Probability and Statistics

FOURTH EDITION

DeGROOT | SCHERVISH

Probability and Statistics

Fourth Edition

This page intentionally left blank

Probability and Statistics

Fourth Edition

MORRIS H. DEGROOT
Carnegie Mellon University

MARK J. SCHERVISH
Carnegie Mellon University

Addison-Wesley
Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montréal Toronto
Delhi Mexico City São Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo

Editor in Chief: Deirdre Lynch
Acquisitions Editor: Christopher Cummings
Associate Content Editors: Leah Goldberg, Dana Jones Bettez
Associate Editor: Christina Lepre
Senior Managing Editor: Karen Wernholm
Production Project Manager: Patty Bergin
Cover Designer: Heather Scott
Design Manager: Andrea Nix
Senior Marketing Manager: Alex Gay
Marketing Assistant: Kathleen DeChavez
Senior Author Support/Technology Specialist: Joe Vetere
Rights and Permissions Advisor: Michael Joyce
Manufacturing Manager: Carol Melville
Project Management, Composition: Windfall Software, using ZzTEX
Cover Photo: Shutterstock/© Marilyn Volan

Library of Congress Cataloging-in-Publication Data

DeGroot, Morris H., 1931–1989.
Probability and statistics / Morris H. DeGroot, Mark J. Schervish.—4th ed.
p. cm.
ISBN 978-0-321-50046-5
1. Probabilities—Textbooks. 2. Mathematical statistics—Textbooks.
I. Schervish, Mark J. II. Title.
QA273.D35 2012
519.2—dc22
2010001486

1 2 3 4 5 6 7 8 9 10—EB—14 13 12 11 10

Addison-Wesley
is an imprint of
PEARSON

www.pearsonhighered.com

ISBN 10: 0-321-50046-6
ISBN 13: 978-0-321-50046-5

To the memory of Morrie DeGroot.

MJS

This page intentionally left blank

Contents

Preface

Changes to the Fourth Edition

- I have reorganized many main results that were included in the body of the text by labeling them as theorems in order to facilitate students in finding and referencing these results.
- I have pulled the important defintions and assumptions out of the body of the text and labeled them as such so that they stand out better.
- When a new topic is introduced, I introduce it with a motivating example before delving into the mathematical formalities. Then I return to the example to illustrate the newly introduced material.
- I moved the material on the law of large numbers and the central limit theorem to a new Chapter 6. It seemed more natural to deal with the main large-sample results together.
- I moved the section on Markov chains into Chapter 3. Every time I cover this material with my own students, I stumble over not being able to refer to random variables, distributions, and conditional distributions. I have actually postponed this material until after introducing distributions, and then gone back to cover Markov chains. I feel that the time has come to place it in a more natural location. I also added some material on stationary distributions of Markov chains.
- I have moved the lengthy proofs of several theorems to the ends of their respective sections in order to improve the flow of the presentation of ideas.
- I rewrote Section 7.1 to make the introduction to inference clearer.
- I rewrote Section 9.1 as a more complete introduction to hypothesis testing, including likelihood ratio tests. For instructors not interested in the more mathematical theory of hypothesis testing, it should now be easier to skip from Section 9.1 directly to Section 9.5.

Some other changes that readers will notice:

- I have replaced the notation in which the intersection of two sets A and B had been represented AB with the more popular $A \cap B$. The old notation, although mathematically sound, seemed a bit arcane for a text at this level.
- I added the statements of Stirling's formula and Jensen's inequality.
- I moved the law of total probability and the discussion of partitions of a sample space from Section 2.3 to Section 2.1.
- I define the cumulative distribution function (c.d.f.) as the prefered name of what used to be called only the distribution function (d.f.).
- I added some discussion of histograms in Chapters 3 and 6.
- I rearranged the topics in Sections 3.8 and 3.9 so that simple functions of random variables appear first and the general formulations appear at the end to make it easier for instructors who want to avoid some of the more mathematically challenging parts.
- I emphasized the closeness of a hypergeometric distribution with a large number of available items to a binomial distribution.

- I gave a brief introduction to Chernoff bounds. These are becoming increasingly important in computer science, and their derivation requires only material that is already in the text.
- I changed the definition of confidence interval to refer to the random interval rather than the observed interval. This makes statements less cumbersome, and it corresponds to more modern usage.
- I added a brief discussion of the method of moments in Section 7.6.
- I added brief introductions to Newton's method and the EM algorithm in Chapter 7.
- I introduced the concept of pivotal quantity to facilitate construction of confidence intervals in general.
- I added the statement of the large-sample distribution of the likelihood ratio test statistic. I then used this as an alternative way to test the null hypothesis that two normal means are equal when it is not assumed that the variances are equal.
- I moved the Bonferroni inequality into the main text (Chapter 1) and later (Chapter 11) used it as a way to construct simultaneous tests and confidence intervals.

How to Use This Book

The text is somewhat long for complete coverage in a one-year course at the undergraduate level and is designed so that instructors can make choices about which topics are most important to cover and which can be left for more in-depth study. As an example, many instructors wish to deemphasize the classical counting arguments that are detailed in Sections 1.7–1.9. An instructor who only wants enough information to be able to cover the binomial and/or multinomial distributions can safely discuss only the definitions and theorems on permutations, combinations, and possibly multinomial coefficients. Just make sure that the students realize what these values count, otherwise the associated distributions will make no sense. The various examples in these sections are helpful, but not necessary, for understanding the important distributions. Another example is Section 3.9 on functions of two or more random variables. The use of Jacobians for general multivariate transformations might be more mathematics than the instructors of some undergraduate courses are willing to cover. The entire section could be skipped without causing problems later in the course, but some of the more straightforward cases early in the section (such as convolution) might be worth introducing. The material in Sections 9.2–9.4 on optimal tests in one-parameter families is pretty mathematics, but it is of interest primarily to graduate students who require a very deep understanding of hypothesis testing theory. The rest of Chapter 9 covers everything that an undergraduate course really needs.

In addition to the text, the publisher has an *Instructor's Solutions Manual,* available for download from the Instructor Resource Center at www.pearsonhighered.com/irc, which includes some specific advice about many of the sections of the text. I have taught a year-long probability and statistics sequence from earlier editions of this text for a group of mathematically well-trained juniors and seniors. In the first semester, I covered what was in the earlier edition but is now in the first five chapters (including the material on Markov chains) and parts of Chapter 6. In the second semester, I covered the rest of the new Chapter 6, Chapters 7–9, Sections 11.1–11.5, and Chapter 12. I have also taught a one-semester probability and random processes course for engineers and computer scientists. I covered what was in the old edition and is now in Chapters 1–6 and 12, including Markov chains, but not Jacobians. This latter course did not emphasize mathematical derivation to the same extent as the course for mathematics students.

A number of sections are designated with an asterisk (*). This indicates that later sections do not rely materially on the material in that section. This designation is not intended to suggest that instructors skip these sections. Skipping one of these sections will not cause the students to miss definitions or results that they will need later. The sections are 2.4, 3.10, 4.8, 7.7, 7.8, 7.9, 8.6, 8.8, 9.2, 9.3, 9.4, 9.8, 9.9, 10.6, 10.7, 10.8, 11.4, 11.7, 11.8, and 12.5. Aside from cross-references between sections within this list, occasional material from elsewhere in the text does refer back to some of the sections in this list. Each of the dependencies is quite minor, however. Most of the dependencies involve references from Chapter 12 back to one of the optional sections. The reason for this is that the optional sections address some of the more difficult material, and simulation is most useful for solving those difficult problems that cannot be solved analytically. Except for passing references that help put material into context, the dependencies are as follows:

- The sample distribution function (Section 10.6) is reintroduced during the discussion of the bootstrap in Section 12.6. The sample distribution function is also a useful tool for displaying simulation results. It could be introduced as early as Example 12.3.7 simply by covering the first subsection of Section 10.6.
- The material on robust estimation (Section 10.7) is revisited in some simulation exercises in Section 12.2 (Exercises 4, 5, 7, and 8).
- Example 12.3.4 makes reference to the material on two-way analysis of variance (Sections 11.7 and 11.8).

Supplements

The text is accompanied by the following supplementary material:

- **Instructor's Solutions Manual** contains fully worked solutions to all exercises in the text. Available for download from the Instructor Resource Center at www.pearsonhighered.com/irc.
- **Student Solutions Manual** contains fully worked solutions to all odd exercises in the text. Available for purchase from MyPearsonStore at www.mypearsonstore .com. (ISBN-13: 978-0-321-71598-2; ISBN-10: 0-321-71598-5)

Acknowledgments

There are many people that I want to thank for their help and encouragement during this revision. First and foremost, I want to thank Marilyn DeGroot and Morrie's children for giving me the chance to revise Morrie's masterpiece.

I am indebted to the many readers, reviewers, colleagues, staff, and people at Addison-Wesley whose help and comments have strengthened this edition. The reviewers were:

Andre Adler, Illinois Institute of Technology; E. N. Barron, Loyola University; Brian Blank, Washington University in St. Louis; Indranil Chakraborty, University of Oklahoma; Daniel Chambers, Boston College; Rita Chattopadhyay, Eastern Michigan University; Stephen A. Chiappari, Santa Clara University; Sheng-Kai Chang, Wayne State University; Justin Corvino, Lafayette College; Michael Evans, University of

Toronto; Doug Frank, Indiana University of Pennsylvania; Anda Gadidov, Kennesaw State University; Lyn Geisler, Randolph–Macon College; Prem Goel, Ohio State University; Susan Herring, Sonoma State University; Pawel Hitczenko, Drexel University; Lifang Hsu, Le Moyne College; Wei-Min Huang, Lehigh University; Syed Kirmani, University of Northern Iowa; Michael Lavine, Duke University; Rich Levine, San Diego State University; John Liukkonen, Tulane University; Sergio Loch, Grand View College; Rosa Matzkin, Northwestern University; Terry McConnell, Syracuse University; Hans-Georg Mueller, University of California–Davis; Robert Myers, Bethel College; Mario Peruggia, The Ohio State University; Stefan Ralescu, Queens University; Krishnamurthi Ravishankar, SUNY New Paltz; Diane Saphire, Trinity University; Steven Sepanski, Saginaw Valley State University; Hen-Siong Tan, Pennsylvania University; Kanapathi Thiru, University of Alaska; Kenneth Troske, Johns Hopkins University; John Van Ness, University of Texas at Dallas; Yehuda Vardi, Rutgers University; Yelena Vaynberg, Wayne State University; Joseph Verducci, Ohio State University; Mahbobeh Vezveai, Kent State University; Brani Vidakovic, Duke University; Karin Vorwerk, Westfield State College; Bette Warren, Eastern Michigan University; Calvin L. Williams, Clemson University; Lori Wolff, University of Mississippi.

The person who checked the accuracy of the book was Anda Gadidov, Kennesaw State University. I would also like to thank my colleagues at Carnegie Mellon University, especially Anthony Brockwell, Joel Greenhouse, John Lehoczky, Heidi Sestrich, and Valerie Ventura.

The people at Addison-Wesley and other organizations that helped produce the book were Paul Anagnostopoulos, Patty Bergin, Dana Jones Bettez, Chris Cummings, Kathleen DeChavez, Alex Gay, Leah Goldberg, Karen Hartpence, and Christina Lepre.

If I left anyone out, it was unintentional, and I apologize. Errors inevitably arise in any project like this (meaning a project in which I am involved). For this reason, I shall post information about the book, including a list of corrections, on my Web page, http://www.stat.cmu.edu/~mark/, as soon as the book is published. Readers are encouraged to send me any errors that they discover.

Mark J. Schervish
October 2010

Chapter 1

Introduction to Probability

1.1 The History of Probability

The use of probability to measure uncertainty and variability dates back hundreds of years. Probability has found application in areas as diverse as medicine, gambling, weather forecasting, and the law.

The concepts of chance and uncertainty are as old as civilization itself. People have always had to cope with uncertainty about the weather, their food supply, and other aspects of their environment, and have striven to reduce this uncertainty and its effects. Even the idea of gambling has a long history. By about the year 3500 B.C., games of chance played with bone objects that could be considered precursors of dice were apparently highly developed in Egypt and elsewhere. Cubical dice with markings virtually identical to those on modern dice have been found in Egyptian tombs dating from 2000 B.C. We know that gambling with dice has been popular ever since that time and played an important part in the early development of probability theory.

It is generally believed that the mathematical theory of probability was started by the French mathematicians Blaise Pascal (1623–1662) and Pierre Fermat (1601–1665) when they succeeded in deriving exact probabilities for certain gambling problems involving dice. Some of the problems that they solved had been outstanding for about 300 years. However, numerical probabilities of various dice combinations had been calculated previously by Girolamo Cardano (1501–1576) and Galileo Galilei (1564–1642).

The theory of probability has been developed steadily since the seventeenth century and has been widely applied in diverse fields of study. Today, probability theory is an important tool in most areas of engineering, science, and management. Many research workers are actively engaged in the discovery and establishment of new applications of probability in fields such as medicine, meteorology, photography from satellites, marketing, earthquake prediction, human behavior, the design of computer systems, finance, genetics, and law. In many legal proceedings involving antitrust violations or employment discrimination, both sides will present probability and statistical calculations to help support their cases.

References

The ancient history of gambling and the origins of the mathematical theory of probability are discussed by David (1988), Ore (1960), Stigler (1986), and Todhunter (1865).

Some introductory books on probability theory, which discuss many of the same topics that will be studied in this book, are Feller (1968); Hoel, Port, and Stone (1971); Meyer (1970); and Olkin, Gleser, and Derman (1980). Other introductory books, which discuss both probability theory and statistics at about the same level as they will be discussed in this book, are Brunk (1975); Devore (1999); Fraser (1976); Hogg and Tanis (1997); Kempthorne and Folks (1971); Larsen and Marx (2001); Larson (1974); Lindgren (1976); Miller and Miller (1999); Mood, Graybill, and Boes (1974); Rice (1995); and Wackerly, Mendenhall, and Schaeffer (2008).

1.2 Interpretations of Probability

This section describes three common operational interpretations of probability. Although the interpretations may seem incompatible, it is fortunate that the calculus of probability (the subject matter of the first six chapters of this book) applies equally well no matter which interpretation one prefers.

In addition to the many formal applications of probability theory, the concept of probability enters our everyday life and conversation. We often hear and use such expressions as "It probably will rain tomorrow afternoon," "It is very likely that the plane will arrive late," or "The chances are good that he will be able to join us for dinner this evening." Each of these expressions is based on the concept of the probability, or the likelihood, that some specific event will occur.

Despite the fact that the concept of probability is such a common and natural part of our experience, no single scientific interpretation of the term *probability* is accepted by all statisticians, philosophers, and other authorities. Through the years, each interpretation of probability that has been proposed by some authorities has been criticized by others. Indeed, the true meaning of probability is still a highly controversial subject and is involved in many current philosophical discussions pertaining to the foundations of statistics. Three different interpretations of probability will be described here. Each of these interpretations can be very useful in applying probability theory to practical problems.

The Frequency Interpretation of Probability

In many problems, the probability that some specific outcome of a process will be obtained can be interpreted to mean the *relative frequency* with which that outcome would be obtained if the process were repeated a large number of times under similar conditions. For example, the probability of obtaining a head when a coin is tossed is considered to be 1/2 because the relative frequency of heads should be approximately 1/2 when the coin is tossed a large number of times under similar conditions. In other words, it is assumed that the proportion of tosses on which a head is obtained would be approximately 1/2.

Of course, the conditions mentioned in this example are too vague to serve as the basis for a scientific definition of probability. First, a "large number" of tosses of the coin is specified, but there is no definite indication of an actual number that would be considered large enough. Second, it is stated that the coin should be tossed each time "under similar conditions," but these conditions are not described precisely. The conditions under which the coin is tossed must not be completely identical for each toss because the outcomes would then be the same, and there would be either all heads or all tails. In fact, a skilled person can toss a coin into the air repeatedly and catch it in such a way that a head is obtained on almost every toss. Hence, the tosses must not be completely controlled but must have some "random" features.

Furthermore, it is stated that the relative frequency of heads should be "approximately 1/2," but no limit is specified for the permissible variation from 1/2. If a coin were tossed 1,000,000 times, we would not expect to obtain exactly 500,000 heads. Indeed, we would be extremely surprised if we obtained exactly 500,000 heads. On the other hand, neither would we expect the number of heads to be very far from 500,000. It would be desirable to be able to make a precise statement of the likelihoods of the different possible numbers of heads, but these likelihoods would of necessity depend on the very concept of probability that we are trying to define.

Another shortcoming of the frequency interpretation of probability is that it applies only to a problem in which there can be, at least in principle, a large number of similar repetitions of a certain process. Many important problems are not of this type. For example, the frequency interpretation of probability cannot be applied directly to the probability that a specific acquaintance will get married within the next two years or to the probability that a particular medical research project will lead to the development of a new treatment for a certain disease within a specified period of time.

The Classical Interpretation of Probability

The classical interpretation of probability is based on the concept of *equally likely outcomes*. For example, when a coin is tossed, there are two possible outcomes: a head or a tail. If it may be assumed that these outcomes are equally likely to occur, then they must have the same probability. Since the sum of the probabilities must be 1, both the probability of a head and the probability of a tail must be 1/2. More generally, if the outcome of some process must be one of n different outcomes, and if these n outcomes are equally likely to occur, then the probability of each outcome is $1/n$.

Two basic difficulties arise when an attempt is made to develop a formal definition of probability from the classical interpretation. First, the concept of equally likely outcomes is essentially based on the concept of probability that we are trying to define. The statement that two possible outcomes are equally likely to occur is the same as the statement that two outcomes have the same probability. Second, no systematic method is given for assigning probabilities to outcomes that are not assumed to be equally likely. When a coin is tossed, or a well-balanced die is rolled, or a card is chosen from a well-shuffled deck of cards, the different possible outcomes can usually be regarded as equally likely because of the nature of the process. However, when the problem is to guess whether an acquaintance will get married or whether a research project will be successful, the possible outcomes would not typically be considered to be equally likely, and a different method is needed for assigning probabilities to these outcomes.

