A First Course in Probability

Tenth Edition

Sheldon Ross

A First Course in Probability

A First Course in Probability

Tenth Edition

Sheldon Ross
University of Southern California

Pearson

Director, Portfolio Management: Deirdre Lynch

Courseware Portfolio Manager: Suzanna Bainbridge

Courseware Portfolio Management Assistant: Morgan Danna

Content Producer: Tara Corpuz

Managing Producer: Scott Disanno

Producer: Jon Wooding

Product Marketing Manager: Yvonne Vannatta

Product Marketing Assistant: Jon Bryant

Field Marketing Manager: Evan St. Cyr

Senior Author Support/Technology Specialist: Joe Vetere

Manager, Rights and Permissions: Gina Cheselka

Cover Design: Studio Montage

Production Coordination, Composition, and Illustrations: Integra Software Services Pvt. Ltd.

Manufacturing Buyer: Carol Melville, LSC Communications

Cover Image: Adrienne Bresnahan/Moment/Getty Images

Credits and acknowledgments borrowed from other sources and reproduced, with permission, in this textbook appear on page **506** within text.

Library of Congress Cataloging-in-Publication Data

Names: Ross, Sheldon M., author.

Title: A first course in probability / Sheldon Ross (University of Southern California).

Other titles: Probability

Description: Tenth edition. | Boston: Pearson, 2018. | Includes index.

Identifiers: LCCN 2018006823 | ISBN 9780134753119 | ISBN 0134753119

Subjects: LCSH: Probabilities—Textbooks.

Classification: LCC QA273 .R83 2018 | DDC 519.2—dc23

LC record available at **https://lccn.loc.gov/2018006823**

1 18

ISBN-10: 0-13-475311-9

ISBN-13: 978-0-13-475311-9

Pearson

For Rebecca

Contents

Preface

“We see that the theory of probability is at bottom only common sense reduced to calculation; it makes us appreciate with exactitude what reasonable minds feel by a sort of instinct, often without being able to account for it... It is remarkable that this science, which originated in the consideration of games of chance, should have become the most important object of human knowledge.... The most important questions of life are, for the most part, really only problems of probability.” So said the famous French mathematician and astronomer (the “Newton of France”) Pierre-Simon, Marquis de Laplace. Although many people believe that the famous marquis, who was also one of the great contributors to the development of probability, might have exaggerated somewhat, it is nevertheless true that probability theory has become a tool of fundamental importance to nearly all scientists, engineers, medical practitioners, jurists, and industrialists. In fact, the enlightened individual had learned to ask not “Is it so?” but rather “What is the probability that it is so?”

General Approach and Mathematical Level

This book is intended as an elementary introduction to the theory of probability for students in mathematics, statistics, engineering, and the sciences (including computer science, biology, the social sciences, and management science) who possess the prerequisite knowledge of elementary calculus. It attempts to present not only the mathematics of probability theory, but also, through numerous examples, the many diverse possible applications of this subject.

Content and Course Planning

Chapter 1 presents the basic principles of combinatorial analysis, which are most useful in computing probabilities.

Chapter 2 handles the axioms of probability theory and shows how they can be applied to compute various probabilities of interest.

Chapter 3 deals with the extremely important subjects of conditional probability and independence of events. By a series of examples, we illustrate how conditional probabilities come into play not only when some partial information is available, but also as a tool to enable us to compute probabilities more easily, even when no partial information is present. This extremely important technique of obtaining probabilities by “conditioning” reappears in **Chapter 7**, where we use it to obtain expectations.

The concept of random variables is introduced in **Chapters 4**, **5**, and **6**. Discrete random variables are dealt with in **Chapter 4**, continuous random variables in **Chapter 5**, and jointly distributed random variables in **Chapter 6**. The important concepts of the expected value and the variance of a random variable are introduced in **Chapters 4** and **5**, and these quantities are then determined for many of the common types of random variables.

Additional properties of the expected value are considered in **Chapter 7**. Many examples illustrating the usefulness of the result that the expected value of a sum of random variables is equal to the sum of their expected values are presented. Sections on conditional expectation, including its use in prediction, and on moment-generating functions are contained in this chapter. In addition, the final section introduces the multivariate normal distribution and presents a simple proof concerning the joint distribution of the sample mean and sample variance of a sample from a normal distribution.

Chapter 8 presents the major theoretical results of probability theory. In particular, we prove the strong law of large numbers and the central limit theorem. Our proof of the strong law is a relatively simple one that assumes that the random variables have a finite fourth moment, and our proof of the central limit theorem assumes Levy’s continuity theorem. This chapter also presents such probability inequalities as Markov’s inequality, Chebyshev's inequality, and Chernoff bounds. The final section of **Chapter 8** gives a bound on the error involved when a probability concerning a sum of independent Bernoulli random variables is approximated by the corresponding probability of a Poisson random variable having the same expected value.

Chapter 9 presents some additional topics, such as Markov chains, the Poisson process, and an introduction to information and coding theory, and **Chapter 10** considers simulation.

As in the previous edition, three sets of exercises are given at the end of each chapter. They are designated as **Problems, Theoretical Exercises**, and **Self-Test Problems and Exercises**. This last set of exercises, for which complete solutions appear in Solutions to Self-Test Problems and Exercises, is designed to help students test their comprehension and study for exams.

Changes for the Tenth Edition

The tenth edition continues the evolution and fine tuning of the text. Aside from a multitude of small changes made to increase the clarity of the text, the new edition includes many new and updated problems, exercises, and text material chosen both for inherent interest and for their use in building student intuition about probability. Illustrative of these goals are Examples 4n of **Chapter 3**, which deals with computing NCAA basketball tournament win probabilities, and Example 5b of **Chapter 4**, which introduces the friendship paradox. There is also new material on the Pareto distribution (introduced in **Section 5.6.5**), on Poisson limit results (in **Section 8.5**), and on the Lorenz curve (in **Section 8.7**).

Acknowledgments

I would like to thank the following people who have graciously taken the time to contact me with comments for improving the text: Amir Ardestani, Polytechnic University of Teheran; Joe Blitzstein, Harvard University; Peter Nuesch, University of Lausaunne; Joseph Mitchell, SUNY, Stony Brook; Alan Chambless, actuary; Robert Kriner; Israel David, Ben-Gurion University; T. Lim, George Mason University; Wei Chen, Rutgers; D. Monrad, University of Illinois; W. Rosenberger, George Mason University; E. Ionides, University of Michigan; J. Corvino, Lafayette College; T. Seppalainen, University of Wisconsin; Jack Goldberg; University of Michigan; Sunil Dhar, New Jersey Institute of Technology; Vladislav Kargin, Stanford University; Marlene Miller; Ahmad Parsian; and Fritz Scholz, University of Washington.

I would also like to especially thank the reviewers of the ninth and tenth editions: Richard Laugesen, University of Illinois; Stacey Hancock, Clark University; Stefan Heinz, University of Wyoming; and Brian Thelen, University of Michigan; Mark Ward, Purdue University. I would like to thank the accuracy checker, Stacey Hancock (Montana State University), for her careful review.

Finally, I would like to thank the following reviewers for their many helpful comments. Reviewers of the tenth edition are marked with an asterisk.

K. B. Athreya, *Iowa State University*

Richard Bass, *University of Connecticut*

Robert Bauer, *University of Illinois at Urbana-Champaign*

Phillip Beckwith, *Michigan Tech*

Arthur Benjamin, *Harvey Mudd College*

Geoffrey Berresford, *Long Island University*

Baidurya Bhattacharya, *University of Delaware*

Howard Bird, *St. Cloud State University*

Shahar Boneh, *Metropolitan State College of Denver*

Jean Cadet, *State University of New York at Stony Brook*

Steven Chiappari, *Santa Clara University*

Nicolas Christou, *University of California, Los Angeles*

James Clay, *University of Arizona at Tucson*

Francis Conlan, *University of Santa Clara*

Justin Corvino, *Lafayette College*

Jay DeVore, *California Polytechnic University, San Luis Obispo*

Scott Emerson, *University of Washington*

Thomas R. Fischer, *Texas A & M University*

Anant Godbole, *Michigan Technical University*

Zakkula Govindarajulu, *University of Kentucky*

Richard Groeneveld, *Iowa State University*

*Stacey Hancock, *Clark University*

Mike Hardy, *Massachusetts Institute of Technology*

Bernard Harris, *University of Wisconsin*

Larry Harris, *University of Kentucky*

David Heath, *Cornell University*

*Stefan Heinz, *University of Wyoming*

Stephen Herschkorn, *Rutgers University*

Julia L. Higle, *University of Arizona*

Mark Huber, *Duke University*

Edward Ionides, *University of Michigan*

Anastasia Ivanova, *University of North Carolina*

Hamid Jafarkhani, *University of California, Irvine*

Chuanshu Ji, *University of North Carolina, Chapel Hill*

Robert Keener, *University of Michigan*

*Richard Laugesen, *University of Illinois*

Fred Leysieffer, *Florida State University*

Thomas Liggett, *University of California, Los Angeles*

Helmut Mayer, *University of Georgia*

Bill McCormick, *University of Georgia*

Ian McKeague, *Florida State University*

R. Miller, *Stanford University*

Ditlev Monrad, *University of Illinois*

Robb J. Muirhead, *University of Michigan*

Joe Naus, *Rutgers University*

Nhu Nguyen, *New Mexico State University*

Ellen O'Brien, *George Mason University*

N. U. Prabhu, *Cornell University*

Kathryn Prewitt, *Arizona State University*

Jim Propp, *University of Wisconsin*

William F. Rosenberger, *George Mason University*

Myra Samuels, *Purdue University*

I. R. Savage, *Yale University*

Art Schwartz, *University of Michigan at Ann Arbor*

Therese Shelton, *Southwestern University*

Malcolm Sherman, *State University of New York at Albany*

Murad Taqqu, *Boston University*

*Brian Thelen, *University of Michigan*

Eli Upfal, *Brown University*

Ed Wheeler, *University of Tennessee*

Allen Webster, *Bradley University*

S. R.
smross@usc.edu

Chapter 1 Combinatorial Analysis

Contents

1.1 Introduction

Here is a typical problem of interest involving probability: A communication system is to consist of n seemingly identical antennas that are to be lined up in a linear order. The resulting system will then be able to receive all incoming signals and will be called *functional* as long as no two consecutive antennas are defective. If it turns out that exactly m of the n antennas are defective, what is the probability that the resulting system will be functional? For instance, in the special case where n=4 and m=2, there are 6 possible system configurations, namely,

011001011010001110011100

where 1 means that the antenna is working and 0 that it is defective. Because the resulting system will be functional in the first 3 arrangements and not functional in the remaining 3, it seems reasonable to take 36=12 as the desired probability. In the case of general n and m, we could compute the probability that the system is functional in a similar fashion. That is, we could count the number of configurations that result in the system's being functional and then divide by the total number of all possible configurations.

From the preceding discussion, we see that it would be useful to have an effective method for counting the number of ways that things can occur. In fact, many problems in probability theory can be solved simply by counting the number of different ways that a certain event can occur. The mathematical theory of counting is formally known as *combinatorial analysis*.

1.2 The Basic Principle of Counting

The basic principle of counting will be fundamental to all our work. Loosely put, it states that if one experiment can result in any of m possible outcomes and if another experiment can result in any of n possible outcomes, then there are *mn* possible outcomes of the two experiments.

> The basic principle of counting
>
> Suppose that two experiments are to be performed. Then if experiment 1 can result in any one of m possible outcomes and if, for each outcome of experiment 1, there are n possible outcomes of experiment 2, then together there are *mn* possible outcomes of the two experiments.

Proof of the Basic Principle: The basic principle may be proven by enumerating all the possible outcomes of the two experiments; that is,

$(1,1),(1,2),\ldots,(1,n)(2,1),(2,2),\ldots,(2,n)\vdots(m,1),(m,2),\ldots,(m,n)$

where we say that the outcome is (i, j) if experiment 1 results in its ith possible outcome and experiment 2 then results in its jth possible outcome. Hence, the set of possible outcomes consists of m rows, each containing n elements. This proves the result.

Example 2a

A small community consists of 10 women, each of whom has 3 children. If one woman and one of her children are to be chosen as mother and child of the year, how many different choices are possible?

Solution

By regarding the choice of the woman as the outcome of the first experiment and the subsequent choice of one of her children as the outcome of the second experiment, we see from the basic principle that there are $10\times3=30$ possible choices.

When there are more than two experiments to be performed, the basic principle can be generalized.

> The generalized basic principle of counting
>
> If r experiments that are to be performed are such that the first one may result in any of n1 possible outcomes; and if, for each of these n1 possible outcomes, there are n2 possible outcomes of the second experiment; and if, for each of the possible outcomes of the first two experiments, there are n3 possible outcomes of the third experiment; and if, then there is a total of $n1\cdot n2\cdots nr$ possible outcomes of the r experiments.

Example 2b

A college planning committee consists of 3 freshmen, 4 sophomores, 5 juniors, and 2 seniors. A subcommittee of 4, consisting of 1 person from each class, is to be chosen. How many different subcommittees are possible?

Solution

We may regard the choice of a subcommittee as the combined outcome of the four separate experiments of choosing a single representative from each of the classes. It then follows from the generalized version of the basic principle that there are $3\times4\times5\times2=120$ possible subcommittees.

Example 2c

How many different 7-place license plates are possible if the first 3 places are to be occupied by letters and the final 4 by numbers?

Solution

By the generalized version of the basic principle, the answer is $26\cdot26\cdot26\cdot10\cdot10\cdot10\cdot10=175,760,000$.

Example 2d

How many functions defined on n points are possible if each functional value is either 0 or 1?

Solution

Let the points be 1,2,. . .,n. Since f(i) must be either 0 or 1 for each i=1,2,. . .,n, it follows that there are 2n possible functions.

Example 2e

In **Example 2c**, how many license plates would be possible if repetition among letters or numbers were prohibited?

Solution

In this case, there would be $26\cdot25\cdot24\cdot10\cdot9\cdot8\cdot7=78,624,000$ possible license plates.

1.3 Permutations

How many different ordered arrangements of the letters a, b, and c are possible? By direct enumeration we see that there are 6, namely, *abc, acb, bac, bca, cab*, and *cba*. Each arrangement is known as a *permutation*. Thus, there are 6 possible permutations of a set of 3 objects. This result could also have been obtained from the basic principle, since the first object in the permutation can be any of the 3, the second object in the permutation can then be chosen from any of the remaining 2, and the third object in the permutation is then the remaining 1. Thus, there are $3\cdot2\cdot1=6$ possible permutations.

> Suppose now that we have n objects. Reasoning similar to that we have just used for the 3 letters then shows that there are
>
> $$n(n-1)(n-2)\cdots3\cdot2\cdot1=n!$$
>
> different permutations of the n objects.

Whereas n! (read as “n factorial”) is defined to equal $1\cdot2\cdots n$ when n is a positive integer, it is convenient to define 0! to equal 1.

Example 3a

How many different batting orders are possible for a baseball team consisting of 9 players?

Solution

There are $9!=362{,}880$ possible batting orders.

Example 3b

A class in probability theory consists of 6 men and 4 women. An examination is given, and the students are ranked according to their performance. Assume that no two students obtain the same score.

a. How many different rankings are possible?
b. If the men are ranked just among themselves and the women just among themselves, how many different rankings are possible?

Solution

a. (a) Because each ranking corresponds to a particular ordered arrangement of the 10 people, the answer to this part is $10!=3{,}628{,}800$.
b. (b) Since there are 6! possible rankings of the men among themselves and 4! possible rankings of the women among themselves, it follows from the basic principle that there are $(6!)(4!)=(720)(24)=17{,}280$ possible rankings in this case.

Example 3c

Ms. Jones has 10 books that she is going to put on her bookshelf. Of these, 4 are mathematics books, 3 are chemistry books, 2 are history books, and 1 is a language book. Ms. Jones wants to arrange her books so that all the books dealing with the same subject are together on the shelf. How many different arrangements are possible?

Solution

There are 4! 3! 2! 1! arrangements such that the mathematics books are first in line, then the chemistry books, then the history books, and then the language book. Similarly, for each possible ordering of the subjects, there are 4! 3! 2! 1! possible arrangements. Hence, as there are 4! possible orderings of the subjects, the desired answer is $4!\,4!\,3!\,2!\,1!=6912$.

We shall now determine the number of permutations of a set of n objects when certain of the objects are indistinguishable from one another. To set this situation straight in our minds, consider the following example.

Example 3d

How many different letter arrangements can be formed from the letters PEPPER?

Solution

We first note that there are 6! permutations of the letters $P_1E_1P_2P_3E_2R$ when the 3P's and the 2E's are distinguished from one another. However, consider any one of these permutations for instance, $P_1P_2E_1P_3E_2R$. If we now permute the P’s among themselves and the E’s among themselves, then the resultant arrangement would still be of the form PPEPER. That is, all 3! 2! permutations

$$P_1P_2E_1P_3E_2R\quad P_1P_2E_2P_3E_1R\quad P_1P_3E_1P_2E_2R\quad P_1P_3E_2P_2E_1R\quad P_2P_1E_1P_3E_2R\quad P_2P_1E_2P_3E_1R\quad P_2P_3E_1P_1E_2R\quad P_2P_3E_2P_1E_1R\quad P_3P_1E_1P_2E_2R\quad P_3P_1E_2P_2E_1R\quad P_3P_2E_1P_1E_2R\quad P_3P_2E_2P_1E_1R$$

are of the form PPEPER. Hence, there are $6!/(3!\,2!)=60$ possible letter arrangements of the letters PEPPER.

> In general, the same reasoning as that used in **Example 3d** shows that there are
>
> $$\frac{n!}{n_1!\,n_2!\cdots n_r!}$$
>
> different permutations of n objects, of which n_1 are alike, n_2 are alike, . . ., n_r are alike.

Example 3e

A chess tournament has 10 competitors, of which 4 are Russian, 3 are from the United States, 2 are from Great Britain, and 1 is from Brazil. If the tournament result lists just the nationalities of the players in the order in which they placed, how many outcomes are possible?

Solution

There are

$$\frac{10!}{4!\,3!\,2!\,1!}=12{,}600$$

possible outcomes.

Example 3f

How many different signals, each consisting of 9 flags hung in a line, can be made from a set of 4 white flags, 3 red flags, and 2 blue flags if all flags of the same color are identical?

Solution

There are

$9!4!\,3!\,2!=1260$

different signals.

1.4 Combinations

We are often interested in determining the number of different groups of r objects that could be formed from a total of n objects. For instance, how many different groups of 3 could be selected from the 5 items A, B, C, D, and E? To answer this question, reason as follows: Since there are 5 ways to select the initial item, 4 ways to then select the next item, and 3 ways to select the final item, there are thus $5\cdot4\cdot3$ ways of selecting the group of 3 when the order in which the items are selected is relevant. However, since every group of 3–say, the group consisting of items A, B, and C will be counted 6 times (that is, all of the permutations *ABC, ACB, BAC, BCA, CAB*, and *CBA* will be counted when the order of selection is relevant), it follows that the total number of groups that can be formed is

$$\frac{5\cdot4\cdot3}{3\cdot2\cdot1}=10$$

In general, as $n(n-1)\cdots(n-r+1)$ represents the number of different ways that a group of r items could be selected from n items when the order of selection is relevant, and as each group of r items will be counted $r!$ times in this count, it follows that the number of different groups of r items that could be formed from a set of n items is

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

> Notation and terminology
>
> We define $\binom{n}{r}$, for $r\le n$, by
>
> $$\binom{n}{r}=\frac{n!}{(n-r)!\,r!}$$
>
> and say that $\binom{n}{r}$ (read as "n choose r") represents the number of possible combinations of n objects taken r at a time.

Thus, $\binom{n}{r}$ represents the number of different groups of size r that could be selected from a set of n objects when the order of selection is not considered relevant.

Equivalently, $\binom{n}{r}$ is the number of subsets of size r that can be chosen from a set of size n. Using that $0!=1$, note that $\binom{n}{n}=\binom{n}{0}=\frac{n!}{0!n!}=1$, which is consistent with the preceding interpretation because in a set of size n there is exactly 1 subset of size n (namely, the entire set), and exactly one subset of size 0 (namely the empty set). A useful convention is to define $\binom{n}{r}$ equal to 0 when either $r>n$ or $r<0$.

Example 4a

A committee of 3 is to be formed from a group of 20 people. How many different committees are possible?

Solution

There are $\binom{20}{3}=\frac{20\cdot19\cdot18}{3\cdot2\cdot1}=1140$ possible committees.

Example 4b

From a group of 5 women and 7 men, how many different committees consisting of 2 women and 3 men can be formed? What if 2 of the men are feuding and refuse to serve on the committee together?

Solution

As there are $\binom{5}{2}$ possible groups of 2 women, and $\binom{7}{3}$ possible groups of 3 men, it follows from the basic principle that there are $\binom{5}{2}\binom{7}{3}=\frac{5\cdot4}{2\cdot1}\cdot\frac{7\cdot6\cdot5}{3\cdot2\cdot1}=350$ possible committees consisting of 2 women and 3 men.

Now suppose that 2 of the men refuse to serve together. Because a total of $\binom{2}{2}\binom{5}{1}=5$ out of the $\binom{7}{3}=35$ possible groups of 3 men contain both of the feuding men, it follows that there are $35-5=30$ groups that do not contain both of the feuding men. Because there are still $\binom{5}{2}=10$ ways to choose the 2 women, there are $30\cdot10=300$ possible committees in this case.

Example 4c

Consider a set of n antennas of which m are defective and $n-m$ are functional and assume that all of the defectives and all of the functionals are considered indistinguishable. How many linear orderings are there in which no two defectives are consecutive?

Solution

Imagine that the $n-m$ functional antennas are lined up among themselves. Now, if no two defectives are to be consecutive, then the spaces between the functional antennas must each contain at most one defective antenna. That is, in the $n-m+1$ possible positions–represented in **Figure 1.1** by carets–between the $n-m$ functional antennas, we must select m of these in which to put the defective antennas. Hence, there are $\binom{n-m+1}{m}$ possible orderings in which there is at least one functional antenna between any two defective ones.

Figure 1.1 No consecutive defectives.

$$\wedge\, 1\, \wedge\, 1\, \wedge\, 1 \ldots \wedge\, 1\, \wedge\, 1\, \wedge$$

$$1 = \text{functional}$$

$$\wedge = \text{place for at most one defective}$$

A useful combinatorial identity, known as *Pascal's identity*, is

$$\binom{n}{r}=\binom{n-1}{r-1}+\binom{n-1}{r} \quad 1\le r\le n \tag{4.1}$$

Equation (4.1) may be proved analytically or by the following combinatorial argument: Consider a group of n objects, and fix attention on some particular one of these objects–call it object 1. Now, there are $\binom{n-1}{r-1}$ groups of size r that contain object 1 (since each such group is formed by selecting $r-1$ from the remaining $n-1$ objects). Also, there are $\binom{n-1}{r}$ groups of size r that do not contain object 1. As there is a total of $\binom{n}{r}$ groups of size r, **Equation (4.1)** follows.

The values $\binom{n}{r}$ are often referred to as *binomial coefficients* because of their prominence in the binomial theorem.

> The binomial theorem
>
> $$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k} x^k y^{n-k} \tag{4.2}$$

We shall present two proofs of the binomial theorem. The first is a proof by mathematical induction, and the second is a proof based on combinatorial considerations.

Proof of the Binomial Theorem by Induction: When $n=1$, **Equation (4.2)** reduces to

$$x+y=\binom{1}{0} x^0y^1+\binom{1}{1} x^1y^0=y+x$$

Assume **Equation (4.2)** for $n-1$. Now,

$$(x+y)^n=(x+y)(x+y)^{n-1}=(x+y)\sum_{k=0}^{n-1}\binom{n-1}{k} x^ky^{n-1-k}=\sum_{k=0}^{n-1}\binom{n-1}{k}x^{k+1}y^{n-1-k}+\sum_{k=0}^{n-1}\binom{n-1}{k} x^ky^{n-k}$$

Letting $i=k+1$ in the first sum and $i=k$ in the second sum, we find that

$$\begin{aligned}(x+y)^n&=\sum_{i=1}^{n}\binom{n-1}{i-1} x^iy^{n-i}+\sum_{i=0}^{n-1}\binom{n-1}{i} x^iy^{n-i}=\sum_{i=1}^{n-1}\binom{n-1}{i-1} x^iy^{n-i}+x^n+y^n+\sum_{i=1}^{n-1}\binom{n-1}{i} x^iy^{n-i}=x^n+\sum_{i=1}^{n-1}\left[\binom{n-1}{i-1}+\binom{n-1}{i}\right] x^iy^{n-i}+y^n\\&=x^n+\sum_{i=1}^{n-1}\binom{n}{i} x^iy^{n-i}+y^n=\sum_{i=0}^{n}\binom{n}{i} x^iy^{n-i}\end{aligned}$$

where the next-to-last equality follows by **Equation (4.1)**. By induction, the theorem is now proved.

