Jgroup/Dreamstime.com

CHAPTER 4

Probability: The Study of Randomness

Introduction

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 or randomized comparative experiments, the laws of probability answer the question, "What would happen if we did this many times?" Games of chance like Texas hold 'em are exciting because the outcomes are determined by the rules of probability.

4.1 Randomness

When you complete this section, you will be able to:

- Identify random phenomena.
- Interpret the term "probability" for particular examples.
- Identify trials as independent or not.

AU/DE/PE: set ital, no quotes for consistency. See pp. 80, 132, 165

Toss a coin, or choose a simple random sample (SRS). The result can't be predicted in advance because the result will vary when you toss the coin or choose the sample repeatedly. But there is, nonetheless, 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

Toss a coin 5000 times. 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 (Trial A and Trial B). For each number of tosses from 1 to 5000, we have plotted the proportion of those tosses that gave a head.

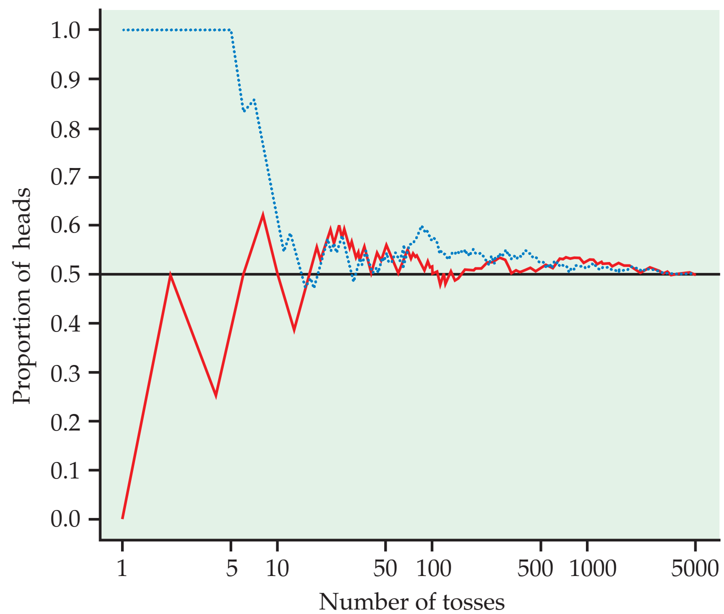


FIGURE 4.1 The proportion of tosses of a coin that give a head varies as we make more tosses. Eventually, however, the proportion approaches 0.5, the probability of a head. This figure shows the results of two trials of 5000 tosses each, Example 4.1.

Trial A (red 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 (blue dotted line), on the other hand, starts with five straight heads, so the proportion of heads is 1 until the sixth toss.

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 each trial gets close to 0.5 and stays there.

probability

If we made yet a third trial at tossing the coin a great many 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 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. You will see that the proportion of heads gradually settles down close to the chosen probability. Equally important, you will also see that the proportion in a small or moderate number of tosses can be far from this probability. *Probability describes only what happens in the long run. Most people expect chance outcomes to show more short-term regularity than is actually true.*

EXAMPLE 4.2

Significance testing and Type I errors. In Chapter 6, we will learn about significance testing and Type I errors. When we perform a significance test, we have the possibility of making a Type I error under certain circumstances. The significance-testing procedure is set up so that the probability of making this kind of error is small, usually 5%. If we perform a large number of significance tests under this set of circumstances, the proportion of times that we will make a Type I error is 0.05.

fair coin

In the coin toss setting, the probability of a head is a characteristic of the coin being tossed. A coin is called **fair** if the probability of a head is 0.5; that is, it is equally likely to come up heads or tails. If we toss a coin five times and it comes up heads for all five tosses, we might suspect that the coin is not fair. Is this outcome likely if, in fact, the coin is fair? This is what happened in Trials A and B of Example 4.1. We will learn a lot more about significance testing in later chapters. For now, we are content with some very general ideas.

When the Type I error of a statistical significance procedure is set at 0.05, this probability is a characteristic of the procedure. If we roll a pair of dice once, we do not know whether the sum of the faces will be seven or not. Similarly, if we perform a significance test once, we do not know if we will make a Type I error or not. However, if the procedure is designed to have a Type I error probability of 0.05, then we are much less likely than not to make a Type I error.

The language of probability

"Random" in statistics is not a synonym for "unpredictable" but a description of a kind of order that emerges 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."

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.

Not all coins are fair. In fact, most real coins have bumps and imperfections that make the probability of heads a little different from 0.5. The probability might be 0.499999 or 0.500002. For our study of probability in this chapter, we will assume that we know the actual values of probabilities. Thus, we assume things like fair coins, even though we know that real coins are not exactly fair. We do this to learn what kinds of outcomes we are likely to see when we make such assumptions. When we study statistical inference in later chapters, we look at the situation from the opposite point of view: given that we have observed certain outcomes, what can we say about the probabilities that generated these outcomes?

USE YOUR KNOWLEDGE

4.1 Use Table B. We can use the random digits in Table B in the back of the book to simulate tossing a fair coin. Start at line 131 and read the numbers from left to right. If the number is 0, 2, 4, 6, or 8, you will say that the coin toss resulted in a head; if the number is a 1, 3, 5, 7, or 9, the outcome is tails. Use the first 10 random digits on line 131 to simulate 10 tosses of a fair coin. What is the actual proportion of heads in your simulated sample? Explain why you did not get exactly five heads.

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 4.3

Many tosses of a coin. 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 statistician John Kerrich tossed a coin 10,000 times. Result: 5067 heads, proportion of heads 0.5067.

Thinking about randomness

That some things are random is an observed fact about the world. The outcome of a coin toss, the time between emissions of particles by a radioactive source, and the sexes of the next litter of lab rats are all random. So is the outcome of a random sample or a randomized experiment. Probability theory is the branch of mathematics that describes random behavior. Of course, we can 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 software 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. Simulations start with given probabilities and imitate random behavior, but we can estimate a real-world probability only by actually observing many trials.

• Nonetheless, 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. The kinds of physical random devices suggested in the exercises are too slow to make performing so many trials practical. Short runs give only rough estimates of a probability.

The uses of probability

Probability theory originated in the study of games of chance. Tossing dice, dealing shuffled cards, and spinning a roulette wheel are examples of deliberate randomization. In that respect, they are similar to random sampling. Although games of chance are ancient, they were not studied by mathematicians until the sixteenth and seventeenth centuries.

It is only a mild simplification to say that probability as a branch of mathematics arose when seventeenth-century French gamblers asked the mathematicians Blaise Pascal and Pierre de Fermat for help. Gambling is still with us, in casinos and state lotteries. We will make use of games of chance as simple examples that illustrate the principles of probability.

Careful measurements in astronomy and surveying led to further advances in probability in the eighteenth and nineteenth centuries because the results of repeated measurements are random and can be described by distributions much like those arising from random sampling. Similar distributions appear in data on human life span (mortality tables) and in data on lengths or weights in a population of skulls, leaves, or cockroaches.[1]

Now, we employ the mathematics of probability to describe the flow of traffic through a highway system, the Internet, or a computer processor; the genetic makeup of individuals or populations; the energy states of subatomic particles; the spread of epidemics or tweets; and the rate of return on risky investments. Although we are interested in probability because of its usefulness in statistics, the mathematics of chance is important in many fields of study.

SECTION 4.1 SUMMARY

• A **random phenomenon** has outcomes that we cannot predict 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 218.

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 your town at noon on Groundhog Day, February 2.

(b) The first digit in your student identification number.

