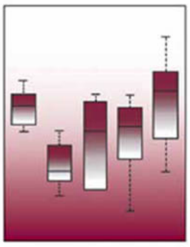

Mathematical Statistics and Data Analysis

THIRD EDITION

John A. Rice

DUXBURY ADVANCED SERIES

DUXBURY

THIRD EDITION

Mathematical Statistics and Data Analysis

John A. Rice
University of California, Berkeley

THOMSON
BROOKS/COLE

Australia • Brazil • Canada • Mexico • Singapore • Spain
United Kingdom • United States

THOMSON
BROOKS/COLE

***Mathematical Statistics and Data Analysis,* Third Edition**
John A. Rice

Acquisitions Editor: *Carolyn Crockett*
Assistant Editor: *Ann Day*
Editorial Assistant: *Elizabeth Gershman*
Technology Project Manager: *Fiona Chong*
Marketing Manager: *Joe Rogove*
Marketing Assistant: *Brian Smith*
Marketing Communications Manager: *Darlene Amidon-Brent*
Project Manager, Editorial Production: *Kelsey McGee*
Creative Director: *Rob Hugel*
Art Director: *Lee Friedman*
Print Buyer: *Karen Hunt*
Permissions Editor: *Bob Kauser*
Production Service: *Interactive Composition Corporation*
Text Designer: *Roy Neuhaus*
Copy Editor: *Victoria Thurman*
Illustrator: *Interactive Composition Corporation*
Cover Designer: *Denise Davidson*
Cover Printer: *Coral Graphic Services*
Compositor: *Interactive Composition Corporation*
Printer: *R.R. Donnelley/Crawfordsville*

Printed in the United States of America
1 2 3 4 5 6 7 10 09 08 07 06

Thomson Higher Education
10 Davis Drive
Belmont, CA 94002-3098
USA

Library of Congress Control Number: 2005938314

Student Edition: ISBN 0-534-39942-8

We must be careful not to confuse data with the abstractions we use to analyze them.

WILLIAM JAMES (1842–1910)

Contents

Preface

Intended Audience

This text is intended for juniors, seniors, or beginning graduate students in statistics, mathematics, natural sciences, and engineering as well as for adequately prepared students in the social sciences and economics. A year of calculus, including Taylor Series and multivariable calculus, and an introductory course in linear algebra are prerequisites.

This Book's Objectives

This book reflects my view of what a first, and for many students a last, course in statistics should be. Such a course should include some traditional topics in mathematical statistics (such as methods based on likelihood), topics in descriptive statistics and data analysis with special attention to graphical displays, aspects of experimental design, and realistic applications of some complexity. It should also reflect the integral role played by computers in statistics. These themes, properly interwoven, can give students a view of the nature of modern statistics. The alternative of teaching two separate courses, one on theory and one on data analysis, seems to me artificial. Furthermore, many students take only one course in statistics and do not have time for two or more.

Analysis of Data and the Practice of Statistics

In order to draw the above themes together, I have endeavored to write a book closely tied to the practice of statistics. It is in the analysis of real data that one sees the roles played by both formal theory and informal data analytic methods. I have organized this book around various kinds of problems that entail the use of statistical methods and have included many real examples to motivate and introduce the theory. Among the advantages of such an approach are that theoretical constructs are presented in meaningful contexts, that they are gradually supplemented and reinforced, and that they are integrated with more informal methods. This is, I think, a fitting approach to statistics, the historical development of which has been spurred on primarily by practical needs rather than by abstract or aesthetic considerations. At the same time, I have not shied away from using the mathematics that the students are supposed to know.

The Third Edition

Eighteen years have passed since the first edition of this book was published and eleven years since the second. Although the basic intent and stucture of the book have not changed, the new editions reflect developments in the discipline of statistics, primarily the computational revolution.

