \sum_j w_j \sum_{k \in S_j} p_X(k) \\
&= \sum_j \sum_{k \in S_j} w_j \cdot p_X(k) \\
&= \sum_j \sum_{k \in S_j} g(k) p_X(k) \quad \text{(why?)} \\
&= \sum_{\text{all } k} g(k) p_X(k)
\end{aligned}$$

Since it is being assumed that $\sum_{\text{all } k} |g(k)| p_X(k) < \infty$, the statement of the theorem holds. □

Corollary. *For any random variable W, $E(aW + b) = aE(W) + b$, where a and b are constants.*

Proof. Suppose W is continuous; the proof for the discrete case is similar. By Theorem 3.5.3, $E(aW + b) = \int_{-\infty}^{\infty} (aw + b) f_W(w)\, dw$, but the latter can be written $a \int_{-\infty}^{\infty} w \cdot f_W(w)\, dw + b \int_{-\infty}^{\infty} f_W(w)\, dw = aE(W) + b \cdot 1 = aE(W) + b$. □

EXAMPLE 3.5.9

Suppose that X is a random variable whose pdf is nonzero only for the three values −2, 1, and +2:

k	$p_X(k)$
−2	$\frac{5}{8}$
1	$\frac{1}{8}$
2	$\frac{2}{8}$
	1

Let $W = g(X) = X^2$. Verify the statement of Theorem 3.5.3 by computing $E(W)$ two ways—first, by finding $p_W(w)$ and summing $w \cdot p_W(w)$ over w and, second, by summing $g(k) \cdot p_X(k)$ over k.

By inspection, the pdf for W is defined for only two values, 1 and 4:

$w\ (= k^2)$	$p_W(w)$
1	$\frac{1}{8}$
4	$\frac{7}{8}$
	1

Taking the first approach to find $E(W)$ gives

$$E(W) = \sum_w w \cdot p_W(w) = 1 \cdot \left(\frac{1}{8}\right) + 4 \cdot \left(\frac{7}{8}\right)$$
$$= \frac{29}{8}$$

To find the expected value via Theorem 3.5.3, we take

$$E[g(X)] = \sum_k k^2 \cdot p_X(k) = (-2)^2 \cdot \frac{5}{8} + (1)^2 \cdot \frac{1}{8} + (2)^2 \cdot \frac{2}{8}$$

with the sum here reducing to the answer we already found, $\frac{29}{8}$.

For this particular situation, neither approach was easier than the other. In general, that will not be the case. Finding $p_W(w)$ is often quite difficult, and on those occasions Theorem 3.5.3 can be of great benefit.

EXAMPLE 3.5.10

Suppose the amount of propellant, Y, put into a can of spray paint is a random variable with pdf

$$f_Y(y) = 3y^2, \quad 0 < y < 1$$

Experience has shown that the largest surface area that can be painted by a can having Y amount of propellant is twenty times the area of a circle generated by a radius of Y ft. If the Purple Dominoes, a newly formed urban gang, have just stolen their first can of spray paint, can they expect to have enough to cover a 5′ × 8′ subway panel with grafitti?

No. By assumption, the maximum area (in ft^2) that can be covered by a can of paint is described by the function

$$g(Y) = 20\pi Y^2$$

According to the second statement in Theorem 3.5.3, though, the average value for $g(Y)$ is slightly less than the desired 40 ft^2:

$$\begin{aligned} E[g(Y)] &= \int_0^1 20\pi y^2 \cdot 3y^2\,dy \\ &= \left.\frac{60\pi y^5}{5}\right|_0^1 \\ &= 12\pi \\ &= 37.7\ \text{ft}^2 \end{aligned}$$

EXAMPLE 3.5.11

A fair coin is tossed until a head appears. You will be given $\left(\frac{1}{2}\right)^k$ dollars if that first head occurs on the kth toss. How much money can you expect to be paid?

Let the random variable X denote the toss at which the first head appears. Then

$$\begin{aligned} p_X(k) = P(X = k) &= P(\text{1st } k\text{-1 tosses are tails and } k\text{th toss is a head}) \\ &= \left(\frac{1}{2}\right)^{k-1} \cdot \frac{1}{2} \\ &= \left(\frac{1}{2}\right)^k, \quad k = 1, 2, \ldots \end{aligned}$$

Moreover,

$$\begin{aligned} E(\text{amount won}) = E\left[\left(\frac{1}{2}\right)^X\right] = E[g(X)] &= \sum_{\text{all } k} g(k) \cdot p_X(k) \\ &= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^k \cdot \left(\frac{1}{2}\right)^k \\ &= \sum_{k=1}^{\infty} \left(\frac{1}{2}\right)^{2k} = \sum_{k=1}^{\infty} \left(\frac{1}{4}\right)^k \\ &= \sum_{k=0}^{\infty} \left(\frac{1}{4}\right)^k - \left(\frac{1}{4}\right)^0 \\ &= \frac{1}{1 - \frac{1}{4}} - 1 \\ &= \$0.33 \end{aligned}$$

EXAMPLE 3.5.12

In one of the early applications of probability to physics, James Clerk Maxwell (1831–1879) showed that the speed S of a molecule in a perfect gas has a density function given by

$$f_S(s) = 4\sqrt{\frac{a^3}{\pi}}s^2e^{-as^2}, \quad s > 0$$

where a is a constant depending on the temperature of the gas and the mass of the particle. What is the average *energy* of a molecule in a perfect gas?

Let m denote the molecule's mass. Recall from physics that energy (W), mass (m), and speed (S) are related through the equation

$$W = \frac{1}{2}mS^2 = g(S)$$

To find $E(W)$ we appeal to the second part of Theorem 3.5.3:

$$\begin{aligned} E(W) &= \int_0^\infty g(s)f_S(s)\,ds \\ &= \int_0^\infty \frac{1}{2}ms^2 \cdot 4\sqrt{\frac{a^3}{\pi}}s^2e^{-as^2}ds \\ &= 2m\sqrt{\frac{a^3}{\pi}}\int_0^\infty s^4e^{-as^2}ds \end{aligned}$$

Make the substitution $t = as^2$. Then

$$E(W) = \frac{m}{a\sqrt{\pi}}\int_0^\infty t^{3/2}e^{-t}dt$$

But

$$\int_0^\infty t^{3/2}e^{-t}dt = \left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi}, \quad \text{(see Section 4.6)}$$

so

$$\begin{aligned} E(\text{energy}) = E(W) &= \frac{m}{a\sqrt{\pi}}\left(\frac{3}{2}\right)\left(\frac{1}{2}\right)\sqrt{\pi} \\ &= \frac{3m}{4a} \end{aligned}$$

EXAMPLE 3.5.13

Consolidated Industries is planning to market a new product and they are trying to decide how many to manufacture. They estimate that each item sold will return a profit of m

dollars; each one not sold represents an n dollar loss. Furthermore, they suspect the demand for the product, V, will have an exponential distribution,

$$f_V(v) = \left(\frac{1}{\lambda}\right)e^{-v/\lambda}, \quad v > 0$$

How many items should the company produce if they want to maximize their expected profit? (Assume that n, m, and λ are known.)

If a total of x items are made, the company's profit can be expressed as a function $Q(v)$, where

$$Q(v) = \begin{cases} mv - n(x - v) & \text{if } v < x \\ mx & \text{if } v \geq x \end{cases}$$

and v is the number of items sold. It follows that their *expected* profit is

$$\begin{aligned} E[Q(V)] &= \int_0^\infty Q(v) \cdot f_V(v)\,dv \\ &= \int_0^x [(m + n)v - nx]\left(\frac{1}{\lambda}\right)e^{-v/\lambda}\,dv + \int_x^\infty mx \cdot \left(\frac{1}{\lambda}\right)e^{-v/\lambda}\,dv \qquad (3.5.7) \end{aligned}$$

The integration here is straightforward, though a bit tedious. Equation 3.5.7 eventually simplifies to

$$E[Q(V)] = \lambda \cdot (m + n) - \lambda \cdot (m + n)e^{-x/\lambda} - nx$$

To find the optimal production level, we need to solve $dE[Q(V)]/dx = 0$ for x. But

$$\frac{dE[Q(V)]}{dx} = (m + n)e^{-x/\lambda} - n$$

and the latter equals zero when

$$x = -\lambda \cdot \ln\left(\frac{n}{m + n}\right)$$

EXAMPLE 3.5.14

A point, y, is selected at random from the interval [0, 1], dividing the line into two segments (see Figure 3.5.5). What is the expected value of the ratio of the shorter segment to the longer segment?

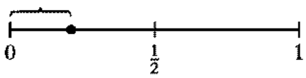


FIGURE 3.5.5

Notice, first, that the function

$$g(Y) = \frac{\text{shorter segment}}{\text{longer segment}}$$

has two expressions, depending on the location of the chosen point:

$$g(Y) = \begin{cases} y/(1-y), & 0 \le y \le \frac{1}{2} \\ (1-y)/y, & \frac{1}{2} < y \le 1 \end{cases}$$

By assumption, $f_Y(y) = 1, 0 \le y \le 1$, so

$$E[g(Y)] = \int_0^{\frac{1}{2}} \frac{y}{1-y} \cdot 1\, dy + \int_{\frac{1}{2}}^1 \frac{1-y}{y} \cdot 1\, dy$$

Writing the second integrand as $(1/y - 1)$ gives

$$\int_{\frac{1}{2}}^1 \frac{1-y}{y} \cdot 1\, dy = \int_{\frac{1}{2}}^1 \left(\frac{1}{y} - 1\right) dy = (\ln y - y)\Big|_{\frac{1}{2}}^1$$
$$= \ln 2 - \frac{1}{2}$$

By symmetry, though, the two integrals are the same, so

$$E\left[\frac{\text{shorter segment}}{\text{longer segment}}\right] = 2\ln 2 - 1$$
$$= 0.39$$

On the average, then, the longer segment will be a little more than $2\frac{1}{2}$ times the length of the shorter segment.

QUESTIONS

3.5.27. Suppose X is a binomial random variable with $n = 10$ and $p = \frac{2}{5}$. What is the expected value of $3X - 4$?

3.5.28. Recall Question 3.2.4. Suppose that each defective component discovered at the work station costs the company \$100. What is the average daily cost to the company for defective components?

3.5.29. Let X have the probability density function

$$f_X(x) = \begin{cases} 2(1-x), & 0 < x < 1 \\ 0, & \text{elsewhere} \end{cases}$$

Suppose that $Y = g(X) = X^3$. Find $E(Y)$ two different ways.

3.5.30. A tool and die company makes castings for steel stress-monitoring gauges. Their annual profit, Q, in hundreds of thousands of dollars, can be expressed as a function of product demand, y:

$$Q(y) = 2(1 - e^{-2y})$$

Suppose that the demand (in thousands) for their castings follows an exponential pdf, $f_Y(y) = 6e^{-6y}$, $y > 0$. Find the company's expected profit.

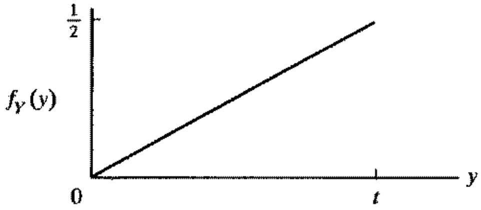
3.5.31. A box is to be constructed so that its height is 5 inches and its base is Y inches by Y inches, where Y is a random variable described by the pdf, $f_Y(y) = 6y(1 - y), 0 < y < 1$. Find the expected volume of the box.

3.5.32. Grades on the last Economics 301 exam were not very good. Graphed, their distribution had a shape similar to the pdf

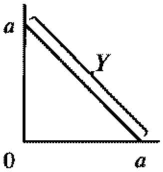
$$f_Y(y) = \frac{1}{5000}(100 - y), \quad 0 \le y \le 100$$

As a way of "curving" the results, the professor announces that he will replace each person's grade, Y, with a new grade, $g(Y)$, where $g(Y) = 10\sqrt{Y}$. Has the professor's strategy been successful in raising the class average above 60?

3.5.33. Find $E(Y^2)$ if the random variable Y has the pdf pictured below:



3.5.34. The hypotenuse, Y, of the isosceles right triangle shown is a random variable having a uniform pdf over the interval [6,10]. Calculate the expected value of the triangle's area. Do not leave the answer as a function of a.



3.5.35. An urn contains n chips numbered 1 through n. Assume that the probability of choosing chip i is equal to $ki, i = 1, 2, \ldots, n$. If one chip is drawn, calculate $E\left(\frac{1}{X}\right)$, where the random variable X denotes the number showing on the chip selected. *Hint:* Recall that the sum of the first n integers is $n(n + 1)/2$.

THE VARIANCE

We saw in Section 3.5 that the location of a distribution is an important characteristic and that it can be effectively measured by calculating either the mean or the median. A second feature of a distribution that warrants further scrutiny is its *dispersion*—that is,

TABLE 3.6.1

k	$p_{X_1}(k)$	k	$p_{X_2}(k)$
-1	$\frac{1}{2}$	$-1{,}000{,}000$	$\frac{1}{2}$
1	$\frac{1}{2}$	$1{,}000{,}000$	$\frac{1}{2}$

the extent to which its values are spread out. The two properties are totally different: Knowing a pdf's location tells us absolutely nothing about its dispersion. Table 3.6.1, for example, shows two simple discrete pdfs with the same expected value (equal to zero) but with vastly different dispersions.

It is not immediately obvious how the dispersion in a pdf should be quantified. Suppose that X is any discrete random variable. One seemingly reasonable approach would be to average the deviations of X from their mean—that is, calculate the expected value of $X - \mu$. As it happens, that strategy will not work because the negative deviations will exactly cancel the positive deviations, making the numerical value of such an average always zero, regardless of the amount of spread present in $p_X(k)$:

$$E(X - \mu) = E(X) - \mu = \mu - \mu = 0 \tag{3.6.1}$$

Another possibility would be to modify Equation 3.6.1 by making all the deviations positive—that is, replace $E(X - \mu)$ with $E(|X - \mu|)$. This does work, and it *is* sometimes used to measure dispersion, but the absolute value is somewhat troublesome mathematically: It does not have a simple arithmetic formula, nor is it a differentiable function. *Squaring* the deviations proves to be a much better approach.

Definition 3.6.1. The *variance* of a random variable is the expected value of its squared deviations from μ. If X is discrete with pdf $p_X(k)$,

$$\text{Var}(X) = \sigma^2 = E[(X - \mu)^2] = \sum_{\text{all } k}(k - \mu)^2 \cdot p_X(k)$$

If Y is continuous with pdf $f_Y(y)$,

$$\text{Var}(Y) = \sigma^2 = E[(Y - \mu)^2] = \int_{-\infty}^{\infty}(y - \mu)^2 \cdot f_Y(y)\,dy$$

[If $E(X^2)$ or $E(Y^2)$ is not finite, the variance is not defined.]

Comment. One unfortunate consequence of Definition 3.6.1 is that the units for the variance are the square of the units for the random variable: If Y is measured in inches, for example, the units for $\text{Var}(Y)$ are inches squared. This causes obvious problems in relating the variance back to the sample values. For that reason, in applied statistics, where unit compatibility is especially important, dispersion is measured not by the variance but by the *standard deviation*, which is defined to be the square root of the variance. That is,

$$\sigma = \text{standard deviation} = \begin{cases} \sqrt{\sum\limits_{\text{all } k} (k - \mu)^2 \cdot p_X(k)} & \text{if } X \text{ is discrete} \\ \sqrt{\int_{-\infty}^{\infty} (y - \mu)^2 \cdot f_Y(y)\, dy} & \text{if } Y \text{ is continuous} \end{cases}$$

Comment. The analogy between the expected value of a random variable and the center of gravity of a physical system was pointed out in Section 3.5. A similar equivalency holds between the variance and what engineers call a *moment of inertia*. If a set of weights having masses $m_1, m_2, \ldots$ are positioned along a (weightless) rigid bar at distances $r_1, r_2, \ldots$ from an axis of rotation (see Figure 3.6.1), the moment of inertia of the system is defined to be value $\sum\limits_i m_i r_i^2$. Notice, though, that if the masses were the probabilities associated with a discrete random variable and if the axis of rotation were actually μ, then $r_1, r_2, \ldots$ could be written $(k_1 - \mu), (k_2 - \mu), \ldots$ and $\sum\limits_i m_i r_i^2$ would be the same as the variance, $\sum\limits_{\text{all } k} (k - \mu)^2 \cdot p_X(k)$.

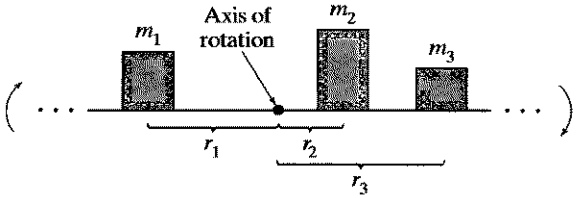


FIGURE 3.6.1

Definition 3.6.1 gives a formula for calculating σ^2 in both the discrete and continuous cases. An equivalent—but easier to use—formula is given in Theorem 3.6.1.

Theorem 3.6.1. *Let W be any random variable, discrete or continuous, having mean μ and for which $E(W^2)$ is finite. Then*

$$\text{Var}(W) = \sigma^2 = E(W^2) - \mu^2$$

Proof. We will prove the theorem for the continuous case. The argument for discrete W is similar. In Theorem 3.5.3, let $g(W) = (W - \mu)^2$. Then

$$\text{Var}(W) = E[(W - \mu)^2] = \int_{-\infty}^{\infty} g(w) f_W(w)\, dw = \int_{-\infty}^{\infty} (w - \mu)^2 f_W(w)\, dw$$

Squaring out the term $(w - \mu)^2$ that appears in the integrand and using the additive property of integrals gives

$$\begin{aligned}\int_{-\infty}^{\infty}(w-\mu)^2 f_W(w)\,dw &= \int_{-\infty}^{\infty}(w^2-2\mu w+\mu^2)f_W(w)\,dw\\ &= \int_{-\infty}^{\infty} w^2 f_W(w)\,dw - 2\mu\int_{-\infty}^{\infty} w f_W(w)\,dw + \int_{-\infty}^{\infty}\mu^2 f_W(w)\,dw\\ &= E(W^2)-2\mu^2+\mu^2 = E(W^2)-\mu^2.\end{aligned}$$

Note that the equality $\int_{-\infty}^{\infty} w^2 f_W(w)\,dw = E(W^2)$ also follows from Theorem 3.5.3. □

EXAMPLE 3.6.1

An urn contains five chips, two red and three white. Suppose that two are drawn out at random, *without replacement*. Let X denote the number of red chips in the sample. Find $\text{Var}(X)$.

Note, first, that since the chips are not being replaced from drawing to drawing X is a hypergeometric random variable. Moreover, we need to find μ, regardless of which formula is used to calculate σ^2. In the notation of Theorem 3.5.2, $r = 2$, $w = 3$, and $n = 2$, so

$$\mu = rn/(r + w) = 2 \cdot 2/(2 + 3) = 0.8$$

To find $\text{Var}(X)$ using Definition 3.6.1, we write

$$\begin{aligned}\text{Var}(X) = E[(X-\mu)^2] &= \sum_{\text{all } x}(x-\mu)^2 \cdot f_X(x)\\ &= (0-0.8)^2 \cdot \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} + (1-0.8)^2 \cdot \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}}\\ &\quad + (2-0.8)^2 \cdot \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}}\\ &= 0.36\end{aligned}$$

To use Theorem 3.6.1, we would first find $E(X^2)$. From Theorem 3.5.3,

$$\begin{aligned}E(X^2) = \sum_{\text{all } x} x^2 \cdot f_X(x) &= 0^2 \cdot \frac{\binom{2}{0}\binom{3}{2}}{\binom{5}{2}} + 1^2 \cdot \frac{\binom{2}{1}\binom{3}{1}}{\binom{5}{2}} + 2^2 \cdot \frac{\binom{2}{2}\binom{3}{0}}{\binom{5}{2}}\\ &= 1.00\end{aligned}$$

Then

$$\begin{aligned}\text{Var}(X) = E(X^2) - \mu^2 &= 1.00 - (0.8)^2 \\ &= 0.36\end{aligned}$$

confirming what we calculated earlier.

In Section 3.5 we encountered a change of scale formula that applied to expected values. For any constants a and b and any random variable W, $E(aW + b) = aE(W) + b$. A similar issue arises in connection with the *variance* of a linear transformation: If $\text{Var}(W) = \sigma^2$, what is the variance of $aW + b$?

Theorem 3.6.2. *Let W be any random variable having mean μ and where $E(W^2)$ is finite. Then $Var(aW + b) = a^2 Var(W)$.*

Proof. Using the same approach taken in the proof of Theorem 3.6.1, it can be shown that $E[(aW + b)^2] = a^2E(W^2) + 2ab\mu + b^2$. We also know from the Corollary to Theorem 3.5.3 that $E(aW + b) = a\mu + b$. Using Theorem 3.6.1, then, we can write

$$\begin{aligned}\text{Var}(aW + b) &= E[(aW + b)^2] - [E(aW + b)]^2 \\ &= [a^2E(W^2) + 2ab\mu + b^2] - [a\mu + b]^2 \\ &= [a^2E(W^2) + 2ab\mu + b^2] - [a^2\mu^2 + 2ab\mu + b^2] \\ &= a^2[E(W^2) - \mu^2] = a^2\text{Var}(W)\end{aligned}$$

☐

EXAMPLE 3.6.2

A random variable Y is described by the pdf

$$f_Y(y) = 2y, \quad 0 < y < 1$$

What is the standard deviation of $3Y + 2$?

First, we need to find the variance of Y. But

$$E(Y) = \int_0^1 y \cdot 2y\,dy = \frac{2}{3}$$

and

$$E(Y^2) = \int_0^1 y^2 \cdot 2y\,dy = \frac{1}{2}$$

so

$$\begin{aligned}\text{Var}(Y) = E(Y^2) - \mu^2 &= \frac{1}{2} - \left(\frac{2}{3}\right)^2 \\ &= \frac{1}{18}\end{aligned}$$

Then, by Theorem 3.6.2,

$$\begin{aligned}\text{Var}(3Y+2) &= (3)^2 \cdot \text{Var}(Y) = 9 \cdot \frac{1}{18} \\ &= \frac{1}{2}\end{aligned}$$

which makes the standard deviation of $3Y + 2$ equal to $\sqrt{\frac{1}{2}}$ or *0.71*.

QUESTIONS

3.6.1. Find Var(X) for the urn problem of Example 3.6.1 if the sampling is done *with* replacement.

3.6.2. Find the variance of Y if

$$f_Y(y) = \begin{cases} \frac{3}{4}, & 0 \le y \le 1 \\ \frac{1}{4}, & 2 \le y \le 3 \\ 0, & \text{elsewhere} \end{cases}$$

3.6.3. Ten equally qualified applicants, six men and four women, apply for three lab technician positions. Unable to justify choosing any of the applicants over all the others, the personnel director decides to select the three at random. Let X denote the number of men hired. Compute the standard deviation of X.

3.6.4. Compute the variance for a uniform random variable defined on the unit interval.

3.6.5. Use Theorem 3.6.1 to find the variance of the random variable Y, where

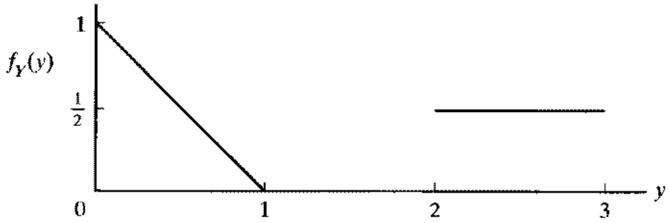
$$f_Y(y) = 3(1-y)^2, \quad 0 < y < 1$$

3.6.6. If

$$f_Y(y) = \frac{2y}{k^2}, \quad 0 \le y \le k$$

for what value of k does Var(Y) = 2?

3.6.7. Calculate the standard deviation, σ, for the random variable Y whose pdf has the graph shown below:



3.6.8. Consider the pdf defined by

$$f_Y(y) = \frac{2}{y^3}, \quad y \geq 1$$

Show that (a) $\int_1^\infty f_Y(y)\,dy = 1$, (b)$E(Y) = 2$, and (c) $\text{Var}(Y)$ is not finite.

3.6.9. Frankie and Johnny play the following game. Frankie selects a number at random from the interval $[a, b]$. Johnny, not knowing Frankie's number, is to pick a second number from that same inverval and pay Frankie an amount, W, equal to the squared difference between the two [so $0 \leq W \leq (b - a)^2$]. What should be Johnny's strategy if he wants to minimize his expected loss?

3.6.10. Let Y be a random variable whose pdf is given by $f_Y(y) = 5y^4, 0 \leq y \leq 1$. Use Theorem 3.6.1 to find $\text{Var}(Y)$.

3.6.11. Suppose that Y is an exponential random variable, so $f_Y(y) = \lambda e^{-\lambda y}$, $y \geq 0$. Show that the variance of Y is $1/\lambda^2$.

3.6.12. Suppose that Y is an exponential random variable with $\lambda = 2$ (recall Question 3.6.11). Find $P(Y > E(Y) + 2\sqrt{\text{Var}(Y)})$.

3.6.13. Let X be a random variable with finite mean μ. Define for every real number $a, g(a) = E[(X - a)^2]$. Show that

$$g(a) = E[(X - \mu)^2] + (\mu - a)^2.$$

What is another name for $\min\limits_a g(a)$?

3.6.14. Let Y have the pdf given in Question 3.6.5. Find the variance of W, where $W = -5Y + 12$.

3.6.15. If Y denotes a temperature recorded in degrees Fahrenheit, then $\frac{5}{9}(Y - 32)$ is the corresponding temperature in degrees Celsius. If the standard deviation for a set of temperatures is 15.7°F, what is the standard deviation of the equivalent Celsius temperatures?

3.6.16. If $E(W) = \mu$ and $\text{Var}(W) = \sigma^2$, show that

$$E\left(\frac{W - \mu}{\sigma}\right) = 0 \quad \text{and} \quad \text{Var}\left(\frac{W - \mu}{\sigma}\right) = 1$$

3.6.17. Suppose U is a uniform random variable over $[0, 1]$.

(a) Show that $Y = (b - a)U + a$ is uniform over $[a, b]$

(b) Use Part (a) and Question 3.6.4 to find the variance of Y.

Higher Moments

The quantities we have identified as the mean and the variance are actually special cases of what are referred to more generally as the *moments* of a random variable. More precisely, $E(W)$ is the *first moment about the origin* and σ^2 is the *second moment about the mean*. As the terminology suggests, we will have occasion to define higher moments of W. Just as $E(W)$ and σ^2 reflect a random variable's location and dispersion, so it is possible to characterize other aspects of a distribution in terms of other moments. We will see, for example, that the skewness of a distribution—that is, the extent to which it is not symmetric around μ—can be effectively measured in terms of a *third* moment. Likewise, there are issues that arise in certain applied statistics problems that require a knowledge of the flatness of a pdf, a property that can be quantified by the *fourth* moment.

Definition 3.6.2. Let W be any random variable with pdf $f_W(w)$. For any positive integer r,

1. The rth *moment of W about the origin*, μ_r, is given by

$$\mu_r = E(W^r)$$

provided $\int_{-\infty}^{\infty} |w|^r \cdot f_W(w)\,dw < \infty$ (or provided the analogous condition on the *summation* of $|w|^r$ holds, if W is discrete). When $r = 1$, we usually delete the subscript and write $E(W)$ as μ rather than μ_1.

2. The rth *moment of W about the mean*, μ'_r, is given by

$$\mu'_r = E[(W - \mu)^r]$$

provided the finiteness conditions of part 1 hold.

Comment. We can express μ'_r in terms of μ_j, $j = 1, 2, \ldots, r$, by simply writing out the binomial expansion of $(W - \mu)^r$:

$$\mu'_r = E[(W - \mu)^r] = \sum_{j=0}^{r} \binom{r}{j} E(W^j)(-\mu)^{r-j}$$

Thus,

$$\begin{aligned}
\mu'_2 &= E[(W - \mu)^2] = \sigma^2 = \mu_2 - \mu_1^2 \\
\mu'_3 &= E[(W - \mu)^3] = \mu_3 - 3\mu_1\mu_2 + 2\mu_1^3 \\
\mu'_4 &= E[(W - \mu)^4] = \mu_4 - 4\mu_1\mu_3 + 6\mu_1^2\mu_2 - 3\mu_1^4
\end{aligned}$$

and so on.

EXAMPLE 3.6.3

The *skewness* of a pdf can be measured in terms of its third moment about the mean. If a pdf is symmetric, $E[(W - \mu)^3]$ will obviously be zero; for pdfs not symmetric, $E[(W - \mu)^3]$ will not be zero. In practice, the symmetry (or lack of symmetry) of a pdf is often measured by the *coefficient of skewness*, γ_1, where

$$\gamma_1 = \frac{E[(W - \mu)^3]}{\sigma^3}$$

Dividing μ'_3 by σ^3 makes γ_1 dimensionless.

A second "shape" parameter in common use is the *coefficient of kurtosis*, γ_2, which involves the *fourth* moment about the mean. Specifically,

$$\gamma_2 = \frac{E[(W - \mu)^4]}{\sigma^4} - 3$$

For certain pdf's, γ_2 is a useful measure of peakedness: relatively flat pdf's are said to be *platykurtic*; more peaked pdf's are called *leptokurtic* [see (97)].

Earlier in this chapter we encountered random variables whose means did not exist—recall, for example, the St. Petersburg paradox. More generally, there are random variables having certain of their higher moments finite and certain others, not finite. Addressing the question of whether or not a given $E(W^j)$ is finite is the following existence theorem.

Theorem 3.6.3. *If the* k*th moment of a random variable exists, all moments of order less than k exist.*

Proof. Let $f_Y(y)$ be the pdf of a continuous random variable Y. By Definition 3.6.2, $E(Y^k)$ exists if and only if

$$\int_{-\infty}^{\infty} |y|^k \cdot f_Y(y)\,dy < \infty \tag{3.6.2}$$

Let $1 \leq j < k$. To prove the theorem we must show that

$$\int_{-\infty}^{\infty} |y|^j \cdot f_Y(y)\,dy < \infty$$

is implied by Inequality 3.6.2. But

$$\begin{aligned}\int_{-\infty}^{\infty} |y|^j \cdot f_Y(y)\,dy &= \int_{|y|\leq 1} |y|^j \cdot f_Y(y)\,dy + \int_{|y|>1} |y|^j \cdot f_Y(y)\,dy \\ &\leq \int_{|y|\leq 1} f_Y(y)\,dy + \int_{|y|>1} |y|^j \cdot f_Y(y)\,dy \\ &\leq 1 + \int_{|y|>1} |y|^j \cdot f_Y(y)\,dy \\ &\leq 1 + \int_{|y|>1} |y|^k \cdot f_Y(y)\,dy < \infty\end{aligned}$$

Therefore, $E(Y^j)$ exists, $j = 1, 2, \ldots, k - 1$. The proof for discrete random variables is similar. ☐

EXAMPLE 3.6.4

Many of the random variables that play a major role in statistics have moments existing for *all k*, as does, for instance, the normal distribution introduced in Example 3.4.3. Still, it is not difficult to find well-known models for which this is *not* true. A case in point is the *Student t distribution,* a probability function widely used in inference procedures. (See Chapter 7.)

The pdf for a Student t random variable is given by

$$f_Y(y) = \frac{c(n)}{\left(1 + \dfrac{y^2}{n}\right)^{(n+1)/2}}, \quad -\infty < y < \infty, \quad n \geq 1$$

where n is referred to as the distribution's "degrees of freedom" and $c(n)$ is a constant. By definition, the $(2k)$th moment is the integral

$$E(Y^{2k}) = c(n) \cdot \int_{-\infty}^{\infty} \frac{y^{2k}}{\left(1 + \frac{y^2}{n}\right)^{(n+1)/2}} \, dy$$

Is $E(Y^{2k})$ finite?

Not necessarily. Recall from calculus that an integral of the form

$$\int_{-\infty}^{\infty} \frac{1}{y^{\alpha}} \, dy$$

will converge only if $\alpha > 1$. Also, the convergence properties for integrals of

$$\frac{y^{2k}}{\left(1 + \frac{y^2}{n}\right)^{(n+1)/2}}$$

are the same as those for

$$\frac{y^{2k}}{(y^2)^{(n+1)/2}} = \frac{1}{y^{n+1-2k}}$$

Therefore, if $E(Y^{2k})$ is to be finite, we must have

$$n + 1 - 2k > 1$$

or, equivalently, $2k < n$. Thus a Student t random variable with, say, $n = 9$ degrees of freedom has $E(X^8) < \infty$, but no moment of order higher than eight exists.

QUESTIONS

3.6.18. Let Y be a uniform random variable defined over the interval $(0, 2)$. Find an expression for the rth moment of Y about the origin. Also, use the binomial expansion as described in the comment to find $E[(Y - \mu)^6]$.

3.6.19. Find the coefficient of skewness for an exponential random variable having the pdf

$$f_Y(y) = e^{-y}, \quad y > 0$$

3.6.20. Calculate the coefficient of kurtosis for a uniform random variable defined over the unit interval, $f_Y(y) = 1$, for $0 \le y \le 1$.

3.6.21. Suppose that W is a random variable for which $E[(W - \mu)^3] = 10$ and $E(W^3) = 4$. Is it possible that $\mu = 2$?

3.6.22. If $Y = aX + b$, show that Y has the same coefficients of skewness and kurtosis as X.

3.6.23. Let Y be the random variable of Question 3.4.6, where for a positive integer n, $f_Y(y) = (n+2)(n+1)y^n(1-y), 0 \leq y \leq 1$.

(a) Find Var(Y)

(b) For any positive integer k, find the kth moment around the origin.

3.6.24. Suppose that the random variable Y is described by the pdf

$$f_Y(y) = c \cdot y^{-6}, \quad y > 1$$

(a) Find c.

(b) What is the highest moment of Y that exists?

JOINT DENSITIES

Sections 3.3. and 3.4 introduced the basic terminology for describing the probabilistic behavior of a *single* random variable. Such information, while adequate for many problems, is insufficient when more than one variable is of interest to the experimenter. Medical researchers, for example, continue to explore the relationship between blood cholesterol and heart disease, and, more recently, between "good" cholesterol and "bad" cholesterol. And more than a little attention—both political and pedagogical—is given to the role played by $K-12$ funding in the performance of would-be high school graduates on exit exams. On a smaller scale, electronic equipment and systems are often designed to have built-in redundancy: Whether or not that equipment functions properly ultimately depends on the reliability of two different components.

The point is, there are many situations where two relevant random variables, say X and Y,[2] are defined on the same sample space. Knowing only $f_X(x)$ and $f_Y(y)$, though, does not necessarily provide enough information to characterize the all-important *simultaneous* behavior of X and Y. The purpose of this section is to introduce the concepts, definitions, and mathematical techniques associated with distributions based on two (or more) random variables.

Discrete Joint Pdfs

As we saw in the single-variable case, the pdf is defined differently, depending on whether the random variable is discrete or continuous. The same distinction applies to joint pdfs. We begin with a discussion of joint pdfs as they apply to two discrete random variables.