(c) You draw an ace from a well-shuffled deck of 52 cards.

note query p. 215

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) You record the outside temperature in your town at noon on Groundhog Day, February 2, 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 Winning at craps. The game of craps starts with a "come-out" roll, in which the shooter rolls a pair of dice. If the total of the "spots" on the up-faces is 7 or 11, the shooter wins immediately (there are ways that the shooter can win on later rolls if other numbers are rolled on the come-out roll). Roll a pair of dice 25 times and estimate the probability that the shooter wins immediately on the come-out roll. For a pair of perfectly made dice, the probability is 0.2222.

AU: this OK here? seems to give answer.

4.6 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 200 tosses. Again record the proportion and count of heads and the difference between the count and 100 (half the number of tosses).

(c) Keep going. Stop at 300 tosses and again at 400 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.

4.7 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?

4.2 Probability Models

When you complete this section, you will be able to:

- Describe a sample space from a description of a random phenomenon.
- Apply the four probability rules.
- Identify random phenomena that have equally likely outcomes and distinguish them from those that do not.

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 probability model

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:

- A list of possible outcomes.
- A probability for each outcome.

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

Sample spaces

A probability model first tells us what outcomes are possible.

SAMPLE SPACE

The **sample space** $\boldsymbol{S}$ of a random phenomenon is the set of all 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 4.4

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 4.5

Sample space for random digits. Let your pencil point fall blindly into Table B of random digits. Record the value of the digit it lands on. The possible outcomes are

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

EXAMPLE 4.6

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 typical 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 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 an opinion poll you select four people at random from a large population and ask each if he or she favors reducing federal spending on low-interest student loans. The answers are Yes or No. The possible outcomes—the sample space—are exactly as in Example 4.6 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 probability model.

USE YOUR KNOWLEDGE

4.8 When were you born? A student is asked “In what month were you born? Set up an appropriate sample space for this setting.

The sample spaces described in Examples 4.4, 4.5, and 4.6 correspond to categorical variables where we can list all the possible values. Other sample spaces correspond to quantitative variables. Here is an example.

EXAMPLE 4.7

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. 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, Minitab generates random numbers like 0.736891, 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.

USE YOUR KNOWLEDGE

4.9 How many hours do you text? You record the number of hours per week that a randomly selected student spends texting. What is the sample space?

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 4.8

Exactly one head 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 one head" is an event. Call this event A. The event A expressed as a set of outcomes is

$$A = \{\text{HTTT, THTT, TTHT, TTTH}\}$$

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 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 in 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 (say) 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 through 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.

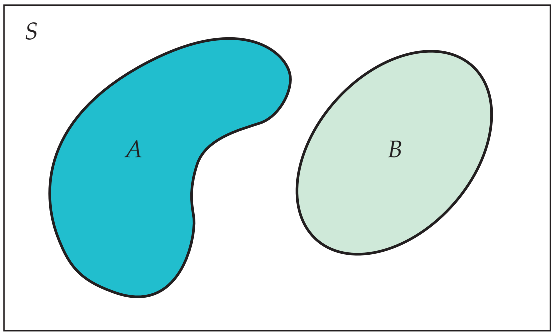


FIGURE 4.2 Venn diagram showing disjoint events *A* and *B*. Disjoint events have no common outcomes.

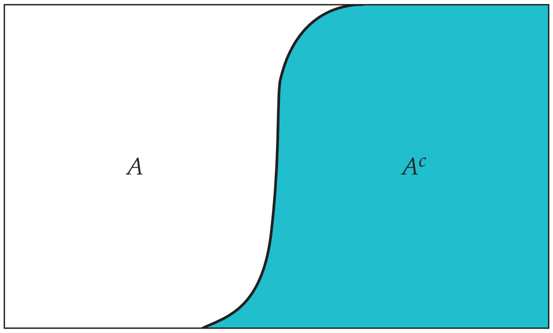


FIGURE 4.3 Venn diagram showing the complement A^c of an event *A*. The complement consists of all outcomes that are not in *A*.

PROBABILITY RULES

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

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

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

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.

Venn diagram

EXAMPLE 4.9

Norito/iStockphoto

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 what is their favorite color.

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

EXAMPLE 4.10

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 their favorite color. Suppose that we want to find the probability that the favorite color is not blue.

EXAMPLE 4.11

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.

USE YOUR KNOWLEDGE

4.10 Red or brown. Find the probability that the favorite color is red or brown.

4.11 White, black, silver, gray, or red. 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.

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.

EXAMPLE 4.12

Benford's law. Faked numbers in tax returns, payment records, invoices, expense account claims, and many other settings often display patterns that aren't present in legitimate records. Some patterns, such as too many round numbers, are obvious and easily avoided by a clever crook. Others are more subtle. 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 a first digit can't be 0):[3]

Benford's law

First digit	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

Benford's law usually applies to the first digits of the sizes of similar quantities, such as invoices, expense account claims, and county populations. Investigators can detect fraud by comparing the first digits in records such as invoices paid by a business with these probabilities.

EXAMPLE 4.13

Find some probabilities for Benford's law. Consider the events

$$A = \{\text{first digit is } 4\}$$
$$B = \{\text{first digit is } 7 \text{ or more}\}$$

From the table of probabilities in Example 4.12,

$$P(A) = P(4) = 0.097$$
$$P(B) = P(7) + P(8) + P(9)$$
$$= 0.058 + 0.051 + 0.046 = 0.155$$

Note that $P(B)$ is not the same as the probability that a first digit is strictly more than 7. The probability $P(7)$ that a first digit is 7 is included in "7 or more" but not in "more than 7."

USE YOUR KNOWLEDGE

4.12 Benford's law. Using the probabilities for Benford's law, find the probability that a first digit is anything other than 1.

4.13 Use the addition rule. Use the addition rule with the probabilities for the events A and B from Example 4.13 to find the probability that a first digit is either 4 or 7 or more.

Be careful to apply the addition rule only to disjoint events.

EXAMPLE 4.14

Find more probabilities for Benford's law. Check that the probability of the event C that a first digit is odd is

$$P(C) = P(1) + P(3) + P(5) + P(7) + P(9) = 0.609$$

The probability

$$P(B \text{ or } C) = P(1) + P(3) + P(5) + P(7) + P(8) + P(9) = 0.660$$

is *not* the sum of $P(B)$ and $P(C)$ because events B and C are not disjoint. Outcomes 7 and 9 are common to both 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 4.15

First digits that are equally likely. You might think that first digits are distributed "at random" among the digits 1 to 9 in business records. 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 7 or more is

$$\begin{aligned} P(B) &= P(7) + P(8) + P(9) \\ &= \frac{1}{9} + \frac{1}{9} + \frac{1}{9} = \frac{3}{9} = 0.333 \end{aligned}$$

Compare this with the Benford's law probability in Example 4.13. A person who fakes data by using "random" digits will end up with too many first digits that are 7 or more.

In Example 4.15, 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 7 or more, the probability of this event is 3/9. In the special situation where 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

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

Most random phenomena do not have equally likely outcomes, so the general rule for finite sample spaces (page 224) is more important than the special rule for equally likely outcomes.

USE YOUR KNOWLEDGE

4.14 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 perfect 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 will occur in the special situation when A and B cannot occur together because they are disjoint. Our final rule describes the probability that *both* events A and B occur, again only in a special situation. More general rules appear in Section 4.5, but in our study of statistics, we will need only the rules that apply to special situations.

