BERNHARD CLASSEN/ALAMY

CHAPTER 4

Probability: The Study of Randomness

Introduction

In this chapter, we study basic concepts of probability. The first two chapters focussed on exploring and describing data in hand. In Chapter 3, we learned how to produce quality data that can be reliably used to infer conclusions about the wider population.

You might then ask yourself, "Where does the study of probability fit in our data journey?" The answer lies in recognizing that the reasoning of statistical inference rests on asking, "How often would this method give a correct answer if I used it very many times?" When we produce data by random sampling a randomized comparative experiment, the laws of probability answer the question, "What would happen if we repeated this process many times?" As such, *probability* can be viewed as the backbone of statistical inference.

The importance of probability ideas for statistical inference is reason enough to delve into this chapter. However, our study of probability is further motivated by the fact that businesses use probability and related concepts as the basis for decision making in a world full of risk and uncertainty.

- As a business student reading this book, there is a good chance you are pursuing an accounting major with the hope to become a certified public accountant (CPA). Did you know that accountants can boost their earnings potential by additional 10% to 25% by adding a certification for fraud detection? Certified fraud accountants must have in their toolkit a probability distribution that we study in this chapter. Liberty Mutual Insurance, Citibank, MasterCard, Deloitte, and the FBI are just a few of the organizations that employ fraud accountants.
- With shrinking product life cycles, what was a "hot" seller quickly becomes obsolete. Imagine the challenge for Nike in its decision of how many Dallas

CHAPTER OUTLINE

- Cowboys jersey replicas to produce with a certain player's name. If Nike makes too many and the player leaves for another team, Nike and shops selling NFL apparel will absorb considerable losses when stuck with a nearly unsellable product. We will explore how probability can help industries with short product life cycles make better decisions.
- Financial advisers at wealth management firms such as Wells Fargo, Fidelity Investments, and J.P. Morgan Chase routinely provide advice to their clients on investments. Which ones (stocks, mutual funds, bonds, etc.) should their clients buy? How much in each possible investment should their clients invest? We will learn that their advice is guided by concepts studied in this chapter.
- Online bookseller **Amazon.com** serves its U.S. customers with inventory consolidated in only a handful of warehouses. Each Amazon warehouse pools demand over a large geographical area, which leads to lower total inventory versus having many smaller warehouses. We will discover the principle as to why this strategy provides Amazon with a competitive edge.

4.1 Randomness

REMINDER simple random sample (SRS), p. 132

Toss a coin, or choose an SRS. The result cannot be predicted with certainty in advance because the result will vary when you toss the coin or choose the sample again. But there is still a regular pattern in the results, a pattern that emerges clearly only after many repetitions. This remarkable fact is the basis for the idea of probability.

EXAMPLE 4.1 Coin Tossing

When you toss a coin, there are only two possible outcomes, heads or tails. Figure 4.1 shows the results of tossing a coin 5000 times twice. For each number of tosses from 1 to 5000, we have plotted the proportion of those tosses that gave a head. Trial A (solid line) begins tail, head, tail, tail. You can see that the proportion of heads for Trial A starts at 0 on the first toss, rises to 0.5 when the second toss gives a head, then falls to 0.33 and 0.25 as we get two more tails. Trial B, on the other hand, starts with five straight heads, so the proportion of heads is 1 until the sixth toss.

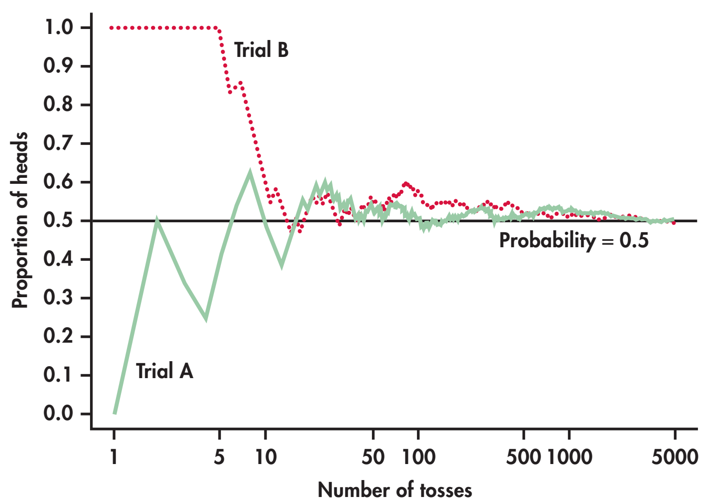


FIGURE 4.1 The proportion of tosses of a coin that give a head changes as we make more tosses. Eventually, however, the proportion approaches 0.5, the probability of a head. This figure shows the results of two trials of 5000 tosses each.

The proportion of tosses that produce heads is quite variable at first. Trial A starts low and Trial B starts high. As we make more and more tosses, however, the proportion of heads for both trials gets close to 0.5 and stays there. If we made yet a third trial at tossing the coin 5000 times, the proportion of heads would again settle down to 0.5 in the long run. We say that 0.5 is the *probability* of a head. The probability 0.5 appears as a horizontal line on the graph.

The *Probability* applet available on the text website animates Figure 4.1. It allows you to choose the probability of a head and simulate any number of tosses of a coin with that probability. Try it. As with Figure 4.1, you will find for your own trial that the proportion of heads gradually settles down close to the probability you chose. Equally important, you will find that the proportion in a small or moderate number of tosses can be far from the probability. *Many people prematurely assess the probability of a phenomenon based only on short-term outcomes.* Probability describes only what happens in the long run.

The language of probability

"Random" in statistics is not a synonym for "haphazard" but a description of a kind of order that emerges only in the long run. We often encounter the unpredictable side of randomness in our everyday experience, but we rarely see enough repetitions of the same random phenomenon to observe the long-term regularity that probability describes. You can see that regularity emerging in Figure 4.1. In the very long run, the proportion of tosses that give a head is 0.5. This is the intuitive idea of probability. Probability 0.5 means "occurs half the time in a very large number of trials."

The idea of probability is *empirical.* That is, it is based on observation rather than theorizing. We might suspect that a coin has probability 0.5 of coming up heads just because the coin has two sides. Probability describes what happens in very many trials, and we must actually observe many trials to pin down a probability. In the case of tossing a coin, some diligent people have, in fact, made thousands of tosses.

EXAMPLE 4.2 Some Coin Tossers

The French naturalist Count Buffon (1707–1788) tossed a coin 4040 times. Result: 2048 heads, or proportion 2048/4040 = 0.5069 for heads.

Around 1900, the English statistician Karl Pearson heroically tossed a coin 24,000 times. Result: 12,012 heads, a proportion of 0.5005.

While imprisoned by the Germans during World War II, the South African mathematician John Kerrich tossed a coin 10,000 times. Result: 5067 heads, a proportion of 0.5067.

The coin-tossing experiments of these individuals did not just result in heads. They also observed the other possible outcome of tails. Pearson, for example, found the proportion of tails to be 0.4995. Their experiments revealed the long-term regularity across all the possible outcomes. In other words, they were able to pin down the **distribution** of outcomes.

distribution

Randomness and Probability

We call a phenomenon **random** if individual outcomes are uncertain but there is, nonetheless, a regular distribution of outcomes in a large number of repetitions.

The **probability** of any outcome of a random phenomenon is the proportion of times the outcome would occur in a very long series of repetitions.

APPLY YOUR KNOWLEDGE

4.1 Not just coins. We introduced this chapter with the most recognizable experiment of chance, the coin toss. The coin has two random outcomes, heads and tails. But, this book is not about coin tossing per se. Provide two examples of business scenarios in which there are two distinct but uncertain outcomes.

Thinking about randomness and probability

Randomness is everywhere. In our personal lives, we observe randomness with varying outdoor temperatures, our blood pressure readings, our commuting times to school or work, and the scores of our favorite sports team. Businesses exist in a world of randomness in the forms of varying dimensions on manufactured parts, customers' waiting times, demand for products or services, prices of a company's stock, injuries in the workplace, and customers' abilities to pay off a loan.

Probability theory is the branch of mathematics that describes random behavior; its advanced study entails high-level mathematics. However, as we will discover, many of the key ideas are basic. Managers who assimilate these key ideas are better able to cope with the stark realities of randomness. They become better decision makers.

Of course, we never observe a probability exactly. We could always continue tossing the coin, for example. Mathematical probability is an idealization based on imagining what would happen in an indefinitely long series of trials. The best way to understand randomness is to observe random behavior—not only the long-run regularity but the unpredictable results of short runs. You can do this with physical devices such as coins and dice, but computer simulations of random behavior allow faster exploration. As you explore randomness, remember:

independence

- You must have a long series of **independent** trials. That is, the outcome of one trial must not influence the outcome of any other. Imagine a crooked gambling house where the operator of a roulette wheel can stop it where she chooses—she can prevent the proportion of "red" from settling down to a fixed number. These trials are not independent.
- The idea of probability is empirical. Computer simulations start with given probabilities and imitate random behavior, but we can estimate a real-world probability only by actually observing many trials.
- Nonetheless, computer simulations are very useful because we need long runs of trials. In situations such as coin tossing, the proportion of an outcome often requires several hundred trials to settle down to the probability of that outcome. Exploration of probability with physical devices is typically too time consuming. Short runs give only rough estimates of a probability.

SECTION 4.1 Summary

- A **random phenomenon** has outcomes that we cannot predict with certainty but that, nonetheless, have a regular distribution in very many repetitions.
- The **probability** of an event is the proportion of times the event occurs in many repeated trials of a random phenomenon.
- Trials are **independent** if the outcome of one trial does not influence the outcome of any other trial.

SECTION 4.1 Exercises

For Exercise 4.1, see page 176.

4.2 Are these phenomena random? Identify each of the following phenomena as random or not. Give reasons for your answers.
(a) The outside temperature in Chicago at noon on New Year's Day.
(b) The first character to the right of the "@" symbol in an employee's company email address.
(c) You draw an ace from a well-shuffled deck of 52 cards.

4.3 Interpret the probabilities. Refer to the previous exercise. In each case, interpret the term "probability" for the phenomena that are random. For those that are not random, explain why the term "probability" does not apply.

4.4 Are the trials independent? For each of the following situations, identify the trials as independent or not. Explain your answers.
(a) The outside temperature in Chicago at noon on New Year's Day, each year for the next five years.
(b) The number of tweets that you receive on the next 10 Mondays.
(c) Your grades in the five courses that you are taking this semester.

4.5 Financial fraud. It has been estimated that around one in six fraud victims knew the perpetrator as a friend or acquaintance. Financial fraud includes crimes such as unauthorized credit card charges, withdrawal of money from a savings or checking account, and opening an account in someone else's name. Suppose you want to use a physical device to simulate the outcome that a fraud victim knew the perpetrator versus the outcome that the fraud victim does not know the perpetrator. What device would you use to conduct a simulation experiment? Explain how you would match the outcomes of the device with the fraud scenario.

4.6 Credit monitoring. In a recent study of consumers, 25% reported purchasing a credit-monitoring product that alerts them to any activity on their credit report. Suppose you want to use a physical device to simulate the outcome of a consumer purchasing the credit-monitoring product versus the outcome of the consumer not purchasing the product. Describe how you could use two fair coins to conduct a simulation experiment to mimic consumer behavior. In particular, what outcomes of the two flipped coins would you associate with purchasing the product versus what outcomes would you associate with not purchasing the product?

4.7 Random digits. As discussed in Chapter 3, generation of random numbers is one approach for obtaining a simple random sample (SRS). If we were to look at the random generation of digits, the mechanism should give each digit probability 0.1. Consider the digit "0" in particular.
(a) The table of random digits (Table B) was produced by a random mechanism that gives each digit probability 0.1 of being a 0. What proportion of the first 200 digits in the table are 0s? This proportion is an estimate, based on 200 repetitions, of the true probability, which in this case is known to be 0.1.
(b) Now use software assigned by your instructor:

- *Excel users:* Enter the formula =**RANDBETWEEN(0,9)** in cell A1. Now, drag and copy the contents of cell A1 into cells A2:A1000. You will find 1000 random digits appear. Any attempt to copy these digits for sorting purposes will result in the digits changing. You will need to "freeze" the generated values. To do so, highlight column 1 and copy the contents and then **Paste Special as Values** the contents into the same or any other column. The values will now not change. Finally, use Excel to sort the values in ascending order.
- *JMP users:* With a new data table, right-click on the header of Column 1 and choose **Column Info**. In the drag-down dialog box named **Initialize Data**, pick **Random** option. Choose the bullet option of **Random Integer**, and set **Minimum/Maximum** to 0 and 9. Input the value of 1000 into the **Number of rows** box, and then click **OK**. The values can then be sorted in ascending order using the **Sort** option found under **Tables**.
- *Minitab users:* Do the following pull-down sequence: **Calc** → **Random Data** → **Integer**. Enter "1000" in the **Number of rows of data to generate** box, type "c1" in the **Store in column(s)** box, enter "0" in the **Minimum value** box, and enter "9" in the **Maximum** box. Click **OK** to find 1000 realizations of X outputted in the worksheet. The values can then be sorted in ascending order using the **Sort** option found under **Data**.

Based on the software you used, what proportion of the 1000 randomly generated digits are 0s? Is this proportion close to 0.1?

4.8 Are McDonald's prices independent? Over time, stock prices are always on the move. Consider a time series of 1126 consecutive daily prices of McDonald's stock from the beginning of January 2010 to the near the end of June 2014.[1] MCD

(a) Using software, plot the prices over time. Are the prices constant over time? Describe the nature of the price movement over time.

(b) Now consider the relationship between price on any given day with the price on the prior day. The previous day's price is sometimes referred to as the *lag* price. You will want to get the lagged prices in another column of your software:

- *Excel users:* Highlight and copy the price values, and paste them in a new column shifted down by one row.
- *JMP users:* Click on the price column header name to highlight the column of price values. Copy the highlighted values. Now click anywhere on the nearest empty column, resulting in the column being filled with missing values. Double-click on the cell in row 2 of the newly formed column. With row 2 cell open, paste the price values to create a column of lagged prices. (Note: A column of lagged values can also be created with JMP's **Lag** function found in the **Formula** option of the column.)
- *Minitab users:* **Stat** → **Time Series** → **Lag**.

Refering back to Chapter 2 and scatterplots, create a scatterplot of McDonald's price on a given day versus the price on the previous day. Does the scatterplot suggest that the price series behaves as a series of independent trials? Explain why or why not.

4.9 Are McDonald's price changes independent? Refer to the daily price series of McDonald's stock in Exercise 4.8. Instead of looking at the prices themselves, consider now the daily *changes* in prices found in the provided data file. MCD

(a) Using software, plot the price changes over time. Describe the nature of the price changes over time.

(b) Now consider the relationship between a given price change and the previous price change. Create a lag of price changes by following the steps of Exercise 4.8(b). Create a scatterplot of price change versus the previous price change. Does the scatterplot seem to suggest that the price-change series behaves essentially as a series of independent trials? Explain why or why not.

(c) This exercise only explored the relationship or lack of it between price changes of successive days. If you want to feel more confident about a conclusion of independence of price changes over time, what additional scatterplots might you consider creating?

APPLET **4.10 Use the *Probability* applet.** The idea of probability is that the *proportion* of heads in many tosses of a balanced coin eventually gets close to 0.5. But does the actual *count* of heads get close to one-half the number of tosses? Let's find out. Set the "Probability of Heads" in the *Probability* applet to 0.5 and the number of tosses to 50. You can extend the number of tosses by clicking "Toss" again to get 50 more. Don't click "Reset" during this exercise.

(a) After 50 tosses, what is the proportion of heads? What is the count of heads? What is the difference between the count of heads and 25 (one-half the number of tosses)?

(b) Keep going to 150 tosses. Again record the proportion and count of heads and the difference between the count and 75 (half the number of tosses).

(c) Keep going. Stop at 300 tosses and again at 600 tosses to record the same facts. Although it may take a long time, the laws of probability say that the proportion of heads will always get close to 0.5 and also that the difference between the count of heads and half the number of tosses will always grow without limit.

APPLET **4.11 A question about dice.** Here is a question that a French gambler asked the mathematicians Fermat and Pascal at the very beginning of probability theory: what is the probability of getting at least one 6 in rolling four dice? The *Law of Large Numbers* applet allows you to roll several dice and watch the outcomes. (Ignore the title of the applet for now.) Because simulation—just like real random phenomena—often takes very many trials to estimate a probability accurately, let's simplify the question: is this probability clearly greater than 0.5, clearly less than 0.5, or quite close to 0.5? Use the applet to roll four dice until you can confidently answer this question. You will have to set "Rolls" to 1 so that you have time to look at the four up-faces. Keep clicking "Roll dice" to roll again and again. How many times did you roll four dice? What percent of your rolls produced at least one 6?

4.12 Proportions of McDonald's price changes. Continue the study of daily price changes of McDonald's stock from the Exercise 4.9. Consider three possible outcomes: (1) positive price change, (2) no price change, and (3) negative price change. MCD

(a) Find the proportions of each of these outcomes. This is most easily done by sorting the price change data into another column of the software and then counting the number of negative, zero, and positive values.

(b) Explain why the proportions found in part (a) are reasonable estimates for the true probabilities.

4.13 Thinking about probability statements. Probability is a measure of how likely an event is to occur. Match one of the probabilities that follow with each statement of likelihood given. (The probability is usually a more exact measure of likelihood than is the verbal statement.)

0 0.01 0.3 0.6 0.99 1

(a) This event is impossible. It can never occur.
(b) This event is certain. It will occur on every trial.
(c) This event is very unlikely, but it will occur once in a while in a long sequence of trials.
(d) This event will occur more often than not.

4.2 Probability Models

probability model

The idea of probability as a proportion of outcomes in very many repeated trials guides our intuition but is hard to express in mathematical form. A description of a random phenomenon in the language of mathematics is called a **probability model.** To see how to proceed, think first about a very simple random phenomenon, tossing a coin once. When we toss a coin, we cannot know the outcome in advance. What do we know? We are willing to say that the outcome will be either heads or tails. Because the coin appears to be balanced, we believe that each of these outcomes has probability 1/2. This description of coin tossing has two parts:

1. a list of possible outcomes
2. a probability for each outcome

This two-part description is the starting point for a probability model. We begin by describing the outcomes of a random phenomenon and then learn how to assign these probabilities ourselves.

Sample spaces

A probability model first tells us what outcomes are possible.

Sample Space
The **sample space** S of a random phenomenon is the set of all distinct possible outcomes.

The name "sample space" is natural in random sampling, where each possible outcome is a sample and the sample space contains all possible samples. To specify S, we must state what constitutes an individual outcome and then state which outcomes can occur. We often have some freedom in defining the sample space, so the choice of S is a matter of convenience as well as correctness. The idea of a sample space, and the freedom we may have in specifying it, are best illustrated by examples.

EXAMPLE 4.3 Sample Space for Tossing a Coin

Toss a coin. There are only two possible outcomes, and the sample space is

$$S = \{\text{heads, tails}\}$$

or, more briefly, $S = \{\text{H, T}\}$.

EXAMPLE 4.4 Sample Space for Random Digits

Type "=RANDBETWEEN(0,9)" into any Excel cell and hit enter. Record the value of the digit that appears in the cell. The possible outcomes are

$$S = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

EXAMPLE 4.5 Sample Space for Tossing a Coin Four Times

Toss a coin four times and record the results. That's a bit vague. To be exact, record the results of each of the four tosses in order. A possible outcome is then HTTH. Counting shows that there are 16 possible outcomes. The sample space S is the set of all 16 strings of four toss results—that is, strings of H's and T's.

Suppose that our only interest is the number of heads in four tosses. Now we can be exact in a simpler fashion. The random phenomenon is to toss a coin four times and count the number of heads. The sample space contains only five outcomes:

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

This example illustrates the importance of carefully specifying what constitutes an individual outcome.

Although these examples seem remote from the practice of statistics, the connection is surprisingly close. Suppose that in conducting a marketing survey, you select four people at random from a large population and ask each if he or she has used a given product. The answers are Yes or No. The possible outcomes—the sample space—are exactly as in Example 4.5 if we replace heads by Yes and tails by No. Similarly, the possible outcomes of an SRS of 1500 people are the same in principle as the possible outcomes of tossing a coin 1500 times. One of the great advantages of mathematics is that the essential features of quite different phenomena can be described by the same mathematical model, which, in our case, is the probability model.

The sample spaces considered so far correspond to situations in which there is a finite list of all the possible values. There are other sample spaces in which, theoretically, the list of outcomes is infinite.

EXAMPLE 4.6 Using Software

Most statistical software has a function that will generate a random number between 0 and 1. The sample space is

$$S = \{\text{all numbers between 0 and 1}\}$$

This S is a mathematical idealization with an infinite number of outcomes. In reality, any specific random number generator produces numbers with some limited number of decimal places so that, strictly speaking, not all numbers between 0 and 1 are possible outcomes. For example, in default mode, Excel reports random numbers like 0.798249, with six decimal places. The entire interval from 0 to 1 is easier to think about. It also has the advantage of being a suitable sample space for different software systems that produce random numbers with different numbers of digits.

APPLY YOUR KNOWLEDGE

4.14 Describing sample spaces. In each of the following situations, describe a sample space S for the random phenomenon. In some cases, you have some freedom in your choice of S.

(a) A new business is started. After two years, it is either still in business or it has closed.
(b) A student enrolls in a business statistics course and, at the end of the semester, receives a letter grade.

(c) A food safety inspector tests four randomly chosen henhouse areas for the presence of Salmonella or not. You record the sequence of results.
(d) A food safety inspector tests four randomly chosen henhouse areas for the presence of Salmonella or not. You record the number of areas that show contamination.

4.15 Describing sample spaces. In each of the following situations, describe a sample space S for the random phenomenon. Explain why, *theoretically*, a list of all possible outcomes is not finite.

(a) You record the number of tosses of a die until you observe a six.
(b) You record the number of tweets per week that a randomly selected student makes.

A sample space S lists the possible outcomes of a random phenomenon. To complete a mathematical description of the random phenomenon, we must also give the probabilities with which these outcomes occur.

The true long-term proportion of any outcome—say, "exactly two heads in four tosses of a coin"—can be found only empirically, and then only approximately. How then can we describe probability mathematically? Rather than immediately attempting to give "correct" probabilities, let's confront the easier task of laying down rules that any assignment of probabilities must satisfy. We need to assign probabilities not only to single outcomes but also to sets of outcomes.

Event
An **event** is an outcome or a set of outcomes of a random phenomenon. That is, an event is a subset of the sample space.

EXAMPLE 4.7 Exactly Two Heads in Four Tosses

Take the sample space S for four tosses of a coin to be the 16 possible outcomes in the form HTHH. Then "exactly two heads" is an event. Call this event A. The event A expressed as a set of outcomes is

$$A = \{\text{TTHH, THTH, THHT, HTTH, HTHT, HHTT}\}$$

In a probability model, events have probabilities. What properties must any assignment of probabilities to events have? Here are some basic facts about any probability model. These facts follow from the idea of probability as "the long-run proportion of repetitions on which an event occurs."

1. Any probability is a number between 0 and 1. Any proportion is a number between 0 and 1, so any probability is also a number between 0 and 1. An event with probability 0 never occurs, and an event with probability 1 occurs on every trial. An event with probability 0.5 occurs in half the trials in the long run.

2. All possible outcomes of the sample space together must have probability 1. Because every trial will produce an outcome, the sum of the probabilities for all possible outcomes must be exactly 1.

3. If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities. If one event occurs in 40% of all trials, a different event occurs in 25% of all trials, and the two can never occur together, then one or the other occurs on 65% of all trials because 40% + 25% = 65%.

4. The probability that an event does not occur is 1 minus the probability that the event does occur. If an event occurs in 70% of all trials, it fails to occur in the other 30%. The probability that an event occurs and the probability that it does not occur always add to 100%, or 1.

Probability rules

Formal probability uses mathematical notation to state Facts 1 to 4 more concisely. We use capital letters near the beginning of the alphabet to denote events. If A is any event, we write its probability as $P(A)$. Here are our probability facts in formal language. As you apply these rules, remember that they are just another form of intuitively true facts about long-run proportions.

Probability Rules

Rule 1. The probability $P(A)$ of any event A satisfies $0 \le P(A) \le 1$.

Rule 2. If S is the sample space in a probability model, then $P(S) = 1$.