Combinatorial Proof of the Binomial Theorem: Consider the product

$$(x_1+y_1)(x_2+y_2)\cdots(x_n+y_n)$$

Its expansion consists of the sum of 2^n terms, each term being the product of n factors. Furthermore, each of the 2^n terms in the sum will contain as a factor either x_i or y_i for each $i=1, 2,\ldots,n$. For example,

$$(x_1+y_1)(x_2+y_2)=x_1x_2+x_1y_2+y_1x_2+y_1y_2$$

Now, how many of the 2^n terms in the sum will have k of the x_i's and $(n-k)$ of the y_i's as factors? As each term consisting of k of the x_i's and $(n-k)$ of the y_i's corresponds to a choice of a group of k from the n values $x_1, x_2\ldots, x_n$, there are $\binom{n}{k}$ such terms. Thus, letting $x_i=x$, $y_i=y$, $i=1,\ldots,n$, we see that

$$(x+y)^n=\sum_{k=0}^{n}\binom{n}{k} x^ky^{n-k}$$

Example 4d

Expand $(x+y)^3$.

Solution

$$(x+y)^3=\binom{3}{0} x^0y^3+\binom{3}{1} x^1y^2+\binom{3}{2} x^2y^1+\binom{3}{3} x^3y^0=y^3+3xy^2+3x^2y+x^3$$

Example 4e

How many subsets are there of a set consisting of n elements?

Solution

Since there are $\binom{n}{k}$ subsets of size k, the desired answer is

$$\sum_{k=0}^{n}\binom{n}{k}=(1+1)^n=2^n$$

This result could also have been obtained by assigning either the number 0 or the number 1 to each element in the set. To each assignment of numbers, there corresponds, in a one-to-one fashion, a subset, namely, that subset consisting of all elements that were assigned the value 1. As there are 2^n possible assignments, the result follows.

Note that we have included the set consisting of 0 elements (that is, the null set) as a subset of the original set. Hence, the number of subsets that contain at least 1 element is 2^n-1.

1.5 Multinomial Coefficients

In this section, we consider the following problem: A set of n distinct items is to be divided into r distinct groups of respective sizes $n_1, n_2, \ldots, n_r$, where $\sum_{i=1}^{r} n_i = n$. How many different divisions are possible? To answer this question, we note that there are $\binom{n}{n_1}$ possible choices for the first group; for each choice of the first group, there are $\binom{n-n_1}{n_2}$ possible choices for the second group; for each choice of the first two groups, there are $\binom{n-n_1-n_2}{n_3}$ possible choices for the third group; and so on. It then follows from the generalized version of the basic counting principle that there are

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\cdots\binom{n-n_1-n_2-\cdots-n_{r-1}}{n_r} = \frac{n!}{(n-n_1)!\,n_1!}\frac{(n-n_1)!}{(n-n_1-n_2)!\,n_2!}\cdots\frac{(n-n_1-n_2-\cdots-n_{r-1})!}{0!\,n_r!} = \frac{n!}{n_1!\,n_2!\cdots n_r!}$$

possible divisions.

Another way to see this result is to consider the n values $1, 1, \ldots, 1, 2, \ldots, 2, \ldots, r, \ldots, r$, where i appears n_i times, for $i = 1, \ldots, r$. Every permutation of these values corresponds to a division of the n items into the r groups in the following manner: Let the permutation $i_1, i_2, \ldots, i_n$ correspond to assigning item 1 to group i_1, item 2 to group i_2, and so on. For instance, if $n=8$ and if $n_1=4, n_2=3$, and $n_3=1$, then the permutation 1, 1, 2, 3, 2, 1, 2, 1 corresponds to assigning items 1, 2, 6, 8 to the first group, items 3, 5, 7 to the second group, and item 4 to the third group. Because every permutation yields a division of the items and every possible division results from some permutation, it follows that the number of divisions of n items into r distinct groups of sizes $n_1, n_2, \ldots, n_r$ is the same as the number of permutations of n items of which n_1 are alike, and n_2 are alike, . . ., and n_r are alike, which was shown in **Section 1.3** to equal $\frac{n!}{n_1!\,n_2!\cdots n_r!}$.

> Notation
>
> If $n_1 + n_2 + \cdots + n_r = n$, we define $\binom{n}{n_1, n_2, \ldots, n_r}$ by
>
> $$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\,n_2!\cdots n_r!}$$
>
> Thus, $\binom{n}{n_1, n_2, \ldots, n_r}$ represents the number of possible divisions of n distinct objects into r distinct groups of respective sizes $n_1, n_2, \ldots, n_r$.

Example 5a

A police department in a small city consists of 10 officers. If the department policy is to have 5 of the officers patrolling the streets, 2 of the officers working full time at the station, and 3 of the officers on reserve at the station, how many different divisions of the 10 officers into the 3 groups are possible?

Solution

There are $\frac{10!}{5!\,2!\,3!} = 2520$ possible divisions.

Example 5b

Ten children are to be divided into an A team and a B team of 5 each. The A team will play in one league and the B team in another. How many different divisions are possible?

Solution

There are $\frac{10!}{5!\,5!} = 252$ possible divisions.

Example 5c

In order to play a game of basketball, 10 children at a playground divide themselves into two teams of 5 each. How many different divisions are possible?

Solution

Note that this example is different from **Example 5b** because now the order of the two teams is irrelevant. That is, there is no A or B team, but just a division consisting of 2 groups of 5 each. Hence, the desired answer is

$$\frac{10!/(5!\,5!)}{2!} = 126$$

The proof of the following theorem, which generalizes the binomial theorem, is left as an exercise.

> The multinomial theorem
>
> $$(x_1 + x_2 + \cdots + x_r)^n = \sum_{\substack{(n_1, \ldots, n_r):\\ n_1 + \cdots + n_r = n}} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$
>
> That is, the sum is over all nonnegative integer-valued vectors $(n_1, n_2, \ldots, n_r)$ such that $n_1 + n_2 + \cdots + n_r = n$.

The numbers $\binom{n}{n_1, n_2, \ldots, n_r}$ are known as *multinomial coefficients*.

Example 5d

In the first round of a knockout tournament involving $n = 2^m$ players, the n players are divided into $n/2$ pairs, with each of these pairs then playing a game. The losers of the games are eliminated while the winners go on to the next round, where the process is repeated until only a single player remains. Suppose we have a knockout tournament of 8 players.

a. How many possible outcomes are there for the initial round? (For instance, one outcome is that 1 beats 2, 3 beats 4, 5 beats 6, and 7 beats 8.)
b. How many outcomes of the tournament are possible, where an outcome gives complete information for all rounds?

Solution

One way to determine the number of possible outcomes for the initial round is to first determine the number of possible pairings for that round. To do so, note that the number of ways to divide the 8 players into a *first* pair, a *second* pair, a *third* pair, and a *fourth* pair is $\binom{8}{2, 2, 2, 2} = \frac{8!}{2^4}$. Thus, the number of possible pairings when there is no ordering of the 4 pairs is $\frac{8!}{2^4 4!}$. For each such pairing, there are 2 possible choices from each pair as to the winner of that game, showing that there are $\frac{8! 2^4}{2^4 4!} = \frac{8!}{4!}$ possible results of round 1. [Another way to see this is to note that there are $\binom{8}{4}$ possible choices of the 4 winners and, for each such choice, there are 4! ways to pair the 4 winners with the 4 losers, showing that there are $4!\binom{8}{4} = \frac{8!}{4!}$ possible results for the first round.]

Similarly, for each result of round 1, there are $\frac{4!}{2!}$ possible outcomes of round 2, and for each of the outcomes of the first two rounds, there are $\frac{2!}{1!}$ possible outcomes of round 3. Consequently, by the generalized basic principle of counting, there are $\frac{8!}{4!}\frac{4!}{2!}\frac{2!}{1!} = 8!$ possible outcomes of the tournament. Indeed, the same argument can be used to show that a knockout tournament of $n = 2^m$ players has $n!$ possible outcomes.

Knowing the preceding result, it is not difficult to come up with a more direct argument by showing that there is a one-to-one correspondence between the set of possible tournament results and the set of permutations of $1, \ldots, n$. To obtain such a correspondence, rank the players as follows for any tournament result: Give the tournament winner rank 1, and give the final-round loser rank 2. For the two players who lost in the next-to-last round, give rank 3 to the one who lost to the player ranked 1 and give rank 4 to the one who lost to the player ranked 2. For the four players who lost in the second-to-last round, give rank 5 to the one who lost to player ranked 1, rank 6 to the one who lost to the player ranked 2, rank 7 to the one who lost to the player ranked 3, and rank 8 to the one who lost to the player ranked 4. Continuing on in this manner gives a rank to each player. (A more succinct description is to give the winner of the tournament rank 1 and let the rank of a player who lost in a round having 2^k matches be 2^k plus the rank of the player who beat him, for $k = 0, \ldots, m-1$.) In this manner, the result of the tournament can be represented by a permutation $i_1, i_2, \ldots, i_n$, where i_j is the player who was given rank j. Because different tournament results give rise to different permutations, and because there is a tournament result for each permutation, it follows that there are the same number of possible tournament results as there are permutations of $1, \ldots, n$.

Example 5e

$$(x_1+x_2+x_3)^2 = \binom{2}{2, 0, 0} x_1^2x_2^0x_3^0 + \binom{2}{0, 2, 0} x_1^0x_2^2x_3^0 + \binom{2}{0, 0, 2} x_1^0x_2^0x_3^2 + \binom{2}{1, 1, 0} x_1^1x_2^1x_3^0 + \binom{2}{1, 0, 1} x_1^1x_2^0x_3^1 + \binom{2}{0, 1, 1} x_1^0x_2^1x_3^1$$
$$= x_1^2+x_2^2+x_3^2+2x_1x_2+2x_1x_3+2x_2x_3$$

* 1.6 The Number of Integer Solutions of Equations

* Asterisks denote material that is optional.

An individual has gone fishing at Lake Ticonderoga, which contains four types of fish: lake trout, catfish, bass, and bluefish. If we take the result of the fishing trip to be the numbers of each type of fish caught, let us determine the number of possible outcomes when a total of 10 fish are caught. To do so, note that we can denote the outcome of the fishing trip by the vector (x1,x2, x3, x4) where x1 is the number of trout that are caught, x2 is the number of catfish, x3 is the number of bass, and x4 is the number of bluefish. Thus, the number of possible outcomes when a total of 10 fish are caught is the number of nonnegative integer vectors (x1,x2, x3, x4) that sum to 10.

More generally, if we supposed there were r types of fish and that a total of n were caught, then the number of possible outcomes would be the number of nonnegative integer-valued vectors x1,. . .,xr such that

x1+x2+. . .+xr=n (6.1)

To compute this number, let us start by considering the number of positive integer-valued vectors x1,. . .,xr that satisfy the preceding. To determine this number, suppose that we have n consecutive zeroes lined up in a row:

0 0 0 . . . 0 0

Note that any selection of r−1 of the n−1 spaces between adjacent zeroes (see **Figure 1.2**) corresponds to a positive solution of **6.1** by letting x1 be the number of zeroes before the first chosen space, x2 be the number of zeroes between the first and second chosen space, . . ., and xn being the number of zeroes following the last chosen space.

Figure 1.2 Number of positive solutions.

$$0 \wedge 0 \wedge 0 \wedge \ldots \wedge 0 \wedge 0$$

n objects 0

Choose $r - 1$ of the spaces $\wedge$.

For instance, if we have n=8 and r=3, then (with the choices represented by dots) the choice

0.0 0 0 0.0 0 0

corresponds to the solution x1=1, x2=4, x3=3. As positive solutions of (6.1) correspond, in a one-to-one fashion, to choices of r−1 of the adjacent spaces, it follows that the number of differerent positive solutions is equal to the number of different selections of r−1 of the n−1 adjacent spaces. Consequently, we have the following proposition.

Proposition 6.1

There are $\binom{n-1}{r-1}$ distinct positive integer-valued vectors $(x_1, x_2, \ldots, x_r)$ satisfying the equation

$$x_1+x_2+\cdots+x_r=n, \quad x_i>0, \quad i=1,\ldots,r$$

To obtain the number of nonnegative (as opposed to positive) solutions, note that the number of nonnegative solutions of $x_1+x_2+\cdots+x_r=n$ is the same as the number of positive solutions of $y_1+\cdots+y_r=n+r$ (seen by letting $y_i=x_i+1$, $i=1,\ldots,r$). Hence, from **Proposition 6.1**, we obtain the following proposition.

Proposition 6.2

There are $\binom{n+r-1}{r-1}$ distinct nonnegative integer-valued vectors $(x_1,x_2,\ldots,x_r)$ satisfying the equation

$$x_1+x_2+\cdots+x_r=n$$

Thus, using **Proposition 6.2**, we see that there are $\binom{13}{3}=286$ possible outcomes when a total of 10 Lake Ticonderoga fish are caught.

Example 6a

How many distinct nonnegative integer-valued solutions of $x_1+x_2=3$ are possible?

Solution

There are $\binom{3+2-1}{2-1}=4$ such solutions: (0, 3), (1, 2), (2, 1), (3, 0).

Example 6b

An investor has $20,000 to invest among 4 possible investments. Each investment must be in units of $1000. If the total $20,000 is to be invested, how many different investment strategies are possible? What if not all the money needs to be invested?

Solution

If we let x_i, $i=1, 2, 3, 4$, denote the number of thousands invested in investment i, then, when all is to be invested, x_1, x_2, x_3, x_4 are integers satisfying the equation

$$x_1+x_2+x_3+x_4=20 \quad x_i\geq 0$$

Hence, by **Proposition 6.2**, there are $\binom{23}{3}=1771$ possible investment strategies. If not all of the money needs to be invested, then if we let x_5 denote the amount kept in reserve, a strategy is a nonnegative integer-valued vector $(x_1, x_2, x_3, x_4, x_5)$ satisfying the equation

$$x_1+x_2+x_3+x_4+x_5=20$$

Hence, by **Proposition 6.2**, there are now $\binom{24}{4}=10{,}626$ possible strategies.

Example 6c

How many terms are there in the multinomial expansion of $(x_1+x_2+\cdots+x_r)^n$?

Solution

$$(x_1+x_2+\cdots+x_r)^n=\sum\binom{n}{n_1,\ldots,n_r} x_1^{n_1}\cdots x_r^{n_r}$$

where the sum is over all nonnegative integer-valued $(n_1,\ldots,n_r)$ such that $n_1+\cdots+n_r=n$. Hence, by

Proposition 6.2, there are $\binom{n+r-1}{r-1}$ such terms.

Example 6d

Let us consider again **Example 4c**, in which we have a set of n items, of which m are (indistinguishable and) defective and the remaining $n-m$ are (also indistinguishable and) functional. Our objective is to determine the number of linear orderings in which no two defectives are next to each other. To determine this number, let us imagine that the defective items are lined up among themselves and the functional ones are now to be put in position. Let us denote x_1 as the number of functional items to the left of the first defective, x_2 as the number of functional items between the first two defectives, and so on. That is, schematically, we have

$$x_1\ 0\ x_2\ 0 \cdots x_m\ 0\ x_{m+1}$$

Now, there will be at least one functional item between any pair of defectives as long as $x_i>0$, $i=2,\ldots,m$. Hence, the number of outcomes satisfying the condition is the number of vectors $x_1,\ldots,x_{m+1}$ that satisfy the equation

$$x_1+\cdots+x_{m+1}=n-m, \quad x_1\geq 0, x_{m+1}\geq 0, x_i>0, i=2,\ldots, m$$

But, on letting $y_1=x_1+1$, $y_i=x_i$, $i=2,\ldots,m$, $y_{m+1}=x_{m+1}+1$, we see that this number is equal to the number of positive vectors $(y_1,\ldots,y_{m+1})$ that satisfy the equation

$$y_1+y_2+\cdots+y_{m+1}=n-m+2$$

Hence, by **Proposition 6.1**, there are $\binom{n-m+1}{m}$ such outcomes, in agreement with the results of **Example 4c**.

Suppose now that we are interested in the number of outcomes in which each pair of defective items is separated by at least 2 functional items. By the same reasoning as that applied previously, this would equal the number of vectors satisfying the equation

$$x_1+\cdots+x_{m+1}=n-m, \quad x_1\geq 0, x_{m+1}\geq 0, x_i\geq 2, i=2,\ldots, m$$

Upon letting $y_1=x_1+1$, $y_i=x_i-1$, $i=2,\ldots,m$, $y_{m+1}=x_{m+1}+1$, we see that this is the same as the number of positive solutions of the equation

$$y_1+\cdots+y_{m+1}=n-2m+3$$

Hence, from **Proposition 6.1**, there are $\binom{n-2m+2}{m}$ such outcomes.

Summary

The basic principle of counting states that if an experiment consisting of two phases is such that there are n possible outcomes of phase 1 and, for each of these n outcomes, there are m possible outcomes of phase 2, then there are *nm* possible outcomes of the experiment.

There are $n! = n(n-1)\cdots 3\cdot 2\cdot 1$ possible linear orderings of n items. The quantity $0!$ is defined to equal 1.

Let

$$\binom{n}{i} = \frac{n!}{(n-i)!\,i!}$$

when $0 \le i \le n$, and let it equal 0 otherwise. This quantity represents the number of different subgroups of size i that can be chosen from a set of size n. It is often called a *binomial coefficient* because of its prominence in the binomial theorem, which states that

$$(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$$

For nonnegative integers $n_1, \ldots, n_r$ summing to n,

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!\,n_2!\cdots n_r!}$$

is the number of divisions of n items into r distinct nonoverlapping subgroups of sizes $n_1, n_2 \ldots, n_r$. These quantities are called *multinomial coefficients*.

Problems

1.
   a. How many different 7-place license plates are possible if the first 2 places are for letters and the other 5 for numbers?
   b. Repeat part (a) under the assumption that no letter or number can be repeated in a single license plate.

2. How many outcome sequences are possible when a die is rolled four times, where we say, for instance, that the outcome is 3, 4, 3, 1 if the first roll landed on 3, the second on 4, the third on 3, and the fourth on 1?

3. Twenty workers are to be assigned to 20 different jobs, one to each job. How many different assignments are possible?

4. John, Jim, Jay, and Jack have formed a band consisting of 4 instruments. If each of the boys can play all 4 instruments, how many different arrangements are possible? What if John and Jim can play all 4 instruments, but Jay and Jack can each play only piano and drums?

5. For years, telephone area codes in the United States and Canada consisted of a sequence of three digits. The first digit was an integer between 2 and 9, the second digit was either 0 or 1, and the third digit was any integer from 1 to 9. How many area codes were possible? How many area codes starting with a 4 were possible?

6. A well-known nursery rhyme starts as follows:

   “As I was going to St. Ives
   I met a man with 7 wives.
   Each wife had 7 sacks.
   Each sack had 7 cats.
   Each cat had 7 kittens. . .”
   How many kittens did the traveler meet?

7.
   a. In how many ways can 3 boys and 3 girls sit in a row?
   b. In how many ways can 3 boys and 3 girls sit in a row if the boys and the girls are each to sit together?
   c. In how many ways if only the boys must sit together?
   d. In how many ways if no two people of the same sex are allowed to sit together?

8. When all letters are used, how many different letter arrangements can be made from the letters
   a. Fluke?
   b. Propose?
   c. Mississippi?
   d. Arrange?

9. A child has 12 blocks, of which 6 are black, 4 are red, 1 is white, and 1 is blue. If the child puts the blocks in a line, how many arrangements are possible?

10. In how many ways can 8 people be seated in a row if
    a. there are no restrictions on the seating arrangement?
    b. persons A and B must sit next to each other?
    c. there are 4 men and 4 women and no 2 men or 2 women can sit next to each other?
    d. there are 5 men and they must sit next to one another?
    e. there are 4 married couples and each couple must sit together?

11. In how many ways can 3 novels, 2 mathematics books, and 1 chemistry book be arranged on a bookshelf if

a. the books can be arranged in any order?
b. the mathematics books must be together and the novels must be together?
c. the novels must be together, but the other books can be arranged in any order?

12. How many 3 digit numbers xyz, with x, y, z all ranging from 0 to 9 have at least 2 of their digits equal. How many have exactly 2 equal digits.

13. How many different letter permutations, of any length, can be made using the letters M O T T O. (For instance, there are 3 possible permutations of length 1.)

14. Five separate awards (best scholarship, best leadership qualities, and so on) are to be presented to selected students from a class of 30. How many different outcomes are possible if

a. a student can receive any number of awards?
b. each student can receive at most 1 award?

15. Consider a group of 20 people. If everyone shakes hands with everyone else, how many handshakes take place?

16. How many 5-card poker hands are there?

17. A dance class consists of 22 students, of which 10 are women and 12 are men. If 5 men and 5 women are to be chosen and then paired off, how many results are possible?

18. A student has to sell 2 books from a collection of 6 math, 7 science, and 4 economics books. How many choices are possible if

a. both books are to be on the same subject?
b. the books are to be on different subjects?

19. Seven different gifts are to be distributed among 10 children. How many distinct results are possible if no child is to receive more than one gift?

20. A committee of 7, consisting of 2 Republicans, 2 Democrats, and 3 Independents, is to be chosen from a group of 5 Republicans, 6 Democrats, and 4 Independents. How many committees are possible?

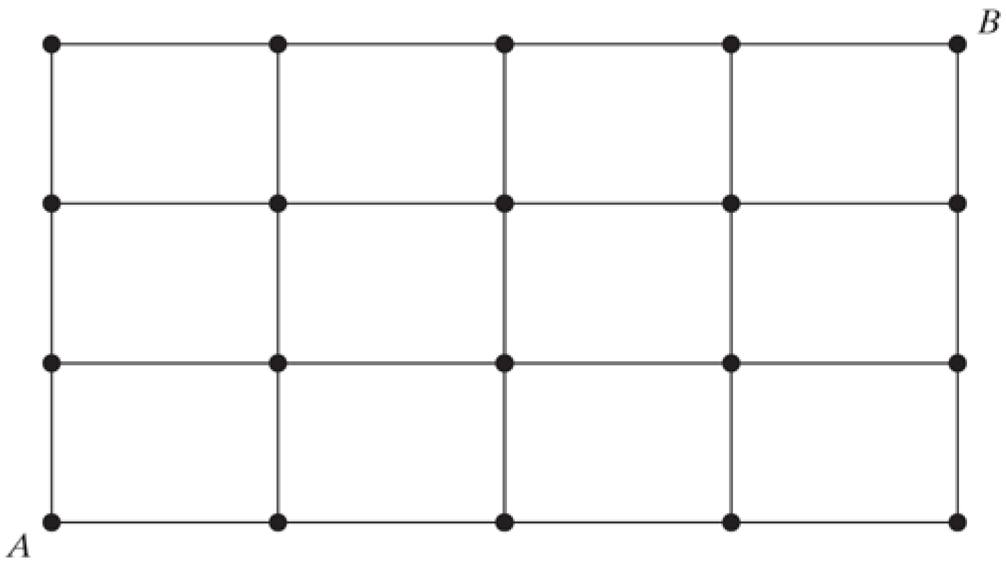
21. From a group of 8 women and 6 men, a committee consisting of 3 men and 3 women is to be formed. How many different committees are possible if

a. 2 of the men refuse to serve together?
b. 2 of the women refuse to serve together?
c. 1 man and 1 woman refuse to serve together?