The most significant change in this edition is the treatment of Bayesian inference. I moved the material from the last chapter, a point that was never reached by many instructors, and integrated it into earlier chapters. Bayesian inference is now first previewed in Chapter 3, in the context of conditional distributions. It is then placed side-by-side with frequentist methods in Chapter 8, where it complements the material on maximum likelihood estimation very naturally. The introductory section on hypothesis testing in Chapter 9 now begins with a Bayesian formulation before moving on to the Neyman-Pearson paradigm. One advantage of this is that the fundamental importance of the likelihood ratio is now much more apparent. In applications, I stress uninformative priors and show the similarity of the qualitative conclusions that would be reached by frequentist and Bayesian methods.

Other new material includes the use of examples from genomics and financial statistics in the probability chapters. In addition to its value as topically relevant, this material naturally reinforces basic concepts. For example, the material on copulas underscores the relationships of marginal and joint distributions. Other changes include the introduction of scatterplots and correlation coefficients within the context of exploratory data analysis in Chapter 10 and a brief introduction to nonparametric smoothing via local linear least squares in Chapter 14. There are nearly 100 new problems, mainly in Chapters 7–14, including several new data sets. Some of the data sets are sufficiently substantial to be the basis for computer lab assignments. I also elucidated many passages that were obscure in earlier editions.

Brief Outline

A complete outline can be found, of course, in the Table of Contents. Here I will just highlight some points and indicate various curricular options for the instructor.

The first six chapters contain an introduction to probability theory, particularly those aspects most relevant to statistics. Chapter 1 introduces the basic ingredients of probability theory and elementary combinatorial methods from a non measure theoretic point of view. In this and the other probability chapters, I tried to use real-world examples rather than balls and urns whenever possible.

The concept of a random variable is introduced in Chapter 2. I chose to discuss discrete and continuous random variables together, instead of putting off the continuous case until later. Several common distributions are introduced. An advantage of this approach is that it provides something to work with and develop in later chapters.

Chapter 3 continues the treatment of random variables by going into joint distributions. The instructor may wish to skip lightly over Jacobians; this can be done with little loss of continuity, since they are rarely used in the rest of the book. The material in Section 3.7 on extrema and order statistics can be omitted if the instructor is willing to do a little backtracking later.

Expectation, variance, covariance, conditional expectation, and moment-generating functions are taken up in Chapter 4. The instructor may wish to pass lightly over conditional expectation and prediction, especially if he or she does not plan to cover sufficiency later. The last section of this chapter introduces the δ method, or the method of propagation of error. This method is used several times in the statistics chapters.

The law of large numbers and the central limit theorem are proved in Chapter 5 under fairly strong assumptions.

Chapter 6 is a compendium of the common distributions related to the normal and sampling distributions of statistics computed from the usual normal random sample. I don't spend a lot of time on this material here but do develop the necessary facts as they are needed in the statistics chapters. It is useful for students to have these distributions collected in one place.

Chapter 7 is on survey sampling, an unconventional, but in some ways natural, beginning to the study of statistics. Survey sampling is an area of statistics with which most students have some vague familiarity, and a set of fairly specific, concrete statistical problems can be naturally posed. It is a context in which, historically, many important statistical concepts have developed, and it can be used as a vehicle for introducing concepts and techniques that are developed further in later chapters, for example:

- The idea of an estimate as a random variable with an associated sampling distribution
- The concepts of bias, standard error, and mean squared error
- Confidence intervals and the application of the central limit theorem
- An exposure to notions of experimental design via the study of stratified estimates and the concept of relative efficiency
- Calculation of expectations, variances, and covariances

One of the unattractive aspects of survey sampling is that the calculations are rather grubby. However, there is a certain virtue in this grubbiness, and students are given practice in such calculations. The instructor has quite a lot of flexibility as to how deeply to cover the concepts in this chapter. The sections on ratio estimation and stratification are optional and can be skipped entirely or returned to at a later time without loss of continuity.

Chapter 8 is concerned with parameter estimation, a subject that is motivated and illustrated by the problem of fitting probability laws to data. The method of moments, the method of maximum likelihood, and Bayesian inference are developed. The concept of efficiency is introduced, and the Cramér-Rao Inequality is proved. Section 8.8 introduces the concept of sufficiency and some of its ramifications. The material on the Cramér-Rao lower bound and on sufficiency can be skipped; to my mind, the importance of sufficiency is usually overstated. Section 8.7.1 (the negative binomial distribution) can also be skipped.