Rule 3. Two events A and B are **disjoint** if they have no outcomes in common and so can never occur together. If A and B are disjoint,

$$P(A \text{ or } B) = P(A) + P(B)$$

This is the **addition rule for disjoint events.**

Rule 4. The **complement** of any event A is the event that A does not occur, written as A^c. The **complement rule** states that

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

Venn diagram

You may find it helpful to draw a picture to remind yourself of the meaning of complements and disjoint events. A picture like Figure 4.2 that shows the sample space S as a rectangular area and events as areas within S is called a **Venn diagram.** The events A and B in Figure 4.2 are disjoint because they do not overlap. As Figure 4.3 shows, the complement A^c contains exactly the outcomes that are not in A.

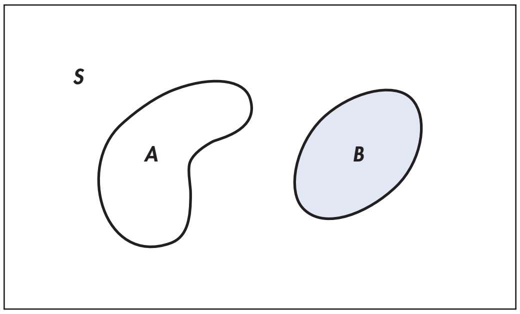


FIGURE 4.2 Venn diagram showing disjoint events *A* and *B*.

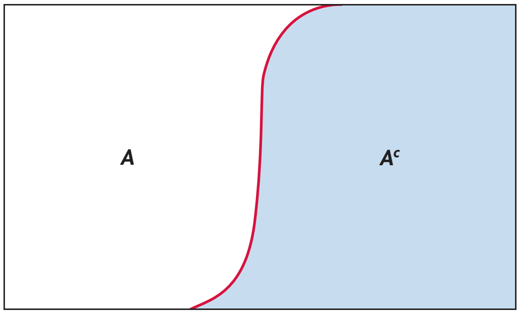


FIGURE 4.3 Venn diagram showing the complement A^c of an event A. The complement consists of all outcomes that are not in A.


CARROTEATER/SHUTTERSTOCK

EXAMPLE 4.8 Favorite Vehicle Colors

What is your favorite color for a vehicle? Our preferences can be related to our personality, our moods, or particular objects. Here is a probability model for color preferences.[2]

Color	White	Black	Silver	Gray
Probability	0.24	0.19	0.16	0.15

Color	Red	Blue	Brown	Other
Probability	0.10	0.07	0.05	0.04

Each probability is between 0 and 1. The probabilities add to 1 because these outcomes together make up the sample space S. Our probability model corresponds to selecting a person at random and asking him or her about a favorite color.

Let's use the probability Rules 3 and 4 to find some probabilities for favorite vehicle colors.

EXAMPLE 4.9 Black or Silver?

What is the probability that a person's favorite vehicle color is black or silver? If the favorite is black, it cannot be silver, so these two events are disjoint. Using Rule 3, we find

$$\begin{aligned} P(\text{black or silver}) &= P(\text{black}) + P(\text{silver}) \\ &= 0.19 + 0.16 = 0.35 \end{aligned}$$

There is a 35% chance that a randomly selected person will choose black or silver as his or her favorite color. Suppose that we want to find the probability that the favorite color is not blue.

EXAMPLE 4.10 Use the Complement Rule

To solve this problem, we could use Rule 3 and add the probabilities for white, black, silver, gray, red, brown, and other. However, it is easier to use the probability that we have for blue and Rule 4. The event that the favorite is not blue is the complement of the event that the favorite is blue. Using our notation for events, we have

$$\begin{aligned} P(\text{not blue}) &= 1 - P(\text{blue}) \\ &= 1 - 0.07 = 0.93 \end{aligned}$$

We see that 93% of people have a favorite vehicle color that is not blue.

APPLY YOUR KNOWLEDGE

4.16 Red or brown. Refer to Example 4.8, and find the probability that the favorite color is red or brown.

4.17 White, black, silver, gray, or red. Refer to Example 4.8, and find the probability that the favorite color is white, black, silver, gray, or red using Rule 4. Explain why this calculation is easier than finding the answer using Rule 3.

4.18 Moving up. An economist studying economic class mobility finds that the probability that the son of a father in the lowest economic class remains in that class is 0.46. What is the probability that the son moves to one of the higher classes?

4.19 Occupational deaths. Government data on job-related deaths assign a single occupation for each such death that occurs in the United States. The data on occupational deaths in 2012 show that the probability is 0.183 that a randomly chosen death was a construction worker and 0.039 that it was miner. What is the probability that a randomly chosen death was either construction related or mining related? What is the probability that the death was related to some other occupation?

4.20 Grading Canadian health care. Annually, the Canadian Medical Association uses the marketing research firm Ipsos Canada to measure public opinion with respect to the Canadian health care system. Between July 17 and July 26 of 2013, Ipsos Canada interviewed a random sample of 1000 adults.[3] The people in the sample were asked to grade the overall quality of health care services as an A, B, C, or F, where an A is the highest grade and an F is a failing grade. Here are the results:

Outcome	Probability
A	0.30
B	0.45
C	?
F	0.06

These proportions are probabilities for choosing an adult at random and asking the person's opinion on the Canadian health care system.

(a) What is the probability that a person chosen at random gives a grade of C? Why?
(b) If a "positive" grade is defined as A or B, what is the probability of a positive grade?

Assigning probabilities: Finite number of outcomes

The individual outcomes of a random phenomenon are always disjoint. So, the addition rule provides a way to assign probabilities to events with more than one outcome: start with probabilities for individual outcomes and add to get probabilities for events. This idea works well when there are only a finite (fixed and limited) number of outcomes.

Probabilities in a Finite Sample Space

Assign a probability to each individual outcome. These probabilities must be numbers between 0 and 1 and must have sum 1.

The probability of any event is the sum of the probabilities of the outcomes making up the event.

CASE 4.1

Uncovering Fraud by Digital Analysis What is the probability that the leftmost digit ("first digit") of a multidigit financial number is 9? Many of us would assume the probability to be 1/9. Surprisingly, this is often not the case for legitimately reported financial numbers. It is a striking fact that the first digits of numbers in legitimate records often follow a distribution known as *Benford's law.* Here it is (note that the first digit can't be 0):

First digit	1	2	3	4	5	6	7	8	9
Proportion	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

It is a regrettable fact that financial fraud permeates business and governmental sectors. In a recent 2014 study, the Association of Certified Fraud Examiners (ACFE) estimates that a typical organization loses 5% of revenues each year to fraud.[4] ACFE projects a global fraud loss of nearly $4 trillion. Common examples of business fraud include:

- *Corporate financial statement fraud:* reporting fictitious revenues, understating expenses, artificially inflating reported assets, and so on.
- *Personal expense fraud:* employee reimbursement claims for fictitious or inflated business expenses (for example, personal travel, meals, etc.).
- *Billing fraud:* submission of inflated invoices or invoices for fictitious goods or services to be paid to an employee-created shell company.
- *Cash register fraud:* false entries on a cash register for fraudulent removal of cash.

In all these situations, the individual(s) committing fraud are needing to "invent" fake financial entry numbers. In whatever means the invented numbers are created, the first digits of the fictitious numbers will most likely not follow the probabilities given by Benford's law. As such, Benford's law serves as an important "digital analysis" tool of auditors, typically CPA accountants, trained to look for fraudulent behavior.

Of course, not all sets of data follow Benford's law. Numbers that are assigned, such as Social Security numbers, do not. Nor do data with a fixed maximum, such as deductible contributions to individual retirement accounts (IRAs). Nor, of course, do random numbers. But given a remarkable number of financial-related data sets do closely obey Benford's law, its role in auditing of financial and accounting statements cannot be ignored.

EXAMPLE 4.11 Find Some Probabilities for Benford's Law

CASE 4.1 Consider the events

$$A = \{\text{first digit is 5}\}$$
$$B = \{\text{first digit is 3 or less}\}$$

From the table of probabilities in Case 4.1,

$$P(A) = P(5) = 0.079$$
$$P(B) = P(1) + P(2) + P(3)$$
$$= 0.301 + 0.176 + 0.125 = 0.602$$

Note that $P(B)$ is not the same as the probability that a first digit is strictly less than 3. The probability $P(3)$ that a first digit is 3 is included in "3 or less" but not in "less than 3."

APPLY YOUR KNOWLEDGE

4.21 Household space heating. Draw a U.S. household at random, and record the primary source of energy to generate heat for warmth of the household using space-heating equipment. "At random" means that we give every household the same chance to be chosen. That is, we choose an SRS of size 1. Here is the distribution of primary sources for U.S. households:[5]

Primary source	Probability
Natural gas	0.50
Electricity	0.35
Distillate fuel oil	0.06
Liquefied petroleum gases	0.05
Wood	0.02
Other	0.02

(a) Show that this is a legitimate probability model.
(b) What is the probability that a randomly chosen U.S. household uses natural gas or electricity as its primary source of energy for space heating?

CASE 4.1 **4.22 Benford's law.** Using the probabilities for Benford's law, find the probability that a first digit is anything other than 4.

CASE 4.1 **4.23 Use the addition rule.** Use the addition rule (page 182) with the probabilities for the events A and B from Example 4.11 to find the probability of A or B.

EXAMPLE 4.12 Find More Probabilities for Benford's Law

CASE 4.1 Check that the probability of the event C that a first digit is even is

$$P(C) = P(2) + P(4) + P(6) + P(8) = 0.391$$

Consider again event B from Example 4.11 (page 185), which had an associated probability of 0.602. The probability

$$P(B \text{ or } C) = P(1) + P(2) + P(3) + P(4) + P(6) + P(8) = 0.817$$

is *not* the sum of $P(B)$ and $P(C)$ because events B and C are not disjoint. The outcome of 2 is common to both events. *Be careful to apply the addition rule only to disjoint events.* In Section 4.3, we expand upon the addition rule given in this section to handle the case of nondisjoint events.

Assigning probabilities: Equally likely outcomes

Assigning correct probabilities to individual outcomes often requires long observation of the random phenomenon. In some circumstances, however, we are willing to assume that individual outcomes are equally likely because of some balance in the phenomenon. Ordinary coins have a physical balance that should make heads and tails equally likely, for example, and the table of random digits comes from a deliberate randomization.

EXAMPLE 4.13 First Digits That Are Equally Likely

You might think that first digits in business records are distributed "at random" among the digits 1 to 9. The nine possible outcomes would then be equally likely. The sample space for a single digit is

$$S = \{1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

Because the total probability must be 1, the probability of each of the nine outcomes must be 1/9. That is, the assignment of probabilities to outcomes is

First digit	1	2	3	4	5	6	7	8	9
Probability	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

The probability of the event B that a randomly chosen first digit is 3 or less is

$$P(B) = P(1) + P(2) + P(3)$$
$$= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{3}{9} = 0.333$$

Compare this with the Benford's law probability in Example 4.11 (page 185). A crook who fakes data by using "random" digits will end up with too few first digits that are 3 or less.

In Example 4.13, all outcomes have the same probability. Because there are nine equally likely outcomes, each must have probability 1/9. Because exactly three of the nine equally likely outcomes are 3 or less, the probability of this event is 3/9. In the special situation in which all outcomes are equally likely, we have a simple rule for assigning probabilities to events.

Equally Likely Outcomes

If a random phenomenon has k possible outcomes, all equally likely, then each individual outcome has probability $1/k$. The probability of any event A is

$$P(A) = \frac{\text{count of outcomes in } A}{\text{count of outcomes in } S} = \frac{\text{count of outcomes in } A}{k}$$

Most random phenomena do not have equally likely outcomes, so the general rule for finite sample spaces (page 184) is more important than the special rule for equally likely outcomes.

APPLY YOUR KNOWLEDGE

4.24 Possible outcomes for rolling a die. A die has six sides with one to six spots on the sides. Give the probability distribution for the six possible outcomes that can result when a fair die is rolled.

Independence and the multiplication rule

Rule 3, the addition rule for disjoint events, describes the probability that *one or the other* of two events A and B occurs when A and B cannot occur together. Now we describe the probability that *both* events A and B occur, again only in a special situation. More general rules appear in Section 4.3.

Suppose that you toss a balanced coin twice. You are counting heads, so two events of interest are

$$A = \{\text{first toss is a head}\}$$
$$B = \{\text{second toss is a head}\}$$

The events A and B are not disjoint. They occur together whenever both tosses give heads. We want to compute the probability of the event $\{A \text{ and } B\}$ that *both* tosses are heads. The Venn diagram in Figure 4.4 illustrates the event $\{A \text{ and } B\}$ as the overlapping area that is common to both A and B.

The coin tossing of Buffon, Pearson, and Kerrich described in Example 4.2 makes us willing to assign probability 1/2 to a head when we toss a coin. So,

$$P(A) = 0.5$$
$$P(B) = 0.5$$

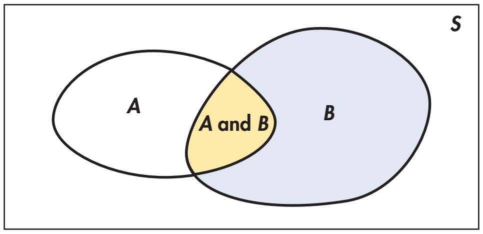


FIGURE 4.4 Venn diagram showing the events *A* and *B* that are not disjoint. The event {*A* and *B*} consists of outcomes common to *A* and *B*.

What is $P(A \text{ and } B)$? Our common sense says that it is 1/4. The first coin will give a head half the time and then the second will give a head on half of those trials, so both coins will give heads on $1/2 \times 1/2 = 1/4$ of all trials in the long run. This reasoning assumes that the second coin still has probability 1/2 of a head after the first has given a head. This is true—we can verify it by tossing two coins many times and observing the proportion of heads on the second toss after the first toss has produced a head. We say that the events "head on the first toss" and "head on the second toss" are *independent.* Here is our final probability rule.

Multiplication Rule for Independent Events

Rule 5. Two events *A* and *B* are **independent** if knowing that one occurs does not change the probability that the other occurs. If *A* and *B* are independent,

$$P(A \text{ and } B) = P(A)P(B)$$

This is the **multiplication rule for independent events.**

Our definition of independence is rather informal. We make this informal idea precise in Section 4.3. In practice, though, we rarely need a precise definition of independence because independence is usually *assumed* as part of a probability model when we want to describe random phenomena that seem to be physically unrelated to each other.

EXAMPLE 4.14 Determining Independence Using the Multiplication Rule

Consider a manufacturer that uses two suppliers for supplying an identical part that enters the production line. Sixty percent of the parts come from one supplier, while the remaining 40% come from the other supplier. Internal quality audits find that there is a 1% chance that a randomly chosen part from the production line is defective. External supplier audits reveal that two parts per 1000 are defective from Supplier 1. Are the events of a part coming from a particular supplier—say, Supplier 1—and a part being defective independent?

Define the two events as follows:

$$S1 = \text{A randomly chosen part comes from Supplier 1}$$
$$D = \text{A randomly chosen part is defective}$$

We have $P(S1) = 0.60$ and $P(D) = 0.01$. The product of these probabilities is

$$P(S1)P(D) = (0.60)(0.01) = 0.006$$

However, supplier audits of Supplier 1 indicate that $P(S1 \text{ and } D) = 0.002$. Given that $P(S1 \text{ and } D) \neq P(S1)P(D)$, we conclude that the supplier and defective part events are not independent.

The multiplication rule $P(A \text{ and } B) = P(A)P(B)$ holds if *A* and *B* are *independent* but not otherwise. The addition rule $P(A \text{ or } B) = P(A) + P(B)$ holds if *A* and *B* are *disjoint* but not otherwise. Resist the temptation to use these simple rules when the circumstances that justify them are not present. *You must also be certain not to confuse disjointness and independence. Disjoint events cannot be independent.* If A and B are disjoint, then the fact that A occurs tells us that B cannot occur—look back at Figure 4.2 (page 182). Thus, disjoint events are not independent. Unlike disjointness, picturing independence with a Venn diagram is not obvious. A mosaic plot introduced in Chapter 2 provides a better way to visualize independence or lack of it. We will see more examples of mosaic plots in Chapter 9.

REMINDER
mosaic plot, p. 109

APPLY YOUR KNOWLEDGE

4.25 High school rank. Select a first-year college student at random and ask what his or her academic rank was in high school. Here are the probabilities, based on proportions from a large sample survey of first-year students:

Rank	Top 20%	Second 20%	Third 20%	Fourth 20%	Lowest 20%
Probability	0.41	0.23	0.29	0.06	0.01

(a) Choose two first-year college students at random. Why is it reasonable to assume that their high school ranks are independent?
(b) What is the probability that both were in the top 20% of their high school classes?
(c) What is the probability that the first was in the top 20% and the second was in the lowest 20%?

4.26 College-educated part-time workers? For people aged 25 years or older, government data show that 34% of employed people have at least four years of college and that 20% of employed people work part-time. Can you conclude that because $(0.34)(0.20) = 0.068$, about 6.8% of employed people aged 25 years or older are college-educated part-time workers? Explain your answer.

Applying the probability rules

If two events A and B are independent, then their complements A^c and B^c are also independent and A^c is independent of B. Suppose, for example, that 75% of all registered voters in a suburban district are Republicans. If an opinion poll interviews two voters chosen independently, the probability that the first is a Republican and the second is not a Republican is $(0.75)(1 - 0.75) = 0.1875$.

The multiplication rule also extends to collections of more than two events, provided that all are independent. Independence of events A, B, and C means that no information about any one or any two can change the probability of the remaining events. The formal definition is a bit messy. Fortunately, independence is usually assumed in setting up a probability model. We can then use the multiplication rule freely.

By combining the rules we have learned, we can compute probabilities for rather complex events. Here is an example.

EXAMPLE 4.15 False Positives in Job Drug Testing

Job applicants in both the public and the private sector are often finding that preemployment drug testing is a requirement. The Society for Human Resource Management found that 71% of larger organizations (25,000 + employees) require drug testing of new job applicants and that 44% of these organizations randomly test hired employees.[6] From an applicant's or employee's perspective, one primary concern with drug testing is a "false-positive" result, that is, an indication of drug use when the individual has indeed not used drugs. If a job applicant tests positive, some companies allow the applicant to pay for a retest. For existing employees, a positive result is sometimes followed up with a more sophisticated and expensive test. Beyond cost considerations, there are issues of defamation, wrongful discharge, and emotional distress.

The enzyme multiplied immunoassay technique, or EMIT, applied to urine samples is one of the most common tests for illegal drugs because it is fast and inexpensive. Applied to people who are free of illegal drugs, EMIT has been reported to have false-positive rates ranging from 0.2% to 2.5%. If 150 employees are tested and all 150 are free of illegal drugs, what is the probability that at least one false positive will occur, assuming a 0.2% false positive rate?

It is reasonable to assume as part of the probability model that the test results for different individuals are independent. The probability that the test is positive for a single person is 0.2%, or 0.002, so the probability of a negative result is $1 - 0.002 = 0.998$ by the complement rule. The probability of at least one false-positive among the 150 people tested is, therefore,

$$\begin{aligned} P(\text{at least 1 positive}) &= 1 - P(\text{no positives}) \\ &= 1 - P(\text{150 negatives}) \\ &= 1 - 0.998^{150} \\ &= 1 - 0.741 = 0.259 \end{aligned}$$

The probability is greater than 1/4 that at least one of the 150 people will test positive for illegal drugs even though no one has taken such drugs.

APPLY YOUR KNOWLEDGE

4.27 Misleading résumés. For more than two decades, Jude Werra, president of an executive recruiting firm, has tracked executive résumés to determine the rate of misrepresenting education credentials and/or employment information. On a biannual basis, Werra reports a now nationally recognized statistic known as the "Liars Index." In 2013, Werra reported that 18.4% of executive job applicants lied on their résumés.[7]

(a) Suppose five résumés are randomly selected from an executive job applicant pool. What is the probability that all of the résumés are truthful?
(b) What is the probability that at least one of five randomly selected résumés has a misrepresentation?

4.28 Failing to detect drug use. In Example 4.15, we considered how drug tests can indicate illegal drug use when no illegal drugs were actually used. Consider now another type of false test result. Suppose an employee is suspected of having used an illegal drug and is given two tests that operate independently of each other. Test A has probability 0.9 of being positive if the illegal drug has been used. Test B has probability 0.8 of being positive if the illegal drug has been used. What is the probability that *neither* test is positive if the illegal drug has been used?

4.29 Bright lights? A string of holiday lights contains 20 lights. The lights are wired in series, so that if any light fails the whole string will go dark. Each light has probability 0.02 of failing during a three-year period. The lights fail independently of each other. What is the probability that the string of lights will remain bright for a three-year period?

SECTION 4.2 Summary

- A **probability model** for a random phenomenon consists of a sample space S and an assignment of probabilities P.
- The **sample space** S is the set of all possible outcomes of the random phenomenon. Sets of outcomes are called **events.** P assigns a number $P(A)$ to an event A as its probability.
- The **complement** A^c of an event A consists of exactly the outcomes that are not in A.
- Events A and B are **disjoint** if they have no outcomes in common.
- Events A and B are **independent** if knowing that one event occurs does not change the probability we would assign to the other event.
- Any assignment of probability must obey the rules that state the basic properties of probability:

  Rule 1. $0 \leq P(A) \leq 1$ for any event A.

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

  Rule 3. Addition rule: If events A and B are **disjoint,** then $P(A \text{ or } B) = P(A) + P(B)$.

  Rule 4. Complement rule: For any event A, $P(A^c) = 1 - P(A)$.

  Rule 5. Multiplication rule: If events A and B are **independent,** then $P(A \text{ and } B) = P(A)P(B)$.

SECTION 4.2 Exercises

For Exercises 4.14 and 4.15, see pages 180–181; for 4.16 to 4.20, see pages 183–184; for 4.21 to 4.23, see pages 185–186; for 4.24, see page 187; for 4.25 and 4.26, see page 189; and for 4.27 to 4.29, see page 190.

4.30 Support for casino in Toronto. In an effort to seek the public's input on the establishment of a casino, Toronto's city council enlisted an independent analytics research company to conduct a public survey. A random sample of 902 adult Toronto residents were asked if they support the casino in Toronto.[8] Here are the results:

Response	Strongly support	Somewhat support	Mixed feelings
Probability	0.16	0.26	?

Response	Somewhat oppose	Strongly oppose	Don't know
Probability	0.14	0.36	0.01

(a) What probability should replace "?" in the distribution?
(b) What is the probability that a randomly chosen adult Toronto resident supports (strongly or somewhat) a casino?

4.31 Confidence in institutions. A Gallup Poll (June 1–4, 2013) interviewed a random sample of 1529 adults (18 years or older). The people in the sample were asked about their level of confidence in a variety of institutions in the United States. Here are the results for small and big businesses:[9]

	Great deal	Quite a lot	Some	Very little	None	No opinion
Small business	0.29	0.36	0.27	0.07	0.00	0.01
Big business	0.09	0.13	0.43	0.31	0.02	0.02

(a) What is the probability that a randomly chosen person has either no opinion, no confidence, or very little confidence in small businesses? Find the similar probability for big businesses.
(b) Using your answer from part (a), determine the probability that a randomly chosen person has *at least* some confidence in small businesses. Again based on part (a), find the similar probability for big businesses.

4.32 Demographics—language. Canada has two official languages, English and French. Choose a

Canadian at random and ask, "What is your mother tongue?" Here is the distribution of responses, combining many separate languages from the broad Asian/Pacific region:[10]

Language	English	French	Sino-Tibetan	Other
Probability	0.581	0.217	0.033	?

(a) What probability should replace "?" in the distribution?
(b) Only English and French are considered official languages. What is the probability that a randomly chosen Canadian's mother tongue is not an official language?

4.33 Online health information. Based on a random sample of 1066 adults (18 years or older), a Harris Poll (July 13–18, 2010) estimates that 175 million U.S. adults have gone online for health information. Such individuals have been labeled as "cyberchondriacs." Cyberchondriacs in the sample were asked about the success of their online search for information about health topics. Here is the distribution of responses:[11]

	Very successful	Somewhat successful	Neither successful nor unsuccessful
Probability	0.41	0.45	0.04

	Somewhat unsuccessful	Very unsuccessful	Decline to answer
Probability	0.05	0.03	0.02

(a) Show that this is a legitimate probability distribution.
(b) What is the probability that a randomly chosen cyberchondriac feels that his or her search for health information was somewhat or very successful?