The Subjective Interpretation of Probability

According to the subjective, or personal, interpretation of probability, the probability that a person assigns to a possible outcome of some process represents her own judgment of the likelihood that the outcome will be obtained. This judgment will be based on each person's beliefs and information about the process. Another person, who may have different beliefs or different information, may assign a different probability to the same outcome. For this reason, it is appropriate to speak of a certain person's *subjective probability* of an outcome, rather than to speak of the *true probability* of that outcome.

As an illustration of this interpretation, suppose that a coin is to be tossed once. A person with no special information about the coin or the way in which it is tossed might regard a head and a tail to be equally likely outcomes. That person would then assign a subjective probability of 1/2 to the possibility of obtaining a head. The person who is actually tossing the coin, however, might feel that a head is much more likely to be obtained than a tail. In order that people in general may be able to assign subjective probabilities to the outcomes, they must express the strength of their belief in numerical terms. Suppose, for example, that they regard the likelihood of obtaining a head to be the same as the likelihood of obtaining a red card when one card is chosen from a well-shuffled deck containing four red cards and one black card. Because those people would assign a probability of 4/5 to the possibility of obtaining a red card, they should also assign a probability of 4/5 to the possibility of obtaining a head when the coin is tossed.

This subjective interpretation of probability can be formalized. In general, if people's judgments of the relative likelihoods of various combinations of outcomes satisfy certain conditions of consistency, then it can be shown that their subjective probabilities of the different possible events can be uniquely determined. However, there are two difficulties with the subjective interpretation. First, the requirement that a person's judgments of the relative likelihoods of an infinite number of events be completely consistent and free from contradictions does not seem to be humanly attainable, unless a person is simply willing to adopt a collection of judgments known to be consistent. Second, the subjective interpretation provides no "objective" basis for two or more scientists working together to reach a common evaluation of the state of knowledge in some scientific area of common interest.

On the other hand, recognition of the subjective interpretation of probability has the salutary effect of emphasizing some of the subjective aspects of science. A particular scientist's evaluation of the probability of some uncertain outcome must ultimately be that person's own evaluation based on all the evidence available. This evaluation may well be based in part on the frequency interpretation of probability, since the scientist may take into account the relative frequency of occurrence of this outcome or similar outcomes in the past. It may also be based in part on the classical interpretation of probability, since the scientist may take into account the total number of possible outcomes that are considered equally likely to occur. Nevertheless, the final assignment of numerical probabilities is the responsibility of the scientist herself.

The subjective nature of science is also revealed in the actual problem that a particular scientist chooses to study from the class of problems that might have been chosen, in the experiments that are selected in carrying out this study, and in the conclusions drawn from the experimental data. The mathematical theory of probability and statistics can play an important part in these choices, decisions, and conclusions.

Note: The Theory of Probability Does Not Depend on Interpretation. The mathematical theory of probability is developed and presented in Chapters 1–6 of this book without regard to the controversy surrounding the different interpretations of the term probability. This theory is correct and can be usefully applied, regardless of which interpretation of probability is used in a particular problem. The theories and techniques that will be presented in this book have served as valuable guides and tools in almost all aspects of the design and analysis of effective experimentation.

1.3 Experiments and Events

Probability will be the way that we quantify how likely something is to occur (in the sense of one of the interpretations in Sec. 1.2). In this section, we give examples of the types of situations in which probability will be used.

Types of Experiments

The theory of probability pertains to the various possible outcomes that might be obtained and the possible events that might occur when an experiment is performed.

Definition 1.3.1 Experiment and Event. An *experiment* is any process, real or hypothetical, in which the possible outcomes can be identified ahead of time. An *event* is a well-defined set of possible outcomes of the experiment.

The breadth of this definition allows us to call almost any imaginable process an experiment whether or not its outcome will ever be known. The probability of each event will be our way of saying how likely it is that the outcome of the experiment is in the event. Not every set of possible outcomes will be called an event. We shall be more specific about which subsets count as events in Sec. 1.4.

Probability will be most useful when applied to a real experiment in which the outcome is not known in advance, but there are many hypothetical experiments that provide useful tools for modeling real experiments. A common type of hypothetical experiment is repeating a well-defined task infinitely often under similar conditions. Some examples of experiments and specific events are given next. In each example, the words following "the probability that" describe the event of interest.

1. In an experiment in which a coin is to be tossed 10 times, the experimenter might want to determine the probability that at least four heads will be obtained.
2. In an experiment in which a sample of 1000 transistors is to be selected from a large shipment of similar items and each selected item is to be inspected, a person might want to determine the probability that not more than one of the selected transistors will be defective.
3. In an experiment in which the air temperature at a certain location is to be observed every day at noon for 90 successive days, a person might want to determine the probability that the average temperature during this period will be less than some specified value.
4. From information relating to the life of Thomas Jefferson, a person might want to determine the probability that Jefferson was born in the year 1741.
5. In evaluating an industrial research and development project at a certain time, a person might want to determine the probability that the project will result in the successful development of a new product within a specified number of months.

The Mathematical Theory of Probability

As was explained in Sec. 1.2, there is controversy in regard to the proper meaning and interpretation of some of the probabilities that are assigned to the outcomes of many experiments. However, once probabilities have been assigned to some simple outcomes in an experiment, there is complete agreement among all authorities that the mathematical theory of probability provides the appropriate methodology for the further study of these probabilities. Almost all work in the mathematical theory of probability, from the most elementary textbooks to the most advanced research, has been related to the following two problems: (i) methods for determining the probabilities of certain events from the specified probabilities of each possible outcome of an experiment and (ii) methods for revising the probabilities of events when additional relevant information is obtained.

These methods are based on standard mathematical techniques. The purpose of the first six chapters of this book is to present these techniques, which, together, form the mathematical theory of probability.

1.4 Set Theory

This section develops the formal mathematical model for events, namely, the theory of sets. Several important concepts are introduced, namely, element, subset, empty set, intersection, union, complement, and disjoint sets.

The Sample Space

Definition 1.4.1 Sample Space. The collection of all possible outcomes of an experiment is called the *sample space* of the experiment.

The sample space of an experiment can be thought of as a *set*, or collection, of different possible outcomes; and each outcome can be thought of as a *point*, or an *element*, in the sample space. Similarly, events can be thought of as *subsets* of the sample space.

Example 1.4.1 Rolling a Die. When a six-sided die is rolled, the sample space can be regarded as containing the six numbers 1, 2, 3, 4, 5, 6, each representing a possible side of the die that shows after the roll. Symbolically, we write

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

One event A is that an even number is obtained, and it can be represented as the subset $A = \{2, 4, 6\}$. The event B that a number greater than 2 is obtained is defined by the subset $B = \{3, 4, 5, 6\}$. ◀

Because we can interpret outcomes as elements of a set and events as subsets of a set, the language and concepts of set theory provide a natural context for the development of probability theory. The basic ideas and notation of set theory will now be reviewed.

Relations of Set Theory

Let S denote the sample space of some experiment. Then each possible outcome s of the experiment is said to be a member of the space S, or to belong to the space S. The statement that s is a member of S is denoted symbolically by the relation $s \in S$.

When an experiment has been performed and we say that some event E has occurred, we mean two equivalent things. One is that the outcome of the experiment satisfied the conditions that specified that event E. The other is that the outcome, considered as a point in the sample space, is an element of E.

To be precise, we should say which sets of outcomes correspond to events as defined above. In many applications, such as Example 1.4.1, it will be clear which sets of outcomes should correspond to events. In other applications (such as Example 1.4.5 coming up later), there are too many sets available to have them all be events. Ideally, we would like to have the largest possible collection of sets called events so that we have the broadest possible applicability of our probability calculations. However, when the sample space is too large (as in Example 1.4.5) the theory of probability simply will not extend to the collection of all subsets of the sample space. We would prefer not to dwell on this point for two reasons. First, a careful handling requires mathematical details that interfere with an initial understanding of the important concepts, and second, the practical implications for the results in this text are minimal. In order to be mathematically correct without imposing an undue burden on the reader, we note the following. In order to be able to do all of the probability calculations that we might find interesting, there are three simple conditions that must be met by the collection of sets that we call events. In every problem that we see in this text, there exists a collection of sets that includes all the sets that we will need to discuss and that satisfies the three conditions, and the reader should assume that such a collection has been chosen as the events. For a sample space S with only finitely many outcomes, the collection of all subsets of S satisfies the conditions, as the reader can show in Exercise 12 in this section.

The first of the three conditions can be stated immediately.

Condition 1 The sample space S must be an event.

That is, we must include the sample space S in our collection of events. The other two conditions will appear later in this section because they require additional definitions. Condition 2 is on page 9, and Condition 3 is on page 10.

Definition 1.4.2 Containment. It is said that a set A *is contained in* another set B if every element of the set A also belongs to the set B. This relation between two events is expressed symbolically by the expression $A \subset B$, which is the set-theoretic expression for saying that A is a subset of B. Equivalently, if $A \subset B$, we may say that B *contains* A and may write $B \supset A$.

For events, to say that $A \subset B$ means that if A occurs then so does B.

The proof of the following result is straightforward and is omitted.

Theorem 1.4.1 Let A, B, and C be events. Then $A \subset S$. If $A \subset B$ and $B \subset A$, then $A = B$. If $A \subset B$ and $B \subset C$, then $A \subset C$. ■

Example 1.4.2 Rolling a Die. In Example 1.4.1, suppose that A is the event that an even number is obtained and C is the event that a number greater than 1 is obtained. Since $A = \{2, 4, 6\}$ and $C = \{2, 3, 4, 5, 6\}$, it follows that $A \subset C$. ◄

The Empty Set Some events are impossible. For example, when a die is rolled, it is impossible to obtain a negative number. Hence, the event that a negative number will be obtained is defined by the subset of S that contains no outcomes.

Definition 1.4.3 Empty Set. The subset of S that contains no elements is called the *empty set*, or *null set*, and it is denoted by the symbol $\emptyset$.

In terms of events, the empty set is any event that cannot occur.

Theorem 1.4.2 Let A be an event. Then $\emptyset \subset A$.

Proof Let A be an arbitrary event. Since the empty set $\emptyset$ contains no points, it is logically correct to say that every point belonging to $\emptyset$ also belongs to A, or $\emptyset \subset A$. ■

Finite and Infinite Sets Some sets contain only finitely many elements, while others have infinitely many elements. There are two sizes of infinite sets that we need to distinguish.

Definition 1.4.4 Countable/Uncountable. An infinite set A is *countable* if there is a one-to-one correspondence between the elements of A and the set of natural numbers $\{1, 2, 3, \ldots\}$. A set is *uncountable* if it is neither finite nor countable. If we say that a set has *at most countably many* elements, we mean that the set is either finite or countable.

Examples of countably infinite sets include the integers, the even integers, the odd integers, the prime numbers, and any infinite sequence. Each of these can be put in one-to-one correspondence with the natural numbers. For example, the following function f puts the integers in one-to-one correspondence with the natural numbers:

$$f(n) = \begin{cases} \frac{n-1}{2} & \text{if } n \text{ is odd,} \\ -\frac{n}{2} & \text{if } n \text{ is even.} \end{cases}$$

Every infinite sequence of distinct items is a countable set, as its indexing puts it in one-to-one correspondence with the natural numbers. Examples of uncountable sets include the real numbers, the positive reals, the numbers in the interval $[0, 1]$, and the set of all ordered pairs of real numbers. An argument to show that the real numbers are uncountable appears at the end of this section. Every subset of the integers has at most countably many elements.

Operations of Set Theory

Definition 1.4.5 Complement. The *complement* of a set A is defined to be the set that contains all elements of the sample space S that *do not* belong to A. The notation for the complement of A is A^c.

In terms of events, the event A^c is the event that A does not occur.

Example 1.4.3 Rolling a Die. In Example 1.4.1, suppose again that A is the event that an even number is rolled; then $A^c = \{1, 3, 5\}$ is the event that an odd number is rolled. ◄

We can now state the second condition that we require of the collection of events.

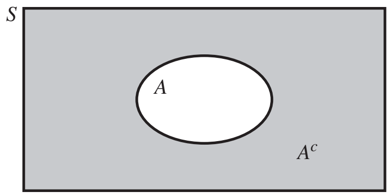


Figure 1.1 The event A^c.

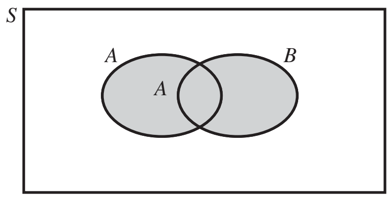


Figure 1.2 The set $A \cup B$.

Condition 2 If A is an event, then A^c is also an event.

That is, for each set A of outcomes that we call an event, we must also call its complement A^c an event.

A generic version of the relationship between A and A^c is sketched in Fig. 1.1. A sketch of this type is called a *Venn diagram*.

Some properties of the complement are stated without proof in the next result.

Theorem 1.4.3 Let A be an event. Then

$$(A^c)^c = A, \qquad \emptyset^c = S, \qquad S^c = \emptyset.$$

The empty event $\emptyset$ is an event. ■

Definition 1.4.6 Union of Two Sets. If A and B are any two sets, the *union* of A and B is defined to be the set containing all outcomes that belong to A alone, to B alone, or to both A and B. The notation for the union of A and B is $A \cup B$.

The set $A \cup B$ is sketched in Fig. 1.2. In terms of events, $A \cup B$ is the event that either A or B or both occur.

The union has the following properties whose proofs are left to the reader.

Theorem 1.4.4 For all sets A and B,

$$A \cup B = B \cup A, \qquad A \cup A = A, \qquad A \cup A^c = S,$$
$$A \cup \emptyset = A, \qquad A \cup S = S.$$

Furthermore, if $A \subset B$, then $A \cup B = B$. ■

The concept of union extends to more than two sets.

Definition 1.4.7 Union of Many Sets. The *union* of n sets $A_1, \ldots, A_n$ is defined to be the set that contains all outcomes that belong to at least one of these n sets. The notation for this union is either of the following:

$$A_1 \cup A_2 \cup \cdots \cup A_n \text{ or } \bigcup_{i=1}^{n} A_i.$$

Similarly, the *union* of an infinite sequence of sets $A_1, A_2, \ldots$ is the set that contains all outcomes that belong to at least one of the events in the sequence. The infinite union is denoted by $\bigcup_{i=1}^{\infty} A_i$.

In terms of events, the union of a collection of events is the event that at least one of the events in the collection occurs.

We can now state the final condition that we require for the collection of sets that we call events.

Condition 3 If $A_1, A_2, \ldots$ is a countable collection of events, then $\bigcup_{i=1}^{\infty} A_i$ is also an event.

In other words, if we choose to call each set of outcomes in some countable collection an event, we are required to call their union an event also. We do *not* require that the union of an arbitrary collection of events be an event. To be clear, let I be an arbitrary set that we use to index a general collection of events $\{A_i : i \in I\}$. The union of the events in this collection is the set of outcomes that are in at least one of the events in the collection. The notation for this union is $\bigcup_{i \in I} A_i$. We do not require that $\bigcup_{i \in I} A_i$ be an event unless I is countable.

Condition 3 refers to a countable collection of events. We can prove that the condition also applies to every finite collection of events.

Theorem 1.4.5 The union of a finite number of events $A_1, \ldots, A_n$ is an event.

Proof For each $m = n + 1, n + 2, \ldots$, define $A_m = \emptyset$. Because $\emptyset$ is an event, we now have a countable collection $A_1, A_2, \ldots$ of events. It follows from Condition 3 that $\bigcup_{m=1}^{\infty} A_m$ is an event. But it is easy to see that $\bigcup_{m=1}^{\infty} A_m = \bigcup_{m=1}^{n} A_m$. ■

The union of three events A, B, and C can be constructed either directly from the definition of $A \cup B \cup C$ or by first evaluating the union of any two of the events and then forming the union of this combination of events and the third event. In other words, the following result is true.

Theorem 1.4.6 Associative Property. For every three events A, B, and C, the following associative relations are satisfied:

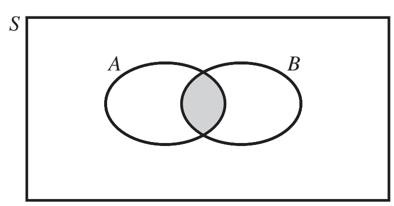
$$A \cup B \cup C = (A \cup B) \cup C = A \cup (B \cup C). \quad ■$$

Definition 1.4.8 Intersection of Two Sets. If A and B are any two sets, the *intersection* of A and B is defined to be the set that contains all outcomes that belong *both to A and to B*. The notation for the intersection of A and B is $A \cap B$.

The set $A \cap B$ is sketched in a Venn diagram in Fig. 1.3. In terms of events, $A \cap B$ is the event that both A and B occur.

The proof of the first part of the next result follows from Exercise 3 in this section. The rest of the proof is straightforward.

Figure 1.3 The set $A \cap B$.



Theorem 1.4.7 If A and B are events, then so is $A \cap B$. For all events A and B,

$$A \cap B = B \cap A, \qquad A \cap A = A, \qquad A \cap A^c = \emptyset,$$
$$A \cap \emptyset = \emptyset, \qquad A \cap S = A.$$

Furthermore, if $A \subset B$, then $A \cap B = A$. ■

The concept of intersection extends to more than two sets.

Definition 1.4.9 Intersection of Many Sets. The *intersection* of n sets $A_1, \ldots, A_n$ is defined to be the set that contains the elements that are common to all these n sets. The notation for this intersection is $A_1 \cap A_2 \cap \ldots \cap A_n$ or $\bigcap_{i=1}^{n} A_i$. Similar notations are used for the intersection of an infinite sequence of sets or for the intersection of an arbitrary collection of sets.

In terms of events, the intersection of a collection of events is the event that every event in the collection occurs.

The following result concerning the intersection of three events is straightforward to prove.

Theorem 1.4.8 Associative Property. For every three events A, B, and C, the following associative relations are satisfied:

$$A \cap B \cap C = (A \cap B) \cap C = A \cap (B \cap C). \quad ■$$

Definition 1.4.10 Disjoint/Mutually Exclusive. It is said that two sets A and B are *disjoint*, or *mutually exclusive*, if A and B have no outcomes in common, that is, if $A \cap B = \emptyset$. The sets $A_1, \ldots, A_n$ or the sets $A_1, A_2, \ldots$ are disjoint if for every $i \neq j$, we have that A_i and A_j are disjoint, that is, $A_i \cap A_j = \emptyset$ for all $i \neq j$. The events in an arbitrary collection are disjoint if no two events in the collection have any outcomes in common.