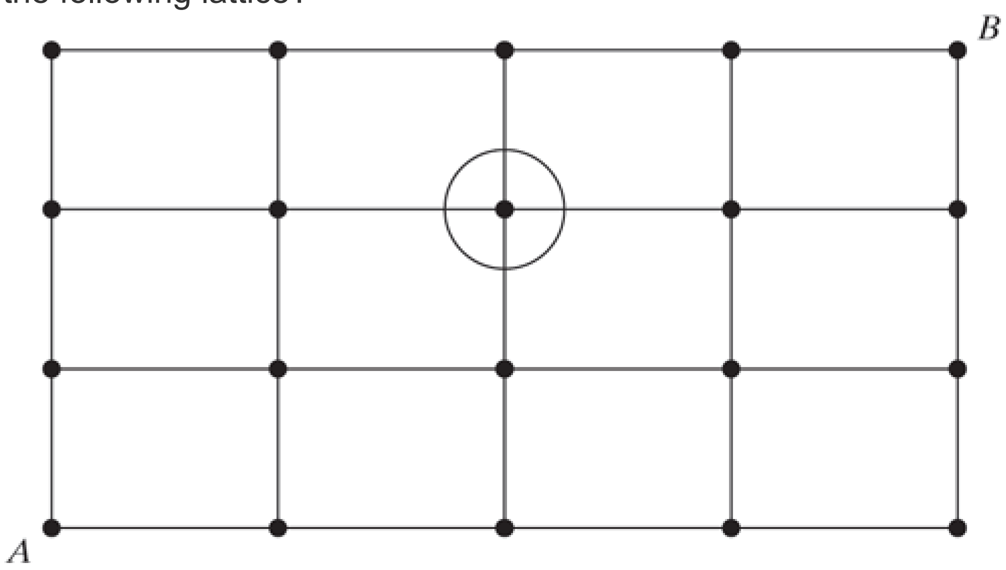
22. A person has 8 friends, of whom 5 will be invited to a party.

a. How many choices are there if 2 of the friends are feuding and will not attend together?
b. How many choices if 2 of the friends will only attend together?

23. Consider the grid of points shown at the top of the next column. Suppose that, starting at the point labeled A, you can go one step up or one step to the right at each move. This procedure is continued until the point labeled B is reached. How many different paths from A to B are possible? **Hint:** Note that to reach B from A, you must take 4 steps to the right and 3 steps upward.



24. In **Problem 23**, how many different paths are there from A to B that go through the point circled in the following lattice?



25. A psychology laboratory conducting dream research contains 3 rooms, with 2 beds in each room. If 3 sets of identical twins are to be assigned to these 6 beds so that each set of twins sleeps in different beds in the same room, how many assignments are possible?

26.

a. Show $\sum_{k=0}^{n}\binom{n}{k}2^k=3^n$

b. Simplify $\sum_{k=0}^{n}\binom{n}{k}x^k$

27. Expand $(3x^2+y)^5$.

28. The game of bridge is played by 4 players, each of whom is dealt 13 cards. How many bridge deals are possible?

29. Expand $(x_1+2x_2+3x_3)^4$.

30. If 12 people are to be divided into 3 committees of respective sizes 3, 4, and 5, how many divisions are possible?

31. If 8 new teachers are to be divided among 4 schools, how many divisions are possible? What if each school must receive 2 teachers?

32. Ten weight lifters are competing in a team weight-lifting contest. Of the lifters, 3 are from the United States, 4 are from Russia, 2 are from China, and 1 is from Canada. If the scoring takes account of the countries that the lifters represent, but not their individual identities, how many different outcomes are possible from the point of view of scores? How many different outcomes correspond to results in which the United States has 1 competitor in the top three and 2 in the bottom three?

33. Delegates from 10 countries, including Russia, France, England, and the United States, are to be seated in a row. How many different seating arrangements are possible if the French and English delegates are to be seated next to each other and the Russian and U.S. delegates are not to be next to each other?
* 34. If 8 identical blackboards are to be divided among 4 schools, how many divisions are possible? How many if each school must receive at least 1 blackboard?
* 35. An elevator starts at the basement with 8 people (not including the elevator operator) and discharges them all by the time it reaches the top floor, number 6. In how many ways could the operator have perceived the people leaving the elevator if all people look alike to him? What if the 8 people consisted of 5 men and 3 women and the operator could tell a man from a woman?
* 36. We have $20,000 that must be invested among 4 possible opportunities. Each investment must be integral in units of $1000, and there are minimal investments that need to be made if one is to invest in these opportunities. The minimal investments are $2000, $2000, $3000, and $4000. How many different investment strategies are available if

a. an investment must be made in each opportunity?
b. investments must be made in at least 3 of the 4 opportunities?

* 37. Suppose that 10 fish are caught at a lake that contains 5 distinct types of fish.

a. How many different outcomes are possible, where an outcome specifies the numbers of caught fish of each of the 5 types?
b. How many outcomes are possible when 3 of the 10 fish caught are trout?
c. How many when at least 2 of the 10 are trout?

Theoretical Exercises

1. Prove the generalized version of the basic counting principle.
2. Two experiments are to be performed. The first can result in any one of m possible outcomes. If the first experiment results in outcome i, then the second experiment can result in any of ni possible outcomes, i=1, 2,. . .,m. What is the number of possible outcomes of the two experiments?
3. In how many ways can r objects be selected from a set of n objects if the order of selection is considered relevant?
4. There are (nr) different linear arrangements of n balls of which r are black and n−r are white. Give a combinatorial explanation of this fact.
5. Determine the number of vectors (x1,. . .,xn), such that each xi is either 0 or 1 and
∑i=1nxi≥ k

6. How many vectors x1,. . .,xk are there for which each xi is a positive integer such that 1≤xi≤n and x1<x2<⋯<xk?
7. Give an analytic proof of **Equation (4.1)**.
8. Prove that
(n+mr)=(n0)(mr)+(n1)(mr−1)+⋯+(nr)(m0)

Hint: Consider a group of n men and m women. How many groups of size r are possible?
9. Use **Theoretical Exercise 8** to prove that
(2nn)=∑k=0n(nk)2

10. From a group of n people, suppose that we want to choose a committee of k, k≤n, one of whom is to be designated as chairperson.
   a. By focusing first on the choice of the committee and then on the choice of the chair, argue that there are (nk) k possible choices.
   b. By focusing first on the choice of the nonchair committee members and then on the choice of the chair, argue that there are (nk−1) (n−k+1) possible choices.
   c. By focusing first on the choice of the chair and then on the choice of the other committee members, argue that there are n (n−1k−1) possible choices.
   d. Conclude from parts (a), (b), and (c) that
   k (nk)=(n−k+1)(nk−1)=n (n−1k−1)

   e. Use the factorial definition of (mr) to verify the identity in part (d).

11. The following identity is known as Fermat’s combinatorial identity:
(nk)=∑i=kn(i−1k−1)   n≥ k

Give a combinatorial argument (no computations are needed) to establish this identity.
Hint: Consider the set of numbers 1 through n. How many subsets of size k have i as their highest numbered member?
12. Consider the following combinatorial identity:
∑k=1nk (nk)=n⋅2n−1

   a. Present a combinatorial argument for this identity by considering a set of n people and determining, in two ways, the number of possible selections of a committee of any size and a chairperson for the committee.
   Hint:
      i. How many possible selections are there of a committee of size k and its chairperson?
      ii. How many possible selections are there of a chairperson and the other committee members?

b. Verify the following identity for $n=1, 2, 3, 4, 5$:

$$\sum_{k=1}^{n} \binom{n}{k} k^2 = 2^{n-2} n(n+1)$$

For a combinatorial proof of the preceding, consider a set of n people and argue that both sides of the identity represent the number of different selections of a committee, its chairperson, and its secretary (possibly the same as the chairperson).

Hint:

i. How many different selections result in the committee containing exactly k people?
ii. How many different selections are there in which the chairperson and the secretary are the same? (ANSWER: $n2^{n-1}$.)
iii. How many different selections result in the chairperson and the secretary being different?

c. Now argue that

$$\sum_{k=1}^{n} \binom{n}{k} k^3 = 2^{n-3} n^2(n+3)$$

13. Show that, for $n > 0$,

$$\sum_{i=0}^{n} (-1)^i \binom{n}{i} = 0$$

Hint: Use the binomial theorem.

14. From a set of n people, a committee of size j is to be chosen, and from this committee, a subcommittee of size $i, i \le j$, is also to be chosen.

a. Derive a combinatorial identity by computing, in two ways, the number of possible choices of the committee and subcommittee–first by supposing that the committee is chosen first and then the subcommittee is chosen, and second by supposing that the subcommittee is chosen first and then the remaining members of the committee are chosen.
b. Use part (a) to prove the following combinatorial identity:

$$\sum_{j=i}^{n} \binom{n}{j}\binom{j}{i} = \binom{n}{i} 2^{n-i} \quad i \le n$$

c. Use part (a) and **Theoretical Exercise 13** to show that

$$\sum_{j=i}^{n} \binom{n}{j}\binom{j}{i} (-1)^{n-j} = 0 \quad i < n$$

15. Let $H_k(n)$ be the number of vectors $x_1, \ldots, x_k$ for which each x_i is a positive integer satisfying $1 \le x_i \le n$ and $x_1 \le x_2 \le \cdots \le x_k$.

a. Without any computations, argue that

$$H_1(n) = n \qquad H_k(n) = \sum_{j=1}^{n} H_{k-1}(j) \quad k > 1$$

Hint: How many vectors are there in which $x_k = j$?

b. Use the preceding recursion to compute $H_3(5)$.
Hint: First compute $H_2(n)$ for $n=1, 2, 3, 4, 5$.

16. Consider a tournament of n contestants in which the outcome is an ordering of these contestants, with ties allowed. That is, the outcome partitions the players into groups, with the first group consisting of the players who tied for first place, the next group being those who tied for the next-best position, and so on. Let $N(n)$ denote the number of different possible outcomes. For instance, $N(2)=3$, since, in a tournament with 2 contestants, player 1 could be uniquely first, player 2 could be uniquely first, or they could tie for first

a. List all the possible outcomes when $n=3$.
b. With $N(0)$ defined to equal 1, argue, without any computations, that

$$N(n) = \sum_{i=1}^{n} \binom{n}{i} N(n-i)$$

Hint: How many outcomes are there in which i players tie for last place?

c. Show that the formula of part (b) is equivalent to the following:
$N(n)=\sum_{i=0}^{n-1}\binom{n}{i} N(i)$

d. Use the recursion to find $N(3)$ and $N(4)$.

17. Present a combinatorial explanation of why $\binom{n}{r}=\binom{n}{r,\ n-r}$
18. Argue that
$\binom{n}{n_1,n_2,\ldots,n_r}=\binom{n-1}{n_1-1,n_2,\ldots,n_r}+\binom{n-1}{n_1,n_2-1,\ldots,n_r}+\cdots+\binom{n-1}{n_1,n_2,\ldots,n_r-1}$

Hint: Use an argument similar to the one used to establish **Equation (4.1)**.
19. Prove the multinomial theorem.
* 20. In how many ways can n identical balls be distributed into r urns so that the ith urn contains at least m_i balls, for each $i=1,\ldots,r$? Assume that $n\geq\sum_{i=1}^{r}m_i$.
* 21. Argue that there are exactly $\binom{r}{k}\binom{n-1}{n-r+k}$ solutions of
$x_1+x_2+\cdots+x_r=n$

for which exactly k of the x_i are equal to 0.
* 22. Consider a function $f(x_1,\ldots,x_n)$ of n variables. How many different partial derivatives of order r does f possess?
* 23. Determine the nu mber of vectors $(x_1,\ldots,x_n)$ such that each x_i is a nonnegative integer and $\sum_{i=1}^{n}x_i\leq k$

Self-Test Problems and Exercises

1. How many different linear arrangements are there of the letters A, B, C, D, E, F for which
    a. A and B are next to each other?
    b. A is before B?
    c. A is before B and B is before C?
    d. A is before B and C is before D?
    e. A and B are next to each other and C and D are also next to each other?
    f. E is not last in line?

2. If 4 Americans, 3 French people, and 3 British people are to be seated in a row, how many seating arrangements are possible when people of the same nationality must sit next to each other?

3. A president, treasurer, and secretary, all different, are to be chosen from a club consisting of 10 people. How many different choices of officers are possible if
    a. there are no restrictions?
    b. A and B will not serve together?
    c. C and D will serve together or not at all?
    d. E must be an officer?
    e. F will serve only if he is president?

4. A student is to answer 7 out of 10 questions in an examination. How many choices has she? How many if she must answer at least 3 of the first 5 questions?

5. In how many ways can a man divide 7 gifts among his 3 children if the eldest is to receive 3 gifts and the others 2 each?

6. How many different 7-place license plates are possible when 3 of the entries are letters and 4 are digits? Assume that repetition of letters and numbers is allowed and that there is no restriction on where the letters or numbers can be placed.

7. Give a combinatorial explanation of the identity
(nr)=(nn−r)

8. Consider n-digit numbers where each digit is one of the 10 integers 0,1,. . .,9. How many such numbers are there for which
    a. no two consecutive digits are equal?
    b. 0 appears as a digit a total of i times, i=0,. . .,n?

9. Consider three classes, each consisting of n students. From this group of 3n students, a group of 3 students is to be chosen.
    a. How many choices are possible?
    b. How many choices are there in which all 3 students are in the same class?
    c. How many choices are there in which 2 of the 3 students are in the same class and the other student is in a different class?
    d. How many choices are there in which all 3 students are in different classes?
    e. Using the results of parts (a) through (d), write a combinatorial identity.

10. How many 5-digit numbers can be formed from the integers 1,2,. . .,9 if no digit can appear more than twice? (For instance, 41434 is not allowed.)

11. From 10 married couples, we want to select a group of 6 people that is not allowed to contain a married couple.
    a. How many choices are there?
    b. How many choices are there if the group must also consist of 3 men and 3 women?

12. A committee of 6 people is to be chosen from a group consisting of 7 men and 8 women. If the

committee must consist of at least 3 women and at least 2 men, how many different committees are possible?
* 13. An art collection on auction consisted of 4 Dalis, 5 van Goghs, and 6 Picassos. At the auction were 5 art collectors. If a reporter noted only the number of Dalis, van Goghs, and Picassos acquired by each collector, how many different results could have been recorded if all of the works were sold?
* 14. Determine the number of vectors $(x_1,\ldots,x_n)$ such that each x_i is a positive integer and

$$\sum_{i=1}^{n} x_i \le k$$

where $k \ge n$.
15. A total of n students are enrolled in a review course for the actuarial examination in probability. The posted results of the examination will list the names of those who passed, in decreasing order of their scores. For instance, the posted result will be “Brown, Cho” if Brown and Cho are the only ones to pass, with Brown receiving the higher score. Assuming that all scores are distinct (no ties), how many posted results are possible?
16. How many subsets of size 4 of the set $S=\{1, 2,\ldots,20\}$ contain at least one of the elements 1, 2, 3, 4, 5?
17. Give an analytic verification of

$$\binom{n}{2}=\binom{k}{2}+k(n-k)+\binom{n-k}{2}, \quad 1\le k\le n$$

Now, give a combinatorial argument for this identity.
18. In a certain community, there are 3 families consisting of a single parent and 1 child, 3 families consisting of a single parent and 2 children, 5 families consisting of 2 parents and a single child, 7 families consisting of 2 parents and 2 children, and 6 families consisting of 2 parents and 3 children. If a parent and child from the same family are to be chosen, how many possible choices are there?
19. If there are no restrictions on where the digits and letters are placed, how many 8-place license plates consisting of 5 letters and 3 digits are possible if no repetitions of letters or digits are allowed? What if the 3 digits must be consecutive?
20. Verify the identity

$$\sum_{x_1+\ldots+x_r=n,\ x_i\ge 0} \frac{n!}{x_1!x_2!\cdots x_r!}=r^n$$

a. by a combinatorial argument that first notes that r^n is the number of different n letter sequences that can be formed from an alphabet consisting of r letters, and then determines how many of these letter sequences have letter 1 a total of x_1 times and letter 2 a total of x_2 times and ... and letter r a total of x_r times;
b. by using the multinomial theorem.

21. Simplify $n-\binom{n}{2}+\binom{n}{3}-\ldots+(-1)^{n+1}\binom{n}{n}$

Chapter 2 Axioms of Probability

Contents

2.1 Introduction

In this chapter, we introduce the concept of the probability of an event and then show how probabilities can be computed in certain situations. As a preliminary, however, we need to discuss the concept of the sample space and the events of an experiment.

2.2 Sample Space and Events

Consider an experiment whose outcome is not predictable with certainty. However, although the outcome of the experiment will not be known in advance, let us suppose that the set of all possible outcomes is known. This set of all possible outcomes of an experiment is known as the *sample space* of the experiment and is denoted by *S*. Following are some examples:

1. If the outcome of an experiment consists of the determination of the sex of a newborn child, then
   $S=\{g, b\}$

   where the outcome g means that the child is a girl and b that it is a boy.
2. If the outcome of an experiment is the order of finish in a race among the 7 horses having post positions 1, 2, 3, 4, 5, 6, and 7, then
   $S=\{\text{all } 7! \text{ permutations of } (1, 2, 3, 4, 5, 6, 7)\}$

   The outcome (2, 3, 1, 6, 5, 4, 7) means, for instance, that the number 2 horse comes in first, then the number 3 horse, then the number 1 horse, and so on.
3. If the experiment consists of flipping two coins, then the sample space consists of the following four points:
   $S=\{(h, h), (h, t), (t, h), (t, t)\}$

   The outcome will be (h, h) if both coins are heads, (h, t) if the first coin is heads and the second tails, (t, h) if the first is tails and the second heads, and (t, t) if both coins are tails.
4. If the experiment consists of tossing two dice, then the sample space consists of the 36 points
   $S=\{(i, j): i, j=1, 2, 3, 4, 5, 6\}$

   where the outcome (i,j) is said to occur if i appears on the leftmost die and j on the other die.
5. If the experiment consists of measuring (in hours) the lifetime of a transistor, then the sample space consists of all nonnegative real numbers; that is,
   $S=\{x: 0\le x<\infty\}$

Any subset E of the sample space is known as an *event*. In other words, an event is a set consisting of possible outcomes of the experiment. If the outcome of the experiment is contained in E, then we say that E has occurred. Following are some examples of events.

In the preceding Example 1, if E={g}, then E is the event that the child is a girl. Similarly, if F={b}, then F is the event that the child is a boy.

In Example 2, if

$E=\{\text{all outcomes in S starting with a 3}\}$

then E is the event that horse 3 wins the race.

In Example 3, if E={(h, h), (h, t)}, then E is the event that a head appears on the first coin.

In Example 4, if E={(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)}, then E is the event that the sum of the dice equals 7.

In Example 5, if $E=\{x: 0\le x\le 5\}$, then E is the event that the transistor does not last longer than 5 hours.

For any two events E and F of a sample space S, we define the new event $E\cup F$ to consist of all outcomes that are either in E or in F or in both E and F. That is, the event $E\cup F$ will occur if *either E* or F occurs. For instance, in Example 1, if $E=\{g\}$ is the event that the child is a girl and $F=\{b\}$ is the event that the child is a boy, then

$$E\cup F=\{g, b\}$$

is the whole sample space S. In Example 3, if $E=\{(h, h), (h, t)\}$ is the event that the first coin lands heads, and $F=\{(t, h), (h, h)\}$ is the event that the second coin lands heads, then

$$E\cup F=\{(h, h), (h, t), (t, h)\}$$

is the event that at least one of the coins lands heads and thus will occur provided that both coins do not land tails.

The event $E\cup F$ is called the *union* of the event E and the event F.

Similarly, for any two events E and F, we may also define the new event *EF*, called the *intersection* of E and F, to consist of all outcomes that are both in E and in F. That is, the event *EF* (sometimes written $E\cap F$) will occur only if both E and F occur. For instance, in Example 3, if $E=\{(h, h), (h, t), (t, h)\}$ is the event that at least 1 head occurs and $F=\{(h, t), (t, h), (t, t)\}$ is the event that at least 1 tail occurs, then

$$EF=\{(h, t), (t, h)\}$$

is the event that exactly 1 head and 1 tail occur. In Example 4, if $E=\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$ is the event that the sum of the dice is 7 and $F=\{(1, 5), (2, 4), (3, 3), (4, 2), (5, 1)\}$ is the event that the sum is 6, then the event *EF* does not contain any outcomes and hence could not occur. To give such an event a name, we shall refer to it as the null event and denote it by $\emptyset$. (That is, $\emptyset$ refers to the event consisting of no outcomes.) If $EF=\emptyset$, then E and F are said to be *mutually exclusive*.

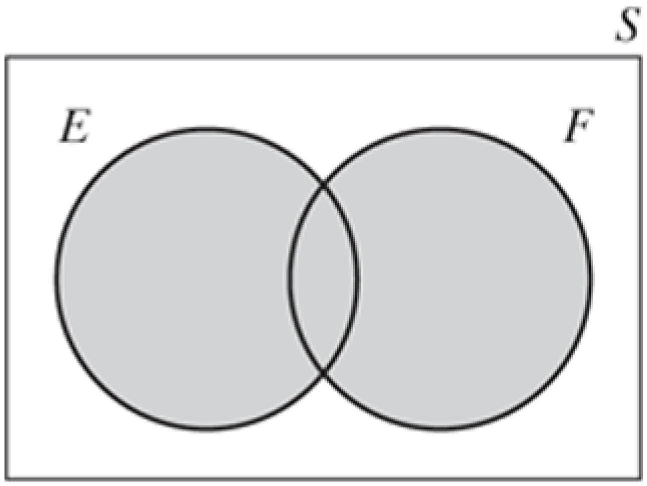
We define unions and intersections of more than two events in a similar manner. If $E_1, E_2, \ldots$ are events, then the union of these events, denoted by $\cup_{n=1}^{\infty} E_n$, is defined to be that event that consists of all outcomes that are in E_n for at least one value of $n=1, 2, \ldots$. Similarly, the intersection of the events E_n, denoted by $\cap_{n=1}^{\infty} E_n$, is defined to be the event consisting of those outcomes that are in all of the events E_n, $n=1, 2, \ldots$.

Finally, for any event E, we define the new event E^c, referred to as the *complement* of E, to consist of all outcomes in the sample space S that are not in E. That is, E^c will occur if and only if E does not occur. In Example 4, if event $E=\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, then E^c will occur when the sum of the dice does not equal 7. Note that because the experiment must result in some outcome, it follows that $S^c=\emptyset$.

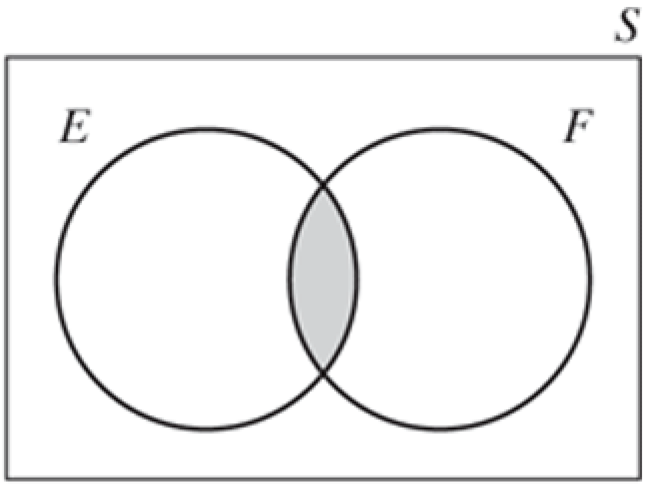
For any two events E and F, if all of the outcomes in E are also in F, then we say that E is *contained* in F, or E is a *subset* of F, and write $E\subset F$ (or equivalently, $F\supset E$, which we sometimes say as F is a *superset* of E). Thus, if $E\subset F$, then the occurrence of E implies the occurrence of F. If $E\subset F$ and $F\subset E$, we say that E and F are equal and write $E=F$.

A graphical representation that is useful for illustrating logical relations among events is the Venn diagram. The sample space S is represented as consisting of all the outcomes in a large rectangle, and the events $E, F, G, \ldots$ are represented as consisting of all the outcomes in given circles within the rectangle. Events of interest can then be indicated by shading appropriate regions of the diagram. For instance, in the three Venn diagrams shown in **Figure 2.1**, the shaded areas represent, respectively, the events $E\cup F$, *EF*, and E^c. The Venn diagram in **Figure 2.2** indicates that $E\subset F$.

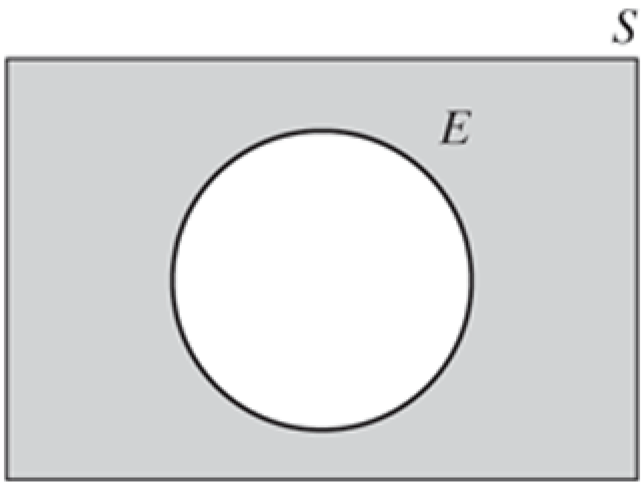
Figure 2.1 Venn diagrams.



(a) Shaded region: E∪F.