Definition 3.7.1. Suppose S is a discrete sample space on which two random variables, X and Y, are defined. The *joint probability density function of X and Y (or joint pdf)* is denoted $p_{X,Y}(x, y)$, where

$$p_{X,Y}(x, y) = P(\{s | X(s) = x \quad \text{and} \quad Y(s) = y\})$$

---

[2] For the next several sections we will suspend our earlier practice of using X to denote a discrete random variable and Y to denote a continuous random variable. The category of the random variables will need to be determined from the context of the problem. Typically, though, X and Y will either both be discrete or both be continuous.

Comment. A convenient shorthand notation for the meaning of $p_{X,Y}(x, y)$, consistent with what was used earlier for pdfs of single discrete random variables, is to write $p_{X,Y}(x, y) = P(X = x, Y = y)$.

EXAMPLE 3.7.1

A supermarket has two express lines. Let X and Y denote the number of customers in the first and in the second, respectively, at any given time. During nonrush hours, the joint pdf of X and Y is summarized by the following table:

		X			
		0	1	2	3
Y	0	0.1	0.2	0	0
	1	0.2	0.25	0.05	0
	2	0	0.05	0.05	0.025
	3	0	0	0.025	0.05

Find $P(|X - Y| = 1)$, the probability that X and Y differ by exactly one.

By definition,

$$\begin{aligned}
P(|X - Y| = 1) &= \sum_{|x-y|=1}\sum p_{X,Y}(x, y) \\
&= p_{X,Y}(0, 1) + p_{X,Y}(1, 0) + p_{X,Y}(1, 2) \\
&\quad + p_{X,Y}(2, 1) + p_{X,Y}(2, 3) + p_{X,Y}(3, 2) \\
&= 0.2 + 0.2 + 0.05 + 0.05 + 0.025 + 0.025 \\
&= 0.55
\end{aligned}$$

[Would you expect $p_{X,Y}(x, y)$ to be symmetric? Would you expect the event $|X - Y| \geq 2$ to have zero probability?]

EXAMPLE 3.7.2

Suppose two fair dice are rolled. Let X be the sum of the numbers showing, and let Y be the larger of the two. So, for example,

$$p_{X,Y}(2, 3) = P(X = 2, Y = 3) = P(\emptyset) = 0$$

$$p_{X,Y}(4, 3) = P(X = 4, Y = 3) = P(\{(1, 3)(3, 1)\}) = \frac{2}{36}$$

and

$$p_{X,Y}(6, 3) = P(X = 6, Y = 3) = P(\{(3, 3)\} = \frac{1}{36}$$

The entire joint pdf is given in Table 3.7.1.

TABLE 3.7.1

y x	1	2	3	4	5	6	Row totals
2	1/36	0	0	0	0	0	1/36
3	0	2/36	0	0	0	0	2/36
4	0	1/36	2/36	0	0	0	3/36
5	0	0	2/36	2/36	0	0	4/36
6	0	0	1/36	2/36	2/36	0	5/36
7	0	0	0	2/36	2/36	2/36	6/36
8	0	0	0	1/36	2/36	2/36	5/36
9	0	0	0	0	2/36	2/36	4/36
10	0	0	0	0	1/36	2/36	3/36
11	0	0	0	0	0	2/36	2/36
12	0	0	0	0	0	1/36	1/36
Col. totals	1/36	3/36	5/36	7/36	9/36	11/36	

Notice that the row totals in the right-hand margin of the table give the pdf for X. Similarly, the column totals along the bottom detail the pdf for Y. Those are not coincidences. Theorem 3.7.1 gives a formal statement of the relationship between the joint pdf and the individual pdfs.

Theorem 3.7.1. *Suppose that $p_{X,Y}(x, y)$ is the joint pdf of the discrete random variables X and Y. Then*

$$p_X(x) = \sum_{\text{all } y} p_{X,Y}(x, y) \quad \text{and} \quad p_Y(y) = \sum_{\text{all } x} p_{X,Y}(x, y)$$

Proof. We will prove the first statement. Note that the collection of sets $(Y = y)$ for all y form a partition of S; that is, they are disjoint and $\bigcup_{\text{all } y}(Y = y) = S$. The set $(X = x) = (X = x) \cap S = (X = x) \cap \bigcup_{\text{all } y}(Y = y) = \bigcup_{\text{all } y}[(X = x) \cap (Y = y)]$, so

$$\begin{aligned} p_X(x) = P(X = x) &= P\left(\bigcup_{\text{all } y}[(X = x) \cap (Y = y)]\right) \\ &= \sum_{\text{all } y} P(X = x, Y = y) = \sum_{\text{all } y} p_{X,Y}(x, y). \end{aligned}$$

□

Definition 3.7.2. An individual pdf obtained by summing a joint pdf over all values of the other random variable is called a *marginal pdf.*

Continuous Joint Pdfs

If X and Y are both continuous random variables, Definition 3.7.1 does not apply because $P(X = x, Y = y)$ will be identically 0 for all (x, y). As was the case in single-variable situations, the joint pdf for two continuous random variables will be defined as a function, which when integrated yields the probability that (X, Y) lies in a specified region of the xy-plane.

Definition 3.7.3. Two random variables defined on the same set of real numbers are *jointly continuous* if there exists a function $f_{X,Y}(x, y)$ such that for any region R in the xy-plane $P((X, Y) \in R) = \int\int_R f_{X,Y}(x, y)\, dx\, dy$. The function $f_{X,Y}(x, y)$ is the *joint pdf of X and Y*.

Note: Any function $f_{X,Y}(x, y)$ for which

1. $f_{X,Y}(x, y) \geq 0$ for all x and y
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dx\, dy = 1$

qualifies as a joint pdf. We shall employ the convention of naming only the domain where the joint pdf is nonzero; everywhere else it will be assumed to be 0. This is analogous, of course, to the notation used earlier in describing the domain of single random variables.

EXAMPLE 3.7.3

Suppose that the variation in two continuous random variables, X and Y, can be modeled by the joint pdf $f_{X,Y}(x, y) = cxy$, for $0 < y < x < 1$. Find c.

By inspection, $f_{X,Y}(x, y)$ will be non-negative as long as $c \geq 0$. The particular c that qualifies $f_{X,Y}(x, y)$ to be a joint pdf, though, is the one that makes the volume under $f_{X,Y}(x, y)$ equal to 1. But

$$\iint_S cxy\, dy\, dx = 1 = c\int_0^1 \left(\int_0^x (xy)\, dy\right) dx = c\int_0^1 x\left(\frac{y^2}{2}\Big|_0^x\right) dx$$

$$= c\int_0^1 \left(\frac{x^3}{2}\right) dx = c\frac{x^4}{8}\Big|_0^1 = \left(\frac{1}{8}\right)c$$

Therefore, $c = 8$.

EXAMPLE 3.7.4

A study claims that the daily number of hours, X, a teenager watches television and the daily number of hours, Y, he works on his homework are approximated by the joint pdf,

$$f_{X,Y}(x, y) = xye^{-(x+y)}, \quad x > 0, \quad y > 0$$

What is the probability a teenager chosen at random spends at least twice as much time watching television as he does working on his homework?

The region, R, in the xy-plane corresponding to the event "$X \geq 2Y$" is shown in Figure 3.7.1. It follows that $P(X \geq 2Y)$ is the volume under $f_{X,Y}(x, y)$ above the region R:

$$P(X \geq 2Y) = \int_0^{\infty}\int_0^{x/2} xye^{-(x+y)}\, dy\, dx$$

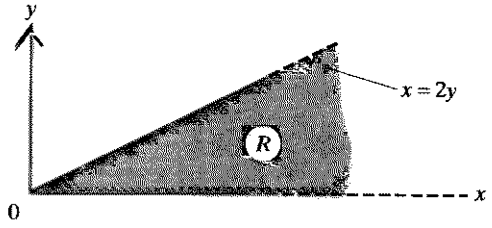


FIGURE 3.7.1

Separating variables, we can write

$$P(X \geq 2Y) = \int_0^{\infty} xe^{-x} \left[\int_0^{x/2} ye^{-y}dy \right] dx$$

and the double integral reduces to $\frac{7}{27}$:

$$\begin{aligned} P(X \geq 2Y) &= \int_0^{\infty} xe^{-x} \left[1 - \left(\frac{x}{2} + 1 \right) e^{-x/2} \right] dx \\ &= \int_0^{\infty} xe^{-x} dx - \int_0^{\infty} \frac{x^2}{2} e^{-3x/2}\, dx - \int_0^{\infty} xe^{-3x/2} dx \\ &= 1 - \frac{16}{54} - \frac{4}{9} \\ &= \frac{7}{27} \end{aligned}$$

Geometric Probability

One particularly important special case of Definition 3.7.3 is the *joint uniform pdf*, which is represented by a surface having a constant height everywhere above a specified rectangle in the xy-plane. That is,

$$f_{X,Y}(x, y) = \frac{1}{(b - a)(d - c)}, \quad a \leq x \leq b, c \leq y \leq d$$

If R is some region in the rectangle where X and Y are defined, $P((X, Y) \in R)$ reduces to a simple ratio of areas:

$$P((X, Y) \in R) = \frac{\text{area of } R}{(b - a)(d - c)} \tag{3.7.1}$$

Calculations based on Equation 3.7.1 are referred to as *geometric probabilities*.

EXAMPLE 3.7.5

Two friends agree to meet on the University Commons "sometime around 12:30." But neither of them is particularly punctual—or patient. What will actually happen is that

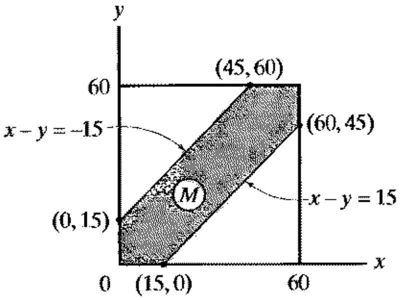


FIGURE 3.7.2

each will arrive at random sometime in the interval from 12:00 to 1:00. If one arrives and the other is not there, the first person will wait fifteen minutes or until 1:00, whichever comes first, and then leave. What is the probability the two will get together?

To simplify notation, we can represent the time period from 12:00 to 1:00 as the interval from zero to sixty minutes. Then if x and y denote the two arrival times, the sample space is the 60×60 square shown in Figure 3.7.2. Furthermore, the event M, "the two friends meet," will occur if and only if $|x - y| \leq 15$ or, equivalently, if and only if $-15 \leq x - y \leq 15$. These inequalities appear as the shaded region in Figure 3.7.2.

Notice that the areas of the two triangles above and below M are each equal to $\frac{1}{2}(45)(45)$. It follows that the two friends have a 44% chance of meeting:

$$\begin{aligned} P(M) &= \frac{\text{area of } M}{\text{area of } S} \\ &= \frac{(60)^2 - 2\left[\frac{1}{2}(45)(45)\right]}{(60)^2} \\ &= 0.44 \end{aligned}$$

EXAMPLE 3.7.6

A carnival operator wants to set up a ringtoss game. Players will throw a ring of diameter d onto a grid of squares, the side of each square being of length s (see Figure 3.7.3). If the ring lands entirely inside a square, the player wins a prize. To ensure a profit, the operator

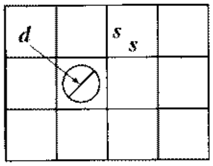


FIGURE 3.7.3

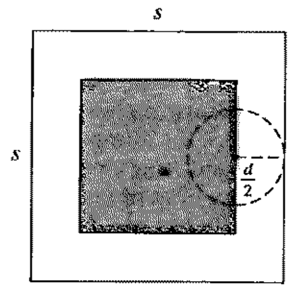


FIGURE 3.7.4

must keep the player's chances of winning down to something less than one in five. How small can the operator make the ratio d/s?

First, it will be assumed that the player is required to stand far enough away so that no skill is involved and the ring is falling at random on the grid. From Figure 3.7.4, we see that in order for the ring not to touch any side of the square, the ring's center must be somewhere in the interior of a smaller square, each side of which is a distance $d/2$ from one of the grid lines.

Since the area of a grid square is s^2 and the area of an interior square is $(s - d)^2$, the probability of a winning toss can be written as the ratio:

$$P(\text{ring touches no lines}) = \frac{(s - d)^2}{s^2}$$

But the operator requires that

$$\frac{(s - d)^2}{s^2} \leq 0.20$$

Solving for d/s gives

$$\frac{d}{s} \geq 1 - \sqrt{0.20} = 0.55$$

That is, if the diameter of the ring is at least 55% as long as the side of one of the squares, the player will have no more than a 20% chance of winning.

QUESTIONS

3.7.1. If $p_{X,Y}(x, y) = cxy$ at the points (1, 1), (2, 1), (2, 2), and (3, 1), and equals 0 elsewhere, find c.

3.7.2. Let X and Y be two continuous random variables defined over the unit square. What does c equal if $f_{X,Y}(x, y) = c(x^2 + y^2)$?

3.7.3. Suppose that random variables X and Y vary in accordance with the joint pdf, $f_{X,Y}(x, y) = c(x + y), 0 < x < y < 1$. Find c.

3.7.4. Find c if $f_{X,Y}(x, y) = cxy$ for X and Y defined over the triangle whose vertices are the points $(0, 0)$, $(0, 1)$, and $(1, 1)$.

3.7.5. An urn contains four red chips, three white chips, and two blue chips. A random sample of size 3 is drawn without replacement. Let X denote the number of white chips in the sample and Y the number of blue. Write a formula for the joint pdf of X and Y.

3.7.6. Four cards are drawn from a standard poker deck. Let X be the number of kings drawn and Y the number of queens. Find $p_{X,Y}(x, y)$.

3.7.7. An advisor looks over the schedules of his 50 students to see how many math and science courses each has registered for in the coming semester. He summarizes his results in a table. What is the probability that a student selected at random will have signed up for more math courses than science courses?

		Number of math courses, X		
		0	1	2
Number or science courses, Y	0	11	6	4
	1	9	10	3
	2	5	0	2

3.7.8. Consider the experiment of tossing a fair coin three times. Let X denote the number of heads on the last flip, and let Y denote the total number of heads on the three flips. Find $p_{X,Y}(x, y)$.

3.7.9. Suppose that two fair dice are tossed one time. Let X denote the number of 2's that appear, and Y the number of 3's. Write the matrix giving the joint probability density function for X and Y. Suppose a third random variable, Z, is defined, where $Z = X + Y$. Use $p_{X,Y}(x, y)$ to find $p_Z(z)$.

3.7.10. Suppose that X and Y have a bivariate uniform density over the unit square:

$$f_{X,Y}(x, y) = \begin{cases} c, & 0 < x < 1, \quad 0 < y < 1 \\ 0, & \text{elsewhere} \end{cases}$$

(a) Find c.
(b) Find $P\left(0 < X < \frac{1}{2}, 0 < Y < \frac{1}{4}\right)$.

3.7.11. Let X and Y have the joint pdf

$$f_{X,Y}(x, y) = 2e^{-(x+y)}, \quad 0 < x < y, \ 0 < y$$

Find $P(Y < 3X)$.

3.7.12. A point is chosen at random from the interior of a circle whose equation is $x^2 + y^2 \leq 4$. Let the random variables X and Y denote the x- and y-coordinates of the sampled point. Find $f_{X,Y}(x, y)$.

3.7.13. Find $P(X < 2Y)$ if $f_{X,Y}(x, y) = x + y$ for X and Y each defined over the unit interval.

3.7.14. Suppose that five independent observations are drawn from the continuous pdf $f_T(t) = 2t, 0 \leq t \leq 1$. Let X denote the number of t's that fall in the interval $0 \leq t < \frac{1}{3}$ and let Y denote the number of t's that fall in the interval $\frac{1}{3} \leq t < \frac{2}{3}$. Find $p_{X,Y}(1, 2)$.

3.7.15. A point is chosen at random from the interior of a right triangle with base b and height h. What is the probability that the y value is between 0 and $h/2$?

Marginal Pdfs for Continuous Random Variables

The notion of marginal pdfs in connection with discrete random variables was introduced in Theorem 3.7.1 and Definition 3.7.2. An analogous relationship holds in the continuous case—*integration,* though, replaces the summation that appears in Theorem 3.7.1.

Theorem 3.7.2. *Suppose X and Y are jointly continuous with joint pdf $f_{X,Y}(x, y)$. Then the* marginal pdfs, *$f_X(x)$ and $f_Y(y)$, are given by*

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

Proof. It suffices to verify the first of the theorem's two equalities. As is often the case with proofs for continuous random variables, we begin with the cdf:

$$F_X(x) = P(X \leq x) = \int_{-\infty}^{\infty}\int_{-\infty}^{x} f_{X,Y}(t, y)\,dt\,dy = \int_{-\infty}^{x}\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\,dt$$

Differentiating both ends of the equation above gives

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

(recall Theorem 3.4.1). □

EXAMPLE 3.7.7

Suppose that two continuous random variables, X and Y, have the joint uniform pdf,

$$f_{X,Y}(x, y) = \frac{1}{6}, \quad 0 \leq x \leq 3, \quad 0 \leq y \leq 2$$

Find $f_X(x)$.

Applying Theorem 3.7.2 gives

$$f_X(x) = \int_0^2 f_{X,Y}(x, y)\,dy = \int_0^2 \frac{1}{6}\,dy = \frac{1}{3}, \quad 0 \leq x \leq 3$$

Notice that X, by itself, is a uniform random variable defined over the interval [0, 3]; similarly, we would find that $f_Y(y)$ has a uniform pdf over the interval [0, 2].

EXAMPLE 3.7.8

Consider the case where X and Y are two continuous random variables, jointly distributed over the first quadrant of the xy-plane according to the joint pdf,

$$f_{X,Y}(x, y) = \begin{cases} y^2 e^{-y(x+1)}, & x \geq 0, \quad y \geq 0 \\ 0, & \text{elsewhere} \end{cases}$$

Find the two marginal pdf's.

First, consider $f_X(x)$. By Theorem 3.7.2,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy = \int_0^{\infty} y^2 e^{-y(x+1)}\,dy$$

In the integrand, substitute

$$u = y(x + 1)$$

making $du = (x + 1)\,dy$. This gives

$$f_X(x) = \frac{1}{x+1}\int_0^{\infty} \frac{u^2}{(x+1)^2} e^{-u}\,du = \frac{1}{(x+1)^3}\int_0^{\infty} u^2 e^{-u}\,du$$

After applying integration by parts (twice) to $\int_0^{\infty} u^2 e^{-u}\,du$, we get

$$\begin{aligned} f_X(x) &= \frac{1}{(x+1)^3}\left[-u^2e^{-u} - 2ue^{-u} - 2e^{-u}\right]\Big|_0^{\infty} \\ &= \frac{1}{(x+1)^3}\left[2 - \lim_{u\to\infty}\left(\frac{u^2}{e^u} + \frac{2u}{e^u} + \frac{2}{e^u}\right)\right] \\ &= \frac{2}{(x+1)^3}, \quad x \geq 0 \end{aligned}$$

Finding $f_Y(y)$ is a bit easier:

$$\begin{aligned} f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx = \int_0^{\infty} y^2 e^{-y(x+1)}\,dx \\ &= y^2e^{-y}\int_0^{\infty} e^{-yx}\,dx = y^2e^{-y}\left(\frac{1}{y}\right)\left(-e^{-yx}\Big|_0^{\infty}\right) \\ &= ye^{-y}, \quad y \geq 0 \end{aligned}$$

QUESTIONS

3.7.16. Find the marginal pdf of X for the joint pdf derived in Question 3.7.5.

3.7.17. Find the marginal pdfs of X and Y for the joint pdf derived in Question 3.7.8.

3.7.18. The campus recruiter for an international conglomerate classifies the large number of students she interviews into three categories—the lower quarter, the middle half, and the upper quarter. If she meets six students on a given morning, what is the probability that they will be evenly divided among the three categories? What is the marginal probability that exactly two will belong to the middle half?

3.7.19. For each of the following joint pdf's, find $f_X(x)$ and $f_Y(y)$.

(a) $f_{X,Y}(x, y) = \frac{1}{2}, 0 \leq x \leq 2, 0 \leq y \leq 1$

(b) $f_{X,Y}(x, y) = \frac{3}{2}y^2, 0 \leq x \leq 2, 0 \leq y \leq 1$

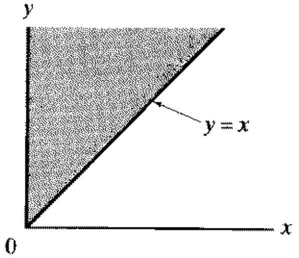
(c) $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y), 0 \le x \le 1, 0 \le y \le 1$
(d) $f_{X,Y}(x, y) = c(x + y), 0 \le x \le 1, 0 \le y \le 1$
(e) $f_{X,Y}(x, y) = 4xy, 0 \le x \le 1, 0 \le y \le 1$
(f) $f_{X,Y}(x, y) = xye^{-(x+y)}, 0 \le x, 0 \le y$
(g) $f_{X,Y}(x, y) = ye^{-xy-y}, 0 \le x, 0 \le y$

3.7.20. For each of the following joint pdf's, find $f_X(x)$ and $f_Y(y)$.

(a) $f_{X,Y}(x, y) = \frac{1}{2}, 0 \le x \le y \le 2$
(b) $f_{X,Y}(x, y) = \frac{1}{x}, 0 \le y \le x \le 1$
(c) $f_{X,Y}(x, y) = 6x, 0 \le x \le 1, 0 \le y \le 1 - x$

3.7.21. Suppose that $f_{X,Y}(x, y) = 6(1 - x - y)$ for x and y defined over the unit square, subject to the restriction that $0 \le x + y \le 1$. Find the marginal pdf for X.

3.7.22. Find $f_Y(y)$ if $f_{X,Y}(x, y) = 2e^{-x}e^{-y}$ for x and y defined over the shaded region pictured.



3.7.23. Suppose that X and Y are discrete random variables with

$$p_{X,Y}(x, y) = \frac{4!}{x!y!(4 - x - y)!}\left(\frac{1}{2}\right)^x\left(\frac{1}{3}\right)^y\left(\frac{1}{6}\right)^{4-x-y}, \quad 0 \le x + y \le 4$$

Find $p_X(x)$ and $p_Y(x)$.

3.7.24. A generalization of the binomial model occurs when there is a sequence of n independent trials with *three* outcomes, where $p_1 = P(\text{Outcome 1})$ and $p_2 = P(\text{Outcome 2})$. Let X and Y denote the number of trials (out of n) resulting in Outcome 1 and Outcome 2, respectively.

(a) Show that $p_{X,Y}(x, y)\dfrac{n!}{x!y!(n - x - y)!}p_1^x p_2^y(1 - p_1 - p_2)^{n-x-y}, 0 \le x + y \le n$
(b) Find $p_X(x)$ and $p_Y(x)$.
Hint: See Question 3.7.23.

Joint Cdfs

For a single random variable X, the cdf of X evaluated at some point x—that is, $F_X(x)$—is the probability that the random variable X takes on a value less than or equal to x. Extended to two variables, a *joint cdf* (evaluated at the point (u, v)) is the probability that $X \le u$ and, simultaneously, $Y \le v$.

Definition 3.7.4. Let X and Y be any two random variables. The *joint cumulative distribution function of X and Y* (or *joint cdf*) is denoted $F_{X,Y}(u, v)$, where

$$F_{X,Y}(u, v) = P(X \le u \quad \text{and} \quad Y \le v)$$

EXAMPLE 3.7.9

Find the joint cdf, $F_{X,Y}(u, v)$, for the two random variables, X and Y, whose joint pdf is given by $f_{X,Y}(x, y) = \frac{4}{3}(x + xy), 0 \leq x \leq 1, 0 \leq y \leq 1$.

If Definition 3.7.4 is applied, the probability that $X \leq u$ and $Y \leq v$ becomes a double integral of $f_{X,Y}(x, y)$:

$$\begin{aligned} F_{X,Y}(u, v) &= \frac{4}{3}\int_0^v \int_0^u (x + xy)\,dx\,dy = \frac{4}{3}\int_0^v \left(\int_0^u (x + xy)\,dx\right) dy \\ &= \frac{4}{3}\int_0^v \left(\frac{x^2}{2}(1 + y)\Bigg|_0^u\right) dy = \frac{4}{3}\int_0^v \frac{u^2}{2}(1 + y)\,dy \\ &= \frac{4}{3}\frac{u^2}{2}\left(y + \frac{y^2}{2}\right)\Bigg|_0^v = \frac{4}{3}\frac{u^2}{2}\left(v + \frac{v^2}{2}\right) \end{aligned}$$

which simplifies to

$$F_{X,Y}(u, v) = \frac{1}{3}u^2(2v + v^2)$$

(For what values of u and v is $F_{X,Y}(u, v)$ defined?)

Theorem 3.7.3. *Let $F_{X,Y}(u, v)$ be the joint cdf associated with the continuous random variables X and Y. Then the joint pdf of X and Y, $f_{X,Y}(x, y)$ is a second partial derivative of the joint cdf—that is, $f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y}F_{X,Y}(x, y)$, provided $F_{X,Y}(x, y)$ has continuous second partial derivatives.*

EXAMPLE 3.7.10

What is the joint pdf of the random variables X and Y whose joint cdf is $F_{X,Y}(x, y) = \frac{1}{3}x^2(2y + y^2)$?

By Theorem 3.7.3,

$$\begin{aligned} f_{X,Y}(x, y) &= \frac{\partial^2}{\partial x\, \partial y}F_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y}\frac{1}{3}x^2(2y + y^2) \\ &= \frac{\partial}{\partial y}\frac{2}{3}x(2y + y^2) = \frac{2}{3}x(2 + 2y) = \frac{4}{3}(x + xy) \end{aligned}$$

Notice the similarity between Examples 3.7.9 and 3.7.10—$f_{X,Y}(x, y)$ is the same in both examples; so is $F_{X,Y}(x, y)$.

Multivariate Densities

The definitions and theorems in this section extend in a very straightforward way to situations involving more than two variables. The joint pdf for n discrete random variables, for example, is denoted $p_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ where

$$p_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 = x_1,\ldots,X_n = x_n)$$

For n continuous random variables, the joint pdf is that function $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ having the property that for any region R in n-space,

$$P((X_1,\ldots,X_n) \in R) = \iint_R \cdots \int f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\,dx_1\cdots dx_n$$

And if $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ is the joint *cdf* of continuous random variables $X_1,\ldots,X_n$—that is, $F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = P(X_1 \le x_1,\ldots,X_n \le x_n)$—then

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = \frac{\partial^n}{\partial X_1\cdots\partial X_n}F_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$$

The notion of a marginal pdf also extends readily, although in the n-variate case, a marginal pdf can, itself, be a joint pdf. Given $X_1,\ldots,X_n$, the marginal pdf of any subset of r of those variables $(X_{i_1}, X_{i_2},\ldots,X_{i_r})$ is derived by integrating (or summing) the joint pdf with respect to the remaining $n - r$ variables $(X_{j_1}, X_{j_2},\ldots,X_{j_{n-r}})$. If the X_i's are all continuous, for example,

$$f_{X_{i_1},\ldots,X_{i_r}}(x_{i_1},\ldots,x_{i_r}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty} f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)\,dx_{j_1}\cdots dx_{j_{n-r}}$$

QUESTIONS

3.7.25. Consider the experiment of simultaneously tossing a fair coin and rolling a fair die. Let X denote the number of heads showing on the coin and Y the number of spots showing on the die.

(a) List the outcomes in S.

(b) Find $F_{X,Y}(1,2)$.

3.7.26. An urn contains 12 chips—4 red, 3 black, and 5 white. A sample of size 4 is to be drawn without replacement. Let X denote the number of white chips in the sample; Y the number of red. Find $F_{X,Y}(1,2)$.

3.7.27. For each of the following joint pdf's, find $F_{X,Y}(u,v)$.

(a) $f_{X,Y}(x,y) = \frac{3}{2}y^2, 0 \le x \le 2, 0 \le y \le 1$

(b) $f_{X,Y}(x,y) = \frac{2}{3}(x + 2y), 0 \le x \le 1, 0 \le y \le 1$

(c) $f_{X,Y}(x,y) = 4xy, 0 \le x \le 1, 0 \le y \le 1$

3.7.28. For each of the following joint pdf's, find $F_{X,Y}(u,v)$.

(a) $f_{X,Y}(x,y) = \frac{1}{2}, 0 \le x \le y \le 2$

(b) $f_{X,Y}(x,y) = \frac{1}{x}, 0 \le y \le x \le 1$

(c) $f_{X,Y}(x,y) = 6x, 0 \le x \le 1, 0 \le y \le 1 - x$

3.7.29. Find and graph $f_{X,Y}(x,y)$ if the joint cdf for random variables X and Y is

$$F_{X,Y}(x,y) = xy, \qquad 0 < x < 1, \quad 0 < y < 1$$

3.7.30. Find the joint pdf associated with two random variables X and Y whose joint cdf is

$$F_{X,Y}(x, y) = (1 - e^{\lambda y})(1 - e^{-\lambda x}), \qquad x > 0, \quad y > 0$$

3.7.31. Given that $F_{X,Y}(x, y) = k(4x^2y^2 + 5xy^4), 0 < x < 1, 0 < y < 1$, find the corresponding pdf and use it to calculate $P(0 < X < \frac{1}{2}, \frac{1}{2} < Y < 1)$.

3.7.32. Prove that

$$P(a < X \leq b, c < Y \leq d) = F_{X,Y}(b, d) - F_{X,Y}(a, d) - F_{X,Y}(b, c) + F_{X,Y}(a, c)$$

3.7.33. A certain brand of fluorescent bulbs will last, on the average, 1000 hours. Suppose that four of these bulbs are installed in an office. What is probability that all four are still functioning after 1050 hours? If X_i denotes the ith bulb's life, assume that

$$f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4) = \prod_{i=1}^{4}\left(\frac{1}{1000}\right)e^{-x/1000}$$

for $x_i > 0, i = 1, 2, 3, 4$.

3.7.34. A hand of six cards is dealt from a standard poker deck. Let X denote the number of aces, Y the number of kings, and Z the number of queens.

(a) Write a formula for $p_{X,Y,Z}(x, y, z)$.

(b) Find $p_{X,Y}(x, y)$ and $p_{X,Z}(x, z)$.

3.7.35. Calculate $p_{X,Y}(0, 1)$ if $p_{X,Y,Z}(x, y, z) = \dfrac{3!}{x!y!z!(3-x-y-z)!}\left(\dfrac{1}{2}\right)^x\left(\dfrac{1}{12}\right)^y\left(\dfrac{1}{6}\right)^z \cdot \left(\dfrac{1}{4}\right)^{3-x-y-z}$ for $x, y, z = 0, 1, 2, 3$ and $0 \leq x + y + z \leq 3$.

3.7.36. Suppose that the random variables X, Y, and Z have the multivariate pdf

$$f_{X,Y,Z}(x, y, z) = (x + y)e^{-z}$$

for $0 < x < 1, 0 < y < 1$, and $z > 0$. Find (a) $f_{X,Y}(x, y)$, (b) $f_{Y,Z}(y, z)$, and (c) $f_Z(z)$.

3.7.37. The four random variables W, X, Y, and Z have the multivariate pdf

$$f_{W,X,Y,Z}(w, x, y, z) = 16wxyz$$

for $0 < w < 1, 0 < x < 1, 0 < y < 1$, and $0 < z < 1$. Find the marginal pdf. $f_{W,X}(w, x)$, and use it to compute $P(0 < W < \frac{1}{2}, \frac{1}{2} < X < 1)$.

Independence of Two Random Variables

The concept of independent events that was introduced in Section 2.5 leads quite naturally to a similar definition for independent random variables.

Definition 3.7.5. Two random variables X and Y are said to be *independent* if for every interval A and every interval B, $P(X \in A \text{ and } Y \in B) = P(X \in A)P(Y \in B)$.

Theorem 3.7.4. *The random variables X and Y are independent if and only if there are functions $g(x)$ and $h(y)$ such that*

$$f_{X,Y}(x, y) = g(x)h(y) \tag{3.7.2}$$

If Equation 3.7.2 holds, there is a constant k such that $f_X(x) = kg(x)$ and $f_Y(y) = (1/k)h(y)$.

Proof. We prove the theorem for the continuous case. The discrete case is similar.

First, suppose that X and Y are independent. Then $F_{X,Y}(x, y) = P(X \leq x$ and $Y \leq y) = P(X \leq x)P(Y \leq y) = F_X(x)F_Y(y)$, and we can write

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_{X,Y}(x, y) = \frac{\partial^2}{\partial x\, \partial y} F_X(x)F_Y(y) = \frac{d}{dx} F_X(x) \frac{d}{dy} F_Y(y) = f_X(x)f_Y(y)$$

Next we need to show that Equation 3.7.2 implies that X and Y are independent. To begin, note that

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y)\, dy = \int_{-\infty}^{\infty} g(x)h(y)\, dy = g(x) \int_{-\infty}^{\infty} h(y)\, dy$$

Set $k = \int_{-\infty}^{\infty} h(y)\, dy$, so $f_X(x) = kg(x)$. Similarly, it can be shown that $f_Y(y) = (1/k)h(y)$. Therefore,

$$\begin{aligned} P(X \in A \text{ and } Y \in B) &= \int_A \int_B f_{X,Y}(x, y)\, dx\, dy = \int_A \int_B g(x)h(y)\, dx\, dy \\ &= \int_A \int_B kg(x)(1/k)h(y)\, dx\, dy = \int_A f_X(x)\, dx \int_B f_Y(y)\, dy \\ &= P(X \in A)P(Y \in B) \end{aligned}$$

and the theorem is proved. □

EXAMPLE 3.7.11

Suppose that the probabilistic behavior of two random variables X and Y is described by the joint pdf $f_{X,Y}(x, y) = 12xy(1 - y), 0 \leq x \leq 1, 0 \leq y \leq 1$. Are X and Y independent? If they are, find $f_X(x)$ and $f_Y(y)$.

According to Theorem 3.7.4, the answer to the independence question is "yes" if $f_{X,Y}(x, y)$ can be factored into a function of x times a function of y. But there are such functions. Let $g(x) = 12x$ and $h(y) = y(1 - y)$.

To find $f_X(x)$ and $f_Y(y)$ requires that the "12" appearing in $f_{X,Y}(x, y)$ be factored in such a way that $g(x) \cdot h(y) = f_X(x) \cdot f_Y(y)$. Let

$$k = \int_{-\infty}^{\infty} h(y)\, dy = \int_0^1 y(1 - y)\, dy = [y^2/2 - y^3/3]\Big|_0^1 = \frac{1}{6}$$

Therefore, $f_X(x) = kg(x) = \frac{1}{6}(12x) = 2x, 0 \leq x \leq 1$ and $f_Y(y) = (1/k)h(y) = 6y(1 - y)$, $0 \leq y \leq 1$.

Independence of *n* (> 2) Random Variables

In Chapter 2, extending the notion of independence from *two* events to *n* events proved to be something of a problem. The independence of each subset of the *n* events had to be checked separately (recall Definition 2.5.2). This is not necessary in the case of n random variables. We simply use the extension of Theorem 3.7.4 to n random variables as the definition of independence in the multidimensional case. The theorem that independence is equivalent to the factorization of the joint pdf holds in the multidimensional case.

Definition 3.7.6. The n random variables $X_1, X_2, \ldots, X_n$ are said to be *independent* if there are functions $g_1(x_1), g_2(x_2), \ldots, g_n(x_n)$ such that for every $x_1, x_2, \ldots, x_n$

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = g_1(x_1)g_2(x_2)\cdots g_n(x_n)$$

A similar statement holds for discrete random variables, in which case f is replaced with p.