4.34 World Internet usage. Approximately 40.4% of the world's population uses the Internet (as of July 2014).[12] Furthermore, a randomly chosen Internet user has the following probabilities of being from the given country of the world:

Region	China	U.S.	India	Japan
Probability	0.2197	0.0958	0.0833	0.0374

(a) What is the probability that a randomly chosen Internet user does not live in one of the four countries listed in this table?
(b) What is the probability that a randomly chosen Internet user does not live in the United States?
(b) At *least* what proportion of Internet users are from Asia?

4.35 Modes of transportation. Governments (local and national) find it important to gather data on modes of transportation for commercial and workplace movement. Such information is useful for policymaking as it pertains to infrastructure (like roads and railways), urban development, energy use, and pollution. Based on 2011 Canadian and 2012 U.S. government data, here are the distributions of the primary means of transportation to work for employees working outside the home:[13]

	Car (self or pool)	Public transportation	Bicycle or motorcycle	Walk	Other
Canada	?	0.120	0.013	0.057	0.014
U.S.	?	0.052	0.006	0.029	0.013

(a) What is the probability that a randomly chosen Canadian employee who works outside the home uses an automobile? What is the probability that a randomly chosen U.S. employee who works outside the home uses an automobile?
(b) Transportation systems primarily based on the automobile are regarded as unsustainable because of the excessive energy consumption and the effects on the health of populations. The Canadian government includes public transit, walking, and cycles as "sustainable" modes of transportation. For both countries, determine the probability that a randomly chosen employee who works outside home uses sustainable transportation. How do you assess the relative status of sustainable transportation for these two countries?

4.36 Car colors. Choose a new car or light truck at random and note its color. Here are the probabilities of the most popular colors for cars purchased in South America in 2012:[14]

Color	Silver	White	Black	Gray	Red	Brown
Probability	0.29	0.21	0.19	0.13	0.09	0.05

(a) What is the probability that a randomly chosen car is either silver or white?
(b) In North America, the probability of a new car being blue is 0.07. What can you say about the probability of a new car in South America being blue?

4.37 Land in Iowa. Choose an acre of land in Iowa at random. The probability is 0.92 that it is farmland and 0.01 that it is forest.
(a) What is the probability that the acre chosen is not farmland?
(b) What is the probability that it is either farmland or forest?
(c) What is the probability that a randomly chosen acre in Iowa is something other than farmland or forest?

4.38 Stock market movements. You watch the price of the Dow Jones Industrial Index for four days. Give a sample space for each of the following random phenomena.
(a) You record the sequence of up-days and down-days.
(b) You record the number of up-days.

4.39 Colors of M&M'S. The colors of candies such as M&M'S are carefully chosen to match consumer preferences. The color of an M&M drawn at random from a bag has a probability distribution determined by the proportions of colors among all M&M'S of that type.
(a) Here is the distribution for plain M&M'S:

Color	Blue	Orange	Green	Brown	Yellow	Red
Probability	0.24	0.20	0.16	0.14	0.14	?

What must be the probability of drawing a red candy?
(b) What is the probability that a plain M&M is any of orange, green, or yellow?

4.40 Almond M&M'S. Exercise 4.39 gives the probabilities that an M&M candy is each of blue, orange, green, brown, yellow, and red. If "Almond" M&M'S are equally likely to be any of these colors, what is the probability of drawing a blue Almond M&M?

4.41 Legitimate probabilities? In each of the following situations, state whether or not the given assignment of probabilities to individual outcomes is legitimate—that is, satisfies the rules of probability. If not, give specific reasons for your answer.
(a) When a coin is spun, $P(\text{H}) = 0.55$ and $P(\text{T}) = 0.45$.
(b) When a coin flipped twice, $P(\text{HH}) = 0.4$, $P(\text{HT}) = 0.4$, $P(\text{TH}) = 0.4$, and $P(\text{TT}) = 0.4$.
(c) Plain M&M'S have not always had the mixture of colors given in Exercise 4.39. In the past there were no red candies and no blue candies. Tan had probability 0.10, and the other four colors had the same probabilities that are given in Exercise 4.39.

4.42 Who goes to Paris? Abby, Deborah, Sam, Tonya, and Roberto work in a firm's public relations office. Their employer must choose two of them to attend a conference in Paris. To avoid unfairness, the choice will be made by drawing two names from a hat. (This is an SRS of size 2.)
(a) Write down all possible choices of two of the five names. This is the sample space.
(b) The random drawing makes all choices equally likely. What is the probability of each choice?
(c) What is the probability that Tonya is chosen?
(d) What is the probability that neither of the two men (Sam and Roberto) is chosen?

4.43 Equally likely events. For each of the following situations, explain why you think that the events are equally likely or not.
(a) The outcome of the next tennis match for Victoria Azarenka is either a win or a loss. (You might want to check the Internet for information about this tennis player.)
(b) You draw a king or a two from a shuffled deck of 52 cards.
(c) You are observing turns at an intersection. You classify each turn as a right turn or a left turn.
(d) For college basketball games, you record the times that the home team wins and the number of times that the home team loses.

4.44 Using Internet sources. Internet sites often vanish or move, so references to them can't be followed. In fact, 13% of Internet sites referenced in major scientific journals are lost within two years after publication.
(a) If a paper contains seven Internet references, what is the probability that all seven are still good two years later?
(b) What specific assumptions did you make in order to calculate this probability?

4.45 Everyone gets audited. Wallen Accounting Services specializes in tax preparation for individual tax returns. Data collected from past records reveals that 9% of the returns prepared by Wallen have been selected for audit by the Internal Revenue Service. Today, Wallen has six new customers. Assume the chances of these six customers being audited are independent.
(a) What is the probability that all six new customers will be selected for audit?
(b) What is the probability that none of the six new customers will be selected for audit?
(c) What is the probability that exactly one of the six new customers will be selected for audit?

4.46 Hiring strategy. A chief executive officer (CEO) has resources to hire one vice president or three managers. He believes that he has probability 0.6 of successfully recruiting the vice president candidate and probability 0.8 of successfully recruiting each of the manager candidates. The three candidates for manager will make their decisions independently of each other. The CEO must successfully recruit either the vice president or all three managers to consider his hiring strategy a success. Which strategy should he choose?

4.47 A random walk on Wall Street? The "random walk" theory of securities prices holds that price

movements in disjoint time periods are independent of each other. Suppose that we record only whether the price is up or down each year and that the probability that our portfolio rises in price in any one year is 0.65. (This probability is approximately correct for a portfolio containing equal dollar amounts of all common stocks listed on the New York Stock Exchange.)
(a) What is the probability that our portfolio goes up for three consecutive years?
(b) If you know that the portfolio has risen in price two years in a row, what probability do you assign to the event that it will go down next year?
(c) What is the probability that the portfolio's value moves in the same direction in both of the next two years?

4.48 The multiplication rule for independent events. The probability that a randomly selected person prefers the vehicle color white is 0.24. Can you apply the multiplication rule for independent events in the situations described in parts (a) and (b)? If your answer is Yes, apply the rule.
(a) Two people are chosen at random from the population. What is the probability that both prefer white?
(b) Two people who are sisters are chosen. What is the probability that both prefer white?
(c) Write a short summary about the multiplication rule for independent events using your answers to parts (a) and (b) to illustrate the basic idea.

4.49 What's wrong? In each of the following scenarios, there is something wrong. Describe what is wrong and give a reason for your answer.
(a) If two events are disjoint, we can multiply their probabilities to determine the probability that they will both occur.
(b) If the probability of A is 0.6 and the probability of B is 0.5, the probability of both A and B happening is 1.1.
(c) If the probability of A is 0.35, then the probability of the complement of A is -0.35.

4.50 What's wrong? In each of the following scenarios, there is something wrong. Describe what is wrong and give a reason for your answer.
(a) If the sample space consists of two outcomes, then each outcome has probability 0.5.
(b) If we select a digit at random, then the probability of selecting a 2 is 0.2.
(c) If the probability of A is 0.2, the probability of B is 0.3, and the probability of A and B is 0.5, then A and B are independent.

4.51 Playing the lottery. An instant lottery game gives you probability 0.02 of winning on any one play. Plays are independent of each other. If you play five times, what is the probability that you win at least once?

4.52 Axioms of probability. Show that any assignment of probabilities to events that obeys Rules 2 and 3 on page 182 automatically obeys the complement rule (Rule 4). This implies that a mathematical treatment of probability can start from just Rules 1, 2, and 3. These rules are sometimes called *axioms* of probability.

4.53 Independence of complements. Show that if events A and B obey the multiplication rule, $P(A \text{ and } B) = P(A)P(B)$, then A and the complement B^c of B also obey the multiplication rule, $P(A \text{ and } B^c) = P(A)P(B^c)$. That is, if events A and B are independent, then A and B^c are also independent. (*Hint:* Start by drawing a Venn diagram and noticing that the events "A and B" and "A and B^c" are disjoint.)

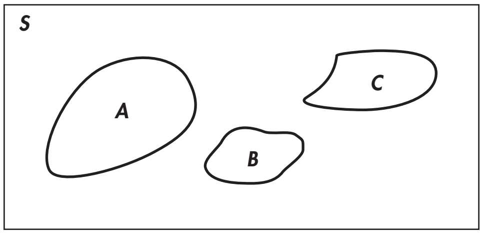
4.3 General Probability Rules

In the previous section, we met and used five basic rules of probability (page 191). To lay the groundwork for probability, we considered simplified settings such as dealing with only one or two events or the making of assumptions that the events are disjoint or independent. In this section, we learn more general laws that govern the assignment of probabilities. We learn that these more general laws of probability allows us to apply probability models to more complex random phenomena.

General addition rules

Probability has the property that if A and B are disjoint events, then $P(A \text{ or } B) = P(A) + P(B)$. What if there are more than two events or the events are not disjoint? These circumstances are covered by more general addition rules for probability.

FIGURE 4.5 The addition rule for disjoint events: $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$ when events A, B, and C are disjoint.



Union

The **union** of any collection of events is the event that at least one of the collection occurs.

For two events A and B, the union is the event $\{A \text{ or } B\}$ that A or B or both occur. From the addition rule for two disjoint events we can obtain rules for more general unions. Suppose first that we have several events—say A, B, and C—that are disjoint in pairs. That is, no two can occur simultaneously. The Venn diagram in Figure 4.5 illustrates three disjoint events. The addition rule for two disjoint events extends to the following law.

Addition Rule for Disjoint Events

If events A, B, and C are disjoint in the sense that no two have any outcomes in common, then

$$P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$$

This rule extends to any number of disjoint events.

EXAMPLE 4.16 Disjoint Events

Generate a random integer in the range of 10 to 59. What is the probability that the 10's digit will be odd? The event that the 10's digit is odd is the union of three disjoint events. These events are

$$A = \{10, 11, \ldots, 19\}$$
$$B = \{30, 31, \ldots, 39\}$$
$$C = \{50, 51, \ldots, 59\}$$

In each of these events, there are 10 outcomes out of the 50 possible outcomes. This implies $P(A) = P(B) = P(C) = 0.2$. As a result, the probability that the 10's digit is odd is

$$\begin{aligned} P(A \text{ or } B \text{ or } C) &= P(A) + P(B) + P(C) \\ &= 0.2 + 0.2 + 0.2 = 0.6 \end{aligned}$$

APPLY YOUR KNOWLEDGE

4.54 Probability that sum of dice is a multiple of 4. Suppose you roll a pair of dice and you record the sum of the dice. What is the probability that the sum is a multiple of 4?

If events A and B are not disjoint, they can occur simultaneously. The probability of their union is then *less* than the sum of their probabilities. As Figure 4.6 suggests, the outcomes common to both are counted twice when we add probabilities, so we must subtract this probability once. Here is the addition rule for the union of any two events, disjoint or not.

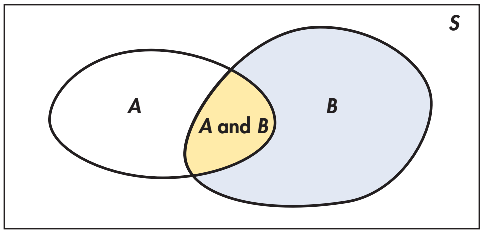


FIGURE 4.6 The general addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ for any events A and B.

General Addition Rule for Unions of Two Events

For any two events A and B,

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$

If A and B are disjoint, the event $\{A \text{ and } B\}$ that both occur has no outcomes in it. This *empty event* is the complement of the sample space S and must have probability 0. So the general addition rule includes Rule 3, the addition rule for disjoint events.

EXAMPLE 4.17 Making Partner

Deborah and Matthew are anxiously awaiting word on whether they have been made partners of their law firm. Deborah guesses that her probability of making partner is 0.7 and that Matthew's is 0.5. (These are personal probabilities reflecting Deborah's assessment of chance.) This assignment of probabilities does not give us enough information to compute the probability that at least one of the two is promoted. In particular, adding the individual probabilities of promotion gives the impossible result 1.2. If Deborah also guesses that the probability that *both* she and Matthew are made partners is 0.3, then by the general addition rule

$$P(\text{at least one is promoted}) = 0.7 + 0.5 - 0.3 = 0.9$$

The probability that *neither* is promoted is then 0.1 by the complement rule.

Venn diagrams are a great help in finding probabilities because you can just think of adding and subtracting areas. Figure 4.7 shows some events and their probabilities for Example 4.17. What is the probability that Deborah is promoted and Matthew is not?

The Venn diagram shows that this is the probability that Deborah is promoted minus the probability that both are promoted, $0.7 - 0.3 = 0.4$. Similarly, the probability that Matthew is promoted and Deborah is not is $0.5 - 0.3 = 0.2$. The four probabilities that appear in the figure add to 1 because they refer to four disjoint events that make up the entire sample space.

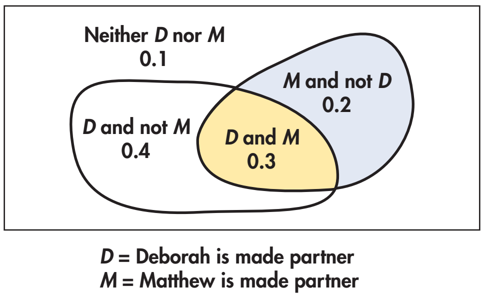


FIGURE 4.7 Venn diagram and probabilities, Example 4.17.

APPLY YOUR KNOWLEDGE

4.55 Probability that sum of dice is even or greater than 8. Suppose you roll a pair of dice and record the sum of the dice. What is the probability that the sum is even or greater than 8?

Conditional probability

The probability we assign to an event can change if we know that some other event has occurred. This idea is the key to many applications of probability. Let's first illustrate this idea with labor-related statistics.

Each month the Bureau of Labor Statistics (BLS) announces a variety of statistics on employment status in the United States. Employment statistics are important gauges of the economy as a whole. To understand the reported statistics, we need to understand how the government defines "labor force." The labor force includes all people who are either currently employed or who are jobless but are looking for jobs and are available for work. The latter group is viewed as unemployed. People who have no job and are not actively looking for one are not considered to be in the labor force. There are a variety of reasons for people not to be in the labor force, including being retired, going to school, having certain disabilities, or being too discouraged to look for a job.

EXAMPLE 4.18 Labor Rates

Averaged over the year 2013, the following table contains counts (in thousands) of persons aged 16 and older in the civilian population, classified by gender and employment status:[15]

Gender	Employed	Unemployed	Not in labor force	Civilian population
Men	76,353	6,314	35,889	118,556
Women	67,577	5,146	54,401	127,124
Total	143,930	11,460	90,290	245,680

The BLS defines the total labor force as the sum of the counts on employed and unemployed. In turn, the total labor force count plus the count of those not in the labor force equals the total civilian population. Depending on the base (total labor force or civilian population), different rates can be computed. For example, the number of people unemployed divided by the total labor force defines the unemployment rate, while the total labor force divided by the civilian population defines labor participation rate.

Randomly choose a person aged 16 or older from the civilian population. What is the probability that person is defined as labor participating? Because "choose at random" gives all 245,680,000 such persons the same chance, the probability is just the proportion that are participating. In thousands,

$$P(\text{participating}) = \frac{143{,}930 + 11{,}460}{245{,}680} = 0.632$$

This calculation does not assume anything about the gender of the person. Suppose now we are told that the person chosen is female. The probability that the person participates, *given the information that the person is female,* is

$$P(\text{participating} \mid \text{female}) = \frac{67{,}577 + 5{,}146}{127{,}124} = 0.572$$

conditional probability

The new notation $P(B \mid A)$ is a **conditional probability.** That is, it gives the probability of one event (person is labor participating) under the condition that we know another event (person is female). You can read the bar | as "given the information that."

APPLY YOUR KNOWLEDGE

4.56 Men labor participating. Refer to Example 4.18. What is the probability that a person is labor participant given the person is male?

CAUTION

Do not confuse the probabilities of $P(B \mid A)$ and $P(A \text{ and } B)$. They are generally not equal. Consider, for example, that the computed probability of 0.572 from Example 4.18 is *not* the probability that a randomly selected person from the civilian population is female and labor participating. Even though these probabilities are different, they are connected in a special way. Find first the proportion of the civilian population who are women. Then, out of the female population, find the proportion who are labor participating. Multiply the two proportions. The actual proportions from Example 4.18 are

$$P(\text{female } \textit{and} \text{ participating}) = P(\text{female}) \times P(\text{participating} \mid \text{female})$$
$$= \left(\frac{127{,}124}{245{,}680}\right)(0.572) = 0.296$$

We can check if this is correct by computing the probability directly as follows:

$$P(\text{female } \textit{and} \text{ participating}) = \frac{67{,}577 + 5{,}146}{245{,}680} = 0.296$$

We have just discovered the general multiplication rule of probability.

Multiplication Rule

The probability that both of two events A and B happen together can be found by

$$P(A \text{ and } B) = P(A)P(B \mid A)$$

Here $P(B \mid A)$ is the conditional probability that B occurs, given the information that A occurs.

EXAMPLE 4.19 Downloading Music from the Internet

The multiplication rule is just common sense made formal. For example, suppose that 29% of Internet users download music files, and 67% of downloaders say they don't care if the music is copyrighted. So the percent of Internet users who download music (event A) *and* don't care about copyright (event B) is 67% of the 29% who download, or

$$(0.67)(0.29) = 0.1943 = 19.43\%$$

The multiplication rule expresses this as

$$P(A \text{ and } B) = P(A) \times P(B \mid A)$$
$$= (0.29)(0.67) = 0.1943$$

APPLY YOUR KNOWLEDGE

4.57 Focus group probabilities. A focus group of 15 consumers has been selected to view a new TV commercial. Even though all of the participants will provide their opinion, two members of the focus group will be randomly selected and asked to answer even more detailed questions about the commercial. The group contains seven men and eight women. What is the probability that the two chosen to answer questions will both be women?

4.58 Buying from Japan. Functional Robotics Corporation buys electrical controllers from a Japanese supplier. The company's treasurer thinks that there is probability 0.4 that the dollar will fall in value against the Japanese yen in the next month. The treasurer also believes that *if* the dollar falls, there is probability 0.8 that the supplier will demand renegotiation of the contract. What probability has the treasurer assigned to the event that the dollar falls and the supplier demands renegotiation?

If $P(A)$ and $P(A \text{ and } B)$ are given, we can rearrange the multiplication rule to produce a *definition* of the conditional probability $P(B \mid A)$ in terms of unconditional probabilities.

Definition of Conditional Probability

When $P(A) > 0$, the **conditional probability** of B given A is

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

CAUTION

Be sure to keep in mind the distinct roles in $P(B \mid A)$ of the event B whose probability we are computing and the event A that represents the information we are given. The conditional probability $P(B \mid A)$ makes no sense if the event A can never occur, so we require that $P(A) > 0$ whenever we talk about $P(B \mid A)$.

EXAMPLE 4.20 College Students

Here is the distribution of U.S. college students classified by age and full-time or part-time status:

Age (years)	Full-time	Part-time
15 to 19	0.21	0.02
20 to 24	0.32	0.07
25 to 39	0.10	0.10
30 and over	0.05	0.13

Let's compute the probability that a student is aged 15 to 19, given that the student is full-time. We know that the probability that a student is full-time *and* aged 15 to 19 is 0.21 from the table of probabilities. But what we want here is a conditional probability, given that a student is full-time. Rather than asking about age among all students, we restrict our attention to the subpopulation of students who are full-time. Let

A = the student is a full-time student
B = the student is between 15 and 19 years of age

Our formula is

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

We read $P(A \text{ and } B) = 0.21$ from the table as mentioned previously. What about $P(A)$? This is the probability that a student is full-time. Notice that there are four groups of students in our table that fit this description. To find the probability needed, we add the entries:

$$P(A) = 0.21 + 0.32 + 0.10 + 0.05 = 0.68$$

We are now ready to complete the calculation of the conditional probability:

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$
$$= \frac{0.21}{0.68}$$
$$= 0.31$$

The probability that a student is 15 to 19 years of age, given that the student is full-time, is 0.31.

Here is another way to give the information in the last sentence of this example: 31% of full-time college students are 15 to 19 years old. Which way do you prefer?

APPLY YOUR KNOWLEDGE

4.59 What rule did we use? In Example 4.20, we calculated $P(A)$. What rule did we use for this calculation? Explain why this rule applies in this setting.

4.60 Find the conditional probability. Refer to Example 4.20. What is the probability that a student is part-time, given that the student is 15 to 19 years old? Explain in your own words the difference between this calculation and the one that we did in Example 4.20.

General multiplication rules

The definition of conditional probability reminds us that, in principle, all probabilities—including conditional probabilities—can be found from the assignment of probabilities to events that describe random phenomena. More often, however, conditional probabilities are part of the information given to us in a probability model, and the multiplication rule is used to compute $P(A \text{ and } B)$. This rule extends to more than two events.

The union of a collection of events is the event that *any* of them occur. Here is the corresponding term for the event that *all* of them occur.

Intersection

The **intersection** of any collection of events is the event that *all* the events occur.

To extend the multiplication rule to the probability that all of several events occur, the key is to condition each event on the occurrence of *all* the preceding events. For example, the intersection of three events A, B, and C has probability

$$P(A \text{ and } B \text{ and } C) = P(A)P(B \mid A)P(C \mid A \text{ and } B)$$

EXAMPLE 4.21 Career in Big Business: NFL

Worldwide, the sports industry has become synonymous with big business. It has been estimated by the United Nations that sports account for nearly 3% of global economic activity. The most profitable sport in the world is professional football under the management of the National Football League (NFL).[16] With multi-million-dollar signing contracts, the economic appeal of pursuing a career as a professional sports athlete is unquestionably strong. But what are the realities? Only 6.5% of high school football players go on to play at the college level. Of these, only 1.2% will play in the NFL.[17] About 40% of the NFL players have a career of more than three years. Define these events for the sport of football:

A = {competes in college}
B = {competes in the NFL}
C = {has an NFL career longer than 3 years}

What is the probability that a high school football player competes in college and then goes on to have an NFL career of more than three years? We know that

$$P(A) = 0.065$$
$$P(B \mid A) = 0.012$$
$$P(C \mid A \text{ and } B) = 0.4$$

The probability we want is, therefore,

$$\begin{aligned} P(A \text{ and } B \text{ and } C) &= P(A)P(B \mid A)P(C \mid A \text{ and } B) \\ &= 0.065 \times 0.012 \times 0.40 = 0.00031 \end{aligned}$$

Only about three of every 10,000 high school football players can expect to compete in college and have an NFL career of more than three years. High school football players would be wise to concentrate on studies rather than unrealistic hopes of fortune from pro football.

Tree diagrams

In Example 4.21, we investigated the likelihood of a high school football player going on to play collegiately and then have an NFL career of more than three years. The sports of football and basketball are unique in that players are prohibited from going straight into professional ranks from high school. Baseball, however, has no such restriction. Some baseball players might make the professional rank through the college route, while others might ultimately make it coming out of high school, often with a journey through the minor leagues.

The calculation of the probability of a baseball player becoming a professional player involves more elaborate calculation than the football scenario. We illustrate with our next example how the use of a **tree diagram** can help organize our thinking.

tree diagram

EXAMPLE 4.22 How Many Go to MLB?