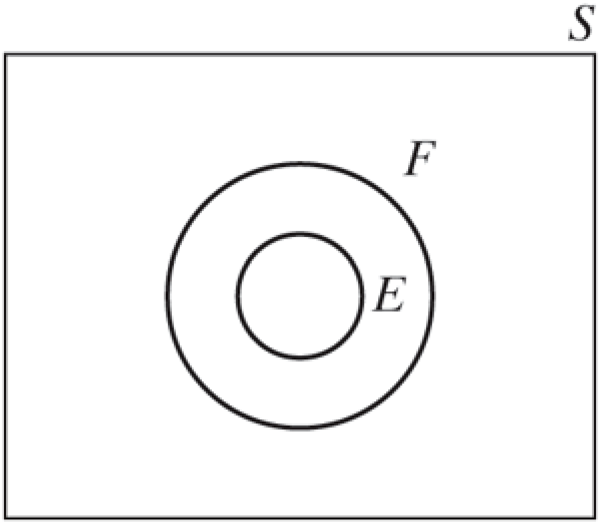


(b) Shaded region: *FF*



(b) Shaded region: E^c

Figure 2.2 E⊂F.

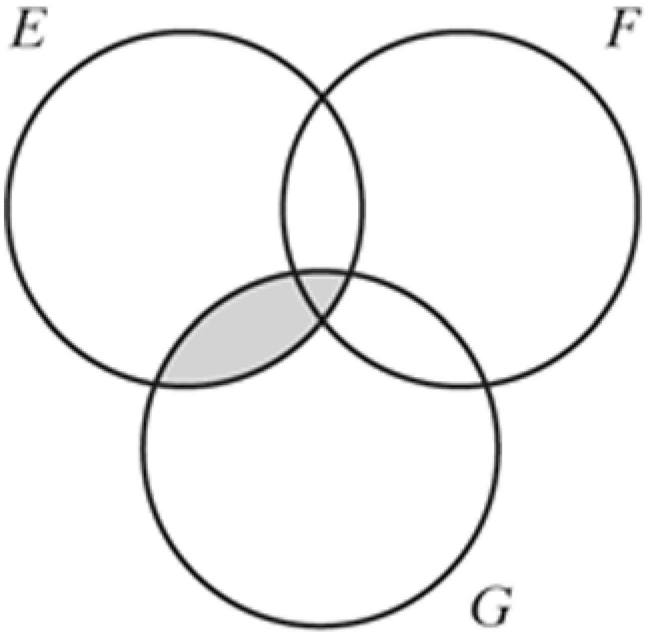


The operations of forming unions, intersections, and complements of events obey certain rules similar to the rules of algebra. We list a few of these rules:

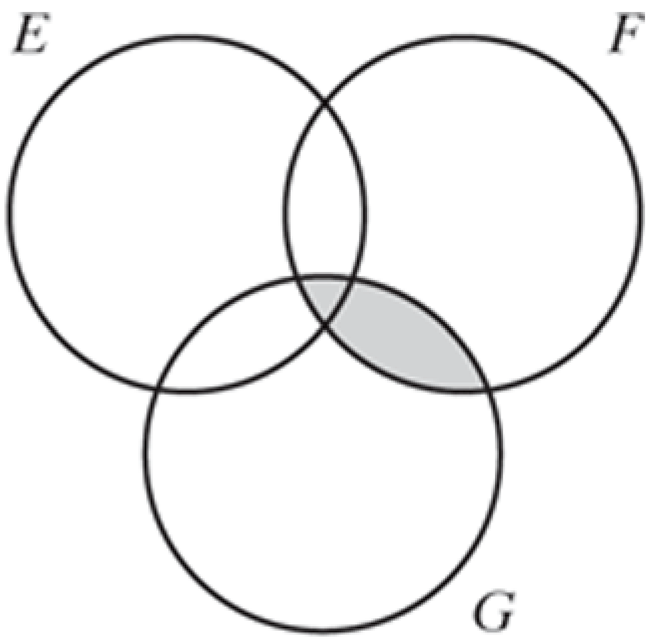
Commutative laws $E \cup F = F \cup E$ $EF = FE$ Associative laws $(E \cup F) \cup G = E \cup (F \cup G)$ $(EF)G = E(FG)$ Distributive laws $(E \cup F)G = EG \cup FG$ $EF \cup G = (E \cup G)(F \cup G)$

These relations are verified by showing that any outcome that is contained in the event on the left side of the equality sign is also contained in the event on the right side, and vice versa. One way of showing this is by means of Venn diagrams. For instance, the distributive law may be verified by the sequence of diagrams in **Figure 2.3**.

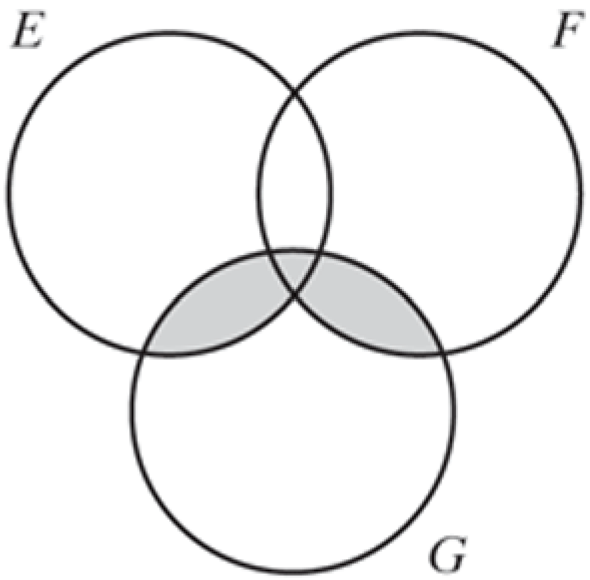
Figure 2.3 $(E \cup F)G = EG \cup FG$.



(a) Shaded region: *EG*.



(b) Shaded region: *FG*.



(c) Shaded region: $(E \cup F)G$.

The following useful relationships among the three basic operations of forming unions, intersections, and complements are known as *DeMorgan's laws*:

$$\left(\bigcup_{i=1}^{n} E_i\right)^c = \bigcap_{i=1}^{n} E_i^c \qquad \left(\bigcap_{i=1}^{n} E_i\right)^c = \bigcup_{i=1}^{n} E_i^c$$

For instance, for two events *E* and *F*, DeMorgan's laws state that

$$(E \cup F)^c = E^c F^c \quad \text{and} \quad (EF)^c = E^c \cup F^c$$

which can be easily proven by using Venn diagrams (see **Theoretical Exercise 7**).

To prove DeMorgan's laws for general n, suppose first that x is an outcome of $\left(\bigcup_{i=1}^{n} E_i\right)^c$. Then x is not contained in $\bigcup_{i=1}^{n} E_i$, which means that *x* is not contained in any of the events E_i, $i=1, 2, \ldots, n$, implying that x is contained in E_i^c for all $i=1, 2, \ldots,$ and thus is contained in $\bigcap_{i=1}^{n} E_i^c$. To go the other way, suppose that *x* is an outcome of $\bigcap_{i=1}^{n} E_i^c$. Then x is contained in E_i^c for all $i=1, 2, \ldots, n$, which means that x is not contained in E_i for any $i=1, 2, \ldots, n$, implying that *x* is not contained in $\bigcup_{i}^{n} E_i$, in turn implying that x is contained in $(\cup_1^n E_i)^c$. This proves the first of DeMorgan’s laws.

To prove the second of DeMorgan’s laws, we use the first law to obtain

$$\left(\cup_{i=1}^{n} E_i^c\right)^c = \cap_{i=1}^{n} (E_i^c)^c$$

which, since $(E^c)^c=E$, is equivalent to

$$(\cup_1^n E_i^c)^c = \cap_1^n E_i$$

Taking complements of both sides of the preceding equation yields the result we seek, namely,

$$\cup_1^n E_i^c = \left(\cap_1^n E_i\right)^c$$

2.3 Axioms of Probability

One way of defining the probability of an event is in terms of its long run *relative frequency*. Such a definition usually goes as follows: We suppose that an experiment, whose sample space is S, is repeatedly performed under exactly the same conditions. For each event E of the sample space S, we define n(E) to be the number of times in the first n repetitions of the experiment that the event E occurs. Then P(E), the *probability* of the event *E*, is defined as

$$P(E)=\lim_{n\to\infty}\frac{n(E)}{n}$$

That is, P(E) is defined as the (limiting) proportion of time that E occurs. It is thus the limiting relative frequency of E.

Although the preceding definition is certainly intuitively pleasing and should always be kept in mind by the reader, it possesses a serious drawback: How do we know that n(E)/n will converge to some constant limiting value that will be the same for each possible sequence of repetitions of the experiment? For example, suppose that the experiment to be repeatedly performed consists of flipping a coin. How do we know that the proportion of heads obtained in the first n flips will converge to some value as n gets large? Also, even if it does converge to some value, how do we know that, if the experiment is repeatedly performed a second time, we shall obtain the same limiting proportion of heads?

Proponents of the relative frequency definition of probability usually answer this objection by stating that the convergence of n(E)/n to a constant limiting value is an assumption, or an *axiom*, of the system. However, to assume that n(E)/n will necessarily converge to some constant value seems to be an extraordinarily complicated assumption. For, although we might indeed hope that such a constant limiting frequency exists, it does not at all seem to be a priori evident that this need be the case. In fact, would it not be more reasonable to assume a set of simpler and more self-evident axioms about probability and then attempt to prove that such a constant limiting frequency does in some sense exist? The latter approach is the modern axiomatic approach to probability theory that we shall adopt in this text. In particular, we shall assume that, for each event E in the sample space S, there exists a value P(E), referred to as the probability of E. We shall then assume that all these probabilities satisfy a certain set of axioms, which, we hope the reader will agree, is in accordance with our intuitive notion of probability.

Consider an experiment whose sample space is S. For each event E of the sample space S, we assume that a number P(E) is defined and satisfies the following three axioms:

> The three axioms of probability
>
> **Axiom 1**
>
> $$0\le P(E)\le 1$$
>
> **Axiom 2**
>
> $$P(S)=1$$
>
> **Axiom 3**
>
> For any sequence of mutually exclusive events $E_1, E_2, \ldots$ (that is, events for which $E_i E_j=\emptyset$ when $i\ne j$),
>
> $$P\left(\bigcup_{i=1}^{\infty} E_i\right)=\sum_{i=1}^{\infty} P(E_i)$$
>
> We refer to P(E) as the *probability* of the event E.

Thus, **Axiom 1** states that the probability that the outcome of the experiment is an outcome in E is some number between 0 and 1. **Axiom 2** states that, with probability 1, the outcome will be a point in the sample space S. **Axiom 3** states that, for any sequence of mutually exclusive events, the probability of at least one of these events occurring is just the sum of their respective probabilities.

If we consider a sequence of events $E_1, E_2, \ldots,$, where $E_1 = S$ and $E_1 = \varnothing$ for $i>1$, then, because the events are mutually exclusive and because $S=\cup_{i=1}^{\infty}E_i$, we have, from **Axiom 3**,

$$P(S)=\sum_{i=1}^{\infty}P(E_i)=P(S)+\sum_{i=2}^{\infty}P(\varnothing)$$

implying that

$$P(\varnothing)=0$$

That is, the null event has probability 0 of occurring.

Note that it follows that, for any finite sequence of mutually exclusive events $E_1, E_2, \ldots, E_n$,

$$P\left(\bigcup_1^n E_i\right)=\sum_{i=1}^{n}P(E_i) \qquad (3.1)$$

This equation follows from **Axiom 3** by defining E_i as the null event for all values of i greater than n. **Axiom 3** is equivalent to **Equation (3.1)** when the sample space is finite. (Why?) However, the added generality of **Axiom 3** is necessary when the sample space consists of an infinite number of points.

Example 3a

If our experiment consists of tossing a coin and if we assume that a head is as likely to appear as a tail, then we would have

$$P(\{H\})=P(\{T\})=\frac{1}{2}$$

On the other hand, if the coin were biased and we believed that a head were twice as likely to appear as a tail, then we would have

$$P(\{H\})=\frac{2}{3} \qquad P(\{T\})=\frac{1}{3}$$

Example 3b

If a die is rolled and we suppose that all six sides are equally likely to appear, then we would have $P(\{1\})=P(\{2\})=P(\{3\})=P(\{4\})=P(\{5\})=P(\{6\})=\frac{1}{6}$. From **Axiom 3**, it would thus follow that the probability of rolling an even number would equal

$$P(\{2,4,6\})=P(\{2\})+P(\{4\})+P(\{6\})=\frac{1}{2}$$

The assumption of the existence of a set function P, defined on the events of a sample space S and satisfying **Axioms 1**, **2**, and **3**, constitutes the modern mathematical approach to probability theory. It is hoped that the reader will agree that the axioms are natural and in accordance with our intuitive concept of probability as related to chance and randomness. Furthermore, using these axioms, we shall be able to prove that if an experiment is repeated over and over again, then, with probability 1, the proportion of time during which any specific event *E* occurs will equal P(E). This result, known as the strong law of large numbers, is presented in **Chapter 8**. In addition, we present another possible interpretation of probability–as being a measure of belief–in **Section 2.7**.

Technical Remark. We have supposed that P(E) is defined for all the events E of the sample space. Actually, when the sample space is an uncountably infinite set, P(E) is defined only for a class of events called *measurable*. However, this restriction need not concern us, as all events of any practical interest are measurable.

2.4 Some Simple Propositions

In this section, we prove some simple propositions regarding probabilities. We first note that since E and Ec are always mutually exclusive and since E∪Ec=S, we have, by **Axioms 2** and **3**,

$$1=P(S)=P(E\cup E^c)=P(E)+P(E^c)$$

Or, equivalently, we have **Proposition 4.1**.

$$P(E^c)=1-P(E)$$

Proposition 4.1

In words, **Proposition 4.1** states that the probability that an event does not occur is 1 minus the probability that it does occur. For instance, if the probability of obtaining a head on the toss of a coin is 38, then the probability of obtaining a tail must be 5 8 .

Our second proposition states that if the event E is contained in the event F, then the probability of E is no greater than the probability of F.

Proposition 4.2

If $E\subset F$, then $P(E)\le P(F)$.

Proof. Since $E\subset F$, it follows that we can express F as

$$F=E\cup E^cF$$

Hence, because E and E^cF are mutually exclusive, we obtain, from **Axiom 3**,

$$P(F)=P(E)+P(E^cF)$$

which proves the result, since $P(E^cF)\ge 0$.

Proposition 4.2 tells us, for instance, that the probability of rolling a 1 with a die is less than or equal to the probability of rolling an odd value with the die.

The next proposition gives the relationship between the probability of the union of two events, expressed in terms of the individual probabilities, and the probability of the intersection of the events.

$$P(E\cup F)=P(E)+P(F)-P(EF)$$

Proposition 4.3

Proof To derive a formula for $P(E\cup F)$, we first note that $E\cup F$ can be written as the union of the two disjoint events E and E^cF. Thus, from **Axiom 3**, we obtain

$$P(E\cup F)=P(E\cup E^cF)=P(E)+P(E^cF)$$

Furthermore, since $F=EF\cup E^cF$, we again obtain from **Axiom 3**

$$P(F)=P(EF)+P(E^cF)$$

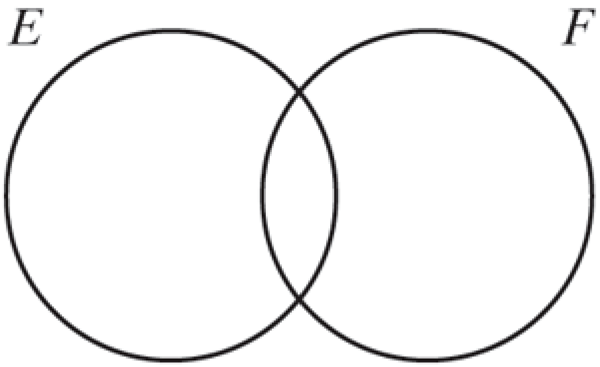
or, equivalently,

$P(EcF)=P(F)-P(EF)$

thereby completing the proof.

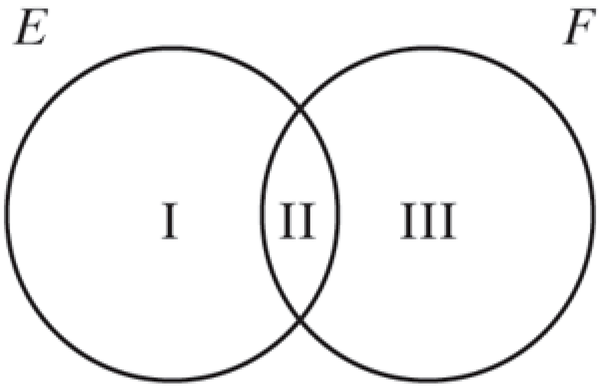
Proposition 4.3 could also have been proved by making use of the Venn diagram in **Figure 2.4**.

Figure 2.4: Venn diagram.



Let us divide $E \cup F$ into three mutually exclusive sections, as shown in **Figure 2.5**. In words, section I represents all the points in E that are not in F (that is, EFc), section II represents all points both in E and in F (that is, *EF*), and section III represents all points in F that are not in E (that is, EcF).

Figure 2.5 Venn diagram in sections.



From **Figure 2.5**, we see that

$E \cup F = \mathrm{I} \cup \mathrm{II} \cup \mathrm{III}$ $E = \mathrm{I} \cup \mathrm{II}$ $F = \mathrm{II} \cup \mathrm{III}$

As I, II, and III are mutually exclusive, it follows from **Axiom 3** that

$P(E \cup F) = P(\mathrm{I})+P(\mathrm{II})+P(\mathrm{III})$ $P(E) = P(\mathrm{I})+P(\mathrm{II})$ $P(F) = P(\mathrm{II})+P(\mathrm{III})$

which shows that

$P(E \cup F)=P(E)+P(F)-P(\mathrm{II})$

and **Proposition 4.3** is proved, since II = *EF*.

Example 4a

J is taking two books along on her holiday vacation. With probability .5, she will like the first book; with probability .4, she will like the second book; and with probability .3, she will like both books. What is the probability that she likes neither book?

Solution

Let B_i denote the event that J likes book i, i=1, 2. Then the probability that she likes at least one of the books is

$$P(B_1 \cup B_2)=P(B_1)+P(B_2)-P(B_1B_2)=.5+.4-.3=.6$$

Because the event that J likes neither book is the complement of the event that she likes at least one of them, we obtain the result

$$P(B_1^cB_2^c)=P((B_1 \cup B_2)^c)=1-P(B_1 \cup B_2)=.4$$

We may also calculate the probability that any one of the three events E, F, and G occurs, namely,

$$P(E \cup F \cup G)=P[(E \cup F) \cup G]$$

which, by **Proposition 4.3**, equals

$$P(E \cup F)+P(G)-P[(E \cup F)G]$$

Now, it follows from the distributive law that the events $(E \cup F)G$ and $EG \cup FG$ are equivalent; hence, from the preceding equations, we obtain

$$\begin{aligned} P(E \cup F \cup G) &= P(E)+P(F)-P(EF)+P(G)-P(EG \cup FG) \\ &= P(E)+P(F)-P(EF)+P(G)-P(EG)-P(FG)+P(EGFG) \\ &= P(E)+P(F)+P(G)+P(EF)-P(EG)-P(FG)+P(EFG) \end{aligned}$$

In fact, the following proposition, known as the *inclusion–exclusion identity*, can be proved by mathematical induction:

Proposition 4.4

$$P(E_1 \cup E_2 \cup \cdots \cup E_n)=\sum_{i=1}^{n} P(E_i)-\sum_{i_1<i_2} P(E_{i_1}E_{i_2})+\cdots+(-1)^{r+1}\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})+\cdots+(-1)^{n+1}P(E_1E_2\cdots E_n)$$

The summation $\sum_{i_1<i_2<\cdots<i_r} P(E_{i_1}E_{i_2}\cdots E_{i_r})$ is taken over all of the $\binom{n}{r}$ possible subsets of size r of the set $\{1, 2 \ldots, n\}$.

In words, **Proposition 4.4** states that the probability of the union of n events equals the sum of the probabilities of these events taken one at a time, minus the sum of the probabilities of these events taken two at a time, plus the sum of the probabilities of these events taken three at a time, and so on.

Remarks 1. For a noninductive argument for **Proposition 4.4**, note first that if an outcome of the sample space is not a member of any of the sets E_i, then its probability does not contribute anything to either side of the equality. Now, suppose that an outcome is in exactly m of the events E_i, where $m>0$. Then, since it is in $\cup_i E_i$, its probability is counted once in $P(\cup_i E_i)$; also, as this outcome is contained in $\binom{m}{k}$ subsets of the type $E_{i_1}E_{i_2}\cdots E_{i_k}$, its probability is counted

$$\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

times on the right of the equality sign in **Proposition 4.4**. Thus, for $m>0$, we must show that

$$1=\binom{m}{1}-\binom{m}{2}+\binom{m}{3}-\cdots\pm\binom{m}{m}$$

However, since $1=\binom{m}{0}$, the preceding equation is equivalent to

$$\sum_{i=0}^{m} \binom{m}{i} (-1)^i = 0$$

and the latter equation follows from the binomial theorem, since

$$0=(-1+1)^m=\sum_{i=0}^{m}\binom{m}{i}(-1)^i(1)^{m-i}$$

2. The following is a succinct way of writing the inclusion–exclusion identity:
   $$P(\cup_{i=1}^{n}E_i)=\sum_{r=1}^{n}(-1)^{r+1}\sum_{i_1<\cdots<i_r}P(E_{i_1}\cdots E_{i_r})$$

3. In the inclusion–exclusion identity, going out one term results in an upper bound on the probability of the union, going out two terms results in a lower bound on the probability, going out three terms results in an upper bound on the probability, going out four terms results in a lower bound, and so on. That is, for events $E_1,\ldots,E_n$, we have
   $$P(\cup_{i=1}^{n} E_i) \le \sum_{i=1}^{n} P(E_i) \tag{4.1}$$

   $$P(\cup_{i=1}^{n}E_i)\ge\sum_{i=1}^{n}P(E_i)-\sum_{j<i}P(E_iE_j) \tag{4.2}$$

   $$P(\cup_{i=1}^{n}E_i)\le\sum_{i=1}^{n}P(E_i)-\sum_{j<i}P(E_iE_j)+\sum_{k<j<i}P(E_iE_jE_k) \tag{4.3}$$

and so on. To prove the validity of these bounds, note the identity

$$\cup_{i=1}^{n}E_i=E_1\cup E_1^cE_2\cup E_1^cE_2^cE_3\cup\cdots\cup E_1^c\cdots E_{n-1}^cE_n$$

That is, at least one of the events E_i occurs if E_1 occurs, or if E_1 does not occur but E_2 does, or if E_1 and E_2 do not occur but E_3 does, and so on. Because the right-hand side is the union of disjoint events, we obtain

$$P(\cup_{i=1}^{n} E_i) = P(E_1)+P(E_1^cE_2)+P(E_1^cE_2^cE_3)+\ldots+P(E_1^c\cdots E_{n-1}^cE_n) = P(E_i)+\sum_{i=2}^{n} P(E_1^c\cdots E_{i-1}^cE_i) \tag{4.4}$$

Now, let $B_i = E_1^c\cdots E_{i-1}^c = (\cup_{j<i} E_j)^c$ be the event that none of the first $i-1$ events occurs. Applying the identity

$$P(E_i)=P(B_iE_i)+P(B_i^cE_i)$$

shows that

$$P(E_i)=P(E_1^c\ldots E_{i-1}^cE_i)=P(E_i)+P(\cup_{j<i} E_iE_j)$$

or, equivalently,

$$P(E_1^c\cdots E_{i-1}^cE_i)=P(E_i)-P(\cup_{j<i}E_iE_j)$$

Substituting this equation into **(4.4)** yields

$$P(\cup_{i=1}^{n}E_i)=\sum_iP(E_i)-\sum_iP(\cup_{j<i}E_iE_j) \tag{4.5}$$

Because probabilities are always nonnegative, Inequality (4.1) follows directly from **Equation (4.5)**. Now, fixing i and applying Inequality (1) to $P(\cup_{j<i}E_iE_j)$ yields

$$P(\cup_{j<i}E_iE_j)\le\sum_{j<i}P(E_iE_j)$$

which, by **Equation (4.5)**, gives Inequality (4.2). Similarly, fixing *i* and applying Inequality (4.2) to $P(\cup_{j<i} E_i E_j)$ yields

$$P\left(\cup_{j<i} E_i E_j\right) \geq \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_i E_k) = \sum_{j<i} P(E_i E_j) - \sum_{k<j<i} P(E_i E_j E_k)$$

which, by **Equation (4.5)**, gives Inequality (4.3). The next inclusion–exclusion inequality is now obtained by fixing i and applying Inequality (4.3) to $P(\cup_{j<i} E_i E_j)$, and so on.