Comment. Analogous to the result for $n = 2$ random variables, the expression on the right-hand side of the equation in Definition 3.7.6 can also be written as the product of the marginal pdfs of $X_1, X_2, \ldots$, and X_n.

EXAMPLE 3.7.12

Consider k urns, each holding n chips, numbered 1 through n. A chip is to be drawn at random from each urn. What is the probability that all k chips will bear the same number?

If $X_1, X_2, \ldots, X_k$ denote the numbers on the 1st, 2nd, ..., and kth chips, respectively, we are looking for the probability that $X_1 = X_2 = \cdots = X_k$. In terms of the joint pdf,

$$P(X_1 = X_2 = \cdots = X_k) = \sum_{x_1=x_2=\cdots=x_k} p_{X_1,X_2,\ldots,X_k}(x_1, x_2, \ldots, x_k)$$

Each of the selections here is obviously independent of all the others so the joint pdf factors according to Definition 3.7.6, and we can write

$$\begin{aligned} P(X_1 = X_2 = \cdots = X_k) &= \sum_{i=1}^{n} p_{X_1}(x_i) \cdot p_{X_2}(x_i) \cdots p_{X_k}(x_i) \\ &= n \cdot \left(\frac{1}{n} \cdot \frac{1}{n} \cdot \cdots \cdot \frac{1}{n}\right) \\ &= \frac{1}{n^{k-1}} \end{aligned}$$

Random Samples

Definition 3.7.6 addresses the question of independence as it applies to n random variables having marginal pdfs—say, $f_{X_1}(x_1)$, $f_{X_2}(x_2), \ldots, f_{X_n}(x_n)$—that might be quite different. A special case of that definition occurs for virtually every set of data collected for statistical analysis. Suppose an experimenter takes a set of n measurements, $x_1, x_2, \ldots, x_n$, under the same conditions. Those X_i's, then, qualify as a set of independent random variables—moreover, each represents the *same* pdf. The special—but familiar—notation for that scenario is given in Definition 3.7.7. We will encounter it often in the chapters ahead.

Definition 3.7.7. Let $X_1, X_2, \ldots, X_n$ be a set of n independent random variables, all having the same pdf. Then $X_1, X_2, \ldots, X_n$ are said to be a *random sample of size n*.

QUESTIONS

3.7.38. Two fair dice are tossed. Let X denote the number appearing on the first die and Y the number on the second. Show that X and Y are independent.

3.7.39. Let $f_{X,Y}(x, y) = \lambda^2 e^{-(x+y)}, 0 \le x, 0 \le y$. Show that X and Y are independent. What are the marginal pdf's in this case?

3.7.40. Suppose that each of two urns has four chips, numbered 1 through 4. A chip is drawn from the first urn and bears the number X. That chip is added to the second urn. A chip is then drawn from the second urn. Call its number Y.

(a) Find $p_{X,Y}(x, y)$.

(b) Show that $p_X(k) = p_Y(k) = \frac{1}{4}, k = 1, 2, 3, 4$

(c) Show that X and Y are not independent

3.7.41. Let X and Y be random variables with joint pdf

$$f_{X,Y}(x, y) = k, \qquad 0 \le x \le 1, \quad 0 \le y \le 1, \quad 0 \le x + y \le 1$$

Give a geometric argument to show that X and Y are not independent.

3.7.42. Are the random variables X and Y independent if $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y), 0 \le x \le 1, 0 \le y \le 1$?

3.7.43. Suppose that random variables X and Y are independent with marginal pdfs, $f_X(x) = 2x, 0 \le x \le 1$ and $f_Y(y) = 3y^2, 0 \le y \le 1$. Find $P(Y < X)$.

3.7.44. Find the joint cdf of the independent random variables X and Y, where $f_X(x) = \frac{x}{2}$, $0 \le x \le 2$ and $f_Y(y) = 2y, 0 \le y \le 1$.

3.7.45. If two random variables X and Y are independent with marginal pdf's $f_X(x) = 2x$, $0 \le x \le 1$ and $f_Y(y) = 1, 0 \le y \le 1$, calculate $P\left(\frac{Y}{X} > 2\right)$.

3.7.46. Suppose $f_{X,Y}(x, y) = xye^{-(x+y)}, x > 0, y > 0$. Prove for any real numbers a, b, c, and d that

$$P(a < X < b, c < Y < d) = P(a < X < b) \cdot P(c < Y < d)$$

thereby establishing the independence of X and Y.

3.7.47. Given the joint pdf $f_{X,Y}(x, y) = 2x + y - 2xy, 0 < x < 1, 0 < y < 1$, find numbers a, b, c, and d such that

$$P(a < X < b, c < Y < d) \neq P(a < X < b) \cdot P(c < Y < d)$$

thus demonstrating that X and Y are not independent.

3.7.48. Prove that if X and Y are two independent random variables, then $U = g(X)$ and $V = h(Y)$ are also independent.

3.7.49. If two random variables X and Y are defined over a region in the XY-plane that is *not* a rectangle (possibly infinite) with sides parallel to the coordinate axes, can X and Y be independent?

3.7.50. Write down the joint probability density function for a random sample of size n drawn from the exponential pdf, $f_X(x) = (1/\lambda)e^{-x/\lambda}, x > 0$.

3.7.51. Suppose that X_1, X_2, X_3, and X_4 are independent random variables, each with pdf $f_{X_i}(x_i) = 4x_i^3, 0 \leq x_i \leq 1$. Find

(a) $P\left(X_1 < \frac{1}{2}\right)$

(b) $P\left(\text{exactly one } X_i < \frac{1}{2}\right)$

(c) $f_{X_1,X_2,X_3,X_4}(x_1, x_2, x_3, x_4)$

(d) $F_{X_2,X_3}(x_2, x_3)$

3.7.52. A random sample of size $n = 2k$ is taken from a uniform pdf defined over the unit interval. Calculate $P\left(X_1 < \frac{1}{2}, X_2 > \frac{1}{2}, X_3 < \frac{1}{2}, X_4 > \frac{1}{2}, \ldots, X_{2k} > \frac{1}{2}\right)$.

3.8 COMBINING RANDOM VARIABLES

In Section 3.4, we derived a linear transformation frequently applied to single random variables—$Y = a + bX$. Now, armed with the multivariable concepts and techniques covered in Section 3.7, we can extend the investigation of transformations to functions defined on sets of random variables. In statistics, the most important combination of a set of random variables is often their sum, so we begin this section with the problem of finding the pdf of $X + Y$.

Finding the pdf of a Sum

Theorem 3.8.1. *Suppose that X and Y are independent random variables. Let* $W = X + Y$. *Then*

1. *If X and Y are discrete random variables with pdfs* $p_X(x)$ *and* $p_Y(y)$, *respectively,*

$$p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w - x)$$

2. *If X and Y are continuous random variables with pdfs* $f_X(x)$ *and* $f_Y(y)$, *respectively,*

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x)f_Y(w - x)\,dx$$

Proof.

1. $p_W(w) = P(W = w) = P(X + Y = W)$

$$= \bigcup_{\text{all } x} P(X = x, Y = w - x) = \sum_{\text{all } x} P(X = x, Y = w - x)$$

$$= \sum_{\text{all } x} P(X = x)P(Y = w - x)$$

$$= \sum_{\text{all } x} p_X(x)p_Y(w - x)$$

where the next-to-last equality derives from the independence of X and Y.

2. Since X and Y are continuous random variables, we can find $f_W(w)$ by differentiating the corresponding cdf, $F_W(w)$. Here, $F_W(w) = P(X + Y \leq w)$ is found by integrating $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ over the shaded region R pictured in Figure 3.8.1.

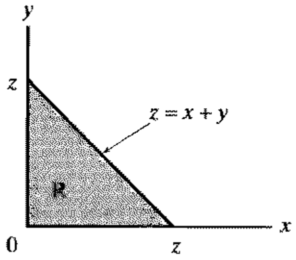


FIGURE 3.8.1

By inspection,

$$F_w(w) = \int_{-\infty}^{\infty} \int_{-\infty}^{w-x} f_X(x) f_Y(y)\, dy\, dx = \int_{-\infty}^{\infty} f_X(x) \left(\int_{-\infty}^{w-x} f_Y(y)\, dy \right) dx$$
$$= \int_{-\infty}^{\infty} f_X(x) F_Y(w - x)\, dx$$

Assume that the integrand in the above equation is sufficiently smooth so that differentiation and integration can be interchanged. Then we can write

$$f_W(w) = \frac{d}{dw} F_W(w) = \frac{d}{dw} \int_{-\infty}^{\infty} f_X(x) F_Y(w - x)\, dx = \int_{-\infty}^{\infty} f_X(x) \left(\frac{d}{dw} F_Y(w - x) \right) dx$$
$$= \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\, dx$$

and the theorem is proved. □

Comment. The integral in part (2) above is referred to as the *convolution* of the functions f_X and f_Y. Besides their frequent appearances in random-variable problems, convolutions turn up in many areas of mathematics and engineering.

EXAMPLE 3.8.1

Suppose that X and Y are two independent binomial random variables, each with the same success probability but defined on m and n trials, respectively. Specifically,

$$p_X(k) = \binom{m}{k} p^k (1 - p)^{m-k}, \quad k = 0, 1, \ldots, m$$

and

$$p_Y(k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Find $p_W(w)$, where $W = X + Y$.

By Theorem 3.8.1, $p_W(w) = \sum_{\text{all } x} p_X(x)p_Y(w - x)$, but the summation over "all x" needs to be interpreted as the set of values for x and $w - x$ such that $p_X(x)$ and $p_Y(w - x)$, respectively, are both nonzero. But that will be true for all integers x from 0 to w. Therefore,

$$p_W(w) = \sum_{x=0}^{w} p_X(x)p_Y(w - x) = \sum_{x=0}^{w} \binom{m}{x} p^x (1 - p)^{m-x} \binom{n}{w - x} p^{w-x}(1 - p)^{n-(w-x)}$$

$$= \sum_{x=0}^{w} \binom{m}{x}\binom{n}{w - x} p^w (1 - p)^{n+m-w}$$

Now, consider an urn having m red chips and n white chips. If w chips are drawn out—without replacement—the probability that exactly x red chips are in the sample is given by the hypergeometric distribution,

$$P(x \text{ reds in sample}) = \frac{\binom{m}{x}\binom{n}{w - x}}{\binom{m + n}{w}} \tag{3.8.1}$$

Summing Equation 3.8.1 from $x = 0$ to $x = w$ must equal one (why?), in which case

$$\sum_{x=0}^{w} \binom{m}{x}\binom{n}{w - x} = \binom{m + n}{w}$$

so

$$p_W(w) = \binom{m + n}{w} p^w (1 - p)^{n+m-w}, \quad w = 0, 1, \ldots, n + m$$

Should we recognize $p_W(w)$? Definitely. Compare the structure of $p_W(w)$ to the statement of Theorem 3.2.1: The random variable W has a binomial distribution where the probability of success at any given trial is p and the total number of trials is $n + m$.

Comment. Example 3.8.1 shows that the binomial distribution "reproduces" itself—that is, if X and Y are independent binomial random variables with the same value for p, their sum is also a binomial random variable. Not all random variables share that property. The sum of two independent uniform random variables, for example, is not a uniform random variable (see Question 3.8.3).

EXAMPLE 3.8.2

Suppose a radiation monitor relies on an electronic sensor, whose lifetime X is modeled by the exponential pdf $f_X(x) = \lambda e^{-\lambda x}, x > 0$. To improve the reliability of the monitor, the manufacturer has included an identical second sensor that is activated only in the event the first sensor malfunctions. (This is called *cold redundancy*.) Let the random variable Y denote the operating lifetime of the second sensor, in which case the lifetime of the monitor can be written as the sum $W = X + Y$. Find $f_W(w)$.

Since X and Y are both continuous random variables,

$$f_W(w) = \int_{-\infty}^{\infty} f_X(x) f_Y(w - x)\,dx \tag{3.8.2}$$

Notice that $f_X(x) > 0$ only if $x > 0$ and $f_Y(w - x) > 0$ only if $x < w$. Therefore, the integral in Equation 3.8.2 that goes from $-\infty$ to ∞ reduces to an integral from 0 to w, and we can write

$$f_W(w) = \int_0^w f_X(x) f_Y(w - x)\,dx = \int_0^w \lambda e^{-\lambda x} \lambda e^{-\lambda(w-x)}\,dx = \lambda^2 \int_0^w e^{-\lambda x} e^{-\lambda(w-x)}\,dx$$
$$= \lambda^2 e^{-\lambda w} \int_0^w dx = \lambda^2 w e^{-\lambda w}, \quad w \geq 0$$

Comment. By integrating $f_X(x)$ and $f_W(w)$, we can assess the improvement in the monitor's reliability afforded by the cold redundancy. Since X is an exponential random variable, $E(X) = 1/\lambda$ (recall Example 3.5.6). How different, for example, are $P(X \geq 1/\lambda)$ and $P(W \geq 1/\lambda)$? A simple calculation shows that the latter is actually *twice* the magnitude of the former:

$$P(X \geq 1/\lambda) = \int_{1/\lambda}^{\infty} \lambda e^{-\lambda x}\,dx = -e^{-u}\Big|_1^{\infty} = e^{-1} = 0.37$$

$$P(W \geq 1/\lambda) = \int_{1/\lambda}^{\infty} \lambda^2 w e^{-\lambda w}\,dw = e^{-u}(-u - 1)\Big|_1^{\infty} = 2e^{-1} = 0.74$$

Finding the pdfs of Quotients and Products

We conclude this section by considering the pdfs for the quotient and product of two independent random variables. That is, given X and Y, we are looking for $f_W(w)$, where 1) $W = Y/X$ and 2) $W = XY$. Neither of the resulting formulas is as important as the pdf for the *sum* of two random variables, but both formulas will play key roles in several derivations in Chapter 7.

Theorem 3.8.2. *Let X and Y be independent continuous random variables, with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Assume that X is zero for at most a set of isolated points. Let $W = Y/X$. Then*

$$f_W(w) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(wx)\,dx$$

Proof.

$$
\begin{aligned}
F_W(w) &= P(Y/X \le w)\\
&= P(Y/X \le w \quad \text{and} \quad X \ge 0) + P(Y/X \ge w \quad \text{and} \quad X < 0)\\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + P(Y \ge wX \quad \text{and} \quad X < 0)\\
&= P(Y \le wX \quad \text{and} \quad X \ge 0) + 1 - P(Y \le wX \quad \text{and} \quad X < 0)\\
&= \int_0^\infty \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx + 1 - \int_{-\infty}^0 \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx
\end{aligned}
$$

Then differentiate $F_W(w)$ to obtain

$$
\begin{aligned}
f_W(w) = \frac{d}{dw} F_W(w) &= \frac{d}{dw} \int_0^\infty \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx - \frac{d}{dw} \int_{-\infty}^0 \int_{-\infty}^{wx} f_X(x) f_Y(y)\, dy\, dx\\
&= \int_0^\infty f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\, dy \right) dx - \int_{-\infty}^0 f_X(x) \left(\frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\, dy \right) dx
\end{aligned}
\tag{3.8.3}
$$

(Note that we are assuming sufficient regularity of the functions to permit interchange of integration and differentiation.)

To proceed, we need to differentiate the function $G(w) = \int_{-\infty}^{wx} f_Y(y)\, dy$ with respect to w. By the Fundamental Theorem of Calculus and the chain rule, we find

$$
\frac{d}{dw} G(w) = \frac{d}{dw} \int_{-\infty}^{wx} f_Y(y)\, dy = f_Y(wx) \frac{d}{dw} wx = x f_Y(wx)
$$

Putting this result into Equation 3.8.3 gives

$$
\begin{aligned}
f_W(w) &= \int_0^\infty x f_X(x) f_Y(wx)\, dx - \int_{-\infty}^0 x f_X(x) f_Y(wx)\, dx\\
&= \int_0^\infty x f_X(x) f_Y(wx)\, dx + \int_{-\infty}^0 (-x) f_X(x) f_Y(wx)\, dx\\
&= \int_0^\infty |x| f_X(x) f_Y(wx)\, dx + \int_{-\infty}^0 |x| f_X(x) f_Y(wx)\, dx\\
&= \int_{-\infty}^\infty |x| f_X(x) f_Y(wx)\, dx
\end{aligned}
$$

which completes the proof. □

EXAMPLE 3.8.3

Let X and Y be independent random variables with pdfs $f_X(x) = \lambda e^{-\lambda x}$, $x > 0$ and $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Define $W = Y/X$. Find $f_W(w)$.

Substituting into the formula given in Theorem 3.8.2, we can write

$$f_W(w) = \int_0^{\infty} x(\lambda e^{-\lambda x})(\lambda e^{-\lambda x w})\,dx = \lambda^2 \int_0^{\infty} x e^{-\lambda(1+w)x}\,dx$$

$$= \frac{\lambda^2}{\lambda(1 + w)} \int_0^{\infty} x\lambda(1 + w)e^{-\lambda(1+w)x}\,dx$$

Notice that the integral is the expected value of an exponential random variable with parameter $\lambda(1 + w)$, so it equals $1/\lambda(1 + w)$ (recall Example 3.5.6). Therefore,

$$f_W(w) = \frac{\lambda^2}{\lambda(1 + w)} \frac{1}{\lambda(1 + w)} = \frac{1}{(1 + w)^2}, \qquad w \geq 0.$$

Theorem 3.8.3. *Let X and Y be independent continuous random variables with pdfs $f_X(x)$ and $f_Y(y)$, respectively. Let $W = XY$. Then*

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(w/x) f_Y(x)\,dx$$

Proof. A line-by-line straightforward modification of the proof of Theorem 3.8.2 will provide a proof of Theorem 3.8.3. The details are left to the reader. □

EXAMPLE 3.8.4

Suppose that X and Y are independent random variables with pdfs $f_X(x) = 1, 0 \leq x \leq 1$ and $f_Y(y) = 2y, 0 \leq y \leq 1$, respectively. Find $f_W(w)$, where $W = XY$.

According to Theorem 3.8.3,

$$f_W(w) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(w/x) f_Y(x)\,dx$$

The region of integration, though, needs to be restricted to values of x for which the integrand is positive. But $f_X(w/x)$ is positive only if $0 \leq w/x \leq 1$, which implies that $x \geq w$. Moreover, for $f_Y(x)$ to be positive requires that $0 \leq x \leq 1$. Any x, then, from w to 1 will yield a positive integrand. Therefore,

$$f_W(w) = \int_w^1 \frac{1}{x}(1)(2x)\,dx = \int_w^1 2\,dx = 2 - 2w, \qquad 0 \leq w \leq 1$$

Comment. Theorems 3.8.1, 3.8.2, and 3.8.3 can be adapted to situations where X and Y are not independent by replacing the product of the marginal pdfs with the joint pdf.

QUESTIONS

3.8.1. Let X and Y be two independent random variables. Given the marginal pdfs shown below, find the pdf of $X + Y$. In each case, check to see if $X + Y$ belongs to the same family of pdfs as do X and Y.

(a) $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ and $p_Y(k) = e^{-\mu}\frac{\mu^k}{k!}$, $k = 0, 1, 2, \ldots$

(b) $p_X(k) = p_Y(k) = (1 - p)^{k-1}p$, $k = 1, 2, \ldots$

3.8.2. Suppose $f_X(x) = xe^{-x}$, $x \geq 0$, and $f_Y(y) = e^{-y}$, $y \geq 0$, where X and Y are independent. Find the pdf of $X + Y$.

3.8.3. Let X and Y be two independent random variables, whose marginal pdfs are given below. Find the pdf of $X + Y$. *Hint:* Consider two cases, $0 \leq w < 1$ and $1 \leq w \leq 2$. $f_X(x) = 1, 0 \leq x \leq 1$, and $f_Y(y) = 1, 0 \leq y \leq 1$

3.8.4. If a random variable V is independent of two independent random variables X and Y, prove that V is independent of $X + Y$.

3.8.5. Let Y be a uniform random variable over the interval $[0, 1]$. Find the pdf of $W = Y^2$. *Hint:* First find $F_W(w)$.

3.8.6. Let Y be a random variable with $f_Y(y) = 6y(1 - y), 0 \leq y \leq 1$. Find the pdf of $W = Y^2$.

3.8.7. Given that X and Y are independent random variables, find the pdf of XY for the following two sets of marginal pdfs:

(a) $f_X(x) = 1, 0 \leq x \leq 1$, and $f_Y(y) = 1, 0 \leq y \leq 1$

(b) $f_X(x) = 2x, 0 \leq x \leq 1$, and $f_Y(y) = 2y, 0 \leq y \leq 1$

3.8.8. Let X and Y be two independent random variables. Given the marginal pdfs indicated below, find the cdf of Y/X. *Hint:* Consider two cases, $0 \leq w \leq 1$ and $1 < w$.

(a) $f_X(x) = 1, 0 \leq x \leq 1$, and $f_Y(y) = 1, 0 \leq y \leq 1$.

(b) $f_X(x) = 2x, 0 \leq x \leq 1$, and $f_Y(y) = 2y, 0 \leq y \leq 1$

3.8.9. Suppose that X and Y are two independent random variables, where $f_X(x) = xe^{-x}$, $x \geq 0$ and $f_Y(y) = e^{-y}$, $y \geq 0$. Find the pdf of Y/X.

3.9 FURTHER PROPERTIES OF THE MEAN AND VARIANCE

Sections 3.5 and 3.6 introduced the basic definitions related to the expected value and variance of *single* random variables. We learned how to calculate $E(W)$, $E[g(W)]$, $E(aW + b)$, $\text{Var}(W)$, and $\text{Var}(aW + b)$, where a and b are any constants and W could be either a discrete or a continuous random variable. The purpose of this section is to examine certain multivariable extensions of those results, based on the joint pdf material covered in Section 3.7.

We begin with a theorem that generalizes $E[g(W)]$. While it is stated here for the case of *two* random variables, it extends in a very straightforward way to include functions of n random variables.

Theorem 3.9.1.

1. *Suppose X and Y are discrete random variables with joint pdf $p_{X,Y}(x, y)$, and let $g(X, Y)$ be a function of X and Y. Then the expected value of the random variable $g(X, Y)$ is given by*

$$E[g(X, Y)] = \sum_{\text{all } x}\sum_{\text{all } y} g(x, y) \cdot p_{X,Y}(x, y)$$

provided $\sum_{\text{all } x}\sum_{\text{all } y} |g(x, y)| \cdot p_{X,Y}(x, y) < \infty$.

2. *Suppose X and Y are continuous random variables with joint pdf $f_{X,Y}(x, y)$, and let $g(X, Y)$ be a continuous function. Then the expected value of the random variable $g(X, Y)$ is given by*

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y)\, dx\, dy$$

provided $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| \cdot f_{X,Y}(x, y)\, dxdy < \infty$

Proof. The basic approach taken in deriving this result is similar to the method followed in the proof of Theorem 3.5.3. See (134) for details. □

EXAMPLE 3.9.1

Consider the two random variables X and Y whose joint pdf is detailed in the 2×4 matrix shown in Table 3.9.1. Let

$$g(X, Y) = 3X - 2XY + Y$$

Find $E[g(X, Y)]$ two ways—first, by using the basic definition of an expected value, and secondly, by using Theorem 3.9.1.

TABLE 3.9.1

$$\begin{array}{cc|cccc|} & & & Y & & \\ & & 0 & 1 & 2 & 3 \\ \hline & 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & 0 \\ X & & & & & \\ & 1 & 0 & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} \\ \hline \end{array}$$

TABLE 3.9.2

z	0	1	2	3
$f_Z(z)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	0

Let $Z = 3X - 2XY + Y$. By inspection, Z takes on the values 0, 1, 2, and 3 according to the pdf $f_Z(z)$ shown in Table 3.9.2. From the basic definition, then, that an expected value is a weighted average, we see that $E[g(X, Y)]$ is equal to one:

$$\begin{aligned} E[g(X, Y)] = E(Z) &= \sum_{\text{all } z} z \cdot f_Z(z) \\ &= 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 3 \cdot 0 \\ &= 1 \end{aligned}$$

The same answer is obtained by applying Theorem 3.9.1 to the joint pdf given in Figure 3.9.1:

$$E[g(X, Y)] = 0 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{8} + 3 \cdot 0 + 3 \cdot 0 + 2 \cdot \frac{1}{8} + 1 \cdot \frac{1}{4} + 0 \cdot \frac{1}{8}$$
$$= 1$$

The advantage, of course, enjoyed by the latter solution is that we avoid the intermediate step of having to determine $f_Z(z)$.

EXAMPLE 3.9.2

An electrical circuit has three resistors, R_X, R_Y, and R_Z, wired in parallel (see Figure 3.9.1). The nominal resistance of each is fifteen ohms, but their *actual* resistances, X, Y, and Z, vary between ten and twenty according to the joint pdf,

$$f_{X,Y,X}(x, y, z) = \frac{1}{675{,}000}(xy + xz + yz), \quad \begin{array}{l} 10 \le x \le 20 \\ 10 \le y \le 20 \\ 10 \le z \le 20 \end{array}$$

What is the expected resistance for the circuit?

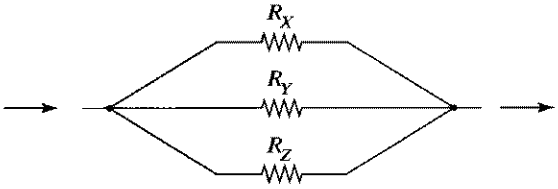


FIGURE 3.9.1

Let R denote the circuit's resistance. A well-known result in physics holds that

$$\frac{1}{R} = \frac{1}{X} + \frac{1}{Y} + \frac{1}{Z}$$

or, equivalently,

$$R = \frac{XYZ}{XY + XZ + YZ} = R(X, Y, Z)$$

Integrating $R(x, y, z) \cdot f_{X,Y,Z}(x, y, z)$ shows that the expected resistance is five:

$$E(R) = \int_{10}^{20}\int_{10}^{20}\int_{10}^{20} \frac{xyz}{xy + xz + yz} \cdot \frac{1}{675{,}000}(xy + xz + yz)\,dx\,dy\,dz$$
$$= \frac{1}{675{,}000}\int_{10}^{20}\int_{10}^{20}\int_{10}^{20} xyz\,dx\,dy\,dz$$
$$= 5.0$$

Theorem 3.9.2. *Let X and Y be any two random variables (discrete or continuous dependent or independent), and let a and b be any two constants. Then*

$$E(aX + bY) = aE(X) + bE(Y)$$

provided $E(X)$ and $E(Y)$ are both finite.

Proof. Consider the continuous case (the discrete case is proved much the same way). Let $f_{X,Y}(x, y)$ be the joint pdf of X and Y, and define $g(X, Y) = aX + bY$. By Theorem 3.9.1,

$$\begin{aligned}
E(aX + bY) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax + by) f_{X,Y}(x, y)\,dx\,dy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (ax) f_{X,Y}(x, y)\,dx\,dy + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (by) f_{X,Y}(x, y)\,dx\,dy \\
&= a\int_{-\infty}^{\infty} x\left(\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dy\right)dx + b\int_{-\infty}^{\infty} y\left(\int_{-\infty}^{\infty} f_{X,Y}(x, y)\,dx\right)dy \\
&= a\int_{-\infty}^{\infty} x f_X(x)\,dx + b\int_{-\infty}^{\infty} y f_Y(y)\,dy \\
&= aE(X) + bE(Y)
\end{aligned}$$

□

Corollary. *Let $W_1, W_2, \ldots, W_n$ be any random variables for which $E(W_i) < \infty$, $i = 1, 2, \ldots, n$, and let $a_1, a_2, \ldots, a_n$ be any set of constants. Then*

$$E(a_1W_1 + a_2W_2 + \cdots + a_nW_n) = a_1E(W_1) + a_2E(W_2) + \cdots + a_nE(W_n)$$

EXAMPLE 3.9.3

Let X be a binomial random variable defined on n independent trials, each trial resulting in success with probability p. Find $E(X)$.

Note, first, that X can be thought of as a sum, $X = X_1 + X_2 + \cdots + X_n$, where X_i represents the number of successes occurring at the ith trial:

$$X_i = \begin{cases} 1 & \text{if the } i\text{th trial produces a success} \\ 0 & \text{if the } i\text{th trial produces a failure} \end{cases}$$

(Any X_i defined in this way on an individual trial is called a *Bernoulli* random variable. Every binomial, then, can be thought of as the sum of n independent Bernoullis.) By assumption, $p_{X_i}(1) = p$ and $p_{X_i}(0) = 1 - p, i = 1, 2, \ldots, n$. Using the corollary,

$$\begin{aligned}
E(X) &= E(X_1) + E(X_2) + \cdots + E(X_n) \\
&= n \cdot E(X_1)
\end{aligned}$$

the last step being a consequence of the X_i's having identical distributions. But

$$E(X_1) = 1 \cdot p + 0 \cdot (1 - p) = p$$

so $E(X) = np$, which is what we found before (recall Theorem 3.5.1).

Comment. The problem-solving implications of Theorem 3.9.2 and its corollary should not be underestimated. There are many real-world events that can be modeled as a linear combination $a_1 W_1 + a_2 W_2 + \cdots + a_n W_n$, where the W_is are relatively simple random variables. Finding $E(a_1 W_1 + a_2 W_2 + \cdots + a_n W_n)$ *directly* may be prohibitively difficult because of the inherent complexity of the linear combination. It may very well be the case, though, that calculating the individual $E(W_i)$'s is easy. Compare, for instance, Example 3.9.1 with Theorem 3.5.1. Both derive the formula that $E(X) = np$ when X is a binomial random variable. The approach taken in Example 3.9.1 (i.e., using Theorem 3.9.2) is *much* easier. The next several examples further explore the technique of using linear combinations to facilitate the calculation of expected values.

EXAMPLE 3.9.4

A disgruntled secretary is upset about having to stuff envelopes. Handed a box of n letters and n envelopes, she vents her frustration by putting the letters into the envelopes *at random*. How many people, on the average, will receive their correct mail?

If X denotes the number of envelopes properly stuffed, what we want is $E(X)$. However, applying Definition 3.5.1 here would prove formidable because of the difficulty in getting a workable expression for $p_X(k)$ [see (97)]. By using the corollary to Theorem 3.9.2, though, we can solve the problem quite easily.

Let X_i denote a random variable equal to the number of correct letters put into the ith envelope, $i = 1, 2, \ldots, n$. Then X_i equals 0 or 1, and

$$p_{X_i}(k) = P(X_i = k) = \begin{cases} \dfrac{1}{n} & \text{for } k = 1 \\ \dfrac{n-1}{n} & \text{for } k = 0 \end{cases}$$

But $X = X_1 + X_2 + \cdots + X_n$ and $E(X) = E(X_1) + E(X_2) + \cdots + E(X_n)$.) Furthermore, each of the X_i's has the same expected value, $1/n$:

$$E(X_i) = \sum_{k=0}^{1} k \cdot P(X_i = k) = 0 \cdot \frac{n-1}{n} + 1 \cdot \frac{1}{n} = \frac{1}{n}$$

It follows that

$$E(X) = \sum_{i=1}^{n} E(X_i) = n \cdot \left(\frac{1}{n}\right)$$
$$= 1$$

showing that, *regardless of n,* the expected number of properly stuffed envelopes is one. (Are the X_i's independent? Does it matter?)

EXAMPLE 3.9.5

Ten fair dice are rolled. Calculate the expected value of the sum of the faces showing.

If the random variable X denotes the sum of the faces showing on the ten dice, then

$$X = X_1 + X_2 + \cdots + X_{10}$$

where X_i is the number showing on the ith die, $i = 1, 2, \dots, 10$. By assumption, $p_{X_i}(k) = \frac{1}{6}$ for $k = 1, 2, 3, 4, 5, 6$, so $E(X_i) = \sum_{k=1}^{6} k \cdot \frac{1}{6} = \frac{1}{6}\sum_{k=1}^{6} k = \frac{1}{6} \cdot \frac{6(7)}{2} = 3.5$. By the corollary to Theorem 3.9.2,

$$\begin{aligned} E(X) &= E(X_1) + E(X_2) + \cdots + E(X_{10}) \\ &= 10(3.5) \\ &= 35 \end{aligned}$$

Notice that $E(X)$ can also be deduced here by appealing to the notion that expected values are centers of gravity. It should be clear from our work with combinatorics that $P(X = 10) = P(X = 60)$, $P(X = 11) = P(X = 59)$, $P(X = 12) = P(X = 58)$, and so on. The probability function $p_X(k)$ is symmetric, in other words, which implies that its center of gravity is the midpoint of the range of its X-values. It must be the case, then, that $E(X)$ equals $\frac{10+60}{2}$ or 35.

EXAMPLE 3.9.6

The honor count in a (thirteen-card) bridge hand can vary from zero to thirty-seven according to the formula:

$$\begin{aligned} \text{honor count} = {} & 4 \cdot (\text{number of aces}) + 3 \cdot (\text{number of kings}) + 2 \cdot (\text{number of queens}) \\ & + 1 \cdot (\text{number of jacks}) \end{aligned}$$

What is the expected honor count of North's hand?

The solution here is a bit unusual in that we use the corollary to Theorem 3.9.2 *backwards*. If X_i, $i = 1, 2, 3, 4$, denotes the honor count for players North, South, East, and West, respectively, and if X denotes the analogous sum for the entire deck, we can write

$$X = X_1 + X_2 + X_3 + X_4$$

But

$$X = E(X) = 4 \cdot 4 + 3 \cdot 4 + 2 \cdot 4 + 1 \cdot 4 = 40$$

By symmetry, $E(X_i) = E(X_j)$, $i \neq j$, so it follows that $40 = 4 \cdot E(X_1)$, which implies that *ten* is the expected honor count of North's hand. (Try doing this problem directly, without making use of the fact that the deck's honor count is forty.)

EXAMPLE 3.9.7

Suppose that a random sequence of 1s and 0s is generated by a computer, where the length of the sequence is n, and

$$p = p(1 \text{ appears in } i\text{th position})$$

and

$$1 - p = p(0 \text{ appears in } i\text{th position}), \qquad i = 1, 2, \ldots, n$$

What is the expected number of *runs* in the sequence? *Note:* A run is a series of consecutive similar outcomes. For example, the sequence

$$1 \quad 1 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0$$

has a total of four runs (1 1, 0, 1, and 0 0 0).

Let X_i denote the outcome appearing in position i, $i = 1, 2, \ldots, n$. The number of runs in the sequence, then, can be expressed in terms of the $n - 1$ transitions from X_i to X_{i+1}, $i = 1, 2, \ldots, n - 1$. Specifically, let

$$Q(X_i, X_{i+1}) = \begin{cases} 0 & \text{if } X_i = X_{i+1} \\ 1 & \text{if } X_i \neq X_{i+1} \end{cases}$$

It follows that

$$R = \text{total number of runs} = 1 + Q(X_1, X_2) + Q(X_2, X_3) + \cdots + Q(X_{n-1}, X_n)$$

and

$$E(R) = 1 + \sum_{i=1}^{n-1} E[Q(X_i, X_{i+1})]$$

But

$$\begin{aligned} E[Q(X_i, X_{i+1})] &= 0 \cdot P(X_i = X_{i+1}) + 1 \cdot P(X_i \neq X_{i+1}) \\ &= P(X_i \neq X_{i+1}) \\ &= P(X_i = 1 \cap X_{i+1} = 0) + P(X_i = 0 \cap X_{i+1} = 1) \\ &= p(1 - p) + (1 - p)p \qquad \text{(because of independence)} \\ &= 2p(1 - p) \end{aligned}$$

Therefore,

$$E(R) = 1 + 2(n - 1)p(1 - p)$$

Expected Values of Products: A Special Case

We know from Theorem 3.9.1 that for any two random variables X and Y,

$$E(XY) = \begin{cases} \displaystyle\sum_{\text{all } x} \sum_{\text{all } y} xyp_{X,Y}(x, y) & \text{if } X \text{ and } Y \text{ are discrete} \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyf_{X,Y}(x, y)\, dx\, dy & \text{if } X \text{ and } Y \text{ are continuous} \end{cases}$$

If, however, X and Y are independent, there is an easier way to calculate $E(XY)$.