For baseball, 6.8% of high school players go on to play at the college level. Of these, 9.4% will play in Major League Baseball (MLB).[18] Borrowing the notation of Example 4.21, the probability of a high school player ultimately playing professionally is $P(B)$. To find $P(B)$, consider the tree diagram shown in Figure 4.8.

Each segment in the tree is one stage of the problem. Each complete branch shows a path that a player can take. The probability written on each segment is the conditional probability that a player follows that segment given that he has reached the point from which it branches. Starting at the left, high school baseball players either do or do not compete in college. We know that the probability of competing in college is $P(A) = 0.068$, so the probability of not competing is $P(A^c) = 0.932$. These probabilities mark the leftmost branches in the tree.

Conditional on competing in college, the probability of playing in MLB is $P(B \mid A) = 0.094$. So the conditional probability of *not* playing in MLB is

$$P(B^c \mid A) = 1 - P(B \mid A) = 1 - 0.094 = 0.906$$

These conditional probabilities mark the paths branching out from A in Figure 4.8.

The lower half of the tree diagram describes players who do not compete in college (A^c). For baseball, in years past, the majority of destined professional players did not take the route through college. However, nowadays it is relatively unusual

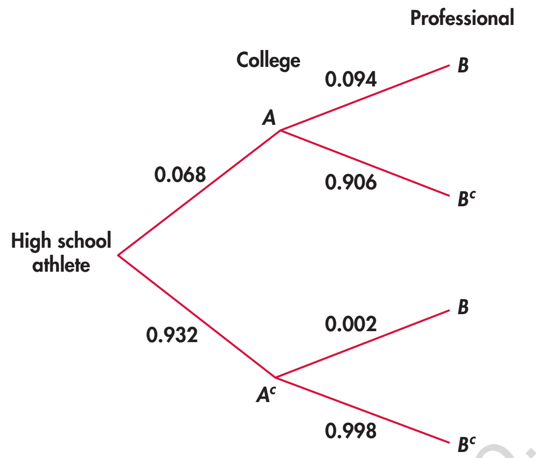


FIGURE 4.8 Tree diagram and probabilities, Example 4.22.

for players to go straight from high school to MLB. Studies have shown that the conditional probability that a high school athlete reaches MLB, given that he does not compete in college, is $P(B \mid A^c) = 0.002$.[19] We can now mark the two paths branching from A^c in Figure 4.8.

There are two disjoint paths to B (MLB play). By the addition rule, $P(B)$ is the sum of their probabilities. The probability of reaching B through college (top half of the tree) is

$$\begin{aligned} P(A \text{ and } B) &= P(A)P(B \mid A) \\ &= 0.068 \times 0.094 = 0.006392 \end{aligned}$$

The probability of reaching B without college is

$$\begin{aligned} P(A^c \text{ and } B) &= P(A^c)P(B \mid A^c) \\ &= 0.932 \times 0.002 = 0.001864 \end{aligned}$$

The final result is

$$P(B) = 0.006392 + 0.001864 = 0.008256$$

About eight high school baseball players out of 1000 will play professionally. Even though this probability is quite small, it is comparatively much greater than the chances of making it to the professional ranks in basketball and football.

It takes longer to explain a tree diagram than it does to use it. Once you have understood a problem well enough to draw the tree, the rest is easy. Tree diagrams combine the addition and multiplication rules. The multiplication rule says that the probability of reaching the end of any complete branch is the product of the probabilities written on its segments. The probability of any outcome, such as the event B that a high school baseball player plays in MLB, is then found by adding the probabilities of all branches that are part of that event.

APPLY YOUR KNOWLEDGE

4.61 Labor rates. Refer to the labor data in Example 4.18 (page 197). Draw a tree diagram with the first-stage branches being gender. Then, off the gender branches, draw two branches as the outcomes being "labor force participating" versus "not in the labor force." Show how the tree would be used to compute the probability that a randomly chosen person is labor force participating.

Bayes's rule

There is another kind of probability question that we might ask in the context of studies of athletes. Our earlier calculations look forward toward professional sports as the final stage of an athlete's career. Now let's concentrate on professional athletes and look back at their earlier careers.

EXAMPLE 4.23 Professional Athletes' Pasts

What proportion of professional athletes competed in college? In the notation of Examples 4.21 and 4.22, this is the conditional probability $P(A \mid B)$. Before we compute this probability, let's take stock of a few facts. First, the multiplication rule tells us

$$P(A \text{ and } B) = P(A)P(B \mid A)$$

We know the probabilities $P(A)$ and $P(A^c)$ that a high school baseball player does and does not compete in college. We also know the conditional probabilities $P(B \mid A)$ and $P(B \mid A^c)$ that a player from each group reaches MLB. Example 4.22 shows how to use this information to calculate $P(B)$. The method can be summarized in a single expression that adds the probabilities of the two paths to B in the tree diagram:

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

Combining these facts, we can now make the following computation:

$$\begin{aligned} P(A \mid B) &= \frac{P(A \text{ and } B)}{P(B)} \\ &= \frac{P(A)\,P(B \mid A)}{P(A)\,P(B \mid A) + P(A^c)\,P(B \mid A^c)} \\ &= \frac{0.068 \times 0.094}{0.068 \times 0.094 + 0.932 \times 0.002} \\ &= 0.774 \end{aligned}$$

About 77% of MLB players competed in college.

In calculating the "reverse" conditional probability of Example 4.23, we had two disjoint events in A and A^c whose probabilities add to exactly 1. We also had the conditional probabilities of event B given each of the disjoint events. More generally, there can be applications in which we have more than two disjoint events whose probabilities add up to 1. Put in general notation, we have another probability law.

Bayes's Rule

Suppose that $A_1, A_2, \ldots, A_k$ are disjoint events whose probabilities are not 0 and add to exactly 1. That is, any outcome is in exactly one of these events. Then, if B is any other event whose probability is not 0 or 1,

$$P(A_i \mid B) = \frac{P(B \mid A_i)\,P(A_i)}{P(B \mid A_1)\,P(A_1) + P(B \mid A_2)\,P(A_2) + \cdots + P(B \mid A_k)\,P(A_k)}$$

The numerator in Bayes's rule is always one of the terms in the sum that makes up the denominator. The rule is named after Thomas Bayes, who wrestled with arguing from outcomes like event B back to the A_i in a book published in 1763. Our next example utilizes Bayes's rule with several disjoint events.

EXAMPLE 4.24 Credit Ratings

Corporate bonds are assigned a credit rating that provides investors with a guide of the general creditworthiness of a corporation as a whole. The most well-known credit rating agencies are Moody's, Standard & Poor's, and Fitch. These rating agencies assign a letter grade to the bond issuer. For example, Fitch uses the letter classifications of AAA, AA, A, BBB, BB, B, CCC, and D. Over time, the credit ratings of the corporation can change. Credit rating specialists use the terms of "credit migration" or "transition rate" to indicate the probability of a corporation going from letter grade to letter grade over some particular span of time. For example, based on a large amount of data from 1990 to 2013, Fitch estimates that the five-year transition rates to be graded AA in the fifth year based on each of the current ("first year") grades to be:[20]

Current rating	AA (in 5th year)
AAA	0.2283
AA	0.6241
A	0.0740
BBB	0.0071
BB	0.0012
B	0.0000
CCC	0.0000
D	0.0000

Recognize that these values represent conditional probabilities. For example, P(AA rating in 5 years | AAA rating currently) = 0.2283. In the financial institution sector, the distribution of grades for year 2013 are

Rating	AAA	AA	A	BBB	BB	B	CCC	D
Proportion	0.010	0.066	0.328	0.358	0.127	0.106	0.004	0.001

The transition rates give us probabilities rating changes moving *forward*. An interesting question is where might a corporation have come from looking back retrospectively. Imagine yourself now in year 2018, and you randomly pick a financial institution that has a AA rating. What is the probability that institution had a AA rating in year 2013? A knee jerk reaction might be to answer 0.6241; however, that would be incorrect. Define these events:

$$AA13 = \{\text{rated AA in year 2013}\}$$
$$AA18 = \{\text{rated AA in year 2018}\}$$

We are seeking $P(AA13 \mid AA18)$ while the transition table gives us $P(AA18 \mid AA13)$. From the distribution of grades for 2013, we have $P(AA13) = 0.066$. Because grades are disjoint and their probabilities add to 1, we can employ Bayes's rule. It will be convenient to present the calculations of the terms in Bayes's rule as a table.

2013 grade	P(2013 grade)	$P(AA18 \mid$ 2013 grade)	$P(AA18 \mid$ 2013 grade) P(2013 grade)
AAA	0.010	0.2283	(0.2283)(0.010) = 0.002283
AA	0.066	0.6241	(0.6241)(0.066) = 0.041191
A	0.328	0.0740	(0.0740)(0.328) = 0.024272
BBB	0.358	0.0071	(0.0071)(0.358) = 0.002542

(Continued)

2013 grade	P(2013 grade)	$P(AA18 \mid 2013 \text{ grade})$	$P(AA18 \mid 2013 \text{ grade})\,P(2013 \text{ grade})$
BB	0.127	0.0012	(0.0012)(0.127) = 0.000152
B	0.106	0.0000	(0.0000)(0.106) = 0
CCC	0.004	0.0000	(0.0000)(0.004) = 0
D	0.001	0.0000	(0.0000)(0.001) = 0

Here is the computation of the desired probability using Bayes's rule along with the preceding computed values:

$$P(AA13 \mid AA18) = \frac{P(AA13)P(AA18 \mid AA13)}{P(AA18)}$$

$$= \frac{0.041191}{0.002283 + 0.041191 + 0.024272 + 0.002542 + 0.000152 + 0 + 0 + 0}$$

$$= \frac{0.041191}{0.07044}$$

$$= 0.5848$$

The probability is 0.5848, *not* 0.6241, that a corporation rated AA in 2018 was rated AA five years earlier in 2013. *This example demonstrates the important general caution that we must not confuse* $P(A \mid B)$ *with* $P(B \mid A)$.

Independence again

The conditional probability $P(B \mid A)$ is generally not equal to the unconditional probability $P(B)$. That is because the occurrence of event A generally gives us some additional information about whether or not event B occurs. If knowing that A occurs gives no additional information about B, then A and B are independent events. The formal definition of independence is expressed in terms of conditional probability.

Independent Events

Two events A and B that both have positive probability are **independent** if

$$P(B \mid A) = P(B)$$

This definition makes precise the informal description of independence given in Section 4.2. We now see that the multiplication rule for independent events, $P(A \text{ and } B) = P(A)P(B)$, is a special case of the general multiplication rule, $P(A \text{ and } B) = P(A)P(B \mid A)$, just as the addition rule for disjoint events is a special case of the general addition rule.

SECTION 4.3 Summary

- The **complement** A^c of an event A contains all outcomes that are not in A. The **union** $\{A \text{ or } B\}$ of events A and B contains all outcomes in A, in B, and in both A and B. The **intersection** $\{A \text{ and } B\}$ contains all outcomes that are in both A and B, but not outcomes in A alone or B alone.
- The **conditional probability** $P(B \mid A)$ of an event B, given an event A, is defined by

$$P(B \mid A) = \frac{P(A \text{ and } B)}{P(A)}$$

when $P(A) > 0$. In practice, conditional probabilities are most often found from directly available information.

- The essential general rules of elementary probability are

  Legitimate values: $0 \le P(A) \le 1$ for any event A
  Total probability 1: $P(S) = 1$
  Complement rule: $P(A^c) = 1 - P(A)$
  Addition rule: $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$
  Multiplication rule: $P(A \text{ and } B) = P(A)P(B \mid A)$

- If A and B are **disjoint,** then $P(A \text{ and } B) = 0$. The general addition rule for unions then becomes the special addition rule, $P(A \text{ or } B) = P(A) + P(B)$.

- A and B are **independent** when $P(B \mid A) = P(B)$. The multiplication rule for intersections then becomes $P(A \text{ and } B) = P(A)P(B)$.

- In problems with several stages, draw a **tree diagram** to organize use of the multiplication and addition rules.

- If $A_1, A_2, \ldots, A_k$ are disjoint events whose probabilities are not 0 and add to exactly 1 and if B is any other event whose probability is not 0 or 1, then **Bayes's rule** can be used to calculate $P(A_i \mid B)$ as follows:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \cdots + P(B \mid A_k)\,P(A_k)}$$

SECTION 4.3 Exercises

For Exercise 4.54, see page 195; for 4.55, see page 197; for 4.56, see page 198; for 4.57 and 4.58, see pages 198–199; for 4.59 and 4.60, see page 200; and for 4.61, see page 202.

4.62 Find and explain some probabilities.
(a) Can we have an event A that has negative probability? Explain your answer.
(b) Suppose $P(A) = 0.2$ and $P(B) = 0.4$. Explain what it means for A and B to be disjoint. Assuming that they are disjoint, find the probability that A or B occurs.
(c) Explain in your own words the meaning of the rule $P(S) = 1$.
(d) Consider an event A. What is the name for the event that A does not occur? If $P(A) = 0.3$, what is the probability that A does not occur?
(e) Suppose that A and B are independent and that $P(A) = 0.2$ and $P(B) = 0.5$. Explain the meaning of the event $\{A \text{ and } B\}$, and find its probability.

4.63 Unions.
(a) Assume that $P(A) = 0.4$, $P(B) = 0.3$, and $P(C) = 0.1$. If the events A, B, and C are disjoint, find the probability that the union of these events occurs.
(b) Draw a Venn diagram to illustrate your answer to part (a).
(c) Find the probability of the complement of the union of A, B, and C.

4.64 Conditional probabilities. Suppose that $P(A) = 0.5$, $P(B) = 0.3$, and $P(B \mid A) = 0.2$.
(a) Find the probability that both A and B occur.
(b) Use a Venn diagram to explain your calculation.
(c) What is the probability of the event that B occurs and A does not?

4.65 Find the probabilities. Suppose that the probability that A occurs is 0.6 and the probability that A and B occur is 0.5.
(a) Find the probability that B occurs given that A occurs.
(b) Illustrate your calculations in part (a) using a Venn diagram.

4.66 What's wrong? In each of the following scenarios, there is something wrong. Describe what is wrong and give a reason for your answer.
(a) $P(A \text{ or } B)$ is always equal to the sum of $P(A)$ and $P(B)$.
(b) The probability of an event minus the probability of its complement is always equal to 1.
(c) Two events are disjoint if $P(B \mid A) = P(B)$.

4.67 Attendance at two-year and four-year colleges. In a large national population of college students, 61% attend four-year institutions and the rest attend two-year institutions. Males make up 44% of the students in the four-year institutions and 41% of the students in the two-year institutions.

(a) Find the four probabilities for each combination of gender and type of institution in the following table. Be sure that your probabilities sum to 1.

	Men	Women
Four-year institution		
Two-year institution		

(b) Consider randomly selecting a female student from this population. What is the probability that she attends a four-year institution?

4.68 Draw a tree diagram. Refer to the previous exercise. Draw a tree diagram to illustrate the probabilities in a situation in which you first identify the type of institution attended and then identify the gender of the student.

4.69 Draw a different tree diagram for the same setting. Refer to the previous two exercises. Draw a tree diagram to illustrate the probabilities in a situation in which you first identify the gender of the student and then identify the type of institution attended. Explain why the probabilities in this tree diagram are different from those that you used in the previous exercise.

4.70 Education and income. Call a household prosperous if its income exceeds \$100,000. Call the household educated if at least one of the householders completed college. Select an American household at random, and let A be the event that the selected household is prosperous and B the event that it is educated. According to the Current Population Survey, $P(A) = 0.138$, $P(B) = 0.261$, and the probability that a household is both prosperous and educated is $P(A \text{ and } B) = 0.082$. What is the probability $P(A \text{ or } B)$ that the household selected is either prosperous or educated?

4.71 Find a conditional probability. In the setting of the previous exercise, what is the conditional probability that a household is prosperous, given that it is educated? Explain why your result shows that events A and B are not independent.

4.72 Draw a Venn diagram. Draw a Venn diagram that shows the relation between the events A and B in Exercise 4.70. Indicate each of the following events on your diagram and use the information in Exercise 4.70 to calculate the probability of each event. Finally, describe in words what each event is.

(a) $\{A \text{ and } B\}$.
(b) $\{A^c \text{ and } B\}$.
(c) $\{A \text{ and } B^c\}$.
(d) $\{A^c \text{ and } B^c\}$.

4.73 Sales of cars and light trucks. Motor vehicles sold to individuals are classified as either cars or light trucks (including SUVs) and as either domestic or imported. In a recent year, 69% of vehicles sold were light trucks, 78% were domestic, and 55% were domestic light trucks. Let A be the event that a vehicle is a car and B the event that it is imported. Write each of the following events in set notation and give its probability.

(a) The vehicle is a light truck.
(b) The vehicle is an imported car.

4.74 Conditional probabilities and independence. Using the information in Exercise 4.73, answer these questions.

(a) Given that a vehicle is imported, what is the conditional probability that it is a light truck?
(b) Are the events "vehicle is a light truck" and "vehicle is imported" independent? Justify your answer.

4.75 Unemployment rates. As noted in Example 4.18 (page 197), in the language of government statistics, you are "in the labor force" if you are available for work and either working or actively seeking work. The unemployment rate is the proportion of the labor force (not of the entire population) who are unemployed. Based on the table given in Example 4.18, find the unemployment rate for people with each gender. How does the unemployment rate change with gender? Explain carefully why your results suggest that gender and being employed are not independent.

4.76 Loan officer decision. A loan officer is considering a loan request from a customer of the bank. Based on data collected from the bank's records over many years, there is an 8% chance that a customer who has overdrawn an account will default on the loan. However, there is only a 0.6% chance that a customer who has never overdrawn an account will default on the loan. Based on the customer's credit history, the loan officer believes there is a 40% chance that this customer will overdraw his account. Let D be the event that the customer defaults on the loan, and let O be the event that the customer overdraws his account.

(a) Express the three probabilities given in the problem in the notation of probability and conditional probability.
(b) What is the probability that the customer will default on the loan?

4.77 Loan officer decision. Considering the information provided in the previous exercise, calculate $P(O \mid D)$. Show your work. Also, express this probability in words in the context of the loan officer's decision. If new information about the customer becomes available before the loan officer makes her decision, and if this information indicates that there is only a 25% chance that this customer will overdraw his account rather than a 40% chance, how does this change $P(O \mid D)$?

4.78 High school football players. Using the information in Example 4.21 (pages 200–201), determine the proportion of high school football players expected to play professionally in the NFL.

4.79 High school baseball players. It is estimated that 56% of MLB players have careers of three or more years. Using the information in Example 4.22 (pages 201–202), determine the proportion of high school players expected to play three or more years in MLB.

4.80 Telemarketing. A telemarketing company calls telephone numbers chosen at random. It finds that 70% of calls are not completed (the party does not answer or refuses to talk), that 20% result in talking to a woman, and that 10% result in talking to a man. After that point, 30% of the women and 20% of the men actually buy something. What percent of calls result in a sale? (Draw a tree diagram.)

4.81 Preparing for the GMAT. A company that offers courses to prepare would-be MBA students for the GMAT examination finds that 40% of its customers are currently undergraduate students and 60% are college graduates. After completing the course, 50% of the undergraduates and 70% of the graduates achieve scores of at least 600 on the GMAT. Use a tree diagram to organize this information.
(a) What percent of customers are undergraduates *and* score at least 600? What percent of customers are graduates *and* score at least 600?
(b) What percent of all customers score at least 600 on the GMAT?

4.82 Sales to women. In the setting of Exercise 4.80, what percent of sales are made to women? (Write this as a conditional probability.)

4.83 Success on the GMAT. In the setting of Exercise 4.81, what percent of the customers who score at least 600 on the GMAT are undergraduates? (Write this as a conditional probability.)

4.84 Successful bids. Consolidated Builders has bid on two large construction projects. The company president believes that the probability of winning the first contract (event A) is 0.6, that the probability of winning the second (event B) is 0.5, and that the probability of winning both jobs (event $\{A \text{ and } B\}$) is 0.3. What is the probability of the event $\{A \text{ or } B\}$ that Consolidated will win at least one of the jobs?

4.85 Independence? In the setting of the previous exercise, are events A and B independent? Do a calculation that proves your answer.

4.86 Successful bids, continued. Draw a Venn diagram that illustrates the relation between events A and B in Exercise 4.84. Write each of the following events in terms of A, B, A^c, and B^c. Indicate the events on your diagram and use the information in Exercise 4.84 to calculate the probability of each.
(a) Consolidated wins both jobs.
(b) Consolidated wins the first job but not the second.
(c) Consolidated does not win the first job but does win the second.
(d) Consolidated does not win either job.

4.87 Credit card defaults. The credit manager for a local department store is interested in customers who default (ultimately failed to pay entire balance). Of those customers who default, 88% were late (by a week or more) with two or more monthly payments. This prompts the manager to suggest that future credit be denied to any customer who is late with two monthly payments. Further study shows that 3% of all credit customers default on their payments and 40% of those who have not defaulted have had at least two late monthly payments in the past.
(a) What is the probability that a customer who has two or more late payments will default?
(b) Under the credit manager's policy, in a group of 100 customers who have their future credit denied, how many would we expect *not* to default on their payments?
(c) Does the credit manager's policy seem reasonable? Explain your response.

4.88 Examined by the IRS. The IRS examines (audits) some tax returns in greater detail to verify that the tax reported is correct. The rates of examination vary depending on the size of the individual's adjusted gross income. In 2014, the IRS reported the percentages of total returns by adjusted gross income categories and the examination coverage (%) of returns within the given income category:[21]

Income ($)	Returns filed (%)	Examination coverage (%)
None	2.08	6.04
1 under 25K	39.91	1.00
25K under 50K	23.55	0.62
50K under 75K	13.02	0.60
75K under 100K	8.12	0.58
100K under 200K	10.10	0.77
200K under 500K	2.60	2.06

(*Continued*)

Income ($)	Returns filed (%)	Examination coverage (%)
500K under 1MM	0.41	3.79
1MM under 5MM	0.19	9.02
5MM under 10MM	0.01	15.98
10MM or more	0.01	24.16

(a) Suppose a 2013 return is randomly selected and it was examined by the IRS. Use Bayes's rule to determine the probability that the individual's adjusted gross income falls in the range of $5 to $10 million. Compute the probability to at least the thousandths place.
(b) The IRS reports that 0.96% of all returns are examined. With the information provided, show how you can arrive at this reported percent.

4.89 Supplier Quality. A manufacturer of an assembly product uses three different suppliers for a particular component. By means of supplier audits, the manufacturer estimates the following percentages of defective parts by supplier:

Supplier	1	2	3
Percent defective	0.4%	0.3%	0.6%

Shipments from the suppliers are continually streaming to the manufacturer in small lots from each of the suppliers. As a result, the inventory of parts held by the manufacturer is a mix of parts representing the relative supplier rate from each supplier. In current inventory, there are 423 parts from Supplier 1, 367 parts from Supplier 2, and 205 parts from Supplier 3. Suppose a part is randomly chosen from inventory. Define "S1" as the event the part came from Supplier 1, "S2" as the event the part came from Supplier 2, and "S3" as the event the part came from Supplier 3. Also, define "D" as the event the part is defective.
(a) Based on the inventory mix, determine $P(S1)$, $P(S2)$, and $P(S3)$.
(b) If the part is found to be defective, use Bayes's rule to determine the probability that it came from Supplier 3.

4.4 Random Variables

Sample spaces need not consist of numbers. When we toss a coin four times, we can record the outcome as a string of heads and tails, such as HTTH. In statistics, however, we are most often interested in numerical outcomes such as the count of heads in the four tosses. It is convenient to use a shorthand notation: Let X be the number of heads. If our outcome is HTTH, then $X = 2$. If the next outcome is TTTH, the value of X changes to $X = 1$. The possible values of X are 0, 1, 2, 3, and 4. Tossing a coin four times will give X one of these possible values. Tossing four more times will give X another and probably different value. We call X a *random variable* because its values vary when the coin tossing is repeated.

Random Variable

A **random variable** is a variable whose value is a numerical outcome of a random phenomenon.

In the preceding coin-tossing example, the random variable is the number of heads in the four tosses.

We usually denote random variables by capital letters near the end of the alphabet, such as X or Y. Of course, the random variables of greatest interest to us are outcomes such as the mean $\bar{x}$ of a random sample, for which we will keep the familiar notation.[22] As we progress from general rules of probability toward statistical inference, we will concentrate on random variables.