Chapter 9 is an introduction to hypothesis testing with particular application to testing for goodness of fit, which ties in with Chapter 8. (This subject is further developed in Chapter 11.) Informal, graphical methods are presented here as well. Several of the last sections of this chapter can be skipped if the instructor is pressed for time. These include Section 9.6 (the Poisson dispersion test), Section 9.7 (hanging rootograms), and Section 9.9 (tests for normality).

A variety of descriptive methods are introduced in Chapter 10. Many of these techniques are used in later chapters. The importance of graphical procedures is stressed, and notions of robustness are introduced. The placement of a chapter on descriptive methods this late in a book may seem strange. I chose to do so because descriptive procedures usually have a stochastic side and, having been through the three chapters preceding this one, students are by now better equipped to study the statistical behavior of various summary statistics (for example, a confidence interval for the median). When I teach the course, I introduce some of this material earlier. For example, I have students make boxplots and histograms from samples drawn in labs on survey sampling. If the instructor wishes, the material on survival and hazard functions can be skipped.

Classical and nonparametric methods for two-sample problems are introduced in Chapter 11. The concepts of hypothesis testing, first introduced in Chapter 9, are further developed. The chapter concludes with some discussion of experimental design and the interpretation of observational studies.

The first eleven chapters are the heart of an introductory course; the theoretical constructs of estimation and hypothesis testing have been developed, graphical and descriptive methods have been introduced, and aspects of experimental design have been discussed.

The instructor has much more freedom in selecting material from Chapters 12 through 14. In particular, it is not necessary to proceed through these chapters in the order in which they are presented.

Chapter 12 treats the one-way and two-way layouts via analysis of variance and nonparametric techniques. The problem of multiple comparisons, first introduced at the end of Chapter 11, is discussed.

Chapter 13 is a rather brief treatment of the analysis of categorical data. Likelihood ratio tests are developed for homogeneity and independence. McNemar's test is presented and finally, estimation of the odds ratio is motivated by a discussion of prospective and retrospective studies.

Chapter 14 concerns linear least squares. Simple linear regression is developed first and is followed by a more general treatment using linear algebra. I chose to employ matrix algebra but keep the level of the discussion as simple and concrete as possible, not going beyond concepts typically taught in an introductory one-quarter course. In particular, I did not develop a geometric analysis of the general linear model or make any attempt to unify regression and analysis of variance. Throughout this chapter, theoretical results are balanced by more qualitative data analytic procedures based on analysis of residuals. At the end of the chapter, I introduce nonparametric regression via local linear least squares.

Computer Use and Problem Solving

Computation is an integral part of contemporary statistics. It is essential for data analysis and can be an aid to clarifying basic concepts. My students use the open-source package R, which they can install on their own computers. Other packages could be used as well but I do not discuss any particular programs in the text. The data in the text are available on the CD that is bound in the U.S. edition or can be downloaded from www.thomsonedu.com/statistics.

This book contains a large number of problems, ranging from routine reinforcement of basic concepts to some that students will find quite difficult. I think that problem solving, especially of nonroutine problems, is very important.

Acknowledgments

I am indebted to a large number of people who contributed directly and indirectly to the first edition. Earlier versions were used in courses taught by Richard Olshen, Yosi Rinnot, Donald Ylvisaker, Len Haff, and David Lane, who made many helpful comments. Students in their classes and in my own had many constructive comments. Teaching assistants, especially Joan Staniswalis, Roger Johnson, Terri Bittner, and Peter Kim, worked through many of the problems and found numerous errors. Many reviewers provided useful suggestions: Rollin Brant, University of Toronto; George Casella, Cornell University; Howard B. Christensen, Brigham Young University; David Fairley, Ohio State University; Peter Guttorp, University of Washington; Hari Iyer, Colorado State University; Douglas G. Kelly, University of North Carolina; Thomas Leonard, University of Wisconsin; Albert S. Paulson, Rensselaer Polytechnic Institute; Charles Peters, University of Houston, University Park; Andrew Rukhin, University of Massachusetts, Amherst; Robert Schaefer, Miami University; and Ruth Williams, University of California, San Diego. Richard Royall and W. G. Cumberland kindly provided the data sets used in Chapter 7 on survey sampling. Several other data sets were brought to my attention by statisticians at the National Bureau of Standards, where I was fortunate to spend a year while on sabbatical. I deeply appreciate the patience, persistence, and faith of my editor, John Kimmel, in bringing this project to fruition.