Theorem 3.9.3. *If X and Y are independent random variables,*

$$E(XY) = E(X) \cdot E(Y)$$

provided $E(X)$ and $E(Y)$ both exist.

Proof. Suppose X and Y are both discrete random variables. Then their joint pdf, $p_{X,Y}(x, y)$, can be replaced by the product of their marginal pdfs, $p_X(x) \cdot p_Y(y)$, and the double summation required by Theorem 3.9.1 can be written as the product of two single summations:

$$\begin{aligned} E(XY) &= \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_{X,Y}(x, y) \\ &= \sum_{\text{all } x} \sum_{\text{all } y} xy \cdot p_X(x) \cdot p_Y(y) \\ &= \sum_{\text{all } x} x \cdot p_X(x) \cdot \left[\sum_{\text{all } y} y \cdot p_Y(y) \right] \\ &= E(X) \cdot E(Y) \end{aligned}$$

The proof when X and Y are both continuous random variables is left as an exercise. □

QUESTIONS

3.9.1. Suppose that r chips are drawn with replacement from an urn containing n chips, numbered 1 through n. Let V denote the sum of the numbers drawn. Find $E(V)$.

3.9.2. Suppose that $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda(x+y)}, 0 \leq x, 0 \leq y$. Find $E(X + Y)$.

3.9.3. Suppose that $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y), 0 \leq x \leq 1, 0 \leq y \leq 1$. (recall Question 3.7.19(c)). Find $E(X + Y)$.

3.9.4. Marksmanship competition at a certain level requires each contestant to take 10 shots with each of two different hand guns. Final scores are computed by taking a weighted average of four times the number of bull's-eyes made with the first gun plus six times the number gotten with the second. If Cathie has a 30% chance of hitting the bull's-eye with each shot from the first gun and a 40% chance with each shot from the second gun, what is her expected score?

3.9.5. Suppose that X_i is a random variable for which $E(X_i) = \mu, i = 1, 2, \ldots, n$. Under what conditions will the following be true?

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \mu$$

3.9.6. Suppose that the daily closing price of stock goes up an eighth of a point with probability p and down an eighth of a point with probability q, where $p > q$. After n days how much gain can we expect the stock to have achieved? Assume that the daily price fluctuations are independent events.

3.9.7. An urn contains r red balls and w white balls. A sample of n balls is drawn *in order* and *without* replacement. Let X_i be 1 if the ith draw is red and 0 otherwise $i = 1, 2, \ldots, n$.

(a) Show that $E(X_i) = E(X_1), i = 2, 3, \ldots, n$

(b) Use the Corollary to Theorem 3.9.2 to show that the expected number of red balls is $nr/(r + w)$.

3.9.8. Suppose two fair dice are tossed. Find the expected value of the product of the faces showing.

3.9.9. Find $E(R)$ for a two-resistor circuit similar to the one described in Example 3.9.2, where $f_{X,Y}(x, y) = k(x + y), 10 \leq x \leq 20, 10 \leq y \leq 20$.

3.9.10. Suppose that X and Y are both uniformly distributed over the interval $[0, 1]$. Calculate the expected value of the square of the distance of the random point (X, Y) from the origin; that is, find $E(X^2 + Y^2)$. *Hint:* See Question 3.8.5.

3.9.11. Suppose X represents a point picked at random from the interval $[0, 1]$ on the x-axis, and Y is a point picked at random from the interval $[0, 1]$ on the y-axis. Assume that X and Y are independent. What is the expected value of the area of the triangle formed by the points $(X, 0)$, $(0, Y)$ and $(0,0)$?

3.9.12. Suppose $Y_1, Y_2, \ldots, Y_n$ is a random sample from the uniform pdf over $[0, 1]$. The geometric mean of the numbers is the random variable $\sqrt[n]{Y_1 Y_2 \cdot \cdots \cdot Y_n}$. Compare the expected value of the geometric mean to that of the arithmetic mean $\bar{Y}$.

Calculating the Variance of a Sum of Random Variables

We know from the corollary to Theorem 3.9.2 that

$$E(W_1 + W_2 + \cdots + W_n) = E(W_1) + E(W_2) + \cdots + E(W_n)$$

for *any* set of random variables $W_1, W_2, \ldots, W_n$, provided $E(W_i)$ exists for all i. A similar result holds for the *variance* of a sum of random variables, *but only if the random variables are independent.*

Theorem 3.9.4. *Let $W_1, W_2, \ldots, W_n$ be a set of independent random variables for which $E(W_i^2)$ is finite for all i. Then*

$$\text{Var}(W_1 + W_2 + \cdots + W_n) = \text{Var}(W_1) + \text{Var}(W_2) + \cdots + \text{Var}(W_n)$$

Proof. The derivation is given for a sum of two random variables, $W_1 + W_2$. A simple induction argument would complete the proof for arbitrary n. From Theorems 3.6.1 and 3.9.2,

$$\text{Var}(W_1 + W_2) = E((W_1 + W_2)^2) - [E(W_1) + E(W_2)]^2$$

Writing out the squares gives

$$\begin{aligned}\text{Var}(W_1 + W_2) &= E\left(W_1^2 + 2W_1W_2 + W_2^2\right) - [E(W_1)]^2 - 2E(W_1)E(W_2) - [E(W_2)]^2 \\ &= E\left(W_1^2\right) - [E(W_1)]^2 + E\left(W_2^2\right) - [E(W_2)]^2 \\ &\quad + 2[E(W_1W_2) - E(W_1)E(W_2)] \qquad (3.9.1)\end{aligned}$$

By the independence of W_1 and W_2, $E(W_1W_2) = E(W_1)E(W_2)$, making the last term in Equation 3.9.1 vanish. The remaining terms combine to give the desired result: $\text{Var}(W_1 + W_2) = \text{Var}(W_1) + \text{Var}(W_2)$. ☐

Corollary. *Let $W_1, W_2, \ldots, W_n$ be any set of independent random variables for which $E(W_i^2) < \infty$ for all i. Let $a_1, a_2, \ldots, a_n$ be any set of constants. Then*

$$\mathrm{Var}(a_1W_1 + a_2W_2 + \cdots + a_nW_n) = a_1^2\mathrm{Var}(W_1) + a_2^2\mathrm{Var}(W_2) + \cdots + a_n^2\mathrm{Var}(X_n)$$

Proof. The derivation is based on Theorems 3.9.4 and 3.6.2. The details will be left as an exercise. ☐

Comment. A more general version of Theorem 3.9.4 can be proved, one that leads to a slightly different formula but does not require the W_i's to be independent. The argument, however, depends on a definition we have not yet introduced. We will return to the problem of finding the variance of a sum of random variables in Section 11.4.

EXAMPLE 3.9.8

The binomial random variable, being a sum of n independent Bernoullis, is an obvious candidate for Theorem 3.9.4. Let X_i denote the number of successes occurring on the ith trial. Then

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

and

$$X = X_1 + X_2 + \cdots + X_n = \text{total number of successes in } n \text{ trials.}$$

Find Var(X).

Note that

$$E(X_i) = 1 \cdot p + 0 \cdot (1 - p)$$

and

$$E(X_i^2) = (1)^2 \cdot p + (0)^2 \cdot (1 - p) = p$$

so

$$\begin{aligned} \mathrm{Var}(X_i) &= E(X_i^2) - [E(X_i)]^2 = p - p^2 \\ &= p(1 - p) \end{aligned}$$

It follows, then, that the *variance of a binomial random variable is $np(1 - p)$*:

$$\mathrm{Var}(X) = \sum_{i=1}^{n} \mathrm{Var}(X_i) = np(1 - p)$$

EXAMPLE 3.9.9

In statistics, it is often necessary to draw inferences based on $\overline{W}$, the average computed from a random sample of n observations. Two properties of $\overline{W}$ are especially important. First, if the W_is come from a population where the mean is μ, the corollary to Theorem 3.9.2 implies that $E(\overline{W}) = \mu$. Second, if the W_is come from a population whose variance is σ^2, then $\text{Var}(\overline{W}) = \sigma^2/n$. To verify the latter, we can appeal to Theorem 3.9.4. Write

$$\overline{W} = \frac{1}{n}\sum_{i=1}^{n} W_i = \frac{1}{n} \cdot W_1 + \frac{1}{n} \cdot W_2 + \cdots + \frac{1}{n} \cdot W_n$$

Then

$$\begin{aligned}\text{Var}(\overline{W}) &= \left(\frac{1}{n}\right)^2 \cdot \text{Var}(W_1) + \left(\frac{1}{n}\right)^2 \cdot \text{Var}(W_2) + \cdots + \left(\frac{1}{n}\right)^2 \cdot \text{Var}(W_n)\\ &= \left(\frac{1}{n}\right)^2 \sigma^2 + \left(\frac{1}{n}\right)^2 \sigma^2 + \cdots + \left(\frac{1}{n}\right)^2 \sigma^2\\ &= \frac{\sigma^2}{n}\end{aligned}$$

QUESTIONS

3.9.13. Suppose that $f_{X,Y}(x, y) = \lambda^2 e^{-\lambda(x+y)}$, $0 \le x$, $0 \le y$. Find $\text{Var}(X + Y)$. *Hint:* See Questions 3.6.11 and 3.9.2.

3.9.14. Suppose that $f_{X,Y}(x, y) = \frac{2}{3}(x + 2y)$, $0 \le x \le 1$, $0 \le y \le 1$. Find $\text{Var}(X + Y)$. *Hint:* See Question 3.9.3.

3.9.15. For the uniform pdf defined over $[0, 1]$, find the variance of the geometric mean when $n = 2$ (see Question 3.9.12).

3.9.16. Let X be a binomial random variable based on n trials and a success probability of p_X; let Y be an independent binomial random variable based on m trials and a success probability of p_Y. Find $E(W)$ and $\text{Var}(W)$, where $W = 4X + 6Y$.

3.9.17. Let the Poisson random variable U be the number of calls for technical assistance received by a computer company during the firm's 9 normal workday hours. Suppose the average number of calls per hour is 7.0 and that each call costs the company \$50. Let V be a Poisson random variable representing the number of calls for technical assistance received during a day's remaining 15 hours. Suppose the average number of calls per hour is 4.0 for that time period and that each such call costs the company \$60. Find the expected cost and the variance of the cost associated with the calls received during a 24-hour day.

3.9.18. A mason is contracted to build a patio retaining wall. Plans call for the base of the wall to be a row of 50 10-inch bricks, each separated by $\frac{1}{2}$-inch-thick mortar. Suppose that the bricks used are randomly chosen from a population of bricks whose mean length is 10 inches and whose standard deviation is $\frac{1}{32}$ inch. Also, suppose that the mason, on the average, will make the mortar $\frac{1}{2}$ inch thick, but the actual dimension varies from brick to brick, the standard deviation of the thicknesses being $\frac{1}{16}$ inch. What is the standard deviation of L, the length of the first row of the wall? What assumption are you making?

3.9.19. An electric circuit has six resistors wired in series, each nominally being 5 ohms. What is the maximum standard deviation that can be allowed in the manufacture of these resistors if the combined circuit resistance is to have a standard deviation no greater than 0.4 ohm?

3.9.20. A gambler plays n hands of poker. If he wins the kth hand, he collects k dollars; if he loses the kth hand, he collects nothing. Let T denote his total winnings in n hands. Assuming that his chances of winning each hand are constant and are independent of his success or failure at any other hand, find $E(T)$ and Var(T).

Approximating the Variance of a Function of Random Variables (Optional)

It is not an uncommon problem for a laboratory scientist to have to measure several quantities, each subject to a certain amount of "error," in order to calculate a final desired result. For example, a physics student trying to determine the acceleration due to gravity, G, knows that the distance, D, traveled by a freely falling body in time, T, is related to G by the equation

$$D = \frac{1}{2}GT^2$$

(assuming the body is initially at rest) or, equivalently,

$$G = \frac{2D}{T^2}$$

Suppose distance and time are to be measured directly with a yardstick and a stopwatch. The values obtained, D and T, will not be exactly correct; rather, we can think of them as being realizations of random variables, with those variables having "true" values μ_D and μ_T and variances Var(D) and Var(T), the latter two numbers reflecting the lack of precision in the measuring process. Suppose we know from past experience the precisions characteristic of the distance and time measurements—what can we then conclude about the precision in the calculated value for G? That is, knowing Var(D) and Var(T), can we find Var(G)?

By way of background, we have already seen one result that bears directly on this sort of "error-propagation" problem. If the quantity to be calculated, W, is the *sum* of n independently measured quantities, $W_1, W_2, \ldots, W_n$, and if the variance associated with each of the W_i's is known, we can appeal to Theorem 3.9.4 and say that

$$\text{Var}(W) = \text{Var}(W_1) + \text{Var}(W_2) + \cdots + \text{Var}(W_n) \tag{3.9.2}$$

In general, extending Equation 3.9.2 in any *exact* way to situations where W is some arbitrary function of a set of W_i's—say, $W = g(W_1, W_2, \ldots, W_n)$—is extremely difficult. It is a relatively simple matter, though, to get an *approximation* for the variance of W.

More specifically, suppose that W is a function of n independent random variables—that is, $W = g(W_1, W_2, \ldots, W_n)$. Assume that μ_i and Var(W_i) are the mean and variance, respectively, of W_i, $i = 1, 2, \ldots, n$. Using the first-order terms in a Taylor expansion of the function $g(W_1, W_2, \ldots, W_n)$ around the point $(\mu_1, \mu_2, \ldots, \mu_n)$, we can write

$$W \doteq g(\mu_1, \mu_2, \ldots, \mu_n) + (W_1 - \mu_1)\left[\left.\frac{\partial g}{\partial W_1}\right|_{(\mu_1,\ldots,\mu_n)}\right] + (W_2 - \mu_2)\left[\left.\frac{\partial g}{\partial W_2}\right|_{(\mu_1,\ldots,\mu_n)}\right] + \cdots + (W_n - \mu_n)\left[\left.\frac{\partial g}{\partial W_n}\right|_{(\mu_1,\ldots,\mu_n)}\right] \tag{3.9.3}$$

Applying the corollary to Theorem 3.9.4 to Equation 3.9.3 yields the sought-after approximation:

$$\operatorname{Var}(W) \doteq \left[\left.\frac{\partial g}{\partial W_1}\right|_{(\mu_1,\mu_2,\ldots,\mu_n)}\right]^2 \operatorname{Var}(W_1) + \left[\left.\frac{\partial g}{\partial W_2}\right|_{(\mu_1,\mu_2,\ldots,\mu_n)}\right]^2 \operatorname{Var}(W_2) + \cdots + \left[\left.\frac{\partial g}{\partial W_n}\right|_{(\mu_1,\mu_2,\ldots,\mu_n)}\right]^2 \operatorname{Var}(W_n). \tag{3.9.4}$$

CASE STUDY 3.9.1

In a typical dental X-ray unit, electrons from the cathode of the X-ray tube are decelerated by nuclei in the anode, thereby producing Bremsstrahlung radiation (X-rays). These emissions, when collimated by a lead-lined tube, effect the desired image on a sheet of film.

Tennessee state regulations (170) require that the distance, W, from the focal spot on the anode of an X-ray tube to the patient's skin be at least 18 cm. On some equipment, particularly older units, that distance cannot be measured directly because the exact location of the focal spot cannot be determined just by looking at the tube's outer housing. When this is the case, state inspectors resort to an indirect measuring procedure. Two films are exposed, one at the unknown distance W and a second at a distance $W + Z$. The two diameters, X and Y, of the resulting circular images are then measured (see Figure 3.9.2).

By similar triangles,

$$\frac{X}{W} = \frac{Y}{W + Z} \quad \text{or} \quad W = \frac{XZ}{Y - X} \tag{3.9.5}$$

Phrased in the context of our previous notation,

$$W = g(X, Y, Z) = XZ(Y - X)^{-1}$$

(Continued on next page)

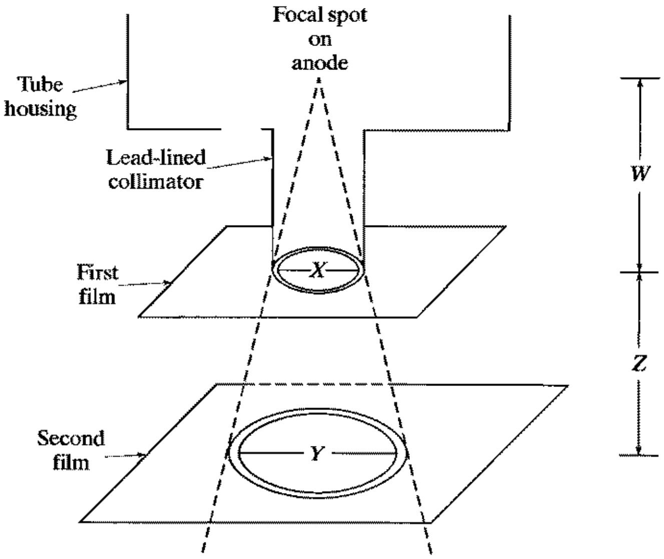


FIGURE 3.9.2

During the course of one such inspection (96), values measured for the two diameters X and Y and the backoff distance Z were 6.4 cm, 9.7 cm, and 10.2 cm, respectively. From Equation 3.9.5, then, the anode-to-patient distance is estimated to be

$$W = \frac{(6.4)(10.2)}{9.7 - 6.4} = 19.8 \text{ cm}$$

indicating that the unit is in compliance. If the error in W, though, were sufficiently large, there might still be a sizable probability that the *true* W was less than 18 cm, meaning the unit was, in fact, out of compliance. It is not unreasonable, therefore, to inquire about the magnitude of Var(W).

To apply Equation 3.9.4, we first need to compute the partial derivatives of $g(X, Y, Z)$. In this case,

$$\frac{\partial g}{\partial X} = \frac{XZ}{(Y - X)^2} + \frac{Z}{(Y - X)}$$

$$\frac{\partial g}{\partial Y} = \frac{-XZ}{(Y - X)^2}$$

and

$$\frac{\partial g}{\partial Z} = \frac{X}{(Y - X)}$$

(Continued on next page)

(Case Study 3.9.1 continued)

Inspectors feel that the standard deviation in any of their measurements is on the order of 0.08 cm, so $\text{Var}(X) = \text{Var}(Y) = \text{Var}(Z) = (0.08)^2$. Substituting the variance estimates and the partial derivatives, evaluated at the point $(\mu_X, \mu_Y, \mu_Z) \doteq (6.4, 9.7, 10.2)$ into Equation 3.9.4 gives

$$\begin{aligned}\text{Var(W)} &\doteq \left[\frac{(6.4)(10.2)}{(9.7-6.4)^2} + \frac{10.2}{(9.7-6.4)}\right]^2 (0.08)^2 \\ &+ \left[\frac{-(6.4)(10.2)}{(9.7-6.4)^2}\right]^2 (0.08)^2 + \left[\frac{6.4}{(9.7-6.4)}\right]^2 (0.08)^2 \\ &= 0.782\end{aligned}$$

Therefore, the estimated standard deviation associated with the calculated value of W is $\sqrt{0.782}$, or 0.88 cm.

QUESTIONS

3.9.21. A physics student is trying to determine the gravitational constant, G, using the expression

$$G = \frac{2D}{T^2}$$

where both distance (D) and time (T) are to be measured. Suppose that the standard deviation of the measurement errors in D is 0.0025 feet and in T, 0.045 seconds. If the experimental apparatus is set up so that D will be 4 feet, then T till be approximately $\frac{1}{2}$ second. If D is set at 16 feet, T will be close to 1 second. Which of these two sets of values for D and T will give a smaller variance for the calculated G?

3.9.22. Suppose that $W_1, W_2, \ldots,$ and W_n are independent random variables with variances σ_1^2, $\sigma_2^2, \ldots,$ and σ_n^2 respectively, and let $W = W_1 + W_2 + \cdots + W_n$. Compare $\text{Var}(W)$ using Theorem 3.9.4 and Equation 3.9.4.

3.9.23. If h is its height and a and b are the lengths of its two parallel sides, the area of a trapezoid is given by

$$A = \frac{1}{2}(a + b)h$$

Find an expression that approximates σ_A if a, b, and h are measured independently with standard deviations σ_a, σ_b, and σ_h, respectively.

3.9.24. In Case Study 3.9.1, notice that the difference between 19.8 cm (the calculated distance) and 18 cm (the state regulation minimum distance) is slightly more than two standard deviations. What does that imply about the probability that this particular X-ray machine is operating safely?

3.10 ORDER STATISTICS

The single-variable transformation taken up in Section 3.4 involved a standard linear operation, $Y = aX + b$. The bivariate transformations in Section 3.8 were similarly arithmetic, typically being concerned with either sums or products. In this section we will consider a different sort of transformation, one involving the *ordering* of an entire *set* of random variables. This particular transformation has wide applicability in many areas of statistics, and we will see some of its consequences in later chapters. Here, though, we will limit our discussion to two basic results: derivations of (1) the marginal pdf of the ith largest observation in a random sample and (2) the joint pdf of the ith and jth largest observations in a random sample.

Definition 3.10.1. Let Y be continuous random variable for which $y_1, y_2, \ldots, y_n$ are the values of a random sample of size n. Reorder the y_is from smallest to largest:

$$y_1' < y_2' < \cdots < y_n'$$

(No two of the y_is are equal, except with probability zero, since Y is continuous.) Define the random variable Y_i' to have the value y_i', $1 \leq i \leq n$. Then Y_i' is called the ith *order statistic*. Sometimes Y_n' and Y_1' are denoted $Y_{\max}$ and $Y_{\min}$, respectively.

EXAMPLE 3.10.1

Suppose that four measurements are made on the random variable Y: $y_1 = 3.4$, $y_2 = 4.6$, $y_3 = 2.6$, and $y_4 = 3.2$. The corresponding ordered sample would be

$$2.6 < 3.2 < 3.4 < 4.6$$

The random variable representing the smallest observation would be denoted Y_1', with its value for this particular sample being 2.6. Similarly, the value for the second order statistic, Y_2', is 3.2, and so on.

The Distribution of Extreme Order Statistics

By definition, every observation in a random sample has the same pdf. For example, if a set of four measurements is taken from a normal distribution with $\mu = 80$ and $\sigma = 15$, then $f_{Y_1}(y)$, $f_{Y_2}(y)$, $f_{Y_3}(y)$, and $f_{Y_4}(y)$ are all the same—each is a normal pdf with $\mu = 80$ and $\sigma = 15$. The pdf describing an *ordered* observation, though, is *not* the same as the pdf describing a *random* observation. Intuitively, that makes sense. If a single observation is drawn from a normal distribution with $\mu = 80$ and $\sigma = 15$, it would not be surprising if that observation were to take on a value near eighty. On the other hand, if a random sample of $n = 100$ observations is drawn from that same distribution, we would not expect the smallest observation—that is, $Y_{\min}$—to be anywhere near eighty. Common sense tells us that that smallest observation is likely to be much smaller than eighty, just as the largest observation, $Y_{\max}$, is likely to be much larger than eighty.

It follows, then, that before we can do any probability calculations—or any applications whatsoever—involving order statistics, we need to know the pdf of Y_i' for $i = 1, 2, \ldots, n$. We begin by investigating the pdfs of the "extreme" order statistics, $f_{Y_{\max}}(y)$ and $f_{Y_{\min}}(y)$. These are the simplest to work with. At the end of the section we return to the more general problems of finding (a) the pdf of Y_i' for any i and (b) the joint pdf of Y_i' and Y_j', where $i < j$.

EXAMPLE 3.10.2

Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of continuous random variables, each having pdf $f_Y(y)$ and cdf $F_Y(y)$. Find

a. $f_{Y_{\max}}(y) = f_{Y'_n}(y)$, the pdf of the largest order statistic

b. $f_{Y_{\min}}(y) = f_{Y'_1}(y)$, the pdf of the smallest order statistic

Finding the pdfs of $Y_{\max}$ and $Y_{\min}$ is accomplished by using the now-familiar technique of differentiating a random variable's cdf. Consider, for example, the case of the largest order statistic, Y'_n:

$$\begin{aligned} F_{Y'_n}(y) = F_{Y_{\max}}(y) &= P(Y_{\max} \leq y) \\ &= P(Y_1 \leq y \cap Y_2 \leq y \cap \cdots \cap Y_n \leq y) \\ &= P(Y_1 \leq y) \cdot P(Y_2 \leq y) \cdots P(Y_n \leq y) \qquad \text{(why?)} \\ &= [F_Y(y)]^n \end{aligned}$$

Therefore,

$$f_{Y'_n}(y) = d/dy[[F_Y(y)]^n] = n[F_Y(y)]^{n-1} f_Y(y)$$

Similarly, for the smallest order statistic ($i = 1$),

$$\begin{aligned} F_{Y'_1}(y) &= F_{Y_{\min}}(y) = P(Y_{\min} \leq y) \\ &= 1 - P(Y_{\min} > y) = 1 - P(Y_1 > y) \cdot P(Y_2 > y) \cdots P(Y_n > y) \\ &= 1 - [1 - F_Y(y)]^n \end{aligned}$$

Therefore,

$$f_{Y'_1}(y) = d/dy[1 - [1 - F_Y(y)]^n] = n[1 - F_Y(y)]^{n-1} f_Y(y)$$

EXAMPLE 3.10.3

Suppose a random sample of $n = 3$ observations—Y_1, Y_2, and Y_3—is taken from the exponential pdf, $f_Y(y) = e^{-y}$, $y \geq 0$. Compare $f_{Y_1}(y)$ with $f_{Y'_1}(y)$. Intuitively, which will be larger, $P(Y_1 < 1)$ or $P(Y'_1 < 1)$?

The pdf for Y_1, of course, is just the pdf of the distribution being sampled—that is,

$$f_{Y_1}(y) = f_Y(y) = e^{-y}, \quad y \geq 0$$

To find the pdf for Y'_1 requires that we apply the formula given in Example 3.10.2 for $f_{Y_{\min}}(y)$. Note, first of all, that

$$F_Y(y) = \int_0^y e^{-t}\,dt = -e^{-t}\Big|_0^y = 1 - e^{-y}$$

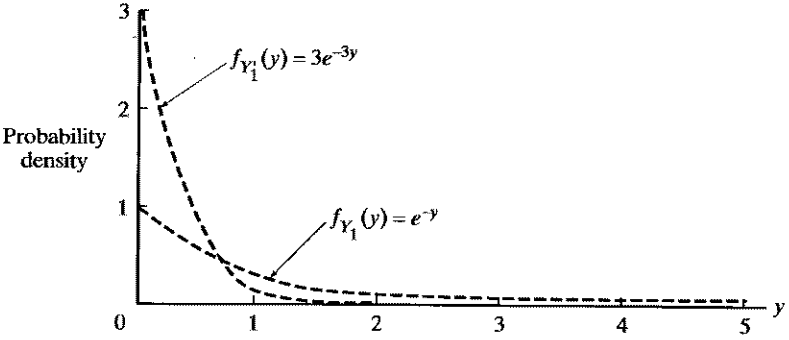


FIGURE 3.10.1

Then, since $n = 3$ (and $i = 1$), we can write

$$\begin{aligned} f_{Y_1'}(y) &= 3(1 - [1 - e^{-y}])^2 e^{-y} \\ &= 3e^{-3y}, \quad y \geq 0 \end{aligned}$$

Figure 3.10.1 shows the two pdfs plotted on the same set of axes. Compared to $f_{Y_1}(y)$, the pdf for Y_1' has more of its area located above the smaller values of y (where Y_1' is more likely to lie). For example, the probability that the smallest observation (out of three) is less than one is 95%, while the probability that a random observation is less than one is only 63%:

$$\begin{aligned} P(Y_1' < 1) &= \int_0^1 3e^{-3y}\,dy = \int_0^3 e^{-u}\,du = -e^{-u}\Big|_0^3 = 1 - e^{-3} \\ &= 0.95 \\ P(Y_1 < 1) &= \int_0^1 e^{-y}\,dy = -e^{-y}\Big|_0^1 = 1 - e^{-1} \\ &= 0.63 \end{aligned}$$

EXAMPLE 3.10.4

Suppose a random sample of size ten is drawn from a continuous pdf $f_Y(y)$. What is the probability that the largest observation, Y_{10}', is less than the pdf's median, m?

Using the formula for $f_{Y_{10}'}(y) = f_{Y_{\max}}(y)$ given in Example 3.10.2, it is certainly true that

$$P(Y_{10}' < m) = \int_{-\infty}^{m} 10 f_Y(y)[F_Y(y)]^9 dy \qquad (3.10.1)$$

but the problem does not specify $f_Y(y)$, so Equation 3.10.1 is of no help.

Fortunately, a much simpler solution is available, even if $f_Y(y)$ were specified: The event "$Y'_{10} < m$" is equivalent to the event "$Y_1 < m \cap Y_2 < m \cap \cdots \cap Y_{10} < m$". Therefore,

$$P(Y'_{10} < m) = P(Y_1 < m, Y_2 < m, \ldots, Y_{10} < m) \tag{3.10.2}$$

But the ten observations here are independent, so the intersection probability implicit on the right-hand side of Equation 3.10.2 factors into a product of ten terms. Moreover, each of those terms equals $\frac{1}{2}$ (by definition of the median), so

$$\begin{aligned} P(Y'_{10} < m) &= P(Y_1 < m) \cdot P(Y_2 < m) \cdots P(Y_{10} < m) \\ &= \left(\tfrac{1}{2}\right)^{10} \\ &= 0.00098 \end{aligned}$$

A General Formula for $f_{Y'_i}(y)$

Having discussed two special cases of order statistics, $Y_{\min}$ and $Y_{\max}$, we now turn to the more general problem of finding the pdf for the ith order statistic, where i can be any integer from 1 through n.

Theorem 3.10.1. *Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of continuous random variables drawn from a distribution having pdf $f_Y(y)$ and cdf $F_Y(y)$. The pdf of the ith order statistic is given by*

$$f_{Y'_i}(y) = \frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1 - F_Y(y)]^{n-i} f_Y(y)$$

for $1 \leq i \leq n$.

Proof. We will give a heuristic argument that draws on the similarity between the statement of Theorem 3.10.1 and the binomial distribution. For a formal induction proof verifying the expression given for $f_{Y'_i}(y)$, see (98).

Recall the derivation of the binomial probability function, $p_X(k) = P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$, where X is the number of successes in n independent trials, and p is the probability that any given trial ends in success. Central to that derivation was the recognition that the event "$X = k$" is actually a union of all the different (mutually exclusive) sequences having exactly k successes and $n - k$ failures. Because the trials are independent, the probability of any such sequence is $p^k(1-p)^{n-k}$ and the number of such sequences (by Theorem 2.6.2) is $n!/[k!(n-k)!]$ (or $\binom{n}{k}$), so the probability that $X = k$ is the product, $\binom{n}{k} p^k (1-p)^{n-k}$.

Here we are looking for the pdf of the ith order statistic at some point y—that is, $f_{Y'_i}(y)$. As was the case with the binomial, that pdf will reduce to a combinatorial

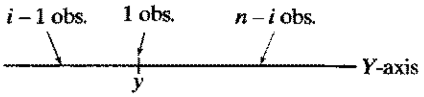


FIGURE 3.10.2

term times the probability associated with an intersection of independent events. The only fundamental difference is that Y_i' is a continuous random variable, whereas the binomial X is discrete, which means that what we find here will be a probability *density* function.

By Theorem 2.6.2, there are $n!/[(i-1)!1!(n-i)!]$ ways that n observations can be parceled into three groups such that the ith largest is at the point y (see Figure 3.10.2). Moreover, the likelihood associated with any particular set of points having the configuration pictured in Figure 3.10.2 will be the probability that $i-1$ (independent) observations are all less than y, $n-i$ observations are greater than y, and one observation is at y. The probability density associated with those constraints for a given set of points would be $[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}f_Y(y)$. The probability density, then, that the ith order statistic is located at the point y is the product,

$$f_{Y_i'}(y) = \frac{n!}{(i-1)!(n-i)!}[F_Y(y)]^{i-1}[1-F_Y(y)]^{n-i}f_Y(y) \qquad \square$$

EXAMPLE 3.10.5

Suppose that many years of observation have confirmed that the annual maximum flood tide Y (in feet) for a certain river can be modeled by the pdf

$$f_Y(y) = \frac{1}{20}, \quad 20 < y < 40$$

(*Note:* It is unlikely that flood tides would be described by anything as simple as a uniform pdf. We are making that choice here solely to facilitate the mathematics.) The Army Corps of Engineers are planning to build a levee along a certain portion of the river, and they want to make it high enough so that there is only a 30% chance that the second worst flood in the next thirty-three years will overflow the embankment. How high should the levee be? (We assume that there will be only one potential flood per year.)

Let h be the desired height. If $Y_1, Y_2, \ldots, Y_{33}$ denote the flood tides for the next $n = 33$ years, what we require of h is that

$$P(Y_{32}' > h) = 0.30$$

As a starting point, notice that for $20 < y < 40$,

$$F_Y(y) = \int_{20}^{y} \frac{1}{20}\,dy = \frac{y}{20} - 1$$

Therefore,

$$f_{Y'_{32}}(y) = \frac{33!}{31!1!}\left(\frac{y}{20} - 1\right)^{31}\left(2 - \frac{y}{20}\right)^{1} \cdot \frac{1}{20}$$

and h is the solution of the integral equation

$$\int_{h}^{40} (33)(32)\left(\frac{y}{20} - 1\right)^{31}\left(2 - \frac{y}{20}\right)^{1} \cdot \frac{dy}{20} = 0.30 \tag{3.10.3}$$

If we make the substitution

$$u = \frac{y}{20} - 1$$

Equation 3.10.3 simplifies to

$$\begin{aligned} P(Y'_{32} > h) &= 33(32)\int_{(h/20)-1}^{1} u^{31}(1 - u)\,du \\ &= 1 - 33\left(\frac{h}{20} - 1\right)^{32} + 32\left(\frac{h}{20} - 1\right)^{33} \end{aligned} \tag{3.10.4}$$

Setting the right-hand side of Equation 3.10.4 equal to 0.30 and solving for h by trial and error gives

$$h = 39.3 \text{ feet}$$

Joint pdfs of Order Statistics

Finding the joint pdf of two or more order statistics is easily accomplished by generalizing the argument that derived from Figure 3.10.2. Suppose, for example, that each of n observations in a random sample has pdf $f_Y(y)$ and cdf $F_Y(y)$. The joint pdf for order statistics Y'_i and Y'_j at points u and v, where $i < j$ and $u < v$, can be deduced from Figure 3.10.3, which shows how the n points must be distributed if the ith and jth order statistics are to be located at points u and v, respectively.