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

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

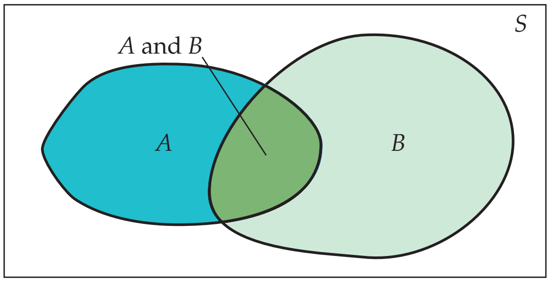
The events A and B are not disjoint. They occur together whenever both tosses give heads. We want to compute the probability of the event $\{A \text{ and } B\}$ that *both* tosses are heads. The Venn diagram in Figure 4.4 illustrates the event $\{A \text{ and } B\}$ as the overlapping area that is common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

The coin tossing of Buffon, Pearson, and Kerrich described in Example 4.3 makes us willing to assign probability 1/2 to a head when we toss a coin. So

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

What is $P(A \text{ and } B)$? Our common sense says that it is 1/4. The first toss will give a head half the time and the second toss will give a head half the time, so both tosses will give heads on $1/2 \times 1/2 = 1/4$ of all trials in the long run. This reasoning assumes that the second toss still has probability 1/2 of a head after the first has given a head. This is true—we can verify it by tossing a coin twice 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.

FIGURE 4.4 Venn diagram showing the event $\{A \text{ and } B\}$. This event consists of outcomes common to A and B.



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 will make this informal idea precise in Section 4.5. 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. Here is an example of independence.

EXAMPLE 4.16

Coins do not have memory. Because a coin has no memory, we assume that coin tosses are independent. For a fair coin, this means that the outcome of the first toss does not influence the outcome of any other toss.

USE YOUR KNOWLEDGE

4.15 A head and then a tail in two tosses. What is the probability of obtaining a head and then a tail on two tosses of a fair coin?

Here is an example of a situation where there are dependent events.

EXAMPLE 4.17

Dependent events in cards. The colors of successive cards dealt from the same deck are not independent. A standard 52-card deck contains 26 red and 26 black cards. For the first card dealt from a shuffled deck, the probability of a red card is $26/52 = 0.50$ because the 52 possible cards are equally likely. Once we see that the first card is red, we know that there are only 25 reds among the remaining 51 cards. The probability that the second card is red is therefore only $25/51 = 0.49$. Knowing the outcome of the first deal changes the probabilities for the second.

USE YOUR KNOWLEDGE

4.16 The probability of a second ace. A deck of 52 cards contains four aces, so the probability that a card drawn from this deck is an ace is 4/52. If we know that the first card drawn is an ace, what is the probability that the second card drawn is also an ace? Using the idea of independence, explain why this probability is not 4/52.

Here is another example of a situation where events are dependent.

EXAMPLE 4.18

Taking a test twice. If you take an IQ test or other mental test twice in succession, the two test scores are not independent. The learning that occurs on the first attempt influences your second attempt. If you learn a lot, then your second test score might be a lot higher than your first test score.

When independence is part of a probability model, the multiplication rule applies. Here is an example.

EXAMPLE 4.19

Profimedia.CZ a.s./Alamy

Mendel's peas. Gregor Mendel used garden peas in some of the experiments that revealed that inheritance operates randomly. The seed color of Mendel's peas can be either green or yellow. Two parent plants are "crossed" (one pollinates the other) to produce seeds.

AU: use decimal rather than fraction, as in equation below?

Each parent plant carries two genes for seed color, and each of these genes has probability 1/2 of being passed to a seed. The two genes that the seed receives, one from each parent, determine its color. The parents contribute their genes independently of each other.

Suppose that both parents carry the G and the Y genes. The seed will be green if both parents contribute a G gene; otherwise, it will be yellow. If M is the event that the male contributes a G gene and F is the event that the female contributes a G gene, then the probability of a green seed is

$$\begin{aligned} P(M \text{ and } F) &= P(M)\,P(F) \\ &= (0.5)(0.5) = 0.25 \end{aligned}$$

In the long run, 1/4 of all seeds produced by crossing these plants will be green.

The multiplication rule applies only to independent events; you cannot use it if events are not independent. Here is a distressing example of misuse of the multiplication rule.

EXAMPLE 4.20

Sudden infant death syndrome. Sudden infant death syndrome (SIDS) causes babies to die suddenly (often in their cribs) with no explanation. Deaths from SIDS have been greatly reduced by placing babies on their backs, but as yet no cause is known.

When more than one SIDS death occurs in a family, the parents are sometimes accused. One "expert witness" popular with prosecutors in England told juries that there is only a 1 in 73 million chance that two children in the same family could have died from SIDS. Here's his calculation: the rate of SIDS in a nonsmoking middle-class family is 1 in 8500. So the probability of two deaths is

$$\frac{1}{8500} \times \frac{1}{8500} = \frac{1}{72{,}250{,}000}$$

Several women were convicted of murder on this basis, without any direct evidence that they harmed their children.

As the Royal Statistical Society said, this reasoning is nonsense. It assumes that SIDS deaths in the same family are independent events. The cause of SIDS is unknown: "There may well be unknown genetic or environmental factors that predispose families to SIDS, so that a second case within the family becomes much more likely."[4] The British government decided to review the cases of 258 parents convicted of murdering their babies.

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 formulas 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 again at Figure 4.2 (page 224). Unlike disjointness or complements, independence cannot be pictured by a Venn diagram because it involves the probabilities of the events rather than just the outcomes that make up the events. However, it could be displayed in a mosaic plot.

LOOK BACK
mosaic plot, p. 143

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^c. 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 $(0.75)(0.25) = 0.1875$.

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 4.21

HIV testing. Many people who come to clinics to be tested for HIV, the virus that causes AIDS, don't come back to learn the test results. Clinics now use "rapid HIV tests" that give a result in a few minutes. The false-positive rate for a diagnostic test is the probability that a person with no disease will have a positive test result. For the rapid HIV tests, the Food and Drug Administration (FDA) has established 2% as the maximum false-positive rate allowed.[5] If a clinic uses a test that matches the FDA standard and tests 50 people who are free of HIV antibodies, what is the probability that at least one false-positive will occur?

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 0.02, so the probability of a negative result is $1 - 0.02 = 0.98$ by the complement rule. The probability of at least one false-positive among the 50 people tested is, therefore,

$$\begin{aligned} P(\text{at least 1 positive}) &= 1 - P(\text{no positives}) \\ &= 1 - P(\text{50 negatives}) \\ &= 1 - 0.98^{50} \\ &= 1 - 0.3642 = 0.6358 \end{aligned}$$

There is approximately a 64% chance that at least 1 of the 50 people will test positive for HIV even though none of them has the virus.

Concern about excessive numbers of false-positives led the New York City Department of Health and Mental Hygiene to suspend the use of one particular rapid HIV test.[6]

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. $0 \le P(A) \le 1$ 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 Exercise 4.8, see page 222; for Exercise 4.9, see page 222; for Exercises 4.10 and 4.11, see page 225; for Exercises 4.12 and 4.13, see page 226; for Exercise 4.14, see page 228; for Exercise 4.15, see page 229; and for Exercise 4.16, see page 229.

4.17 What is the sample space? For each of the following questions, define a sample space for the associated random phenomenon. Explain your answers. Be sure to specify units if that is appropriate.

(a) Will it rain tomorrow?

(b) How many times do you tweet in a typical day?

(c) What is the average age of your Facebook friends?

(d) What are the majors for students at your college?

4.18 Probability rules. For each of the following situations, state the probability rule or rules that you would use and apply it or them. Write a sentence explaining how the situation illustrates the use of the probability rules.

(a) The probability of event A is 0.417. What is the probability that event A does not occur?

(b) A coin is tossed four times. The probability of zero heads is 1/16 and the probability of zero tails is 1/16. What is the probability that all four tosses result in the same outcome?

(c) Refer to part (b). What is the probability that there is at least one head and at least one tail?

(d) The probability of event A is 0.4 and the probability of event B is 0.8. Events A and B are disjoint. Can this happen?

(e) Event A is very rare. Its probability is -0.04. Can this happen?

4.19 Equally likely events. For each of the following situations, explain why you think that the events are equally likely or not. Explain your answers.

(a) The outcome of the next tennis match for Sloane Stevens is either a win or a loss. (You might want to check the Internet for information about this tennis player.)