In terms of events, A and B are disjoint if they cannot both occur.

As an illustration of these concepts, a Venn diagram for three events A_1, A_2, and A_3 is presented in Fig. 1.4. This diagram indicates that the various intersections of A_1, A_2, and A_3 and their complements will partition the sample space S into eight disjoint subsets.

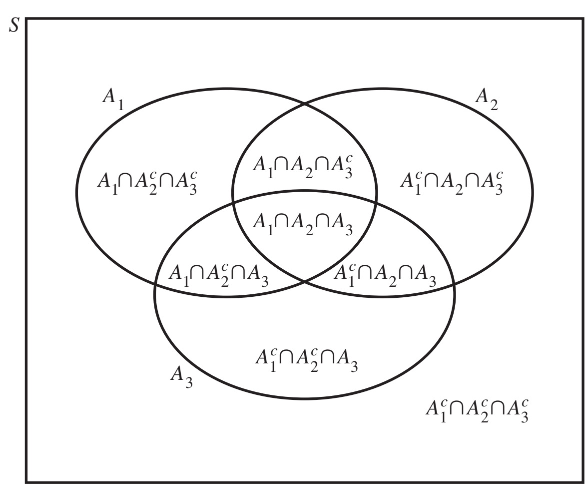


Figure 1.4 Partition of S determined by three events A_1, A_2, A_3.

Example 1.4.4

Tossing a Coin. Suppose that a coin is tossed three times. Then the sample space S contains the following eight possible outcomes $s_1, \ldots, s_8$:

s_1: HHH,
s_2: THH,
s_3: HTH,
s_4: HHT,
s_5: HTT,
s_6: THT,
s_7: TTH,
s_8: TTT.

In this notation, H indicates a head and T indicates a tail. The outcome s_3, for example, is the outcome in which a head is obtained on the first toss, a tail is obtained on the second toss, and a head is obtained on the third toss.

To apply the concepts introduced in this section, we shall define four events as follows: Let A be the event that at least one head is obtained in the three tosses; let B be the event that a head is obtained on the second toss; let C be the event that a tail is obtained on the third toss; and let D be the event that *no* heads are obtained. Accordingly,

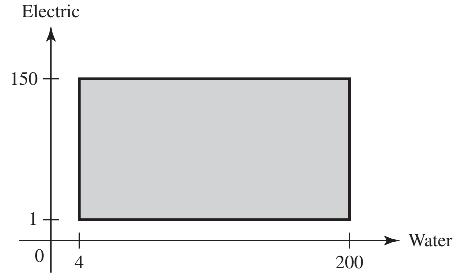
$$A = \{s_1, s_2, s_3, s_4, s_5, s_6, s_7\},$$
$$B = \{s_1, s_2, s_4, s_6\},$$
$$C = \{s_4, s_5, s_6, s_8\},$$
$$D = \{s_8\}.$$

Various relations among these events can be derived. Some of these relations are $B \subset A$, $A^c = D$, $B \cap D = \emptyset$, $A \cup C = S$, $B \cap C = \{s_4, s_6\}$, $(B \cup C)^c = \{s_3, s_7\}$, and $A \cap (B \cup C) = \{s_1, s_2, s_4, s_5, s_6\}$. ◀

Example 1.4.5

Demands for Utilities. A contractor is building an office complex and needs to plan for water and electricity demand (sizes of pipes, conduit, and wires). After consulting with prospective tenants and examining historical data, the contractor decides that the demand for electricity will range somewhere between 1 million and 150 million kilowatt-hours per day and water demand will be between 4 and 200 (in thousands of gallons per day). All combinations of electrical and water demand are considered possible. The shaded region in Fig. 1.5 shows the sample space for the experiment, consisting of learning the actual water and electricity demands for the office complex. We can express the sample space as the set of ordered pairs $\{(x, y) : 4 \le x \le 200, 1 \le y \le 150\}$, where x stands for water demand in thousands of gallons per day and y

Figure 1.5 Sample space for water and electric demand in Example 1.4.5



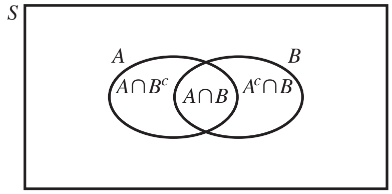


Figure 1.6 Partition of $A \cup B$ in Theorem 1.4.11.

stands for the electric demand in millions of kilowatt-hours per day. The types of sets that we want to call events include sets like

$$\{\text{water demand is at least } 100\} = \{(x, y) : x \geq 100\}, \text{ and}$$
$$\{\text{electric demand is no more than } 35\} = \{(x, y) : y \leq 35\},$$

along with intersections, unions, and complements of such sets. This sample space has infinitely many points. Indeed, the sample space is uncountable. There are many more sets that are difficult to describe and which we will have no need to consider as events. ◂

Additional Properties of Sets The proof of the following useful result is left to Exercise 3 in this section.

Theorem 1.4.9 De Morgan's Laws. For every two sets A and B,

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

The generalization of Theorem 1.4.9 is the subject of Exercise 5 in this section.

The proofs of the following distributive properties are left to Exercise 2 in this section. These properties also extend in natural ways to larger collections of events.

Theorem 1.4.10 Distributive Properties. For every three sets A, B, and C,

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C) \quad \text{and} \quad A \cup (B \cap C) = (A \cup B) \cap (A \cup C). \quad \blacksquare$$

The following result is useful for computing probabilities of events that can be partitioned into smaller pieces. Its proof is left to Exercise 4 in this section, and is illuminated by Fig. 1.6.

Theorem 1.4.11 Partitioning a Set. For every two sets A and B, $A \cap B$ and $A \cap B^c$ are disjoint and

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

In addition, B and $A \cap B^c$ are disjoint, and

$$A \cup B = B \cup (A \cap B^c). \quad \blacksquare$$

Proof That the Real Numbers Are Uncountable

We shall show that the real numbers in the interval $[0, 1)$ are uncountable. Every larger set is a fortiori uncountable. For each number $x \in [0, 1)$, define the sequence $\{a_n(x)\}_{n=1}^{\infty}$ as follows. First, $a_1(x) = \lfloor 10x \rfloor$, where $\lfloor y \rfloor$ stands for the greatest integer less than or equal to y (round nonintegers down to the closest integer below). Then

$$
\begin{array}{cccccccc}
\underline{0} & 2 & 3 & 0 & 7 & 1 & 3 & \ldots \\
1 & \underline{9} & 9 & 2 & 1 & 0 & 0 & \ldots \\
2 & 7 & \underline{3} & 6 & 0 & 1 & 1 & \ldots \\
8 & 0 & 2 & \underline{1} & 2 & 7 & 9 & \ldots \\
7 & 0 & 1 & 6 & \underline{0} & 1 & 3 & \ldots \\
1 & 5 & 1 & 5 & \underline{1} & 5 & 1 & \ldots \\
2 & 3 & 4 & 5 & 6 & \underline{7} & 8 & \ldots \\
0 & 1 & 7 & 3 & 2 & 9 & \underline{8} & \ldots \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}
$$

Figure 1.7 An array of a countable collection of sequences of digits with the diagonal underlined.

set $b_1(x) = 10x - a_1(x)$, which will again be in $[0, 1)$. For $n > 1$, $a_n(x) = \lfloor 10b_{n-1}(x) \rfloor$ and $b_n(x) = 10b_{n-1}(x) - a_n(x)$. It is easy to see that the sequence $\{a_n(x)\}_{n=1}^{\infty}$ gives a decimal expansion for x in the form

$$x = \sum_{n=1}^{\infty} a_n(x)10^{-n}. \tag{1.4.1}$$

By construction, each number of the form $x = k/10^m$ for some nonnegative integers k and m will have $a_n(x) = 0$ for $n > m$. The numbers of the form $k/10^m$ are the only ones that have an alternate decimal expansion $x = \sum_{n=1}^{\infty} c_n(x)10^{-n}$. When k is not a multiple of 10, this alternate expansion satisfies $c_n(x) = a_n(x)$ for $n = 1, \ldots, m-1$, $c_m(x) = a_m(x) - 1$, and $c_n(x) = 9$ for $n > m$. Let $C = \{0, 1, \ldots, 9\}^{\infty}$ stand for the set of all infinite sequences of digits. Let B denote the subset of C consisting of those sequences that don't end in repeating 9's. Then we have just constructed a function a from the interval $[0, 1)$ onto B that is one-to-one and whose inverse is given in (1.4.1). We now show that the set B is uncountable, hence $[0, 1)$ is uncountable. Take any countable subset of B and arrange the sequences into a rectangular array with the kth sequence running across the kth row of the array for $k = 1, 2, \ldots$. Figure 1.7 gives an example of part of such an array.

In Fig. 1.7, we have underlined the kth digit in the kth sequence for each k. This portion of the array is called the *diagonal* of the array. We now show that there must exist a sequence in B that is not part of this array. This will prove that the whole set B cannot be put into such an array, and hence cannot be countable. Construct the sequence $\{d_n\}_{n=1}^{\infty}$ as follows. For each n, let $d_n = 2$ if the nth digit in the nth sequence is 1, and $d_n = 1$ otherwise. This sequence does not end in repeating 9's; hence, it is in B. We conclude the proof by showing that $\{d_n\}_{n=1}^{\infty}$ does not appear anywhere in the array. If the sequence did appear in the array, say, in the kth row, then its kth element would be the kth diagonal element of the array. But we constructed the sequence so that for every n (including $n = k$), its nth element never matched the nth diagonal element. Hence, the sequence can't be in the kth row, no matter what k is. The argument given here is essentially that of the nineteenth-century German mathematician Georg Cantor.

Summary

We will use set theory for the mathematical model of events. Outcomes of an experiment are elements of some sample space S, and each event is a subset of S. Two events both occur if the outcome is in the intersection of the two sets. At least one of a collection of events occurs if the outcome is in the union of the sets. Two events cannot both occur if the sets are disjoint. An event fails to occur if the outcome is in the complement of the set. The empty set stands for every event that cannot possibly occur. The collection of events is assumed to contain the sample space, the complement of each event, and the union of each countable collection of events.

Exercises

1. Suppose that $A \subset B$. Show that $B^c \subset A^c$.

2. Prove the distributive properties in Theorem 1.4.10.

3. Prove De Morgan's laws (Theorem 1.4.9).

4. Prove Theorem 1.4.11.

5. For every collection of events A_i $(i \in I)$, show that

$$\left(\bigcup_{i\in I} A_i\right)^c = \bigcap_{i\in I} A_i^c \quad \text{and} \quad \left(\bigcap_{i\in I} A_i\right)^c = \bigcup_{i\in I} A_i^c.$$

6. Suppose that one card is to be selected from a deck of 20 cards that contains 10 red cards numbered from 1 to 10 and 10 blue cards numbered from 1 to 10. Let A be the event that a card with an even number is selected, let B be the event that a blue card is selected, and let C be the event that a card with a number less than 5 is selected. Describe the sample space S and describe each of the following events both in words and as subsets of S:

a. $A \cap B \cap C$ **b.** $B \cap C^c$ **c.** $A \cup B \cup C$
d. $A \cap (B \cup C)$ **e.** $A^c \cap B^c \cap C^c$.

7. Suppose that a number x is to be selected from the real line S, and let A, B, and C be the events represented by the following subsets of S, where the notation $\{x: - - -\}$ denotes the set containing every point x for which the property presented following the colon is satisfied:

$$A = \{x: 1 \le x \le 5\},$$
$$B = \{x: 3 < x \le 7\},$$
$$C = \{x: x \le 0\}.$$

Describe each of the following events as a set of real numbers:

a. A^c **b.** $A \cup B$ **c.** $B \cap C^c$
d. $A^c \cap B^c \cap C^c$ **e.** $(A \cup B) \cap C$.

8. A simplified model of the human blood-type system has four blood types: A, B, AB, and O. There are two antigens, anti-A and anti-B, that react with a person's blood in different ways depending on the blood type. Anti-A reacts with blood types A and AB, but not with B and O. Anti-B reacts with blood types B and AB, but not with A and O. Suppose that a person's blood is sampled and tested with the two antigens. Let A be the event that the blood reacts with anti-A, and let B be the event that it reacts with anti-B. Classify the person's blood type using the events A, B, and their complements.

9. Let S be a given sample space and let $A_1, A_2, \ldots$ be an infinite sequence of events. For $n = 1, 2, \ldots$, let $B_n = \bigcup_{i=n}^{\infty} A_i$ and let $C_n = \bigcap_{i=n}^{\infty} A_i$.

- **a.** Show that $B_1 \supset B_2 \supset \cdots$ and that $C_1 \subset C_2 \subset \cdots$.
- **b.** Show that an outcome in S belongs to the event $\bigcap_{n=1}^{\infty} B_n$ if and only if it belongs to an infinite number of the events $A_1, A_2, \ldots$.
- **c.** Show that an outcome in S belongs to the event $\bigcup_{n=1}^{\infty} C_n$ if and only if it belongs to all the events $A_1, A_2, \ldots$ except possibly a finite number of those events.

10. Three six-sided dice are rolled. The six sides of each die are numbered 1–6. Let A be the event that the first die shows an even number, let B be the event that the second die shows an even number, and let C be the event that the third die shows an even number. Also, for each $i = 1, \ldots, 6$, let A_i be the event that the first die shows the number i, let B_i be the event that the second die shows the number i, and let C_i be the event that the third die shows the number i. Express each of the following events in terms of the named events described above:

- **a.** The event that all three dice show even numbers
- **b.** The event that no die shows an even number
- **c.** The event that at least one die shows an odd number
- **d.** The event that at most two dice show odd numbers
- **e.** The event that the sum of the three dices is no greater than 5

11. A power cell consists of two subcells, each of which can provide from 0 to 5 volts, regardless of what the other subcell provides. The power cell is functional if and only if the sum of the two voltages of the subcells is at least 6 volts. An experiment consists of measuring and recording the voltages of the two subcells. Let A be the event that the power cell is functional, let B be the event that two subcells have the same voltage, let C be the event that the first subcell has a strictly higher voltage than the second subcell, and let D be the event that the power cell is not functional but needs less than one additional volt to become functional.

a. Define a sample space S for the experiment as a set of ordered pairs that makes it possible for you to express the four sets above as events.

b. Express each of the events A, B, C, and D as sets of ordered pairs that are subsets of S.

c. Express the following set in terms of A, B, C, and/or D: $\{(x, y) : x = y \text{ and } x + y \le 5\}$.

d. Express the following event in terms of A, B, C, and/or D: the event that the power cell is not functional and the second subcell has a strictly higher voltage than the first subcell.

12. Suppose that the sample space S of some experiment is finite. Show that the collection of all subsets of S satisfies the three conditions required to be called the collection of events.

13. Let S be the sample space for some experiment. Show that the collection of subsets consisting solely of S and $\emptyset$ satisfies the three conditions required in order to be called the collection of events. Explain why this collection would not be very interesting in most real problems.

14. Suppose that the sample space S of some experiment is countable. Suppose also that, for every outcome $s \in S$, the subset $\{s\}$ is an event. Show that every subset of S must be an event. *Hint:* Recall the three conditions required of the collection of subsets of S that we call events.

1.5 The Definition of Probability

We begin with the mathematical definition of probability and then present some useful results that follow easily from the definition.

Axioms and Basic Theorems

In this section, we shall present the mathematical, or axiomatic, definition of probability. In a given experiment, it is necessary to assign to each event A in the sample space S a number $\Pr(A)$ that indicates the probability that A will occur. In order to satisfy the mathematical definition of probability, the number $\Pr(A)$ that is assigned must satisfy three specific axioms. These axioms ensure that the number $\Pr(A)$ will have certain properties that we intuitively expect a probability to have under each of the various interpretations described in Sec. 1.2.

The first axiom states that the probability of every event must be nonnegative.

Axiom 1 For every event A, $\Pr(A) \ge 0$.

The second axiom states that if an event is certain to occur, then the probability of that event is 1.

Axiom 2 $\Pr(S) = 1$.

Before stating Axiom 3, we shall discuss the probabilities of disjoint events. If two events are disjoint, it is natural to assume that the probability that one or the other will occur is the sum of their individual probabilities. In fact, it will be assumed that this *additive property* of probability is also true for every finite collection of disjoint events and even for every infinite sequence of disjoint events. If we assume that this additive property is true only for a finite number of disjoint events, we cannot then be certain that the property will be true for an infinite sequence of disjoint events as well. However, if we assume that the additive property is true for every infinite sequence of disjoint events, then (as we shall prove) the property must also be true for every finite number of disjoint events. These considerations lead to the third axiom.

Axiom 3 For every infinite sequence of disjoint events $A_1, A_2, \ldots,$

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

Example 1.5.1 Rolling a Die. In Example 1.4.1, for each subset A of $S = \{1, 2, 3, 4, 5, 6\}$, let $\Pr(A)$ be the number of elements of A divided by 6. It is trivial to see that this satisfies the first two axioms. There are only finitely many distinct collections of nonempty disjoint events. It is not difficult to see that Axiom 3 is also satisfied by this example. ◂

Example 1.5.2 A Loaded Die. In Example 1.5.1, there are other choices for the probabilities of events. For example, if we believe that the die is loaded, we might believe that some sides have different probabilities of turning up. To be specific, suppose that we believe that 6 is twice as likely to come up as each of the other five sides. We could set $p_i = 1/7$ for $i = 1, 2, 3, 4, 5$ and $p_6 = 2/7$. Then, for each event A, define $\Pr(A)$ to be the sum of all p_i such that $i \in A$. For example, if $A = \{1, 3, 5\}$, then $\Pr(A) = p_1 + p_3 + p_5 = 3/7$. It is not difficult to check that this also satisfies all three axioms. ◂

We are now prepared to give the mathematical definition of probability.

Definition 1.5.1 Probability. A *probability measure*, or simply a *probability*, on a sample space S is a specification of numbers $\Pr(A)$ for all events A that satisfy Axioms 1, 2, and 3.

We shall now derive two important consequences of Axiom 3. First, we shall show that if an event is impossible, its probability must be 0.

Theorem 1.5.1 $\Pr(\emptyset) = 0$.