By Theorem 2.6.2, the number of ways to divide a set of n observations into groups of sizes $i - 1, 1, j - i - 1, 1$, and $n - j$ is the quotient

$$\frac{n!}{(i - 1)!1!(j - i - 1)!1!(n - j)!}$$

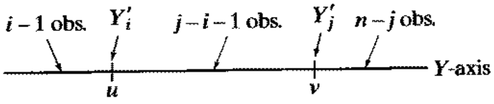


FIGURE 3.10.3

Also, given the independence of the n observations, the probability that $i - 1$ are less than u is $[F_Y(u)]^{i-1}$, the probability that $j - i - 1$ are between u and v is $[F_Y(v) - F_Y(u)]^{j-i-1}$, and the probability that $n - j$ are greater than v is $[1 - F_Y(v)]^{n-j}$. Multiplying, then, by the pdfs describing the likelihoods that Y_i' and Y_j' would be at points u and v, respectively, gives the joint pdf of the two order statistics:

$$\begin{aligned} f_{Y_i',Y_j'}(u,v) = &\frac{n!}{(i-1)!(j-i-1)!(n-j)!}[F_Y(u)]^{i-1}[F_Y(v) - F_Y(u)]^{j-i-1}\cdot \\ &[1 - F_Y(v)]^{n-j} f_Y(u) f_Y(v) \end{aligned} \tag{3.10.5}$$

for $i < j$ and $u < v$.

EXAMPLE 3.10.6

Let Y_1, Y_2, and Y_3 be a random sample of size $n = 3$ from the uniform pdf defined over the unit interval, $f_Y(y) = 1$, $0 \le y \le 1$. By definition, the *range*, R, of a sample is the difference between the largest and smallest order statistics—in this case,

$$R = \text{range} = Y_{\max} - Y_{\min} = Y_3' - Y_1'$$

Find $f_R(r)$, the pdf for the range.

We will begin by finding the joint pdf of Y_1' and Y_3'. Then $f_{Y_1',Y_3'}(u,v)$ is integrated over the region $Y_3' - Y_1' \le r$ to find the cdf, $F_R(r) = P(R \le r)$. The final step is to differentiate the cdf and make use of the fact that $f_R(r) = F_R'(r)$.

If $f_Y(y) = 1, 0 \le y \le 1$, it follows that

$$F_Y(y) = P(Y \le y) = \begin{cases} 0, & y < 0 \\ y, & 0 \le y \le 1 \\ 1. & y > 1 \end{cases}$$

Applying Equation 3.10.5, then, with $n = 3$, $i = 1$, and $j = 3$, gives the joint pdf of Y_1' and Y_3'. Specifically,

$$\begin{aligned} f_{Y_1',Y_3'}(u,v) &= \frac{3!}{0!1!0!}u^0(v-u)^1(1-v)^0 \cdot 1 \cdot 1 \\ &= 6(v-u), \quad 0 \le u < v \le 1 \end{aligned}$$

Moreover, we can write the cdf for R in terms of Y_1' and Y_3':

$$F_R(r) = P(R \le r) = P(Y_3' - Y_1' \le r) = P(Y_3' \le Y_1' + r)$$

Figure 3.10.4 shows the region in the $Y_1'Y_3'$-plane corresponding to the event that $R \le r$. Integrating the joint pdf of Y_1' and Y_3' over the shaded region gives

$$F_R(r) = P(R \le r) = \int_0^{1-r}\int_u^{u+r} 6(v-u)\,dv\,du + \int_{1-r}^{1}\int_u^{1} 6(v-u)\,dv\,du$$

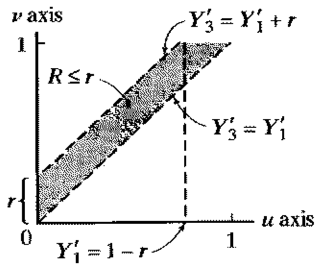


FIGURE 3.10.4

The first double integral equals $3r^2 - 3r^3$; the second equals r^3. Therefore,

$$F_R(r) = 3r^2 - 3r^3 + r^3 = 3r^2 - 2r^3$$

which implies that

$$f_R(r) = F_R'(r) = 6r - 6r^2, \quad 0 \le r \le 1$$

QUESTIONS

3.10.1. Suppose the length of time, in minutes, that you have to wait at a bank teller's window is uniformly distributed over the interval (0, 10). If you go to the bank four times during the next month, what is the probability that your second longest wait will be less than 5 minutes?

3.10.2. A random sample of size $n = 6$ is taken from the pdf $f_Y(y) = 3y^2$, $0 \le y \le 1$. Find $P(Y_5' > 0.75)$.

3.10.3. What is the probability that the larger of two random observations drawn from any continuous pdf will exceed the sixtieth percentile?

3.10.4. A random sample of size 5 is drawn from the pdf $f_Y(y) = 2y$, $0 \le y \le 1$. Calculate $P(Y_1' < 0.6 < Y_5')$. *Hint:* Consider the complement.

3.10.5. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n drawn from a continuous pdf, $f_Y(y)$, whose median is m. Is $P(Y_1' > m)$ less than, equal to, or greater than $P(Y_n' > m)$?

3.10.6. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the exponential pdf $f_y(y) = e^{-y}$, $y > 0$. What is the smallest n for which $P(Y_{\min} < 0.2) > 0.9$?

3.10.7. Calculate $P(0.6 < Y_4' < 0.7)$ if a random sample of size 6 is drawn from the uniform pdf defined over the interval (0,1).

3.10.8. A random sample of size $n = 5$ is drawn from the pdf $f_Y(y) = 2y$, $0 < y < 1$. On the same set of axes, graph the pdfs for Y_2, Y_1', and Y_5'.

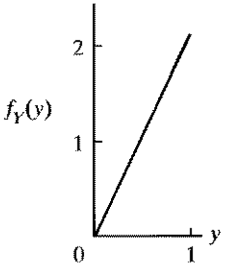
3.10.9. Suppose that n observations are taken at random from the pdf

$$f_Y(y) = \frac{1}{\sqrt{2\pi}(6)} e^{-\frac{1}{2}\left(\frac{y-20}{6}\right)^2}, \quad -\infty < y < \infty$$

What is the probability that the smallest observation is larger than 20?

3.10.10. Suppose that n observations are chosen at random from a continuous pdf $f_Y(y)$. What is the probability that the last observation recorded will be the smallest number in the entire sample?

3.10.11. In a certain large metropolitan area the proportion, Y, of students bused varies widely from school to school. The distribution of proportions is roughly described by the following pdf:



Suppose the enrollment figures for five schools selected at random are examined. What is the probability that the school with the fourth highest proportion of bused children will have a Y-value in excess of 0.75? What is the probability that none of the schools will have fewer than 10% of their student bused?

3.10.12. Consider a system containing n components, where the lifetimes of the components are independent random variables and each has pdf $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Show that the average time elapsing before the first component failure occurs is $1/n\lambda$.

3.10.13. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a uniform pdf over [0,1]. Use Theorem 3.10.1 to show that $\int_0^1 y^{i-1}(1 - y)^{n-i}dy = \frac{(i - 1)!(n - i)!}{n!}$.

3.10.14. Use Question 3.10.13 to find the expected value of Y_i', where $Y_1, Y_2, \ldots, Y_n$ is a random sample from a uniform pdf defined over the interval [0, 1].

3.10.15. Suppose three points are picked randomly from the unit interval. What is the probability that the three are within a half unit of one another?

3.10.16. Suppose a device has three independent components, all of whose lifetimes (in months) are modeled by the exponential pdf, $f_Y(y) = e^{-y}$, $y > 0$. What is the probability that all three components will fail within two months of one another?

CONDITIONAL DENSITIES

We have already seen that many of the concepts defined in Chapter 2 relating to the probabilities of *events*—for example, independence—have their random-variable counterparts. Another of these carryovers is the notion of a conditional probability, or, in what will be our present terminology, a *conditional probability density function.* Applications of conditional pdfs are not uncommon. The height and girth of a tree, for instance, can be considered a pair of random variables. While it is easy to measure girth, it can be difficult to determine height; thus it might be of interest to a lumberman to know the probabilities of a Ponderosa pine's attaining certain heights given a known value for its girth. Or consider the plight of a school board member agonizing over which way to vote on a proposed budget increase. Her task would be that much easier if she knew the conditional probability that x additional tax dollars would stimulate an average increase of y points among twelfth-graders taking a standardized proficiency exam.

Finding Conditional pdfs for Discrete Random Variables

In the case of discrete random variables, a conditional pdf can be treated in the same way as a conditional probability. Note the similarity between Definitions 3.11.1 and 2.4.1.

Definition 3.11.1. Let X and Y be discrete random variables. The *conditional probability density function of Y given x*—that is, the probability that Y takes on the value y given that X is equal to x—is denoted $p_{Y|x}(y)$ and given by

$$p_{Y|x}(y) = P(Y = y \mid X = x) = \frac{p_{X,Y}(x,y)}{p_X(x)}$$

for $p_X(x) \neq 0$.

EXAMPLE 3.11.1

A fair coin is tossed five times. Let the random variable Y denote the total number of heads that occur, and let X denote the number of heads occurring on the last two tosses. Find the conditional pdf $p_{Y|x}(y)$ for all x and y.

Clearly, there will be three different conditional pdfs, one for each possible value of X ($x = 0, x = 1$, and $x = 2$). Moreover, for each value of x there will be four possible values of Y, based on whether the first three tosses yield 0, 1, 2, or 3 heads.

For example, suppose no heads occur on the last two tosses. Then $X = 0$, and

$$\begin{aligned} p_{Y|0}(y) = P(Y = y \mid X = 0) &= P(y \text{ Heads occur on first three tosses}) \\ &= \binom{3}{y}\left(\frac{1}{2}\right)^{y}\left(1-\frac{1}{2}\right)^{3-y} \\ &= \binom{3}{y}\left(\frac{1}{2}\right)^{3}, \quad y = 0, 1, 2, 3 \end{aligned}$$

Now, suppose that $X = 1$. The corresponding conditional pdf in that case becomes

$$p_{Y|x}(y) = P(Y = y \mid X = 1)$$

Notice that $Y = 1$ if zero heads occur in the first three tosses, $Y = 2$ if one head occurs in the first three trials, and so on. Therefore,

$$\begin{aligned} p_{Y|1}(y) &= \binom{3}{y-1}\left(\frac{1}{2}\right)^{y-1}\left(1-\frac{1}{2}\right)^{3-(y-1)} \\ &= \binom{3}{y-1}\left(\frac{1}{2}\right)^{3}, \quad y = 1, 2, 3, 4 \end{aligned}$$

Similarly,

$$p_{Y|2}(y) = P(Y = y \mid X = 2) = \binom{3}{y-2}\left(\frac{1}{2}\right)^{3}, \quad y = 2, 3, 4, 5$$

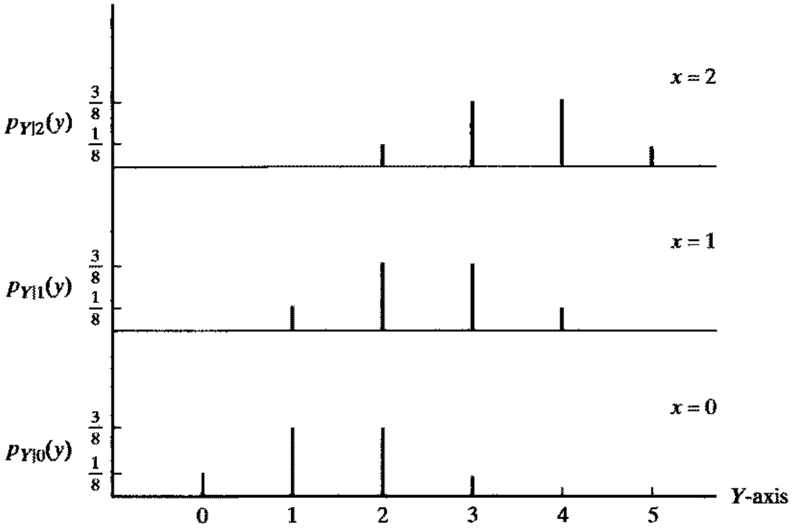


FIGURE 3.11.1

Figure 3.11.1 shows the three conditional pdfs. Each has the same shape, but the possible values of Y are different for each value of X

EXAMPLE 3.11.2

Assume that the probabilistic behavior of a pair of discrete random variables X and Y is described by the joint pdf

$$p_{X,Y}(x, y) = xy^2/39$$

defined over the four points (1, 2), (1, 3), (2, 2), and (2, 3). Find the conditional probability that $X = 1$ given that $Y = 2$.

By definition,

$$\begin{aligned} p_{X|2}(1) &= P(X = 1 \text{ given that } Y = 2) \\ &= \frac{p_{X,Y}(1, 2)}{p_Y(2)} \\ &= \frac{1 \cdot 2^2/39}{1 \cdot 2^2/39 + 2 \cdot 2^2/39} \\ &= 1/3 \end{aligned}$$

EXAMPLE 3.11.3

Suppose that X and Y are two independent binomial random variables, each defined on n trials and each having the same success probability p. Let $Z = X + Y$. Show that the conditional pdf $p_{X|z}(x)$ is a hypergeometric distribution.

We know from Example 3.8.1 that Z has a binomial distribution with parameters $2n$ and p. That is,

$$p_Z(z) = P(Z = z) = \binom{2n}{z} p^z (1 - p)^{2n-z}, \quad z = 0, 1, \ldots, 2n.$$

By Definition 3.11.1,

$$\begin{aligned}
p_{X|z}(x) = P(X = x|Z = z) &= \frac{p_{X,Z}(x, z)}{p_Z(z)} \\
&= \frac{P(X = x \text{ and } Z = z)}{P(Z = z)} \\
&= \frac{P(X = x \text{ and } Y = z - x)}{P(Z = z)} \\
&= \frac{P(X = x) \cdot P(Y = z - x)}{P(Z = z)} \quad \text{(because } X \text{ and } Y \text{ are independent)} \\
&= \frac{\binom{n}{x} p^x (1 - p)^{n-x} \cdot \binom{n}{z - x} p^{z-x} (1 - p)^{n-(z-x)}}{\binom{2n}{z} p^z (1 - p)^{2n-z}} \\
&= \frac{\binom{n}{x}\binom{n}{z - x}}{\binom{2n}{z}}
\end{aligned}$$

which we recognize as being the hypergeometric distribution.

Comment. The notion of a conditional pdf generalizes easily to situations involving more than two discrete random variables. For example, if X, Y, and Z have the joint pdf $p_{X,Y,Z}(x, y, z)$, the *joint conditional pdf* of, say, X and Y given that $Z = z$ is the ratio

$$p_{X,Y|z}(x, y) = \frac{p_{X,Y,Z}(x, y, z)}{p_Z(z)}$$

EXAMPLE 3.11.4

Suppose that random variables X, Y, and Z have the joint pdf

$$p_{X,Y,Z}(x, y, z) = xy/9z$$

for points (1, 1, 1), (2, 1, 2), (1, 2, 2), (2, 2, 2), and (2, 2, 1). Find $p_{X,Y|z}(x, y)$ for all values of z.

To begin, we see from the points for which $p_{X,Y,Z}(x, y, z)$ is defined that Z has two possible values, 1 and 2. Suppose $z = 1$. Then

$$p_{X,Y|1}(x, y) = \frac{p_{X,Y,Z}(x, y, 1)}{p_Z(1)}$$

But

$$\begin{aligned} p_Z(1) = P(Z = 1) &= P[(1, 1, 1) \cup (2, 2, 1)] \\ &= 1 \cdot \frac{1}{9} \cdot 1 + 2 \cdot \frac{2}{9} \cdot 1 \\ &= \frac{5}{9} \end{aligned}$$

Therefore,

$$p_{X,Y|1}(x, y) = \frac{xy/9}{\frac{5}{9}} = xy/5 \quad \text{for} \quad (x, y) = (1, 1) \quad \text{and} \quad (2, 2)$$

Suppose $z = 2$. Then

$$\begin{aligned} p_Z(2) = P(Z = 2) &= P[(2, 1, 2) \cup (1, 2, 2) \cup (2, 2, 2)] \\ &= 2 \cdot \frac{1}{18} + 1 \cdot \frac{2}{18} + 2 \cdot \frac{2}{18} \\ &= \frac{8}{18} \end{aligned}$$

so

$$\begin{aligned} p_{X,Y|2}(x, y) &= \frac{p_{X,Y,Z}(x, y, 2)}{p_Z(2)} \\ &= \frac{x \cdot y/18}{\frac{8}{18}} \\ &= \frac{xy}{8} \quad \text{for} \quad (x, y) = (2, 1), (1, 2), \quad \text{and} \quad (2, 2) \end{aligned}$$

QUESTIONS

3.11.1. Suppose X and Y have the joint pdf $p_{X,Y}(x, y) = \dfrac{x + y + xy}{21}$ for the points (1, 1), (1, 2), (2, 1), (2, 2), where X denotes a "message" sent (either $x = 1$ or $x = 2$) and Y denotes a "message" received. Find the probability that the message sent was the message received—that is, find $p_{Y|x}(x)$.

3.11.2. Suppose a die is rolled six times. Let X be the total number of 4's that occur and let Y be the number of 4's in the first two tosses. Find $p_{Y|x}(y)$.

3.11.3. An urn contains eight red chips, six white chips, and four blue chips. A sample of size 3 is drawn without replacement. Let X denote the number of red chips in the sample and Y, the number of white chips. Find an expression for $p_{Y|x}(y)$.

3.11.4. Five cards are dealt from a standard poker deck. Let X be the number of aces received, and Y, the number of kings. Compute $P(X = 2|Y = 2)$.

3.11.5. Given that two discrete random variables X and Y follow the joint pdf $p_{X,Y}(x, y) = k(x + y)$, for $x = 1, 2, 3$ and $y = 1, 2, 3$,

(a) Find k.

(b) Evaluate $p_{Y|x}(1)$ for all values of x for which $p_x(x) > 0$.

3.11.6. Let X denote the number on a chip drawn at random from an urn containing three chips, numbered 1, 2, and 3. Let Y be the number of heads that occur when a fair coin is tossed X times.

(a) Find $p_{X,Y}(x, y)$.

(b) Find the marginal pdf of Y by summing out the x-values.

3.11.7. Suppose X, Y, and Z have a trivariate distribution described by the joint pdf

$$p_{X,Y,Z}(x, y, z) = \frac{xy + xz + yz}{54}$$

where x, y, and z can be 1 or 2. Tabulate the joint conditional pdf of X and Y given each of the two values of z.

3.11.8. In Question 3.11.7 define the random variable W to be the "majority" of x, y, and z. For example, $W(2, 2, 1) = 2$ and $W(1, 1, 1) = 1$. Find the pdf of $W|x$.

3.11.9. Let X and Y be independent random variables where $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$ and $p_Y(k) = e^{-\mu}\frac{\mu^k}{k!}$ for $k = 0, 1, \ldots$. Show that the conditional pdf of X given $X + Y = n$ is binomial with parameters n and $\frac{\lambda}{\lambda + \mu}$. *Hint:* See Question 3.8.1.

3.11.10. Suppose Compositor A is preparing a manuscript to be published. Assume that she makes X errors on a given page, where X has the Poisson pdf, $p_X(k) = e^{-2}2^k/k!$, $k = 0, 1, 2, \ldots$ A second compositor, B, is also working on the book. He makes Y errors on a page, where $p_Y(k) = e^{-3}3^k/k!$, $k = 0, 1, 2, \ldots$ Assume that Compositor A prepares the first 100 pages of the text and Compositor B, the last 100 pages. After the book is completed, reviewers (with too much time on their hands!) find that the text contains a total of 520 errors. Write a formula for the exact probability that fewer than half of the errors are due to Compositor A.

Finding Conditional pdfs for Continuous Random Variables

If the variables X and Y are continuous, we can still appeal to the quotient $f_{X,Y}(x, y)/f_X(x)$ as the definition of $f_{Y|x}(y)$ and argue its propriety by analogy. A more satisfying approach, though, is to arrive at the same conclusion by taking the limit of Y's "conditional" *cdf.*

If X is continuous, a direct evaluation of $F_{Y|x}(y) = P(Y \le y|X = x)$, via Definition 2.4.1, is impossible, since the denominator would be 0. Alternatively, we can think of $P(Y \le y|X = x)$ as a limit:

$$\begin{aligned} P(Y \le y|X = x) &= \lim_{h\to 0} P(Y \le y|x \le X \le x + h) \\ &= \lim_{h\to\infty} \frac{\int_x^{x+h}\int_{-\infty}^{y} f_{X,Y}(t, u)\,du\,dt}{\int_x^{x+h} f_X(t)\,dt} \end{aligned}$$

Evaluating the quotient of the limits gives $\frac{0}{0}$, so l'Hôpital's rule is indicated:

$$P(Y \le y|X = x) = \lim_{h\to 0} \frac{\frac{d}{dh}\int_x^{x+h}\int_{-\infty}^{y} f_{X,Y}(t,u)\,du\,dt}{\frac{d}{dh}\int_x^{x+h} f_X(t)\,dt} \tag{3.11.1}$$

By the fundamental theorem of calculus,

$$\frac{d}{dh}\int_x^{x+h} g(t)\,dt = g(x+h)$$

which simplifies Equation 3.11.1 to

$$P(Y \le y|X = x) = \lim_{h\to 0} \frac{\int_{-\infty}^{y} f_{X,Y}[(x+h),u]\,du}{f_X(x+h)}$$

$$= \frac{\int_{-\infty}^{y} \lim_{h\to 0} f_{X,Y}(x+h,u)\,du}{\lim_{h\to 0} f_X(x+h)} = \int_{-\infty}^{y} \frac{f_{X,Y}(x,u)}{f_X(x)}\,du$$

provided that the limit operation and the integration can be interchanged [see (9) for a discussion of when such an interchange is valid]. It follows from this last expression that $f_{X,Y}(x,y)/f_X(x)$ behaves as a conditional probability density function should, and we are justified in extending Definition 3.11.1 to the continuous case.

EXAMPLE 3.11.5

Let X and Y be continuous random variables with joint pdf

$$f_{X,Y}(x,y) = \begin{cases} \left(\frac{1}{8}\right)(6-x-y), & 0 < x < 2, \quad 2 < y < 4 \\ 0, & \text{elsewhere} \end{cases}$$

Find (a) $f_X(x)$, (b) $f_{Y|x}(y)$, and (c) $P(2 < Y < 3|x = 1)$.

a. From Theorem 3.7.2,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy = \int_2^4 \left(\frac{1}{8}\right)(6-x-y)\,dy$$
$$= \left(\frac{1}{8}\right)(6-2x), \quad 0 < x < 2$$

b. Substituting into the "continuous" statement of Definition 3.11.1, we can write

$$f_{Y|x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\left(\frac{1}{8}\right)(6-x-y)}{\left(\frac{1}{8}\right)(6-2x)}$$
$$= \frac{6-x-y}{6-2x}, \quad 0 < x < 2, \quad 2 < y < 4$$

c. To find $P(2 < Y < 3|x = 1)$ we simply integrate $f_{Y|1}(y)$ over the interval $2 < Y < 3$:

$$P(2 < Y < 3|x = 1) = \int_2^3 f_{Y|1}(y)\,dy$$
$$= \int_2^3 \frac{5 - y}{4}\,dy$$
$$= \frac{5}{8}$$

[A partial check that the derivation of a conditional pdf is correct can be performed by integrating $f_{Y|x}(y)$ over the entire range of Y. That integral should be one. Here, for example, when $x = 1$, $\int_{-\infty}^{\infty} f_{Y|1}(y)\,dy = \int_2^4 [(5 - y)/4]\,dy$ *does* equal one.]

QUESTIONS

3.11.11. Let X be a nonnegative random variable. We say that X *is memoryless* if

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

Show that a random variable with pdf $f_X(x) = (1/\lambda)e^{-x/\lambda}$, $x > 0$, is memoryless.

3.11.12. Given the joint pdf

$$f_{X,Y}(x, y) = 2e^{-(x+y)}, \quad 0 < x < y, \quad y > 0$$

find

(a) $P(Y < 1|X < 1)$
(b) $P(Y < 1|X = 1)$
(c) $f_{Y|x}(y)$
(d) $E(Y|x)$

3.11.13. Find the conditional pdf of Y given x if

$$f_{X,Y}(x, y) = x + y$$

for $0 \leq x \leq 1$ and $0 \leq y \leq 1$.

3.11.14. If

$$f_{X,Y}(x, y) = 2, \quad x \geq 0, \quad y \geq 0, \quad x + y \leq 1$$

show that the conditional pdf of Y given x is uniform.

3.11.15. Suppose that

$$f_{Y|x}(y) = \frac{2y + 4x}{1 + 4x} \quad \text{and} \quad f_X(x) = \frac{1}{3} \cdot (1 + 4x)$$

for $0 < x < 1$ and $0 < y < 1$. Find the marginal pdf for Y.

3.11.16. Suppose that X and Y are distributed according to the joint pdf

$$f_{X,Y}(x, y) = \frac{2}{5} \cdot (2x + 3y), \quad 0 \leq x \leq 1, \quad 0 \leq y \leq 1$$

By Definition 3.12.1,

$$M_X(t) = E(e^{tX}) = \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1 - p)^{n-k}$$

$$= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1 - p)^{n-k} \qquad (3.12.2)$$

To get a closed-form expression for $M_X(t)$—that is, to evaluate the sum indicated in Equation 3.12.2—requires a (hopefully) familiar formula from algebra: According to *Newton's binomial expansion,*

$$(x + y)^n = \sum_{k=0}^{n}\binom{n}{k}x^k y^{n-k} \qquad (3.12.3)$$

for any x and y. Suppose we let $x = pe^t$ and $y = 1 - p$. It follows from Equations 3.12.2 and 3.12.3, then, that

$$M_X(t) = (1 - p + pe^t)^n$$

(Notice in this case that $M_X(t)$ is defined for all values of t).

EXAMPLE 3.12.3

Suppose that Y has an exponential pdf, where $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. Find $M_Y(t)$.

Since the exponential pdf describes a continuous random variable, $M_Y(t)$ is an integral:

$$M_Y(t) = E(e^{tY}) = \int_0^\infty e^{ty} \cdot \lambda e^{-\lambda y}\,dy$$

$$= \int_0^\infty \lambda e^{-(\lambda - t)y}\,dy$$

After making the substitution $u = (\lambda - t)y$, we can write

$$M_Y(t) = \int_{u=0}^{\infty} \lambda e^{-u}\frac{du}{\lambda - t}$$

$$= \frac{\lambda}{\lambda - t}\left[-e^{-u}\Big|_{u=0}^{\infty}\right]$$

$$= \frac{\lambda}{\lambda - t}\left[1 - \lim_{u\to\infty} e^{-u}\right] = \frac{\lambda}{\lambda - t}$$

Here, $M_Y(t)$ is finite and nonzero only when $u = (\lambda - t)y > 0$, which implies that t must be less than λ. For $t \geq \lambda$, $M_Y(t)$ fails to exist.

EXAMPLE 3.12.1

Suppose the random variable X has a *geometric pdf,*

$$p_X(k) = (1 - p)^{k-1}p, \quad k = 1, 2, \ldots$$

(In practice, this is the pdf that models the occurrence of the first success in a series of independent trials, where each trial has a probability p of ending in success [recall Example 3.3.2]). Find $M_X(t)$, the moment-generating function for X.

Since X is discrete, the first part of Definition 3.12.1 applies, so

$$\begin{aligned} M_X(t) = E(e^{tX}) &= \sum_{k=1}^{\infty} e^{tk}(1 - p)^{k-1}p \\ &= \frac{p}{1-p}\sum_{k=1}^{\infty} e^{tk}(1 - p)^k \\ &= \frac{p}{1-p}\sum_{k=1}^{\infty}[(1 - p)e^t]^k \qquad (3.12.1) \end{aligned}$$

The t in $M_X(t)$ can be any number in a neighborhood of zero, as long as $M_X(t) < \infty$. Here, $M_X(t)$ is an infinite sum of the terms $[(1 - p)e^t]^k$, and that sum will be finite only if $(1 - p)e^t < 1$, or, equivalently, if $t < \ln(1/(1 - p))$. It will be assumed, then, in what follows that $0 < t < \ln(1/(1 - p))$.

Recall that

$$\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$$

provided $0 < r < 1$. This formula can be used on Equation 3.12.1, where $r = (1 - p)e^t$ and $0 < t < \ln\left(\frac{1}{(1-p)}\right)$. Specifically,

$$\begin{aligned} M_X(t) &= \frac{p}{1-p}\left(\sum_{k=0}^{\infty}[(1 - p)e^t]^k - [(1 - p)e^t]^0\right) \\ &= \frac{p}{1-p}\left(\frac{1}{1-(1 - p)e^t} - 1\right) \\ &= \frac{pe^t}{1-(1 - p)e^t} \end{aligned}$$

EXAMPLE 3.12.2

Suppose that X is a binomial random variable with pdf

$$p_X(k) = \binom{n}{k}p^k(1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

Find $M_X(t)$.

Find

(a) $f_X(x)$,

(b) $f_{Y|x}(y)$, and

(c) $P\left(\frac{1}{4} \le Y \le \frac{3}{4} | X = \frac{1}{2}\right)$

(d) $E(Y|x)$

3.11.17. If X and Y have the joint pdf

$$f_{X,Y}(x, y) = 2, \quad 0 < x < y < 1$$

find $P\left(0 < X < \frac{1}{2} | Y = \frac{3}{4}\right)$.

3.11.18. Find $P\left(X < 1 | Y = 1\frac{1}{2}\right)$ if X and Y have the joint pdf

$$f_{X,Y}(x, y) = xy/2, \quad 0 < x < y < 2$$

3.11.19. Suppose that X_1, X_2, X_3, X_4, and X_5 have the joint pdf

$$f_{X_1,X_2,X_3,X_4,X_5}(x_1, x_2, x_3, x_4, x_5) = 32x_1x_2x_3x_4x_5$$

for $0 < x_i < 1, i = 1, 2, \ldots, 5$. Find the joint conditional pdf of X_1, X_2, and X_3 given that $X_4 = x_4$ and $X_5 = x_5$.

3.11.20. Suppose the random variables X and Y are jointly distributed according to the pdf

$$f_{X,Y}(x, y) = \frac{6}{7}\left(x^2 + \frac{xy}{2}\right), \quad 0 < x < 1, \quad 0 < y < 2$$

Find

(a) $f_X(x)$

(b) $P(X > 2Y)$

(c) $P\left(Y > 1 | X > \frac{1}{2}\right)$

3.12 MOMENT-GENERATING FUNCTIONS

Finding moments of random variables directly, particularly the higher moments defined in Section 3.6, is conceptually straightforward but can be quite problematic: Depending on the nature of the pdf, integrals and sums of the form $\int_{-\infty}^{\infty} y^r f_Y(y)\,dy$ and $\sum_{\text{all } k} k^r p_X(k)$ can be very difficult to evaluate. Fortunately, an alternative method is available. For many pdfs, we can find a *moment-generating function* (or *mgf*), $M_W(t)$, one of whose properties is that the rth derivative of $M_W(t)$ evaluated at zero is equal to $E(W^r)$.

Calculating a Random Variable's Moment-Generating Function

In principle, what we call a moment-generating function is a direct application of Theorem 3.5.3.

Definition 3.12.1. Let W be a random variable. The *moment-generating function (mgf) for W* is denoted $M_W(t)$ and given by

$$M_W(t) = E(e^{tW}) = \begin{cases} \sum_{\text{all } k} e^{tk} p_W(k) & \text{if } W \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tw} f_W(w)\,dw & \text{if } W \text{ is continuous} \end{cases}$$

at all values of t for which the expected value exists.

EXAMPLE 3.12.4

The normal (or bell-shaped) curve was introduced in Example 3.4.3. Its pdf is the rather cumbersome function

$$f_Y(y) = \left(1/\sqrt{2\pi}\sigma\right)\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \qquad -\infty < y < \infty$$

where $\mu = E(Y)$ and $\sigma^2 = \operatorname{Var}(Y)$. Derive the moment-generating function for this most important of all probability models.

Since Y is a continuous random variable,

$$\begin{aligned} M_Y(t) = E(e^{tY}) &= \left(1/\sqrt{2\pi}\sigma\right)\int_{-\infty}^{\infty} \exp(ty)\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]dy \\ &= \left(1/\sqrt{2\pi}\sigma\right)\int_{-\infty}^{\infty} \exp\left[-\frac{y^2-2\mu y-2\sigma^2 ty+\mu^2}{2\sigma^2}\right]dy \qquad (3.12.4) \end{aligned}$$

Evaluating the integral in Equation 3.12.4 is best accomplished by completing the square of the numerator of the exponent (which means that the square of half the coefficient of y is added and subtracted). That is, we can write

$$\begin{aligned} &y^2 - (2\mu + 2\sigma^2 t)y + (\mu + \sigma^2 t)^2 - (\mu + \sigma^2 t)^2 + \mu^2 \\ &\qquad = (y - (\mu + \sigma^2 t))^2 - \sigma^4 t^2 + 2\mu t\sigma^2 \qquad (3.12.5) \end{aligned}$$

The last two terms on the right-hand side of Equation 3.12.5, though, do not involve y, so they can be factored out of the integral, and Equation 3.12.4 reduces to

$$M_Y(t) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)\left(1/\sqrt{2\pi}\sigma\right)\int_{-\infty}^{\infty}\exp\left[-\frac{1}{2}\left(\frac{y-(\mu+t\sigma^2)}{\sigma}\right)^2\right]dy$$

But, together, the latter two factors equal one (why?), implying that the moment-generating function for a normally distributed random variable is given by

$$M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$$

QUESTIONS

3.12.1. Let X be a random variable with pdf $p_X(k) = 1/n$, for $k = 0, 1, 2, \ldots, n-1$ and 0 otherwise. Show that $M_X(t) = \dfrac{1-e^{nt}}{n(1-e^t)}$.

3.12.2. Two chips are drawn at random and without replacement from an urn that contains five chips, numbered 1 through 5. If the sum of the chips drawn is even, the random variable X equals 5; if the sum of the chips drawn is odd, $X = -3$. Find the moment-generating function for X.

3.12.3. Find the expected value of e^{3X} if X is a binominal random variable with $n = 10$ and $p = \frac{1}{3}$.

3.12.4. Find the moment-generating function for the discrete random variable X whose probability function is given by

$$p_X(k) = \left(\frac{3}{4}\right)^k \left(\frac{1}{4}\right), \quad k = 0, 1, 2, \ldots$$

3.12.5. Which pdfs would have the following moment-generating functions:

(a) $M_Y(t) = e^{6t^2}$

(b) $M_Y(t) = 2/(2 - t)$

(c) $M_X(t) = \left(\frac{1}{2} + \frac{1}{2}e^t\right)^4$

(d) $M_X(t) = 0.3e^t/(1 - 0.7e^t)$

3.12.6. Let X have pdf

$$f_Y(y) = \begin{cases} y, & 0 \le y \le 1 \\ 2 - y, & 1 \le y \le 2 \\ 0, & \text{elsewhere} \end{cases}$$

Find $M_Y(t)$.