(b) You roll a fair die and get a 3 or a 4.

(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.

4.20 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.

4.21 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 0.7 and the probability of B is 0.5, the probability of both A and B happening is 1.2.

(c) If the probability of A is 0.45, then the probability of the complement of A is -0.45.

4.22 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 3 is 0.3.

(c) If the probability of A is 0.3, the probability of B is 0.4, and the probability of A and B is 0.5, then A and B are independent.

4.23 Evaluating web page designs. You are a web page designer and you set up a page with four different links. A user of the page can click on one of the links or he or she can leave that page. Describe the sample space for the outcome of someone visiting your web page.

4.24 Record the length of time spent on the page. Refer to the previous exercise. You also decide to measure the length of time a visitor spends on your page. Give the sample space for this measure.

4.25 Distribution of blood types. All human blood can be "ABO-typed" as one of O, A, B, or AB, but the distribution of the types varies a bit among groups of people. Here is the distribution of blood types for a randomly chosen person in the United States:[7]

Blood type	A	B	AB	O
U.S. probability	0.42	0.11	?	0.44

(a) What is the probability of type AB blood in the United States?

(b) Maria has type B blood. She can safely receive blood transfusions from people with blood types O and B. What is the probability that a randomly chosen person from the United States can donate blood to Maria?

4.26 Blood types in Ireland. The distribution of blood types in Ireland differs from the U.S. distribution given in the previous exercise:

Blood type	A	B	AB	O
Ireland probability	0.35	0.10	0.03	0.52

Choose a person from the United States and a person from Ireland at random, independently of each other. What is the probability that both have type O blood? What is the probability that both have the same blood type?

4.27 Are the probabilities legitimate? In each of the following situations, state whether or not the given assignment of probabilities to individual outcomes is legitimate—that is, it satisfies the rules of probability. If not, give specific reasons for your answer.

(a) Choose a college student at random and record gender and enrollment status: $P(\text{female full-time}) = 0.44$, $P(\text{female part-time}) = 0.56$, $P(\text{male full-time}) = 0.46$, $P(\text{male part-time}) = 0.54$.

(b) Deal a card from a shuffled deck: $P(\text{clubs}) = 16/52$, $P(\text{diamonds}) = 12/52$, $P(\text{hearts}) = 12/52$, $P(\text{spades}) = 12/52$.

(c) Roll a die and record the count of spots on the up-face: $P(1) = 1/3$, $P(2) = 0$, $P(3) = 1/6$, $P(4) = 1/3$, $P(5) = 1/6$, $P(6) = 0$.

4.28 French and English in Canada. 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:[8]

Language	English	French	Asian/Pacific	Other
Probability	0.59	?	0.07	0.11

(a) What probability should replace "?" in the distribution?

(b) What is the probability that a Canadian's mother tongue is not English? Explain how you computed your answer.

4.29 Education levels of young adults. Choose a young adult (age 25 to 34 years) at random. The probability is 0.12 that the person chosen did not complete high school, 0.31 that the person has a high school diploma but no further education, and 0.29 that the person has at least a bachelor's degree.

(a) What must be the probability that a randomly chosen young adult has some education beyond high school but does not have a bachelor's degree?

(b) What is the probability that a randomly chosen young adult has at least a high school education?

4.30 Loaded dice. There are many ways to produce crooked dice. To *load* a die so that 6 comes up too often and 1 (which is opposite 6) comes up too seldom, add a bit of lead to the filling of the spot on the 1 face. Because the spot is solid plastic, this works even with transparent dice. If a die is loaded so that 6 comes up with probability 0.24 and the probabilities of the 2, 3, 4, and 5 faces are not affected, what is the assignment of probabilities to the six faces?

4.31 Rh blood types. Human blood is typed as O, A, B, or AB and also as Rh-positive or Rh-negative. ABO type and Rh-factor type are independent because they are governed by different genes. In the American population, 84% of people are Rh-positive. Use the information about ABO type in Exercise 4.25 to give the probability distribution of blood type (ABO and Rh) for a randomly chosen American.

4.32 Roulette. A roulette wheel has 38 slots, numbered 0, 00, and 1 to 36. The slots 0 and 00 are colored green, 18 of the others are red, and 18 are black. The dealer spins the wheel and, at the same time, rolls a small ball along the wheel in the opposite direction. The wheel is carefully balanced so that the ball is equally likely to land in any slot when the wheel slows. Gamblers can bet on various combinations of numbers and colors.

(a) What is the probability that the ball will land in any one slot?

(b) If you bet on "red," you win if the ball lands in a red slot. What is the probability of winning?

(c) The slot numbers are laid out on a board on which gamblers place their bets. One column of numbers on the board contains all multiples of 3, that is, 3, 6, 9, . . . , 36. You place a "column bet" that wins if any of these numbers comes up. What is your probability of winning?

4.33 Winning the lottery. A state lottery's Pick 3 game asks players to choose a three-digit number, 000 to 999. The state chooses the winning three-digit number at random so that each number has probability 1/1000. You win if the winning number contains the digits in your number, in any order.

(a) Your number is 059. What is your probability of winning?

(b) Your number is 223. What is your probability of winning?

4.34 PINs. The personal identification numbers (PINs) for automatic teller machines usually consist of four digits. You notice that most of your PINs have at least one 0, and you wonder if the issuers use lots of 0s to make the numbers easy to remember. Suppose that PINs are assigned at random, so that all four-digit numbers are equally likely.

(a) How many possible PINs are there?

(b) What is the probability that a PIN assigned at random has at least one 0?

4.35 Universal blood donors. People with type O-negative blood are universal donors. That is, any patient can receive a transfusion of O-negative blood. Only 7% of the American population have O-negative blood. If eight people appear at random to give blood, what is the probability that at least one of them is a universal donor?

4.36 Axioms of probability. Show that any assignment of probabilities to events that obeys Rules 2 and 3 on page 224 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.

4.37 Independence of complements. Show that if events A and B obey the multiplication rule, $P(\text{A and B}) = P(A)P(B)$, 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.)

Mendelian inheritance. *Some traits of plants and animals depend on inheritance of a single gene. This is called Mendelian inheritance, after Gregor Mendel (1822–1884). Exercises 4.38 through 4.41 are based on the following information about Mendelian inheritance of blood type.*

Each of us has an ABO blood type, which describes whether two characteristics, called A and B, are present. Every one of us has two blood type alleles (gene forms), one inherited from our mother and one from our father. Each of these alleles can be A, B, or O. Which two we inherit determines our blood type. Here is a table that *shows what our blood type is for each combination of two alleles:*

Alleles inherited	Blood type
A and A	A
A and B	AB
A and O	A
B and B	B
B and O	B
O and O	O

We inherit each of a parent's two alleles with probability 0.5. We inherit independently from our mother and father.

4.38 Blood types of children. Emily and Michael both have alleles O and O.

(a) What blood types can their children have?

(b) What is the probability that their next child has each of these blood types?

4.39 Parents with alleles B and O. Andreona and Caleb both have alleles B and O.

(a) What blood types can their children have?

(b) What is the probability that their next child has each of these blood types?

4.40 Two children. Samantha has alleles B and O. Dylan has alleles A and B. They have two children. What is the probability that both children have blood type A? What is the probability that both children have the same blood type?

4.41 Three children. Anna has alleles B and O. Nathan has alleles A and O.

(a) What is the probability that a child of these parents has blood type O?

(b) If Anna and Nathan have three children, what is the probability that all three have blood type O? What is the probability that the first child has blood type O and the next two do not?

4.3 Random Variables

When you complete this section, you will be able to:

- Describe the probability distribution of a discrete random variable.
- Use a probability histogram to provide a graphical description of the probability distribution of a discrete random variable.
- Use the distribution of a discrete random variable to calculate probabilities of events.
- Find probabilities of events for the uniform and normal distributions.