With a random variable X, the sample space S just lists the possible values of the random variable. We usually do not mention S separately. There remains the second part of any probability model, the assignment of probabilities to events. There are two main ways of assigning probabilities to the values of a random variable. The two types of probability models that result will dominate our application of probability to statistical inference.

Discrete random variables

We have learned several rules of probability, but only one method of assigning probabilities: state the probabilities of the individual outcomes and assign probabilities to events by summing over the outcomes. The outcome probabilities must be between 0 and 1 and have sum 1. When the outcomes are numerical, they are values of a random variable. We now attach a name to random variables having probability assigned in this way.

Discrete Random Variable

A **discrete random variable** X has possible values that can be given in an ordered list. The **probability distribution** of X lists the values and their probabilities:

Value of X	x_1	x_2	x_3	. . .
Probability	p_1	p_2	p_3	. . .

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. The sum of the probabilities is 1; $p_1 + p_2 + \cdots = 1$.

Find the probability of any event by adding the probabilities p_i of the particular values x_i that make up the event.

In most of the situations that we will study, the number of possible values is a finite number, k. Think about the number of heads in four tosses of a coin. In this case, $k = 5$ with X taking the possible values of 0, 1, 2, 3, and 4.

However, there are settings in which the number of possible values can be infinite. Think about counting the number of tosses of a coin until you get a head. In this case, the set of possible values for X is given by $\{1, 2, 3, \ldots\}$. As another example, suppose X represents the number of complaining customers to a retail store during a certain time period. Now, the set of possible values for X is given by $\{0, 1, 2, \ldots\}$. In both of these examples, we say that there is a **countably infinite** number of possible values. Simply defined, *countably infinite* means that we can correspond each possible outcome to the counting or natural numbers of $\{0, 1, 2, \ldots\}$.

countably infinite

In summary, a discrete random variable either has a finite number of possible values or has a countably infinite number of possible values.


CHAIKOM/SHUTTERSTOCK

CASE 4.2

Tracking Perishable Demand Whether a business is in manufacturing, retailing, or service, there is inevitably the need to hold inventory to meet demand on the items held in stock. One of most basic decisions in the control of an inventory management system is the decision of how many items should be ordered to be stocked. Ordering too much leads to unnecessary inventory costs, while ordering too little risks the organization to stock-out situations.

Hospitals have a unique challenge in the inventory management of blood. Blood is a perishable product, and hence a blood inventory management is a trade-off between shortage and wastage. The demand for blood and its components fluctuates. Hospitals routinely track daily blood demand to estimate rates of usage so that they can manage their blood inventory.

For this case, we consider the daily usage of red blood cells (RBC) O + transfusion blood bags collected from a Midwest hospital.[23] These transfusion data are categorized as "new-aged" blood cells, which are used for the most critical patients, such as cancer and immune-deficient patients. If these blood cells are unused by day's end, then they are downgraded to the category of medium-aged blood cells. Here is the distribution of the number of bags X used in a day:

Bags used	0	1	2	3	4	5	6
Probability	0.202	0.159	0.201	0.125	0.088	0.087	0.056

Bags used	7	8	9	10	11	12
Probability	0.025	0.022	0.018	0.008	0.006	0.003

probability histogram

We can use histograms to show probability distributions as well as distributions of data. Figure 4.9 displays the **probability histogram** of the blood bag probabilities. The height of each bar shows the probability of the outcome at its base. Because the heights are probabilities, they add to 1. As usual, all the bars in a histogram have the same width. So the areas also display the assignment of probability to outcomes. For the blood bag distribution, we can visually see that more than 50% of the distribution is less than or equal to two bags and the distribution is generally skewed to the right. Histograms can also make it easy to quickly compare the two distributions. For example, Figure 4.10 compares the probability model for equally likely random digits (Example 4.13) (pages 186–187) with the model given by Benford's law (Case 4.1) (pages 184–185).

EXAMPLE 4.25 Demand of at Least One Bag?

CASE 4.2 Consider the event that daily demand is at least one bag. In the language of random variables,

$$P(X \geq 1) = P(X = 1) + P(X = 2) + \cdots + P(X = 11) + P(X = 12)$$
$$= 0.159 + 0.201 + \cdots + 0.006 + 0.003 = 0.798$$

The adding of 12 probabilities is a bit of a tedious affair. But there is a much easier way to get at the ultimate probability when we think about the complement rule. The probability of at least one bag demanded is more simply found as follows:

$$P(X \geq 1) = 1 - P(X = 0)$$
$$= 1 - 0.202 = 0.798$$

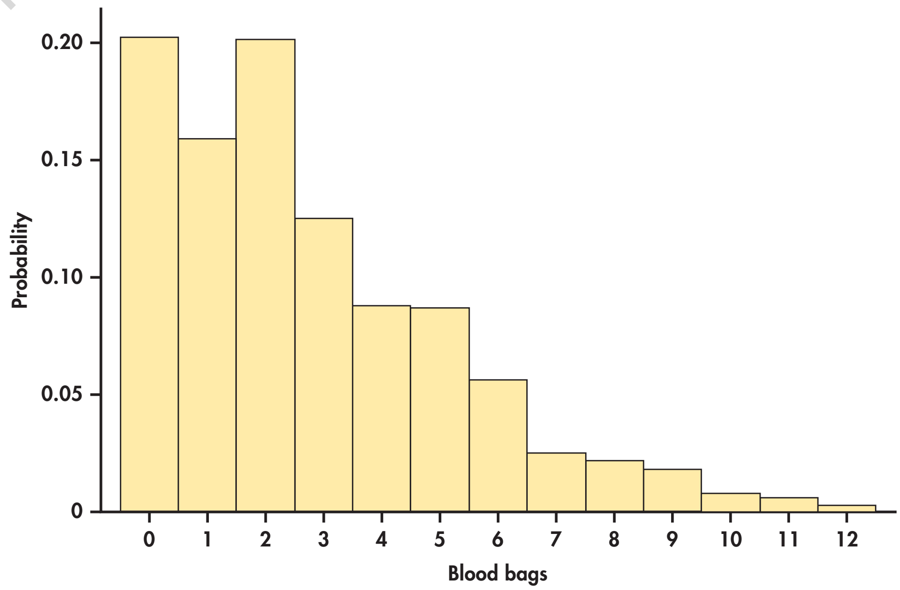


FIGURE 4.9 Probability histogram for blood bag demand probabilities. The height of each bar shows the probability assigned to a single outcome.

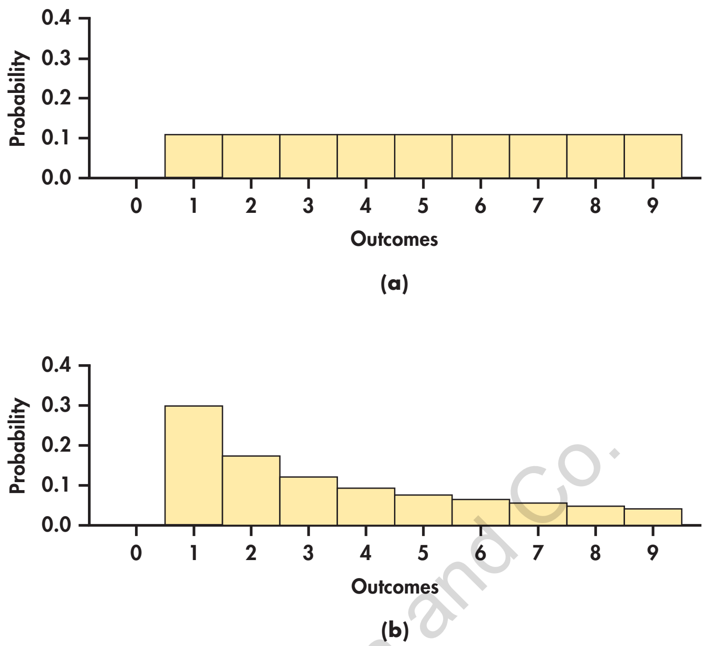


FIGURE 4.10 Probability histograms: (a) equally likely random digits 1 to 9; and (b) Benford's law.

With our discussions of discrete random variables in this chapter, it is important to note that our goal is for you to gain a base understanding of discrete random variables and how to work with them. In Chapter 5, we introduce you to two important discrete distributions, known as the binomial and Poisson distributions, that have wide application in business.

APPLY YOUR KNOWLEDGE

CASE 4.2 **4.90 High demand.** Refer to Case 4.2 for the probability distribution on daily demand for blood transfusion bags.

(a) What is the probability that the hospital will face a high demand of either 11 or 12 bags? Compute this probability directly using the respective probabilities for 11 and 12 bags.
(b) Now show how the complement rule would be used to find the same probability of part (a).
(c) Consider the calculations of parts (a) and (b) and the calculations of Example 4.25 (page 211). Explain under what circumstances does the use of the complement rule ease computations?

4.91 How many cars? Choose an American household at random and let the random variable X be the number of cars (including SUVs and light trucks) they own. Here is the probability model if we ignore the few households that own more than five cars:

Number of cars X	0	1	2	3	4	5
Probability	0.09	0.36	0.35	0.13	0.05	0.02

(a) Verify that this is a legitimate discrete distribution. Display the distribution in a probability histogram.
(b) Say in words what the event $\{X \geq 1\}$ is. Find $P(X \geq 1)$.
(c) Your company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

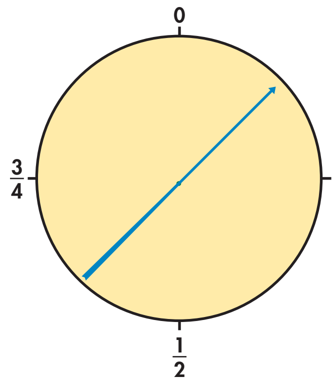


FIGURE 4.11 A spinner that generates a random number between 0 and 1.

Continuous random variables

When we use the table of random digits to select a digit between 0 and 9, the result is a discrete random variable. The probability model assigns probability 1/10 to each of the 10 possible outcomes. Suppose that we want to choose a number at random between 0 and 1, allowing *any* number between 0 and 1 as the outcome. Software random number generators will do this.

You can visualize such a random number by thinking of a spinner (Figure 4.11) that turns freely on its axis and slowly comes to a stop. The pointer can come to rest anywhere on a circle that is marked from 0 to 1. The sample space is now an interval of numbers:

$$S = \{\text{all numbers } x \text{ such that } 0 \leq x \leq 1\}$$

How can we assign probabilities to events such as $\{0.3 \leq x \leq 0.7\}$? As in the case of selecting a random digit, we would like all possible outcomes to be equally likely. But we cannot assign probabilities to each individual value of x and then sum, because there are infinitely many possible values.

Earlier, we noted that there are situations in which discrete random variables can take on an infinite number of possible values corresponding to the set of counting numbers $\{0, 1, 2, \ldots\}$. However, the infinity associated with the spinner's possible outcomes is a different infinity. There is no way to correspond the infinite number of decimal values in range from 0 to 1 to the counting numbers. We are dealing with the possible outcomes being associated with the *real numbers* as opposed to the counting numbers. As such, we say here that there is an **uncountably infinite** number of possible values.

uncountably infinite

In light of these facts, we need to use a new way of assigning probabilities directly to events—as *areas under a density curve.* Any density curve has area exactly 1 underneath it, corresponding to total probability 1.

EXAMPLE 4.26 Uniform Random Numbers

The random number generator will spread its output uniformly across the entire interval from 0 to 1 as we allow it to generate a long sequence of numbers. The results of many trials are represented by the density curve of a **uniform distribution.**

uniform distribution

This density curve appears in red in Figure 4.12. It has height 1 over the interval from 0 to 1, and height 0 everywhere else. The area under the density curve is 1: the area of a rectangle with base 1 and height 1. The probability of any event is the area under the density curve and above the event in question.

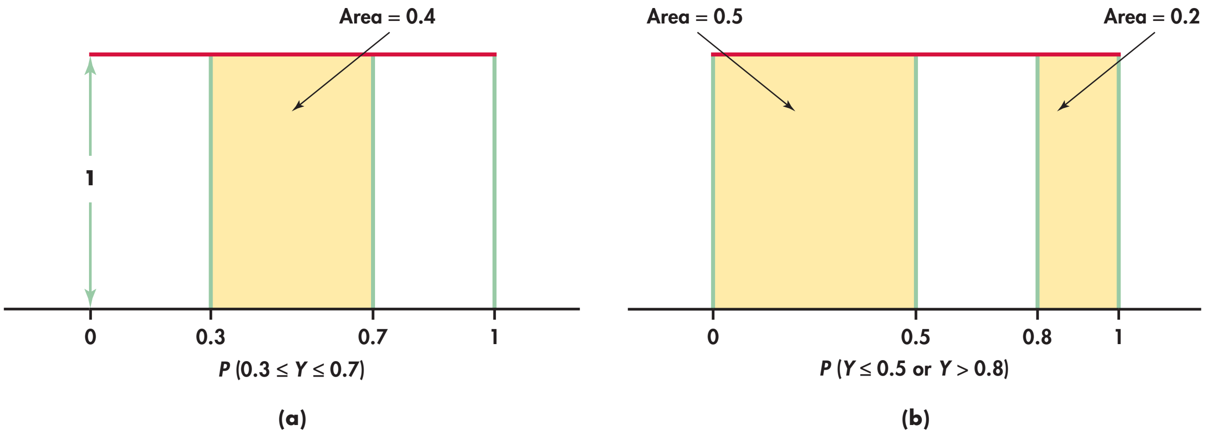


FIGURE 4.12 Assigning probabilities for generating a random number between 0 and 1, Example 4.26. The probability of any interval of numbers is the area above the interval and under the density curve.

As Figure 4.12(a) illustrates, the probability that the random number generator produces a number X between 0.3 and 0.7 is

$$P(0.3 \leq X \leq 0.7) = 0.4$$

because the area under the density curve and above the interval from 0.3 to 0.7 is 0.4. The height of the density curve is 1, and the area of a rectangle is the product of height and length, so the probability of any interval of outcomes is just the length of the interval.

Similarly,

$$P(X \leq 0.5) = 0.5$$
$$P(X > 0.8) = 0.2$$
$$P(X \leq 0.5 \text{ or } X > 0.8) = 0.7$$

Notice that the last event consists of two nonoverlapping intervals, so the total area above the event is found by adding two areas, as illustrated by Figure 4.12(b). This assignment of probabilities obeys all of our rules for probability.

APPLY YOUR KNOWLEDGE

4.92 Find the probability. For the uniform distribution described in Example 4.26, find the probability that X is between 0.2 and 0.7.

Probability as area under a density curve is a second important way of assigning probabilities to events. Figure 4.13 illustrates this idea in general form. We call X in Example 4.26 a *continuous random variable* because its values are not isolated numbers but an interval of numbers.

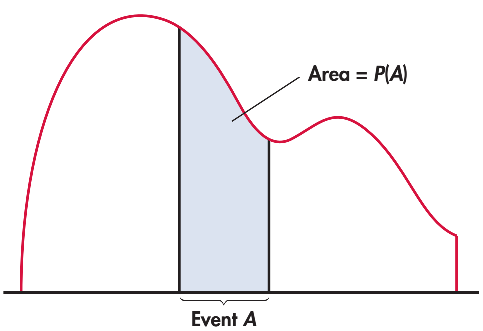


FIGURE 4.13 The probability distribution of a continuous random variable assigns probabilities as areas under a density curve. The total area under any density curve is 1.

Continuous Random Variable

A **continuous random variable** X takes all values in an interval of numbers. The **probability distribution** of X is described by a density curve. The probability of any event is the area under the density curve and above the values of X that make up the event.

The probability model for a continuous random variable assigns probabilities to intervals of outcomes rather than to individual outcomes. In fact, **all continuous probability distributions assign probability 0 to every individual outcome.** Only intervals of values have positive probability. To see that this is true, consider a specific outcome such as $P(X = 0.8)$ in the context of Example 4.26. The probability of any interval is the same as its length. The point 0.8 has no length, so its probability is 0.

Although this fact may seem odd, it makes intuitive, as well as mathematical, sense. The random number generator produces a number between 0.79 and 0.81 with probability 0.02. An outcome between 0.799 and 0.801 has probability 0.002. A result between 0.799999 and 0.800001 has probability 0.000002. You see that as we approach 0.8 the probability gets closer to 0.

To be consistent, the probability of an outcome *exactly* equal to 0.8 must be 0. Because there is no probability exactly at $X = 0.8$, the two events $\{X > 0.8\}$ and $\{X \geq 0.8\}$ have the same probability. In general, we can ignore the distinction between $>$ and $\geq$ when finding probabilities for continuous random variables. Similarly, we can also ignore the distinction between $<$ and $\leq$ in the continuous case. *However, when dealing with discrete random variables, we cannot ignore these distinctions. Thus, it is important to be alert as to whether you are dealing with continuous or discrete random variables when doing probability calculations.*

CAUTION

Normal distributions as probability distributions

The density curves that are most familiar to us are the Normal curves. Because any density curve describes an assignment of probabilities, *Normal distributions are probability distributions.* Recall from Section 1.4 (page 44) that $N(\mu, \sigma)$ is our shorthand for the Normal distribution having mean μ and standard deviation σ. In the language of random variables, if X has the $N(\mu, \sigma)$ distribution, then the standardized variable

$$Z = \frac{X - \mu}{\sigma}$$

is a standard Normal random variable having the distribution $N(0, 1)$.

REMINDER standard Normal distribution, p. 46

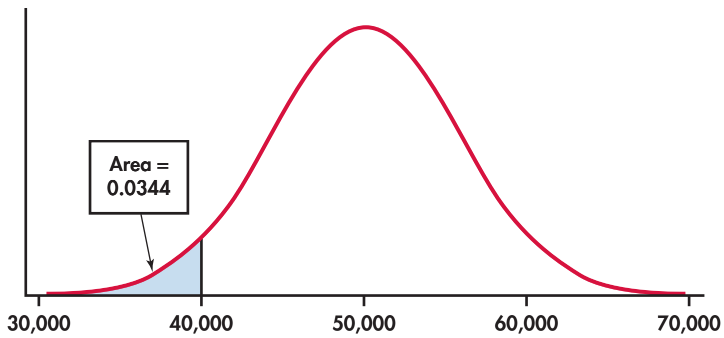
EXAMPLE 4.27 Tread Life

The actual tread life X of a 40,000-mile automobile tire has a Normal probability distribution with $\mu = 50{,}000$ miles and $\sigma = 5500$ miles. We say X has an $N(50{,}000, 5500)$ distribution. From a manufacturer's perspective, it would be useful to know the probability that a tire fails to meet the guaranteed wear life of 40,000 miles. Figure 4.14 shows this probability as an area under a Normal density curve. You can find it by software or by standardizing and using Table A. From Table A,

$$\begin{aligned} P(X < 40{,}000) &= P\left(\frac{X-50{,}000}{5500} < \frac{40{,}000-50{,}000}{5500}\right) \\ &= P(Z < -1.82) \\ &= 0.0344 \end{aligned}$$

The manufacturer should expect to incur warranty costs for about 3.4% of its tires.

FIGURE 4.14 The Normal distribution with $\mu = 50{,}000$ and $\sigma = 5500$. The shaded area is $P(X < 40{,}000)$, calculated in Example 4.27.



APPLY YOUR KNOWLEDGE

4.93 Normal probabilities. Example 4.27 gives the Normal distribution $N(50{,}000, 5500)$ for the tread life X of a type of tire (in miles). Calculate the following probabilities:

(a) The probability that a tire lasts more than 50,000 miles.
(b) $P(X > 60{,}000)$.
(c) $P(X \geq 60{,}000)$.

We began this chapter with a general discussion of the idea of probability and the properties of probability models. Two very useful specific types of probability models are distributions of discrete and continuous random variables. In our study of statistics, we employ only these two types of probability models.

SECTION 4.4 Summary

- A **random variable** is a variable taking numerical values determined by the outcome of a random phenomenon. The **probability distribution** of a random variable X tells us what the possible values of X are and how probabilities are assigned to those values.
- A random variable X and its distribution can be **discrete** or **continuous.**
- A **discrete random variable** has possible values that can be given in an ordered list. The probability distribution assigns each of these values a probability between 0 and 1 such that the sum of all the probabilities is 1. The probability of any event is the sum of the probabilities of all the values that make up the event.
- A **continuous random variable** takes all values in some interval of numbers. A **density curve** describes the probability distribution of a continuous random variable. The probability of any event is the area under the curve and above the values that make up the event.
- **Normal distributions** are one type of continuous probability distribution.
- You can picture a probability distribution by drawing a **probability histogram** in the discrete case or by graphing the density curve in the continuous case.

SECTION 4.4 Exercises

For Exercises 4.90 and 4.91, see page 212; for 4.92, see page 214; and for 4.93, see page 216.

CASE 4.2 **4.94 Two day demand.** Refer to the distribution of daily demand for blood bags X in Case 4.2 (pages 210–211). Let Y be the total demand over two days. Assume that demand is independent from day to day.
(a) List the possible values for Y.
(b) From the distribution of daily demand, we find that the probability that no bags are demanded on a given day is 0.202. In that light, suppose a hospital manager states, "The chances that no bags are demanded over two consecutive days is 0.404." Provide a simple argument to the manager explaining the mistake in probability conclusion. (*Hint*: Use more than two days as the basis for your argument.)
(c) What is the probability that the total demand over two days is 0? In terms of the random variable, what is $P(Y = 0)$?

4.95 How many courses? At a small liberal arts college, students can register for one to six courses. In a typical fall semester, 5% take one course, 5% take two courses, 13% take three courses, 26% take four courses, 36% take five courses, and 15% take six courses. Let X be the number of courses taken in the fall by a randomly selected student from this college. Describe the probability distribution of this random variable.

4.96 Make a graphical display. Refer to the previous exercise. Use a probability histogram to provide a graphical description of the distribution of X.

4.97 Find some probabilities. Refer to Exercise 4.95.
(a) Find the probability that a randomly selected student takes three or fewer courses.
(b) Find the probability that a randomly selected student takes four or five courses.
(c) Find the probability that a randomly selected student takes eight courses.

4.98 Texas hold 'em. The game of Texas hold 'em starts with each player receiving two cards. Here is the probability distribution for the number of aces in two-card hands:

Number of aces	0	1	2
Probability	0.8507	0.1448	0.0045

(a) Verify that this assignment of probabilities satisfies the requirement that the sum of the probabilities for a discrete distribution must be 1.
(b) Make a probability histogram for this distribution.
(c) What is the probability that a hand contains at least one ace? Show two different ways to calculate this probability.

4.99 How large are households? Choose an American household at random, and let X be the number of persons living in the household. If we ignore the few households with more than seven inhabitants, the probability model for X is as follows:

Household size X	1	2	3	4	5	6	7
Probability	0.27	0.33	0.16	0.14	0.06	0.03	0.01

(a) Verify that this is a legitimate probability distribution.
(b) What is $P(X \geq 5)$?
(c) What is $P(X > 5)$?
(d) What is $P(2 < X \leq 4)$?
(e) What is $P(X \neq 1)$?
(f) Write the event that a randomly chosen household contains more than two persons in terms of X. What is the probability of this event?

CASE 4.2 **4.100 How much to order?** Faced with the demand for the perishable product in blood, hospital managers need to establish an ordering policy that deals with the trade-off between shortage and wastage. As it turns out, this scenario, referred to as a single-period inventory problem, is well known in the area of operations management, and there is an optimal policy. What we need to know is the per item cost of being short (C_S) and the per item cost of being in excess (C_E). In terms of the blood example, the hospital estimates that for every bag short, there is a cost of \$80 per bag, which includes expediting and emergency delivery costs. Any transfusion blood bags left in excess at day's end are associated with \$20 per bag cost, which includes the original cost of purchase along with end-of-day handling costs. With the objective of minimizing long-term average costs, the following critical ratio (CR) needs to be computed:

$$CR = \frac{C_S}{C_S + C_E}$$

Recognize that CR will always be in the range of 0 to 1. It turns out that the optimal number of items to order is the *smallest* value of k such that $P(X \leq k)$ is at *least* the CR value.
(a) Based on the given values of C_S and C_E, what is the value of CR?