3.12.7. A random variable X is said to have a *Poisson distribution* if $p_X(k) = P(X = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$. Find the moment-generating function for a Poisson random variable. *Hint:* Use the fact that

$$e^r = \sum_{k=0}^{\infty} \frac{r^k}{k!}$$

3.12.8. Let Y be a continuous random variable with $f_Y(y) = ye^{-y}$, $0 \le y$. Show that $M_Y(t) = \dfrac{1}{(1 - t)^2}$.

Using Moment-Generating Functions to Find Moments

Having practiced *finding* the functions $M_X(t)$ and $M_Y(t)$, we now turn to the theorem that spells out their relationship to X^r and Y^r.

Theorem 3.12.1. *Let W be a random variable with probability density function $f_W(w)$. [If W is continuous, $f_W(w)$ must be sufficiently smooth to allow the order of differentiation and integration to be interchanged.] Let $M_W(t)$ be the moment-generating function for W. Then, provided the rth moment exists,*

$$M_W^{(r)}(0) = E(W^r)$$

Proof. We will verify the theorem for the continuous case where r is either 1 or 2. The extensions to discrete random variables and to an arbitrary positive integer r are straightforward.

For $r = 1$,

$$M_Y^{(1)}(0) = \frac{d}{dt}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} \frac{d}{dt} e^{ty} f_Y(y)\,dy\bigg|_{t=0}$$

$$= \int_{-\infty}^{\infty} ye^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} ye^{0\cdot y} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} y f_Y(y)\,dy = E(Y)$$

For $r = 2$,

$$M_Y^{(2)}(0) = \frac{d^2}{dt^2}\int_{-\infty}^{\infty} e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} \frac{d^2}{dt^2} e^{ty} f_Y(y)\,dy\bigg|_{t=0}$$

$$= \int_{-\infty}^{\infty} y^2 e^{ty} f_Y(y)\,dy\bigg|_{t=0} = \int_{-\infty}^{\infty} y^2 e^{0\cdot y} f_Y(y)\,dy$$

$$= \int_{-\infty}^{\infty} y^2 f_Y(y)\,dy = E(Y^2)$$

☐

EXAMPLE 3.12.5

For a geometric random variable X with pdf

$$p_X(k) = (1 - p)^{k-1} p, \quad k = 1, 2, \ldots$$

we saw in Example 3.12.1 that

$$M_X(t) = pe^t[1 - (1 - p)e^t]^{-1}$$

Find the expected value of X by differentiating its moment-generating function.

Using the product rule, we can write the first derivative of $M_X(t)$ as

$$M_X^{(1)}(t) = pe^t(-1)(1 - (1 - p)e^t)^{-2}(-1)(1 - p)e^t + [1 - (1 - p)e^t]^{-1}pe^t$$

$$= \frac{p(1 - p)e^{2t}}{[1 - (1 - p)e^t]^2} + \frac{pe^t}{1 - (1 - p)e^t}$$

Setting $t = 0$ shows that $E(X) = \frac{1}{p}$:

$$M_X^{(1)}(0) = E(X) = \frac{p(1 - p)e^{2\cdot 0}}{[1 - (1 - p)e^0]^2} + \frac{pe^0}{1 - (1 - p)e^0}$$

$$= \frac{p(1 - p)}{p^2} + \frac{p}{p}$$

$$= \frac{1}{p}$$

EXAMPLE 3.12.6

Find the expected value of an exponential random variable with pdf

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0$$

Use the fact that

$$M_Y(t) = \lambda(\lambda - t)^{-1}$$

(as shown in Example 3.12.3).

Differentiating $M_Y(t)$ gives

$$\begin{aligned} M_Y^{(1)}(t) &= \lambda(-1)(\lambda - t)^{-2}(-1) \\ &= \frac{\lambda}{(\lambda - t)^2} \end{aligned}$$

Set $t = 0$. Then

$$M_Y^{(1)}(0) = \frac{\lambda}{(\lambda - 0)^2}$$

implying that

$$E(Y) = \frac{1}{\lambda}$$

EXAMPLE 3.12.7

Find an expression for $E(X^k)$ if the moment-generating function for X is given by

$$M_X(t) = (1 - p_1 - p_2) + p_1e^t + p_2e^{2t}$$

The only way to deduce a formula for an arbitrary moment such as $E(X^k)$ is to calculate the first couple moments and look for a pattern that can be generalized. Here,

$$M_X^{(1)}(t) = p_1e^t + 2p_2e^{2t}$$

so

$$\begin{aligned} E(X) = M_X^{(1)}(0) &= p_1e^0 + 2p_2e^{2\cdot 0} \\ &= p_1 + 2p_2 \end{aligned}$$

Taking the second derivative, we see that

$$M_X^{(2)}(t) = p_1e^t + 2^2p_2e^{2t}$$

implying that

$$E(X^2) = M_X^{(2)}(0) = p_1 e^0 + 2^2 p_2 e^{2 \cdot 0}$$
$$= p_1 + 2^2 p_2$$

Clearly, each successive differentiation will leave the p_1 term unaffected but will multiply the p_2 term by two. Therefore,

$$E(X^k) = M_X^{(k)}(0) = p_1 + 2^k p_2$$

Using Moment-Generating Functions to Find Variances

In addition to providing a useful technique for calculating $E(W^r)$, moment-generating functions can also find variances, because

$$\operatorname{Var}(W) = E(W^2) - [E(W)]^2 \tag{3.12.6}$$

for any random variable W (recall Theorem 3.6.1). Other useful "descriptors" of pdfs can also be reduced to combinations of moments. The *skewness* of a distribution, for example, is a function of $E[W - \mu)^3]$, where $\mu = E(W)$. But

$$E[(W - \mu)^3] = E(W^3) - 3E(W^2)E(W) + 2[E(W)]^3$$

In many cases, finding $E[(W - \mu)^2]$ or $E[(W - \mu)^3]$ could be quite difficult if moment-generating functions were not available.

EXAMPLE 3.12.8

We know from Example 3.12.2 that if X is a binomial random variable with parameters n and p, then

$$M_X(t) = (1 - p + pe^t)^n$$

Use $M_X(t)$ to find the variance of X.

The first two derivatives of $M_X(t)$ are

$$M_X^{(1)}(t) = n(1 - p + pe^t)^{n-1} \cdot pe^t$$

and

$$M_X^{(2)}(t) = pe^t \cdot n(n - 1)(1 - p + pe^t)^{n-2} \cdot pe^t + n(1 - p + pe^t)^{n-1} \cdot pe^t$$

Setting $t = 0$ gives

$$M_X^{(1)}(0) = np = E(X)$$

and

$$M_X^{(2)}(0) = n(n - 1)p^2 + np = E(X^2)$$

From Equation 3.12.6, then,

$$\begin{aligned}\text{Var}(X) &= n(n-1)p^2 + np - (np)^2 \\ &= np(1-p)\end{aligned}$$

(the same answer we found in Example 3.9.8).

EXAMPLE 3.12.9

A discrete random variable X is said to have a *Poisson* distribution if

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

(An example of such a distribution is the mortality data described in Case Study 3.3.1.) It can be shown (see Question 3.12.7) that the moment-generating function for a Poisson random variable is given by

$$M_X(t) = e^{-\lambda+\lambda e^t}$$

Use $M_X(t)$ to find $E(X)$ and $\text{Var}(X)$.

Taking the first derivative of $M_X(t)$ gives

$$M_X^{(1)}(t) = e^{-\lambda+\lambda e^t} \cdot \lambda e^t$$

so

$$\begin{aligned}E(X) = M_X^{(1)}(0) &= e^{-\lambda+\lambda e^0} \cdot \lambda e^0 \\ &= \lambda\end{aligned}$$

Applying the product rule to $M_X^{(1)}(t)$ yields the second derivative,

$$M_X^{(2)}(t) = e^{-\lambda+\lambda e^t} \cdot \lambda e^t + \lambda e^t e^{-\lambda+\lambda e^t} \cdot \lambda e^t$$

For $t = 0$,

$$\begin{aligned}M_X^{(2)}(0) = E(X^2) &= e^{-\lambda+\lambda e^0} \cdot \lambda e^0 + \lambda e^0 \cdot e^{-\lambda+\lambda e^0} \cdot \lambda e^0 \\ &= \lambda + \lambda^2\end{aligned}$$

The variance of a Poisson random variable, then, proves to be the same as its mean:

$$\begin{aligned}\text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= M_X^{(2)}(0) - \left[M_X^{(1)}(0)\right]^2 \\ &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda\end{aligned}$$

QUESTIONS

3.12.9. Calculate $E(Y^3)$ for a random variable whose moment-generating function is $M_Y(t) = e^{t^2/2}$.

3.12.10. Find $E(Y^4)$ if Y is an exponential random variable with $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$.

3.12.11. The form of the moment-generating function for a normal random variable is $M_Y(t) = e^{at+b^2t^2/2}$ (recall Example 3.12.4). Differentiate $M_Y(t)$ to verify that $a = E(Y)$ and $b^2 = \text{Var}(Y)$.

3.12.12. What is $E(Y^4)$ if the random variable Y has moment-generating function $M_Y(t) = (1 - \alpha t)^{-k}$?

3.12.13. Find $E(Y^2)$ if the moment-generating function for Y is given by $M_Y(t) = e^{-t+4t^2}$. Use Example 3.12.4 to find $E(Y^2)$ without taking any derivatives. *Hint*: Recall Theorem 3.6.1

3.12.14. Find an expression for $E(Y^k)$ if $M_Y(t) = (1 - t/\lambda)^{-r}$, where λ is any positive real number and r is a positive integer.

3.12.15. Use $M_Y(t)$ to find the expected value of the uniform random variable described in Question 3.12.1.

3.12.16. Find the variance of Y if $M_Y(t) = e^{2t}/(1 - t^2)$.

Using Moment-Generating Functions to Identify pdf's

Finding moments is not the only application of moment-generating functions. They are also used to identify the pdf of *sums* of random variables—that is, finding $f_W(w)$, where $W = W_1 + W_2 + \cdots + W_n$. Their assistance in the latter is particularly important for two reasons: (1) Many statistical procedures are defined in terms of sums, and (2) alternative methods for deriving $f_{W_1+W_2+\cdots+W_n}(w)$ are extremely cumbersome.

The next two theorems give the background results necessary for deriving $f_W(w)$. Theorem 3.12.2 states a key uniqueness property of moment-generating functions: If W_1 and W_2 are random variables with the same mgfs, they must necessarily have the same pdfs. In practice, applications of Theorem 3.12.2 typically rely on one or both of the algebraic properties cited in Theorem 3.12.3.

Theorem 3.12.2. *Suppose that W_1 and W_2 are random variables for which $M_{W_1}(t) = M_{W_2}(t)$ for some interval of t's containing 0. Then $f_{W_1}(w) = f_{W_2}(w)$.*

Proof. See (97). □

Theorem 3.12.3.

a. *Let W be a random variable with moment-generating function $M_W(t)$. Let $V = aW + b$. Then*

$$M_V(t) = e^{bt} M_W(at)$$

b. *Let $W_1, W_2, \ldots, W_n$ be independent random variables with moment-generating functions $M_{W_1}(t), M_{W_2}(t), \ldots,$ and $M_{W_n}(t)$, respectively. Let $W = W_1 + W_2 + \cdots + W_n$. Then*

$$M_W(t) = M_{W_1}(t) \cdot M_{W_2}(t) \cdots M_{W_n}(t)$$

Proof. The proof is left as an exercise. □

EXAMPLE 3.12.10

Suppose that X_1 and X_2 are two independent Poisson random variables with parameters λ_1 and λ_2, respectively. That is,

$$p_{X_1}(k) = P(X_1 = k) = \frac{e^{-\lambda_1}\lambda_1 k}{k!}, \quad k = 0, 1, 2, \ldots$$

and

$$p_{X_2}(k) = P(X_2 = k) = \frac{e^{-\lambda_2}\lambda_2 k}{k!}, \quad k = 0, 1, 2\ldots$$

Let $X = X_1 + X_2$. What is the pdf for X?

According to Example 3.12.9, the moment-generating functions for X_1 and X_2 are

$$M_{X_1}(t) = e^{-\lambda_1+\lambda_1 e^t}$$

and

$$M_{X_2}(t) = e^{-\lambda_2+\lambda_2 e^t}$$

Moreover, if $X = X_1 + X_2$, then by Part b of Theorem 3.12.3,

$$\begin{aligned} M_X(t) &= M_{X_1}(t) \cdot M_{X_2}(t) \\ &= e^{-\lambda_1+\lambda_1 e^t} \cdot e^{-\lambda_2+\lambda_2 e^t} \\ &= e^{-(\lambda_1+\lambda_2)+(\lambda_1+\lambda_2)e^t} \end{aligned} \tag{3.12.7}$$

But, by inspection, Equation 3.12.7 is the moment-generating function that a Poisson random variable with $\lambda = \lambda_1 + \lambda_2$ would have. It follows, then, by Theorem 3.12.2 that

$$p_X(k) = \frac{e^{-(\lambda_1+\lambda_2)}(\lambda_1 + \lambda_2)^k}{k!}, \quad k = 0, 1, 2, \ldots$$

Comment. The Poisson random variable reproduces itself in the sense that the sum of independent Poissons is also a Poisson. A similar property holds for independent normal random variables (see Question 3.12.19) and, under certain conditions, for independent binomial random variables (recall Example 3.8.1).

EXAMPLE 3.12.11

We saw in Example 3.12.4 that a normal random variable, Y, with mean μ and variance σ^2 has pdf

$$f_Y(y) = (1/\sqrt{2\pi}\sigma)\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right], \quad -\infty < y < \infty$$

and mgf

$$M_Y(t) = e^{\mu t + \sigma^2 t^2/2}$$

By definition, a *standard normal random variable* is a normal random variable for which $\mu = 0$ and $\sigma = 1$. Denoted Z, the pdf and mgf for a standard normal random variable are $f_Z(z) = (1/\sqrt{2\pi})e^{-z^2/2}, -\infty < z < \infty$ and $M_Z(t) = e^{t^2/2}$, respectively. Show that the ratio

$$\frac{Y - \mu}{\sigma}$$

is a standard normal random variable, Z.

Write $\frac{Y - \mu}{\sigma}$ as $\frac{1}{\sigma}Y - \frac{\mu}{\sigma}$. By Part a of Theorem 3.12.3,

$$\begin{aligned} M_{(Y-\mu)/\sigma}(t) &= e^{-\mu t/\sigma} M_Y\left(\frac{t}{\sigma}\right) \\ &= e^{-\mu t/\sigma} e^{(\mu t/\sigma + \sigma^2 (t/\sigma)^2/2)} \\ &= e^{t^2/2} \end{aligned}$$

But $M_Z(t) = e^{t^2/2}$ so it follows from Theorem 3.12.2 that the pdf for $\frac{Y - \mu}{\sigma}$ is the same as $f_z(z)$. (We call $\frac{Y - \mu}{\sigma}$ a *Z transformation*. Its importance will become evident in Chapter 4.)

QUESTIONS

3.12.17. Use Theorem 3.12.3(a) and Question 3.12.8 to find the moment-generating function of the random variable Y, where $f_Y(y) = \lambda y e^{-\lambda y}$, $y \geq 0$.

3.12.18. Let Y_1, Y_2, and Y_3 be independent random variables, each having the pdf of Question 3.12.17. Use Theorem 3.12.3(b) to find the moment-generating function of $Y_1 + Y_2 + Y_3$. Compare your answer to the moment-generating function in Question 3.12.14.

3.12.19. Use Theorems 3.12.2 and 3.12.3 to determine which of the following statements is true:

(a) The sum of two independent Poisson random variables has a Poisson distribution.

(b) The sum of two independent exponential random variables has an exponential distribution.

(c) The sum of two independent normal random variables has a normal distribution.

3.12.20. Calculate $P(X \leq 2)$ if $M_X(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$.

3.12.21. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a normal distribution with mean μ and standard deviation σ. Use moment-generating functions to deduce the pdf of $\overline{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$.

3.12.22. Suppose the moment-generating function for a random variable W is given by

$$M_W(t) = e^{-3+3e^t} \cdot \left(\frac{2}{3} + \frac{1}{3}e^t\right)^4$$

Calculate $P(W \leq 1)$. *Hint:* Write W as a sum.

3.12.23. Suppose that X is a Poisson random variable, where $p_X(k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, \ldots$.

(a) Does the random variable $W = 3X$ have a Poisson distribution?

(b) Does the random variable $W = 3X + 1$ have a Poisson distribution?

3.12.24. Suppose that Y is a normal variable, where $f_Y(y) = (1/\sqrt{2\pi}\sigma)\exp\left[-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right]$, $-\infty < y < \infty$.

(a) Does the random variable $W = 3Y$ have a normal distribution?

(b) Does the random variable $W = 3Y + 1$ have a normal distribution?

TAKING A SECOND LOOK AT STATISTICS (INTERPRETING MEANS)

One of the most important ideas coming out of Chapter 3 is the notion of the *expected value* (or *mean*) of a random variable. Defined in Section 3.5 as a number that reflects the "center" of a pdf, the expected value (μ) was originally introduced for the benefit of gamblers. It spoke directly to one of their most fundamental questions—How much will I win or lose, *on the average,* if I play a certain game? (Actually, the real question they probably had in mind was "How much are *you* going to *lose,* on the average?") Despite having had such a selfish, materialistic, gambling-oriented *raison d'etre,* the expected value was quickly embraced by (respectable) scientists and researchers of all persuasions as a preeminently useful descriptor of a distribution. Today, it would not be an exaggeration to claim that the majority of *all* statistical analyses focus on either (1) the expected value of a single random variable or (2) comparing the expected values of two or more random variables.

In the lingo of applied statistics, there are actually two fundamentally different types of "means"—*population means* and *sample means.* The term "population mean" is a synonym for what mathematical statisticians would call an expected value—that is, a population mean (μ) is a weighted average of the possible values associated with a theoretical probability model, either $p_X(k)$ or $f_Y(y)$, depending on whether the underlying random variable is discrete or continuous. A *sample mean* is the arithmetic average of a set of measurements. If, for example, n observations—$y_1, y_2, \ldots, y_n$—are taken on a continuous random variable Y, the sample mean is denoted $\bar{y}$, where

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Conceptually, sample means are *estimates* of population means, where the "quality" of the estimation is a function of (1) the sample size and (2) the standard deviation (σ) associated with the individual measurements. Intuitively, as the sample size gets larger and/or the standard deviation gets smaller, the approximation will tend to get better.

Interpreting means (either $\bar{y}$ or μ) is not always easy. To be sure, what they imply *in principle* is clear enough—both $\bar{y}$ and μ are measuring the centers of their respective distributions. Still, many a wrong conclusion can be traced directly to researchers misunderstanding the value of a mean. Why? Because the distributions that $\bar{y}$ and/or μ are *actually* representing may be dramatically different than the distributions we *think* they are representing.

An interesting case in point arises in connection with SAT scores. Each Fall the average SATs earned by students in each of the fifty states and the District of Columbia are released by the Educational Testing Service (ETS). With "accountability" being one of the new paradigms and buzz words associated with K-12 education, SAT scores have become highly politicized. At the national level, Democrats and Republicans each campaign on their own versions of education reform, fueled in no small measure by scores on standardized exams, SATs included; at the state level, legislatures often modify education budgets in response to how well or how poorly their students performed the year before. Does it make sense, though, to use SAT averages to characterize the quality of a state's education system? Absolutely not! Averages of this sort refer to very different distributions from state to state. Any attempt to interpret them at face value will necessarily be misleading.

One such state-by-state SAT comparison that appeared in the mid-90s is reproduced in Table 3.13.1 (128). Notice that Tennessee's entry is 1023, which is the tenth highest average listed. Does it follow that Tennessee's educational system is among the best in the nation? Probably not. Most independent assessments of K-12 education rank Tennessee's schools among the weakest in the nation, not among the best. If those opinions are accurate, why do Tennessee's students do so well on the SAT?

The answer to that question lies in the academic profiles of the students who take the SAT in Tennessee. Most college-bound students in that state apply exlusively to schools in the South and the Midwest, where admissions are based on the ACT, not the SAT. The SAT is primarily used by private schools, where admissions tend to be more competitive. As a result, the students in Tennessee who take the SAT are not representative of the entire population of students in that state. A disproportionate number are exceptionally strong academically, those being the students who feel that they have the ability to be competitive at Ivy League-type schools. The number 1023, then, is the average of *something* (in this case, an elite subset of all Tennessee students), but it does not correspond to the center of the SAT distribution for *all* Tennessee students.

The moral here is that analyzing data effectively requires that we look beyond the obvious. What we learn in Chapter 3 about random variables and probability distributions and expected values is helpful only if we take the time to learn about the context and

TABLE 3.13.1

State	Average SAT Score	State	Average SAT Score
AK	911	MT	986
AL	1011	NE	1025
AZ	939	NV	913
AR	935	NH	924
CA	895	NJ	893
CO	969	NM	1003
CT	898	NY	888
DE	892	NC	860
DC	849	ND	1056
FL	879	OH	966
GA	844	OK	1019
HI	881	OR	927
ID	969	PA	879
IL	1024	RI	882
IN	876	SC	838
IA	1080	SD	1031
KS	1044	TN	1023
KY	997	TX	886
LA	1011	UT	1067
ME	883	VT	899
MD	908	VA	893
MA	901	WA	922
MI	1009	WV	921
MN	1057	WI	1044
MS	1013	WY	980
MO	1017		

the idiosyncracies of the phenomenon being studied. To do otherwise is likely to lead to conclusions that are, at best, superficial and, at worst, incorrect.

NDIX 3.A.1 MINITAB APPLICATIONS

Numerous software packages are available for doing a variety of probability and statistical calculations. Among the first to be developed and one that continues to be very popular is MINITAB. Beginning here, we will include at the ends of certain chapters a short discussion of MINITAB solutions to some of the problems that were discussed in that chapter. What other software packages can do and the ways their outputs are formatted are likely to be quite similar.

Contained in MINITAB are subroutines that can do some of the more important pdf and cdf computations described in Sections 3.3 and 3.4. In the case of binomial random variables, for instance, the statements

```
MTB  > pdf k;
SUBC > binomial n p.
```

and

```
MTB  > cdf k;
SUBC > binomial n p.
```

will calculate $\binom{n}{k}p^k(1-p)^{n-k}$ and $\sum_{r=0}^{k}\binom{n}{r}p^r(1-p)^{n-r}$, respectively. Figure 3.A.1.1 shows the MINITAB program for doing the cdf calculation ($= P(X \leq 15)$) asked for in Part a of Example 3.2.2.

The commands `pdf k` and `cdf k` can be run on many of the probability models most likely to be encountered in real-world problems. Those on the list that we have already seen are the binomial, Poisson, normal, uniform, and exponential distributions.

```
MTB > cdf 15;
SUBC > binomial 30 0.60.
Cumulative Distribution Function
Binomial with n = 30 and p = 0.600000
        x     P(X <= x)
    15.00      0.1754
```

FIGURE 3.A.1.1

For discrete random variables, the cdf can be printed out in its entirety (that is, for every integer) by deleting the argument k and using the command `MTB < cdf;`. Typical is the output in Figure 3.A.1.2, corresponding to the cdf for a binomial random variable with $n = 4$ and $p = \frac{1}{6}$.

```
MTB > cdf;
SUBC > binomial 4 0.167.
Cumulative Distribution Function
Binomial with n = 4 and p = 0.167000
         x     P( X <= x)
         0         0.4815
         1         0.8676
         2         0.9837
         3         0.9992
         4         1.0000
```

FIGURE 3.A.1.2

Also available is an *inverse cdf* command, which in the case of a continuous random variable Y and a specified probability p identifies the value y having the property that $P(Y \le y) = F_Y(Y) = p$. For example, if $p = 0.60$ and Y is an exponential random variable with pdf $f_Y(y) = e^{-y}$, $y > 0$, the value $y = 0.9163$ has the property that $P(Y \le 0.9163) = F_Y(0.9163) = 0.60$. That is,

$$F_Y(0.9163) = \int_0^{0.9163} e^{-y}\,dy = 0.60$$

With MINITAB the number 0.9163 is found by using the command `MTB > invcdf 0.60` (see Figure 3.A.1.3).

```
MTB > invcdf 0.60;
SUBC> exponential 1.
Inverse Cumulative Distribution Function
Exponential with mean = 1.00000
P(X <= x)       x
    0.6000      0.9163
```

FIGURE 3.A.1.3

CHAPTER 4

Special Distributions

L. A. J. Quetelet

Quételet.

Although he maintained lifelong literary and artistic interests, Quetelet's mathematical talents led him to a doctorate from the University of Ghent and from there to a college teaching position in Brussels. In 1833 he was appointed astronomer at the Brussels Royal Observatory, after having been largely responsible for its founding. His work with the Belgian census marked the beginning of his pioneering efforts in what today would be called mathematical sociology. Quetelet was well-known throughout Europe in scientific and literary circles: At the time of his death he was a member of more than one hundred learned societies.

—Lambert Adolphe Jacques Quetelet (1796–1874)

INTRODUCTION

To "qualify" as a probability model, a function defined over a sample space S needs to satisfy only two criteria: (1) It must be nonnegative for all outcomes in S, and (2) it must sum or integrate to one. That means, for example, that $f_Y(y) = \frac{y}{4} + \frac{7y^3}{2}, 0 \le y \le 1$ can be considered a pdf because $f_Y(y) \ge 0$ for all $0 \le y \le 1$ and $\int_0^1 \left(\frac{y}{4} + \frac{7y^3}{2}\right) dy = 1$.

It certainly does not follow, though, that every $f_Y(y)$ and $p_X(k)$ that satisfy these two criteria would actually be used as probability models. A pdf has practical significance only if it does, indeed, model the probabilistic behavior of real-world phenomena. In point of fact, only a handful of functions do [and $f_Y(y) = \frac{y}{4} + \frac{7y^3}{2}, 0 \le y \le 1$ is not one of them!].

Whether a probability function—say, $f_Y(y)$—adequately models a given phenomenon ultimately depends on whether the physical factors that influence the value of Y parallel the mathematical assumptions implicit in $f_Y(y)$. Surprisingly, many measurements (i.e., random variables) that seem to be very different are actually the consequence of the same set of assumptions (and will, therefore, be modeled by the same pdf). That said, it makes sense to single out these "real-world" pdf's and investigate their properties in more detail. This, of course, is not an idea we are seeing for the first time—recall the attention given to the binomial and hypergeometric distributions in Section 3.2.

Chapter 4 continues in the spirit of Section 3.2 by examining five other widely used models. Three of the five are discrete; the other two are continuous. One of the continuous pdf's is the normal (or Gaussian) distribution, which, by far, is the most important of all probability models. As we will see, the normal "curve" figures prominently in every chapter from this point on.

Examples play a major role in Chapter 4. The only way to appreciate fully the generality of a probability model is to look at some of its specific applications. Included in this chapter are case studies ranging from the discovery of alpha-particle radiation to an early ESP experiment to an analysis of pregnancy durations to counting bug parts in peanut butter.

THE POISSON DISTRIBUTION

The binomial distribution problems that appeared in Section 3.2 all had relatively small values for n, so evaluating $p_X(k) = P(X = k) = \binom{n}{k} p^k(1 - p)^{n-k}$ was not particularly difficult. But suppose n were 1000 and k, 500. Evaluating $p_X(500)$ would be a formidable task for many handheld calculators, even today. Two hundred years ago, the prospect of doing cumbersome binomial calculations *by hand* was a catalyst for mathematicians to develop some easy-to-use approximations. One of the first such approximations was the *Poisson limit*, which eventually gave rise to the *Poisson distribution*. Both are described in Section 4.2.

Simeon Denis Poisson (1781–1840) was an eminent French mathematician and physicist, an academic administrator of some note, and, according to an 1826 letter from the mathematician Abel to a friend, Poisson was a man who knew "how to behave with a great deal of dignity." One of Poisson's many interests was the application of probability to the law, and in 1837 he wrote *Recherches sur la Probabilite de Jugements*. Included in the latter is a limit for $p_X(k) = \binom{n}{k} p^k(1 - p)^{n-k}$ that holds when n approaches ∞, p approaches 0, and np remains constant. In practice, Poisson's limit is used to approximate hard-to-calculate binomial probabilities where the values of n and p reflect the conditions of the limit—that is, when n is large and p is small.

The Poisson Limit

Deriving an asymptotic expression for the binomial probability model is a straightforward exercise in calculus, given that np is to remain fixed as n increases.

Theorem 4.2.1. *Suppose X is a binomial random variable, where*

$$P(X = k) = p_X(k) = \binom{n}{k} p^k(1 - p)^{n-k}, \quad k = 0, 1, \ldots, n$$

If $n \to \infty$ and $p \to 0$ in such a way that $\lambda = np$ remains constant, then

$$\lim_{\substack{n\to\infty \\ p\to 0 \\ np=\text{const.}}} P(X = k) = \lim_{\substack{n\to\infty \\ p\to 0 \\ np=\text{const.}}} \binom{n}{k} p^k(1 - p)^{n-k} = \frac{e^{-np}(np)^k}{k!}$$

Proof. We begin by rewriting the binomial probability in terms of λ:

$$\begin{aligned}\lim_{n\to\infty} \binom{n}{k} p^k(1 - p)^{n-k} &= \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{n\to\infty} \frac{n!}{k!(n - k)!} \lambda^k \left(\frac{1}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^{-k} \left(1 - \frac{\lambda}{n}\right)^n \\ &= \frac{\lambda^k}{k!} \lim_{n\to\infty} \frac{n!}{(n - k)!} \frac{1}{(n - \lambda)^k} \left(1 - \frac{\lambda}{n}\right)^n\end{aligned}$$

But since $[1 - (\lambda/n)]^n \to e^{-\lambda}$ as $n \to \infty$, we need only show that

$$\frac{n!}{(n - k)!(n - \lambda)^k} \to 1$$

to prove the theorem. However, note that

$$\frac{n!}{(n - k)!(n - \lambda)^k} = \frac{n(n - 1)\cdots(n - k + 1)}{(n - \lambda)(n - \lambda)\cdots(n - \lambda)}$$

a quantity that, indeed, tends to 1 as $n \to \infty$ (since λ remains constant). ☐

EXAMPLE 4.2.1

Theorem 4.2.1 is an *asymptotic* result. Left unanswered is the question of the relevance of the Poisson limit for *finite* n and p. That is, how large does n have to be and how small does p have to be before $e^{-np}(np)^k/k!$ becomes a good approximation to the binomial probability, $p_X(k)$?

Since "good approximation" is undefined, there is no way to answer that question in any completely specific way. Tables 4.2.1 and 4.2.2, though, offer a partial solution by comparing the closeness of the approximation for two particular sets of values for n and p. In both cases $\lambda = np$ is equal to one, but in the former, n is set equal to five—in the latter, to one hundred. We see in Table 4.2.1 ($n = 5$) that for some k the agreement between the binomial probability and Poisson's limit is not very good. If n is as large as one hundred, though (Table 4.2.2), the agreement is remarkably good for all k.

TABLE 4.2.1: Binomial Probabilities and Poisson Limits; $n = 5$ and $p = \frac{1}{5}$ ($\lambda = 1$)

k	$\binom{5}{k}(0.2)^k(0.8)^{5-k}$	$\frac{e^{-1}(1)^k}{k!}$
0	0.328	0.368
1	0.410	0.368
2	0.205	0.184
3	0.051	0.061
4	0.006	0.015
5	0.000	0.003
6+	0	0.001
	1.000	1.000

TABLE 4.2.2: Binomial Probabilities and Poisson Limits; $n = 100$ and $p = \frac{1}{100}$ ($\lambda = 1$)

k	$\binom{100}{k}(0.01)^k(0.99)^{100-k}$	$\frac{e^{-1}(1)^k}{k!}$
0	0.366032	0.367879
1	0.369730	0.367879
2	0.184865	0.183940
3	0.060999	0.061313
4	0.014942	0.015328
5	0.002898	0.003066
6	0.000463	0.000511
7	0.000063	0.000073
8	0.000007	0.000009
9	0.000001	0.000001
10	0.000000	0.000000
	1.000000	0.999999

EXAMPLE 4.2.2

Shadyrest Hospital draws its patients from a rural area that has twelve thousand elderly residents. The probability that any one of the twelve thousand will have a heart attack on any given day and will need to be connected to a special cardiac monitoring machine has been estimated to be one in eight thousand. Currently, the hospital has three such machines. What is the probability that equipment will be inadequate to meet tomorrow's emergencies?

Let X denote the number of residents who will need the cardiac machine tomorrow. Note that X is a binomial random variable based on a large $n(= 12{,}000)$ and a small $p\left(= \frac{1}{8000}\right)$. As such, Poisson's limit can be used to approximate $p_X(k)$ for any k. In particular,

$$\begin{aligned}
P(\text{Shadyrest's cardiac facilities are inadequate}) &= P(X > 3) \\
&= 1 - P(X \leq 3) \\
&= 1 - \sum_{k=0}^{3} \binom{12{,}000}{k} \left(\frac{1}{8000}\right)^k \left(\frac{7999}{8000}\right)^{12{,}000-k} \\
&\doteq 1 - \sum_{k=0}^{3} \frac{e^{-1.5}(1.5)^k}{k!} \\
&= 0.0656
\end{aligned}$$

where $\lambda = np = 12{,}000(\frac{1}{8000}) = 1.5$. On the average, then, Shadyrest will not be able to meet all the cardiac needs of its clientele once every fifteen or sixteen days. (Based on the binomial and Poisson limit comparisons shown on page 276, we would expect the approximation here to be excellent—n (= 12,000) is much larger and $p(= \frac{1}{8000})$ is much smaller than their counterparts in Table 4.2.2, so the conditions of Theorem 4.2.1 are more nearly satisfied.)

CASE STUDY 4.2.1

Leukemia is a rare form of cancer whose cause and mode of transmission remain largely unknown. While evidence abounds that excessive exposure to radiation can increase a person's risk of contracting the disease, it is at the same time true that most cases occur among persons whose history contains no such overexposure. A related issue, one maybe even more basic than the causality question, concerns the *spread* of the disease. It is safe to say that the prevailing medical opinion is that most forms of leukemia are not contagious—still, the hypothesis persists that some forms of the disease, particularly the childhood variety, may be. What continues to fuel this speculation are the discoveries of so-called "leukemia clusters," aggregations in time and space of unusually large numbers of cases.

To date, one of the most frequently cited leukemia clusters in the medical literature occurred during the late 1950s and early 1960s in Niles, Illinois, a suburb of Chicago (74). In the $5\frac{1}{3}$-year period from 1956 to the first four months of 1961, physicians in Niles reported a total of eight cases of leukemia among children less than fifteen years of age. The number at risk (that is, the number of residents in that age range) was 7076. To assess the likelihood of that many cases occurring in such a small population, it is necessary to look first at the leukemia incidence in neighboring towns. For all of Cook county, excluding Niles, there were 1,152,695 children less than 15 years of age—and among those, 286 diagnosed cases of leukemia. That gives an average $5\frac{1}{3}$-year leukemia rate of 24.8 cases per 100,000:

$$\frac{286 \text{ cases for } 5\frac{1}{3} \text{ years}}{1{,}152{,}695 \text{ children}} \times \frac{100{,}000}{100{,}000} = 24.8 \text{ cases}/100{,}000 \text{ children in } 5\frac{1}{3} \text{ years}$$

(Continued on next page)

Now, imagine the 7076 children in Niles to be a series of $n = 7076$ (independent) Bernoulli trials, each having a probability of $p = 24.8/100{,}000 = 0.000248$ of contracting leukemia. The question then becomes, given an n of 7076 and a p of 0.000248, how likely is it that eight "successes" would occur? (The expected number, of course, would be $7076 \times 0.000248 = 1.75$.) Actually, for reasons that will be elaborated on in Chapter 6, it will prove more meaningful to consider the related event, eight *or more* cases occurring in a $5\frac{1}{3}$-year span. If the probability associated with the latter is very small, it could be argued that leukemia did not occur randomly in Niles and that, perhaps, contagion was a factor.