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 process.

In our earlier coin-tossing example, the process is the tossing of a coin four times. 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.[9] As we progress from general rules of probability toward statistical inference, we will concentrate on random variables.

When a random variable X describes a random process, 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 will now attach a name to random variables having probability assigned in this way.[10]

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	$\cdots$
Probability	p_1	p_2	p_3	$\cdots$

The probabilities p_i must satisfy two requirements:

1. Every probability p_i is a number between 0 and 1.
2. $p_1 + p_2 + \cdots = 1$.

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 discrete random variable situations that we will study, the number of possible values is a finite number, k. For example, in our example on the number of heads in four tosses of a coin, there are $k = 5$ possible values: 0, 1, 2, 3, and 4.

There are, however, settings in which the number of possible values can be infinite. Think about tossing a fair coin until you get a head. The number of possible tosses is any positive integer.

EXAMPLE 4.22

Grade distributions. A liberal arts college posts the grade distributions for its courses. In a recent semester, students in one section of English 130 received 32% A's, 42% B's, 19% C's, 3% D's, and 4% F's. Choose an English 130 student at random. To "choose at random" means to give every student the same chance to be chosen. The student's grade on a five-point scale (with A = 4) is a random variable X.

The value of X changes when we repeatedly choose students at random, but it is always one of 0, 1, 2, 3, or 4. Here is the distribution of X:

Value of X	0	1	2	3	4
Probability	0.04	0.03	0.19	0.42	0.32

The probability that the student got a B or better is the sum of the probabilities of an A and a B. In the language of random variables,

$$\begin{aligned} P(X \geq 3) &= P(X = 3) + P(X = 4) \\ &= 0.42 + 0.32 = 0.74 \end{aligned}$$

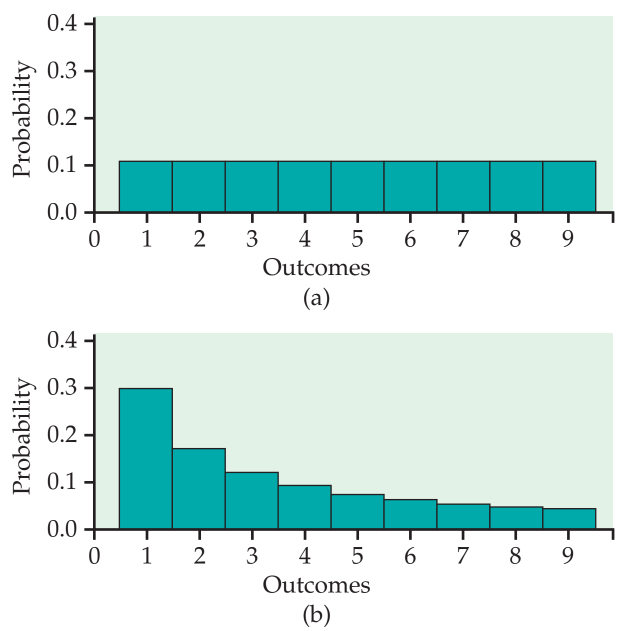
USE YOUR KNOWLEDGE

4.42 Will the course satisfy the requirement? Refer to Example 4.22. Suppose that a grade of D or F in English 130 does not satisfy a requirement for a major in linguistics. What is the probability that a randomly selected student will not satisfy this requirement?

probability histogram

We can use histograms to show probability distributions as well as distributions of data. Figure 4.5 displays **probability histograms** that compare the probability model for equally likely random digits (Example 4.15, page 227) with the model given by Benford's law (Example 4.12, page 226). The height of each bar shows the probability of the outcome at its base. Because the heights

FIGURE 4.5 Probability histograms for (a) equally likely random digits 1 to 9 and (b) Benford's law. The height of each bar shows the probability assigned to a single outcome.



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. Think of these histograms as idealized pictures of the results of very many trials. The histograms make it easy to quickly compare the two distributions.

EXAMPLE 4.23

Number of heads in four tosses of a coin. What is the probability distribution of the discrete random variable X that counts the number of heads in four tosses of a coin? We can derive this distribution if we make two reasonable assumptions:

- The coin is balanced, so it is fair and each toss is equally likely to give H or T.
- The coin has no memory, so tosses are independent.

The outcome of four tosses is a sequence of heads and tails such as HTTH. There are 16 possible outcomes in all. Figure 4.6 lists these outcomes along with the value of X for each outcome. The multiplication rule for independent events tells us that, for example,

$$P(\text{HTTH}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$$

$X = 0$	$X = 1$	$X = 2$	$X = 3$	$X = 4$
TTTT	HTTT	HTTH	HHHT	HHHH
	THTT	HTHT	HHTH	
	TTHT	THTH	HTHH	
	TTTH	HHTT	THHH	
		THHT		
		TTHH		

FIGURE 4.6 Possible outcomes in four tosses of a coin, Example 4.23. The outcomes are arranged by the values of the random variable X, the number of heads.

Each of the 16 possible outcomes similarly has probability 1/16. That is, these outcomes are equally likely.

The number of heads X has possible values 0, 1, 2, 3, and 4. These values are *not* equally likely. As Figure 4.6 shows, there is only one way that $X = 0$ can occur: namely, when the outcome is TTTT. So

$$P(X = 0) = \frac{1}{16} = 0.0625$$

The event $\{X = 2\}$ can occur in six different ways, so that

$$P(X = 2) = \frac{\text{count of ways } X = 2 \text{ can occur}}{16}$$

$$= \frac{6}{16} = 0.375$$

We can find the probability of each value of X from Figure 4.6 in the same way. Here is the result:

Value of X	0	1	2	3	4
Probability	0.0625	0.25	0.375	0.25	0.0625

AU/DE/PE: xref Example 4.23 on Fig. 4.7 cap?

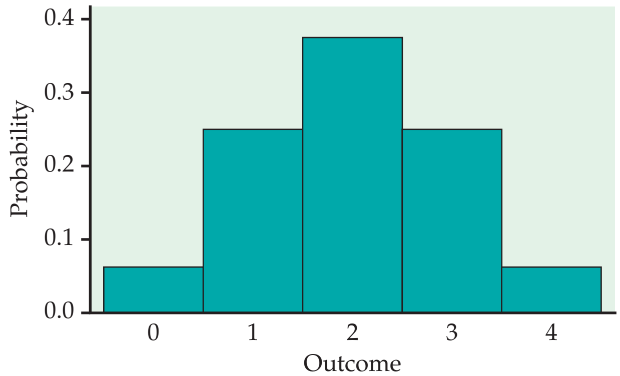


FIGURE 4.7 Probability histogram for the number of heads in four tosses of a coin.

Figure 4.7 is a probability histogram for the distribution in Example 4.23. The probability distribution is exactly symmetric. The probabilities (bar heights) are idealizations of the proportions after very many tosses of four coins. The actual distribution of proportions observed would be nearly symmetric but is unlikely to be exactly symmetric.

EXAMPLE 4.24

Probability of at least three heads. Any event involving the number of heads observed can be expressed in terms of X, and its probability can be found from the distribution of X. For example, the probability of tossing at least three heads is

$$P(X \geq 3) = 0.25 + 0.0625 = 0.3125$$

The probability of at least one head is most simply found by use of the complement rule:

$$\begin{aligned} P(X \geq 1) &= 1 - P(X = 0) \\ &= 1 - 0.0625 = 0.9375 \end{aligned}$$

Recall that tossing a coin n times is similar to choosing an SRS of size n from a large population and asking a Yes or No question. We will extend the results of Example 4.23 when we return to sampling distributions in the next chapter.

USE YOUR KNOWLEDGE

4.43 Two tosses of a fair coin. Find the probability distribution for the number of heads that appear in two tosses of a fair coin.

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.