The first inclusion-exclusion inequality, namely that

$$P\left(\cup_{i=1}^{n} E_i\right) \leq \sum_{i=1}^{n} P(E_i)$$

is known as *Boole's inequality*.

2.5 Sample Spaces Having Equally Likely Outcomes

In many experiments, it is natural to assume that all outcomes in the sample space are equally likely to occur. That is, consider an experiment whose sample space S is a finite set, say, $S=\{1, 2, \ldots, N\}$. Then, it is often natural to assume that

$$P(\{1\})=P(\{2\})=\cdots=P(\{N\})$$

which implies, from **Axioms 2** and **3** (why?), that

$$P(\{i\})=\frac{1}{N} \quad i=1, 2, \ldots, N$$

From this equation, it follows from **Axiom 3** that, for any event E,

$$P(E)=\frac{\text{number of outcomes in } E}{\text{number of outcomes in } S}$$

In words, if we assume that all outcomes of an experiment are equally likely to occur, then the probability of any event E equals the proportion of outcomes in the sample space that are contained in E.

Example 5a

If two dice are rolled, what is the probability that the sum of the upturned faces will equal 7?

Solution

We shall solve this problem under the assumption that all of the 36 possible outcomes are equally likely. Since there are 6 possible outcomes–namely, (1, 6), (2, 5), (3, 4), (4, 3), (5, 2), and (6, 1)–that result in the sum of the dice being equal to 7, the desired probability is $\frac{6}{36}=\frac{1}{6}$.

Example 5b

If 3 balls are “randomly drawn” from a bowl containing 6 white and 5 black balls, what is the probability that one of the balls is white and the other two black?

Solution

If we regard the balls as being distinguishable and the order in which they are selected as being relevant, then the sample space consists of $11\cdot10\cdot9=990$ outcomes. Furthermore, there are $6\cdot5\cdot4=120$ outcomes in which the first ball selected is white and the other two are black; $5\cdot6\cdot4=120$ outcomes in which the first is black, the second is white, and the third is black; and $5\cdot4\cdot6=120$ in which the first two are black and the third is white. Hence, assuming that “randomly drawn” means that each outcome in the sample space is equally likely to occur, we see that the desired probability is

$$\frac{120+120+120}{990}=\frac{4}{11}$$

This problem could also have been solved by regarding the outcome of the experiment as the unordered set of drawn balls. From this point of view, there are $\binom{11}{3}=165$ outcomes in the sample space. Now, each set of 3 balls corresponds to $3!$ outcomes when the order of selection is noted. As a result, if all outcomes are assumed equally likely when the order of selection is noted, then it follows that they remain equally likely when the outcome is taken to be the unordered set of selected balls. Hence, using the latter representation of the experiment, we see that the desired probability is

$$\frac{\binom{6}{1}\binom{5}{2}}{\binom{11}{3}}=\frac{4}{11}$$

which, of course, agrees with the answer obtained previously.

When the experiment consists of a random selection of k items from a set of n items, we have the flexibility of either letting the outcome of the experiment be the ordered selection of the k items or letting it be the unordered set of items selected. In the former case, we would assume that each new selection is equally likely to be any of the so far unselected items of the set, and in the latter case, we would assume that all $\binom{n}{k}$ possible subsets of k items are equally likely to be the set selected. For instance, suppose 5 people are to be randomly selected from a group of 20 individuals consisting of 10 married couples, and we want to determine P(N), the probability that the 5 chosen are all unrelated. (That is, no two are married to each other.) If we regard the sample space as the set of 5 people chosen, then there are $\binom{20}{5}$ equally likely outcomes. An outcome that does not contain a married couple can be thought of as being the result of a six-stage experiment: In the first stage, 5 of the 10 couples to have a member in the group are chosen; in the next 5 stages, 1 of the 2 members of each of these couples is selected. Thus, there are $\binom{10}{5}2^5$ possible outcomes in which the 5 members selected are unrelated, yielding the desired probability of

$$P(N)=\frac{\binom{10}{5}2^5}{\binom{20}{5}}$$

In contrast, we could let the outcome of the experiment be the *ordered* selection of the 5 individuals. In this setting, there are 20·19·18·17·16 equally likely outcomes, of which 20·18·16·14·12 outcomes result in a group of 5 unrelated individuals, yielding the result

$$P(N)=\frac{20\cdot 18\cdot 16\cdot 14\cdot 12}{20\cdot 19\cdot 18\cdot 17\cdot 16}$$

We leave it for the reader to verify that the two answers are identical.

Example 5c

A committee of 5 is to be selected from a group of 6 men and 9 women. If the selection is made randomly, what is the probability that the committee consists of 3 men and 2 women?

Solution

Because each of the $\binom{15}{5}$ possible committees is equally likely to be selected, the desired probability is

$$\frac{\binom{6}{3}\binom{9}{2}}{\binom{15}{5}}=\frac{240}{1001}$$

Example 5d

An urn contains n balls, one of which is special. If k of these balls are withdrawn one at a time, with each selection being equally likely to be any of the balls that remain at the time, what is the probability that the special ball is chosen?

Solution

Since all of the balls are treated in an identical manner, it follows that the set of k balls selected is equally likely to be any of the $\binom{n}{k}$ sets of k balls. Therefore,

$$P\{\text{special ball is selected}\}=\frac{\binom{1}{1}\binom{n-1}{k-1}}{\binom{n}{k}}=\frac{k}{n}$$

We could also have obtained this result by letting A_i denote the event that the special ball is the ith ball to be chosen, $i=1, \ldots, k$. Then, since each one of the n balls is equally likely to be the ith ball chosen, it follows that $P(A_i)=1/n$. Hence, because these events are clearly mutually exclusive, we have

$$P\{\text{special ball is selected}\}=P\left(\bigcup_{i=1}^{k} A_i\right)=\sum_{i=1}^{k} P(A_i)=\frac{k}{n}$$

We could also have argued that $P(A_i)=1/n$, by noting that there are $n(n-1)\cdots(n-k+1)=n!/(n-k)!$ equally likely outcomes of the experiment, of which $(n-1)(n-2)\cdots(n-i+1)(1)(n-i)\cdots(n-k+1)=(n-1)!/(n-k)!$ result in the special ball being the ith one chosen. From this reasoning, it follows that

$P(Ai)=(n-1)!n!=1n$

Example 5e

Suppose that n+m balls, of which n are red and m are blue, are arranged in a linear order in such a way that all (n+m)! possible orderings are equally likely. If we record the result of this experiment by listing only the colors of the successive balls, show that all the possible results remain equally likely.

Solution

Consider any one of the (n+m)! possible orderings, and note that any permutation of the red balls among themselves and of the blue balls among themselves does not change the sequence of colors. As a result, every ordering of colorings corresponds to n! m! different orderings of the n+m balls, so every ordering of the colors has probability n!m! (n+m)! of occurring.

For example, suppose that there are 2 red balls, numbered r1,r2, and 2 blue balls, numbered b1,b2. Then, of the 4! possible orderings, there will be 2! 2! orderings that result in any specified color combination. For instance, the following orderings result in the successive balls alternating in color, with a red ball first:

r 1 , b 1 , r 2 , b 2 r 1 , b 2 , r 2 , b 1 r 2 , b 1 , r 1 , b 2 r 2 , b 2 , r 1 , b 1

Therefore, each of the possible orderings of the colors has probability 424=16 of occurring.

Example 5f

A poker hand consists of 5 cards. If the cards have distinct consecutive values and are not all of the same suit, we say that the hand is a straight. For instance, a hand consisting of the five of spades, six of spades, seven of spades, eight of spades, and nine of hearts is a straight. What is the probability that one is dealt a straight?

Solution

We start by assuming that all (525) possible poker hands are equally likely. To determine the number of outcomes that are straights, let us first determine the number of possible outcomes for which the poker hand consists of an ace, two, three, four, and five (the suits being irrelevant). Since the ace can be any 1 of the 4 possible aces, and similarly for the two, three, four, and five, it follows that there are 45 outcomes leading to exactly one ace, two, three, four, and five. Hence, since in 4 of these outcomes all the cards will be of the same suit (such a hand is called a straight flush), it follows that there are 45−4 hands that make up a straight of the form ace, two, three, four, and five. Similarly, there are 45−4 hands that make up a straight of the form ten, jack, queen, king, and ace. Thus, there are 10(45−4) hands that are straights, and it follows that the desired probability is

$10(45-4)(525)\approx.0039$

Example 5g

A 5-card poker hand is said to be a full house if it consists of 3 cards of the same denomination and 2 other cards of the same denomination (of course, different from the first denomination). Thus, a full house is three of a kind plus a pair. What is the probability that one is dealt a full house?

Solution

Again, we assume that all (525) possible hands are equally likely. To determine the number of possible full houses, we first note that there are (42)(43) different combinations of, say, 2 tens and 3 jacks. Because there are 13 different choices for the kind of pair and, after a pair has been chosen, there are 12 other choices for the denomination of the remaining 3 cards, it follows that the probability of a full house is

13·12·(42)(43)(525)≈.0014

Example 5h

In the game of bridge, the entire deck of 52 cards is dealt out to 4 players. What is the probability that

a. one of the players receives all 13 spades;
b. each player receives 1 ace?

Solution

a. Letting Ei be the event that hand i has all 13 spades, then
P(E i)= 1 (52 13) ,   i=1,2,3,4

   Because the events Ei, i=1, 2, 3, 4, are mutually exclusive, the probability that one of the hands is dealt all 13 spades is
   P(∪ i=1 4 E i)= ∑ i=1 4 P(E i)=4/(52 13)≈6.3× 10 −12

b. Let the outcome of the experiment be the sets of 13 cards of each of the players 1, 2, 3, 4. To determine the number of outcomes in which each of the distinct players receives exactly 1 ace, put aside the aces and note that there are (4812, 12, 12, 12) possible divisions of the other 48 cards when each player is to receive 12. Because there are 4! ways of dividing the 4 aces so that each player receives 1, we see that the number of possible outcomes in which each player receives exactly 1 ace is 4!(4812, 12, 12, 12).

As there are (52 13, 13,13,13) possible hands, the desired probability is thus

4!(4812,12,12,12)(5213,13,13,13)≈.1055

Some results in probability are quite surprising when initially encountered. Our next two examples illustrate this phenomenon.

Example 5i

If n people are present in a room, what is the probability that no two of them celebrate their birthday on the same day of the year? How large need n be so that this probability is less than 12?

Solution

As each person can celebrate his or her birthday on any one of 365 days, there are a total of (365)n possible outcomes. (We are ignoring the possibility of someone having been born on February 29.) Assuming that each outcome is equally likely, we see that the desired probability is (365)(364)(363) ... (365−n+1)/(365)n. It is a rather surprising fact that when n≥23, this probability is less than 12. That is, if there are 23 or more people in a room, then the probability that at least two of them have the same birthday exceeds 12. Many people are initially surprised by this result, since 23 seems so small in relation to 365, the number of days of the year. However, every pair of individuals has probability 365 (365) 2 = 1 365 of having the same birthday, and in a group of 23 people, there are

(232)=253 different pairs of individuals. Looked at this way, the result no longer seems so surprising.

When there are 50 people in the room, the probability that at least two share the same birthday is approximately .970, and with 100 persons in the room, the odds are better than 3,000,000:1. (That is, the probability is greater than 3×1063×106+1 that at least two people have the same birthday.)

Example 5j

A deck of 52 playing cards is shuffled, and the cards are turned up one at a time until the first ace appears. Is the next card–that is, the card following the first ace–more likely to be the ace of spades or the two of clubs?

Solution

To determine the probability that the card following the first ace is the ace of spades, we need to calculate how many of the (52)! possible orderings of the cards have the ace of spades immediately following the first ace. To begin, note that each ordering of the 52 cards can be obtained by first ordering the 51 cards different from the ace of spades and then inserting the ace of spades into that ordering. Furthermore, for each of the (51)! orderings of the other cards, there is only one place where the ace of spades can be placed so that it follows the first ace. For instance, if the ordering of the other 51 cards is

4c, 6h, Jd, 5s, Ac, 7d, ... , Kh

then the only insertion of the ace of spades into this ordering that results in its following the first ace is

4c, 6h, Jd, 5s, Ac, As, 7d, ... , Kh

Therefore, there are (51)! orderings that result in the ace of spades following the first ace, so

P{the ace of spades follows the first ace}=(51)!(52)!=152

In fact, by exactly the same argument, it follows that the probability that the two of clubs (or any other specified card) follows the first ace is also 152. In other words, each of the 52 cards of the deck is equally likely to be the one that follows the first ace!

Many people find this result rather surprising. Indeed, a common reaction is to suppose initially that it is more likely that the two of clubs (rather than the ace of spades) follows the first ace, since that first ace might itself be the ace of spades. This reaction is often followed by the realization that the two of clubs might itself appear before the first ace, thus negating its chance of immediately following the first ace. However, as there is one chance in four that the ace of spades will be the first ace (because all 4 aces are equally likely to be first) and only one chance in five that the two of clubs will appear before the first ace (because each of the set of 5 cards consisting of the two of clubs and the 4 aces is equally likely to be the first of this set to appear), it again appears that the two of clubs is more likely. However, this is not the case, and our more complete analysis shows that they are equally likely.

Example 5k

A football team consists of 20 offensive and 20 defensive players. The players are to be paired in groups of 2 for the purpose of determining roommates. If the pairing is done at random, what is the probability that there are no offensive–defensive roommate pairs? What is the probability that there are 2i offensive–defensive roommate pairs, i=1, 2, ... , 10?

Solution

There are

$$\binom{40}{2, 2, \ldots , 2}=\frac{(40)!}{(2!)^{20}}$$

ways of dividing the 40 players into 20 *ordered* pairs of two each. (That is, there are $(40)!/2^{20}$ ways of dividing the players into a *first* pair, a *second* pair, and so on.) Hence, there are $(40)!/2^{20}(20)!$ ways of dividing the players into (unordered) pairs of 2 each. Furthermore, since a division will result in no offensive–defensive pairs if the offensive (and defensive) players are paired among themselves, it follows that there are $[(20)!/2^{10}(10)!]^2$ such divisions. Hence, the probability of no offensive–defensive roommate pairs, call it P_0, is given by

$$P_0=\frac{\left(\frac{(20)!}{2^{10}(10)!}\right)^2}{\frac{(40)!}{2^{20}(20)!}}=\frac{[(20)!]^3}{[(10)!]^2(40)!}$$

To determine P_{2i}, the probability that there are $2i$ offensive–defensive pairs, we first note that there are $\binom{20}{2i}^2$ ways of selecting the $2i$ offensive players and the $2i$ defensive players who are to be in the offensive–defensive pairs. These $4i$ players can then be paired up into $(2i)!$ possible offensive–defensive pairs. (This is so because the first offensive player can be paired with any of the $2i$ defensive players, the second offensive player with any of the remaining $2i-1$ defensive players, and so on.) As the remaining $20-2i$ offensive (and defensive) players must be paired among themselves, it follows that there are

$$\binom{20}{2i}^2(2i)!\left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2$$

divisions that lead to $2i$ offensive–defensive pairs. Hence,

$$P_{2i}=\frac{\binom{20}{2i}^2(2i)!\left[\frac{(20-2i)!}{2^{10-i}(10-i)!}\right]^2}{\frac{(40)!}{2^{20}(20)!}} \qquad i=0, 1, \ldots , 10$$

The P_{2i}, $i=0, 1, \ldots, 10$, can now be computed, or they can be approximated by making use of a result of Stirling, which shows that $n!$ can be approximated by $n^{n+1/2}e^{-n}\sqrt{2\pi}$. For instance, we obtain

$$P_0 \approx 1.3403\times 10^{-6} \quad P_{10} \approx .345861 \quad P_{20} \approx 7.6068\times 10^{-6}$$

Our next three examples illustrate the usefulness of the inclusion–exclusion identity (**Proposition 4.4**). In **Example 51**, the introduction of probability enables us to obtain a quick solution to a counting problem.

Example 5I

A total of 36 members of a club play tennis, 28 play squash, and 18 play badminton. Furthermore, 22 of the members play both tennis and squash, 12 play both tennis and badminton, 9 play both squash and badminton, and 4 play all three sports. How many members of this club play at least one of three sports?

Solution

Let N denote the number of members of the club, and introduce probability by assuming that a member of the club is randomly selected. If, for any subset C of members of the club, we let $P(C)$ denote the probability that the selected member is contained in C, then

$$P(C)=\frac{\text{number of members in } C}{N}$$

Now, with T being the set of members that plays tennis, S being the set that plays squash, and B being the set that plays badminton, we have, from **Proposition 4.4**,

$$P(T\cup S\cup B) = P(T)+P(S)+P(B)-P(TS)-P(TB)-P(SB)+P(TSB) = \frac{36+28+18-22-12-9+4}{N} = \frac{43}{N}$$

Hence, we can conclude that 43 members play at least one of the sports.

The next example in this section not only possesses the virtue of giving rise to a somewhat surprising answer, but is also of theoretical interest.

Example 5m The Matching Problem

Suppose that each of N men at a party throws his hat into the center of the room. The hats are first mixed up, and then each man randomly selects a hat. What is the probability that none of the men selects his own hat?

Solution

We first calculate the complementary probability of at least one man selecting his own hat. Let us denote by E_i, $i=1, 2, \ldots, N$ the event that the ith man selects his own hat. Now, by the inclusion-exclusion identity $P(\cup_{i=1}^{N} E_i)$, the probability that at least one of the men selects his own hat, is given by

$$P\left(\cup_{i=1}^{N} E_i\right)=\sum_{i=1}^{N} P(E_i)-\sum_{i_1<i_2} P(E_{i_1}E_{i_2})+\cdots+(-1)^{n+1}\sum_{i_1<i_2\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n})+\cdots+(-1)^{N+1}P(E_1E_2\cdots E_N)$$

If we regard the outcome of this experiment as a vector of *N* numbers, where the ith element is the number of the hat drawn by the ith man, then there are N! possible outcomes. [The outcome (1, 2, 3, ... , N) means, for example, that each man selects his own hat.] Furthermore, $E_{i_1}E_{i_2}\ldots E_{i_n}$, the event that each of the *n* men $i_1, i_2, \ldots, i_n$ selects his own hat, can occur in any of $(N-n)(N-n-1)\cdots 3\cdot 2\cdot 1=(N-n)!$ possible ways; for, of the remaining $N-n$ men, the first can select any of $N-n$ hats, the second can then select any of $N-n-1$ hats, and so on. Hence, assuming that all N! possible outcomes are equally likely, we see that

$$P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{(N-n)!}{N!}$$

Also, as there are $\binom{N}{n}$ terms in $\sum_{i_1<i_2\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n})$, it follows that

$$\sum_{i_1<i_2\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{N!(N-n)!}{(N-n)!\,n!\,N!}=\frac{1}{n!}$$

Thus,

$$P\left(\cup_{i=1}^{N} E_i\right)=1-\frac{1}{2!}+\frac{1}{3!}-\cdots+(-1)^{N+1}\frac{1}{N!}$$

Hence, the probability that none of the men selects his own hat is

$$1-1+\frac{1}{2!}-\frac{1}{3!}+\ldots+\frac{(-1)^N}{N!}=\sum_{i=0}^{N}(-1)^i/i!$$

Upon letting $x=-1$ in the identity $e^x=\sum_{i=0}^{\infty} x^i/i!$, the preceding probability when N is large is seen to be approximately equal to $e^{-1}\approx .3679$. In other words, for N large, the probability that none of the men selects his own hat is approximately .37. (How many readers would have incorrectly thought that this probability would go to 1 as $N\to\infty$?)

For another illustration of the usefulness of the inclusion-exclusion identity, consider the following example.

Example 5n

Compute the probability that if 10 married couples are seated at random at a round table, then no wife sits next to her husband.

Solution

If we let $E_i, i=1, 2, \ldots, 10$ denote the event that the ith couple sit next to each other, it follows that the desired probability is $1-P(\cup_{i=1}^{10} E_i)$. Now, from the inclusion-exclusion identity,

$$P\left(\cup_1^{10} E_i\right)=\sum_1^{10} P(E_i)-\cdots+(-1)^{n+1}\sum_{i_1<i_2<\cdots<i_n} P(E_{i_1}E_{i_2}\cdots E_{i_n})+\cdots-P(E_1E_2\cdots E_{10})$$

To compute $P(E_{i_1}E_{i_2}\cdots E_{i_n})$, we first note that there are 19! ways of arranging 20 people around a round table. (Why?) The number of arrangements that result in a specified set of n men sitting next to their wives can most easily be obtained by first thinking of each of the n married couples as being single entities. If this were the case, then we would need to arrange $20-n$ entities around a round table, and there are clearly $(20-n-1)!$ such arrangements. Finally, since each of the n married couples can be arranged next to each other in one of two possible ways, it follows that there are $2^n(20-n-1)!$ arrangements that result in a specified set of n men each sitting next to their wives. Therefore,

$$P(E_{i_1}E_{i_2}\cdots E_{i_n})=\frac{2^n(19-n)!}{(19)!}$$

Thus, from **Proposition 4.4**, we obtain that the probability that at least one married couple sits together is

$$\binom{10}{1}2^1\frac{(18)!}{(19)!}-\binom{10}{2}2^2\frac{(17)!}{(19)!}+\binom{10}{3}2^3\frac{(16)!}{(19)!}-\cdots-\binom{10}{10}2^{10}\frac{9!}{(19)!}\approx .6605$$

and the desired probability is approximately .3395.

***Example 5o Runs**

Consider an athletic team that had just finished its season with a final record of n wins and m losses. By examining the sequence of wins and losses, we are hoping to determine whether the team had stretches of games in which it was more likely to win than at other times. One way to gain some insight into this question is to count the number of runs of wins and then see how likely that result would be when all $(n+m)!/(n!\,m!)$ orderings of the n wins and m losses are assumed equally likely. By a run of wins, we mean a consecutive sequence of wins. For instance, if $n=10$, $m=6$, and the sequence of outcomes was *WWLLWWWLWLLLWWWW*, then there would be 4 runs of wins–the first run being of size 2, the second of size 3, the third of size 1, and the fourth of size 4.

Suppose now that a team has n wins and m losses. Assuming that all $(n+m)!/(n!\,m!)=\binom{n+m}{n}$ orderings are equally likely, let us determine the probability that there will be exactly r runs of wins. To do so, consider first any vector of positive integers $x_1, x_2, \ldots, x_r$ with $x_1+\cdots+x_r=n$, and let us see how many outcomes result in *r* runs of wins in which the ith run is of size x_i, $i=1, \ldots, r$. For any such outcome, if we let y_1 denote the number of losses before the first run of wins, y_2 the number of losses between the first 2 runs of wins, ... , y_{r+1} the number of losses after the last run of wins, then the y_i satisfy

$$y_1+y_2+\cdots+y_{r+1}=m \qquad y_1\ge 0, y_{r+1}\ge 0, y_i>0, i=2, \ldots, r$$

and the outcome can be represented schematically as

$$\underbrace{LL\ldots L}_{y_1}\underbrace{WW\ldots W}_{x_1}\underbrace{L\ldots L}_{y_2}\underbrace{WW\ldots W}_{x_2}\cdots\underbrace{WW}_{x_r}\underbrace{L\ldots L}_{y_{r+1}}$$

Hence, the number of outcomes that result in r runs of wins the ith of size x_i, $i=1, \ldots r-$ is equal to the number of integers $y_1, \ldots, y_{r+1}$ that satisfy the foregoing, or, equivalently, to the number of positive integers

$$\bar{y}_1 = y_1+1 \quad \bar{y}_1 = y_i, i=2,\ldots,r, \quad \bar{y}_{r+1} = y_{r+1}+1$$

that satisfy

$$\bar{y}_1+\bar{y}_2+\cdots+\bar{y}_{r+1}=m+2$$

By **Proposition 6.1** in **Chapter 1**, there are $\binom{m+1}{r}$ such outcomes. Hence, the total number of outcomes that result in r runs of wins is $\binom{m+1}{r}$ multiplied by the number of positive integral solutions of $x_1+\cdots+x_r=n$. Thus, again from **Proposition 6.1**, there are $\binom{m+1}{r}\binom{n-1}{r-1}$ outcomes resulting in r runs of wins. As there are $\binom{n+m}{n}$ equally likely outcomes, it follows that

$$P(\{r \text{ runs of wins}\})=\frac{\binom{m+1}{r}\binom{n-1}{r-1}}{\binom{m+n}{n}} \quad r\ge 1$$

For example, if $n=8$ and $m=6$, then the probability of 7 runs is $\binom{7}{7}\binom{7}{6}/\binom{14}{8}=1/429$ if all $\binom{14}{8}$ outcomes are equally likely. Hence, if the outcome was *WLWLWLWLWWLWLW*, then we might suspect that the team's probability of winning was changing over time. (In particular, the probability that the team wins seems to be quite high when it lost its last game and quite low when it won its last game.) On the other extreme, if the outcome were *WWWWWWWWLLLLLL*, then there would have been only 1 run, and as $P(\{1\text{ run}\})=\binom{7}{1}\binom{7}{0}/\binom{14}{8}=1/429$, it would thus again seem unlikely that the team's probability of winning remained unchanged over its 14 games.