The candid comments of many students and faculty who used the first edition of the book were influential in the creation of the second edition. In particular I would like to thank Ian Abramson, Edward Bedrick, Jon Frank, Richard Gill, Roger Johnson, Torgny Lindvall, Michael Martin, Deb Nolan, Roger Pinkham, Yosi Rinott, Philip Stark, and Bin Yu; I apologize to any individuals who have inadvertently been left off this list. Finally, I would like to thank Alex Kugushev for his encouragement and support in carrying out the revision and the work done by Terri Bittner in carefully reading the manuscript for accuracy and in the solutions of the new problems.

Many people contributed to the third edition. I would like to thank the reviewers of this edition: Marten Wegkamp, Yale University; Aparna Huzurbazar, University of New Mexico; Laura Bernhofen, Clark University; Joe Glaz, University of Connecticut; and Michael Minnotte, Utah State University. I deeply appreciate many readers for generously taking the time to point out errors and make suggestions on improving the exposition. In particular, Roger Pinkham sent many helpful email messages and Nick Cox provided a very long list of grammatical errors. Alice Hsiaw made detailed comments on Chapters 7–14. I also wish to thank Ani Adhikari, Paulo Berata, Patrick Brewer, Sang-Hoon Cho Gier Eide, John Einmahl, David Freedman, Roger Johnson, Paul van der Laan, Patrick Lee, Yi Lin, Jim Linnemann, Rasaan Moshesh, Eugene Schuster, Dylan Small, Luis Tenorio, Richard De Veaux, and Ping Zhang. Bob Stine contributed financial data, Diane Cook provided data on Italian olive oils, and Jim Albert provided a baseball data set that nicely illustrates regression toward the mean. Rainer Sachs provided the lovely data on chromatin separations. I thank my editor, Carolyn Crockett, for her graceful persistence and patience in bringing about this revision, and also the energetic production team. I apologize to any others whose names have inadvertently been left off this list.

CHAPTER **1**

Probability

1.1 Introduction

The idea of probability, chance, or randomness is quite old, whereas its rigorous axiomatization in mathematical terms occurred relatively recently. Many of the ideas of probability theory originated in the study of games of chance. In this century, the mathematical theory of probability has been applied to a wide variety of phenomena; the following are some representative examples:

- Probability theory has been used in genetics as a model for mutations and ensuing natural variability, and plays a central role in bioinformatics.
- The kinetic theory of gases has an important probabilistic component.
- In designing and analyzing computer operating systems, the lengths of various queues in the system are modeled as random phenomena.
- There are highly developed theories that treat noise in electrical devices and communication systems as random processes.
- Many models of atmospheric turbulence use concepts of probability theory.
- In operations research, the demands on inventories of goods are often modeled as random.
- Actuarial science, which is used by insurance companies, relies heavily on the tools of probability theory.
- Probability theory is used to study complex systems and improve their reliability, such as in modern commercial or military aircraft.
- Probability theory is a cornerstone of the theory of finance.

The list could go on and on.

This book develops the basic ideas of probability and statistics. The first part explores the theory of probability as a mathematical model for chance phenomena. The second part of the book is about statistics, which is essentially concerned with procedures for analyzing data, especially data that in some vague sense have a random character. To comprehend the theory of statistics, you must have a sound background in probability.

1.2 Sample Spaces

Probability theory is concerned with situations in which the outcomes occur randomly. Generically, such situations are called *experiments*, and the set of all possible outcomes is the **sample space** corresponding to an experiment. The sample space is denoted by Ω, and an element of Ω is denoted by ω. The following are some examples.