Proof Consider the infinite sequence of events $A_1, A_2, \ldots$ such that $A_i = \emptyset$ for $i = 1, 2, \ldots$. In other words, each of the events in the sequence is just the empty set $\emptyset$. Then this sequence is a sequence of disjoint events, since $\emptyset \cap \emptyset = \emptyset$. Furthermore, $\bigcup_{i=1}^{\infty} A_i = \emptyset$. Therefore, it follows from Axiom 3 that

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

This equation states that when the number $\Pr(\emptyset)$ is added repeatedly in an infinite series, the sum of that series is simply the number $\Pr(\emptyset)$. The only real number with this property is zero. ■

We can now show that the additive property assumed in Axiom 3 for an infinite sequence of disjoint events is also true for every finite number of disjoint events.

Theorem 1.5.2 For every finite sequence of n disjoint events $A_1, \ldots, A_n$,

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

Proof Consider the infinite sequence of events $A_1, A_2, \ldots,$ in which $A_1, \ldots, A_n$ are the n given disjoint events and $A_i = \emptyset$ for $i > n$. Then the events in this infinite sequence are disjoint and $\bigcup_{i=1}^{\infty} A_i = \bigcup_{i=1}^{n} A_i$. Therefore, by Axiom 3,

$$\begin{aligned}\Pr\left(\bigcup_{i=1}^{n} A_i\right) &= \Pr\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \Pr(A_i) \\ &= \sum_{i=1}^{n} \Pr(A_i) + \sum_{i=n+1}^{\infty} \Pr(A_i) \\ &= \sum_{i=1}^{n} \Pr(A_i) + 0 \\ &= \sum_{i=1}^{n} \Pr(A_i).\end{aligned}$$ ■

Further Properties of Probability

From the axioms and theorems just given, we shall now derive four other general properties of probability measures. Because of the fundamental nature of these four properties, they will be presented in the form of four theorems, each one of which is easily proved.

Theorem 1.5.3 For every event A, $\Pr(A^c) = 1 - \Pr(A)$.

Proof Since A and A^c are disjoint events and $A \cup A^c = S$, it follows from Theorem 1.5.2 that $\Pr(S) = \Pr(A) + \Pr(A^c)$. Since $\Pr(S) = 1$ by Axiom 2, then $\Pr(A^c) = 1 - \Pr(A)$. ■

Theorem 1.5.4 If $A \subset B$, then $\Pr(A) \le \Pr(B)$.

Proof As illustrated in Fig. 1.8, the event B may be treated as the union of the two disjoint events A and $B \cap A^c$. Therefore, $\Pr(B) = \Pr(A) + \Pr(B \cap A^c)$. Since $\Pr(B \cap A^c) \ge 0$, then $\Pr(B) \ge \Pr(A)$. ■

Theorem 1.5.5 For every event A, $0 \le \Pr(A) \le 1$.

Proof It is known from Axiom 1 that $\Pr(A) \ge 0$. Since $A \subset S$ for every event A, Theorem 1.5.4 implies $\Pr(A) \le \Pr(S) = 1$, by Axiom 2. ■

Theorem 1.5.6 For every two events A and B,

$$\Pr(A \cap B^c) = \Pr(A) - \Pr(A \cap B).$$

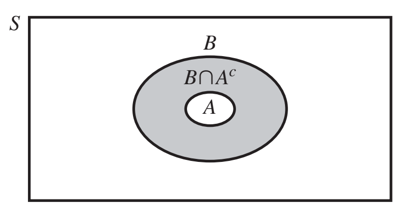


Figure 1.8 $B = A \cup (B \cap A^c)$ in the proof of Theorem 1.5.4.

Proof According to Theorem 1.4.11, the events $A \cap B^c$ and $A \cap B$ are disjoint and

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

It follows from Theorem 1.5.2 that

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

Subtract $\Pr(A \cap B)$ from both sides of this last equation to complete the proof. ■

Theorem 1.5.7 For every two events A and B,

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \tag{1.5.1}$$

Proof From Theorem 1.4.11, we have

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

and the two events on the right side of this equation are disjoint. Hence, we have

$$\begin{aligned}\Pr(A \cup B) &= \Pr(B) + \Pr(A \cap B^c)\\ &= \Pr(B) + \Pr(A) - \Pr(A \cap B),\end{aligned}$$

where the first equation follows from Theorem 1.5.2, and the second follows from Theorem 1.5.6. ■

Example 1.5.3 Diagnosing Diseases. A patient arrives at a doctor's office with a sore throat and low-grade fever. After an exam, the doctor decides that the patient has either a bacterial infection or a viral infection or both. The doctor decides that there is a probability of 0.7 that the patient has a bacterial infection and a probability of 0.4 that the person has a viral infection. What is the probability that the patient has both infections?

Let B be the event that the patient has a bacterial infection, and let V be the event that the patient has a viral infection. We are told $\Pr(B) = 0.7$, that $\Pr(V) = 0.4$, and that $S = B \cup V$. We are asked to find $\Pr(B \cap V)$. We will use Theorem 1.5.7, which says that

$$\Pr(B \cup V) = \Pr(B) + \Pr(V) - \Pr(B \cap V). \tag{1.5.2}$$

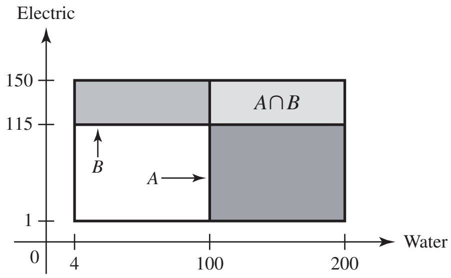
Since $S = B \cup V$, the left-hand side of (1.5.2) is 1, while the first two terms on the right-hand side are 0.7 and 0.4. The result is

$$1 = 0.7 + 0.4 - \Pr(B \cap V),$$

which leads to $\Pr(B \cap V) = 0.1$, the probability that the patient has both infections. ◀

Example 1.5.4 Demands for Utilities. Consider, once again, the contractor who needs to plan for water and electricity demands in Example 1.4.5. There are many possible choices for how to spread the probability around the sample space (pictured in Fig. 1.5 on page 12). One simple choice is to make the probability of an event E proportional to the area of E. The area of S (the sample space) is $(150 - 1) \times (200 - 4) = 29{,}204$, so $\Pr(E)$ equals the area of E divided by 29,204. For example, suppose that the contractor is interested in high demand. Let A be the set where water demand is at least 100, and let B be the event that electric demand is at least 115, and suppose that these values are considered high demand. These events are shaded with different patterns in Fig. 1.9. The area of A is $(150 - 1) \times (200 - 100) = 14{,}900$, and the area

Figure 1.9 The two events of interest in utility demand sample space for Example 1.5.4.



of B is $(150 - 115) \times (200 - 4) = 6{,}860$. So,

$$\Pr(A) = \frac{14{,}900}{29{,}204} = 0.5102, \quad \Pr(B) = \frac{6{,}860}{29{,}204} = 0.2349.$$

The two events intersect in the region denoted by $A \cap B$. The area of this region is $(150 - 115) \times (200 - 100) = 3{,}500$, so $\Pr(A \cap B) = 3{,}500/29{,}204 = 0.1198$. If the contractor wishes to compute the probability that at least one of the two demands will be high, that probability is

$$\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = 0.5102 + 0.2349 - 0.1198 = 0.6253,$$

according to Theorem 1.5.7. ◀

The proof of the following useful result is left to Exercise 13.

Theorem 1.5.8 Bonferroni Inequality. For all events $A_1, \ldots, A_n$,

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \Pr(A_i) \text{ and } \Pr\left(\bigcap_{i=1}^{n} A_i\right) \ge 1 - \sum_{i=1}^{n} \Pr(A_i^c).$$

(The second inequality above is known as the *Bonferroni inequality*.) ■

Note: Probability Zero Does Not Mean Impossible. When an event has probability 0, it does not mean that the event is impossible. In Example 1.5.4, there are many events with 0 probability, but they are not all impossible. For example, for every x, the event that water demand equals x corresponds to a line segment in Fig. 1.5. Since line segments have 0 area, the probability of every such line segment is 0, but the events are not all impossible. Indeed, if every event of the form {water demand equals x} were impossible, then water demand could not take any value at all. If $\epsilon > 0$, the event

$$\{\text{water demand is between } x - \epsilon \text{ and } x + \epsilon\}$$

will have positive probability, but that probability will go to 0 as ϵ goes to 0.

Summary

We have presented the mathematical definition of probability through the three axioms. The axioms require that every event have nonnegative probability, that the whole sample space have probability 1, and that the union of an infinite sequence of disjoint events have probability equal to the sum of their probabilities. Some important results to remember include the following:

- If $A_1, \dots, A_k$ are disjoint, $\Pr\left(\cup_{i=1}^k A_i\right) = \sum_{i=1}^k \Pr(A_i)$.
- $\Pr(A^c) = 1 - \Pr(A)$.
- $A \subset B$ implies that $\Pr(A) \le \Pr(B)$.
- $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B)$.

It does not matter how the probabilities were determined. As long as they satisfy the three axioms, they must also satisfy the above relations as well as all of the results that we prove later in the text.

Exercises

1. One ball is to be selected from a box containing red, white, blue, yellow, and green balls. If the probability that the selected ball will be red is 1/5 and the probability that it will be white is 2/5, what is the probability that it will be blue, yellow, or green?

2. A student selected from a class will be either a boy or a girl. If the probability that a boy will be selected is 0.3, what is the probability that a girl will be selected?

3. Consider two events A and B such that $\Pr(A) = 1/3$ and $\Pr(B) = 1/2$. Determine the value of $\Pr(B \cap A^c)$ for each of the following conditions: **(a)** A and B are disjoint; **(b)** $A \subset B$; **(c)** $\Pr(A \cap B) = 1/8$.

4. If the probability that student A will fail a certain statistics examination is 0.5, the probability that student B will fail the examination is 0.2, and the probability that both student A and student B will fail the examination is 0.1, what is the probability that at least one of these two students will fail the examination?

5. For the conditions of Exercise 4, what is the probability that neither student A nor student B will fail the examination?

6. For the conditions of Exercise 4, what is the probability that exactly one of the two students will fail the examination?

7. Consider two events A and B with $\Pr(A) = 0.4$ and $\Pr(B) = 0.7$. Determine the maximum and minimum possible values of $\Pr(A \cap B)$ and the conditions under which each of these values is attained.

8. If 50 percent of the families in a certain city subscribe to the morning newspaper, 65 percent of the families subscribe to the afternoon newspaper, and 85 percent of the families subscribe to at least one of the two newspapers, what percentage of the families subscribe to both newspapers?

9. Prove that for every two events A and B, the probability that exactly one of the two events will occur is given by the expression

$$\Pr(A) + \Pr(B) - 2\Pr(A \cap B).$$

10. For two arbitrary events A and B, prove that

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

11. A point (x, y) is to be selected from the square S containing all points (x, y) such that $0 \le x \le 1$ and $0 \le y \le 1$. Suppose that the probability that the selected point will belong to each specified subset of S is equal to the area of that subset. Find the probability of each of the following subsets: **(a)** the subset of points such that $(x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \ge \frac{1}{4}$; **(b)** the subset of points such that $\frac{1}{2} < x + y < \frac{3}{2}$; **(c)** the subset of points such that $y \le 1 - x^2$; **(d)** the subset of points such that $x = y$.

12. Let $A_1, A_2, \dots$ be an arbitrary infinite sequence of events, and let $B_1, B_2, \dots$ be another infinite sequence of events defined as follows: $B_1 = A_1$, $B_2 = A_1^c \cap A_2$, $B_3 = A_1^c \cap A_2^c \cap A_3$, $B_4 = A_1^c \cap A_2^c \cap A_3^c \cap A_4, \dots$. Prove that

$$\Pr\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n \Pr(B_i) \text{ for } n = 1, 2, \dots,$$

and that

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

13. Prove Theorem 1.5.8. *Hint:* Use Exercise 12.

14. Consider, once again, the four blood types A, B, AB, and O described in Exercise 8 in Sec. 1.4 together with the two antigens anti-A and anti-B. Suppose that, for a given person, the probability of type O blood is 0.5, the probability of type A blood is 0.34, and the probability of type B blood is 0.12.

a. Find the probability that each of the antigens will react with this person's blood.

b. Find the probability that both antigens will react with this person's blood.

1.6 Finite Sample Spaces

The simplest experiments in which to determine and derive probabilities are those that involve only finitely many possible outcomes. This section gives several examples to illustrate the important concepts from Sec. 1.5 in finite sample spaces.

Example 1.6.1

Current Population Survey. Every month, the Census Bureau conducts a survey of the United States population in order to learn about labor-force characteristics. Several pieces of information are collected on each of about 50,000 households. One piece of information is whether or not someone in the household is actively looking for employment but currently not employed. Suppose that our experiment consists of selecting three households at random from the 50,000 that were surveyed in a particular month and obtaining access to the information recorded during the survey. (Due to the confidential nature of information obtained during the Current Population Survey, only researchers in the Census Bureau would be able to perform the experiment just described.) The outcomes that make up the sample space S for this experiment can be described as lists of three three distinct numbers from 1 to 50,000. For example (300, 1, 24602) is one such list where we have kept track of the order in which the three households were selected. Clearly, there are only finitely many such lists. We can assume that each list is equally likely to be chosen, but we need to be able to count how many such lists there are. We shall learn a method for counting the outcomes for this example in Sec. 1.7. ◀

Requirements of Probabilities

In this section, we shall consider experiments for which there are only a finite number of possible outcomes. In other words, we shall consider experiments for which the sample space S contains only a finite number of points $s_1, \dots, s_n$. In an experiment of this type, a probability measure on S is specified by assigning a probability p_i to each point $s_i \in S$. The number p_i is the probability that the outcome of the experiment will be s_i $(i = 1, \dots, n)$. In order to satisfy the axioms of probability, the numbers $p_1, \dots, p_n$ must satisfy the following two conditions:

$$p_i \geq 0 \quad \text{for } i = 1, \dots, n$$

and

$$\sum_{i=1}^{n} p_i = 1.$$

The probability of each event A can then be found by adding the probabilities p_i of all outcomes s_i that belong to A. This is the general version of Example 1.5.2.

Example 1.6.2

Fiber Breaks. Consider an experiment in which five fibers having different lengths are subjected to a testing process to learn which fiber will break first. Suppose that the lengths of the five fibers are 1, 2, 3, 4, and 5 inches, respectively. Suppose also that the probability that any given fiber will be the first to break is proportional to the length of that fiber. We shall determine the probability that the length of the fiber that breaks first is not more than 3 inches.

In this example, we shall let s_i be the outcome in which the fiber whose length is i inches breaks first $(i = 1, \dots, 5)$. Then $S = \{s_1, \dots, s_5\}$ and $p_i = \alpha i$ for $i = 1, \dots, 5$, where α is a proportionality factor. It must be true that $p_1 + \cdots + p_5 = 1$, and we know that $p_1 + \cdots + p_5 = 15\alpha$, so $\alpha = 1/15$. If A is the event that the length of the fiber that breaks first is not more than 3 inches, then $A = \{s_1, s_2, s_3\}$. Therefore,

$$\Pr(A) = p_1 + p_2 + p_3 = \frac{1}{15} + \frac{2}{15} + \frac{3}{15} = \frac{2}{5}. \quad ◀$$

Simple Sample Spaces

A sample space S containing n outcomes $s_1, \dots, s_n$ is called a simple sample space if the probability assigned to each of the outcomes $s_1, \dots, s_n$ is $1/n$. If an event A in this simple sample space contains exactly m outcomes, then

$$\Pr(A) = \frac{m}{n}.$$

Example 1.6.3

Tossing Coins. Suppose that three fair coins are tossed simultaneously. We shall determine the probability of obtaining exactly two heads.

Regardless of whether or not the three coins can be distinguished from each other by the experimenter, it is convenient for the purpose of describing the sample space to assume that the coins can be distinguished. We can then speak of the result for the first coin, the result for the second coin, and the result for the third coin; and the sample space will comprise the eight possible outcomes listed in Example 1.4.4 on page 12.

Furthermore, because of the assumption that the coins are fair, it is reasonable to assume that this sample space is simple and that the probability assigned to each of the eight outcomes is 1/8. As can be seen from the listing in Example 1.4.4, exactly two heads will be obtained in three of these outcomes. Therefore, the probability of obtaining exactly two heads is 3/8. ◀

It should be noted that if we had considered the only possible outcomes to be no heads, one head, two heads, and three heads, it would have been reasonable to assume that the sample space contained just these four outcomes. This sample space would not be simple because the outcomes *would not be equally probable*.

Example 1.6.4

Genetics. Inherited traits in humans are determined by material in specific locations on chromosomes. Each normal human receives 23 chromosomes from each parent, and these chromosomes are naturally paired, with one chromosome in each pair coming from each parent. For the purposes of this text, it is safe to think of a *gene* as a portion of each chromosome in a pair. The genes, either one at a time or in combination, determine the inherited traits, such as blood type and hair color. The material in the two locations that make up a gene on the pair of chromosomes comes in forms called *alleles*. Each distinct combination of alleles (one on each chromosome) is called a *genotype*.

Consider a gene with only two different alleles A and a. Suppose that both parents have genotype Aa, that is, each parent has allele A on one chromosome and allele a on the other. (We do not distinguish the same alleles in a different order as a different genotype. For example, aA would be the same genotype as Aa. But it can be convenient to distinguish the two chromosomes during intermediate steps in probability calculations, just as we distinguished the three coins in Example 1.6.3.) What are the possible genotypes of an offspring of these two parents? If all possible results of the parents contributing pairs of alleles are equally likely, what are the probabilities of the different genotypes?

To begin, we shall distinguish which allele the offspring receives from each parent, since we are assuming that pairs of contributed alleles are equally likely.

Afterward, we shall combine those results that produce the same genotype. The possible contributions from the parents are:

	Mother	
Father	A	a
A	AA	Aa
a	aA	aa

So, there are three possible genotypes AA, Aa, and aa for the offspring. Since we assumed that every combination was equally likely, the four cells in the table all have probability 1/4. Since two of the cells in the table combined into genotype Aa, that genotype has probability 1/2. The other two genotypes each have probability 1/4, since they each correspond to only one cell in the table. ◀

Example 1.6.5 Rolling Two Dice. We shall now consider an experiment in which two balanced dice are rolled, and we shall calculate the probability of each of the possible values of the sum of the two numbers that may appear.