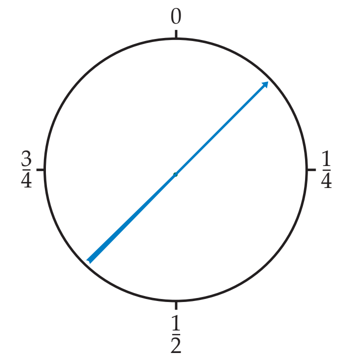


FIGURE 4.8 A spinner that generates a random number between 0 and 1.

You can visualize such a random number by thinking of a spinner (Figure 4.8) 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 entire interval of numbers:

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

LOOK BACK

density curve, p. 51

How can we assign probabilities to events such as $\{0.3 \le x \le 0.7\}$? 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 too many possible values. Instead, we 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 4.25

uniform distribution

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**.

This density curve appears in red in Figure 4.9. 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 square with base 1 and height 1. The probability of any event is the area under the density curve and above the event in question.

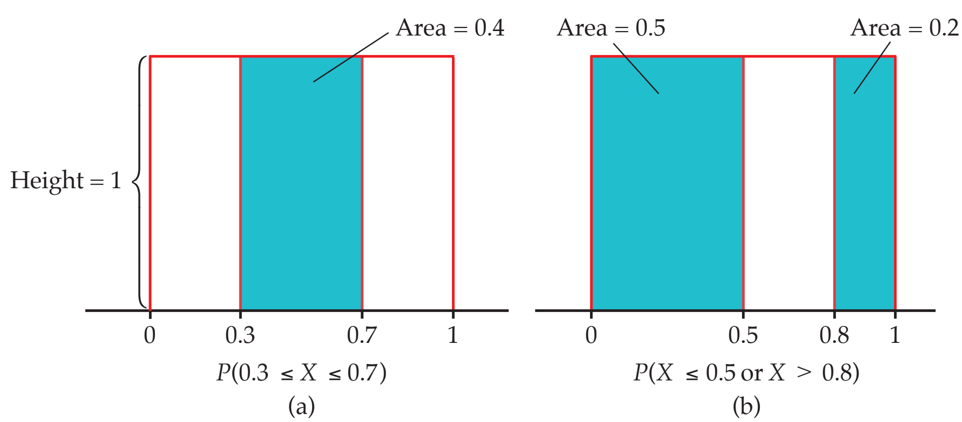


FIGURE 4.9 Assigning probabilities for generating a random number between 0 and 1, Example 4.25. The probability of any interval of numbers is the area above the interval and under the density curve.

As Figure 4.9(a) illustrates, the probability that the random number generator produces a number X between 0.3 and 0.7 is

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

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,

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

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 4.9(b). This assignment of probabilities obeys all of our rules for probability.

USE YOUR KNOWLEDGE

4.44 Find the probability. For the uniform distribution described in Example 4.25, find the probability that X is between 0.3 and 0.9.

Probability as area under a density curve is a second important way of assigning probabilities to events. Figure 4.10 illustrates this idea in general form. We call X in Example 4.25 a *continuous random variable* because its values are not isolated numbers but an entire interval of numbers.

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,

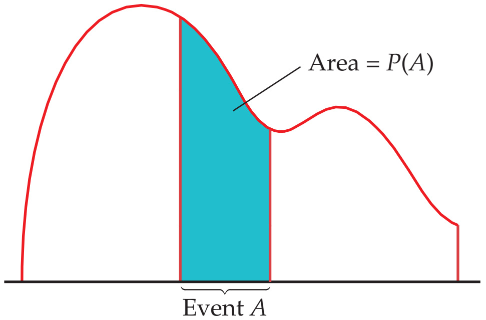


FIGURE 4.10 The probability distribution of a continuous random variable assigns probabilities as areas under a density curve. The total area under any density curve is 1.

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 4.25. 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.

CAUTION

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. *We can ignore the distinction between $>$ and $\geq$ when finding probabilities for continuous (but not discrete) random variables.*

Normal distributions as probability distributions

LOOK BACK

Normal distributions, p. 56

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 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 $N(0,1)$.

EXAMPLE 4.26

Texting while driving. Texting while driving can be dangerous, but young people want to remain connected. Suppose that 26% of teen drivers text while driving. If we take a sample of 500 teen drivers, what percent would we expect to say that they text while driving?[11]

The proportion $p = 0.26$ is a number that describes the population of teen drivers. The proportion $\hat{p}$ of the sample who say that they text while driving is used to estimate p. The proportion $\hat{p}$ is a random variable because repeating the SRS would give a different sample of 500 teen drivers and a different value of $\hat{p}$.

We will see in the next chapter that in this setting, with teen drivers answering honestly, $\hat{p}$ has approximately the $N(0.26, 0.0196)$ distribution. The mean 0.26 of this distribution is the same as the population proportion because $\hat{p}$ is an unbiased estimate of p. The standard deviation is controlled mainly by the size of the sample.

What is the probability that the survey result differs from the truth about the population by no more than 3 percentage points? We can use what we learned about Normal distribution calculations to answer this question. Because $p = 0.26$, the survey misses by no more than 3 percentage points if the sample proportion is between 0.23 and 0.29.

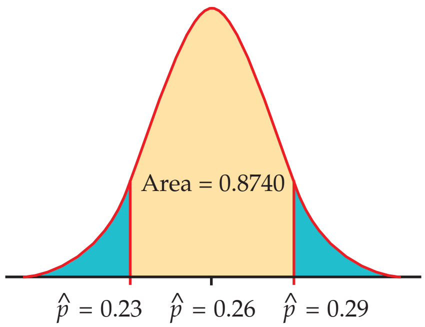


FIGURE 4.11 Probability as area under a Normal density curve, Example 4.26.

Figure 4.11 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,

$$P(0.23 \le \hat{p} \le 0.29) = P\left(\frac{0.23 - 0.26}{0.0196} \le \frac{\hat{p} - 0.26}{0.0196} \le \frac{0.29 - 0.26}{0.0196}\right)$$
$$= P(-1.53 \le Z \le 1.53)$$
$$= 0.9370 - 0.0630 = 0.8740$$

About 87% of the time, the sample $\hat{p}$ will be within 3 percentage points of the proportion p.

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 will employ only these two types of probability models.

SECTION 4.3 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 exactly 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.
- **Uniform distributions** are continuous probability distributions that are very similar to equally likely discrete distributions.
- **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 4.3 EXERCISES

For Exercise 4.42, see page 237; for Exercise 4.43, see page 239; and for Exercise 4.44, see page 241.

4.45 How many courses? At a small liberal arts college, students can register for one to six courses. Let X be the number of courses taken in the fall by a randomly selected student from this college. 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.

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

4.47 Find some probabilities. Refer to Exercise 4.45.

(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.

4.48 Use the uniform distribution. Suppose that a random variable X follows the uniform distribution described in Example 4.25 (page 240). 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 4.9 (page 240).

(a) The probability that X is less than 0.2.

(b) The probability that X is greater than or equal to 0.7.

(c) The probability that X is less than 0.8 and greater than 0.4.

(d) The probability that X is 0.7.

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) The possible values for a discrete random variable can't be negative.

(b) A continuous random variable can take any value between 0 and 1.

(c) Normal distributions are discrete random variables.

4.50 Use of Twitter. Suppose that the population proportion of Internet users who say that they use Twitter or another service to post updates about themselves or to see updates about others is 19%.[12] Think about selecting random samples from a population in which 19% are Twitter users.

(a) Describe the sample space for selecting a single person.

(b) If you select three people, describe the sample space.

(c) Using the results of part (b), define the sample space for the random variable that expresses the number of Twitter users in the sample of size 3.

(d) What information is contained in the sample space for part (b) that is not contained in the sample space for part (c)? Do you think this information is important? Explain your answer.