EXAMPLE **A** Driving to work, a commuter passes through a sequence of three intersections with traffic lights. At each light, she either stops, s, or continues, c. The sample space is the set of all possible outcomes:

$$\Omega = \{ccc, ccs, css, csc, sss, ssc, scc, scs\}$$

where csc, for example, denotes the outcome that the commuter continues through the first light, stops at the second light, and continues through the third light. ■

EXAMPLE **B** The number of jobs in a print queue of a mainframe computer may be modeled as random. Here the sample space can be taken as

$$\Omega = \{0, 1, 2, 3, \ldots\}$$

that is, all the nonnegative integers. In practice, there is probably an upper limit, N, on how large the print queue can be, so instead the sample space might be defined as

$$\Omega = \{0, 1, 2, \ldots, N\}$$ ■

EXAMPLE **C** Earthquakes exhibit very erratic behavior, which is sometimes modeled as random. For example, the length of time between successive earthquakes in a particular region that are greater in magnitude than a given threshold may be regarded as an experiment. Here Ω is the set of all nonnegative real numbers:

$$\Omega = \{t \mid t \geq 0\}$$ ■

We are often interested in particular subsets of Ω, which in probability language are called **events.** In Example A, the event that the commuter stops at the first light is the subset of Ω denoted by

$$A = \{sss, ssc, scc, scs\}$$

(Events, or subsets, are usually denoted by italic uppercase letters.) In Example B, the event that there are fewer than five jobs in the print queue can be denoted by

$$A = \{0, 1, 2, 3, 4\}$$

The algebra of set theory carries over directly into probability theory. The **union** of two events, A and B, is the event C that either A occurs or B occurs or both occur: $C = A \cup B$. For example, if A is the event that the commuter stops at the first light (listed before), and if B is the event that she stops at the third light,

$$B = \{sss, scs, ccs, css\}$$

then C is the event that she stops at the first light or stops at the third light and consists of the outcomes that are in A or in B or in both:

$$C = \{sss, ssc, scc, scs, ccs, css\}$$

The **intersection** of two events, $C = A \cap B$, is the event that both A and B occur. If A and B are as given previously, then C is the event that the commuter stops at the first light and stops at the third light and thus consists of those outcomes that are common to both A and B:

$$C = \{sss, scs\}$$

The **complement** of an event, A^c, is the event that A does not occur and thus consists of all those elements in the sample space that are not in A. The complement of the event that the commuter stops at the first light is the event that she continues at the first light:

$$A^c = \{ccc, ccs, css, csc\}$$

You may recall from previous exposure to set theory a rather mysterious set called the empty set, usually denoted by $\emptyset$. The **empty set** is the set with no elements; it is the event with no outcomes. For example, if A is the event that the commuter stops at the first light and C is the event that she continues through all three lights, $C = \{ccc\}$, then A and C have no outcomes in common, and we can write

$$A \cap C = \emptyset$$

In such cases, A and C are said to be **disjoint.**

Venn diagrams, such as those in Figure 1.1, are often a useful tool for visualizing set operations.

The following are some laws of set theory.

Commutative Laws:

$$A \cup B = B \cup A$$
$$A \cap B = B \cap A$$

Associative Laws:

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

Distributive Laws:

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

Of these, the distributive laws are the least intuitive, and you may find it instructive to illustrate them with Venn diagrams.

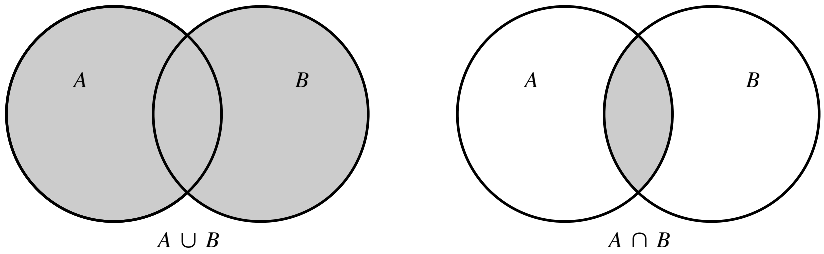

FIGURE **1.1** Venn diagrams of $A \cup B$ and $A \cap B$.