*2.6 Probability as a Continuous Set Function

A sequence of events $\{E_n, n \geq 1\}$ is said to be an increasing sequence if

$$E_1 \subset E_2 \subset \cdots \subset E_n \subset E_{n+1} \subset \cdots$$

whereas it is said to be a decreasing sequence if

$$E_1 \supset E_2 \supset \cdots \supset E_n \supset E_{n+1} \supset \cdots$$

If $\{E_n, n \geq 1\}$ is an increasing sequence of events, then we define a new event, denoted by $\lim_{n\to\infty} E_n$, by

$$\lim_{n\to\infty} E_n = \bigcup_{i=1}^{\infty} E_i$$

Similarly, if $\{E_n, n \geq 1\}$ is a decreasing sequence of events, we define $\lim_{n\to\infty} E_n$, by

$$\lim_{n\to\infty} E_n = \bigcap_{i=1}^{\infty} E_i$$

We now prove the following **Proposition 6.1**:

Proposition 6.1

If $\{E_n, n \geq 1\}$ is either an increasing or a decreasing sequence of events, then

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

Proof Suppose, first, that $\{E_n, n \geq 1\}$ is an increasing sequence, and define the events F_n, $n \geq 1$, by

$$F_1 = E_1 \quad F_n = E_n \left(\bigcup_1^{n-1} E_i \right)^c = E_n E_{n-1}^c \quad n > 1$$

where we have used the fact that $\bigcup_1^{n-1} E_i = E_{n-1}$, since the events are increasing. In words, F_n consists of those outcomes in E_n that are not in any of the earlier $E_i, i<n$. It is easy to verify that the F_n are mutually exclusive events such that

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} E_i \quad \text{and} \quad \bigcup_{i=1}^{n} F_i = \bigcup_{i=1}^{n} E_i \quad \text{for all } n \geq 1$$

Thus,

$$P\left(\bigcup_1^{\infty} E_i \right) = P\left(\bigcup_1^{\infty} F_i \right) = \sum_1^{\infty} P(F_i) \quad \text{(by Axiom 3)} = \lim_{n\to\infty} \sum_1^{n} P(F_i) = \lim_{n\to\infty} P\left(\bigcup_1^{n} F_i \right) = \lim_{n\to\infty} P\left(\bigcup_1^{n} E_i \right) = \lim_{n\to\infty} P(E_n)$$

which proves the result when $\{E_n, n \geq 1\}$ is increasing.

If $\{E_n, n \geq 1\}$ is a decreasing sequence, then $\{E_n^c, n \geq 1\}$ is an increasing sequence; hence, from the preceding equations,

$$P\left(\bigcup_1^{\infty} E_i^c \right) = \lim_{n\to\infty} P(E_n^c)$$

However, because $\bigcup_1^{\infty} E_i^c = \left(\bigcap_1^{\infty} E_i \right)^c$, it follows that

$$P\left(\left(\bigcap_1^{\infty} E_i \right)^c \right) = \lim_{n\to\infty} p(E_n^c)$$

or, equivalently,

$$1-P\left(\bigcap_1^\infty E_i\right)=\lim_{n\to\infty}[1-P(E_n)]=1-\lim_{n\to\infty}P(E_n)$$

or

$$P\left(\bigcap_1^\infty E_i\right)=\lim_{n\to\infty}P(E_n)$$

which proves the result.

Example 6a Probability and a “Paradox”

Suppose that we possess an infinitely large urn and an infinite collection of balls labeled ball number 1, number 2, number 3, and so on. Consider an experiment performed as follows: At 1 minute to 12 P.M., balls numbered 1 through 10 are placed in the urn and ball number 10 is withdrawn. (Assume that the withdrawal takes no time.) At 12 minute to 12 P.M., balls numbered 11 through 20 are placed in the urn and ball number 20 is withdrawn. At 14 minute to 12 P.M., balls numbered 21 through 30 are placed in the urn and ball number 30 is withdrawn. At 18 minute to 12 P.M., and so on. The question of interest is, How many balls are in the urn at 12 P.M.?

The answer to this question is clearly that there is an infinite number of balls in the urn at 12 P.M., since any ball whose number is not of the form $10n$, $n\geq 1$, will have been placed in the urn and will not have been withdrawn before 12 P.M. Hence, the problem is solved when the experiment is performed as described.

However, let us now change the experiment and suppose that at 1 minute to 12 P.M., balls numbered 1 through 10 are placed in the urn and ball number 1 is withdrawn; at 12 minute to 12 P.M., balls numbered 11 through 20 are placed in the urn and ball number 2 is withdrawn; at 14 minute to 12 P.M, balls numbered 21 through 30 are placed in the urn and ball number 3 is withdrawn; at 18 minute to 12 P.M., balls numbered 31 through 40 are placed in the urn and ball number 4 is withdrawn, and so on. For this new experiment, how many balls are in the urn at 12 P.M.?

Surprisingly enough, the answer now is that the urn is *empty* at 12 P.M. For, consider any ball–say, ball number n. At some time prior to 12 P.M. [in particular, at $(12)n-1$ minutes to 12 P.M.], this ball would have been withdrawn from the urn. Hence, for each n, ball number n is not in the urn at 12 P.M.; therefore, the urn must be empty at that time.

Because for all n, the number of balls in the urn after the *n*th interchange is the same in both variations of the experiment, most people are surprised that the two scenarios produce such different results in the limit. It is important to recognize that the reason the results are different is not because there is an actual *paradox*, or mathematical contradiction, but rather because of the logic of the situation, and also that the surprise results because one’s initial intuition when dealing with infinity is not always correct. (This latter statement is not surprising, for when the theory of the infinite was first developed by the mathematician Georg Cantor in the second half of the nineteenth century, many of the other leading mathematicians of the day called it nonsensical and ridiculed Cantor for making such claims as that the set of all integers and the set of all even integers have the same number of elements.)

We see from the preceding discussion that the manner in which the balls are withdrawn makes a difference. For, in the first case, only balls numbered $10n, n\geq 1$, are ever withdrawn, whereas in the second case all of the balls are eventually withdrawn. Let us now suppose that whenever a ball is to be withdrawn, that ball is randomly selected from among those present. That is, suppose that at 1 minute to 12 P.M. balls numbered 1 through 10 are placed in the urn and a ball is randomly selected and withdrawn, and so on. In this case, how many balls are in the urn at 12 P.M.?

Solution

We shall show that, with probability 1, the urn is empty at 12 P.M. Let us first consider ball number 1. Define E_n to be the event that ball number 1 is still in the urn after the first n withdrawals have been made. Clearly,

$$P(E_n)=\frac{9\cdot 18\cdot 27 \cdots (9n)}{10\cdot 19\cdot 28 \cdots (9n+1)}$$

[To understand this equation, just note that if ball number 1 is still to be in the urn after the first n withdrawals, the first ball withdrawn can be any one of 9, the second any one of 18 (there are 19 balls in the urn at the time of the second withdrawal, one of which must be ball number 1), and so on. The denominator is similarly obtained.]

Now, the event that ball number 1 is in the urn at 12 P.M. is just the event $\bigcap_{n=1}^{\infty} E_n$. Because the events $E_n, n\geq 1$, are decreasing events, it follows from **Proposition 6.1** that

$$\begin{aligned} P\{\text{ball number 1 is in the urn at 12 P.M.}\} &= P\left(\bigcap_{n=1}^{\infty} E_n\right) \\ &= \lim_{n\to\infty} P(E_n) \\ &= \prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right) \end{aligned}$$

We now show that

$$\prod_{n=1}^{\infty}\frac{9n}{9n+1}=0$$

Since

$$\prod_{n=1}^{\infty}\left(\frac{9n}{9n+1}\right)=\left[\prod_{n=1}^{\infty}\left(\frac{9n+1}{9n}\right)\right]^{-1}$$

this is equivalent to showing that

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right)=\infty$$

Now, for all $m\geq 1$,

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right) \geq \prod_{n=1}^{m}\left(1+\frac{1}{9n}\right) = \left(1+\frac{1}{9}\right)\left(1+\frac{1}{18}\right)\left(1+\frac{1}{27}\right)\cdots\left(1+\frac{1}{9m}\right) > \frac{1}{9}+\frac{1}{18}+\frac{1}{27}+\cdots+\frac{1}{9m} = \frac{1}{9}\sum_{i=1}^{m}\frac{1}{i}$$

Hence, letting $m\to\infty$ and using the fact that $\sum_{i=1}^{\infty} 1/i=\infty$ yields

$$\prod_{n=1}^{\infty}\left(1+\frac{1}{9n}\right)=\infty$$

Thus, letting F_i denote the event that ball number i is in the urn at 12 P.M., we have shown that $P(F_1)=0$. Similarly, we can show that $P(F_i)=0$ for all i.

(For instance, the same reasoning shows that $P(F_i)=\prod_{n=2}^{\infty}[9n/(9n+1)]$ for $i=11,12,\ldots,20$.) Therefore, the probability that the urn is not empty at 12 P.M., $P\left(\bigcup_1^{\infty} F_i\right)$, satisfies

$$P\left(\bigcup_1^{\infty} F_i\right)\leq \sum_1^{\infty} P(F_i)=0$$

by Boole's inequality.

Thus, with probability 1, the urn will be empty at 12 P.M.

2.7 Probability as a Measure of Belief

Thus far we have interpreted the probability of an event of a given experiment as being a measure of how frequently the event will occur when the experiment is continually repeated. However, there are also other uses of the term *probability*. For instance, we have all heard such statements as “It is 90 percent probable that Shakespeare actually wrote *Hamlet*” or “The probability that Oswald acted alone in assassinating Kennedy is .8.” How are we to interpret these statements?

The most simple and natural interpretation is that the probabilities referred to are measures of the individual’s degree of belief in the statements that he or she is making. In other words, the individual making the foregoing statements is quite certain that Oswald acted alone and is even more certain that Shakespeare wrote *Hamlet*. This interpretation of probability as being a measure of the degree of one’s belief is often referred to as the *personal* or *subjective* view of probability.

It seems logical to suppose that a “measure of the degree of one’s belief” should satisfy all of the axioms of probability. For example, if we are 70 percent certain that Shakespeare wrote *Julius Caesar* and 10 percent certain that it was actually Marlowe, then it is logical to suppose that we are 80 percent certain that it was either Shakespeare or Marlowe. Hence, whether we interpret probability as a measure of belief or as a long-run frequency of occurrence, its mathematical properties remain unchanged.

Example 7a

Suppose that in a 7-horse race, you believe that each of the first 2 horses has a 20 percent chance of winning, horses 3 and 4 each have a 15 percent chance, and the remaining 3 horses have a 10 percent chance each. Would it be better for you to wager at even money that the winner will be one of the first three horses or to wager, again at even money, that the winner will be one of the horses 1, 5, 6, and 7?

Solution

On the basis of your personal probabilities concerning the outcome of the race, your probability of winning the first bet is $.2+.2+.15=.55$, whereas it is $.2+.1+.1+.1=.5$ for the second bet. Hence, the first wager is more attractive.

Note that in supposing that a person’s subjective probabilities are always consistent with the axioms of probability, we are dealing with an idealized rather than an actual person. For instance, if we were to ask someone what he thought the chances were of

a. rain today,
b. rain tomorrow,
c. rain both today and tomorrow,
d. rain either today or tomorrow,

it is quite possible that, after some deliberation, he might give 30 percent, 40 percent, 20 percent, and 60 percent as answers. Unfortunately, such answers (or such subjective probabilities) are not consistent with the axioms of probability. (Why not?) We would of course hope that after this was pointed out to the respondent, he would change his answers. (One possibility we could accept is 30 percent, 40 percent, 10 percent, and 60 percent.)

Summary

Let S denote the set of all possible outcomes of an experiment. S is called the *sample space* of the experiment. An *event* is a subset of S. If $A_i, i=1,\ldots,n$, are events, then $\cup_{i=1}^{n} A_i$, called the *union* of these events, consists of all outcomes that are in at least one of the events $A_i, i=1,\ldots,n$. Similarly, $\cap_{i=1}^{n} A_i$, sometimes written as $A_1 \cdots A_n$, is called the *intersection* of the events A_i and consists of all outcomes that are in all of the events $A_i, i=1,\ldots,n$.

For any event A, we define A^c to consist of all outcomes in the sample space that are not in A. We call A^c the *complement* of the event *A*. The event S^c, which is empty of outcomes, is designated by $\varnothing$ and is called the *null* set. If $AB=\varnothing$, then we say that A and B are *mutually exclusive*.

For each event A of the sample space S, we suppose that a number $P(A)$, called the probability of A, is defined and is such that

i. $0 \le P(A) \le 1$
ii. $P(S)=1$
iii. For mutually exclusive events $A_i, i \ge 1$,

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

$P(A)$ represents the *probability* that the outcome of the experiment is in A.

It can be shown that

$$P(A^c)=1-P(A)$$

A useful result is that

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

which can be generalized to give

$$P\left(\cup_{i=1}^{n} A_i\right)=\sum_{i=1}^{n} P(A_i)-\sum\sum_{i<j} P(A_i A_j)+\sum\sum\sum_{i<j<k} P(A_i A_j A_k)+\cdots+(-1)^{n+1} P(A_1 \cdots A_n)$$

This result is known as the *inclusion–exclusion identity*.

If S is finite and each one point set is assumed to have equal probability, then

$$P(A)=\frac{|A|}{|S|}$$

where $|E|$ denotes the number of outcomes in the event E.

$P(A)$ can be interpreted either as a long-run relative frequency or as a measure of one's degree of belief.

Problems

1. A box contains 3 marbles: 1 red, 1 green, and 1 blue. Consider an experiment that consists of taking 1 marble from the box and then replacing it in the box and drawing a second marble from the box. Describe the sample space. Repeat when the second marble is drawn without replacing the first marble.
2. In an experiment, die is rolled continually until a 6 appears, at which point the experiment stops. What is the sample space of this experiment? Let En denote the event that n rolls are necessary to complete the experiment. What points of the sample space are contained in En? What is $(\cup 1\infty En)c$?
3. Two dice are thrown. Let E be the event that the sum of the dice is odd, let F be the event that at least one of the dice lands on 1, and let G be the event that the sum is 5. Describe the events EF,E∪F,FG,EFc, and *EFG*.
4. *A*, *B*, and *C* take turns flipping a coin. The first one to get a head wins. The sample space of this experiment can be defined by
S={1,01,001,0001,...,0000⋯

   a. Interpret the sample space.
   b. Define the following events in terms of S: 10pt
      i. *A* wins = *A*.
      ii. *B* wins = *B*.
      iii. (A∪B)c.
      Assume that *A* flips first, then *B*, then *C*, then *A*, and so on.

5. A system is composed of 5 components, each of which is either working or failed. Consider an experiment that consists of observing the status of each component, and let the outcome of the experiment be given by the vector (x1, x2, x3, x4, x5), where xi is equal to 1 if component i is working and is equal to 0 if component i is failed.
   a. How many outcomes are in the sample space of this experiment?
   b. Suppose that the system will work if components 1 and 2 are both working, or if components 3 and 4 are both working, or if components 1, 3, and 5 are all working. Let W be the event that the system will work. Specify all the outcomes in W.
   c. Let A be the event that components 4 and 5 are both failed. How many outcomes are contained in the event A?
   d. Write out all the outcomes in the event *AW*.

6. A hospital administrator codes incoming patients suffering gunshot wounds according to whether they have insurance (coding 1 if they do and 0 if they do not) and according to their condition, which is rated as good (g), fair (f), or serious (s). Consider an experiment that consists of the coding of such a patient.
   a. Give the sample space of this experiment.
   b. Let A be the event that the patient is in serious condition. Specify the outcomes in A.
   c. Let B be the event that the patient is uninsured. Specify the outcomes in B.
   d. Give all the outcomes in the event B c ∪A.

7. Consider an experiment that consists of determining the type of job–either blue collar or white collar–and the political affiliation Republican, Democratic, or Independent–of the 15 members of an adult soccer team. How many outcomes are
   a. in the sample space?
   b. in the event that at least one of the team members is a blue-collar worker?
   c. in the event that none of the team members considers himself or herself an Independent?

8. Suppose that A and B are mutually exclusive events for which P(A)=.3 and P(B)=.5. What is the

probability that

a. either A or B occurs?
b. A occurs but B does not?
c. both A and B occur?

9. A retail establishment accepts either the American Express or the VISA credit card. A total of 24 percent of its customers carry an American Express card, 61 percent carry a VISA card, and 11 percent carry both cards. What percentage of its customers carry a credit card that the establishment will accept?

10. Sixty percent of the students at a certain school wear neither a ring nor a necklace. Twenty percent wear a ring and 30 percent wear a necklace. If one of the students is chosen randomly, what is the probability that this student is wearing

a. a ring or a necklace?
b. a ring and a necklace?

11. A total of 28 percent of American males smoke cigarettes, 7 percent smoke cigars, and 5 percent smoke both cigars and cigarettes.

a. What percentage of males smokes neither cigars nor cigarettes?
b. What percentage smokes cigars but not cigarettes?

12. An elementary school is offering 3 language classes: one in Spanish, one in French, and one in German. The classes are open to any of the 100 students in the school. There are 28 students in the Spanish class, 26 in the French class, and 16 in the German class. There are 12 students who are in both Spanish and French, 4 who are in both Spanish and German, and 6 who are in both French and German. In addition, there are 2 students taking all 3 classes.

a. If a student is chosen randomly, what is the probability that he or she is not in any of the language classes?
b. If a student is chosen randomly, what is the probability that he or she is taking exactly one language class?
c. If 2 students are chosen randomly, what is the probability that at least 1 is taking a language class?

13. A certain town with a population of 100,000 has 3 newspapers: I, II, and III. The proportions of townspeople who read these papers are as follows:
I: 10 percent I and II: 8 percent I and II and III: 1 percent
II: 30 percent I and III: 2 percent
III: 5 percent II and III: 4 percent
(The list tells us, for instance, that 8000 people read newspapers I and II.)

a. Find the number of people who read only one newspaper.
b. How many people read at least two newspapers?
c. If I and III are morning papers and II is an evening paper, how many people read at least one morning paper plus an evening paper?
d. How many people do not read any newspapers?
e. How many people read only one morning paper and one evening paper?

14. The following data were given in a study of a group of 1000 subscribers to a certain magazine: In reference to job, marital status, and education, there were 312 professionals, 470 married persons, 525 college graduates, 42 professional college graduates, 147 married college graduates, 86 married professionals, and 25 married professional college graduates. Show that the numbers reported in the study must be incorrect.
Hint: Let M, W, and G denote, respectively, the set of professionals, married persons, and college graduates. Assume that one of the 1000 persons is chosen at random, and use **Proposition 4.4** to show that if the given numbers are correct, then $P(M \cup W \cup G) > 1$.

15. If it is assumed that all (525) poker hands are equally likely, what is the probability of being dealt

a. a flush? (A hand is said to be a flush if all 5 cards are of the same suit.)
b. one pair? (This occurs when the cards have denominations a, a, b, c, d, where a, b, c, and d are all distinct.)
c. two pairs? (This occurs when the cards have denominations a, a, b, b, c, where a, b, and c are all distinct.)
d. three of a kind? (This occurs when the cards have denominations a, a, a, b, c, where a, b, and c are all distinct.)
e. four of a kind? (This occurs when the cards have denominations a, a, a, a, b.)

16. Poker dice is played by simultaneously rolling 5 dice. Show that

a. P{no two alike}=.0926;
b. P{one pair}=.4630;
c. P{two pair}=.2315;
d. P{three alike}=.1543;
e. P{full house}=.0386;
f. P{four alike}=.0193;
g. P{five alike}=.0008.

17. Twenty five people, consisting of 15 women and 10 men are lined up in a random order. Find the probability that the ninth woman to appear is in position 17. That is, find the probability there are 8 women in positions 1 thru 16 and a woman in position 17.
18. Two cards are randomly selected from an ordinary playing deck. What is the probability that they form a blackjack? That is, what is the probability that one of the cards is an ace and the other one is either a ten, a jack, a queen, or a king?
19. Two symmetric dice have had two of their sides painted red, two painted black, one painted yellow, and the other painted white. When this pair of dice is rolled, what is the probability that both dice land with the same color face up?
20. Suppose that you are playing blackjack against a dealer. In a freshly shuffled deck, what is the probability that neither you nor the dealer is dealt a blackjack?
21. A small community organization consists of 20 families, of which 4 have one child, 8 have two children, 5 have three children, 2 have four children, and 1 has five children.

a. If one of these families is chosen at random, what is the probability it has i children, i=1, 2, 3, 4, 5?
b. If one of the children is randomly chosen, what is the probability that child comes from a family having i children, i=1, 2, 3, 4, 5?

22. Consider the following technique for shuffling a deck of n cards: For any initial ordering of the cards, go through the deck one card at a time and at each card, flip a fair coin. If the coin comes up heads, then leave the card where it is; if the coin comes up tails, then move that card to the end of the deck. After the coin has been flipped n times, say that one round has been completed. For instance, if n=4 and the initial ordering is 1, 2, 3, 4, then if the successive flips result in the outcome h,t,t,h, then the ordering at the end of the round is 1, 4, 2, 3. Assuming that all possible outcomes of the sequence of n coin flips are equally likely, what is the probability that the ordering after one round is the same as the initial ordering?
23. A pair of fair dice is rolled. What is the probability that the second die lands on a higher value than does the first?
24. If two dice are rolled, what is the probability that the sum of the upturned faces equals i? Find it for i=2, 3, ... , 11, 12.
25. A pair of dice is rolled until a sum of either 5 or 7 appears. Find the probability that a 5 occurs first.
Hint: Let En denote the event that a 5 occurs on the nth roll and no 5 or 7 occurs on the first n−1 rolls. Compute P(En) and argue that ∑n=1∞P(En) is the desired probability.
26. The game of craps is played as follows: A player rolls two dice. If the sum of the dice is either a 2,

3, or 12, the player loses; if the sum is either a 7 or an 11, the player wins. If the outcome is anything else, the player continues to roll the dice until she rolls either the initial outcome or a 7. If the 7 comes first, the player loses, whereas if the initial outcome reoccurs before the 7 appears, the player wins. Compute the probability of a player winning at craps.
Hint: Let Ei denote the event that the initial outcome is i and the player wins. The desired probability is ∑i=212P(Ei). To compute P(Ei), define the events Ei,n to be the event that the initial sum is *i* and the player wins on the nth roll. Argue that P(E i)= ∑ n=1 ∞ P(E i,n).
27. An urn contains 3 red and 7 black balls. Players A and B withdraw balls from the urn consecutively until a red ball is selected. Find the probability that A selects the red ball. (A draws the first ball, then B, and so on. There is no replacement of the balls drawn.)
28. An urn contains 5 red, 6 blue, and 8 green balls. If a set of 3 balls is randomly selected, what is the probability that each of the balls will be (a) of the same color? (b) of different colors? Repeat under the assumption that whenever a ball is selected, its color is noted and it is then replaced in the urn before the next selection. This is known as *sampling with replacement*.
29. An urn contains n white and m black balls, where n and m are positive numbers.

a. If two balls are randomly withdrawn, what is the probability that they are the same color?
b. If a ball is randomly withdrawn and then replaced before the second one is drawn, what is the probability that the withdrawn balls are the same color?
c. Show that the probability in part (b) is always larger than the one in part (a).

30. The chess clubs of two schools consist of, respectively, 8 and 9 players. Four members from each club are randomly chosen to participate in a contest between the two schools. The chosen players from one team are then randomly paired with those from the other team, and each pairing plays a game of chess. Suppose that Rebecca and her sister Elise are on the chess clubs at different schools. What is the probability that

a. Rebecca and Elise will be paired?
b. Rebecca and Elise will be chosen to represent their schools but will not play each other?
c. either Rebecca or Elise will be chosen to represent her school?

31. A 3-person basketball team consists of a guard, a forward, and a center.

a. If a person is chosen at random from each of three different such teams, what is the probability of selecting a complete team?
b. What is the probability that all 3 players selected play the same position?