Although the experimenter need not be able to distinguish the two dice from one another in order to observe the value of their sum, the specification of a simple sample space in this example will be facilitated if we assume that the two dice are distinguishable. If this assumption is made, each outcome in the sample space S can be represented as a pair of numbers (x, y), where x is the number that appears on the first die and y is the number that appears on the second die. Therefore, S comprises the following 36 outcomes:

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

It is natural to assume that S is a simple sample space and that the probability of each of these outcomes is 1/36.

Let P_i denote the probability that the sum of the two numbers is i for $i = 2, 3, \ldots, 12$. The only outcome in S for which the sum is 2 is the outcome (1, 1). Therefore, $P_2 = 1/36$. The sum will be 3 for either of the two outcomes (1, 2) and (2, 1). Therefore, $P_3 = 2/36 = 1/18$. By continuing in this manner, we obtain the following probability for each of the possible values of the sum:

$$P_2 = P_{12} = \frac{1}{36}, \qquad P_5 = P_9 = \frac{4}{36},$$
$$P_3 = P_{11} = \frac{2}{36}, \qquad P_6 = P_8 = \frac{5}{36},$$
$$P_4 = P_{10} = \frac{3}{36}, \qquad P_7 = \frac{6}{36}.$$

◀

Summary

A simple sample space is a finite sample space S such that every outcome in S has the same probability. If there are n outcomes in a simple sample space S, then each one must have probability $1/n$. The probability of an event E in a simple sample space is the number of outcomes in E divided by n. In the next three sections, we will present some useful methods for counting numbers of outcomes in various events.

Exercises

1. If two balanced dice are rolled, what is the probability that the sum of the two numbers that appear will be odd?

2. If two balanced dice are rolled, what is the probability that the sum of the two numbers that appear will be even?

3. If two balanced dice are rolled, what is the probability that the difference between the two numbers that appear will be less than 3?

4. A school contains students in grades 1, 2, 3, 4, 5, and 6. Grades 2, 3, 4, 5, and 6 all contain the same number of students, but there are twice this number in grade 1. If a student is selected at random from a list of all the students in the school, what is the probability that she will be in grade 3?

5. For the conditions of Exercise 4, what is the probability that the selected student will be in an odd-numbered grade?

6. If three fair coins are tossed, what is the probability that all three faces will be the same?

7. Consider the setup of Example 1.6.4 on page 23. This time, assume that two parents have genotypes Aa and aa. Find the possible genotypes for an offspring and find the probabilities for each genotype. Assume that all possible results of the parents contributing pairs of alleles are equally likely.

8. Consider an experiment in which a fair coin is tossed once and a balanced die is rolled once.

a. Describe the sample space for this experiment.

b. What is the probability that a head will be obtained on the coin and an odd number will be obtained on the die?

1.7 Counting Methods

In simple sample spaces, one way to calculate the probability of an event involves counting the number of outcomes in the event and the number of outcomes in the sample space. This section presents some common methods for counting the number of outcomes in a set. These methods rely on special structure that exists in many common experiments, namely, that each outcome consists of several parts and that it is relatively easy to count how many possibilities there are for each of the parts.

We have seen that in a simple sample space S, the probability of an event A is the ratio of the number of outcomes in A to the total number of outcomes in S. In many experiments, the number of outcomes in S is so large that a complete listing of these outcomes is too expensive, too slow, or too likely to be incorrect to be useful. In such an experiment, it is convenient to have a method of determining the total number of outcomes in the space S and in various events in S without compiling a list of all these outcomes. In this section, some of these methods will be presented.

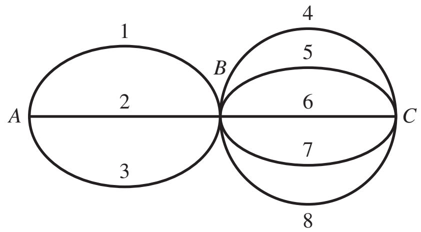


Figure 1.10 Three cities with routes between them in Example 1.7.1.

Multiplication Rule

Example 1.7.1 Routes between Cities. Suppose that there are three different routes from city A to city B and five different routes from city B to city C. The cities and routes are depicted in Fig. 1.10, with the routes numbered from 1 to 8. We wish to count the number of different routes from A to C that pass through B. For example, one such route from Fig. 1.10 is 1 followed by 4, which we can denote (1, 4). Similarly, there are the routes (1, 5), (1, 6), . . . , (3, 8). It is not difficult to see that the number of different routes $3 \times 5 = 15$. ◀

Example 1.7.1 is a special case of a common form of experiment.

Example 1.7.2 Experiment in Two Parts. Consider an experiment that has the following two characteristics:

i. The experiment is performed in two parts.
ii. The first part of the experiment has m possible outcomes $x_1, \ldots, x_m$, and, regardless of which one of these outcomes x_i occurs, the second part of the experiment has n possible outcomes $y_1, \ldots, y_n$.

Each outcome in the sample space S of such an experiment will therefore be a pair having the form (x_i, y_j), and S will be composed of the following pairs:

$$\begin{array}{l} (x_1, y_1)(x_1, y_2) \cdots (x_1, y_n) \\ (x_2, y_1)(x_2, y_2) \cdots (x_2, y_n) \\ \quad \vdots \qquad \vdots \qquad\quad \vdots \\ (x_m, y_1)(x_m, y_2) \cdots (x_m, y_n). \end{array}$$ ◀

Since each of the m rows in the array in Example 1.7.2 contains n pairs, the following result follows directly.

Theorem 1.7.1 Multiplication Rule for Two-Part Experiments. In an experiment of the type described in Example 1.7.2, the sample space S contains exactly mn outcomes. ■

Figure 1.11 illustrates the multiplication rule for the case of $n = 3$ and $m = 2$ with a tree diagram. Each end-node of the tree represents an outcome, which is the pair consisting of the two parts whose names appear along the branch leading to the end-node.

Example 1.7.3 Rolling Two Dice. Suppose that two dice are rolled. Since there are six possible outcomes for each die, the number of possible outcomes for the experiment is $6 \times 6 = 36$, as we saw in Example 1.6.5. ◀

The multiplication rule can be extended to experiments with more than two parts.

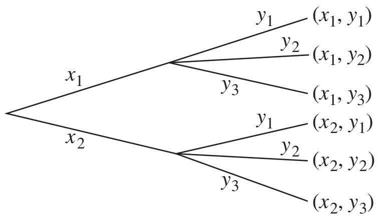


Figure 1.11 Tree diagram in which end-nodes represent outcomes.

Theorem 1.7.2 Multiplication Rule. Suppose that an experiment has k parts ($k \geq 2$), that the ith part of the experiment can have n_i possible outcomes ($i = 1, \ldots, k$), and that all of the outcomes in each part can occur regardless of which specific outcomes have occurred in the other parts. Then the sample space S of the experiment will contain all vectors of the form $(u_1, \ldots, u_k)$, where u_i is one of the n_i possible outcomes of part i ($i = 1, \ldots, k$). The total number of these vectors in S will be equal to the product $n_1 n_2 \cdots n_k$. ■

Example 1.7.4 Tossing Several Coins. Suppose that we toss six coins. Each outcome in S will consist of a sequence of six heads and tails, such as HTTHHH. Since there are two possible outcomes for each of the six coins, the total number of outcomes in S will be $2^6 = 64$. If head and tail are considered equally likely for each coin, then S will be a simple sample space. Since there is only one outcome in S with six heads and no tails, the probability of obtaining heads on all six coins is 1/64. Since there are six outcomes in S with one head and five tails, the probability of obtaining exactly one head is $6/64 = 3/32$. ◄

Example 1.7.5 Combination Lock. A standard combination lock has a dial with tick marks for 40 numbers from 0 to 39. The combination consists of a sequence of three numbers that must be dialed in the correct order to open the lock. Each of the 40 numbers may appear in each of the three positions of the combination regardless of what the other two positions contain. It follows that there are $40^3 = 64{,}000$ possible combinations. This number is supposed to be large enough to discourage would-be thieves from trying every combination. ◄

Note: The Multiplication Rule Is Slightly More General. In the statements of Theorems 1.7.1 and 1.7.2, it is assumed that each possible outcome in each part of the experiment can occur regardless of what occurs in the other parts of the experiment. Technically, all that is necessary is that the *number* of possible outcomes for each part of the experiment not depend on what occurs on the other parts. The discussion of permutations below is an example of this situation.

Permutations

Example 1.7.6 Sampling without Replacement. Consider an experiment in which a card is selected and removed from a deck of n different cards, a second card is then selected and removed from the remaining $n - 1$ cards, and finally a third card is selected from the remaining $n - 2$ cards. Each outcome consists of the three cards in the order selected. A process of this kind is called *sampling without replacement*, since a card that is drawn is not replaced in the deck before the next card is selected. In this experiment, any one of the n cards could be selected first. Once this card has been removed, any one of the other $n - 1$ cards could be selected second. Therefore, there are $n(n - 1)$

possible outcomes for the first two selections. Finally, for every given outcome of the first two selections, there are $n-2$ other cards that could possibly be selected third. Therefore, the total number of possible outcomes for all three selections is $n(n-1)(n-2)$. ◀

The situation in Example 1.7.6 can be generalized to any number of selections without replacement.

Definition 1.7.1 Permutations. Suppose that a set has n elements. Suppose that an experiment consists of selecting k of the elements one at a time without replacement. Let each outcome consist of the k elements in the order selected. Each such outcome is called a *permutation of n elements taken k at a time*. We denote the number of distinct such permutations by the symbol $P_{n,k}$.

By arguing as in Example 1.7.6, we can figure out how many different permutations there are of n elements taken k at a time. The proof of the following theorem is simply to extend the reasoning in Example 1.7.6 to selecting k cards without replacement. The proof is left to the reader.

Theorem 1.7.3 Number of Permutations. The number of permutations of n elements taken k at a time is $P_{n,k} = n(n-1)\cdots(n-k+1)$. ■

Example 1.7.7 Current Population Survey. Theorem 1.7.3 allows us to count the number of points in the sample space of Example 1.6.1. Each outcome in S consists of a permutation of $n = 50{,}000$ elements taken $k = 3$ at a time. Hence, the sample space S in that example consisits of

$$50{,}000 \times 49{,}999 \times 49{,}998 = 1.25 \times 10^{14}$$

outcomes. ◀

When $k = n$, the number of possible permutations will be the number $P_{n,n}$ of different permutations of all n cards. It is seen from the equation just derived that

$$P_{n,n} = n(n-1)\cdots 1 = n!$$

The symbol $n!$ is read *n factorial*. In general, the number of permutations of n different items is $n!$.

The expression for $P_{n,k}$ can be rewritten in the following alternate form for $k = 1, \ldots, n-1$:

$$P_{n,k} = n(n-1)\cdots(n-k+1)\frac{(n-k)(n-k-1)\cdots 1}{(n-k)(n-k-1)\cdots 1} = \frac{n!}{(n-k)!}.$$

Here and elsewhere in the theory of probability, it is convenient to define $0!$ by the relation

$$0! = 1.$$

With this definition, it follows that the relation $P_{n,k} = n!/(n-k)!$ will be correct for the value $k = n$ as well as for the values $k = 1, \ldots, n-1$. To summarize:

Theorem 1.7.4 Permutations. The number of distinct orderings of k items selected without replacement from a collection of n different items $(0 \le k \le n)$ is

$$P_{n,k} = \frac{n!}{(n-k)!}.$$ ■

Example 1.7.8

Choosing Officers. Suppose that a club consists of 25 members and that a president and a secretary are to be chosen from the membership. We shall determine the total possible number of ways in which these two positions can be filled.

Since the positions can be filled by first choosing one of the 25 members to be president and then choosing one of the remaining 24 members to be secretary, the possible number of choices is $P_{25,2} = (25)(24) = 600$. ◀

Example 1.7.9

Arranging Books. Suppose that six different books are to be arranged on a shelf. The number of possible permutations of the books is $6! = 720$. ◀

Example 1.7.10

Sampling with Replacement. Consider a box that contains n balls numbered $1, \ldots, n$. First, one ball is selected at random from the box and its number is noted. This ball is then put back in the box and another ball is selected (it is possible that the same ball will be selected again). As many balls as desired can be selected in this way. This process is called *sampling with replacement*. It is assumed that each of the n balls is equally likely to be selected at each stage and that all selections are made independently of each other.

Suppose that a total of k selections are to be made, where k is a given positive integer. Then the sample space S of this experiment will contain all vectors of the form $(x_1, \ldots, x_k)$, where x_i is the outcome of the ith selection $(i = 1, \ldots, k)$. Since there are n possible outcomes for each of the k selections, the total number of vectors in S is n^k. Furthermore, from our assumptions it follows that S is a simple sample space. Hence, the probability assigned to each vector in S is $1/n^k$. ◀

Example 1.7.11

Obtaining Different Numbers. For the experiment in Example 1.7.10, we shall determine the probability of the event E that each of the k balls that are selected will have a different number.

If $k > n$, it is impossible for all the selected balls to have different numbers because there are only n different numbers. Suppose, therefore, that $k \leq n$. The number of outcomes in the event E is the number of vectors for which all k components are different. This equals $P_{n,k}$, since the first component x_1 of each vector can have n possible values, the second component x_2 can then have any one of the other $n - 1$ values, and so on. Since S is a simple sample space containing n^k vectors, the probability p that k different numbers will be selected is

$$p = \frac{P_{n,k}}{n^k} = \frac{n!}{(n-k)!n^k}.$$ ◀

Note: Using Two Different Methods in the Same Problem. Example 1.7.11 illustrates a combination of techniques that might seem confusing at first. The method used to count the number of outcomes in the sample space was based on sampling with replacement, since the experiment allows repeat numbers in each outcome. The method used to count the number of outcomes in the event E was permutations (sampling without replacement) because E consists of those outcomes without repeats. It often happens that one needs to use different methods to count the numbers of outcomes in different subsets of the sample space. The birthday problem, which follows, is another example in which we need more than one counting method in the same problem.

The Birthday Problem

In the following problem, which is often called the birthday problem, it is required to determine the probability p that at least two people in a group of k people will have the same birthday, that is, will have been born on the same day of the same month but not necessarily in the same year. For the solution presented here, we assume that the birthdays of the k people are unrelated (in particular, we assume that twins are not present) and that each of the 365 days of the year is equally likely to be the birthday of any person in the group. In particular, we ignore the fact that the birth rate actually varies during the year and we assume that anyone actually born on February 29 will consider his birthday to be another day, such as March 1.

When these assumptions are made, this problem becomes similar to the one in Example 1.7.11. Since there are 365 possible birthdays for each of k people, the sample space S will contain 365^k outcomes, all of which will be equally probable. If $k > 365$, there are not enough birthdays for every one to be different, and hence at least two people *must* have the same birthday. So, we assume that $k \le 365$. Counting the number of outcomes in which at least two birthdays are the same is tedious. However, the number of outcomes in S for which all k birthdays will be different is $P_{365,k}$, since the first person's birthday could be any one of the 365 days, the second person's birthday could then be any of the other 364 days, and so on. Hence, the probability that all k persons will have different birthdays is

$$\frac{P_{365,k}}{365^k}.$$

The probability p that at least two of the people will have the same birthday is therefore

$$p = 1 - \frac{P_{365,k}}{365^k} = 1 - \frac{(365)!}{(365-k)!365^k}.$$

Numerical values of this probability p for various values of k are given in Table 1.1. These probabilities may seem surprisingly large to anyone who has not thought about them before. Many persons would guess that in order to obtain a value of p greater than 1/2, the number of people in the group would have to be about 100. However, according to Table 1.1, there would have to be only 23 people in the group. As a matter of fact, for $k = 100$ the value of p is 0.9999997.

Table 1.1 The probability p that at least two people in a group of k people will have the same birthday

k	p	k	p
5	0.027	25	0.569
10	0.117	30	0.706
15	0.253	40	0.891
20	0.411	50	0.970
22	0.476	60	0.994
23	0.507		

The calculation in this example illustrates a common technique for solving probability problems. If one wishes to compute the probability of some event A, it might be more straightforward to calculate $\Pr(A^c)$ and then use the fact that $\Pr(A) = 1 - \Pr(A^c)$. This idea is particularly useful when the event A is of the form "at least n things happen" where n is small compared to how many things could happen.

Stirling's Formula

For large values of n, it is nearly impossible to compute $n!$. For $n \ge 70$, $n! > 10^{100}$ and cannot be represented on many scientific calculators. In most cases for which $n!$ is needed with a large value of n, one only needs the ratio of $n!$ to another large number a_n. A common example of this is $P_{n,k}$ with large n and not so large k, which equals $n!/(n-k)!$. In such cases, we can notice that

$$\frac{n!}{a_n} = e^{\log(n!) - \log(a_n)}.$$

Compared to computing $n!$, it takes a much larger n before $\log(n!)$ becomes difficult to represent. Furthermore, if we had a simple approximation s_n to $\log(n!)$ such that $\lim_{n\to\infty} |s_n - \log(n!)| = 0$, then the ratio of $n!/a_n$ to s_n/a_n would be close to 1 for large n. The following result, whose proof can be found in Feller (1968), provides such an approximation.

Theorem 1.7.5 Stirling's Formula. Let

$$s_n = \frac{1}{2}\log(2\pi) + \left(n + \frac{1}{2}\right)\log(n) - n.$$

Then $\lim_{n\to\infty} |s_n - \log(n!)| = 0$. Put another way,

$$\lim_{n\to\infty} \frac{(2\pi)^{1/2} n^{n+1/2} e^{-n}}{n!} = 1.$$ ■

Example 1.7.12 Approximating the Number of Permutations. Suppose that we want to compute $P_{70,20} = 70!/50!$. The approximation from Stirling's formula is

$$\frac{70!}{50!} \approx \frac{(2\pi)^{1/2} 70^{70.5} e^{-70}}{(2\pi)^{1/2} 50^{50.5} e^{-50}} = 3.940 \times 10^{35}.$$

The exact calculation yields 3.938×10^{35}. The approximation and the exact calculation differ by less than 1/10 of 1 percent. ◄

Summary

Suppose that the following conditions are met:

- Each element of a set consists of k distinguishable parts $x_1, \ldots, x_k$.
- There are n_1 possibilities for the first part x_1.
- For each $i = 2, \ldots, k$ and each combination $(x_1, \ldots, x_{i-1})$ of the first $i-1$ parts, there are n_i possibilities for the ith part x_i.