1.3 Probability Measures

A **probability measure** on Ω is a function P from subsets of Ω to the real numbers that satisfies the following axioms:

1. $P(\Omega) = 1$.
2. If $A \subset \Omega$, then $P(A) \geq 0$.
3. If A_1 and A_2 are disjoint, then

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

More generally, if $A_1, A_2, \ldots, A_n, \ldots$ are mutually disjoint, then

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

The first two axioms are obviously desirable. Since Ω consists of all possible outcomes, $P(\Omega) = 1$. The second axiom simply states that a probability is nonnegative. The third axiom states that if A and B are disjoint—that is, have no outcomes in common—then $P(A \cup B) = P(A) + P(B)$ and also that this property extends to limits. For example, the probability that the print queue contains either one or three jobs is equal to the probability that it contains one plus the probability that it contains three.

The following properties of probability measures are consequences of the axioms.

Property A $P(A^c) = 1 - P(A)$. This property follows since A and A^c are disjoint with $A \cup A^c = \Omega$ and thus, by the first and third axioms, $P(A) + P(A^c) = 1$. In words, this property says that the probability that an event does not occur equals one minus the probability that it does occur.

Property B $P(\emptyset) = 0$. This property follows from Property A since $\emptyset = \Omega^c$. In words, this says that the probability that there is no outcome at all is zero.

Property C If $A \subset B$, then $P(A) \leq P(B)$. This property states that if B occurs whenever A occurs, then $P(A) \leq P(B)$. For example, if whenever it rains (A) it is cloudy (B), then the probability that it rains is less than or equal to the probability that it is cloudy. Formally, it can be proved as follows: B can be expressed as the union of two disjoint sets:

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

Then, from the third axiom,

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

and thus

$$P(A) = P(B) - P(B \cap A^c) \leq P(B)$$

Property D Addition Law $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. This property is easy to see from the Venn diagram in Figure 1.2. If $P(A)$ and $P(B)$ are added together, $P(A \cap B)$ is counted twice. To prove it, we decompose $A \cup B$ into three disjoint subsets, as shown in Figure 1.2:

$$\begin{aligned} C &= A \cap B^c \\ D &= A \cap B \\ E &= A^c \cap B \end{aligned}$$

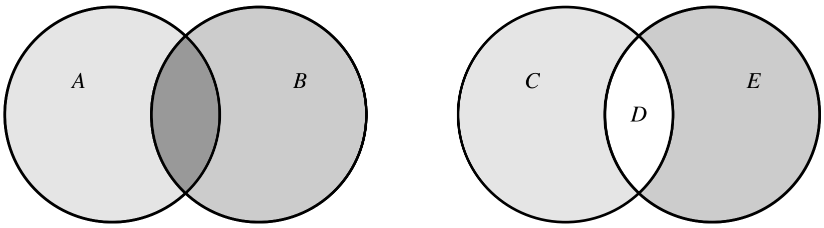

FIGURE **1.2** Venn diagram illustrating the addition law.

We then have, from the third axiom,

$$P(A \cup B) = P(C) + P(D) + P(E)$$

Also, $A = C \cup D$, and C and D are disjoint; so $P(A) = P(C) + P(D)$. Similarly, $P(B) = P(D) + P(E)$. Putting these results together, we see that

$$\begin{aligned} P(A) + P(B) &= P(C) + P(E) + 2P(D) \\ &= P(A \cup B) + P(D) \end{aligned}$$

or

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

EXAMPLE **A** Suppose that a fair coin is thrown twice. Let A denote the event of heads on the first toss, and let B denote the event of heads on the second toss. The sample space is

$$\Omega = \{hh, ht, th, tt\}$$

We assume that each elementary outcome in Ω is equally likely and has probability $\frac{1}{4}$. $C = A \cup B$ is the event that heads comes up on the first toss or on the second toss. Clearly, $P(C) \neq P(A) + P(B) = 1$. Rather, since $A \cap B$ is the event that heads comes up on the first toss and on the second toss,

$$P(C) = P(A) + P(B) - P(A \cap B) = .5 + .5 - .25 = .75 \qquad \blacksquare$$

EXAMPLE **B** An article in the *Los Angeles Times* (August 24, 1987) discussed the statistical risks of AIDS infection:

> Several studies of sexual partners of people infected with the virus show that a single act of unprotected vaginal intercourse has a surprisingly low risk of infecting the uninfected partner—perhaps one in 100 to one in 1000. For an average, consider the risk to be one in 500. If there are 100 acts of intercourse with an infected partner, the odds of infection increase to one in five.
>
> Statistically, 500 acts of intercourse with one infected partner or 100 acts with five partners lead to a 100% probability of infection (statistically, not necessarily in reality).