Using the binomial distribution, we can express the probability of eight or more cases as

$$P(8 \text{ or more cases}) = \sum_{k=8}^{7076} \binom{7076}{k} (0.000248)^k (0.999752)^{7076-k} \tag{4.2.1}$$

Much of the computational unpleasantness implicit in Equation 4.2.1 can be avoided by appealing to Theorem 4.2.1. Given that $np = 7076 \times 0.000248 = 1.75$,

$$\begin{aligned} P(X \geq 8) &= 1 - P(X \leq 7) \\ &\doteq 1 - \sum_{k=0}^{7} \frac{e^{-1.75}(1.75)^k}{k!} \\ &= 1 - 0.99951 \\ &= 0.00049 \end{aligned}$$

How close can we expect 0.00049 to be to the "true" binomial sum? Very close. Considering the accuracy of the Poisson limit when n is as small as one hundred (recall Table 4.2.2), we should feel very confident here, where n is 7076.

Interpreting the 0.00049 probability is not nearly as easy as assessing its accuracy. The fact that the probability is so very small tends to denigrate the hypothesis that leukemia in Niles occurred at random. On the other hand, rare events, such as clusters, *do* happen by chance. The basic difficulty in putting the probability associated with a given cluster in any meaningful perspective is not knowing in how many similar communities leukemia did *not* exhibit a tendency to cluster. That there is no obvious way to do this is one reason the leukemia controversy is still with us.

QUESTIONS

4.2.1. If a typist averages one misspelling in every 3250 words, what are the chances that a 6000-word report is free of all such errors? Answer the question two ways—first, by using an exact binomial analysis, and second, by using a Poisson approximation. Does the similarity (or dissimilarity) of the two answers surprise you? Explain.

4.2.2. A medical study recently documented that 905 mistakes were made among the 289,411 prescriptions written during one year at a large metropolitan teaching hospital. Suppose a patient is admitted with a condition serious enough to warrant 10 different prescriptions. Approximate the probability that at least one will contain an error.

4.2.3. Five hundred people are attending the first annual "I was Hit by Lighting" Club. Approximate the probability that at most one of the 500 was born on Poisson's birthday.

4.2.4. A chromosome mutation linked with colorblindness is known to occur, on the average, once in every 10,000 births.

- **(a)** Approximate the probability that exactly 3 of the next 20,000 babies born will have the mutation.
- **(b)** How many babies out of the next 20,000 would have to be born with the mutation to convince you that the "1 in 10,000" estimate is too low? *Hint:* Calculate $P(X \geq k) = 1 - P(X \leq k - 1)$ for various k. (Recall Case Study 4.2.1.)

4.2.5. Suppose that 1% of all items in a supermarket are not priced properly. A customer buys 10 items. What is the probability that she will be delayed by the cashier because one or more of her items requires a price check? Calculate both a binomial answer and a Poisson answer. Is the binomial model "exact" in this case? Explain.

4.2.6. A newly formed life insurance company has underwritten term policies on 120 women between the ages of 40 and 44. Suppose that each woman has a 1/150 probability of dying during the next calendar year, and each death requires the company to pay out $50,000 in benefits. Approximate the probability that the company will have to pay at least $150,000 in benefits next year.

4.2.7. According to an airline industry report (187), roughly 1 piece of luggage out of every 200 that are checked is lost. Suppose that a frequent-flying businesswoman will be checking 120 bags over the course of the next year. Approximate the probability that she will lose 2 of more pieces of luggage.

4.2.8. Electromagnetic fields generated by power transmission lines are suspected by some researchers to be a cause of cancer. Especially at risk would be telephone linemen because of their frequent proximity to high-voltage wires. According to one study, two cases of a rare form of cancer were detected among a group of 9500 linemen (181). In the general population, the incidence of that particular condition is on the order of one in a million. What would you conclude? *Hint:* Recall the approach taken in Case Study 4.2.1.

4.2.9. Astronomers estimate that as many as 100 billion stars in the Milky Way galaxy are encircled by planets. If so, we may have a plethora of cosmic neighbors. Let p denote the probability that any such solar system contains intelligent life. How small can p be and still give a 50-50 chance that we are not alone?

The Poisson Distribution

The real significance of Poisson's limit theorem went unrecognized for more than fifty years. For most of the latter part of the nineteenth century, Theorem 4.2.1 was taken strictly at face value: It provided a convenient approximation for $p_X(k)$ when X is binomial, n is large, and p is small. But then in 1898 a German professor, Ladislaus von Bortkiewicz, published a monograph entitled *Das Gesetz der Kleinen Zahlen (The Law of Small Numbers)* that would quickly transform Poisson's "limit" into Poisson's "distribution."

What is best remembered about Bortkiewicz's monograph is the curious set of data described in Question 4.2.10. The measurements recorded were the numbers of Prussian cavalry soldiers who were kicked to death by their horses. In analyzing those figures, Bortkiewicz was able to show that the function $e^{-\lambda}\lambda^k/k!$ is a useful probability model in its own right, even when (1) no explicit binomial random variable is present and (2) values for n and p are unavailable. Other researchers were quick to follow Bortkiewicz's lead, and a steady stream of Poisson distribution applications began showing up in technical journals. Today the function $p_X(k) = e^{-\lambda}\lambda^k/k!$ is universally recognized as being among the three or four most important data models in all of statistics.

Theorem 4.2.2. *The random variable X is said to have a Poisson distribution if*

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

where λ is a positive constant. Also, for any Poisson random variable, $E(X) = \lambda$ and $\text{Var}(X) = \lambda$.

Proof. To show that $p_X(k)$ qualifies as a probability function, note, first of all, that $p_X(k) \geq 0$ for all nonnegative integers k. Also, $p_X(k)$ sums to one:

$$\sum_{k=0}^{\infty} p_X(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot e^{\lambda} = 1$$

since $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}$ is the Taylor series expansion of e^{λ}. Verifying that $E(X) = \lambda$ and $\text{Var}(X) = \lambda$ has already been done in Example 3.12.9, using moment-generating functions. □

Fitting the Poisson Distribution to Data

Poisson data invariably refer to the numbers of times a certain event occurs during each of a series of "units" (often *time* or *space*). For example, X might be the weekly number of traffic accidents reported at a given intersection. If such records are kept for an entire year, the resulting data would be the sample $k_1, k_2, \ldots, k_{52}$, where each k_i is a nonnegative integer.

Whether or not a set of k_is can be viewed as Poisson data depends on whether the proportions of 0s, 1s, 2s, and so on *in the sample* are numerically similar to the probabilities that $X = 0, 1, 2$, and so on, as predicted by $p_X(k) = e^{-\lambda}\lambda^k/k!$. The next two case studies show data sets where the variability in the observed k_is *is* consistent with the probabilities predicted by the Poisson distribution. Notice in each case that the λ in $p_X(k)$ is replaced by the sample mean of the k_is—that is by $\bar{k} = (1/n)\sum_{c=1}^{n} k_i$. The reason for making that substitution will be taken up in Chapter 5.

CASE STUDY 4.2.2

Among the early research projects investigating the nature of radiation was a 1910 study of α-particle emission by Ernest Rutherford and Hans Geiger (156). For each of 2608 eighth-minute intervals, the two physicists recorded the number of α-particles emitted from a polonium source (as detected by what would eventually be called a Geiger counter). The numbers and proportions of times that k such particles were detected in a given eighth-minute ($k = 0, 1, 2, \dots$) are detailed in the first three columns of Table 4.2.3. Two α particles, for example, were detected in each of 383 eighth-minute intervals, meaning that $X = 2$ was the observation recorded 15% ($= 383/2608 \times 100$) of the time.

To see whether a probability function of the form $p_X(k) = e^{-\lambda}\lambda^k/k!$ can adequately model the observed proportions in the third column, we first need to replace λ with the sample's average value for X. Suppose the six observations comprising the "11+" category are each assigned the value eleven. Then

TABLE 4.2.3

No. Detected, k	Frequency	Proportion	$p_X(k) = e^{-3.87}(3.87)^k/k!$
0	57	0.02	0.02
1	203	0.08	0.08
2	383	0.15	0.16
3	525	0.20	0.20
4	532	0.20	0.20
5	408	0.16	0.15
6	273	0.10	0.10
7	139	0.05	0.05
8	45	0.02	0.03
9	27	0.01	0.01
10	10	0.00	0.00
11+	6	0.00	0.00
	2608	1.0	1.0

$$\bar{k} = \frac{57(0) + 203(1) + 383(2) + \cdots + 6(11)}{2608} = \frac{10{,}092}{2608}$$
$$= 3.87$$

and the presumed model is $p_X(k) = e^{-3.87}(3.87)^k/k!$, $k = 0, 1, 2, \dots$. Notice how closely the entries in the fourth column [i.e., $p_X(0), p_X(1), p_X(2), \dots$] agree with the sample proportions appearing in the third column. The conclusion here is inescapable: The phenomenon of radiation can be modeled very effectively by the Poisson distribution.

CASE STUDY 4.2.3

Table 4.2.4 gives the numbers of fumbles made by 110 Division IA football teams during a recent weekend's slate of fifty-five games (107). Do the data support the contention that the number of fumbles, X, that a team makes during a game is a Poisson random variable?

TABLE 4.2.4

2	1	2	2	3	1	3	4	3	4	5
5	2	1	3	2	5	2	4	1	2	2
1	0	4	2	4	1	2	0	2	0	3
0	1	2	0	1	2	2	3	5	1	3
2	3	4	5	4	3	6	0	3	1	2
1	2	2	1	2	1	3	2	4	2	4
4	2	0	5	4	3	6	5	3	5	1
3	1	1	3	1	4	3	1	5	1	2
1	3	4	4	4	2	7	4	2	5	3
1	3	6	2	1	1	4	1	2	3	0

The first step in summarizing these data is to tally the frequencies and calculate the sample proportions associated with each value of X (see Columns 1–3 of Table 4.2.5). Notice, also, that the average number of fumbles per team is 2.55:

TABLE 4.2.5

No. of Fumbles, k	Frequency	Proportion	$p_X(k) = e^{-2.55}(2.55)^k/k!$
0	8	0.07	0.08
1	24	0.22	0.20
2	27	0.25	0.25
3	20	0.18	0.22
4	17	0.16	0.14
5	10	0.09	0.07
6	3	0.03	0.03
7+	1	0.01	0.01
	110	1.0	1.0

$$\bar{k} = \frac{8(0) + 24(1) + 27(2) + \cdots + 1(7)}{110}$$
$$= 2.55$$

Substituting 2.55 for λ, then, gives $p_X(k) = e^{-2.55}(2.55)^k/k!$ as the particular Poisson model most likely to fit the data.

(Continued on next page)

(Case Study 4.2.3 continued)

The fourth column of Table 4.2.5 shows $p_X(k)$ evaluated for each of the eight values listed for k: $p_X(0) = e^{-2.55}(2.55)^0/0! = 0.08$, and so on. Once again, the row-by-row agreement is quite strong. There appears to be nothing in these data that would refute the presumption that the number of fumbles a team makes is a Poisson random variable.

The Poisson Model: The Law of Small Numbers

Given that the expression $e^{-\lambda}\lambda^k/k!$ models phenomena as diverse as α-radiation and football fumbles raises an obvious question: *Why* is that same $p_X(k)$ describing such different random variables? The answer, of course, is that the underlying physical conditions that produce those two sets of measurements are actually much the same, despite how superficially different the resulting data may seem to be. Both phenomena are examples of a set of mathematical assumptions known as the *Poisson model.* Any measurements that are derived from conditions that mirror those assumptions will necessarily vary in accordance with the Poisson distribution.

Consider, for example, the number of fumbles that a football team makes during the course of a game. Imagine dividing a time interval of length T into n nonoverlapping subintervals, each of length $\frac{T}{n}$, where n is large (see Figure 4.2.1). Suppose that

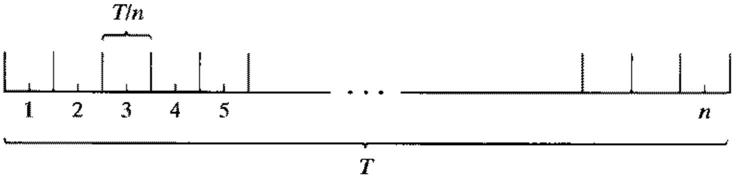


FIGURE 4.2.1

1. The probability that two or more fumbles occur in any given subinterval is essentially 0.
2. Fumbles are independent events.
3. The probability that a fumble occurs during a given subinterval is constant over the entire interval from 0 to T.

The n subintervals, then, are analogous to the n independent trials that form the backdrop for the "binomial model": In each subinterval there will be either zero fumbles or one fumble, where

$$p_n = P(\text{fumble occurs in a given subinterval})$$

remains constant from subinterval to subinterval.

Let the random variable X denote the total number of fumbles a team makes during time T, and let λ denote the *rate* at which a team fumbles (e.g., λ might be expressed as 0.10 fumbles per minute). Then

$$E(X) = \lambda T = np_n \quad \text{(why?)}$$

which implies that $p_n = \frac{\lambda T}{n}$. From Theorem 4.2.1, then,

$$\begin{aligned} p_X(k) = P(X = k) &= \binom{n}{k}\left(\frac{\lambda T}{n}\right)^k\left(1 - \frac{\lambda T}{n}\right)^{n-k} \\ &\doteq \frac{e^{-n(\lambda T/n)}\,[n(\lambda T/n)]^k}{k!} \\ &= \frac{e^{-\lambda T}(\lambda T)^k}{k!} \end{aligned} \tag{4.2.2}$$

So, if a team fumbles at the rate of, say, 0.10 times per minute and they have the ball for 30 minutes during a game, $\lambda T = (0.1)(30) = 3.0$, and the probability that they fumble exactly k times is approximated by the pdf, $p_X(k) = e^{-3.0}(3.0)^k/k!, k = 0, 1, 2 \ldots.$

Now we can see more clearly why Poisson's "limit," as given in Theorem 4.2.1, is so important. The three Poisson model assumptions listed at the top of the page for football fumbles are so unexceptional that they apply to countless real-world phenomena. Each time they do, the pdf $p_X(k) = e^{-\lambda T}(\lambda T)^k/k!$ finds another application.

Calculating Poisson Probabilities

In practice, calculating Poisson probabilities is an exercise in choosing T so that λT represents the expected number of occurrences in whatever "unit" is associated with the random variable X. They look different, but the pdf's $p_X(k) = e^{-\lambda}\lambda^k/k!$ and $p_X(k) = e^{-\lambda T}(\lambda T)^k/k!$ are exactly the same and will give identical values for $P(X = k)$ once λ and T are properly defined.

EXAMPLE 4.2.3

Suppose that typographical errors are made at the rate of 0.4 per page in State Tech's campus newspaper. If next Tuesday's edition is sixteen pages long, what is the probability that fewer than three typos will appear?

We start by defining X to be the number of errors that will appear *in sixteen pages*. The assumptions of independence and constant probability are not unreasonable in this setting, so X is likely to be a Poisson random variable. To answer the question using the formula in Theorem 4.2.2, we need to set λ equal to $E(X)$. But if the error rate is 0.4 errors/page, the expected number of typos in sixteen pages will be *6.4:*

$$0.4\frac{\text{errors}}{\text{page}} \times 16 \text{ pages} = 6.4 \text{ errors}$$

It follows, then, that

$$\begin{aligned} P(X < 3) = P(X \leq 2) &= \sum_{k=0}^{2} \frac{e^{-6.4}(6.4)^k}{k!} \\ &= \frac{e^{-6.4}(6.4)^0}{0!} + \frac{e^{-6.4}(6.4)^1}{1!} + \frac{e^{-6.4}(6.4)^2}{2!} \\ &= 0.046 \end{aligned}$$

If Equation 4.2.2 is used, we would define

$$\lambda = 0.4 \text{ errors/page}$$

and

$$T = 16 \text{ pages}$$

Then $\lambda T = E(X) = 6.4$ and $P(X < 3)$ would be $\sum_{k=0}^{2} e^{-6.4}(6.4)^k/k!$, the same numerical value found from Theorem 4.2.2.

EXAMPLE 4.2.4

Entomologists estimate that an average person consumes almost a pound of bug parts each year (180). There are that many insect eggs, larvae, and miscellaneous body pieces in the foods we eat and the liquids we drink. The Food and Drug Administration (FDA) sets a Food Defect Action Level (FDAL) for each product: Bug-part concentrations below the FDAL are considered acceptable. The legal limit for peanut butter, for example, is thirty insect fragments per hundred grams. Suppose the crackers you just bought from a vending machine are spread with twenty grams of peanut butter. What are the chances that snack will include at least five crunchy critters?

Let X denote the number of bug parts in twenty grams of peanut butter. Assuming the worst, we will set the contamination level equal to the FDA limit—that is, thirty fragments per hundred grams (or 0.30 fragments/g). Notice that $E(X) = 6.0$:

$$\frac{0.30 \text{ fragments}}{\text{g}} \times 20 \text{ g} = 6.0 \text{ fragments}$$

It follows, then, that the probability that your snack contains five or more bug parts is a disgusting *0.71*:

$$\begin{aligned} P(X \geq 5) = 1 - P(X \leq 4) = 1 &- \sum_{k=0}^{4} \frac{e^{-6.0}(6.0)^k}{k!} \\ &= 1 - 0.2851 \\ &= 0.71 \end{aligned}$$

Bon appetit!

QUESTIONS

4.2.10. During the latter part of the nineteenth century, Prussian officials gathered information relating to the hazards that horses posed to cavalry soldiers. A total of 10 cavalry corps were monitored over a period of 20 years. Recorded for each year and each corps was X, the annual number of fatalities due to kicks. Summarized in the following table are the 200 values recorded for X (14). Show that these data can be modeled by a Poisson pdf. Follow the procedure illustrated in Case Studies 4.2.2 and 4.2.3.

No. of Deaths, k	Observed Number of Corps-Years in Which k Fatalities Occurred
0	109
1	65
2	22
3	3
4	1
	200

4.2.11. A random sample of 356 seniors enrolled at the University of West Florida was categorized according to X, the number of times they had changed majors (114). Based on the summary of that information shown in the following table, would you conclude that X can be treated as a Poisson random variable?

Number of Major Changes	Frequency
0	237
1	90
2	22
3	7

4.2.12. Midwestern Skies books 10 commuter flights each week. Passenger totals are much the same from week to week, as are the numbers of pieces of luggage that are checked. Listed in the following table are the numbers of bags that were lost during each of the first 40 weeks in 2004. Do these figures support the presumption that the number of bags lost by Midwestern during a typical week is a Poisson random variable?

Week	Bags Lost	Week	Bags Lost	Week	Bags Lost
1	1	14	2	27	1
2	0	15	1	28	2
3	0	16	3	29	0
4	3	17	0	30	0
5	4	18	2	31	1
6	1	19	5	32	3
7	0	20	2	33	1
8	2	21	1	34	2
9	0	22	1	35	0
10	2	23	1	36	1
11	3	24	2	37	4
12	1	25	1	38	2
13	2	26	3	39	1
				40	0

4.2.13. In 1893, New Zealand became the first country to permit women to vote. Scattered over the ensuing 113 years, various countries joined this movement to grant this right to women. The table below (127) shows how many countries took this step in a given year. Do these data seem to follow a Poisson distribution?

Yearly Number of Countries Granting Women the Vote	Frequency
0	82
1	25
2	4
3	0
4	2

4.2.14. The following are the daily numbers of death notices for women over the age of 80 that appeared in the *London Times* over a three-year period (73).

Number of Deaths	Observed Frequency
0	162
1	267
2	271
3	185
4	111
5	61
6	27
7	8
8	3
9	1
	1096

(a) Does the Poisson pdf provide a good description of the variability pattern evident in these data?

(b) If your answer to Part (a) is "no," which of the Poisson model assumptions do you think might not be holding?

4.2.15. A certain species of European mite is capable of damaging the bark on orange trees. The following are the results of inspections done on 100 saplings chosen at random from a large orchard. The measurement recorded, X, is the number of mite infestations found on the trunk of each tree. Is it reasonable to assume that X is a Poisson random variable? If not, which of the Poisson model assumptions is likely not to be true?

No. of Infestations, k	No. of Trees
0	55
1	20
2	21
3	1
4	1
5	1
6	0
7	1

4.2.16. A tool and die press that stamps out cams used in small gasoline engines tends to break down once every five hours. The machine can be repaired and put back on line quickly, but each such incident costs \$50. What is the probability that maintenance expenses for the press will be no more than \$100 on a typical eight-hour workday?

4.2.17. In a new fiber optic communication system, transmission errors occur at the rate of 1.5 per 10 seconds. What is the probability that more than two errors will occur during the next half-minute?

4.2.18. Assume that the number of hits, X, that a baseball team makes in a nine-inning game has a Poisson distribution. If the probability that a team makes zero hits is $\frac{1}{3}$, what are their chances of getting two or more hits?

4.2.19. Flaws in metal sheeting produced by a high-temperature roller occur at the rate of one per 10 square feet. What is the probability that three or more flaws will appear in a 5-by-8-foot panel?

4.2.20. Suppose a radioactive source is metered for two hours, during which time the total number of alpha particles counted is 482. What is the probability that exactly three particles will be counted in the next two minutes? Answer the question two ways—first, by defining X to be the number of particles counted in two minutes, and second, by defining X to be the number of particles counted in one minute.

4.2.21. Suppose that on-the-job injuries in a textile mill occur at the rate of 0.1 per day.

(a) What is the probability that two accidents will occur during the next (five-day) work week?

(b) Is the probability that four accidents will occur over the next two work weeks the square of your answer to Part (a)? Explain.

4.2.22. Find $P(X = 4)$ if the random variable X has a Poisson distribution such that $P(X = 1) = P(X = 2)$.

4.2.23. Let X be a Poisson random variable with parameter λ. Show that the probability that X is even is $\frac{1}{2}(1 + e^{-2\lambda})$.

4.2.24. Let X and Y be independent Poisson random variables with parameters λ and μ, respectively. Example 3.12.10 established that $X + Y$ is also Poisson with parameter $\lambda + \mu$. Prove that same result using Theorem 3.8.1.

4.2.25. If X_1 is a Poisson random variable for which $E(X_1) = \lambda$ and if the conditional pdf of X_2 given that $X_1 = x_1$ is binomial with parameters x_1 and p, show that the marginal pdf of X_2 is Poisson with $E(X_2) = \lambda p$.

Intervals Between Events: The Poisson/Exponential Relationship

Situations sometimes arise where the time interval between consecutively occurring events is an important random variable. Imagine being responsible for the maintenance on a network of computers. Clearly, the number of technicians you would need to employ in order to be capable of responding to service calls in a timely fashion would be a function of the "waiting time" from one breakdown to another.

Figure 4.2.2 shows the relationship between the random variables X and Y, where X denotes the number of occurrences in a unit of time and Y denotes the interval between consecutive occurrences. Pictured are six intervals: $X = 0$ on one occasion, $X = 1$ on three occasions, $X = 2$ once, and $X = 3$ once. Resulting from those eight occurrences are seven measurements on the random variable Y. Obviously, the pdf for Y will depend on the pdf for X. One particularly important special case of that dependence is the Poisson/exponential relationship outlined in Theorem 4.2.3.

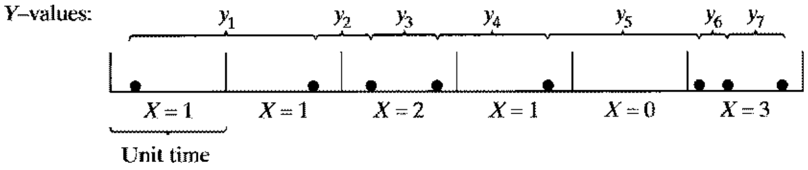


FIGURE 4.2.2

Theorem 4.2.3. *Suppose a series of events satisfying the Poisson model are occurring at the rate of λ per unit time. Let the random variable Y denote the interval between consecutive events. Then Y has the exponential distribution*

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0$$

Proof. Suppose an event has occurred at time a. Consider the interval that extends from a to $a + y$. Since the (Poisson) events are occurring at the rate of λ per unit time, the probability that no outcomes will occur in the interval $(a, a + y)$ is $\dfrac{e^{-\lambda y}(\lambda y)^0}{0!} = e^{-\lambda y}$. Define the random variable Y to denote the interval between consecutive occurrences. Notice that there will be no occurrences in the interval $(a, a + y)$ only if $Y > y$. Therefore,

$$P(Y > y) = e^{-\lambda y}$$

or, equivalently,

$$P(Y \leq y) = 1 - P(Y > y) = 1 - e^{-\lambda y}$$

Let $f_Y(y)$ be the (unknown) pdf for Y. It must be true that

$$P(Y \leq y) = \int_0^y f_Y(t)\,dt$$

Taking derivatives of the two expressions for $P(Y \leq y)$, we can write

$$\frac{d}{dy}\int_0^y f_Y(t)\,dt = \frac{d}{dy}(1 - e^{-\lambda y})$$

which implies that

$$f_Y(y) = \lambda e^{-\lambda y}, \quad y > 0$$

□

CASE STUDY 4.2.4

Over "short" geological periods, a volcano's eruptions are believed to be Poisson events—that is, they are thought to occur independently and at a constant rate. If so, the pdf describing the intervals between eruptions should have the form $f_Y(y) = \lambda e^{-\lambda y}$. Collected for the purpose of testing that presumption are the data in Table 4.2.6, showing the intervals (in months) that elapsed between thirty-seven consecutive

(Continued on next page)

eruptions of Mauna Loa, a fourteen thousand-foot volcano in Hawaii (110). During the period covered—1832 to 1950—eruptions were occurring at the rate of $\lambda = 0.027$ per month (or once every 3.1 years). Is the variability in these thirty-six y_is consistent with the statement of Theorem 4.2.3?

TABLE 4.2.6

126	73	3	6	37	23
73	23	2	65	94	51
26	21	6	68	16	20
6	18	6	41	40	18
41	11	12	38	77	61
26	3	38	50	91	12

To answer that question requires that the data be reduced to a density-scaled histogram and superimposed on a graph of the predicted exponential pdf (recall Case Study 3.4.1). Table 4.2.7 details the construction of the histogram. Notice in Figure 4.2.3 that the shape of that histogram *is* entirely consistent with the theoretical model—$f_Y(y) = 0.027e^{-0.027y}$—stated in Theorem 4.2.3.

TABLE 4.2.7

Interval (mos), y	Frequency	Density
$0 \le y < 20$	13	0.0181
$20 \le y < 40$	9	0.0125
$40 \le y < 60$	5	0.0069
$60 \le y < 80$	6	0.0083
$80 \le y < 100$	2	0.0028
$100 \le y < 120$	0	0.0000
$120 \le y < 140$	1	0.0014
	36	

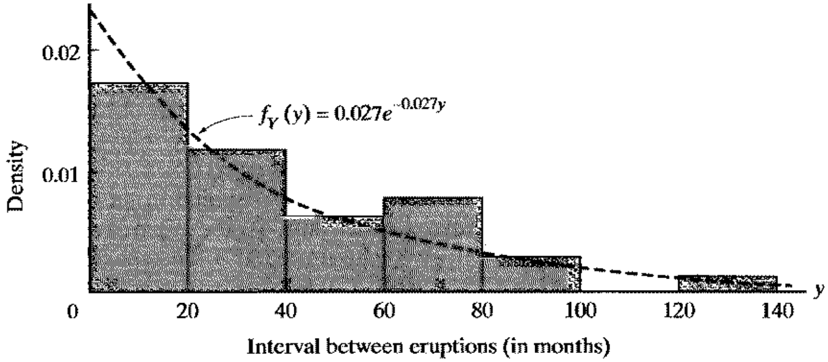


FIGURE 4.2.3

EXAMPLE 4.2.5

Among the most famous of all meteor showers are the Perseids, which occur each year in early August. In some areas the frequency of visible Perseids can be as high as forty per hour. Given that such sightings are Poisson events, calculate the probability that an observer who has just seen a meteor will have to wait at least five minutes before seeing another.

Let the random variable Y denote the interval (in minutes) between consecutive sightings. Expressed in the units of Y, the *forty per hour* rate of visible Perseids becomes *0.67 per minute.* A straightforward integration, then, shows that the probability is *0.036* that an observer will have to wait five minutes or more to see another meteor:

$$\begin{aligned}
P(Y > 5) &= \int_5^\infty 0.67e^{-0.67y}\,dy \\
&= \int_{3.33}^\infty e^{-u}\,du \quad (\text{where } u = 0.67y) \\
&= -e^{-u}\Big|_{3.33}^{\infty} = e^{-3.33} \\
&= 0.036
\end{aligned}$$

QUESTIONS

4.2.26. Suppose that commercial airplane crashes in a certain country occur at the rate of 2.5 per year.

- **(a)** Is it reasonable to assume that such crashes are Poisson events? Explain.
- **(b)** What is the probability that four or more crashes will occur next year?
- **(c)** What is the probability that the next two crashes will occur within three months of one another?

4.2.27. Records show that deaths occur at the rate of 0.1 per day among patients residing in a large nursing home. If someone dies today, what are the chances that a week or more will elapse before another death occurs?

4.2.28. Suppose that Y_1 and Y_2 are independent exponential random variables, each having pdf $f_Y(y) = \lambda e^{-\lambda y}$, $y > 0$. If $Y = Y_1 + Y_2$, it can be shown that

$$f_{Y_1+Y_2}(y) = \lambda^2 y e^{-\lambda y}, \quad y > 0$$

Recall Case Study 4.2.4. What is the probability that the next three eruptions of Mauna Loa will be less than 40 months apart?

4.2.29. Fifty spotlights have just been installed in an outdoor security system. According to the manufacturer's specifications, these particular lights are expected to burn out at the rate of 1.1 per 100 hours. What is the expected number of bulbs that will fail to last for at least 75 hours?

4.3 THE NORMAL DISTRIBUTION

The Poisson limit described in Section 4.2 was not the only, or the first, approximation developed for the purpose of facilitating the calculation of binomial probabilities. Early in the eighteenth century, Abraham DeMoivre proved that areas under the curve

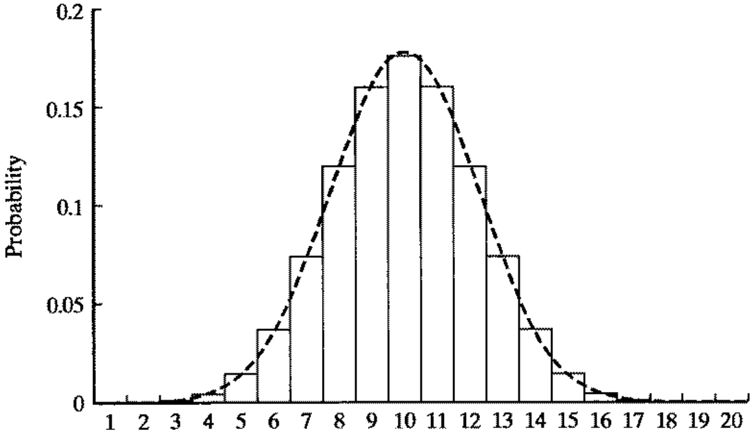


FIGURE 4.3.1

$f_z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$, $-\infty < z < \infty$ can be used to estimate $P\left(a \le \frac{X - n(\frac{1}{2})}{\sqrt{n(\frac{1}{2})(\frac{1}{2})}} \le b\right)$, where X is a binomial random variable with a large n and $p = \frac{1}{2}$.

Figure 4.3.1 illustrates the central idea in DeMoivre's discovery. Pictured is a probability histogram of the binomial distribution with $n = 20$ and $p = \frac{1}{2}$. Superimposed over the histogram is the function $f_z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$. Notice how closely the area under the curve approximates the area of the bar, even for this relatively small value of n. The French mathematician Pierre-Simon Laplace generalized DeMoivre's original idea to binomial approximations for arbitrary p and brought this theorem to the full attention of the mathematical community by including it in his influential 1812 book, *Theorie Analytique des Probabilities.*

Theorem 4.3.1. *Let X be a binomial random variable defined on n independent trials for which $p = P(\text{success})$. For any numbers a and b,*

$$\lim_{n\to\infty} P\left(a \le \frac{X - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}\,dz$$

Proof. One of the ways to verify Theorem 4.3.1 is to show that the limit of the moment-generating function for $\frac{X - np}{\sqrt{np(1-p)}}$ as $n \to \infty$ is $e^{t^2/2}$ and that $e^{t^2/2}$ is also the value of $\int_{-\infty}^{\infty} e^{tz} \cdot \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz$. By Theorem 3.12.2, then, the limiting pdf of $Z = \frac{X - np}{\sqrt{np(1-p)}}$ is the function $f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$, $-\infty < z < \infty$. See Appendix 4.A.2 for the proof of a more general result. □

Comment. We saw in Section 4.2 that Poisson's *limit* is actually a special case of Poisson's *distribution*, $p_X(k) = \frac{e^{-\lambda}\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$. Similarly, the DeMoivre-Laplace limit is a pdf in its own right. Justifying that assertion, of course, requires proving that $f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ integrates to 1 for $-\infty < z < \infty$.

Curiously, there is no algebraic or trigonometric substitution that can be used to demonstrate that the area under $f_Z(z)$ is 1. However, by using polar coordinates, we can verify a necessary and sufficient alternative—namely, that the *square* of $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz$ equals one.

To begin, note that

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\,dx \cdot \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-y^2/2}\,dy = \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy$$

Let $x = r\cos\theta$ and $y = r\sin\theta$, so $dx\,dy = r\,dr\,d\theta$. Then

$$\begin{aligned}\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2+y^2)}\,dx\,dy &= \frac{1}{2\pi}\int_{0}^{2\pi}\int_{0}^{\infty} e^{-r^2/2}\,r\,dr\,d\theta\\ &= \frac{1}{2\pi}\int_{0}^{\infty} re^{-r^2/2}\,dr \cdot \int_{0}^{2\pi} d\theta\\ &= 1\end{aligned}$$

Comment. The function $f_Z(z) = \frac{1}{\sqrt{2\pi}}e^{-z^2/2}$ is referred to as the *standard normal* (or *Gaussian*) *curve*. By convention, any random variable whose probabilistic behavior is described by a standard normal curve is denoted by Z (rather than X, Y, or W). Since $M_Z(t) = e^{t^2/2}$, it follows readily that $E(Z) = 0$ and $\text{Var}(Z) = 1$.

Finding Areas Under the Standard Normal Curve

In order to use Theorem 4.3.1, we need to be able to find the area under the graph of $f_Z(z)$ above an arbitrary interval $[a, b]$. In practice, such values are obtained in one of two ways—either by using a *normal table*, a copy of which appears at the back of every statistics book, or by running a computer software package. Typically, both approaches give the *cdf*, $F_Z(z) = P(Z \le z)$, associated with Z (and from the cdf we can deduce the desired area).