4.51 Use of Twitter. Refer to the previous exercise. Find the probabilities for the number of Twitter users in a sample of size 2.

4.52 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	0.44	0.22	0.20	0.09	0.03	0.02

Make probability histograms for these two discrete distributions, using the same scales. What are the most important differences between the sizes of households and families?

4.53 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) You record the number of hits per day on your web page.

(c) You record the yearly income of a visitor to your web page.

4.54 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.

4.55 Tossing two dice. Some games of chance rely on tossing two dice. Each die has six faces, marked with one, two, . . . , six spots called pips. The dice used in casinos are carefully balanced so that each face is equally likely to come up. When two dice are tossed, each of the 36 possible pairs of faces is equally likely to come up. The outcome of interest to a gambler is the sum of the pips on the two up-faces. Call this random variable X.

(a) Write down all 36 possible pairs of up-faces.

(b) If all pairs have the same probability, what must be the probability of each pair?

(c) Write the value of X next to each pair of up-faces and use this information with the result of part (b) to give the probability distribution of X. Draw a probability histogram to display the distribution.

(d) One bet available in the game called craps wins if a 7 or an 11 comes up on the next roll of two dice. What is the probability of rolling a 7 or an 11 on the next roll?

(e) Several bets in craps lose if a 7 is rolled. If any outcome other than 7 occurs, these bets either win or continue to the next roll. What is the probability that anything other than a 7 is rolled?

4.56 Nonstandard dice. Nonstandard dice can produce interesting distributions of outcomes. You have two balanced, six-sided dice. One is a standard die, with faces having one, two, three, four, five, and six spots. The other die has three faces with one spot and three faces with six spots. Find the probability distribution for the total number of spots Y on the up-faces when you roll these two dice.

4.57 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 $X \leq 3$ in words. What is its probability? What is the probability that $X < 3$?

4.58 Find the probabilities. Let the random variable X be a random number with the uniform density curve in Figure 4.9 (page 240). Find the following probabilities:

(a) $P(X \geq 0.35)$.

(b) $P(X = 0.35)$.

(c) $P(0.35 < X < 1.35)$.

(d) $P(0.18 \leq X \leq 0.25 \text{ or } 0.4 \leq X \leq 0.5)$.

(e) X is not in the interval 0.4 to 0.8.

4.59 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 $P(Y \geq 0.95)$.

4.60 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 4.12.

(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 1.5?

(d) What is the probability that Y is greater than 0.5?

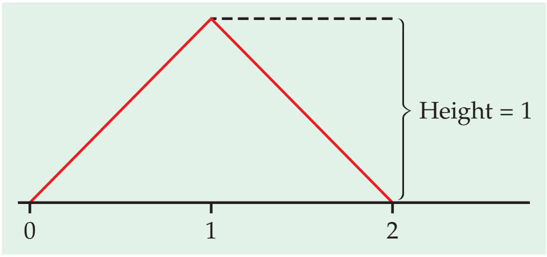


FIGURE 4.12 The density curve for the sum Y of two random numbers, Exercise 4.60.

4.61 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 will see in the next chapter 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 $P(8 \le \bar{x} \le 10)$, the probability that $\bar{x}$ estimates μ to within ± 1?

4.62 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 will see in the next chapter that, in this situation, the proportion $\hat{p}$ of the sample 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 $\hat{p}$ estimates 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.

4.4 Means and Variances of Random Variables

When you complete this section, you will be able to:

- Use a probability distribution to find the mean of a discrete random variable.
- Apply the law of large numbers to describe the behavior of the sample mean as the sample size increases.
- Find means using the rules for means of linear transformations, sums, and differences.
- Use a probability distribution to find the variance and the standard deviation of a discrete random variable.
- Find variances and standard deviations using the rules for variances and standard deviations for linear transformations.
- Find variances and standard deviations using the rules for variances and standard deviations for sums of and differences between two random variables and for uncorrelated and for correlated 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 will 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 will 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 28), 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. You can think of the value of a random variable as the average of a very large sample where the relative frequencies of the values are the same as their probabilities.

If we think of the random process as corresponding to the population, then the mean of the random variable is a characteristic of this population. Here is an example.

EXAMPLE 4.27

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

AU: Please check x-ref

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 191), we can view the random process as selecting an SRS of size 1 from the population. The random variable X is 1 if the selected number is equal to the one that you chose and 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 4.27 is a different quantity—it is the long-run average winnings you expect if you play a very large number of times.

USE YOUR KNOWLEDGE

4.63 Find the mean of the probability distribution. You toss a fair coin. If the outcome is heads, you win \$10.00; 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 mean μ

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.

To remind ourselves that we are talking about the mean of X, we often write μ_X rather than simply μ. In Example 4.27, $\mu_X = \$0.50$. 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 one observation on X to be close to its expected value.

expected value

The mean of any discrete random variable is found just as in Example 4.27. It is 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:

$$\begin{aligned}\mu_X &= x_1p_1 + x_2p_2 + \cdots \\ &= \sum x_ip_i\end{aligned}$$

EXAMPLE 4.28

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

$$\begin{aligned}\mu_X = 1 \times \frac{1}{9} + 2 \times \frac{1}{9} + 3 \times \frac{1}{9} + 4 \times \frac{1}{9} + 5 \times \frac{1}{9} \\ + 6 \times \frac{1}{9} + 7 \times \frac{1}{9} + 8 \times \frac{1}{9} + 9 \times \frac{1}{9} \\ = 45 \times \frac{1}{9} = 5\end{aligned}$$

Suppose that the random digits in Example 4.28 had a different probability distribution. In Example 4.12 (page 226), 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 4.29

The mean of first digits that follow Benford's law. 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 4.28. 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

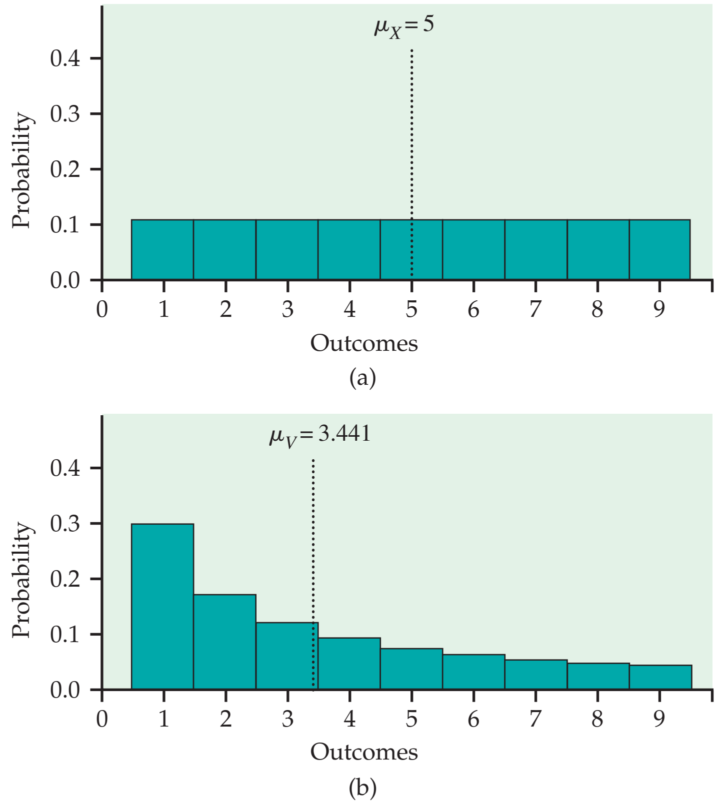
$$\begin{aligned}\mu_V &= (1)(0.301) + (2)(0.176) + (3)(0.125) + (4)(0.097) + (5)(0.079) \\ &\quad + (6)(0.067) + (7)(0.058) + (8)(0.051) + (9)(0.046) \\ &= 3.441\end{aligned}$$

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 4.13 locates the means of X and V on the two probability histograms. Because the discrete uniform distribution of Figure 4.13(a) is symmetric, the mean lies at the center of symmetry. We can't locate the mean of the right-skewed distribution of Figure 4.13(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 (page 54) showed how to find the mean of the distribution: it is the point at

FIGURE 4.13 Locating the mean of a discrete random variable on the probability histogram for (a) digits between 1 and 9 chosen at random; (b) digits between 1 and 9 chosen from records that obey Benford's law.