Under these conditions, there are $n_1 \cdots n_k$ elements of the set. The third condition requires only that the number of possibilities for x_i be n_i no matter what the earlier parts are. For example, for $i = 2$, it does *not* require that the same n_2 possibilities be available for x_2 regardless of what x_1 is. It only requires that the *number* of possibilities for x_2 be n_2 no matter what x_1 is. In this way, the general rule includes the multiplication rule, the calculation of permutations, and sampling with replacement as special cases. For permutations of m items k at a time, we have $n_i = m - i + 1$ for $i = 1, \ldots, k$, and the n_i possibilities for part i are just the n_i items that have not yet appeared in the first $i - 1$ parts. For sampling with replacement from m items, we have $n_i = m$ for all i, and the m possibilities are the same for every part. In the next section, we shall consider how to count elements of sets in which the parts of each element are not distinguishable.

Exercises

1. Each year starts on one of the seven days (Sunday through Saturday). Each year is either a leap year (i.e., it includes February 29) or not. How many different calendars are possible for a year?

2. Three different classes contain 20, 18, and 25 students, respectively, and no student is a member of more than one class. If a team is to be composed of one student from each of these three classes, in how many different ways can the members of the team be chosen?

3. In how many different ways can the five letters a, b, c, d, and e be arranged?

4. If a man has six different sportshirts and four different pairs of slacks, how many different combinations can he wear?

5. If four dice are rolled, what is the probability that each of the four numbers that appear will be different?

6. If six dice are rolled, what is the probability that each of the six different numbers will appear exactly once?

7. If 12 balls are thrown at random into 20 boxes, what is the probability that no box will receive more than one ball?

8. An elevator in a building starts with five passengers and stops at seven floors. If every passenger is equally likely to get off at each floor and all the passengers leave independently of each other, what is the probability that no two passengers will get off at the same floor?

9. Suppose that three runners from team A and three runners from team B participate in a race. If all six runners have equal ability and there are no ties, what is the probability that the three runners from team A will finish first, second, and third, and the three runners from team B will finish fourth, fifth, and sixth?

10. A box contains 100 balls, of which r are red. Suppose that the balls are drawn from the box one at a time, at random, without replacement. Determine **(a)** the probability that the first ball drawn will be red; **(b)** the probability that the 50th ball drawn will be red; and **(c)** the probability that the last ball drawn will be red.

11. Let n and k be positive integers such that both n and $n - k$ are large. Use Stirling's formula to write as simple an approximation as you can for $P_{n,k}$.

1.8 Combinatorial Methods

Many problems of counting the number of outcomes in an event amount to counting how many subsets of a certain size are contained in a fixed set. This section gives examples of how to do such counting and where it can arise.

Combinations

Example 1.8.1 Choosing Subsets. Consider the set $\{a, b, c, d\}$ containing the four different letters. We want to count the number of distinct subsets of size two. In this case, we can list all of the subsets of size two:

$$\{a, b\}, \quad \{a, c\}, \quad \{a, d\}, \quad \{b, c\}, \quad \{b, d\}, \quad \text{and} \quad \{c, d\}.$$

We see that there are six distinct subsets of size two. This is different from counting permutaions because $\{a, b\}$ and $\{b, a\}$ are the same subset. ◀

For large sets, it would be tedious, if not impossible, to enumerate all of the subsets of a given size and count them as we did in Example 1.8.1. However, there is a connection between counting subsets and counting permutations that will allow us to derive the general formula for the number of subsets.

Suppose that there is a set of n distinct elements from which it is desired to choose a subset containing k elements ($1 \le k \le n$). We shall determine the number of different subsets that can be chosen. In this problem, the arrangement of the elements in a subset is irrelevant and each subset is treated as a unit.

Definition 1.8.1 Combinations. Consider a set with n elements. Each subset of size k chosen from this set is called a *combination of n elements taken k at a time*. We denote the number of distinct such combinations by the symbol $C_{n,k}$.

No two combinations will consist of exactly the same elements because two subsets with the same elements are the same subset.

At the end of Example 1.8.1, we noted that two different permutations (a, b) and (b, a) both correspond to the same combination or subset $\{a, b\}$. We can think of permutations as being constructed in two steps. First, a combination of k elements is chosen out of n, and second, those k elements are arranged in a specific order. There are $C_{n,k}$ ways to choose the k elements out of n, and for each such choice there are $k!$ ways to arrange those k elements in different orders. Using the multiplication rule from Sec. 1.7, we see that the number of permutations of n elements taken k at a time is $P_{n,k} = C_{n,k}k!$; hence, we have the following.

Theorem 1.8.1 Combinations. The number of distinct subsets of size k that can be chosen from a set of size n is

$$C_{n,k} = \frac{P_{n,k}}{k!} = \frac{n!}{k!(n-k)!}. \quad \blacksquare$$

In Example 1.8.1, we see that $C_{4,2} = 4!/[2!2!] = 6$.

Example 1.8.2 Selecting a Committee. Suppose that a committee composed of eight people is to be selected from a group of 20 people. The number of different groups of people that might be on the committee is

$$C_{20,8} = \frac{20!}{8!12!} = 125{,}970. \quad ◀$$

Example 1.8.3 Choosing Jobs. Suppose that, in Example 1.8.2, the eight people in the committee each get a different job to perform on the committee. The number of ways to choose eight people out of 20 and assign them to the eight different jobs is the number of permutations of 20 elements taken eight at a time, or

$$P_{20,8} = C_{20,8} \times 8! = 125{,}970 \times 8! = 5{,}078{,}110{,}400. \quad ◀$$

Examples 1.8.2 and 1.8.3 illustrate the difference and relationship between combinations and permutations. In Example 1.8.3, we count the same group of people in a different order as a different outcome, while in Example 1.8.2, we count the same group in different orders as the same outcome. The two numerical values differ by a factor of 8!, the number of ways to reorder each of the combinations in Example 1.8.2 to get a permutation in Example 1.8.3.

Binomial Coefficients

Definition 1.8.2 Binomial Coefficients. The number $C_{n,k}$ is also denoted by the symbol $\binom{n}{k}$. That is, for $k = 0, 1, \ldots, n$,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}. \tag{1.8.1}$$

When this notation is used, this number is called a *binomial coefficient*.

The name *binomial coefficient* derives from the appearance of the symbol in the binomial theorem, whose proof is left as Exercise 20 in this section.

Theorem 1.8.2 Binomial Theorem. For all numbers x and y and each positive integer n,

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

There are a couple of useful relations between binomial coefficients.

Theorem 1.8.3 For all n,

$$\binom{n}{0} = \binom{n}{n} = 1.$$

For all n and all $k = 0, 1, \ldots, n$,

$$\binom{n}{k} = \binom{n}{n-k}.$$

Proof The first equation follows from the fact that $0! = 1$. The second equation follows from Eq. (1.8.1). The second equation can also be derived from the fact that selecting k elements to form a subset is equivalent to selecting the remaining $n - k$ elements to form the complement of the subset. ■

It is sometimes convenient to use the expression "n choose k" for the value of $C_{n,k}$. Thus, the same quantity is represented by the two different notations $C_{n,k}$ and $\binom{n}{k}$, and we may refer to this quantity in three different ways: as the number of combinations of n elements taken k at a time, as the binomial coefficient of n and k, or simply as "n choose k."

Example 1.8.4 Blood Types. In Example 1.6.4 on page 23, we defined genes, alleles, and genotypes. The gene for human blood type consists of a pair of alleles chosen from the three alleles commonly called O, A, and B. For example, two possible combinations of alleles (called genotypes) to form a blood-type gene would be BB and AO. We will not distinguish the same two alleles in different orders, so OA represents the same genotype as AO. How many genotypes are there for blood type?

The answer could easily be found by counting, but it is an example of a more general calculation. Suppose that a gene consists of a pair chosen from a set of n different alleles. Assuming that we cannot distinguish the same pair in different orders, there are n pairs where both alleles are the same, and there are $\binom{n}{2}$ pairs where the two alleles are different. The total number of genotypes is

$$n + \binom{n}{2} = n + \frac{n(n-1)}{2} = \frac{n(n+1)}{2} = \binom{n+1}{2}.$$

For the case of blood type, we have $n = 3$, so there are

$$\binom{4}{2} = \frac{4 \times 3}{2} = 6$$

genotypes, as could easily be verified by counting. ◀

Note: Sampling with Replacement. The counting method described in Example 1.8.4 is a type of sampling with replacement that is different from the type described in Example 1.7.10. In Example 1.7.10, we sampled with replacement, but we distinguished between samples having the same balls in different orders. This could be called *ordered sampling with replacement*. In Example 1.8.4, samples containing the same genes in different orders were considered the same outcome. This could be called *unordered sampling with replacement*. The general formula for the number of unordered samples of size k with replacement from n elements is $\binom{n+k-1}{k}$, and can be derived in Exercise 19. It is possible to have k larger than n when sampling with replacement.

Example 1.8.5

Selecting Baked Goods. You go to a bakery to select some baked goods for a dinner party. You need to choose a total of 12 items. The baker has seven different types of items from which to choose, with lots of each type available. How many different boxfuls of 12 items are possible for you to choose? Here we will not distinguish the same collection of 12 items arranged in different orders in the box. This is an example of unordered sampling with replacement because we can (indeed we must) choose the same type of item more than once, but we are not distinguishing the same items in different orders. There are $\binom{7+12-1}{12} = 18{,}564$ different boxfuls. ◀

Example 1.8.5 raises an issue that can cause confusion if one does not carefully determine the elements of the sample space and carefully specify which outcomes (if any) are equally likely. The next example illustrates the issue in the context of Example 1.8.5.

Example 1.8.6

Selecting Baked Goods. Imagine two different ways of choosing a boxful of 12 baked goods selected from the seven different types available. In the first method, you choose one item at random from the seven available. Then, without regard to what item was chosen first, you choose the second item at random from the seven available. Then you continue in this way choosing the next item at random from the seven available without regard to what has already been chosen until you have chosen 12. For this method of choosing, it is natural to let the outcomes be the possible sequences of the 12 types of items chosen. The sample space would contain $7^{12} = 1.38 \times 10^{10}$ different outcomes that would be equally likely.

In the second method of choosing, the baker tells you that she has available 18,564 different boxfuls freshly packed. You then select one at random. In this case, the sample space would consist of 18,564 different equally likely outcomes.

In spite of the different sample spaces that arise in the two methods of choosing, there are some verbal descriptions that identify an event in both sample spaces. For example, both sample spaces contain an event that could be described as {all 12 items are of the same type} even though the outcomes are different types of mathematical objects in the two sample spaces. The probability that all 12 items are of the same type will actually be different depending on which method you use to choose the boxful.

In the first method, seven of the 7^{12} equally likely outcomes contain 12 of the same type of item. Hence, the probability that all 12 items are of the same type is

$7/7^{12} = 5.06 \times 10^{-10}$. In the second method, there are seven equally liklely boxes that contain 12 of the same type of item. Hence, the probability that all 12 items are of the same type is $7/18{,}564 = 3.77 \times 10^{-4}$. Before one can compute the probability for an event such as {all 12 items are of the same type}, one must be careful about defining the experiment and its outcomes. ◀

Arrangements of Elements of Two Distinct Types When a set contains only elements of two distinct types, a binomial coefficient can be used to represent the number of different arrangements of all the elements in the set. Suppose, for example, that k similar red balls and $n - k$ similar green balls are to be arranged in a row. Since the red balls will occupy k positions in the row, each different arrangement of the n balls corresponds to a different choice of the k positions occupied by the red balls. Hence, the number of different arrangements of the n balls will be equal to the number of different ways in which k positions can be selected for the red balls from the n available positions. Since this number of ways is specified by the binomial coefficient $\binom{n}{k}$, the number of different arrangements of the n balls is also $\binom{n}{k}$. In other words, the number of different arrangements of n objects consisting of k similar objects of one type and $n - k$ similar objects of a second type is $\binom{n}{k}$.

Example 1.8.7

Tossing a Coin. Suppose that a fair coin is to be tossed 10 times, and it is desired to determine (a) the probability p of obtaining exactly three heads and (b) the probability p' of obtaining three or fewer heads.

(a) The total possible number of different sequences of 10 heads and tails is 2^{10}, and it may be assumed that each of these sequences is equally probable. The number of these sequences that contain exactly three heads will be equal to the number of different arrangements that can be formed with three heads and seven tails. Here are some of those arrangements:

HHHTTTTTTT, HHTHTTTTTT, HHTTHTTTTT, TTHTHTHTTT, etc.

Each such arrangement is equivalent to a choice of where to put the 3 heads among the 10 tosses, so there are $\binom{10}{3}$ such arrangements. The probability of obtaining exactly three heads is then

$$p = \frac{\binom{10}{3}}{2^{10}} = 0.1172.$$

(b) Using the same reasoning as in part (a), the number of sequences in the sample space that contain exactly k heads ($k = 0, 1, 2, 3$) is $\binom{10}{k}$. Hence, the probability of obtaining three or fewer heads is

$$p' = \frac{\binom{10}{0} + \binom{10}{1} + \binom{10}{2} + \binom{10}{3}}{2^{10}}$$
$$= \frac{1 + 10 + 45 + 120}{2^{10}} = \frac{176}{2^{10}} = 0.1719.$$ ◀

Note: Using Two Different Methods in the Same Problem. Part (a) of Example 1.8.7 is another example of using two different counting methods in the same problem. Part (b) illustrates another general technique. In this part, we broke the event of interest into several disjoint subsets and counted the numbers of outcomes separately for each subset and then added the counts together to get the total. In many problems, it can require several applications of the same or different counting methods in order to count the number of outcomes in an event. The next example is one in which the elements of an event are formed in two parts (multiplication rule), but we need to perform separate combination calculations to determine the numbers of outcomes for each part.

Example 1.8.8

Sampling without Replacement. Suppose that a class contains 15 boys and 30 girls, and that 10 students are to be selected at random for a special assignment. We shall determine the probability p that exactly three boys will be selected.

The number of different combinations of the 45 students that might be obtained in the sample of 10 students is $\binom{45}{10}$, and the statement that the 10 students are selected at random means that each of these $\binom{45}{10}$ possible combinations is equally probable. Therefore, we must find the number of these combinations that contain exactly three boys and seven girls.

When a combination of three boys and seven girls is formed, the number of different combinations in which three boys can be selected from the 15 available boys is $\binom{15}{3}$, and the number of different combinations in which seven girls can be selected from the 30 available girls is $\binom{30}{7}$. Since each of these combinations of three boys can be paired with each of the combinations of seven girls to form a distinct sample, the number of combinations containing exactly three boys is $\binom{15}{3}\binom{30}{7}$. Therefore, the desired probability is

$$p = \frac{\binom{15}{3}\binom{30}{7}}{\binom{45}{10}} = 0.2904.$$ ◂

Example 1.8.9

Playing Cards. Suppose that a deck of 52 cards containing four aces is shuffled thoroughly and the cards are then distributed among four players so that each player receives 13 cards. We shall determine the probability that each player will receive one ace.

The number of possible different combinations of the four positions in the deck occupied by the four aces is $\binom{52}{4}$, and it may be assumed that each of these $\binom{52}{4}$ combinations is equally probable. If each player is to receive one ace, then there must be exactly one ace among the 13 cards that the first player will receive and one ace among each of the remaining three groups of 13 cards that the other three players will receive. In other words, there are 13 possible positions for the ace that the first player is to receive, 13 other possible positions for the ace that the second player is to receive, and so on. Therefore, among the $\binom{52}{4}$ possible combinations of the positions for the four aces, exactly 13^4 of these combinations will lead to the desired result. Hence, the probability p that each player will receive one ace is

$$p = \frac{13^4}{\binom{52}{4}} = 0.1055.$$ ◂

Ordered versus Unordered Samples Several of the examples in this section and the previous section involved counting the numbers of possible samples that could arise using various sampling schemes. Sometimes we treated the same collection of elements in different orders as different samples, and sometimes we treated the same elements in different orders as the same sample. In general, how can one tell which is the correct way to count in a given problem? Sometimes, the problem description will make it clear which is needed. For example, if we are asked to find the probability that the items in a sample arrive in a specified order, then we cannot even specify the event of interest unless we treat different arrangements of the same items as different outcomes. Examples 1.8.5 and 1.8.6 illustrate how different problem descriptions can lead to very different calculations.

However, there are cases in which the problem description does not make it clear whether or not one must count the same elements in different orders as different outcomes. Indeed, there are some problems that can be solved correctly both ways. Example 1.8.9 is one such problem. In that problem, we needed to decide what we would call an outcome, and then we needed to count how many outcomes were in the whole sample space S and how many were in the event E of interest. In the solution presented in Example 1.8.9, we chose as our outcomes the positions in the 52-card deck that were occupied by the four aces. We did not count different arrangements of the four aces in those four positions as different outcomes when we counted the number of outcomes in S. Hence, when we calculated the number of outcomes in E, we also did not count the different arrangements of the four aces in the four possible positions as different outcomes. In general, this is the principle that should guide the choice of counting method. If we have the choice between whether or not to count the same elements in different orders as different outcomes, then we need to make our choice and be consistent throughout the problem. If we count the same elements in different orders as different outcomes when counting the outcomes in S, we must do the same when counting the elements of E. If we do not count them as different outcomes when counting S, we should not count them as different when counting E.

Example 1.8.10

Playing Cards, Revisited. We shall solve the problem in Example 1.8.9 again, but this time, we shall distinguish outcomes with the same cards in different orders. To go to the extreme, let each outcome be a complete ordering of the 52 cards. So, there are 52! possible outcomes. How many of these have one ace in each of the four sets of 13 cards received by the four players? As before, there are 13^4 ways to choose the four positions for the four aces, one among each of the four sets of 13 cards. No matter which of these sets of positions we choose, there are 4! ways to arrange the four aces in these four positions. No matter how the aces are arranged, there are 48! ways to arrange the remaining 48 cards in the 48 remaining positions. So, there are $13^4 \times 4! \times 48!$ outcomes in the event of interest. We then calculate

$$p = \frac{13^4 \times 4! \times 48!}{52!} = 0.1055.$$ ◄

In the following example, whether one counts the same items in different orders as different outcomes is allowed to depend on which events one wishes to use.