32. A group of individuals containing b boys and g girls is lined up in random order; that is, each of the (b+g)! permutations is assumed to be equally likely. What is the probability that the person in the ith position, 1≤i≤b+g, is a girl?
33. A forest contains 20 elk, of which 5 are captured, tagged, and then released. A certain time later, 4 of the 20 elk are captured. What is the probability that 2 of these 4 have been tagged? What assumptions are you making?
34. The second Earl of Yarborough is reported to have bet at odds of 1000 to 1 that a bridge hand of 13 cards would contain at least one card that is ten or higher. (By *ten or higher* we mean that a card is either a ten, a jack, a queen, a king, or an ace.) Nowadays, we call a hand that has no cards higher than 9 a *Yarborough*. What is the probability that a randomly selected bridge hand is a Yarborough?
35. Seven balls are randomly withdrawn from an urn that contains 12 red, 16 blue, and 18 green balls. Find the probability that

a. 3 red, 2 blue, and 2 green balls are withdrawn;
b. at least 2 red balls are withdrawn;
c. all withdrawn balls are the same color;
d. either exactly 3 red balls or exactly 3 blue balls are withdrawn.

36. Two cards are chosen at random from a deck of 52 playing cards. What is the probability that they

a. are both aces?

b. have the same value?

37. An instructor gives her class a set of 10 problems with the information that the final exam will consist of a random selection of 5 of them. If a student has figured out how to do 7 of the problems, what is the probability that he or she will answer correctly

a. all 5 problems?
b. at least 4 of the problems?

38. There are *n* socks, 3 of which are red, in a drawer. What is the value of n if, when 2 of the socks are chosen randomly, the probability that they are both red is 12?

39. There are 5 hotels in a certain town. If 3 people check into hotels in a day, what is the probability that they each check into a different hotel? What assumptions are you making?

40. If 4 balls are randomly chosen from an urn containing 4 red, 5 white, 6 blue, and 7 green balls, find the probability that

a. at least one of the 4 balls chosen is green;
b. one ball of each color is chosen.

41. If a die is rolled 4 times, what is the probability that 6 comes up at least once?

42. Two dice are thrown *n* times in succession. Compute the probability that double 6 appears at least once. How large need n be to make this probability at least 12?

43.

a. If N people, including A and B, are randomly arranged in a line, what is the probability that A and B are next to each other?
b. What would the probability be if the people were randomly arranged in a circle?

44. Five people, designated as A, B, C, D, E, are arranged in linear order. Assuming that each possible order is equally likely, what is the probability that

a. there is exactly one person between A and B?
b. there are exactly two people between A and B?
c. there are three people between A and B?

45. A woman has n keys, of which one will open her door.

a. If she tries the keys at random, discarding those that do not work, what is the probability that she will open the door on her kth try?
b. What if she does not discard previously tried keys?

46. How many people have to be in a room in order that the probability that at least two of them celebrate their birthday in the same month is at least 12? Assume that all possible monthly outcomes are equally likely.

47. Suppose that 5 of the numbers 1, 2, ... , 14 are chosen. Find the probability that 9 is the third smallest value chosen.

48. Given 20 people, what is the probability that among the 12 months in the year, there are 4 months containing exactly 2 birthdays and 4 containing exactly 3 birthdays?

49. A group of 6 men and 6 women is randomly divided into 2 groups of size 6 each. What is the probability that both groups will have the same number of men?

50. In a hand of bridge, find the probability that you have 5 spades and your partner has the remaining 8.

51. Suppose that n balls are randomly distributed into N compartments. Find the probability that m balls will fall into the first compartment. Assume that all Nn arrangements are equally likely.

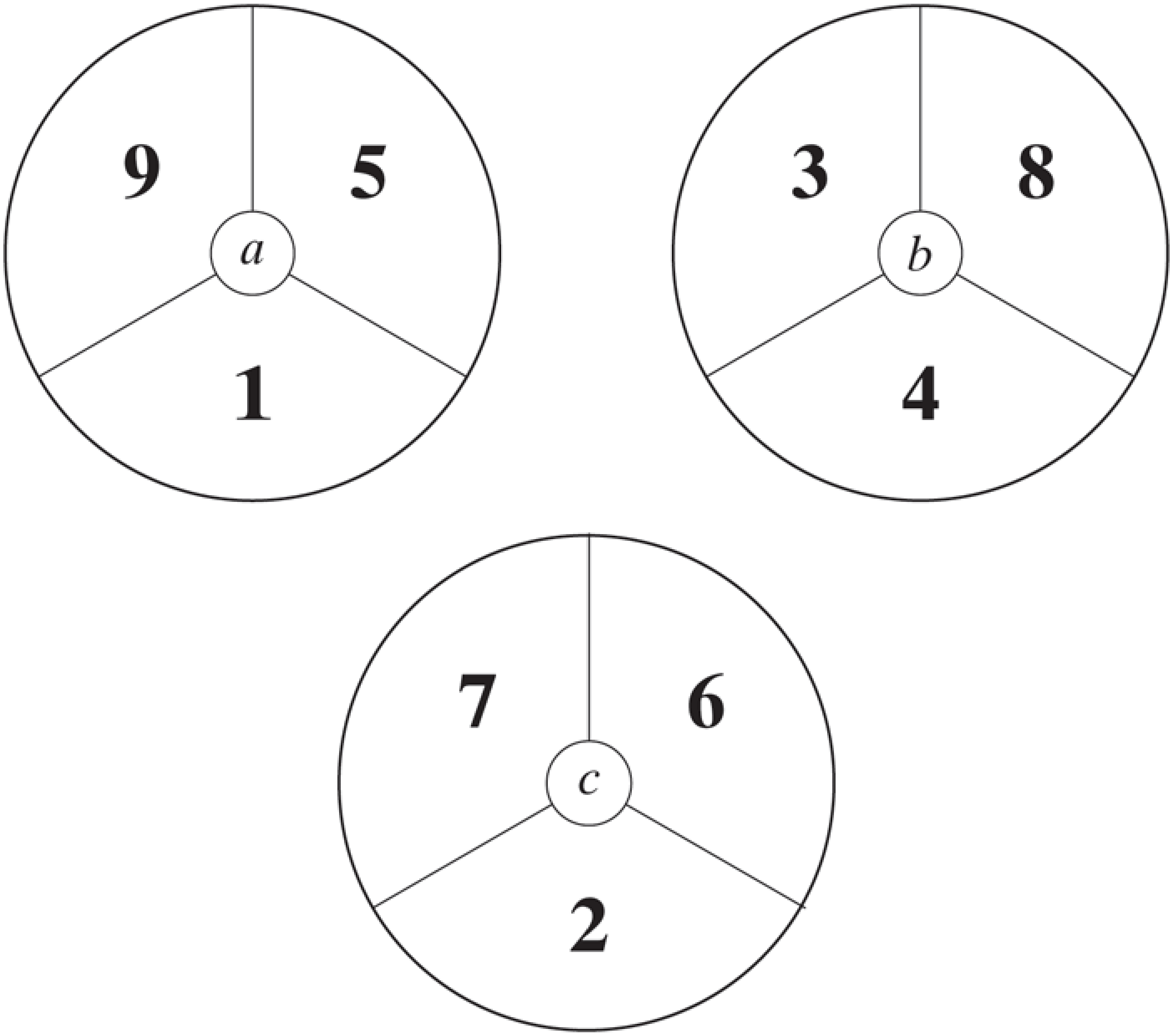
52. A closet contains 10 pairs of shoes. If 8 shoes are randomly selected, what is the probability that there will be

a. no complete pair?
b. exactly 1 complete pair?

53. If 8 people, consisting of 4 couples, are randomly arranged in a row, find the probability that no person is next to their partner.
54. Compute the probability that a bridge hand is void in at least one suit. Note that the answer is not (41)(3913)(5213)

(Why not?)
Hint: Use **Proposition 4.4**.
55. Compute the probability that a hand of 13 cards contains
   a. the ace and king of at least one suit;
   b. all 4 of at least 1 of the 13 denominations.

56. Two players play the following game: Player A chooses one of the three spinners pictured in **Figure 2.6**, and then player B chooses one of the remaining two spinners. Both players then spin their spinner, and the one that lands on the higher number is declared the winner. Assuming that each spinner is equally likely to land in any of its 3 regions, would you rather be player *A* or player B? Explain your answer!
Figure 2.6 Spinners



Theoretical Exercises

Prove the following relations:

1. $EF \subset E \subset E \cup F$.
2. If $E \subset F$, then $F^c \subset E^c$.
3. $F = FE \cup FE^c$ and $E \cup F = E \cup E^c F$.
4. $(\cup_1^\infty E_i)F = \cup_1^\infty E_i F$ and
$(\cap_1^\infty E_i) \cup F = \cap_1^\infty (E_i \cup F)$.

5. For any sequence of events $E_1, E_2, \ldots$, define a new sequence $F_1, F_2, \ldots$ of disjoint events (that is, events such that $F_i F_j = \varnothing$ whenever $i \neq j$) such that for all $n \geq 1$,
$\cup_1^n F_i = \cup_1^n E_i$

6. Let E, F, and G be three events. Find expressions for the events so that, of E, F, and G,
   a. only E occurs;
   b. both E and G, but not F, occur;
   c. at least one of the events occurs;
   d. at least two of the events occur;
   e. all three events occur;
   f. none of the events occurs;
   g. at most one of the events occurs;
   h. at most two of the events occur;
   i. exactly two of the events occur;
   j. at most three of the events occur.

7. Use Venn diagrams
   a. to simplify the expression $(E \cup F)(E \cup F^c)$;
   b. to prove DeMorgan's laws for events E and F. [That is, prove $(E \cup F)^c = E^c F^c$, and $(EF)^c = E^c \cup F^c$.]

8. Let S be a given set. If, for some $k > 0$, $S_1, S_2, \ldots, S_k$ are mutually exclusive nonempty subsets of S such that $\cup_{i=1}^k S_i = S$,, then we call the set $\{S_1, S_2, \ldots, S_k\}$ a *partition* of S. Let T_n denote the number of different partitions of $\{1,2,\ldots,n\}$. Thus, $T_1 = 1$ (the only partition being $S_1 = \{1\}$) and $T_2 = 2$ (the two partitions being $\{\{1,2,\}\}$, $\{\{1\}, \{2\}\}$).
   a. Show, by computing all partitions, that $T_3 = 5, T_4 = 15$.
   b. Show that
   $$T_{n+1} = 1 + \sum_{k=1}^{n} \binom{n}{k} T_k$$
   and use this equation to compute T_{10}.
   Hint: One way of choosing a partition of n+1 items is to call one of the items *special*. Then we obtain different partitions by first choosing $k, k = 0, 1, \ldots, n$, then a subset of size $n-k$ of the nonspecial items, and then any of the T_k partitions of the remaining k nonspecial items. By adding the special item to the subset of size $n-k$, we obtain a partition of all n+1 items.

9. Suppose that an experiment is performed n times. For any event E of the sample space, let $n(E)$ denote the number of times that event E occurs and define $f(E) = n(E)/n$. Show that $f(\cdot)$ satisfies Axioms 1, 2, and 3.
10. Prove that $P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E^c FG) - P(EF^c G) - P(EFG^c) - P(EFG^c) - 2P(EFG)$.
11. If $P(E) = .9$ and $P(F) = .8$, show that $P(EF) \geq .7$. In general, prove *Bonferroni's inequality*, namely, $P(EF) \geq P(E) + P(F) - 1$

12. Show that the probability that exactly one of the events E or F occurs equals $P(E)+P(F)-2P(EF)$.
13. Prove that $P(EF^c)=P(E)-P(EF)$.
14. Prove **Proposition 4.4** by mathematical induction.
15. An urn contains M white and N black balls. If a random sample of size r is chosen, what is the probability that it contains exactly k white balls?
16. Use induction to generalize *Bonferroni's inequality* to n events. That is, show that
$P(E_1E_2\cdots E_n)\geq P(E_1)+\cdots+P(E_n)-(n-1)$

17. Consider the matching problem, **Example 5m**, and define A_N to be the number of ways in which the N men can select their hats so that no man selects his own. Argue that
$A_N=(N-1)(A_{N-1}+A_{N-2})$

This formula, along with the boundary conditions $A_1=0$, $A_2=1$, can then be solved for A_N, and the desired probability of no matches would be $A_N/N!$.
Hint: After the first man selects a hat that is not his own, there remain $N-1$ men to select among a set of $N-1$ hats that does not contain the hat of one of these men. Thus, there is one extra man and one extra hat. Argue that we can get no matches either with the extra man selecting the extra hat or with the extra man not selecting the extra hat.
18. Let f_n denote the number of ways of tossing a coin n times such that successive heads never appear. Argue that
$f_n=f_{n-1}+f_{n-2} \qquad n\geq 2$, where $f_0\equiv 1$, $f_1\equiv 2$

Hint: How many outcomes are there that start with a head, and how many start with a tail? If P_n denotes the probability that successive heads never appear when a coin is tossed n times, find P_n (in terms of f_n) when all possible outcomes of the n tosses are assumed equally likely. Compute P_{10}.
19. An urn contains n red and m blue balls. They are withdrawn one at a time until a total of $r, r\leq n$, red balls have been withdrawn. Find the probability that a total of k balls are withdrawn.
Hint: A total of k balls will be withdrawn if there are $r-1$ red balls in the first $k-1$ withdrawals and the kth withdrawal is a red ball.
20. Consider an experiment whose sample space consists of a countably infinite number of points. Show that not all points can be equally likely. Can all points have a positive probability of occurring?
*21. Consider **Example 5o**, which is concerned with the number of runs of wins obtained when n wins and m losses are randomly permuted. Now consider the total number of runs–that is, win runs plus loss runs–and show that
$P\{2k \text{ runs}\}=2\dfrac{\binom{m-1}{k-1}\binom{n-1}{k-1}}{\binom{m+n}{n}}$ $P\{2k+1 \text{ runs}\} = \dfrac{\binom{m-1}{k-1}\binom{n-1}{k}+\binom{m-1}{k}\binom{n-1}{k-1}}{\binom{m+n}{n}}$

Self-Test Problems and Exercises

1. A cafeteria offers a three-course meal consisting of an entree, a starch, and a dessert. The possible choices are given in the following table:

Course	Choices
Entree	Chicken or roast beef
Starch	Pasta or rice or potatoes
Dessert	Ice cream or Jello or apple pie or a peach

A person is to choose one course from each category.

a. How many outcomes are in the sample space?
b. Let A be the event that ice cream is chosen. How many outcomes are in A?
c. Let B be the event that chicken is chosen. How many outcomes are in B?
d. List all the outcomes in the event *AB*.
e. Let C be the event that rice is chosen. How many outcomes are in C?
f. List all the outcomes in the event *ABC*.

2. A customer visiting the suit department of a certain store will purchase a suit with probability .22, a shirt with probability .30, and a tie with probability .28. The customer will purchase both a suit and a shirt with probability .11, both a suit and a tie with probability .14, and both a shirt and a tie with probability .10. A customer will purchase all 3 items with probability .06. What is the probability that a customer purchases

a. none of these items?
b. exactly 1 of these items?

3. A deck of cards is dealt out. What is the probability that the 14th card dealt is an ace? What is the probability that the first ace occurs on the 14th card?

4. Let A denote the event that the midtown temperature in Los Angeles is 70∘F, and let B denote the event that the midtown temperature in New York is 70 ∘ F . Also, let C denote the event that the maximum of the midtown temperatures in New York and in Los Angeles is 70∘F. If P(A)=.3,P(B)=.4, and P(C)=.2, find the probability that the minimum of the two midtown temperatures is 70∘F.

5. An ordinary deck of 52 cards is shuffled. What is the probability that the top four cards have

a. different denominations?
b. different suits?

6. Urn A contains 3 red and 3 black balls, whereas urn B contains 4 red and 6 black balls. If a ball is randomly selected from each urn, what is the probability that the balls will be the same color?

7. In a state lottery, a player must choose 8 of the numbers from 1 to 40. The lottery commission then performs an experiment that selects 8 of these 40 numbers. Assuming that the choice of the lottery commission is equally likely to be any of the (40 8) combinations, what is the probability that a player has

a. all 8 of the numbers selected by the lottery commission?
b. 7 of the numbers selected by the lottery commission?
c. at least 6 of the numbers selected by the lottery commission?

8. From a group of 3 first-year students, 4 sophomores, 4 juniors, and 3 seniors, a committee of size 4 is randomly selected. Find the probability that the committee will consist of

a. 1 from each class;
b. 2 sophomores and 2 juniors;
c. only sophomores or juniors.

9. For a finite set A, let N(A) denote the number of elements in A.
   a. Show that
      $N(A\cup B)=N(A)+N(B)-N(AB)$

   b. More generally, show that

$$N\left(\cup_{i=1}^{n} A_i\right)=\sum_i N(A_i)-\sum\sum_{i<j} N(A_i A_j)+\cdots+(-1)^{n+1} N(A_1\cdots A_n)$$

10. Consider an experiment that consists of 6 horses, numbered 1 through 6, running a race, and suppose that the sample space consists of the 6! possible orders in which the horses finish. Let A be the event that the number-1 horse is among the top three finishers, and let B be the event that the number-2 horse comes in second. How many outcomes are in the event $A\cup B$?
11. A 5-card hand is dealt from a well-shuffled deck of 52 playing cards. What is the probability that the hand contains at least one card from each of the four suits?
12. A basketball team consists of 6 frontcourt and 4 backcourt players. If players are divided into roommates at random, what is the probability that there will be exactly two roommate pairs made up of a backcourt and a frontcourt player?
13. Suppose that a person chooses a letter at random from R E S E R V E and then chooses one at random from V E R T I C A L. What is the probability that the same letter is chosen?
14. Prove *Boole's inequality*:
$P(\cup_{i=1}^{\infty}A_i)\le\sum_{i=1}^{\infty}P(A_i)$

15. Show that if $P(A_i)=1$ for all $i\ge1$, then $P(\cap_{i=1}^{\infty}A_i)=1$.
16. Let $T_k(n)$ denote the number of partitions of the set $\{1,\ldots,n\}$ into k nonempty subsets, where $1\le k\le n$. (See **Theoretical Exercise 8** for the definition of a partition.) Argue that
$T_k(n)=kT_k(n-1)+T_{k-1}(n-1)$

Hint: In how many partitions is {1} a subset, and in how many is 1 an element of a subset that contains other elements?
17. Five balls are randomly chosen, without replacement, from an urn that contains 5 red, 6 white, and 7 blue balls. Find the probability that at least one ball of each color is chosen.
18. Four red, 8 blue, and 5 green balls are randomly arranged in a line.
   a. What is the probability that the first 5 balls are blue?
   b. What is the probability that none of the first 5 balls is blue?
   c. What is the probability that the final 3 balls are of different colors?
   d. What is the probability that all the red balls are together?

19. Ten cards are randomly chosen from a deck of 52 cards that consists of 13 cards of each of 4 different suits. Each of the selected cards is put in one of 4 piles, depending on the suit of the card.
   a. What is the probability that the largest pile has 4 cards, the next largest has 3, the next largest has 2, and the smallest has 1 card?
   b. What is the probability that two of the piles have 3 cards, one has 4 cards, and one has no cards?

20. Balls are randomly removed from an urn initially containing 20 red and 10 blue balls. What is the probability that all of the red balls are removed before all of the blue ones have been removed?

Chapter 3 Conditional Probability and Independence

Contents

3.1 Introduction

In this chapter, we introduce one of the most important concepts in probability theory, that of conditional probability. The importance of this concept is twofold. In the first place, we are often interested in calculating probabilities when some partial information concerning the result of an experiment is available; in such a situation, the desired probabilities are conditional. Second, even when no partial information is available, conditional probabilities can often be used to compute the desired probabilities more easily.

3.2 Conditional Probabilities

Suppose that we toss 2 dice, and suppose that each of the 36 possible outcomes is equally likely to occur and hence has probability 136. Suppose further that we observe that the first die is a 3. Then, given this information, what is the probability that the sum of the 2 dice equals 8? To calculate this probability, we reason as follows: Given that the initial die is a 3, there can be at most 6 possible outcomes of our experiment, namely, (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), and (3, 6). Since each of these outcomes originally had the same probability of occurring, the outcomes should still have equal probabilities. That is, given that the first die is a 3, the (conditional) probability of each of the outcomes (3, 1), (3, 2), (3, 3), (3, 4), (3, 5), and (3, 6) is 16, whereas the (conditional) probability of the other 30 points in the sample space is 0. Hence, the desired probability will be 16.

If we let E and F denote, respectively, the event that the sum of the dice is 8 and the event that the first die is a 3, then the probability just obtained is called the *conditional probability that E occurs given that F has occurred* and is denoted by

$P(E|F)$

A general formula for P(E|F) that is valid for all events E and F is derived in the same manner: If the event F occurs, then, in order for E to occur, it is necessary that the actual occurrence be a point both in E and in F; that is, it must be in EF. Now, since we know that F has occurred, it follows that F becomes our new, or reduced, sample space; hence, the probability that the event EF occurs will equal the probability of *EF* relative to the probability of F. That is, we have the following definition.

> Definition
>
> If $P(F)>0$, then
>
> $$P(E|F)=P(EF)P(F) \tag{2.1}$$

Example 2a

Joe is 80 percent certain that his missing key is in one of the two pockets of his hanging jacket, being 40 percent certain it is in the left-hand pocket and 40 percent certain it is in the right-hand pocket. If a search of the left-hand pocket does not find the key, what is the conditional probability that it is in the other pocket?

Solution

If we let L be the event that the key is in the left-hand pocket of the jacket, and R be the event that it is in the right-hand pocket, then the desired probability P(R|Lc) can be obtained as follows:

$$P(R|Lc)=P(RLc)P(Lc)=P(R)1-P(L)=2/3$$

If each outcome of a finite sample space S is equally likely, then, conditional on the event that the outcome lies in a subset $F\subset S$, all outcomes in F become equally likely. In such cases, it is often convenient to compute conditional probabilities of the form P(E|F) by using F as the sample space. Indeed, working with this reduced sample space often results in an easier and better understood solution. Our next two examples illustrate this point.

Example 2b

A coin is flipped twice. Assuming that all four points in the sample space $S=\{(h,h),(h,t),(t,h),(t,t)\}$ are equally likely, what is the conditional probability that both flips land on heads, given that (a) the first flip lands on heads? (b) at least one flip lands on heads?

Solution

Let $B=\{(h,h)\}$ be the event that both flips land on heads; let $F=\{(h,h),(h,t)\}$ be the event that the first flip lands on heads; and let $A=\{(h,h),(h,t),(t,h)\}$ be the event that at least one flip lands on heads. The probability for (a) can be obtained from

$$P(B|F)=P(BF)P(F)=P(\{(h,h)\})P(\{(h,h),(h,t)\})=1/42/4=1/2$$

For (b), we have

$$P(B|A)=P(BA)P(A)=P(\{(h,h)\})P(\{(h,h),(h,t),(t,h)\})=1/43/4=1/3$$

Thus, the conditional probabilityhat both flips land on heads given that the first one does is 1/2, whereas the conditional probability that both flips land on heads given that at least one does is only 1/3. Many students initially find this latter result surprising. They reason that given that at least one flip lands on heads, there are two possible results: Either they both land on heads or only one does. Their mistake, however, is in assuming that these two possibilities are equally likely. Initially there are 4 equally likely outcomes. Because the information that at least one flip lands on heads is equivalent to the information that the outcome is not (t,t), we are left with the 3 equally likely outcomes (h,h),(h,t),(t,h), only one of which results in both flips landing on heads.

Example 2c

In the card game bridge, the 52 cards are dealt out equally to 4 players–called East, West, North, and South. If North and South have a total of 8 spades among them, what is the probability that East has 3 of the remaining 5 spades?

Solution

Probably the easiest way to compute the desired probability is to work with the reduced sample space. That is, given that North–South have a total of 8 spades among their 26 cards, there remains a total of 26 cards, exactly 5 of them being spades, to be distributed among the East–West hands. Since each distribution is equally likely, it follows that the conditional probability that East will have exactly 3 spades among his or her 13 cards is

$$(53)(2110)(2613)\approx .339$$

Multiplying both sides of **Equation (2.1)** by P(F), we obtain

$$P(EF)=P(F)P(E|F) \tag{2.2}$$

In words, **Equation (2.2)** states that the probability that both E and F occur is equal to the probability that F occurs multiplied by the conditional probability of E given that F occurred. **Equation (2.2)** is often quite useful in computing the probability of the intersection of events.