Following this reasoning, 1000 acts of intercourse with one infected partner would lead to a probability of infection equal to 2 (statistically, but not necessarily in reality). To see the flaw in the reasoning that leads to this conclusion, consider two acts of intercourse. Let A_1 denote the event that infection occurs on the first act and let A_2 denote the event that infection occurs on the second act. Then the event that infection occurs is $B = A_1 \cup A_2$ and

$$P(B) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \leq P(A_1) + P(A_2) = \frac{2}{500} \qquad \blacksquare$$

1.4 Computing Probabilities: Counting Methods

Probabilities are especially easy to compute for finite sample spaces. Suppose that $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ and that $P(\omega_i) = p_i$. To find the probability of an event A, we simply add the probabilities of the ω_i that constitute A.

EXAMPLE **A** Suppose that a fair coin is thrown twice and the sequence of heads and tails is recorded. The sample space is

$$\Omega = \{hh, ht, th, tt\}$$

As in Example A of the previous section, we assume that each outcome in Ω has probability .25. Let A denote the event that at least one head is thrown. Then $A = \{hh, ht, th\}$, and $P(A) = .75$. ■

This is a simple example of a fairly common situation. The elements of Ω all have equal probability; so if there are N elements in Ω, each of them has probability $1/N$. If A can occur in any of n mutually exclusive ways, then $P(A) = n/N$, or

$$P(A) = \frac{\text{number of ways } A \text{ can occur}}{\text{total number of outcomes}}$$

Note that this formula holds only if all the outcomes are equally likely. In Example A, if only the number of heads were recorded, then Ω would be $\{0, 1, 2\}$. These outcomes are not equally likely, and $P(A)$ is not $\frac{2}{3}$. ■

EXAMPLE **B** *Simpson's Paradox*

A black urn contains 5 red and 6 green balls, and a white urn contains 3 red and 4 green balls. You are allowed to choose an urn and then choose a ball at random from the urn. If you choose a red ball, you get a prize. Which urn should you choose to draw from? If you draw from the black urn, the probability of choosing a red ball is $\frac{5}{11} = .455$ (the number of ways you can draw a red ball divided by the total number of outcomes). If you choose to draw from the white urn, the probability of choosing a red ball is $\frac{3}{7} = .429$, so you should choose to draw from the black urn.

Now consider another game in which a second black urn has 6 red and 3 green balls, and a second white urn has 9 red and 5 green balls. If you draw from the black urn, the probability of a red ball is $\frac{6}{9} = .667$, whereas if you choose to draw from the white urn, the probability is $\frac{9}{14} = .643$. So, again you should choose to draw from the black urn.

In the final game, the contents of the second black urn are added to the first black urn, and the contents of the second white urn are added to the first white urn. Again, you can choose which urn to draw from. Which should you choose? Intuition says choose the black urn, but let's calculate the probabilities. The black urn now contains 11 red and 9 green balls, so the probability of drawing a red ball from it is $\frac{11}{20} = .55$. The white urn now contains 12 red and 9 green balls, so the probability of drawing a red ball from it is $\frac{12}{21} = .571$. So, you should choose the white urn. This counterintuitive result is an example of *Simpson's paradox*. For an example that occurred in real life, see Section 11.4.7. For more amusing examples, see Gardner (1976). ■

In the preceding examples, it was easy to count the number of outcomes and calculate probabilities. To compute probabilities for more complex situations, we must develop systematic ways of counting outcomes, which are the subject of the next two sections.

1.4.1 The Multiplication Principle

The following is a statement of the very useful multiplication principle.