(b) Given the *CR* found in part (a) and the distribution of blood bag demand (page 211), determine the optimal order quantity of blood bags per day.
(c) Keeping C_E at \$20, for what range of values of C_S does the hospital order three bags?

4.101 Discrete or continuous? In each of the following situations, decide whether the random variable is discrete or continuous, and give a reason for your answer.
(a) Your web page has five different links, and a user can click on one of the links or can leave the page. You record the length of time that a user spends on the web page before clicking one of the links or leaving the page.
(b) The number of hits on your web page.
(c) The yearly income of a visitor to your web page.

4.102 Use the uniform distribution. Suppose that a random variable X follows the uniform distribution described in Example 4.26 (pages 213–214). For each of the following events, find the probability and illustrate your calculations with a sketch of the density curve similar to the ones in Figure 4.12 (page 214).
(a) The probability that X is less than 0.1.
(b) The probability that X is greater than or equal to 0.8.
(c) The probability that X is less than 0.7 and greater than 0.5.
(d) The probability that X is 0.5.

4.103 Spell-checking software. Spell-checking software catches "nonword errors," which are strings of letters that are not words, as when "the" is typed as "eth." When undergraduates are asked to write a 250-word essay (without spell-checking), the number X of nonword errors has the following distribution:

Value of X	0	1	2	3	4
Probability	0.1	0.3	0.3	0.2	0.1

(a) Sketch the probability distribution for this random variable.
(b) Write the event "at least one nonword error" in terms of X. What is the probability of this event?
(c) Describe the event $X \leq 2$ in words. What is its probability? What is the probability that $X < 2$?

4.104 Find the probabilities. Let the random variable X be a random number with the uniform density curve in Figure 4.12 (page 214). Find the following probabilities:
(a) $P(X \geq 0.30)$.
(b) $P(X = 0.30)$.
(c) $P(0.30 < X < 1.30)$.
(d) $P(0.20 \leq X \leq 0.25 \text{ or } 0.7 \leq X \leq 0.9)$.
(e) X is not in the interval 0.4 to 0.7.

4.105 Uniform numbers between 0 and 2. Many random number generators allow users to specify the range of the random numbers to be produced. Suppose that you specify that the range is to be all numbers between 0 and 2. Call the random number generated Y. Then the density curve of the random variable Y has constant height between 0 and 2, and height 0 elsewhere.
(a) What is the height of the density curve between 0 and 2? Draw a graph of the density curve.
(b) Use your graph from part (a) and the fact that probability is area under the curve to find $P(Y \leq 1.6)$.
(c) Find $P(0.5 < Y < 1.7)$.
(d) Find $P(Y \geq 0.95)$.

4.106 The sum of two uniform random numbers. Generate *two* random numbers between 0 and 1 and take Y to be their sum. Then Y is a continuous random variable that can take any value between 0 and 2. The density curve of Y is the triangle shown in Figure 4.15.
(a) Verify by geometry that the area under this curve is 1.
(b) What is the probability that Y is less than 1? (Sketch the density curve, shade the area that represents the probability, then find that area. Do this for part (c) also.)
(c) What is the probability that Y is greater than 0.6?

4.107 How many close friends? How many close friends do you have? Suppose that the number of close friends adults claim to have varies from person to person with mean $\mu = 9$ and standard deviation $\sigma = 2.4$. An opinion poll asks this question of an SRS of 1100 adults. We see in Chapter 6 that, in this situation, the sample mean response $\bar{x}$ has approximately the Normal distribution with mean 9 and standard deviation 0.0724. What is $P(8 \leq \bar{x} \leq 10)$, the probability that the statistic $\bar{x}$ estimates μ to within ± 1?

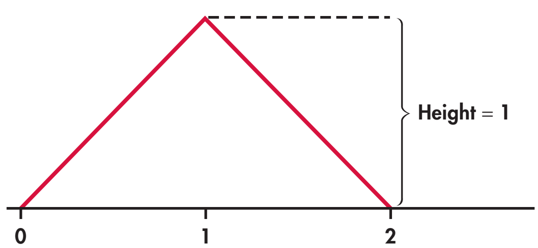


FIGURE 4.15 The density curve for the sum of two random numbers, Exercise 4.106. This density curve spreads probability between 0 and 2.

4.108 Normal approximation for a sample proportion. A sample survey contacted an SRS of 700 registered voters in Oregon shortly after an election and asked respondents whether they had voted. Voter records show that 56% of registered voters had actually voted. We see in the next chapter that in this situation the proportion of the sample $\hat{p}$ who voted has approximately the Normal distribution with mean $\mu = 0.56$ and standard deviation $\sigma = 0.019$.

(a) If the respondents answer truthfully, what is $P(0.52 \le \hat{p} \le 0.60)$? This is the probability that the sample proportion $\hat{p}$ estimates the mean of 0.56 within plus or minus 0.04.

(b) In fact, 72% of the respondents said they had voted ($\hat{p} = 0.72$). If respondents answer truthfully, what is $P(\hat{p} \ge 0.72)$? This probability is so small that it is good evidence that some people who did not vote claimed that they did vote.

4.5 Means and Variances of Random Variables

The probability histograms and density curves that picture the probability distributions of random variables resemble our earlier pictures of distributions of data. In describing data, we moved from graphs to numerical measures such as means and standard deviations. Now we make the same move to expand our descriptions of the distributions of random variables. We can speak of the mean winnings in a game of chance or the standard deviation of the randomly varying number of calls a travel agency receives in an hour. In this section, we learn more about how to compute these descriptive measures and about the laws they obey.

The mean of a random variable

In Chapter 1 (page 24), we learned that the mean $\bar{x}$ is the average of the observations in a *sample*. Recall that a random variable X is a numerical outcome of a random process. Think about repeating the random process many times and recording the resulting values of the random variable. In general, you can think of the mean of a random variable as the average of a very large sample. In the case of discrete random variables, the relative frequencies of the values in the very large sample are the same as their probabilities.

Here is an example for a discrete random variable.

EXAMPLE 4.28 The Tri-State Pick 3 Lottery

Most states and Canadian provinces have government-sponsored lotteries. Here is a simple lottery wager from the Tri-State Pick 3 game that New Hampshire shares with Maine and Vermont. You choose a three-digit number, 000 to 999. The state chooses a three-digit winning number at random and pays you $500 if your number is chosen.

Because there are 1000 three-digit numbers, you have probability 1/1000 of winning. Taking X to be the amount your ticket pays you, the probability distribution of X is

Payoff X	$0	$500
Probability	0.999	0.001

The random process consists of drawing a three-digit number. The population consists of the numbers 000 to 999. Each of these possible outcomes is equally likely in this example. In the setting of sampling in Chapter 3 (page 132), we can view the random process as selecting an SRS of size 1 from the population. The random variable X is 500 if the selected number is equal to the one that you chose and is 0 if it is not.

What is your average payoff from many tickets? The ordinary average of the two possible outcomes $0 and $500 is $250, but that makes no sense as the average because $500 is much less likely than $0. In the long run, you receive $500 once in every 1000 tickets and \$0 on the remaining 999 of 1000 tickets. The long-run average payoff is

$$\$500\frac{1}{1000} + \$0\frac{999}{1000} = \$0.50$$

or 50 cents. That number is the mean of the random variable X. (Tickets cost \$1, so in the long run, the state keeps half the money you wager.)

If you play Tri-State Pick 3 several times, we would, as usual, call the mean of the actual amounts you win $\bar{x}$. The mean in Example 4.28 is a different quantity—it is the long-run average winnings you expect if you play a very large number of times.

APPLY YOUR KNOWLEDGE

4.109 Find the mean of the probability distribution. You toss a fair coin. If the outcome is heads, you win \$5.00; if the outcome is tails, you win nothing. Let X be the amount that you win in a single toss of a coin. Find the probability distribution of this random variable and its mean.

Just as probabilities are an idealized description of long-run proportions, the mean of a probability distribution describes the long-run average outcome. We can't call this mean $\bar{x}$, so we need a different symbol. The common symbol for the **mean of a probability distribution** is μ, the Greek letter mu. We used μ in Chapter 1 for the mean of a Normal distribution, so this is not a new notation. We will often be interested in several random variables, each having a different probability distribution with a different mean.

mean μ

To remind ourselves that we are talking about the mean of X, we often write μ_X rather than simply μ. In Example 4.28, $\mu_X = \$0.50$. Notice that, as often happens, the mean is not a possible value of X. You will often find the mean of a random variable X called the **expected value** of X. *This term can be misleading because we don't necessarily expect an observation on X to equal its expected value.*

expected value

The mean of any discrete random variable is found just as in Example 4.28. It is not simply an average of the possible outcomes, but a weighted average in which each outcome is weighted by its probability. Because the probabilities add to 1, we have total weight 1 to distribute among the outcomes. An outcome that occurs half the time has probability one-half and gets one-half the weight in calculating the mean. Here is the general definition.

Mean of a Discrete Random Variable

Suppose that X is a **discrete random variable** whose distribution is

Value of X	x_1	x_2	x_3	$\cdots$
Probability	p_1	p_2	p_3	$\cdots$

To find the **mean** of X, multiply each possible value by its probability, then add all the products:

$$\mu_X = x_1p_1 + x_2p_2 + \cdots$$
$$= \sum x_ip_i$$

EXAMPLE 4.29 The Mean of Equally Likely First Digits

If first digits in a set of data all have the same probability, the probability distribution of the first digit X is then

First digit X	1	2	3	4	5	6	7	8	9
Probability	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9	1/9

The mean of this distribution is

$$\mu_X = 1 \times \frac{1}{9} + 2 \times \frac{1}{9} + 3 \times \frac{1}{9} + 4 \times \frac{1}{9} + 5 \times \frac{1}{9} + 6 \times \frac{1}{9} + 7 \times \frac{1}{9} + 8 \times \frac{1}{9} + 9 \times \frac{1}{9}$$

$$= 45 \times \frac{1}{9} = 5$$

Suppose that the random digits in Example 4.29 had a different probability distribution. In Case 4.1 (pages 184–185), we described Benford's law as a probability distribution that describes first digits of numbers in many real situations. Let's calculate the mean for Benford's law.

EXAMPLE 4.30 The Mean of First Digits That Follow Benford's Law

CASE 4.1 Here is the distribution of the first digit for data that follow Benford's law. We use the letter V for this random variable to distinguish it from the one that we studied in Example 4.29. The distribution of V is

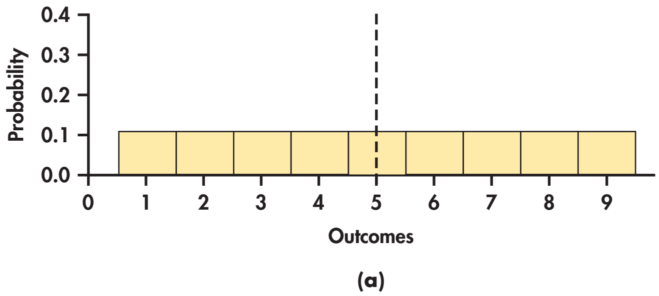
First digit V	1	2	3	4	5	6	7	8	9
Probability	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

The mean of V is

$$\mu_V = (1)(0.301) + (2)(0.176) + (3)(0.125) + (4)(0.097) + (5)(0.079) + (6)(0.067) + (7)(0.058) + (8)(0.051) + (9)(0.046) = 3.441$$

The mean reflects the greater probability of smaller first digits under Benford's law than when first digits 1 to 9 are equally likely.

Figure 4.16 locates the means of X and V on the two probability histograms. Because the discrete uniform distribution of Figure 4.16(a) is symmetric, the mean



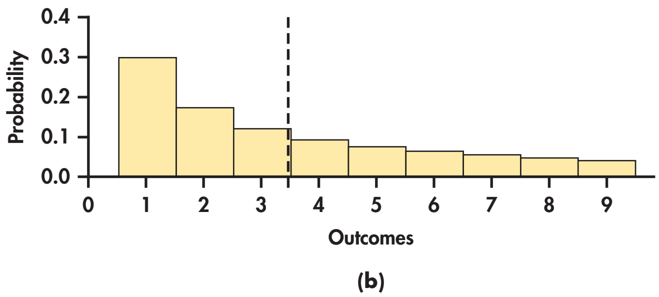


FIGURE 4.16 Locating the mean of a discrete random variable on the probability histogram: (a) digits between 1 and 9 chosen at random; and (b) digits between 1 and 9 chosen from records that obey Benford's law.

lies at the center of symmetry. We can't locate the mean of the right-skewed distribution of Figure 4.16(b) by eye—calculation is needed.

What about continuous random variables? The probability distribution of a continuous random variable X is described by a density curve. Chapter 1 showed how to find the mean of the distribution: it is the point at which the area under the density curve would balance if it were made out of solid material. The mean lies at the center of symmetric density curves such as the Normal curves. Exact calculation of the mean of a distribution with a skewed density curve requires advanced mathematics.[24] The idea that the mean is the balance point of the distribution applies to discrete random variables as well, but in the discrete case, we have a formula that gives us this point.

REMINDER
mean as balance point, p. 41

Mean and the law of large numbers

With probabilities in hand, we have shown that, for discrete random variables, the mean of the distribution (μ) can be determined by computing a weighted average in which each possible value of the random variable is weighted by its probability. For example, in Example 4.30, we found the mean of the first digit of numbers obeying Benford's law is 3.441.

Suppose, however, we are unaware of the probabilities of Benford's law but we still want to determine the mean of the distribution. To do so, we choose an SRS of financial statements and record the first digits of entries known to follow Benford's law. We then calculate the sample mean $\bar{x}$ to estimate the unknown population mean μ. In the vocabulary of statistics, μ is referred to as a *parameter* and $\bar{x}$ is called a *statistic*. These terms and their definitions are more formally described in Section 5.3 when we introduce the ideas of statistical inference.

It seems reasonable to use $\bar{x}$ to estimate μ. An SRS should fairly represent the population, so the mean $\bar{x}$ of the sample should be somewhere near the mean μ of the population. Of course, we don't expect $\bar{x}$ to be exactly equal to μ, and we realize that if we choose another SRS, the luck of the draw will probably produce a different $\bar{x}$. How can we control the variability of the sample means? The answer is to increase the sample size. If we keep on adding observations to our random sample, the statistic $\bar{x}$ is *guaranteed* to get as close as we wish to the parameter μ and then stay that close. We have the comfort of knowing that if we gather up more financial statements and keep recording more first digits, eventually we will estimate the mean value of the first digit very accurately. This remarkable fact is called the *law of large numbers*. It is remarkable because it holds for *any* population, not just for some special class such as Normal distributions.

Law of Large Numbers

Draw independent observations at random from any population with finite mean μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values becomes progressively closer to the population mean μ.

The behavior of $\bar{x}$ is similar to the idea of probability. In the long run, the *proportion* of outcomes taking any value gets close to the *probability* of that value, and the *average outcome* gets close to the distribution *mean*. Figure 4.1 (page 174) shows how proportions approach probability in one example. Here is an example of how sample means approach the distribution mean.

EXAMPLE 4.31 Applying the Law of Large Numbers

CASE 4.1 With a clipboard, we begin our sampling. The first randomly drawn financial statement entry has an 8 as its first digit. Thus, the initial sample mean is 8. We proceed to select a second financial statement entry, and find the first digit to be 3, so for $n = 2$ the mean is now

$$\bar{x} = \frac{8 + 3}{2} = 5.5$$

As this stage, we might be tempted to think that digits are equally likely because we have observed a large and a small digit. The flaw in this thinking is obvious. We are believing that short-run results accurately reflect long-run behavior. With clear mind, we proceed to collect more observations and continue to update the sample mean. Figure 4.17 shows that the sample mean changes as we increase the sample size. Notice that the first point is 8 and the second point is the previously calculated mean of 5.5. More importantly, notice that the mean of the observations gets close to the distribution mean $\mu = 3.441$ and settles down to that value. The law of large numbers says that this *always* happens.

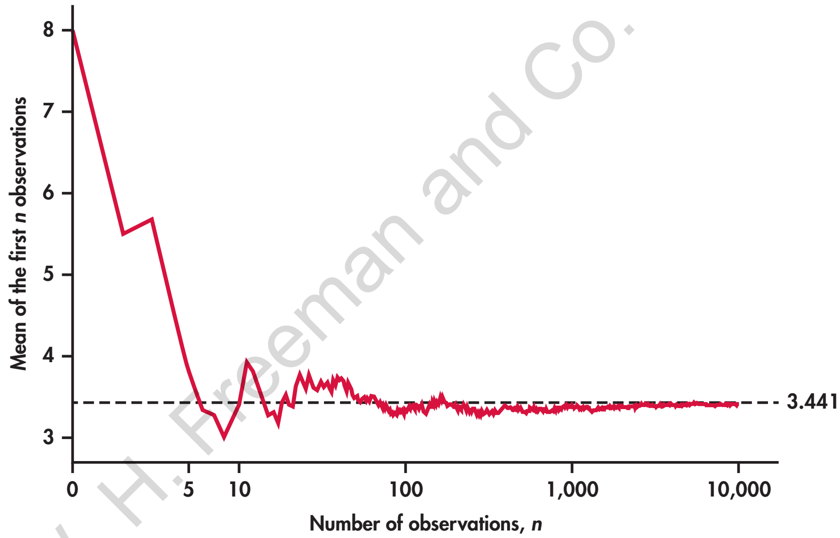


FIGURE 4.17 The law of large numbers in action. As we take more observations, the sample mean $\bar{x}$ always approaches the mean (μ) of the population.

APPLY YOUR KNOWLEDGE

4.110 Use the *Law of Large Numbers* applet. The *Law of Large Numbers* applet animates a graph like Figure 4.17 for rolling dice. Use it to better understand the law of large numbers by making a similar graph.

The mean μ of a random variable is the average value of the variable in two senses. By its definition, μ is the average of the possible values, weighted by their probability of occurring. The law of large numbers says that μ is also the long-run average of many independent observations on the variable. The law of large numbers can be proved mathematically starting from the basic laws of probability.

Thinking about the law of large numbers

The law of large numbers says broadly that the average results of many independent observations are stable and predictable. The gamblers in a casino may win or lose, but the casino will win in the long run because the law of large numbers says what the average outcome of many thousands of bets will be. An insurance company deciding how much to charge for life insurance and a fast-food restaurant deciding how many beef patties to prepare also rely on the fact that averaging over many individuals produces a stable result. It is worth the effort to think a bit more closely about so important a fact.

The “law of small numbers”

Both the rules of probability and the law of large numbers describe the regular behavior of chance phenomena *in the long run.* Psychologists have discovered that our intuitive understanding of randomness is quite different from the true laws of chance.[25] For example, most people believe in an incorrect “law of small numbers.” That is, we expect even short sequences of random events to show the kind of average behavior that, in fact, appears only in the long run.

Some teachers of statistics begin a course by asking students to toss a coin 50 times and bring the sequence of heads and tails to the next class. The teacher then announces which students just wrote down a random-looking sequence rather than actually tossing a coin. The faked tosses don’t have enough “runs” of consecutive heads or consecutive tails. Runs of the same outcome don’t look random to us but are, in fact, common. For example, the probability of a run of three or more consecutive heads or tails in just 10 tosses is greater than 0.8.[26] The runs of consecutive heads or consecutive tails that appear in real coin tossing (and that are predicted by the mathematics of probability) seem surprising to us. Because we don’t expect to see long runs, we may conclude that the coin tosses are not independent or that some influence is disturbing the random behavior of the coin.

EXAMPLE 4.32 The “Hot Hand” in Basketball

Belief in the law of small numbers influences behavior. If a basketball player makes several consecutive shots, both the fans and her teammates believe that she has a “hot hand” and is more likely to make the next shot. This is doubtful.

Careful study suggests that runs of baskets made or missed are no more frequent in basketball than would be expected if each shot were independent of the player’s previous shots. Baskets made or missed are just like heads and tails in tossing a coin. (Of course, some players make 30% of their shots in the long run and others make 50%, so a coin-toss model for basketball must allow coins with different probabilities of a head.) Our perception of hot or cold streaks simply shows that we don’t perceive random behavior very well.[27]

CAUTION

Our intuition doesn’t do a good job of distinguishing random behavior from systematic influences. This is also true when we look at data. We need statistical inference to supplement exploratory analysis of data because probability calculations can help verify that what we see in the data is more than a random pattern.

How large is a large number?

The law of large numbers says that the actual mean outcome of many trials gets close to the distribution mean μ as more trials are made. It doesn’t say how many trials are needed to guarantee a mean outcome close to μ. That depends on the *variability* of the random outcomes. The more variable the outcomes, the more trials are needed to ensure that the mean outcome $\bar{x}$ is close to the distribution mean μ. Casinos understand this: the outcomes of games of chance are variable enough to hold the interest of gamblers. Only the casino plays often enough to rely on the law of large numbers. Gamblers get entertainment; the casino has a business.

Rules for means

Imagine yourself as a financial adviser who must provide advice to clients regarding how to distribute their assets among different investments such as individual stocks, mutual funds, bonds, and real estate. With data available on all these financial instruments, you are able to gather a variety of insights, such as the proportion of the time a particular stock outperformed the market index, the average performance of the different investments, the consistency or inconsistency of performance of the different investments, and relationships among the investments. In other words, you are seeking measures of probability, mean, standard deviation, and correlation. In general, the discipline of finance relies heavily on a solid understanding of probability and statistics. In the next case, we explore how the concepts of this chapter play a fundamental role in constructing an investment portfolio.

CASE 4.3

Portfolio Analysis One of the fundamental measures of performance of an investment is its *rate of return*. For a stock, rate of return of an investment over a time period is basically the percent change in the share price during the time period. However, corporate actions such as dividend payments and stock splits can complicate the calculation. A stock's closing price can be amended to include any distributions and corporate actions to give us an adjusted closing price. The percent change of adjusted closing prices can then serve as a reasonable calculation of return.

For example, the closing adjusted price of the well-known S&P 500 market index was \$1,923.57 for April 2014 and was \$1,960.96 for May 2014. So, the index's *monthly* rate of return for that time period was

$$\frac{\text{change in price}}{\text{starting price}} = \frac{1{,}960.96 - 1{,}923.57}{1{,}923.57} = 0.0194, \text{ or } 1.94\%$$

Investors want high positive returns, but they also want safety. Since 2000 to mid-2014, the S&P 500's monthly returns have swung to as low as −17% and to as high as +11%. The variability of returns, called *volatility* in finance, is a measure of the risk of an investment. A highly volatile stock, which may often go either up or down, is more risky than a Treasury bill, whose return is very predictable.

A *portfolio* is a collection of investments held by an individual or an institution. *Portfolio analysis* begins by studying how the risk and return of a portfolio are determined by the risk and return of the individual investments it contains. That's where statistics comes in: the return on an investment over some period of time is a random variable. We are interested in the *mean* return, and we measure volatility by the *standard deviation* of returns. Indeed, investment firms will report online the historical mean and standard deviation of returns of individual stocks or funds.[28]

Suppose that we are interested in building a simple portfolio based on allocating funds into one of two investments. Let's take one of the investments to be the commonly chosen S&P 500 index. The key now is to pick another investment that does *not* have a high positive correlation with the market index. Investing in two investments that have very high positive correlation with each other is tantamount to investing in just one.

Possible choices against the S&P 500 index are different asset classes like real estate, gold, energy, and utilities. For example, suppose we build a portfolio with 70% of funds invested in the S&P 500 index and 30% in a well-known utilities sector fund (XLU). If X is the monthly return on the S&P 500 index and Y the monthly return on the utilities fund, the portfolio rate of return is

$$R = 0.7X + 0.3Y$$

How can we find the mean and standard deviation of the portfolio return R starting from information about X and Y? We must now develop the machinery to do this.

Think first not about investments but about making refrigerators. You are studying flaws in the painted finish of refrigerators made by your firm. Dimples and paint sags are two kinds of surface flaw. Not all refrigerators have the same number of dimples: many have none, some have one, some two, and so on. You ask for the average number of imperfections on a refrigerator. The inspectors report finding an average of 0.7 dimple and 1.4 sags per refrigerator. How many total imperfections of both kinds (on the average) are there on a refrigerator? That's easy: if the average number of dimples is 0.7 and the average number of sags is 1.4, then counting both gives an average of $0.7 + 1.4 = 2.1$ flaws.