Example 2d

Celine is undecided as to whether to take a French course or a chemistry course. She estimates that her probability of receiving an A grade would be 12 in a French course and 23 in a chemistry course. If Celine decides to base her decision on the flip of a fair coin, what is the probability that she gets an A in chemistry?

Solution

Let *C* be the event that Celine takes chemistry and A denote the event that she receives an A in whatever course she takes, then the desired probability is P (*CA*), which is calculated by using **Equation (2.2)** as follows:

$$P(CA)=P(C)P(A|C)=(12)(23)=13$$

Example 2e

Suppose that an urn contains 8 red balls and 4 white balls. We draw 2 balls from the urn without replacement. (a) If we assume that at each draw, each ball in the urn is equally likely to be chosen, what is the probability that both balls drawn are red? (b) Now suppose that the balls have different weights, with each red ball having weight r and each white ball having weight w. Suppose that the probability that a given ball in the urn is the next one selected is its weight divided by the sum of the weights of all balls currently in the urn. Now what is the probability that both balls are red?

Solution

Let R1 and R2 denote, respectively, the events that the first and second balls drawn are red. Now, given that the first ball selected is red, there are 7 remaining red balls and 4 white balls, so P(R2|R1)=711. As P(R1) is clearly 812, the desired probability is

$$P(R_1R_2) = P(R_1)P(R_2|R_1) = \left(\frac{2}{3}\right)\left(\frac{7}{11}\right) = \frac{14}{33}$$

Of course, this probability could have been computed by $P(R_1R_2) = \binom{8}{2} \Big/ \binom{12}{2}$.

For part (b), we again let R_i be the event that the ith ball chosen is red and use

$$P(R_1R_2) = P(R_1)P(R_2|R_1)$$

Now, number the red balls, and let $B_i, i = 1, \ldots, 8$ be the event that the first ball drawn is red ball number i. Then

$$P(R_1) = P\left(\bigcup_{i=1}^{8} B_i\right) = \sum_{i=1}^{8} P(B_i) = 8\,\frac{r}{8r+4w}$$

Moreover, given that the first ball is red, the urn then contains 7 red and 4 white balls. Thus, by an argument similar to the preceding one,

$$P(R_2|R_1) = \frac{7r}{7r+4w}$$

Hence, the probability that both balls are red is

$$P(R_1R_2) = \frac{8r}{8r+4w}\,\frac{7r}{7r+4w}$$

A generalization of **Equation (2.2)**, which provides an expression for the probability of the intersection of an arbitrary number of events, is sometimes referred to as the *multiplication rule*.

> The multiplication rule
>
> $$P(E_1E_2E_3\cdots E_n) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)\cdots P(E_n|E_1\cdots E_{n-1})$$

In words, the multiplication rule states that $P(E_1E_2\cdots E_n)$, the probability that all of the events $E_1, E_2, \ldots, E_n$ occur, is equal to $P(E_1)$, the probability that E_1 occurs, multiplied by $P(E_2|E_1)$, the conditional probability that E_2 occurs given that E_1 has occurred, multiplied by $P(E_3|E_1E_2)$, the conditional probability that E_3 occurs given that both E_1 and E_2 have occurred, and so on.

To prove the multiplication rule, just apply the definition of conditional probability to its right-hand side, giving

$$P(E_1)\frac{P(E_1E_2)}{P(E_1)}\frac{P(E_1E_2E_3)}{P(E_1E_2)}\cdots\frac{P(E_1E_2\cdots E_n)}{P(E_1E_2\cdots E_{n-1})} = P(E_1E_2\cdots E_n)$$

Example 2f

In the match problem stated in **Example 5m** of **Chapter 2**, it was shown that P_N, the probability that there are no matches when N people randomly select from among their own N hats, is given by

$$P_N = \sum_{i=0}^{N} (-1)^i / i!$$

What is the probability that exactly k of the N people have matches?

Solution

Let us fix our attention on a particular set of k people and determine the probability that these k individuals have matches and no one else does. Letting E denote the event that everyone in this set has a match, and letting G be the event that none of the other $N-k$ people have a match, we have

$$P(EG) = P(E)P(G|E)$$

Now, let $F_i, i = 1, \ldots, k$, be the event that the ith member of the set has a match. Then

$$P(E) = P(F_1F_2\cdots F_k) = P(F_1)P(F_2|F_1)P(F_3|F_1F_2)\cdots P(F_k|F_1\cdots F_{k-1}) = \frac{1}{N}\,\frac{1}{N-1}\,\frac{1}{N-2}\cdots\frac{1}{N-k+1} = \frac{(N-k)!}{N!}$$

Given that everyone in the set of k has a match, the other $N-k$ people will be randomly choosing among their own $N-k$ hats, so the probability that none of them has a match is equal to the probability of no matches in a problem having $N-k$ people choosing among their own $N-k$ hats. Therefore,

$$P(G|E) = P_{N-k} = \sum_{i=0}^{N-k} (-1)^i / i!$$

showing that the probability that a specified set of k people have matches and no one else does is

$$P(EG) = \frac{(N-k)!}{N!}\,P_{N-k}$$

Because there will be exactly k matches if the preceding is true for any of the $\binom{N}{k}$ sets of k individuals, the desired probability is

$$P(\text{exactly } k \text{ matches}) = \binom{N}{k}P(EG) = P_{N-K}/K! \approx e^{-1}/k! \quad \text{when } N \text{ is large}$$

Example 2g

An ordinary deck of 52 playing cards is randomly divided into 4 piles of 13 cards each. Compute the probability that each pile has exactly 1 ace.

Solution

Define events $E_i, i = 1, 2, 3, 4$, as follows:

$$\begin{aligned} E_1 &= \{\text{the ace of spades is in any one of the piles}\} \\ E_2 &= \{\text{the ace of spades and the ace of hearts are in different piles}\} \\ E_3 &= \{\text{the aces of spades, hearts, and diamonds are all in different piles}\} \\ E_4 &= \{\text{all 4 aces are in different piles}\} \end{aligned}$$

The desired probability is $P(E_1E_2E_3E_4)$, and by the multiplication rule,

$$P(E_1E_2E_3E_4) = P(E_1)P(E_2|E_1)P(E_3|E_1E_2)P(E_4|E_1E_2E_3)$$

Now,

$$P(E_1) = 1$$

since E_1 is the sample space S. To determine $P(E_2|E_1)$, consider the pile that contains the ace of spades. Because its remaining 12 cards are equally likely to be any 12 of the remaining 51 cards, the probability that the ace of hearts is among them is 12/51, giving that

$$P(E_2|E_1) = 1 - \frac{12}{51} = \frac{39}{51}$$

Also, given that the ace of spades and ace of hearts are in different piles, it follows that the set of the remaining 24 cards of these two piles is equally likely to be any set of 24 of the remaining 50 cards. As the probability that the ace of diamonds is one of these 24 is 24/50, we see that

$$P(E_3|E_1E_2) = 1 - \frac{24}{50} = \frac{26}{50}$$

Because the same logic as used in the preceding yields that

$$P(E_4|E_1E_2E_3) = 1 - \frac{36}{49} = \frac{13}{49}$$

the probability that each pile has exactly 1 ace is

$$P(E_1E_2E_3E_4) = \frac{39 \cdot 26 \cdot 13}{51 \cdot 50 \cdot 49} \approx .105$$

That is, there is approximately a 10.5 percent chance that each pile will contain an ace. (**Problem 13** gives another way of using the multiplication rule to solve this problem.)

Example 2h

Four of the eight teams in the quarterfinal round of the 2016 European Champions League Football (soccer) tournament were the acknowledged strong teams Barcelona, Bayern Munich, Real Madrid, and Paris St-Germain. The pairings in this round are supposed to be totally random, in the sense that all possible pairings are equally likely. Assuming this is so, find the probability that none of the strong teams play each other in this round. (Surprisingly, it seems to be a common occurrence in this tournament that, even though the pairings are supposedly random, the very strong teams are rarely matched against each other in this round.)

Solution

If we number the four strong teams 1 through 4, and then let $W_i, i=1,2,3,4$, be the event that team i plays one of the four weak teams, then the desired probability is $P(W_1W_2W_3W_4)$. By the multiplication rule

$$P(W_1W_2W_3W_4)=P(W_1)P(W_2|W_1)P(W_3|W_1W_2)P(W_4|W_1W_2W_3)=(4/7)(3/5)(2/3)(1)=8/35$$

The preceding follows by first noting that because team 1 is equally likely to be matched with any of the other 7 teams, we have that $P(W_1)=4/7$. Now, given that W_1 occurs, team 2 is equally likely to be matched with any of five teams: namely, teams 3, 4, or any of the three weak teams not matched with team 1. As three of these five teams are weak, we see that $P(W_2|W_1)=3/5$. Similarly, given that events W_1 and W_2 have occurred, team 3 is equally likely to be matched with any from a set of three teams, consisting of team 4 and the remaining two weaker teams not matched with 1 or 2. Hence, $P(W_3|W_1W_2)=2/3$. Finally, given that W_1, W_2, and W_3 all occur, team 4 will be matched with the remaining weak team not matched with any of 1,2,3, giving that $P(W_4|W_1W_2W_3)=1$.

Remarks Our definition of $P(E|F)$ is consistent with the interpretation of probability as being a long-run relative frequency. To see this, suppose that n repetitions of the experiment are to be performed, where n is large. We claim that if we consider only those experiments in which F occurs, then $P(E|F)$ will equal the long-run proportion of them in which E also occurs. To verify this statement, note that since $P(F)$ is the long-run proportion of experiments in which F occurs, it follows that in the n repetitions of the experiment, F will occur approximately $nP(F)$ times. Similarly, in approximately $nP(EF)$ of these experiments, both E and F will occur. Hence, out of the approximately $nP(F)$ experiments in which F occurs, the proportion of them in which E also occurs is approximately equal to

$$\frac{nP(EF)}{nP(F)}=\frac{P(EF)}{P(F)}$$

Because this approximation becomes exact as n becomes larger and larger, we have the appropriate definition of $P(E|F)$.

3.3 Bayes's Formula

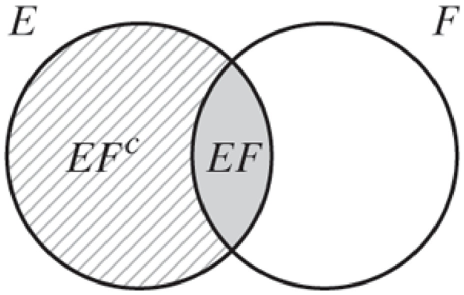
Let E and F be events. We may express E as

$$E=EF\cup EFc$$

for, in order for an outcome to be in E, it must either be in both E and F or be in E but not in F. (See **Figure 3.1**.) As *EF* and EFc are clearly mutually exclusive, we have, by Axiom 3,

$$P(E)=P(EF)+P(EFc)=P(E|F)P(F)+P(E|Fc)P(Fc)=P(E|F)P(F)+P(E|Fc)[1-P(F)] \tag{3.1}$$

Figure 3.1 E=EF∪EFc. EF=Shaded Area; EFc=Striped Area.



Equation (3.1) states that the probability of the event E is a weighted average of the conditional probability of E given that F has occurred and the conditional probability of E given that F has not occurred—each conditional probability being given as much weight as the event on which it is conditioned has of occurring. This is an extremely useful formula, because its use often enables us to determine the probability of an event by first "conditioning" upon whether or not some second event has occurred. That is, there are many instances in which it is difficult to compute the probability of an event directly, but it is straightforward to compute it once we know whether or not some second event has occurred. We illustrate this idea with some examples.

Example 3a

(Part 1)

An insurance company believes that people can be divided into two classes: those who are accident prone and those who are not. The company's statistics show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability .4, whereas this probability decreases to .2 for a person who is not accident prone. If we assume that 30 percent of the population is accident prone, what is the probability that a new policyholder will have an accident within a year of purchasing a policy?

Solution

We shall obtain the desired probability by first conditioning upon whether or not the policyholder is accident prone. Let A1 denote the event that the policyholder will have an accident within a year of purchasing the policy, and let A denote the event that the policyholder is accident prone. Hence, the desired probability is given by

$$P(A1)=P(A1|A)P(A)+P(A1|Ac)P(Ac)=(.4)(.3)+(.2)(.7)=.26$$

Example 3a

(Part 2)

Suppose that a new policyholder has an accident within a year of purchasing a policy. What is the probability that he or she is accident prone?

Solution

The desired probability is

$$P(A|A1)=P(AA1)P(A1)=P(A)P(A1|A)P(A1)=(.3)(.4).26=613$$

Example 3b

Consider the following game played with an ordinary deck of 52 playing cards: The cards are shuffled and then turned over one at a time. At any time, the player can guess that the next card to be turned over will be the ace of spades; if it is, then the player wins. In addition, the player is said to win if the ace of spades has not yet appeared when only one card remains and no guess has yet been made. What is a good strategy? What is a bad strategy?

Solution

Every strategy has probability 1/52 of winning! To show this, we will use induction to prove the stronger result that for an n card deck, one of whose cards is the ace of spades, the probability of winning is 1/n, no matter what strategy is employed. Since this is clearly true for n=1, assume it to be true for an n−1 card deck, and now consider an n card deck. Fix any strategy, and let p denote the probability that the strategy guesses that the first card is the ace of spades. Given that it does, the player's probability of winning is 1/n. If, however, the strategy does not guess that the first card is the ace of spades, then the probability that the player wins is the probability that the first card is not the ace of spades, namely, (n−1)/n, multiplied by the conditional probability of winning given that the first card is not the ace of spades. But this latter conditional probability is equal to the probability of winning when using an n−1 card deck containing a single ace of spades; it is thus, by the induction hypothesis, 1/(n−1). Hence, given that the strategy does not guess the first card, the probability of winning is

$$n-1n1n-1=1n$$

Thus, letting G be the event that the first card is guessed, we obtain

$$P\{win\}=P\{win|G\}P(G)+P\{win|Gc\}(1-P(G))=1np+1n(1-p)=1n$$

Example 3c

In answering a question on a multiple-choice test, a student either knows the answer or guesses. Let p be the probability that the student knows the answer and 1−p be the probability that the student guesses. Assume that a student who guesses at the answer will be correct with probability 1/m, where m is the number of multiple-choice alternatives. What is the conditional probability that a student knew the answer to a question given that he or she answered it correctly?

Solution

Let C and K denote, respectively, the events that the student answers the question correctly and the event that he or she actually knows the answer. Now,

$$P(K|C)=P(KC)P(C)=P(C|K)P(K)P(C|K)P(K)+P(C|Kc)P(Kc)=pp+(1/m)(1-p)=mp1+(m-1)p$$

For example, if m=5,p=12, then the probability that the student knew the answer to a question he or she answered correctly is 56.

Example 3d

A laboratory blood test is 95 percent effective in detecting a certain disease when it is, in fact, present. However, the test also yields a "false positive" result for 1 percent of the healthy persons tested. (That is, if a healthy person is tested, then, with probability .01, the test result will imply that he or she has the disease.) If .5 percent of the population actually has the disease, what is the probability that a person has the disease given that the test result is positive?

Solution

Let D be the event that the person tested has the disease and E the event that the test result is positive. Then the desired probability is

$$P(D|E)=P(DE)P(E)=P(E|D)P(D)P(E|D)P(D)+P(E|Dc)P(Dc)=(.95)(.005)(.95)(.005)+(.01)(.995)=95294\approx.323$$

Thus, only 32 percent of those persons whose test results are positive actually have the disease. Many students are often surprised at this result (they expect the percentage to be much higher, since the blood test seems to be a good one), so it is probably worthwhile to present a second argument that, although less rigorous than the preceding one, is probably more revealing. We now do so.

Since .5 percent of the population actually has the disease, it follows that, on the average, 1 person out of every 200 tested will have it. The test will correctly confirm that this person has the disease with probability .95. Thus, on the average, out of every 200 persons tested, the test will correctly confirm that .95 person has the disease. On the other hand, out of the (on the average) 199 healthy people, the test will incorrectly state that (199)(.01) of these people have the disease. Hence, for every .95 diseased persons that the test correctly states is ill, there are (on the average) (199)(.01) healthy persons who the test incorrectly states are ill. Thus, the proportion of time that the test result is correct when it states that a person is ill is

$$.95.95+(199)(.01)=95294\approx.323$$

Equation (3.1) is also useful when one has to reassess one's personal probabilities in the light of additional information. For instance, consider the examples that follow.

Example 3e

Consider a medical practitioner pondering the following dilemma: "If I'm at least 80 percent certain that my patient has this disease, then I always recommend surgery, whereas if I'm not quite as certain, then I recommend additional tests that are expensive and sometimes painful. Now, initially I was only 60 percent certain that Jones had the disease, so I ordered the series A test, which always gives a positive result when the patient has the disease and almost never does when he is healthy. The test result was positive, and I was all set to recommend surgery when Jones informed me, for the first time, that he was diabetic. This information complicates matters because, although it doesn't change my original 60 percent estimate of his chances of having the disease in question, it does affect the interpretation of the results of the A test. This is so because the A test, while never yielding a positive result when the patient is healthy, does unfortunately yield a positive result 30 percent of the time in the case of *diabetic* patients who are not suffering from the disease. Now what do I do? More tests or immediate surgery?"

Solution

In order to decide whether or not to recommend surgery, the doctor should first compute her updated probability that Jones has the disease given that the A test result was positive. Let D denote the event that Jones has the disease and E the event that the A test result is positive. The desired conditional probability is then

$$P(D|E)=\frac{P(DE)}{P(E)}=\frac{P(D)P(E|D)}{P(E|D)P(D)+P(E|D^c)P(D^c)}=\frac{(.6)1}{1(.6)+(.3)(.4)}=.833$$

Note that we have computed the probability of a positive test result by conditioning on whether or not Jones has the disease and then using the fact that because Jones is a diabetic, his conditional probability of a positive result given that he does not have the disease, $P(E|D^c)$, equals .3. Hence, as the doctor should now be more than 80 percent certain that Jones has the disease, she should recommend surgery.

Example 3f

At a certain stage of a criminal investigation, the inspector in charge is 60 percent convinced of the guilt of a certain suspect. Suppose, however, that a *new* piece of evidence which shows that the criminal has a certain characteristic (such as left-handedness, baldness, or brown hair) is uncovered. If 20 percent of the population possesses this characteristic, how certain of the guilt of the suspect should the inspector now be if it turns out that the suspect has the characteristic?

Solution

Letting G denote the event that the suspect is guilty and C the event that he possesses the characteristic of the criminal, we have

$$P(G|C)=\frac{P(GC)}{P(C)}=\frac{P(C|G)P(G)}{P(C|G)P(G)+P(C|G^c)P(G^c)}=\frac{1(.6)}{1(.6)+(.2)(.4)}\approx .882$$

where we have supposed that the probability of the suspect having the characteristic if he is, in fact, innocent is equal to .2, the proportion of the population possessing the characteristic.

Example 3g

In the world bridge championships held in Buenos Aires in May 1965, the famous British bridge partnership of Terrence Reese and Boris Schapiro was accused of cheating by using a system of finger signals that could indicate the number of hearts held by the players. Reese and Schapiro denied the accusation, and eventually a hearing was held by the British bridge league. The hearing was in the form of a legal proceeding with prosecution and defense teams, both having the power to call and cross-examine witnesses. During the course of the proceeding, the prosecutor examined specific hands played by Reese and Schapiro and claimed that their playing these hands was consistent with the hypothesis that they were guilty of having illicit knowledge of the heart suit. At this point, the defense attorney pointed out that their play of these hands was also perfectly consistent with their standard line of play. However, the prosecution then argued that as long as their play was consistent with the hypothesis of guilt, it must be counted as evidence toward that hypothesis. What do you think of the reasoning of the prosecution?

Solution

The problem is basically one of determining how the introduction of new evidence (in this example, the playing of the hands) affects the probability of a particular hypothesis. If we let H denote a particular hypothesis (such as the hypothesis that Reese and Schapiro are guilty) and E the new evidence, then

$$P(H|E)=\frac{P(HE)}{P(E)}=\frac{P(E|H)P(H)}{P(E|H)P(H)+P(E|H^c)[1-P(H)]} \tag{3.2}$$

where *P(H)* is our evaluation of the likelihood of the hypothesis before the introduction of the new evidence. The new evidence will be in support of the hypothesis whenever it makes the hypothesis more likely—that is, whenever $P(H|E)\geq P(H)$. From **Equation (3.2)**, this will be the case whenever

$$P(E|H)\geq P(E|H)P(H)+P(E|H^c)[1-P(H)]$$

or, equivalently, whenever

$$P(E|H)\geq P(E|H^c)$$

In other words, any new evidence can be considered to be in support of a particular hypothesis only if its occurrence is more likely when the hypothesis is true than when it is false. In fact, the new probability of the hypothesis depends on its initial probability and the ratio of these conditional probabilities, since, from **Equation (3.2)**,

$$P(H|E) = \frac{P(H)}{P(H) + [1 - P(H)]\frac{P(E|H^c)}{P(E|H)}}$$

Hence, in the problem under consideration, the play of the cards can be considered to support the hypothesis of guilt only if such play would have been more likely if the partnership were cheating than if it were not. As the prosecutor never made this claim, his assertion that the evidence is in support of the guilt hypothesis is invalid.

Example 3h

Twins can be either identical or fraternal. Identical, also called monozygotic, twins form when a single fertilized egg splits into two genetically identical parts. Consequently, identical twins always have the same set of genes. Fraternal, also called dizygotic, twins develop when two eggs are fertilized and implant in the uterus. The genetic connection of fraternal twins is no more or less the same as siblings born at separate times. A Los Angeles County, California, scientist wishing to know the current fraction of twin pairs born in the county that are identical twins has assigned a county statistician to study this issue. The statistician initially requested each hospital in the county to record all twin births, indicating whether or not the resulting twins were identical. The hospitals, however, told her that to determine whether newborn twins were identical was not a simple task, as it involved the permission of the twins' parents to perform complicated and expensive DNA studies that the hospitals could not afford. After some deliberation, the statistician just asked the hospitals for data listing all twin births along with an indication as to whether the twins were of the same sex. When such data indicated that approximately 64 percent of twin births were same-sexed, the statistician declared that approximately 28 percent of all twins were identical. How did she come to this conclusion?

Solution

The statistician reasoned that identical twins are always of the same sex, whereas fraternal twins, having the same relationship to each other as any pair of siblings, will have probability 1/2 of being of the same sex. Letting I be the event that a pair of twins is identical, and SS be the event that a pair of twins is of the same sex, she computed the probability $P(SS)$ by conditioning on whether the twin pair was identical. This gave

$$P(SS) = P(SS|I)P(I) + P(SS|I^c)P(I^c)$$

or

$$P(SS) = 1 \times P(I) + \frac{1}{2} \times [1 - P(I)] = \frac{1}{2} + \frac{1}{2}P(I)$$

which, using that $P(SS) \approx .64$ yielded the result

$$P(I) \approx .28$$

The change in the probability of a hypothesis when new evidence is introduced can be expressed compactly in terms of the change in the *odds* of that hypothesis, where the concept of odds is defined as follows.

Definition

The odds of an event A are defined by

$$\frac{P(A)}{P(A^c)} = \frac{P(A)}{1 - P(A)}$$

That is, the odds of an event A tell how much more likely it is that the event A occurs than it is that it does not occur. For instance, if $P(A) = \frac{2}{3}$, then $P(A) = 2P(A^c)$, so the odds are 2. If the odds are equal to α, then it is common to say that the odds are "α to 1" in favor of the hypothesis.

Consider now a hypothesis H that is true with probability $P(H)$, and suppose that new evidence E is introduced. Then, the conditional probabilities, given the evidence E, that H is true and that H is not true are respectively given by

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)} \qquad P(H^c|E) = \frac{P(E|H^c)P(H^c)}{P(E)}$$

Therefore, the new odds after the evidence E has been introduced are

$$\frac{P(H|E)}{P(H^c|E)} = \frac{P(H)}{P(H^c)}\frac{P(E|H)}{P(E|H^c)} \tag{3.3}$$

That is, the new value of the odds of H is the old value multiplied by the ratio of the conditional probability of the new evidence given that H is true to the conditional probability given that H is not true. Thus, **Equation (3.3)** verifies the result of **Example 3f**, since the odds, and thus the probability of H, increase whenever the new evidence is more likely when H is true than when it is false. Similarly, the odds decrease whenever the new evidence is more likely when H is false than when it is true.

Example 3i

An urn contains two type A coins and one type B coin. When a type A coin is flipped, it comes up heads with probability 1/4, whereas when a type B coin is flipped, it comes up heads with probability 3/4. A coin is randomly chosen from the urn and flipped. Given that the flip landed on heads, what is the probability that it was a type A coin?