Table 4.3.1 shows a portion of the normal table that appears in Appendix A.1. Each row under the Z heading represents a number along the horizontal axis of $f_Z(z)$ rounded off to the nearest tenth; columns 0 through 9 allow that number to be written to the hundredths place. Entries in the body of the table are areas under the graph of $f_Z(z)$ *to the left* of the number indicated by the entry's row and column. For example, the number listed at the intersection of the "1.1" row and the "4" column is 0.8729, which means that the area under $f_Z(z)$ from $-\infty$ to 1.14 is 0.8729. That is,

$$\int_{-\infty}^{1.14} \frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,dz = 0.8729 = P(-\infty < Z \le 1.14) = F_Z(1.14)$$

(see Figure 4.3.2).

statement for Theorem 4.3.1: If X is a binomial random variable with parameters n and p,

$$P(a \leq X \leq b) \doteq F_Z\left(\frac{b + 0.5 - np}{\sqrt{np(1 - p)}}\right) - F_Z\left(\frac{a - 0.5 - np}{\sqrt{np(1 - p)}}\right)$$

Comment. Even with the continuity correction refinement, normal curve approximations can be inadequate if n is too small, especially when p is close to 0 or to 1. As a rule of thumb, the DeMoivre-Laplace limit should be used only if the magnitudes of n and p are such that $n > 9\frac{p}{1 - p}$ and $n > 9\frac{1 - p}{p}$.

EXAMPLE 4.3.1

Boeing 757s flying certain routes are configured to have 168 economy class seats. Experience has shown that only 90% of all ticket-holders on those flights will actually show up in time to board the plane. Knowing that, suppose an airline sells 178 tickets for the 168 seats. What is the probability that not everyone who arrives at the gate on time can be accommodated?

Let the random variable X denote the number of would-be passengers who show up for a flight. Since travelers are sometimes with their families, not every ticket-holder constitutes an independent event. Still, we can get a useful approximation to the probability that the flight is overbooked by assuming that X is binomial with $n = 178$ and $p = 0.9$. What we are looking for is $P(169 \leq X \leq 178)$, the probability that more ticket-holders show up than there are seats on the plane. According to Theorem 4.3.2 (and using the continuity correction),

$$\begin{aligned}
P(\text{flight is overbooked}) &= P(169 \leq X \leq 178) \\
&= P\left(\frac{169 - 0.5 - np}{\sqrt{np(1 - p)}} \leq \frac{X - np}{\sqrt{np(1 - p)}} \leq \frac{178 + 0.5 - np}{\sqrt{np(1 - p)}}\right) \\
&= P\left(\frac{168.5 - 178(0.9)}{\sqrt{178(0.9)(0.1)}} \leq \frac{X - 178(0.9)}{\sqrt{178(0.9)(0.1)}} \leq \frac{178.5 - 178(0.9)}{\sqrt{178(0.9)(0.1)}}\right) \\
&\doteq P(2.07 \leq Z \leq 4.57) = F_z(4.57) - F_z(2.07)
\end{aligned}$$

From Appendix A.1, $F_Z(4.57) = P(Z \leq 4.57)$ is equal to one, for all practical purposes, and the area under $f_Z(z)$ to the left of 2.07 is 0.9808. Therefore,

$$\begin{aligned}
P(\text{flight is overbooked}) &= 1.0000 - 0.9808 \\
&= 0.0192
\end{aligned}$$

implying that the chances are about one in fifty that not every ticket-holder will have a seat.

CASE STUDY 4.3.1

Research in extrasensory perception has ranged from the slightly unconventional to the downright bizarre. Toward the latter part of the nineteenth century and even well into the twentieth century, much of what was done involved spiritualists and mediums. But beginning around 1910, experimenters moved out of the seance parlors and into the laboratory, where they began setting up controlled studies that could be analyzed statistically. In 1938, Pratt and Woodruff, working out of Duke University, did an experiment that became a prototype for an entire generation of ESP research (70).

The investigator and a subject sat at opposite ends of a table. Between them was a screen with a large gap at the bottom. Five blank cards, visible to both participants, were placed side by side on the table beneath the screen. On the subject's side of the screen one of the standard ESP symbols (see Figure 4.3.4) was hung over each of the blank cards.

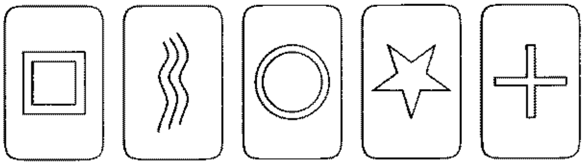


FIGURE 4.3.4

The experimenter shuffled a deck of ESP cards, picked up the top one, and concentrated on it. The subject tried to guess its identity: If he thought it was a *circle*, he would point to the blank card on the table that was beneath the circle card hanging on his side of the screen. The procedure was then repeated. Altogether, a total of thirty-two subjects, all students, took part in the experiment. They made a total of sixty thousand guesses—and were correct 12,489 times.

With five denominations involved, the probability of a subject's making a correct identification just by chance was $\frac{1}{5}$. Assuming a binomial model, the expected number of correct guesses would be $60{,}000 \times \frac{1}{5}$, or 12,000. The question is, how "near" to 12,000 is 12,489? Should we write off the observed excess of 489 as nothing more than luck, or can we conclude that ESP has been demonstrated?

To effect a resolution between the conflicting "luck" and "ESP" hypotheses, we need to compute the probability of the subjects' getting 12,489 or more correct answers *under the presumption that* $p = \frac{1}{5}$. Only if that probability is very small can 12,489 be construed as evidence in support of ESP.

Let the random variable X denote the number of correct responses in sixty thousand tries. Then

$$P(X \geq 12{,}489) = \sum_{k=12{,}489}^{60{,}000} \binom{60{,}000}{k} \left(\frac{1}{5}\right)^k \left(\frac{4}{5}\right)^{60{,}000-k} \tag{4.3.1}$$

(Continued on next page)

At this point the DeMoivre-Laplace limit theorem becomes a welcome alternative to computing the 47,512 binomial probabilities implicit in Equation 4.3.1. First we apply the continuity correction and rewrite $P(X \geq 12{,}489)$ as $P(X \geq 12{,}488.5)$. Then

$$\begin{aligned} P(X \geq 12{,}489) &= P\left(\frac{X - np}{\sqrt{np(1-p)}} \geq \frac{12{,}488.5 - 60{,}000(1/5)}{\sqrt{60{,}000(1/5)(4/5)}}\right) \\ &= P\left(\frac{X - np}{\sqrt{np(1-p)}} \geq 4.99\right) \\ &\doteq \frac{1}{\sqrt{2\pi}} \int_{4.99}^{\infty} e^{-z^2/2}\, dz \\ &= 0.0000003 \end{aligned}$$

this last value being obtained from a more extensive version of Table A.1 in the Appendix.

Here, the fact that $P(X \geq 12{,}489)$ is so extremely small makes the "luck" hypothesis $\left(p = \frac{1}{5}\right)$ untenable. It would appear that something other than chance had to be responsible for the occurrence of so many correct guesses. Still, it does not follow that ESP has necessarily been demonstrated. Flaws in the experimental setup as well as errors in reporting the scores could have inadvertently produced what appears to be a statistically significant result. Suffice it to say that a great many scientists remain highly skeptical of ESP research in general and of the Pratt-Woodruff experiment in particular. [For a more thorough critique of the data we have just described, see (45).]

Comment. This is a good set of data for illustrating why we need formal mathematical methods for interpreting data. The fact is, our intuitions, when left unsupported by probability calculations, can often be deceived. A typical first reaction to the Pratt–Woodruff results is to dismiss as inconsequential the 489 additional correct answers. To many, it seems entirely believable that 60,000 guesses could produce, by chance, an extra 489 correct responses. Only after making the $P(X \geq 12{,}489)$ computation do we see the utter implausibility of that conclusion. What statistics is doing here is what we would like it to do in general—rule out hypotheses that are not supported by the data and point us in the direction of inferences that are more likely to be true.

QUESTIONS

4.3.1. Use Appendix Table A.1 to evaluate the following integrals. In each case, draw a diagram of $f_Z(z)$ and shade the area that corresponds to the integral.

(a) $\displaystyle\int_{-0.44}^{1.33} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

(b) $\displaystyle\int_{-\infty}^{0.94} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

(c) $\int_{-1.48}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

(d) $\int_{-\infty}^{-4.32} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$

4.3.2. Let Z be a standard normal random variable. Use Appendix Table A.1 to find the numerical value for each of the following probabilities. Show each of your answers as an area under $f_Z(z)$.

(a) $P(0 \le Z \le 2.07)$
(b) $P(-0.64 \le Z < -0.11)$
(c) $P(Z > -1.06)$
(d) $P(Z < -2.33)$
(e) $P(Z \ge 4.61)$

4.3.3. (a) Let $0 < a < b$. Which number is larger?

$$\int_a^b \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad \text{or} \quad \int_{-b}^{-a} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

(b) Let $a > 0$. Which number is larger?

$$\int_a^{a+1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz \quad \text{or} \quad \int_{a-1/2}^{a+1/2} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

4.3.4. (a) Evaluate $\int_0^{1.24} e^{-z^2/2}\, dz$.

(b) Evaluate $\int_{-\infty}^{\infty} 6e^{-z^2/2}\, dz$.

4.3.5. Assume that the random variable Z is described by a standard normal curve $f_Z(z)$. For what values of z are the following statements true?

(a) $P(Z \le z) = 0.33$
(b) $P(Z \ge z) = 0.2236$
(c) $P(-1.00 \le Z \le z) = 0.5004$
(d) $P(-z < Z < z) = 0.80$
(e) $P(z \le Z \le 2.03) = 0.15$

4.3.6. Let z_α denote the value of Z for which $P(Z \ge z_\alpha) = \alpha$. By definition, the *interquartile range, Q*, for the standard normal curve is the difference

$$Q = z_{.25} - z_{.75}$$

Find Q.

4.3.7. Oak Hill has 74,806 registered automobiles. A city ordinance requires each to display a bumper decal showing that the owner paid an annual wheel tax of \$50. By law, new decals need to be purchased during the month of the owner's birthday. This year's budget assumes that at least \$306,000 in decal revenue will be collected in November. What is the probability that taxes reported in that month will be less than anticipated and produce a budget shortfall?

4.3.8. Hertz Brothers, a small, family-owned radio manufacturer, produces electronic components domestically but subcontracts the cabinets to a foreign supplier. Although inexpensive, the foreign supplier has a quality control program that leaves much to be desired. On the average, only 80% of the standard 1600-unit shipment that Hertz receives is usable. Currently, Hertz has back orders for 1260 radios but storage space for no more than 1310 cabinets. What are the chances that the number of usable units in Hertz's latest shipment will be large enough to allow Hertz to fill all the orders already on hand, yet small enough to avoid causing any inventory problems?

4.3.9. Fifty-five percent of the registered voters in Sheridanville favor their incumbent mayor in her bid for reelection. If 400 voters go to the polls, approximate the probability that

(a) the race ends in a tie

(b) the challenger scores an upset victory

4.3.10. State Tech's basketball team, the Fighting Logarithms, have a 70% foul-shooting percentage.

(a) Write a formula for the exact probability that out of their next 100 free throws they will make between 75 and 80, inclusive.

(b) Approximate the probability asked for in Part (a).

4.3.11. A random sample of 747 obituaries published recently in Salt Lake City newspapers revealed that 344 (or 46%) of the decedents died in the three-month period following their birthdays (129). Assess the statistical significance of that finding by approximating the probability that 46% or more would die in that particular interval if deaths occurred randomly throughout the year. What would you conclude on the basis of your answer?

4.3.12. There is a theory embraced by certain parapsychologists that hypnosis can enhance a person's ESP ability. To test that hypothesis, an experiment was set up with 15 hypnotized subjects (22). Each was asked to make 100 guesses using the same sort of ESP cards and protocol that were described in Case Study 4.3.1. A total of 326 correct identifications were made. Can it be argued on the basis of those results that hypnosis does have an effect on a person's ESP ability? Explain.

4.3.13. If $p_X(k) = \binom{10}{k}(0.7)^k(0.3)^{10-k}$, $k = 0, 1, \ldots, 10$, is it appropriate to approximate $P(4 \leq X \leq 8)$ by computing

$$P\left(\frac{3.5 - 10(0.7)}{\sqrt{10(0.7)(0.3)}} \leq Z \leq \frac{8.5 - 10(0.7)}{\sqrt{10(0.7)(0.3)}}\right)$$

Explain.

4.3.14. A sell-out crowd of 42,200 is expected at Cleveland's Jacobs Field for next Tuesday's game with the Baltimore Orioles, the last before a long road trip. The ballpark's concession manager is trying to decide how much food to have on hand. Looking at records from games played earlier in the season, she knows that, on the average, 38% of all those in attendance will buy a hot dog. How large an order should she place if she wants to have no more that a 20% chance of demand exceeding supply?

Central Limit Theorem

It was pointed out in Example 3.9.3 that every binomial random variable X can be written as the sum of n independent Bernoulli random variables $X_1, X_2, \ldots, X_n$, where

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p \end{cases}$$

But if $X = X_1 + X_2 + \cdots + X_n$, Theorem 4.3.1 can be reexpressed as

$$\lim_{n\to\infty} P\left(a \le \frac{X_1 + X_2 + \cdots + X_n - np}{\sqrt{np(1-p)}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}\,dz \tag{4.3.2}$$

Implicit in Equation 4.3.2 is an obvious question: Does the DeMoivre-Laplace limit apply to sums of other types of random variables as well? Remarkably, the answer is "yes." Efforts to extend Equation 4.3.2 have continued for more than one hundred and fifty years. Russian probabilists—A. M. Lyapunov, in particular—made many of the key advances. In 1920, George Polya gave these new generalizations a name that has been associated with the result ever since: He called it the *central limit theorem* (141).

Theorem 4.3.2 (Central Limit Theorem). *Let $W_1, W_2, \ldots$ be an infinite sequence of independent random variables, each with the same distribution. Suppose that the mean μ and the variance σ^2 of $f_W(w)$ are both finite. For any numbers a and b,*

$$\lim_{n\to\infty} P\left(a \le \frac{W_1 + \cdots + W_n - n\mu}{\sqrt{n}\sigma} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}\,dz$$

Proof. See Appendix 4.A.2. □

Comment. The central limit theorem is often stated in terms of the *average of* W_1, $W_2, \ldots$, and W_n, rather than their sum. Since

$$E\left[\frac{1}{n}(W_1 + \cdots + W_n)\right] = E(\overline{W}) = \mu \quad \text{and} \quad \text{Var}\left[\frac{1}{n}(W_1 + \cdots + W_n)\right] = \sigma^2/n,$$

Theorem 4.3.2 can be stated in the equivalent form

$$\lim_{n\to\infty} P\left(a \le \frac{\overline{W} - \mu}{\sigma/\sqrt{n}} \le b\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-z^2/2}\,dz$$

We will use both formulations, the choice depending on which is more convenient for the problem at hand.

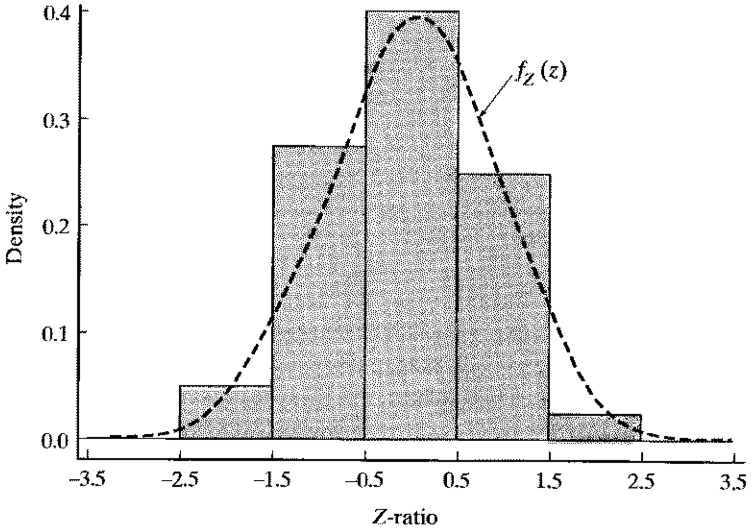
EXAMPLE 4.3.2

The top of Table 4.3.2 shows a MINITAB simulation where forty random samples of size five were drawn from a uniform pdf defined over the interval [0, 1]. Each row corresponds to a different sample. The sum of the five numbers appearing in a given sample is denoted "y" and is listed in column C6. For this particular uniform pdf, $\mu = \frac{1}{2}$ and $\sigma^2 = \frac{1}{12}$ (recall Question 3.6.4), so

$$\frac{W_1 + \cdots + W_n - n\mu}{\sqrt{n}\sigma} = \frac{Y - \frac{5}{2}}{\sqrt{\frac{5}{12}}}$$

TABLE 4.3.2

	C1 y1	C2 y2	C3 y3	C4 y4	C5 y5	C6 y	C7 Z ratio
1	0.556099	0.646873	0.354373	0.673821	0.233126	2.46429	−0.05532
2	0.497846	0.588979	0.272095	0.956614	0.819901	3.13544	0.98441
3	0.284027	0.209458	0.414743	0.614309	0.439456	1.96199	−0.83348
4	0.599286	0.667891	0.194460	0.839481	0.694474	2.99559	0.76777
5	0.280689	0.692159	0.036593	0.728826	0.314434	2.05270	−0.69295
6	0.462741	0.349264	0.471254	0.613070	0.489125	2.38545	−0.17745
7	0.556940	0.246789	0.719907	0.711414	0.918221	3.15327	1.01204
8	0.102855	0.679119	0.559210	0.014393	0.518450	1.87403	−0.96975
9	0.642859	0.004636	0.728131	0.299165	0.801093	2.47588	−0.03736
10	0.017770	0.568188	0.416351	0.908079	0.075108	1.98550	−0.79707
11	0.331291	0.410705	0.118571	0.979254	0.242582	2.08240	−0.64694
12	0.355047	0.961126	0.920597	0.575467	0.585492	3.39773	1.39076
13	0.626197	0.304754	0.530345	0.933018	0.675899	3.07021	0.88337
14	0.211714	0.404505	0.045544	0.213012	0.520614	1.39539	−1.71125
15	0.535199	0.130715	0.603642	0.333023	0.405782	2.00836	−0.76164
16	0.810374	0.153955	0.082226	0.827269	0.897901	2.77172	0.42095
17	0.687550	0.185393	0.620878	0.013395	0.819712	2.32693	−0.26812
18	0.424193	0.529199	0.201554	0.157073	0.090455	1.40248	−1.70028
19	0.397373	0.143507	0.973991	0.234845	0.681147	2.43086	−0.10711
20	0.413788	0.653468	0.017335	0.556255	0.900568	2.54141	0.06416
21	0.602607	0.094162	0.247676	0.638875	0.653910	2.23723	−0.40708
22	0.963678	0.375850	0.909377	0.307358	0.828882	3.38515	1.37126
23	0.967499	0.868809	0.940770	0.405564	0.814348	3.99699	2.31913
24	0.439913	0.446679	0.075227	0.983295	0.554581	2.49970	−0.00047
25	0.215774	0.407494	0.002307	0.971140	0.437144	2.03386	−0.72214
26	0.108881	0.271860	0.972351	0.604762	0.210347	2.16820	−0.51402
27	0.337798	0.173911	0.309916	0.300208	0.666831	1.78866	−1.10200
28	0.635017	0.187311	0.365419	0.831417	0.463567	2.48273	−0.02675
29	0.563097	0.065293	0.841320	0.518055	0.685137	2.67290	0.26786
30	0.687242	0.544286	0.980337	0.649507	0.077364	2.93874	0.67969
31	0.784501	0.745614	0.459559	0.565875	0.529171	3.08472	0.90584
32	0.505460	0.355340	0.163285	0.352540	0.896521	2.27315	−0.35144
33	0.336992	0.734869	0.824409	0.321047	0.682283	2.89960	0.61906
34	0.784279	0.194038	0.323756	0.430020	0.459238	2.19133	−0.47819
35	0.548008	0.788351	0.831117	0.200790	0.823102	3.19137	1.07106
36	0.096383	0.844281	0.680927	0.656946	0.050867	2.32940	−0.26429
37	0.161502	0.972933	0.038113	0.515530	0.553788	2.24187	−0.39990
38	0.677552	0.232181	0.307234	0.588927	0.365403	2.17130	−0.50922
39	0.470454	0.267230	0.652802	0.633286	0.410964	2.43474	−0.10111
40	0.104377	0.819950	0.047036	0.189226	0.399502	1.56009	−1.45610



At the bottom of Table 4.3.2 is a density-scaled histogram of the forty "Z ratios," $\frac{y-5/2}{\sqrt{5/12}}$ (as listed in column C7). Notice the close agreement between the distribution of those ratios and $f_Z(z)$: What we see there is entirely consistent with the statement of Theorem 4.3.2.

Comment. Theorem 4.3.2 is an asymptotic result, yet it can provide surprisingly good approximations *even when n is very small.* Example 4.3.2 is a typical case in point: The uniform pdf over [0, 1] looks nothing like a bell-shaped curve, yet random samples as small as $n = 5$ yield sums that behave probabilistically much like the theoretical limit.

In general, samples from symmetric pdfs will produce sums that "converge" quickly to the theoretical limit. On the other hand, if the underlying pdf is sharply skewed—for example, $f_Y(y) = 10e^{-10y}$, $y > 0$—it would take a larger n to achieve the level of agreement present in Figure 4.3.2.

EXAMPLE 4.3.3

A random sample of size $n = 15$ is drawn from the pdf $f_Y(y) = 3(1-y)^2, 0 \leq y \leq 1$. Let $\overline{Y} = \left(\frac{1}{15}\right) \sum_{i=1}^{15} Y_i$. Use the central limit theorem to approximate $P\left(\frac{1}{8} \leq \overline{Y} \leq \frac{3}{8}\right)$.

Note, first of all, that

$$E(Y) = \int_0^1 y \cdot 3(1-y)^2 \, dy = \frac{1}{4}$$

and

$$\sigma^2 = \text{Var}(Y) = E(Y^2) - \mu^2 = \int_0^1 y^2 \cdot 3(1-y)^2 \, dy - \left(\frac{1}{4}\right)^2 = \frac{3}{80}$$

According, then, to the central limit theorem formulation that appears in the comment on page 302, the probability that $\overline{Y}$ will lie between $\frac{1}{8}$ and $\frac{3}{8}$ is approximately *0.99:*

$$\begin{aligned} P\left(\frac{1}{8} \leq \overline{Y} \leq \frac{3}{8}\right) &= P\left(\frac{\frac{1}{8}-\frac{1}{4}}{\sqrt{\frac{3}{80}}/\sqrt{15}} \leq \frac{\overline{Y}-\frac{1}{4}}{\sqrt{\frac{3}{80}}/\sqrt{15}} \leq \frac{\frac{3}{8}-\frac{1}{4}}{\sqrt{\frac{3}{80}}/\sqrt{15}}\right) \\ &= P(-2.50 \leq Z \leq 2.50) \\ &= 0.9876 \end{aligned}$$

EXAMPLE 4.3.4

In preparing next quarter's budget, the accountant for a small business has one hundred different expenditures to account for. Her predecessor listed each entry to the penny, but doing so grossly overstates the precision of the process. As a more truthful alternative, she intends to record each budget allocation to the nearest \$100. What is the probability that her total estimated budget will end up differing from the actual cost by more than

\$500? Assume that $Y_1, Y_2, \ldots, Y_{100}$, the rounding errors she makes on the one hundred items, are independent and uniformly distributed over the interval $[-\$50, +\$50]$.

Let

$$\begin{aligned} S_{100} &= Y_1 + Y_2 + \cdots + Y_{100} \\ &= \text{total rounding error} \end{aligned}$$

What the accountant wants to estimate is $P(|S_{100}| > \$500)$. By the distribution assumption made for each Y_i,

$$E(Y_i) = 0, \quad i = 1, 2, \ldots, 100$$

and

$$\begin{aligned} \text{Var}(Y_i) = E\left(Y_i^2\right) &= \int_{-50}^{50} \frac{1}{100} y^2\, dy \\ &= \frac{2500}{3} \end{aligned}$$

Therefore,

$$E(S_{100}) = E(Y_1 + Y_2 + \cdots + Y_{100}) = 0$$

and

$$\begin{aligned} \text{Var}(S_{100}) &= \text{Var}(Y_1 + Y_2 + \cdots + Y_{100}) = 100\left(\frac{2500}{3}\right) \\ &= \frac{250{,}000}{3} \end{aligned}$$

Applying Theorem 4.3.2, then, shows that her strategy has roughly an *8%* chance of being in error by more than \$500:

$$\begin{aligned} P(|S_{100}| > \$500) &= 1 - P(-500 \le S_{100} \le 500) \\ &= 1 - P\left(\frac{-500 - 0}{500/\sqrt{3}} \le \frac{S_{100} - 0}{500/\sqrt{3}} \le \frac{500 - 0}{500/\sqrt{3}}\right) \\ &\doteq 1 - P(-1.73 < Z < 1.73) \\ &= 0.0836. \end{aligned}$$

EXAMPLE 4.3.5

The annual number of earthquakes registering 2.5 or higher on the Richter scale and having an epicenter within forty miles of downtown Memphis follows a Poisson distribution with $\lambda = 6.5$. Calculate the exact probability that nine or more such earthquakes will strike next year, and compare that value to an approximation based on the central limit theorem.

If X denotes the number of earthquakes of that magnitude that will hit Memphis next year, the exact probability that $X \geq 9$ is a Poisson sum:

$$\begin{aligned} P(X \geq 9) = 1 - P(X \leq 8) = 1 - \sum_{x=0}^{8} \frac{e^{-6.5}(6.5)^x}{x!} \\ = 1 - 0.7916 \\ = 0.2084 \end{aligned}$$

For Poisson random variables, the ratio $\dfrac{W_1 + \cdots + W_n - n\mu}{\sqrt{n}\sigma}$ that appears in the central limit theorem reduces to $\dfrac{X - \lambda}{\sqrt{\lambda}}$ (see Question 4.3.18). Therefore,

$$\begin{aligned} P(X \geq 9) = 1 - P(X \leq 8) &= 1 - P(X \leq 8.5) \\ &= 1 - P\left(\frac{X - 6.5}{\sqrt{6.5}} \leq \frac{8.5 - 6.5}{\sqrt{6.5}}\right) \\ &\doteq 1 - P(Z \leq 0.78) \\ &= 0.2170 \end{aligned}$$

(Notice that the event "$X \leq 8$" is replaced with "$X \leq 8.5$" before applying the central limit theorem transformation. As always, the *continuity correction* is appropriate whenever a discrete probability model is being approximated by the area under a curve.)

QUESTIONS

4.3.15. A fair coin is tossed 200 times. Let $X_i = 1$ if the ith toss comes up heads and $X_i = 0$ otherwise, $i = 1, 2, \ldots, 200$. Calculate the central limit theorem approximation for $P(|X - E(X)| \leq 5)$. How does this differ from the DeMoivre-Laplace approximation?

4.3.16. Suppose that 100 fair dice are tossed. Estimate the probability that the sum of the faces showing exceeds 370. Include a continuity correction in your analysis.

4.3.17. Let X be the amount won or loss in betting \$5 on red in roulette. Then $p_X(5) = \frac{18}{38}$ and $p_X(-5) = \frac{20}{38}$. If a gambler bets on red 100 times, use the central limit theorem to estimate the probability that those wagers result in less than \$50 in losses.

4.3.18. If $X_1, X_2, \ldots, X_n$ are independent Poisson random variables with parameters $\lambda_1, \lambda_2, \ldots, \lambda_n$, respectively, and if $X = X_1 + X_2 + \cdots + X_n$, then X is a Poisson random variable with parameter $\lambda = \sum_{i=1}^{n} \lambda_i$ (recall Example 3.12.10). What specific form does the ratio in Theorem 4.3.2 take if the X_i's are Poisson random variables?

4.3.19. An electronics firm receives, on the average, 50 orders per week for a particular silicon chip. If the company has 60 chips on hand, use the central limit theorem to approximate the probability that they will be unable to fill all their orders for the upcoming week. Assume that weekly demands follow a Poisson distribution. *Hint:* See Question 4.3.18.

4.3.20. Considerable controversy has arisen over the possible aftereffects of a nuclear weapons test conducted in Nevada in 1957. Included as part of the test were some 3000 military and civilian "observers." Now, more than 40 years later, eight cases of leukemia have been diagnosed among those 3000. The expected number of cases, based on the demographic characteristics of the observers, was three. Assess the statistical significance of those findings. Calculate both an exact answer using the Poisson distribution as well as an approximation based on the central limit theorem.

The Normal Curve as a Model for Individual Measurements

Because of the central limit theorem, we know that sums (or averages) of virtually any set of random variables, when suitably scaled, have distributions that can be approximated by a standard normal curve. Perhaps even more surprising is the fact that many *individual* measurements, when suitably scaled, also have a standard normal distribution. Why should the latter be true? What do single observations have in common with samples of size n?

Astronomers in the early nineteenth century were among the first to understand the connection. Imagine looking through a telescope for the purpose of determining the location of a star. Conceptually, the data point, Y, eventually recorded is the sum of two components: (1) the star's *true* location μ^* (which remains unknown) and (2) measurement error. By definition, measurement error is the net effect of all those factors that cause the random variable Y to have a different value than μ^*. Typically, these effects will be additive, in which case the random variable can be written as a sum:

$$Y = \mu^* + W_1 + W_2 + \cdots + W_t \tag{4.3.3}$$

where W_1, for example, might represent the effect of atmospheric irregularities, W_2 the effect of seismic vibrations, W_3 the effect of parallax distortions, and so on.

If Equation 4.3.3 is a valid representation of the random variable Y, then it would follow that the central limit theorem applies to the *individual* Y_is. Moreover, if

$$E(Y) = E(\mu^* + W_1 + W_2 + \cdots + W_t) = \mu$$

and

$$\text{Var}(Y) = \text{Var}(\mu^* + W_1 + W_2 + \cdots + W_t) = \sigma^2$$

the ratio in Theorem 4.3.2 takes the form $\frac{Y - \mu}{\sigma}$. Furthermore, t is likely to be very large, so the approximation implied by the central limit theorem is essentially an equality—that is, *we take the pdf of* $\frac{Y - \mu}{\sigma}$ *to be* $f_Z(z)$.

Finding an actual formula for $f_Y(y)$, then, becomes an exercise in applying Theorem 3.4.3. Given that $\frac{Y - \mu}{\sigma} = Z$,

$$Y = \mu + \sigma Z$$

and

$$\begin{aligned} f_Y(y) &= \frac{1}{\sigma} f_Z\left(\frac{y - \mu}{\sigma}\right) \\ &= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \quad -\infty < y < \infty \end{aligned}$$

Definition 4.3.1. A random variable Y is said to be normally distributed with mean μ and variance σ^2 if

$$f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}, \qquad -\infty < y < \infty$$

The symbol $Y \sim N(\mu, \sigma^2)$ will sometimes be used to denote the fact that Y has a normal distribution with mean μ and variance σ^2.

Comment. Areas under an "arbitrary" normal distribution, $f_Y(y)$, are calculated by finding the equivalent area under the standard normal distribution, $f_Z(z)$:

$$P(a \leq Y \leq b) = P\left(\frac{a-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \leq Z \leq \frac{b-\mu}{\sigma}\right)$$

The ratio $\frac{Y-\mu}{\sigma}$ is often referred to as either a Z *transformation* or a Z *score*.

EXAMPLE 4.3.6

In many states a motorist is legally drunk or driving under the influence (DUI), if his or her blood alcohol concentration, Y, is 0.10% or higher. When a suspected DUI offender is pulled over, police often request a sobriety test. Although the breath analyzers used for that purpose are remarkably precise, the machines do exhibit a certain amount of measurement error. Because of that variability, the possibility exists that a driver's *true* blood alcohol concentration may be *under* 0.10% even though the analyzer gives a reading *over* 0.10%.

Experience has shown that repeated breath analyzer measurements taken on the same person produce a distribution of responses that can be described by a normal pdf with μ equal to the person's true blood alcohol concentration and σ equal to 0.004%. Suppose a driver is stopped at a roadblock on his way home from a party. Having celebrated a bit more than he should have, he has a true blood alcohol concentration of 0.095%, just barely under the legal limit. If he takes the breath analyzer test, what are the chances that he will be incorrectly booked on a DUI charge?

Since a DUI arrest occurs when $Y \geq 0.10\%$, we need to find $P(Y \geq 0.10)$ when $\mu = 0.095$ and $\sigma = 0.004$ (the percentage is irrelevant to any probability calculation and can be ignored). An application of the Z transformation shows that the driver has almost an *11*% chance of being falsely accused:

$$\begin{aligned} P(Y \geq 0.10) &= P\left(\frac{Y-0.095}{0.004} \geq \frac{0.10-0.095}{0.004}\right) \\ &= P(Z \geq 1.25) = 1 - P(Z < 1.25) \\ &= 1 - 0.8944 = 0.1056 \end{aligned}$$

Figure 4.3.5 shows $f_Y(y)$, $f_Z(z)$, and the two areas that are equal.

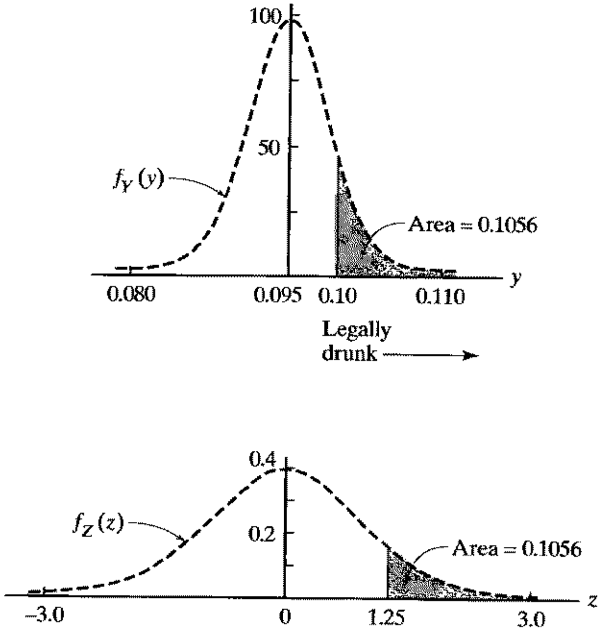


FIGURE 4.3.5

EXAMPLE 4.3.7

Mensa (from the Latin word for "mind") is an international society devoted to intellectual pursuits. Any person who has an IQ in the upper 2% of the general population is eligible to join. What is the *lowest* IQ that will qualify a person for membership? Assume that IQs are normally distributed with $\mu = 100$ and $\sigma = 16$.

Let the random variable Y denote a person's IQ, and let the constant y_L be the lowest IQ that qualifies someone to be a card-carrying Mensan. The two are related by a probability equation:

$$P(Y \geq y_L) = 0.02$$

or, equivalently,

$$P(Y < y_L) = 1 - 0.02 = 0.98 \tag{4.3.4}$$

(see Figure 4.3.6).

Applying the Z transformation to Equation 4.3.4 gives

$$P(Y < y_L) = P\left(\frac{Y - 100}{16} < \frac{y_L - 100}{16}\right) = P\left(Z < \frac{y_L - 100}{16}\right) = 0.98$$