In more formal language, the number of dimples on a refrigerator is a random variable X that varies as we inspect one refrigerator after another. We know only that the mean number of dimples is $\mu_X = 0.7$. The number of paint sags is a second random variable Y having mean $\mu_Y = 1.4$. (As usual, the subscripts keep straight which variable we are talking about.) The total number of both dimples and sags is another random variable, the sum $X + Y$. Its mean μ_{X+Y} is the average number of dimples and sags together. It is just the sum of the individual means μ_X and μ_Y. That's an important rule for how means of random variables behave.

Here's another rule. A large lot of plastic coffee-can lids has a mean diameter of 4.2 inches. What is the mean in centimeters? There are 2.54 centimeters in an inch, so the diameter in centimeters of any lid is 2.54 times its diameter in inches. If we multiply every observation by 2.54, we also multiply their average by 2.54. The mean in centimeters must be 2.54×4.2, or about 10.7 centimeters. More formally, the diameter in inches of a lid chosen at random from the lot is a random variable X with mean μ_X. The diameter in centimeters is $2.54X$, and this new random variable has mean $2.54\mu_X$.

The point of these examples is that means behave like averages. Here are the rules we need.

Rules for Means

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$

Rule 2. If X and Y are random variables, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$

Rule 3. If X and Y are random variables, then

$$\mu_{X-Y} = \mu_X - \mu_Y$$

EXAMPLE 4.33 Aggregating Demand in a Supply Chain

To remain competitive, companies worldwide are increasingly recognizing the need to effectively manage their supply chains. Let us consider a simple but realistic supply chain scenario. ElectroWorks is a company that manufactures and distributes electronic parts to various regions in the United States. To serve the Chicago–Milwaukee region, the company has a warehouse in Milwaukee and another in Chicago. Because the company produces thousands of parts, it is considering an alternative strategy of locating a single, centralized warehouse between the two markets—say, in Kenosha, Wisconsin—that will serve all customer orders. Delivery time, referred to as *lead time,* from manufacturing to warehouse(s) and ultimately to customers is unaffected by the new strategy.

To illustrate the implications of the centralized warehouse, let us focus on one specific part: SurgeArrester. The lead time for this part from manufacturing to warehouses is one week. Based on historical data, the lead time demands for the part in each of the markets are Normally distributed with

$$X = \text{Milwaukee warehouse} \quad \mu_X = 415 \text{ units} \quad \sigma_X = 48 \text{ units}$$
$$Y = \text{Chicago warehouse} \quad \mu_Y = 2689 \text{ units} \quad \sigma_Y = 272 \text{ units}$$

If the company were to centralize, what would be the mean of the total aggregated lead time demand $X + Y$? Using Rule 2, we can easily find the mean overall lead time demand is

$$\mu_{X+Y} = \mu_X + \mu_Y = 415 + 2689 = 3104$$

At this stage, we only have part of the picture on the aggregated demand random variable—namely, its mean value. In Example 4.39 (pages 232–233), we continue our study of aggregated demand to include the variability dimension that, in turn, will reveal operational benefits from the proposed strategy of centralizing. Let's now consider the portfolio scenario of Case 4.3 (page 225) to demonstrate the use of a combination of the mean rules.

EXAMPLE 4.34 Portfolio Analysis

CASE 4.3 The past behavior of the two securities in the portfolio is pictured in Figure 4.18, which plots the monthly returns for S&P 500 market index against the utility sector index from January 2000 to May 2014. We can see that the returns on the two indices have a moderate level of positive correlation. This fact will be used later for gaining a complete assessment of the expected performance of the portfolio. For now, we can calculate mean returns from the 173 data points shown on the plot:[29]

$$X = \text{monthly return for S\&P 500 index} \quad \mu_X = 0.298\%$$
$$Y = \text{monthly return for Utility index} \quad \mu_Y = 0.675\%$$

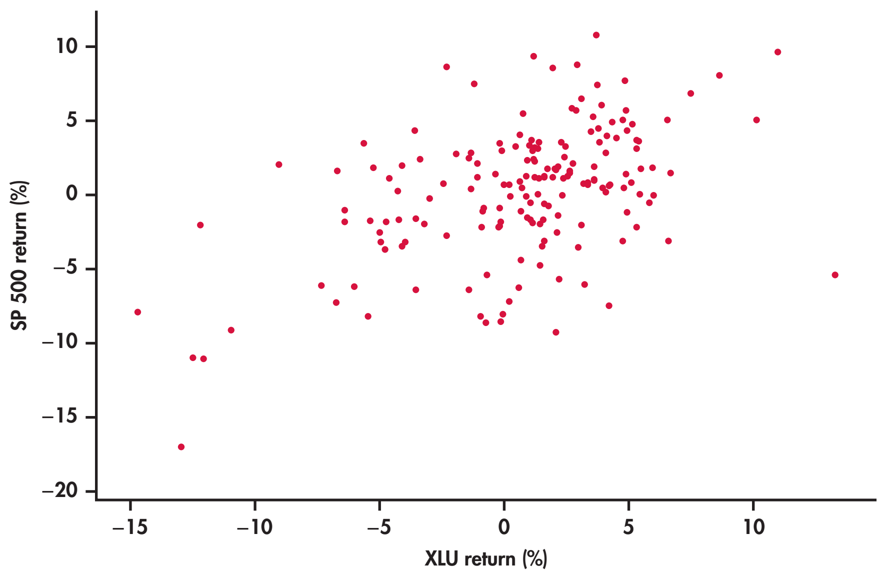


FIGURE 4.18 Monthly returns on S&P 500 index versus returns on Utilities Sector index (January 2000 to May 2014), Example 4.34.

By combining Rules 1 and 2, we can find the mean return on the portfolio based on a 70/30 mix of S&P index shares and utility shares:

$$R = 0.7X + 0.3Y$$
$$\mu_R = 0.7\mu_X + 0.3\mu_Y$$
$$= (0.7)(0.298) + (0.3)(0.675) = 0.411\%$$

This calculation uses historical data on returns. Next month may, of course, be very different. It is usual in finance to use the term *expected return* in place of mean return.

APPLY YOUR KNOWLEDGE

4.111 Find μ_Y. The random variable X has mean $\mu_X = 8$. If $Y = 12 + 7X$, what is μ_Y?

4.112 Find μ_W. The random variable U has mean $\mu_U = 22$, and the random variable V has mean $\mu_V = 22$. If $W = 0.5U + 0.5V$, find μ_W.

4.113 Managing a new-product development process. Managers often have to oversee a series of related activities directed to a desired goal or output. As a new-product development manager, you are responsible for two sequential steps of the product development process—namely, the development of product specifications followed by the design of the manufacturing process. Let X be the number of weeks required to complete the development of product specifications, and let Y be the number of weeks required to complete the design of the manufacturing process. Based on experience, you estimate the following probability distribution for the first step:

Weeks (X)	1	2	3
Probability	0.3	0.5	0.2

For the second step, your estimated distribution is

Weeks (Y)	1	2	3	4	5
Probability	0.1	0.15	0.4	0.30	0.05

(a) Calculate μ_X and μ_Y.
(b) The cost per week for the activity of developing product specifications is $8000, while the cost per week for the activity of designing the manufacturing process is $30,000. Calculate the mean cost for each step.
(c) Calculate the mean completion time and mean cost for the two steps combined.

CASE 4.3 **4.114 Mean return on portfolio.** The addition rule for means extends to sums of any number of random variables. Let's look at a portfolio containing three mutual funds from three different industrial sectors: biotechnology, information services, and defense. The monthly returns on Fidelity Select Biotechnology Fund (FBIOX), Fidelity National Information Services Fund (FIX), and Fidelity Select Defense and Aerospace Fund (FSDAX) for the 60 months ending in July 2014 had approximately these means:[30]

X = Biotechnology monthly return $\mu_X = 2.282\%$
Y = Information services monthly return $\mu_Y = 1.669\%$
Z = Defense and aerospace monthly return $\mu_Z = 1.653\%$

What is the mean monthly return for a portfolio consisting of 50% biotechnology, 30% information services, and 20% defense and aerospace?

The variance of a random variable

The mean is a measure of the center of a distribution. Another important characteristic of a distribution is its spread. The variance and the standard deviation are the standard measures of spread that accompany the choice of the mean to measure center. Just as for the mean, we need a distinct symbol to distinguish the variance of a random variable from the variance s^2 of a data set. We write the variance of a random variable X as σ_X^2. Once again, the subscript reminds us which variable we have in mind. The definition of the variance σ_X^2 of a random variable is similar to the definition of the sample variance s^2 given in Chapter 1. That is, the variance is an average value of the squared deviation $(X - \mu_X)^2$ of the variable X from its mean μ_X.

As for the mean of a discrete random variable, we use a weighted average of these squared deviations based on the probability of each outcome. Calculating this weighted average is straightforward for discrete random variables but requires advanced mathematics in the continuous case. Here is the definition.

Variance of a Discrete Random Variable

Suppose that X is a **discrete random variable** whose distribution is

Value of X	x_1	x_2	x_3	$\cdots$
Probability	p_1	p_2	p_3	$\cdots$

and that μ_X is the mean of X. The **variance** of X is

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots$$
$$= \sum (x_i - \mu_X)^2 p_i$$

The **standard deviation** σ_X of X is the square root of the variance.

EXAMPLE 4.35 Find the Mean and the Variance

CASE 4.2 In Case 4.2 (pages 210–211), we saw that the distribution of the daily demand X of transfusion blood bags is

Bags used	0	1	2	3	4	5	6
Probability	0.202	0.159	0.201	0.125	0.088	0.087	0.056

Bags used	7	8	9	10	11	12
Probability	0.025	0.022	0.018	0.008	0.006	0.003

We can find the mean and variance of X by arranging the calculation in the form of a table. Both μ_X and σ_X^2 are sums of columns in this table.

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$		
0	0.202	0.00	$(0 - 2.754)^2(0.202)$	=	1.53207
1	0.159	0.159	$(1 - 2.754)^2(0.159)$	=	0.48917
2	0.201	0.402	$(2 - 2.754)^2(0.201)$	=	0.11427
3	0.125	0.375	$(3 - 2.754)^2(0.125)$	=	0.00756
4	0.088	0.352	$(4 - 2.754)^2(0.088)$	=	0.13662
5	0.087	0.435	$(5 - 2.754)^2(0.087)$	=	0.43887
6	0.056	0.336	$(6 - 2.754)^2(0.056)$	=	0.59004
7	0.025	0.175	$(7 - 2.754)^2(0.025)$	=	0.45071

(*Continued*)

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$		
8	0.022	0.176	$(8 - 2.754)^2(0.022)$	=	0.60545
9	0.018	0.162	$(9 - 2.754)^2(0.018)$	=	0.70223
10	0.008	0.080	$(10 - 2.754)^2(0.008)$	=	0.42004
11	0.006	0.066	$(11 - 2.754)^2(0.006)$	=	0.40798
12	0.003	0.036	$(12 - 2.754)^2(0.003)$	=	0.25647
		$\mu_X = 2.754$	σ_X^2	=	6.151

REMINDER
68–95–99.7 rule, p. 43

We see that $\sigma_X^2 = 6.151$. The standard deviation of X is $\sigma_X = \sqrt{6.151} = 2.48$. The standard deviation is a measure of the variability of the daily demand of blood bags. As in the case of distributions for data, the connection of standard deviation to probability is easiest to understand for Normal distributions (for example, 68–95–99.7 rule). For general distributions, we are content to understand that the standard deviation provides us with a basic measure of variability.

APPLY YOUR KNOWLEDGE

4.115 Managing new-product development process. Exercise 4.113 (page 228) gives the distribution of time to complete two steps in the new-product development process.

(a) Calculate the variance and the standard deviation of the number of weeks to complete the development of product specifications.

(b) Calculate σ_Y^2 and σ_Y for the design of the manufacturing-process step.

Rules for variances and standard deviations

What are the facts for variances that parallel Rules 1, 2, and 3 for means? *The mean of a sum of random variables is always the sum of their means, but this addition rule is true for variances only in special situations.* To understand why, take X to be the percent of a family's after-tax income that is spent, and take Y to be the percent that is saved. When X increases, Y decreases by the same amount. Though X and Y may vary widely from year to year, their sum $X + Y$ is always 100% and does not vary at all. It is the association between the variables X and Y that prevents their variances from adding.

If random variables are independent, this kind of association between their values is ruled out and their variances do add. As defined earlier for general events A and B (page 205), two random variables X and Y are **independent** if knowing that any event involving X alone did or did not occur tells us nothing about the occurrence of any event involving Y alone.

Probability models often assume independence when the random variable outcomes appear unrelated to each other. *You should ask in each instance whether the assumption of independence seems reasonable.*

correlation

When random variables are not independent, the variance of their sum depends on the **correlation** between them as well as on their individual variances. In Chapter 2, we met the correlation r between two observed variables measured on the same individuals. We defined the correlation r (page 75) as an average of the products of the standardized x and y observations. The correlation between two random variables is defined in the same way, once again using a weighted average with probabilities as weights in the case of discrete random variables. We won't give the details—it is enough to know that the correlation between two random variables has the same basic properties as the correlation r calculated from data. We use ρ, the Greek letter rho, for the correlation between two random variables. The correlation ρ is a number between -1 and 1 that measures the direction and strength of the linear relationship between two variables. **The correlation between two independent random variables is zero.**

Returning to family finances, if X is the percent of a family's after-tax income that is spent and Y is the percent that is saved, then $Y = 100 - X$. This is a perfect linear relationship with a negative slope, so the correlation between X and Y is $\rho = -1$. With the correlation at hand, we can state the rules for manipulating variances.

Rules for Variances and Standard Deviations of Linear Transformations, Sums, and Differences

Rule 1. If X is a random variable and a and b are fixed numbers, then

$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

Rule 2. If X and Y are independent random variables, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

This is the **addition rule for variances of independent random variables.**

Rule 3. If X and Y have correlation ρ, then

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$

This is the **general addition rule for variances of random variables.**

To find the standard deviation, take the square root of the variance.

CAUTION

Because a variance is the average of squared deviations from the mean, multiplying X by a constant b multiplies σ^2_X by the square of the constant. Adding a constant a to a random variable changes its mean but does not change its variability. The variance of $X + a$ is, therefore, the same as the variance of X. Because the square of -1 is 1, the addition rule says that the variance of a difference between independent random variables is the *sum* of the variances. For independent random variables, the difference $X - Y$ is more variable than either X or Y alone because variations in both X and Y contribute to variation in their difference.

As with data, we prefer the standard deviation to the variance as a measure of the variability of a random variable. Rule 2 for variances implies that standard deviations of independent random variables do not add. To work with standard deviations, use the rules for variances rather than trying to remember separate rules for standard deviations. For example, the standard deviations of $2X$ and $-2X$ are both equal to $2\sigma_X$ because this is the square root of the variance $4\sigma^2_X$.

EXAMPLE 4.36 Payoff in the Tri-State Pick 3 Lottery

The payoff X of a \$1 ticket in the Tri-State Pick 3 game is \$500 with probability 1/1000 and 0 the rest of the time. Here is the combined calculation of mean and variance:

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$		
0	0.999	0	$(0 - 0.5)^2(0.999)$	=	0.24975
500	0.001	0.5	$(500 - 0.5)^2(0.001)$	=	249.50025
	$\mu_X = 0.5$		σ^2_X	=	249.75

The mean payoff is 50 cents. The standard deviation is $\sigma_X = \sqrt{249.75} = \$15.80$. It is usual for games of chance to have large standard deviations because large variability makes gambling exciting.

If you buy a Pick 3 ticket, your winnings are $W = X - 1$ because the dollar you paid for the ticket must be subtracted from the payoff. Let's find the mean and variance for this random variable.

EXAMPLE 4.37 Winnings in the Tri-State Pick 3 Lottery

By the rules for means, the mean amount you win is

$$\mu_W = \mu_X - 1 = -\$0.50$$

That is, you lose an average of 50 cents on a ticket. The rules for variances remind us that the variance and standard deviation of the winnings $W = X - 1$ are the same as those of X. Subtracting a fixed number changes the mean but not the variance.

Suppose now that you buy a \$1 ticket on each of two different days. The payoffs X and Y on the two tickets are independent because separate drawings are held each day. Your total payoff is $X + Y$. Let's find the mean and standard deviation for this payoff.

EXAMPLE 4.38 Two Tickets

The mean for the payoff for the two tickets is

$$\mu_{X+Y} = \mu_X + \mu_Y = \$0.50 + \$0.50 = \$1.00$$

Because X and Y are independent, the variance of $X + Y$ is

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y = 249.75 + 249.75 = 499.5$$

The standard deviation of the total payoff is

$$\sigma_{X+Y} = \sqrt{499.5} = \$22.35$$

CAUTION This is not the same as the sum of the individual standard deviations, which is \$15.80 + \$15.80 = \$31.60. *Variances of independent random variables add; standard deviations generally do not.*

When we add random variables that are correlated, we need to use the correlation for the calculation of the variance, but not for the calculation of the mean. Here are two examples.

EXAMPLE 4.39 Aggregating Demand in a Supply Chain

In Example 4.33, we learned that the lead time demands for SurgeArresters in two markets are Normally distributed with

X = Milwaukee warehouse	$\mu_X = 415$ units	$\sigma_X = 48$ units
Y = Chicago warehouse	$\mu_Y = 2689$ units	$\sigma_Y = 272$ units

Based on the given means, we found that the mean aggregated demand μ_{X+Y} is 3104. The variance and standard deviation of the aggregated *cannot be computed* from the information given so far. Not surprisingly, demands in the two markets are not independent because of the proximity of the regions. Therefore, Rule 2 for variances does not apply. We need to know ρ, the correlation between X and Y, to apply Rule 3. Historically, the correlation between Milwaukee demand and Chicago demand is about $\rho = 0.52$. To find the variance of the overall demand, we use Rule 3:

$$\begin{aligned}\sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y \\ &= (48)^2 + (272)^2 + (2)(0.52)(48)(272) \\ &= 89{,}866.24\end{aligned}$$

The variance of the sum $X + Y$ is greater than the sum of the variances $\sigma^2_X + \sigma^2_Y$ because of the positive correlation between the two markets. We find the standard deviation from the variance,

$$\sigma_{X+Y} = \sqrt{89{,}866.24} = 299.78$$

Notice that even though the variance of the sum is greater than the sum of the variances, the standard deviation of the sum is less than the sum of the standard deviations. Here lies the potential benefit of a centralized warehouse. To protect against stockouts, ElectroWorks maintains safety stock for a given product at each warehouse. Safety stock is extra stock in hand over and above the mean demand. For example, if ElectroWorks has a policy of holding two standard deviations of safety stock, then the amount of safety stock (rounded to the nearest integer) at warehouses would be

Location	Safety Stock
Milwaukee warehouse	2(48) = 96 units
Chicago warehouse	2(272) = 544 units
Centralized warehouse	2(299.78) = 600 units

The combined safety stock for the Milwaukee and Chicago warehouses is 640 units, which is 40 more units required than if distribution was operated out of a centralized warehouse. Now imagine the implication for safety stock when you take into consideration not just one part but *thousands* of parts that need to be stored.

risk pooling

This example illustrates the important supply chain concept known as **risk pooling**. Many companies such as Walmart and e-commerce retailer Amazon take advantage of the benefits of risk pooling as illustrated by this example.

EXAMPLE 4.40 Portfolio Analysis

CASE 4.3 Now we can complete our initial analysis of the portfolio constructed on a 70/30 mix of S&P 500 index shares and utility sector shares. Based on monthly returns between 2000 and 2014, we have

X = monthly return for S&P 500 index	$\mu_X = 0.298\%$	$\sigma_X = 4.453\%$
Y = monthly return for Utility index	$\mu_Y = 0.675\%$	$\sigma_Y = 4.403\%$
Correlation between X and Y:	$\rho = 0.495$	

In Example 4.34 (pages 227–228), we found that the mean return R is 0.411%. To find the variance of the portfolio return, combine Rules 1 and 3:

$$\begin{aligned}\sigma^2_R &= \sigma^2_{0.7X} + \sigma^2_{0.3Y} + 2\rho\sigma_{0.7X}\sigma_{0.3Y} \\ &= (0.7)^2\sigma^2_X + (0.3)^2\sigma^2_Y + 2\rho(0.7 \times \sigma_X)\,(0.3 \times \sigma_X) \\ &= (0.7)^2(4.453)^2 + (0.3)^2(4.403)^2 + (2)(0.495)(0.7 \times 4.453)(0.3 \times 4.403) \\ &= 15.54\end{aligned}$$

$$\sigma_R = \sqrt{15.54} = 3.942\%$$

We see that portfolio has a smaller mean return than investing all in the utility index. However, what is gained is that the portfolio has less variability (or volatility) than investing all in one or the other index.

Example 4.40 illustrates the first step in modern finance, using the mean and standard deviation to describe the behavior of a portfolio. We illustrated a particular mix (70/30), but what is needed is an exploration of different combinations to seek the best construction of the portfolio.

EXAMPLE 4.41 Portfolio Analysis

CASE 4.3 By doing the mean computations of Example 4.34 (pages 227–228) and the standard deviation computations of Example 4.40 for different mixes, we find the following values.

S&P 500 proportion	μ_R	σ_R
0.0	0.675	4.403
0.1	0.637	4.201
0.2	0.600	4.038
0.3	0.562	3.919
0.4	0.524	3.848
0.5	0.487	3.828
0.6	0.449	3.860
0.7	0.411	3.942
0.8	0.373	4.071
0.9	0.336	4.243
1.0	0.298	4.453

minimum variance portfolio

From Figure 4.19, we see that the plot of the portfolio mean returns against the corresponding standard deviations forms a parabola. The point on the parabola where the portfolio standard deviation is lowest is the **minimum variance portfolio** (MVP). From the preceding table, we see that the MVP is somewhere near a 50/50 allocation between the two investments. The solid curve of the parabola provides the preferable options in that the expected return is, for a given level of risk, higher than the dashed line option.

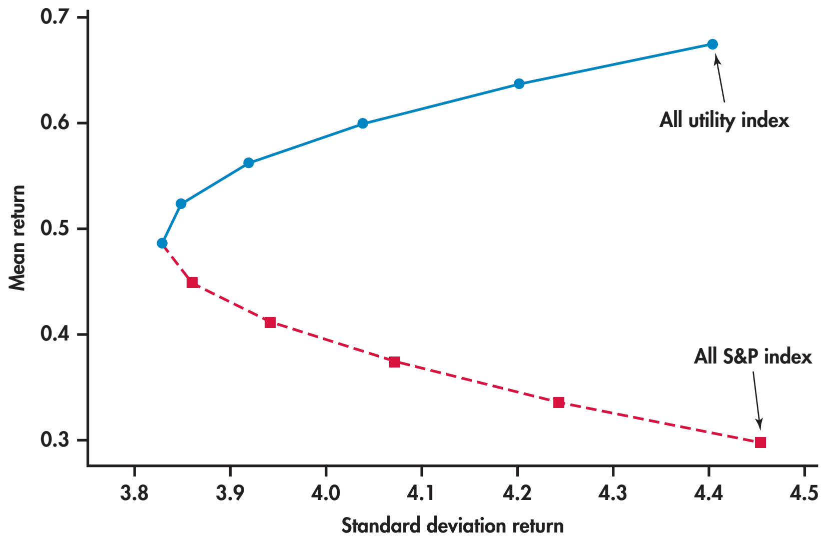


FIGURE 4.19 Mean return of portfolio versus standard deviation of portfolio, Example 4.41.

APPLY YOUR KNOWLEDGE

4.116 Comparing sales. Tamara and Derek are sales associates in a large electronics and appliance store. Their store tracks each associate's daily sales in dollars. Tamara's sales total X varies from day to day with mean and standard deviation

$$\mu_X = \$1100 \text{ and } \sigma_X = \$100$$

Derek's sales total Y also varies, with

$$\mu_Y = \$1000 \text{ and } \sigma_Y = \$80$$

Because the store is large and Tamara and Derek work in different departments, we might assume that their daily sales totals vary independently of each other. What are the mean and standard deviation of the difference $X - Y$ between Tamara's daily sales and Derek's daily sales? Tamara sells more on the average. Do you think she sells more every day? Why?

4.117 Comparing sales. It is unlikely that the daily sales of Tamara and Derek in the previous problem are uncorrelated. They will both sell more during the weekends, for example. Suppose that the correlation between their sales is $p = 0.4$. Now what are the mean and standard deviation of the difference $X - Y$? Can you explain conceptually why positive correlation between two variables reduces the variability of the difference between them?