Example 1.8.11

Lottery Tickets. In a lottery game, six numbers from 1 to 30 are drawn at random from a bin without replacement, and each player buys a ticket with six different numbers from 1 to 30. If all six numbers drawn match those on the player's ticket, the player wins. We assume that all possible draws are equally likely. One way to construct a sample space for the experiment of drawing the winning combination is to consider the possible sequences of draws. That is, each outcome consists of an ordered subset of six numbers chosen from the 30 available numbers. There are $P_{30,6} = 30!/24!$ such outcomes. With this sample space S, we can calculate probabilities for events such as

$$A = \{\text{the draw contains the numbers 1, 14, 15, 20, 23, and 27}\},$$
$$B = \{\text{one of the numbers drawn is 15}\}, \text{ and}$$
$$C = \{\text{the first number drawn is less than 10}\}.$$

There is another natural sample space, which we shall denote S', for this experiment. It consists solely of the different combinations of six numbers drawn from the 30 available. There are $\binom{30}{6} = 30!/(6!24!)$ such outcomes. It also seems natural to consider all of these outcomes equally likely. With this sample space, we can calculate the probabilities of the events A and B above, but C is not a subset of the sample space S', so we cannot calculate its probability using this smaller sample space. When the sample space for an experiment could naturally be constructed in more than one way, one needs to choose based on for which events one wants to compute probabilities. ◂

Example 1.8.11 raises the question of whether one will compute the same probabilities using two different sample spaces when the event, such as A or B, exists in both sample spaces. In the example, each outcome in the smaller sample space S' corresponds to an event in the larger sample space S. Indeed, each outcome s' in S' corresponds to the event in S containing the 6! permutations of the single combination s'. For example, the event A in the example has only one outcome $s' = (1, 14, 15, 20, 23, 27)$ in the sample space S', while the corresponding event in the sample space S has 6! permutations including

$$(1, 14, 15, 20, 23, 27),\ (14, 20, 27, 15, 23, 1),\ (27, 23, 20, 15, 14, 1),\ \text{etc.}$$

In the sample space S, the probability of the event A is

$$\Pr(A) = \frac{6!}{P_{30,6}} = \frac{6!24!}{30!} = \frac{1}{\binom{30}{6}}.$$

In the sample space S', the event A has this same probability because it has only one of the $\binom{30}{6}$ equally likely outcomes. The same reasoning applies to every outcome in S'. Hence, if the same event can be expressed in both sample spaces S and S', we will compute the same probability using either sample space. This is a special feature of examples like Example 1.8.11 in which each outcome in the smaller sample space corresponds to an event in the larger sample space with the same number of elements. There are examples in which this feature is not present, and one cannot treat both sample spaces as simple sample spaces.

Example 1.8.12

Tossing Coins. An experiment consists of tossing a coin two times. If we want to distinguish H followed by T from T followed by H, we should use the sample space $S = \{HH, HT, TH, TT\}$, which might naturally be assumed a simple sample space. On the other hand, we might be interested solely in the number of H's tossed. In this case, we might consider the smaller sample space $S' = \{0, 1, 2\}$ where each outcome merely counts the number of H's. The outcomes 0 and 2 in S' each correspond to a single outcome in S, but $1 \in S'$ corresponds to the event $\{HT, TH\} \subset S$ with two outcomes. If we think of S as a simple sample space, then S' will not be a simple sample space, because the outcome 1 will have probability 1/2 while the other two outcomes each have probability 1/4.

There are situations in which one would be justified in treating S' as a simple sample space and assigning each of its outcomes probability 1/3. One might do this if one believed that the coin was not fair, but one had no idea how unfair it was or which side were more likely to land up. In this case, S would not be a simple sample space, because two of its outcomes would have probability 1/3 and the other two would have probabilities that add up to 1/3. ◂

Example 1.8.6 is another case of two different sample spaces in which each outcome in one sample space corresponds to a different number of outcomes in the other space. See Exercise 12 in Sec. 1.9 for a more complete analysis of Example 1.8.6.

The Tennis Tournament

We shall now present a difficult problem that has a simple and elegant solution. Suppose that n tennis players are entered in a tournament. In the first round, the players are paired one against another at random. The loser in each pair is eliminated from the tournament, and the winner in each pair continues into the second round. If the number of players n is odd, then one player is chosen at random before the pairings are made for the first round, and that player automatically continues into the second round. All the players in the second round are then paired at random. Again, the loser in each pair is eliminated, and the winner in each pair continues into the third round. If the number of players in the second round is odd, then one of these players is chosen at random before the others are paired, and that player automatically continues into the third round. The tournament continues in this way until only two players remain in the final round. They then play against each other, and the winner of this match is the winner of the tournament. We shall assume that all n players have equal ability, and we shall determine the probability p that two specific players A and B will ever play against each other during the tournament.

We shall first determine the total number of matches that will be played during the tournament. After each match has been played, one player—the loser of that match—is eliminated from the tournament. The tournament ends when everyone has been eliminated from the tournament except the winner of the final match. Since exactly $n-1$ players must be eliminated, it follows that exactly $n-1$ matches must be played during the tournament.

The number of possible pairs of players is $\binom{n}{2}$. Each of the two players in every match is equally likely to win that match, and all initial pairings are made in a random manner. Therefore, before the tournament begins, every possible pair of players is equally likely to appear in each particular one of the $n-1$ matches to be played during the tournament. Accordingly, the probability that players A and B will meet in some particular match that is specified in advance is $1/\binom{n}{2}$. If A and B do meet in that particular match, one of them will lose and be eliminated. Therefore, these same two players cannot meet in more than one match.

It follows from the preceding explanation that the probability p that players A and B will meet at some time during the tournament is equal to the product of the probability $1/\binom{n}{2}$ that they will meet in any particular specified match and the total number $n-1$ of different matches in which they might possibly meet. Hence,

$$p = \frac{n-1}{\binom{n}{2}} = \frac{2}{n}.$$

Summary

We showed that the number of size k subsets of a set of size n is $\binom{n}{k} = n!/[k!(n-k)!]$. This turns out to be the number of possible samples of size k drawn without replacement from a population of size n as well as the number of arrangements of n items of two types with k of one type and $n-k$ of the other type. We also saw several examples in which more than one counting technique was required at different points in the same problem. Sometimes, more than one technique is required to count the elements of a single set.

Exercises

1. Two pollsters will canvas a neighborhood with 20 houses. Each pollster will visit 10 of the houses. How many different assignments of pollsters to houses are possible?

2. Which of the following two numbers is larger: $\binom{93}{30}$ or $\binom{93}{31}$?

3. Which of the following two numbers is larger: $\binom{93}{30}$ or $\binom{93}{63}$?

4. A box contains 24 light bulbs, of which four are defective. If a person selects four bulbs from the box at random, without replacement, what is the probability that all four bulbs will be defective?

5. Prove that the following number is an integer:

$$\frac{4155 \times 4156 \times \cdots \times 4250 \times 4251}{2 \times 3 \times \cdots \times 96 \times 97}.$$

6. Suppose that n people are seated in a random manner in a row of n theater seats. What is the probability that two particular people A and B will be seated next to each other?

7. If k people are seated in a random manner in a row containing n seats $(n > k)$, what is the probability that the people will occupy k adjacent seats in the row?

8. If k people are seated in a random manner in a circle containing n chairs $(n > k)$, what is the probability that the people will occupy k adjacent chairs in the circle?

9. If n people are seated in a random manner in a row containing $2n$ seats, what is the probability that no two people will occupy adjacent seats?

10. A box contains 24 light bulbs, of which two are defective. If a person selects 10 bulbs at random, without replacement, what is the probability that both defective bulbs will be selected?

11. Suppose that a committee of 12 people is selected in a random manner from a group of 100 people. Determine the probability that two particular people A and B will both be selected.

12. Suppose that 35 people are divided in a random manner into two teams in such a way that one team contains 10 people and the other team contains 25 people. What is the probability that two particular people A and B will be on the same team?

13. A box contains 24 light bulbs of which four are defective. If one person selects 10 bulbs from the box in a random manner, and a second person then takes the remaining 14 bulbs, what is the probability that all four defective bulbs will be obtained by the same person?

14. Prove that, for all positive integers n and k $(n \ge k)$,

$$\binom{n}{k} + \binom{n}{k-1} = \binom{n+1}{k}.$$

15.

a. Prove that

$$\binom{n}{0} + \binom{n}{1} + \binom{n}{2} + \cdots + \binom{n}{n} = 2^n.$$

b. Prove that

$$\binom{n}{0} - \binom{n}{1} + \binom{n}{2} - \binom{n}{3} + \cdots + (-1)^n \binom{n}{n} = 0.$$

Hint: Use the binomial theorem.

16. The United States Senate contains two senators from each of the 50 states. **(a)** If a committee of eight senators is selected at random, what is the probability that it will contain at least one of the two senators from a certain specified state? **(b)** What is the probability that a group of 50 senators selected at random will contain one senator from each state?

17. A deck of 52 cards contains four aces. If the cards are shuffled and distributed in a random manner to four players so that each player receives 13 cards, what is the probability that all four aces will be received by the same player?

18. Suppose that 100 mathematics students are divided into five classes, each containing 20 students, and that awards are to be given to 10 of these students. If each student is equally likely to receive an award, what is the probability that exactly two students in each class will receive awards?

19. A restaurant has n items on its menu. During a particular day, k customers will arrive and each one will choose one item. The manager wants to count how many different collections of customer choices are possible without regard to the order in which the choices are made. (For example, if $k = 3$ and $a_1, \ldots, a_n$ are the menu items, then $a_1a_3a_1$ is not distinguished from $a_1a_1a_3$.) Prove that the number of different collections of customer choices is $\binom{n+k-1}{k}$. *Hint:* Assume that the menu items are $a_1, \ldots, a_n$. Show that each collection of customer choices, arranged with the a_1's first, the a_2's second, etc., can be identified with a sequence of k zeros and $n-1$ ones, where each 0 stands for a customer choice and each 1 indicates a point in the sequence where the menu item number increases by 1. For example, if $k=3$ and $n=5$, then $a_1a_1a_3$ becomes 0011011.

20. Prove the binomial theorem 1.8.2. *Hint:* You may use an *induction* argument. That is, first prove that the result is true if $n=1$. Then, under the assumption that there is n_0 such that the result is true for all $n \le n_0$, prove that it is also true for $n=n_0+1$.

21. Return to the birthday problem on page 30. How many different sets of birthdays are available with k people and 365 days when we don't distinguish the same birthdays in different orders? For example, if $k=3$, we would count (Jan. 1, Mar. 3, Jan.1) the same as (Jan. 1, Jan. 1, Mar. 3).

22. Let n be a large even integer. Use Stirlings' formula (Theorem 1.7.5) to find an approximation to the binomial coefficient $\binom{n}{n/2}$. Compute the approximation with $n = 500$.

1.9 Multinomial Coefficients

We learn how to count the number of ways to partition a finite set into more than two disjoint subsets. This generalizes the binomial coefficients from Sec. 1.8. The generalization is useful when outcomes consist of several parts selected from a fixed number of distinct types.

We begin with a fairly simple example that will illustrate the general ideas of this section.

Example 1.9.1 Choosing Committees. Suppose that 20 members of an organization are to be divided into three committees A, B, and C in such a way that each of the committees A and B is to have eight members and committee C is to have four members. We shall determine the number of different ways in which members can be assigned to these committees. Notice that each of the 20 members gets assigned to one and only one committee.

One way to think of the assignments is to form committee A first by choosing its eight members and then split the remaining 12 members into committees B and C. Each of these operations is choosing a combination, and every choice of committee A can be paired with every one of the splits of the remaining 12 members into committees B and C. Hence, the number of assignments into three committees is the product of the numbers of combinations for the two parts of the assignment. Specifically, to form committee A, we must choose eight out of 20 members, and this can be done in $\binom{20}{8}$ ways. Then to split the remaining 12 members into committees B and C there are are $\binom{12}{8}$ ways to do it. Here, the answer is

$$\binom{20}{8}\binom{12}{8} = \frac{20!}{8!12!}\frac{12!}{8!4!} = \frac{20!}{8!8!4!} = 62{,}355{,}150. \blacktriangleleft$$

Notice how the 12! that appears in the denominator of $\binom{20}{8}$ divides out with the 12! that appears in the numerator of $\binom{12}{8}$. This fact is the key to the general formula that we shall derive next.

In general, suppose that n distinct elements are to be divided into k different groups ($k \ge 2$) in such a way that, for $j = 1, \ldots, k$, the jth group contains exactly n_j elements, where $n_1 + n_2 + \cdots + n_k = n$. It is desired to determine the number of different ways in which the n elements can be divided into the k groups. The

n_1 elements in the first group can be selected from the n available elements in $\binom{n}{n_1}$ different ways. After the n_1 elements in the first group have been selected, the n_2 elements in the second group can be selected from the remaining $n - n_1$ elements in $\binom{n-n_1}{n_2}$ different ways. Hence, the total number of different ways of selecting the elements for both the first group and the second group is $\binom{n}{n_1}\binom{n-n_1}{n_2}$. After the $n_1 + n_2$ elements in the first two groups have been selected, the number of different ways in which the n_3 elements in the third group can be selected is $\binom{n-n_1-n_2}{n_3}$. Hence, the total number of different ways of selecting the elements for the first three groups is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}.$$

It follows from the preceding explanation that, for each $j = 1, \ldots, k-2$ after the first j groups have been formed, the number of different ways in which the n_{j+1} elements in the next group $(j+1)$ can be selected from the remaining $n - n_1 - \cdots - n_j$ elements is $\binom{n-n_1-\cdots-n_j}{n_{j+1}}$. After the elements of group $k-1$ have been selected, the remaining n_k elements must then form the last group. Hence, the total number of different ways of dividing the n elements into the k groups is

$$\binom{n}{n_1}\binom{n-n_1}{n_2}\binom{n-n_1-n_2}{n_3}\cdots\binom{n-n_1-\cdots-n_{k-2}}{n_{k-1}} = \frac{n!}{n_1!n_2!\cdots n_k!},$$

where the last formula follows from writing the binomial coefficients in terms of factorials.

Definition 1.9.1 Multinomial Coefficients. The number

$$\frac{n!}{n_1!n_2!\cdots n_k!}, \quad \text{which we shall denote by} \quad \binom{n}{n_1, n_2, \ldots, n_k},$$

is called a *multinomial coefficient*.

The name *multinomial coefficient* derives from the appearance of the symbol in the multinomial theorem, whose proof is left as Exercise 11 in this section.

Theorem 1.9.1 Multinomial Theorem. For all numbers $x_1, \ldots, x_k$ and each positive integer n,

$$(x_1 + \cdots + x_k)^n = \sum \binom{n}{n_1, n_2, \ldots, n_k} x_1^{n_1} x_2^{n_2} \cdots x_k^{n_k},$$

where the summation extends over all possible combinations of nonnegative integers $n_1, \ldots, n_k$ such that $n_1 + n_2 + \cdots + n_k = n$. ■

A multinomial coefficient is a generalization of the binomial coefficient discussed in Sec. 1.8. For $k = 2$, the multinomial theorem is the same as the binomial theorem, and the multinomial coefficient becomes a binomial coefficient. In particular,

$$\binom{n}{k, n-k} = \binom{n}{k}.$$

Example 1.9.2 Choosing Committees. In Example 1.9.1, we see that the solution obtained there is the same as the multinomial coefficient for which $n = 20$, $k = 3$, $n_1 = n_2 = 8$, and $n_3 = 4$, namely,

$$\binom{20}{8, 8, 4} = \frac{20!}{(8!)^2 4!} = 62{,}355{,}150.$$ ◄

Arrangements of Elements of More Than Two Distinct Types Just as binomial coefficients can be used to represent the number of different arrangements of the elements of a set containing elements of only two distinct types, multinomial coefficients can be used to represent the number of different arrangements of the elements of a set containing elements of k different types ($k \geq 2$). Suppose, for example, that n balls of k different colors are to be arranged in a row and that there are n_j balls of color j ($j = 1, \ldots, k$), where $n_1 + n_2 + \cdots + n_k = n$. Then each different arrangement of the n balls corresponds to a different way of dividing the n available positions in the row into a group of n_1 positions to be occupied by the balls of color 1, a second group of n_2 positions to be occupied by the balls of color 2, and so on. Hence, the total number of different possible arrangements of the n balls must be

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

Example 1.9.3 Rolling Dice. Suppose that 12 dice are to be rolled. We shall determine the probability p that each of the six different numbers will appear twice.

Each outcome in the sample space S can be regarded as an ordered sequence of 12 numbers, where the ith number in the sequence is the outcome of the ith roll. Hence, there will be 6^{12} possible outcomes in S, and each of these outcomes can be regarded as equally probable. The number of these outcomes that would contain each of the six numbers $1, 2, \ldots, 6$ exactly twice will be equal to the number of different possible arrangements of these 12 elements. This number can be determined by evaluating the multinomial coefficient for which $n = 12$, $k = 6$, and $n_1 = n_2 = \cdots = n_6 = 2$. Hence, the number of such outcomes is

$$\binom{12}{2, 2, 2, 2, 2, 2} = \frac{12!}{(2!)^6},$$

and the required probability p is

$$p = \frac{12!}{2^6 6^{12}} = 0.0034. \qquad ◀$$

Example 1.9.4 Playing Cards. A deck of 52 cards contains 13 hearts. Suppose that the cards are shuffled and distributed among four players A, B, C, and D so that each player receives 13 cards. We shall determine the probability p that player A will receive six hearts, player B will receive four hearts, player C will receive two hearts, and player D will receive one heart.

The total number N of different ways in which the 52 cards can be distributed among the four players so that each player receives 13 cards is

$$N = \binom{52}{13, 13, 13, 13} = \frac{52!}{(13!)^4}.$$

It may be assumed that each of these ways is equally probable. We must now calculate the number M of ways of distributing the cards so that each player receives the required number of hearts. The number of different ways in which the hearts can be distributed to players A, B, C, and D so that the numbers of hearts they receive are 6, 4, 2, and 1, respectively, is

$$\binom{13}{6, 4, 2, 1} = \frac{13!}{6!4!2!1!}.$$

Also, the number of different ways in which the other 39 cards can then be distributed to the four players so that each will have a total of 13 cards is

$$\binom{39}{7, 9, 11, 12} = \frac{39!}{7!9!11!12!}.$$

Therefore,

$$M = \frac{13!}{6!4!2!1!} \cdot \frac{39!}{7!9!11!12!},$$

and the required probability p is

$$p = \frac{M}{N} = \frac{13!39!(13!)^4}{6!4!2!1!7!9!11!12!52!} = 0.00196.$$

There is another approach to this problem along the lines indicated in Example 1.8.9 on page 37. The number of possible different combinations of the 13 positions in the deck occupied by the hearts is $\binom{52}{13}$. If player A is to receive six hearts, there are $\binom{13}{6}$ possible combinations of the six positions these hearts occupy among the 13 cards that A will receive. Similarly, if player B is to receive four hearts, there are $\binom{13}{4}$ possible combinations of their positions among the 13 cards that B will receive. There are $\binom{13}{2}$ possible combinations for player C, and there are $\binom{13}{1}$ possible combinations for player D. Hence,

$$p = \frac{\binom{13}{6}\binom{13}{4}\binom{13}{2}\binom{13}{1}}{\binom{52}{13}},$$

which produces the same value as the one obtained by the first method of solution. ◀

Summary

Multinomial coefficients generalize binomial coefficients. The coefficient $\binom{n}{n_1, \ldots, n_k}$ is the number of ways to partition a set of n items into distinguishable subsets of sizes $n_1, \ldots, n_k$ where $n_1 + \cdots + n_k = n$. It is also the number of arrangements of n items of k different types for which n_i are of type i for $i = 1, \ldots, k$. Example 1.9.4 illustrates another important point to remember about computing probabilities: There might be more than one correct method for computing the same probability.

Exercises

1. Three pollsters will canvas a neighborhood with 21 houses. Each pollster will visit seven of the houses. How many different assignments of pollsters to houses are possible?

2. Suppose that 18 red beads, 12 yellow beads, eight blue beads, and 12 black beads are to be strung in a row. How many different arrangements of the colors can be formed?

3. Suppose that two committees are to be formed in an organization that has 300 members. If one committee is to have five members and the other committee is to have eight members, in how many different ways can these committees be selected?

4. If the letters $s, s, s, t, t, t, i, i, a, c$ are arranged in a random order, what is the probability that they will spell the word "statistics"?

5. Suppose that n balanced dice are rolled. Determine the probability that the number j will appear exactly n_j times $(j = 1, \ldots, 6)$, where $n_1 + n_2 + \ldots + n_6 = n$.

6. If seven balanced dice are rolled, what is the probability that each of the six different numbers will appear at least once?

7. Suppose that a deck of 25 cards contains 12 red cards. Suppose also that the 25 cards are distributed in a random manner to three players A, B, and C in such a way that player A receives 10 cards, player B receives eight cards, and player C receives seven cards. Determine the probability that player A will receive six red cards, player B will receive two red cards, and player C will receive four red cards.

8. A deck of 52 cards contains 12 picture cards. If the 52 cards are distributed in a random manner among four players in such a way that each player receives 13 cards, what is the probability that each player will receive three picture cards?

9. Suppose that a deck of 52 cards contains 13 red cards, 13 yellow cards, 13 blue cards, and 13 green cards. If the 52 cards are distributed in a random manner among four players in such a way that each player receives 13 cards, what is the probability that each player will receive 13 cards of the same color?

10. Suppose that two boys named Davis, three boys named Jones, and four boys named Smith are seated at random in a row containing nine seats. What is the probability that the Davis boys will occupy the first two seats in the row, the Jones boys will occupy the next three seats, and the Smith boys will occupy the last four seats?

11. Prove the multinomial theorem 1.9.1. (You may wish to use the same hint as in Exercise 20 in Sec. 1.8.)

12. Return to Example 1.8.6. Let S be the larger sample space (first method of choosing) and let S' be the smaller sample space (second method). For each element s' of S', let $N(s')$ stand for the number of elements of S that lead to the same boxful s' when the order of choosing is ignored.

a. For each $s' \in S'$, find a formula for $N(s')$. *Hint:* Let n_i stand for the number of items of type i in s' for $i = 1, \ldots, 7$.

b. Verify that $\sum_{s' \in S'} N(s')$ equals the number of outcomes in S.

1.10 The Probability of a Union of Events

The axioms of probability tell us directly how to find the probability of the union of disjoint events. Theorem 1.5.7 showed how to find the probability for the union of two arbitrary events. This theorem is generalized to the union of an arbitrary finite collection of events.

We shall now consider again an arbitrary sample space S that may contain either a finite number of outcomes or an infinite number, and we shall develop some further general properties of the various probabilities that might be specified for the events in S. In this section, we shall study in particular the probability of the union $\bigcup_{i=1}^{n} A_i$ of n events $A_1, \ldots, A_n$.

If the events $A_1, \ldots, A_n$ are disjoint, we know that

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

Furthermore, for every two events A_1 and A_2, regardless of whether or not they are disjoint, we know from Theorem 1.5.7 of Sec. 1.5 that

$$\Pr(A_1 \cup A_2) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2).$$

In this section, we shall extend this result, first to three events and then to an arbitrary finite number of events.

The Union of Three Events

Theorem 1.10.1 For every three events A_1, A_2, and A_3,

$$\Pr(A_1 \cup A_2 \cup A_3) = \Pr(A_1) + \Pr(A_2) + \Pr(A_3) - [\Pr(A_1 \cap A_2) + \Pr(A_2 \cap A_3) + \Pr(A_1 \cap A_3)] + \Pr(A_1 \cap A_2 \cap A_3). \quad (1.10.1)$$

Proof By the associative property of unions (Theorem 1.4.6), we can write

$$A_1 \cup A_2 \cup A_3 = (A_1 \cup A_2) \cup A_3.$$

Apply Theorem 1.5.7 to the two events $A = A_1 \cup A_2$ and $B = A_3$ to obtain

$$\Pr(A_1 \cup A_2 \cup A_3) = \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \quad (1.10.2)$$

We next compute the three probabilities on the far right side of (1.10.2) and combine them to get (1.10.1). First, apply Theorem 1.5.7 to the two events A_1 and A_2 to obtain

$$\Pr(A) = \Pr(A_1) + \Pr(A_2) - \Pr(A_1 \cap A_2). \quad (1.10.3)$$

Next, use the first distributive property in Theorem 1.4.10 to write

$$A \cap B = (A_1 \cup A_2) \cap A_3 = (A_1 \cap A_3) \cup (A_2 \cap A_3). \quad (1.10.4)$$

Apply Theorem 1.5.7 to the events on the far right side of (1.10.4) to obtain

$$\Pr(A \cap B) = \Pr(A_1 \cap A_3) + \Pr(A_2 \cap A_3) - \Pr(A_1 \cap A_2 \cap A_3). \quad (1.10.5)$$

Substitute (1.10.3), $\Pr(B) = \Pr(A_3)$, and (1.10.5) into (1.10.2) to complete the proof. ■

Example 1.10.1 Student Enrollment. Among a group of 200 students, 137 students are enrolled in a mathematics class, 50 students are enrolled in a history class, and 124 students are enrolled in a music class. Furthermore, the number of students enrolled in both the mathematics and history classes is 33, the number enrolled in both the history and music classes is 29, and the number enrolled in both the mathematics and music classes is 92. Finally, the number of students enrolled in all three classes is 18. We shall determine the probability that a student selected at random from the group of 200 students will be enrolled in at least one of the three classes.

Let A_1 denote the event that the selected student is enrolled in the mathematics class, let A_2 denote the event that he is enrolled in the history class, and let A_3 denote the event that he is enrolled in the music class. To solve the problem, we must determine the value of $\Pr(A_1 \cup A_2 \cup A_3)$. From the given numbers,

$$\Pr(A_1) = \frac{137}{200}, \quad \Pr(A_2) = \frac{50}{200}, \quad \Pr(A_3) = \frac{124}{200},$$

$$\Pr(A_1 \cap A_2) = \frac{33}{200}, \quad \Pr(A_2 \cap A_3) = \frac{29}{200}, \quad \Pr(A_1 \cap A_3) = \frac{92}{200},$$

$$\Pr(A_1 \cap A_2 \cap A_3) = \frac{18}{200}.$$

It follows from Eq. (1.10.1) that $\Pr(A_1 \cup A_2 \cup A_3) = 175/200 = 7/8$. ◂

The Union of a Finite Number of Events

A result similar to Theorem 1.10.1 holds for any arbitrary finite number of events, as shown by the following theorem.

Theorem 1.10.2 For every n events $A_1, \ldots, A_n$,

$$\begin{aligned}\Pr\left(\bigcup_{i=1}^{n} A_i\right) = &\sum_{i=1}^{n} \Pr(A_i) - \sum_{i<j} \Pr(A_i \cap A_j) + \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) \\ &- \sum_{i<j<k<l} \Pr(A_i \cap A_j \cap A_k \cap A_l) + \cdots \\ &+ (-1)^{n+1} \Pr(A_1 \cap A_2 \cap \cdots \cap A_n).\end{aligned} \tag{1.10.6}$$

Proof The proof proceeds by induction. In particular, we first establish that (1.10.6) is true for $n = 1$ and $n = 2$. Next, we show that if there exists m such that (1.10.6) is true for all $n \le m$, then (1.10.6) is also true for $n = m + 1$. The case of $n = 1$ is trivial, and the case of $n = 2$ is Theorem 1.5.7. To complete the proof, assume that (1.10.6) is true for all $n \le m$. Let $A_1, \ldots, A_{m+1}$ be events. Define $A = \bigcup_{i=1}^{m} A_i$ and $B = A_{m+1}$. Theorem 1.5.7 says that

$$\Pr\left(\bigcup_{i=1}^{n} A_i\right) = \Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B). \tag{1.10.7}$$

We have assumed that $\Pr(A)$ equals (1.10.6) with $n = m$. We need to show that when we add $\Pr(A)$ to $\Pr(B) - \Pr(A \cap B)$, we get (1.10.6) with $n = m + 1$. The difference between (1.10.6) with $n = m + 1$ and $\Pr(A)$ is all of the terms in which one of the subscripts (i, j, k, etc.) equals $m + 1$. Those terms are the following:

$$\begin{aligned}&\Pr(A_{m+1}) - \sum_{i=1}^{m} \Pr(A_i \cap A_{m+1}) + \sum_{i<j} \Pr(A_i \cap A_j \cap A_{m+1}) \\ &- \sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k \cap A_{m+1}) + \cdots \\ &+ (-1)^{m+2} \Pr(A_1 \cap A_2 \cap \cdots \cap A_m \cap A_{m+1}).\end{aligned} \tag{1.10.8}$$

The first term in (1.10.8) is $\Pr(B) = \Pr(A_{m+1})$. All that remains is to show that $-\Pr(A \cap B)$ equals all but the first term in (1.10.8).

Use the natural generalization of the distributive property (Theorem 1.4.10) to write

$$A \cap B = \left(\bigcup_{i=1}^{m} A_i\right) \cap A_{m+1} = \bigcup_{i=1}^{m} (A_i \cap A_{m+1}). \tag{1.10.9}$$

The union in (1.10.9) contains m events, and hence we can apply (1.10.6) with $n = m$ and each A_i replaced by $A_i \cap A_{m+1}$. The result is that $-\Pr(A \cap B)$ equals all but the first term in (1.10.8). ■

The calculation in Theorem 1.10.2 can be outlined as follows: First, take the sum of the probabilities of the n individual events. Second, subtract the sum of the probabilities of the intersections of all possible pairs of events; in this step, there will be $\binom{n}{2}$ different pairs for which the probabilities are included. Third, add the probabilities of the intersections of all possible groups of three of the events; there will be $\binom{n}{3}$ intersections of this type. Fourth, subtract the sum of the probabilities of the intersections of all possible groups of four of the events; there will be $\binom{n}{4}$ intersections of this type. Continue in this way until, finally, the probability of the intersection of all n events is either added or subtracted, depending on whether n is an odd number or an even number.

The Matching Problem

Suppose that all the cards in a deck of n different cards are placed in a row, and that the cards in another similar deck are then shuffled and placed in a row on top of the cards in the original deck. It is desired to determine the probability p_n that there will be at least one match between the corresponding cards from the two decks. The same problem can be expressed in various entertaining contexts. For example, we could suppose that a person types n letters, types the corresponding addresses on n envelopes, and then places the n letters in the n envelopes in a random manner. It could be desired to determine the probability p_n that at least one letter will be placed in the correct envelope. As another example, we could suppose that the photographs of n famous film actors are paired in a random manner with n photographs of the same actors taken when they were babies. It could then be desired to determine the probability p_n that the photograph of at least one actor will be paired correctly with this actor's own baby photograph.

Here we shall discuss this matching problem in the context of letters being placed in envelopes. Thus, we shall let A_i be the event that letter i is placed in the correct envelope ($i = 1, \dots, n$), and we shall determine the value of $p_n = \Pr\left(\bigcup_{i=1}^{n} A_i\right)$ by using Eq. (1.10.6). Since the letters are placed in the envelopes at random, the probability $\Pr(A_i)$ that any particular letter will be placed in the correct envelope is $1/n$. Therefore, the value of the first summation on the right side of Eq. (1.10.6) is

$$\sum_{i=1}^{n} \Pr(A_i) = n \cdot \frac{1}{n} = 1.$$

Furthermore, since letter 1 could be placed in any one of n envelopes and letter 2 could then be placed in any one of the other $n-1$ envelopes, the probability $\Pr(A_1 \cap A_2)$ that both letter 1 and letter 2 will be placed in the correct envelopes is $1/[n(n-1)]$. Similarly, the probability $\Pr(A_i \cap A_j)$ that any two specific letters i and j ($i \neq j$) will both be placed in the correct envelopes is $1/[n(n-1)]$. Therefore, the value of the second summation on the right side of Eq. (1.10.6) is

$$\sum_{i<j} \Pr(A_i \cap A_j) = \binom{n}{2} \frac{1}{n(n-1)} = \frac{1}{2!}.$$

By similar reasoning, it can be determined that the probability $\Pr(A_i \cap A_j \cap A_k)$ that any three specific letters i, j, and k ($i < j < k$) will be placed in the correct envelopes is $1/[n(n-1)(n-2)]$. Therefore, the value of the third summation is

$$\sum_{i<j<k} \Pr(A_i \cap A_j \cap A_k) = \binom{n}{3} \frac{1}{n(n-1)(n-2)} = \frac{1}{3!}.$$

This procedure can be continued until it is found that the probability $\Pr(A_1 \cap A_2 \cdots \cap A_n)$ that all n letters will be placed in the correct envelopes is $1/(n!)$. It now follows from Eq. (1.10.6) that the probability p_n that at least one letter will be placed in the correct envelope is

$$p_n = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots + (-1)^{n+1}\frac{1}{n!}. \qquad (1.10.10)$$

This probability has the following interesting features. As $n \to \infty$, the value of p_n approaches the following limit:

$$\lim_{n\to\infty} p_n = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \cdots.$$

It is shown in books on elementary calculus that the sum of the infinite series on the right side of this equation is $1 - (1/e)$, where $e = 2.71828\ldots$. Hence, $1 - (1/e) = 0.63212\ldots$. It follows that for a large value of n, the probability p_n that at least one letter will be placed in the correct envelope is approximately 0.63212.

The exact values of p_n, as given in Eq. (1.10.10), will form an oscillating sequence as n increases. As n increases through the even integers $2, 4, 6, \ldots$, the values of p_n will increase toward the limiting value 0.63212; and as n increases through the odd integers $3, 5, 7, \ldots$, the values of p_n will decrease toward this same limiting value.

The values of p_n converge to the limit very rapidly. In fact, for $n = 7$ the exact value p_7 and the limiting value of p_n agree to four decimal places. Hence, regardless of whether seven letters are placed at random in seven envelopes or seven million letters are placed at random in seven million envelopes, the probability that at least one letter will be placed in the correct envelope is 0.6321.

Summary

We generalized the formula for the probability of the union of two arbitrary events to the union of finitely many events. As an aside, there are cases in which it is easier to compute $\Pr(A_1 \cup \ldots \cup A_n)$ as $1 - \Pr(A_1^c \cap \cdots \cap A_n^c)$ using the fact that $(A_1 \cup \ldots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c$.

Exercises

1. Three players are each dealt, in a random manner, five cards from a deck containing 52 cards. Four of the 52 cards are aces. Find the probability that at least one person receives exactly two aces in their five cards.

2. In a certain city, three newspapers A, B, and C are published. Suppose that 60 percent of the families in the city subscribe to newspaper A, 40 percent of the families subscribe to newspaper B, and 30 percent subscribe to newspaper C. Suppose also that 20 percent of the families subscribe to both A and B, 10 percent subscribe to both A and C, 20 percent subscribe to both B and C, and 5 percent subscribe to all three newspapers A, B, and C. What percentage of the families in the city subscribe to at least one of the three newspapers?

3. For the conditions of Exercise 2, what percentage of the families in the city subscribe to exactly one of the three newspapers?

4. Suppose that three compact discs are removed from their cases, and that after they have been played, they are put back into the three empty cases in a random manner. Determine the probability that at least one of the CD's will be put back into the proper cases.

5. Suppose that four guests check their hats when they arrive at a restaurant, and that these hats are returned to them in a random order when they leave. Determine the probability that no guest will receive the proper hat.

6. A box contains 30 red balls, 30 white balls, and 30 blue balls. If 10 balls are selected at random, without replacement, what is the probability that at least one color will be missing from the selection?

7. Suppose that a school band contains 10 students from the freshman class, 20 students from the sophomore class, 30 students from the junior class, and 40 students from the senior class. If 15 students are selected at random from the band, what is the probability that at least one student will be selected from each of the four classes? *Hint:* First determine the probability that at least one of the four classes will not be represented in the selection.

8. If n letters are placed at random in n envelopes, what is the probability that exactly $n - 1$ letters will be placed in the correct envelopes?

9. Suppose that n letters are placed at random in n envelopes, and let q_n denote the probability that no letter is placed in the correct envelope. For which of the following four values of n is q_n largest: $n = 10$, $n = 21$, $n = 53$, or $n = 300$?