4.118 Managing new-product development process. Exercise 4.113 (page 228) gives the distributions of X, the number of weeks to complete the development of product specifications, and Y, the number of weeks to complete the design of the manufacturing process. You did some useful variance calculations in Exercise 4.115 (page 230). The cost per week for developing product specifications is \$8000, while the cost per week for designing the manufacturing process is \$30,000.

(a) Calculate the standard deviation of the cost for each of the two activities using Rule 1 for variances (page 231).
(b) Assuming the activity times are independent, calculate the standard deviation for the total cost of both activities combined.
(c) Assuming $\rho = 0.8$, calculate the standard deviation for the total cost of both activities combined.
(d) Assuming $\rho = 0$, calculate the standard deviation for the total cost of both activities combined. How does this compare with your result in part (b)? In part (c)?
(e) Assuming $\rho = -0.8$, calculate the standard deviation for the total cost of both activities combined. How does this compare with your result in part (b)? In part (c)? In part (d)?

SECTION 4.5 Summary

- The probability distribution of a random variable X, like a distribution of data, has a **mean** $\boldsymbol{\mu_X}$ and a **standard deviation** $\boldsymbol{\sigma_X}$.
- The **law of large numbers** says that the average of the values of X observed in many trials must approach μ.
- The **mean** $\boldsymbol{\mu}$ is the balance point of the probability histogram or density curve. If X is **discrete** with possible values x_i having probabilities p_i, the mean is the average of the values of X, each weighted by its probability:

$$\mu_X = x_1p_1 + x_2p_2 + \cdots$$

- The **variance** $\boldsymbol{\sigma_X^2}$ is the average squared deviation of the values of the variable from their mean. For a discrete random variable,

$$\sigma_X^2 = (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots$$

- The **standard deviation** $\boldsymbol{\sigma_X}$ is the square root of the variance. The standard deviation measures the variability of the distribution about the mean. It is easiest to interpret for Normal distributions.
- The **mean and variance of a continuous random variable** can be computed from the density curve, but to do so requires more advanced mathematics.
- The means and variances of random variables obey the following rules. If a and b are fixed numbers, then

$$\mu_{a+bX} = a + b\mu_X$$
$$\sigma^2_{a+bX} = b^2\sigma^2_X$$

If X and Y are any two random variables having correlation ρ, then

$$\mu_{X+Y} = \mu_X + \mu_Y$$
$$\mu_{X-Y} = \mu_X - \mu_Y$$
$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y - 2\rho\sigma_X\sigma_Y$$

If X and Y are **independent,** then $\rho = 0$. In this case,

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

SECTION 4.5 Exercises

For Exercise 4.109, see page 220; for 4.110, see page 223; for 4.111 to 4.114, see page 228; for 4.115, see page 230; and for 4.116 to 4.118, see page 235.

CASE 4.3 **4.119 Portfolio analysis.** Show that if 20% of the portfolio is based on the S&P 500 index, then the mean and standard deviation of the portfolio are indeed the values given in Example 4.41 (page 234).

4.120 Find some means. Suppose that X is a random variable with mean 20 and standard deviation 5. Also suppose that Y is a random variable with mean 40 and standard deviation 10. Find the mean of the random variable Z for each of the following cases. Be sure to show your work.
(a) $Z = 2 + 10X$.
(b) $Z = 10X - 2$.
(c) $Z = X + Y$.
(d) $Z = X - Y$.
(e) $Z = -3X - 2Y$.

4.121 Find the variance and the standard deviation. A random variable X has the following distribution.

X	−1	0	1	2
Probability	0.3	0.2	0.2	0.3

Find the variance and the standard deviation for this random variable. Show your work.

4.122 Find some variances and standard deviations. Suppose that X is a random variable with mean 20 and standard deviation 5. Also suppose that Y is a random variable with mean 40 and standard deviation 10. Assume that X and Y are independent. Find the variance and the standard deviation of the random variable Z for each of the following cases. Be sure to show your work.
(a) $Z = 2 + 10X$.
(b) $Z = 10X - 2$.
(c) $Z = X + Y$.
(d) $Z = X - Y$.
(e) $Z = -3X - 2Y$.

4.123 What happens if the correlation is not zero? Suppose that X is a random variable with mean 20 and standard deviation 5. Also suppose that Y is a

random variable with mean 40 and standard deviation 10. Assume that the correlation between X and Y is 0.5. Find the variance and standard deviation of the random variable Z for each of the following cases. Be sure to show your work.
(a) $Z = X + Y$.
(b) $Z = X - Y$.
(c) $Z = -3X - 2Y$.

4.124 What's wrong? In each of the following scenarios, there is something wrong. Describe what is wrong, and give a reason for your answer.
(a) If you toss a fair coin three times and get heads all three times, then the probability of getting a tail on the next toss is much greater than one-half.
(b) If you multiply a random variable by 10, then the mean is multiplied by 10 and the variance is multiplied by 10.
(c) When finding the mean of the sum of two random variables, you need to know the correlation between them.

4.125 Difference between heads and tails. Suppose a fair coin is tossed three times.
(a) Using the labels of "H" and "T," list all the possible outcomes in the sample space.
(b) For each outcome in the sample space, define the random variable D as the number of heads minus the number of tails observed. Use the fact that all outcomes of part (a) are equally likely to find the probability distribution of D.
(c) Use the probability distribution found in (b) to find the mean and standard deviation of D.

4.126 Mean of the distribution for the number of aces. In Exercise 4.98 (page 217), you examined the probability distribution for the number of aces when you are dealt two cards in the game of Texas hold 'em. Let X represent the number of aces in a randomly selected deal of two cards in this game. Here is the probability distribution for the random variable X:

Value of X	0	1	2
Probability	0.8507	0.1448	0.0045

Find μ_X, the mean of the probability distribution of X.

4.127 Standard deviation of the number of aces. Refer to the previous exercise. Find the standard deviation of the number of aces.

4.128 Difference between heads and tails. In Exercise 4.125, the mean and standard deviation were computed directly from the probability distribution of random variable D. Instead, define X as the number of heads in the three flips, and define Y as the number of tails in the three flips.
(a) Find the probability distribution for X along with the mean μ_X and standard deviation σ_X.
(b) Find the probability distribution for Y along with the mean μ_Y and standard deviation σ_Y.
(c) Explain why the correlation ρ between X and Y is -1.
(d) Define D as $X - Y$. Use the rules of means and variances along with $\rho = 1$ to find the mean and standard deviation of D. Confirm the values are the same as found in Exercise 4.125.

4.129 Pick 3 and law of large numbers. In Example 4.28 (pages 219–220), the mean payoff for the Tri-State Pick 3 lottery was found to be \$0.50. In our discussion of the law of large numbers, we learned that the mean of a probability distribution describes the long-run average outcome. In this exercise, you will explore this concept using technology.

- *Excel users:* Input the values "0" and "500" in the first two rows of column A. Now input the corresponding probabilities of 0.999 and 0.001 in the first two rows of column B. Now choose "Random Number Generation" from the **Data Analysis** menu box. Enter "1" in the **Number of Variables** box, enter "20000" in the **Number of Random Numbers** box, choose "Discrete" for the **Distribution** option, enter the cell range of the X-values and their probabilities (\$A\$1:\$B\$2) in **Value and Probability Input Range** box, and finally select Row 1 of any empty column for the **Output Range**. Click **OK** to find 20,000 realizations of X outputted in the worksheet. Using Excel's AVERAGE() function, find the average of the 20,000 X-values.

- *JMP users:* With a new data table, right-click on header of Column 1 and choose **Column Info**. In the drag-down dialog box named **Initialize Data**, pick **Random** option. Choose the bullet option of **Random Indicator**. Put the values of "0" and "500" in the first two **Value** dialog boxes, and put the values of 0.999 and 0.001 in the corresponding **Proportion** dialog boxes. Input the Enter "20000" into the **Number of rows** box, and then click **OK**. Find the average of the 20,000 X-values.

- *Minitab users:* Input the values "0" and "500" in the first two rows of column 1 (c1). Now input the corresponding probabilities of 0.999 and 0.001 in the first two rows of column 2 (c2). Do the following pull-down sequence: Calc → Random Data → Discrete. Enter "20000" in the **Number of rows of data to generate** box, type "c3" in the **Store in**

column(s) box, click-in "c1" in the **Values in** box, and click-in "c2" in the **Probabilities in** box. Click **OK** to find 20,000 realizations of X outputted in the worksheet. Find the average of the 20,000 X-values.

Whether you used Excel, JMP, or Minitab, how does the average value of the 20,000 X-values compare with the mean reported in Example 4.28?

4.130 Households and families in government data. In government data, a household consists of all occupants of a dwelling unit, while a family consists of two or more persons who live together and are related by blood or marriage. So all families form households, but some households are not families. Here are the distributions of household size and of family size in the United States:

Number of persons	1	2	3	4	5	6	7
Household probability	0.27	0.33	0.16	0.14	0.06	0.03	0.01
Family probability	0.00	0.44	0.22	0.20	0.09	0.03	0.02

Compare the two distributions using probability histograms on the same scale. Also compare the two distributions using means and standard deviations. Write a summary of your comparisons using your calculations to back up your statements.

CASE 4.3 **4.131 Perfectly negatively correlated investments.** Consider the following quote from an online site providing investment guidance: "Perfectly negatively correlated investments would provide 100% diversification, as they would form a portfolio with zero variance, which translates to zero risk." Consider a portfolio based on two investments (X and Y) with standard deviations of σ_X and σ_Y. In line with the quote, assume that the two investments are perfectly negatively correlated ($\rho = -1$).

(a) Suppose $\sigma_X = 4$, $\sigma_Y = 2$, and the portfolio mix is 70/30 of X to Y. What is the standard deviation of the portfolio? Does the portfolio have zero risk?

(b) Suppose $\sigma_X = 4$, $\sigma_Y = 2$, and the portfolio mix is 50/50. What is the standard deviation of the portfolio? Does the portfolio have zero risk?

(c) Suppose $\sigma_X = 4$, $\sigma_Y = 4$, and the portfolio mix is 50/50. What is the standard deviation of the portfolio? Does the portfolio have zero risk?

(d) Is the online quote a universally true statement? If not, how would you modify it so that it can be stated that the portfolio has zero risk?

4.132 What happens when the correlation is 1? We know that variances add if the random variables involved are uncorrelated ($\rho = 0$), but not otherwise. The opposite extreme is perfect positive correlation ($\rho = 1$). Show by using the general addition rule for variances that in this case the standard deviations add. That is, $\sigma_{X+Y} = \sigma_X + \sigma_Y$ if $\rho = 1$.

4.133 Making glassware. In a process for manufacturing glassware, glass stems are sealed by heating them in a flame. The temperature of the flame varies. Here is the distribution of the temperature X measured in degrees Celsius:

Temperature	540°	545°	550°	555°	560°
Probability	0.1	0.25	0.3	0.25	0.1

(a) Find the mean temperature μ_X and the standard deviation σ_X.

(b) The target temperature is 550°C. Use the rules for means and variances to find the mean and standard deviation of the number of degrees off target, $X - 550$.

(c) A manager asks for results in degrees Fahrenheit. The conversion of X into degrees Fahrenheit is given by

$$Y = \frac{9}{5}X + 32$$

What are the mean μ_Y and standard deviation σ_Y of the temperature of the flame in the Fahrenheit scale?

CASE 4.3 ***Portfolio analysis.*** *Here are the means, standard deviations, and correlations for the monthly returns from three Fidelity mutual funds for the 60 months ending in July 2014. Because there are three random variables, there are three correlations. We use subscripts to show which pair of random variables a correlation refers to.*

X = *Biotechnology monthly return*	$\mu_X = 2.282\%$	$\sigma_X = 6.089\%$
Y = *Information services monthly return*	$\mu_Y = 1.669\%$	$\sigma_Y = 5.882\%$
Z = *Defense and aerospace monthly return*	$\mu_Z = 1.653\%$	$\sigma_Z = 4.398\%$

Correlations

$\rho_{XY} = 0.392$ $\quad\rho_{XZ} = 0.613$ $\quad\rho_{YZ} = 0.564$

Exercises 4.134 through 4.136 make use of these historical data.

CASE 4.3 **4.134 Diversification.** Currently, Michael is exclusively invested in the Fidelity Biotechnology fund. Even though the mean return for this biotechnology fund is quite high, it comes with greater volatility and risk. So, he decides to diversify his portfolio by constructing a portfolio of 80% biotechnology fund and 20% information services fund. Based on the provided historical performance, what is the expected return and standard deviation of the portfolio? Relative to his original investment scheme, what is the percentage reduction in his risk level (as measured by standard deviation) by going to this particular portfolio?

CASE 4.3 **4.135 More on diversification.** Continuing with the previous exercise, suppose Michael's primary goal is to seek a portfolio mix of the biotechnology and information services funds that will give him *minimal* risk as measured by standard deviation of the portfolio. Compute the standard deviations for portfolios based on the proportion of biotechnology fund in the portfolio ranging from 0 to 1 in increments of 0.1. You may wish to do these calculations in Excel. What is your recommended mix of biotechnology and information services funds for Michael? What is the standard deviation for your recommended portfolio?

CASE 4.3 **4.136 Larger portfolios.** Portfolios often contain more than two investments. The rules for means and variances continue to apply, though the arithmetic gets messier. A portfolio containing proportions a of Biotechnology Fund, b of Information Services Fund, and c of Defense and Aerospace Fund has return $R = aX + bY + cZ$. Because a, b, and c are the proportions invested in the three funds, $a + b + c = 1$. The mean and variance of the portfolio return R are

$$\mu_R = a\mu_X + b\mu_Y + c\mu_Z$$

$$\sigma_R^2 = a^2\sigma_X^2 + b^2\sigma_Y^2 + c^2\sigma_Z^2 + 2ab\rho_{XY}\sigma_X\sigma_Y + 2ac\rho_{XZ}\sigma_X\sigma_Z + 2bc\rho_{YZ}\sigma_Y\sigma_Z$$

Having seen the advantages of diversification, Michael decides to invest his funds 20% in biotechnology, 35% in information services, and 45% in defense and aerospace. What are the (historical) mean and standard deviation of the monthly returns for this portfolio?

CHAPTER 4 Review Exercises

4.137 Using probability rules. Let $P(A) = 0.7$, $P(B) = 0.6$, and $P(C) = 0.2$.
(a) Explain why it is not possible that events A and B can be disjoint.
(b) What is the smallest possible value for $P(A \text{ and } B)$? What is the largest possible value for $P(A \text{ and } B)$? It might be helpful to draw a Venn diagram.
(c) If events A and C are independent, what is $P(A \text{ or } C)$?

4.138 Work with a transformation. Here is a probability distribution for a random variable X:

Value of X	1	2
Probability	0.4	0.6

(a) Find the mean and the standard deviation of this distribution.
(b) Let $Y = 4X - 2$. Use the rules for means and variances to find the mean and the standard deviation of the distribution of Y.
(c) For part (b), give the rules that you used to find your answer.

4.139 A different transformation. Refer to the previous exercise. Now let $Y = 4X^2 - 2$.
(a) Find the distribution of Y.
(b) Find the mean and standard deviation for the distribution of Y.
(c) Explain why the rules that you used for part (b) of the previous exercise do not work for this transformation.

4.140 Roll a pair of dice two times. Consider rolling a pair of fair dice two times. For a given roll, consider the total on the up-faces. For each of the following pairs of events, tell whether they are disjoint, independent, or neither.
(a) $A = 2$ on the first roll, $B = 8$ or more on the first roll.
(b) $A = 2$ on the first roll, $B = 8$ or more on the second roll.
(c) $A = 5$ or less on the second roll, $B = 4$ or less on the first roll.
(d) $A = 5$ or less on the second roll, $B = 4$ or less on the second roll.

4.141 Find the probabilities. Refer to the previous exercise. Find the probabilities for each event.

4.142 Some probability distributions. Here is a probability distribution for a random variable X:

Value of X	2	3	4
Probability	0.2	0.4	0.4

(a) Find the mean and standard deviation for this distribution.
(b) Construct a different probability distribution with the same possible values, the same mean, and a larger standard deviation. Show your work and report the standard deviation of your new distribution.

(c) Construct a different probability distribution with the same possible values, the same mean, and a smaller standard deviation. Show your work and report the standard deviation of your new distribution.

4.143 Wine tasters. Two wine tasters rate each wine they taste on a scale of 1 to 5. From data on their ratings of a large number of wines, we obtain the following probabilities for both tasters' ratings of a randomly chosen wine:

	Taster 2				
Taster 1	**1**	**2**	**3**	**4**	**5**
1	0.03	0.02	0.01	0.00	0.00
2	0.02	0.07	0.06	0.02	0.01
3	0.01	0.05	0.25	0.05	0.01
4	0.00	0.02	0.05	0.20	0.02
5	0.00	0.01	0.01	0.02	0.06

(a) Why is this a legitimate assignment of probabilities to outcomes?
(b) What is the probability that the tasters agree when rating a wine?
(c) What is the probability that Taster 1 rates a wine higher than 3? What is the probability that Taster 2 rates a wine higher than 3?

4.144 Slot machines. Slot machines are now video games, with winning determined by electronic random number generators. In the old days, slot machines were like this: you pull the lever to spin three wheels; each wheel has 20 symbols, all equally likely to show when the wheel stops spinning; the three wheels are independent of each other. Suppose that the middle wheel has eight bells among its 20 symbols, and the left and right wheels have one bell each.
(a) You win the jackpot if all three wheels show bells. What is the probability of winning the jackpot?
(b) What is the probability that the wheels stop with exactly two bells showing?

4.145 Bachelor's degrees by gender. Of the 2,325,000 bachelor's, master's, and doctoral degrees given by U.S. colleges and universities in a recent year, 69% were bachelor's degrees, 28% were master's degrees, and the rest were doctorates. Moreover, women earned 57% of the bachelor's degrees, 60% of the master's degrees, and 52% of the doctorates.[31] You choose a degree at random and find that it was awarded to a woman. What is the probability that it is a bachelor's degree?

4.146 Higher education at two-year and four-year institutions. The following table gives the counts of U.S. institutions of higher education classified as public or private and as two-year or four-year:[32]

	Public	**Private**
Two-year	1000	721
Four-year	2774	672

Convert the counts to probabilities, and summarize the relationship between these two variables using conditional probabilities.

4.147 Wine tasting. In the setting of Exercise 4.143, Taster 1's rating for a wine is 3. What is the conditional probability that Taster 2's rating is higher than 3?

4.148 An interesting case of independence. Independence of events is not always obvious. Toss two balanced coins independently. The four possible combinations of heads and tails in order each have probability 0.25. The events

A = head on the first toss
B = both tosses have the same outcome

may seem intuitively related. Show that $P(B \mid A) = P(B)$ so that A and B are, in fact, independent.

4.149 Find some conditional probabilities. Choose a point at random in the square with sides $0 \le x \le 1$ and $0 \le y \le 1$. This means that the probability that the point falls in any region within the square is the area of that region. Let X be the x coordinate and Y the y coordinate of the point chosen. Find the conditional probability $P(Y < 1/3 \mid Y > X)$. (*Hint:* Sketch the square and the events $Y < 1/3$ and $Y > X$.)

4.150 Sample surveys for sensitive issues. It is difficult to conduct sample surveys on sensitive issues because many people will not answer questions if the answers might embarrass them. ***Randomized response*** is an effective way to guarantee anonymity while collecting information on topics such as student cheating or sexual behavior. Here is the idea. To ask a sample of students whether they have plagiarized a term paper while in college, have each student toss a coin in private. If the coin lands heads *and* they have not plagiarized, they are to answer No. Otherwise, they are to give Yes as their answer. Only the student knows whether the answer reflects the truth or just the coin toss, but the researchers can use a proper random sample with follow-up for nonresponse and other good sampling practices.

Suppose that, in fact, the probability is 0.3 that a randomly chosen student has plagiarized a paper. Draw a tree diagram in which the first stage is tossing the coin and the second is the truth about plagiarism.

The outcome at the end of each branch is the answer given to the randomized-response question. What is the probability of a No answer in the randomized-response poll? If the probability of plagiarism were 0.2, what would be the probability of a No response on the poll? Now suppose that you get 39% No answers in a randomized-response poll of a large sample of students at your college. What do you estimate to be the percent of the population who have plagiarized a paper?

CASE 4.2 **4.151 Blood bag demand.** Refer to the distribution of daily demand for blood bags X in Case 4.2 (pages 210–211). Assume that demand is independent from day to day.
(a) What is the probability at least one bag will be demanded every day of a given month? Assume 30 days in the month.
(b) What is the interpretation of one minus the probability found part (a)?
(c) What is the probability that the bank will go a whole year (365 days) without experiencing a demand of 12 bags on a given day?

4.152 Risk pooling in a supply chain. Example 4.39 (pages 232–233) compares a decentralized versus a centralized inventory system as it ultimately relates to the amount of safety stock (extra inventory over and above mean demand) held in the system. Suppose that the CEO of ElectroWorks requires a 99% customer service level. This means that the probability of satisfying customer demand during the lead time is 0.99. Assume that lead time demands for the Milwaukee warehouse, Chicago warehouse, and centralized warehouse are Normally distributed with the means and standard deviations found in the example.
(a) For a 99% service level, how much safety stock of the part SurgeArrester does the Milwaukee warehouse need to hold? Round your answer to the nearest integer.
(b) For a 99% service level, how much safety stock of the part SurgeArrester does the Chicago warehouse need to hold? Round your answer to the nearest integer.
(c) For a 99% service level, how much safety stock of the part SurgeArrester does the centralized warehouse need to hold? Round your answer to the nearest integer. How many more units of the part need to be held in the decentralized system than in the centralized system?

4.153 Life insurance. Assume that a 25-year-old man has these probabilities of dying during the next five years:

Age at death	25	26	27	28	29
Probability	0.00039	0.00044	0.00051	0.00057	0.00060

(a) What is the probability that the man does not die in the next five years?
(b) An online insurance site offers a term insurance policy that will pay $100,000 if a 25-year-old man dies within the next five years. The cost is $175 per year. So the insurance company will take in $875 from this policy if the man does not die within five years. If he does die, the company must pay $100,000. Its loss depends on how many premiums the man paid, as follows:

Age at death	25	26	27	28	29
Loss	$99,825	$99,650	$99,475	$99,300	$99,125

What is the insurance company's mean cash intake (income) from such polices?

4.154 Risk for one versus many life insurance policies. It would be quite risky for an insurance company to insure the life of only one 25-year-old man under the terms of Exercise 4.153. There is a high probability that person would live and the company would gain $875 in premiums. But if he were to die, the company would lose almost $100,000. We have seen that the risk of an investment is often measured by the standard deviation of the return on the investment. The more variable the return is (the larger σ is), the riskier the investment.
(a) Suppose only one person's life is insured. Compute standard deviation of the income X that the insurer will receive. Find σ_X, using the distribution and mean you found in Exercise 4.153.
(b) Suppose that the insurance company insures two men. Define the total income as $T = X_1 + X_2$ where X_i is the income made from man i. Find the mean and standard deviation of T.
(c) You should have found that the standard deviation computed in part (b) is greater than that found in part (a). But this does not necessarily imply that insuring two people is riskier than insuring one person. What needs to be recognized is that the mean income has also gone up. So, to measure the riskiness of each scenario we need to scale the standard deviation values relative to the mean values. This is simply done by computing σ/μ, which is called the ***coefficient of variation*** (CV). Compute the coefficients of variation for insuring one person and for insuring two people. What do the CV values suggest about the relative riskiness of the two scenarios?
(d) Compute the mean total income, standard deviation of total income, and the CV of total income when 30 people are insured.

(e) Compute the mean total income, standard deviation of total income, and the CV of total income when 1000 people are insured.

(f) There is a remarkable result in probability theory that states that the sum of a large number of independent random variables follows approximately the Normal distribution even if the random variables themselves are not Normal. In most cases, 30 is sufficiently "large." Given this fact, use the mean and standard deviation from part (d) to compute the probability that the insurance company will lose money from insuring 30 people—that is, compute $P(T < 0)$. Compute now the probability of a loss to the company if 1000 people are insured. What did you learn from these probability computations?