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.[13] 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.

Statistical estimation and the law of large numbers

AU/DE/PE: there was a missing MN item (see ms. 4-75):
[Look Back icon]
sampling distributions, p. xxx
This already appears on p. 250. Is it also needed here?

We would like to estimate the mean height μ of the population of all American women between the ages of 18 and 24 years. This μ is the mean μ_X of the random variable X obtained by choosing a young woman at random and measuring her height. To estimate μ, we choose an SRS of young women and use the sample mean $\bar{x}$ to estimate the unknown population mean μ. In the language of Section 5.1 (page 282), μ is a *parameter* and $\bar{x}$ is a *statistic*.

Statistics obtained from probability samples are random variables because their values vary in repeated sampling. The sampling distributions of statistics are just the probability distributions of these random variables.

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 $\bar{x}$.

If $\bar{x}$ is rarely exactly right and varies from sample to sample, why is it nonetheless a reasonable estimate of the population mean μ? 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 can afford to keep on measuring more women, eventually we will estimate the mean height of all young women 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. Decide how accurately you would like to estimate μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close.

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 216) shows how proportions approach probability in one example. Here is an example of how sample means approach the distribution mean.

EXAMPLE 4.30

Heights of young women. The distribution of the heights of all young women is close to the Normal distribution with mean 64.5 inches and standard deviation 2.5 inches. Suppose that $\mu = 64.5$ were exactly true.

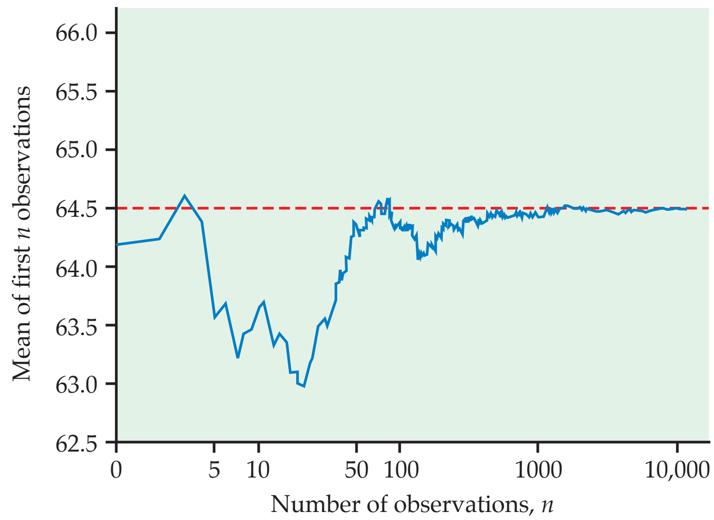


FIGURE 4.14 The law of large numbers in action, Example 4.30. As we take more observations, the sample mean always approaches the mean of the population.

Figure 4.14 shows the behavior of the mean height $\bar{x}$ of n women chosen at random from a population whose heights follow the $N(64.5, 2.5)$ distribution. The graph plots the values of $\bar{x}$ as we add women to our sample. The first woman drawn had height 64.21 inches, so the line starts there. The second had height 64.35 inches, so for $n = 2$ the mean is

$$\bar{x} = \frac{64.21 + 64.35}{2} = 64.28$$

This is the second point on the line in the graph.

At first, the graph shows that the mean of the sample changes as we take more observations. Eventually, however, the mean of the observations gets close to the population mean $\mu = 64.5$ and settles down at that value. The law of large numbers says that this *always* happens.

USE YOUR KNOWLEDGE

4.64 Use the *Law of Large Numbers* applet. The *Law of Large Numbers* applet animates a graph like Figure 4.14 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.[14] 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.[15] 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 4.31

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.[16]

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.

BEYOND THE BASICS

More Laws of Large Numbers

The law of large numbers is one of the central facts about probability. It helps us understand the mean μ of a random variable. It explains why gambling casinos and insurance companies make money. It assures us that statistical estimation will be accurate if we can afford enough observations. The basic law of large numbers applies to independent observations that all have the same distribution. Mathematicians have extended the law to many more general settings. Here are two of these.

Is there a winning system for gambling? Serious gamblers often follow a system of betting in which the amount bet on each play depends on the outcome of previous plays. You might, for example, double your bet on each spin of the roulette wheel until you win—or, of course, until your fortune is exhausted. Such a system tries to take advantage of the fact that you have a memory even though the roulette wheel does not. Can you beat the odds with a system based on the outcomes of past plays? No. Mathematicians have established a stronger version of the law of large numbers that says that, if you do not have an infinite fortune to gamble with, your long-run average winnings μ remain the same as long as successive trials of the game (such as spins of the roulette wheel) are independent.

What if observations are not independent? You are in charge of a process that manufactures video screens for computer monitors. Your equipment measures the tension on the metal mesh that lies behind each screen and is critical to its image quality. You want to estimate the mean tension μ for the process by the average $\bar{x}$ of the measurements. Alas, the tension measurements are not independent. If the tension on one screen is a bit too high, the tension on the next is more likely to also be high. Many real-world processes are like this—the process stays stable in the long run, but two observations made close together are likely to both be above or both be below the long-run mean. Again the mathematicians come to the rescue: as long as the dependence dies out fast enough as we take measurements farther and farther apart in time, the law of large numbers still holds.

Rules for means

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 $\mu_X = 1.4$. (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. The crickets living in a field have mean length of 1.2 inches. What is the mean in centimeters? There are 2.54 centimeters in an inch, so the length of a cricket in centimeters is 2.54 times its length in inches. If we multiply every observation by 2.54, we also multiply their average by 2.54. The mean in centimeters must be 2.54×1.2, or about 3.05 centimeters. More formally, the length in inches of a cricket chosen at random from the field is a random variable X with mean μ_X. The length 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 OF LINEAR TRANSFORMATIONS, SUMS, AND DIFFERENCES

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

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

LOOK BACK
linear transformation, p. 44

Note that $a + bX$ is a linear transformation of the random variable X.

EXAMPLE 4.32


skynesher/iStockphoto

How many courses? In Exercise 4.45 (page 244) you described the probability distribution of the number of courses taken in the fall by students at a small liberal arts college. Here is the distribution:

Courses in the fall	1	2	3	4	5	6
Probability	0.05	0.05	0.13	0.26	0.36	0.15

For the spring semester, the distribution is a little different.

Courses in the spring	1	2	3	4	5	6
Probability	0.06	0.08	0.15	0.25	0.34	0.12

For a randomly selected student, let X be the number of courses taken in the fall semester, and let Y be the number of courses taken in the spring semester. The means of these random variables are

$$\mu_X = (1)(0.05) + (2)(0.05) + (3)(0.13) + (4)(0.26) + (5)(0.36) + (6)(0.15) = 4.28$$

$$\mu_Y = (1)(0.06) + (2)(0.08) + (3)(0.15) + (4)(0.25) + (5)(0.34) + (6)(0.12) = 4.09$$

The mean course load for the fall is 4.28 courses, and the mean course load for the spring is 4.09 courses. We assume that these distributions apply to students who earned credit for courses taken in the fall and the spring semesters. The mean of the total number of courses taken for the academic year is $X + Y$. Using Rule 2, we calculate the mean of the total number of courses:

$$\begin{aligned}\mu_Z &= \mu_X + \mu_Y \\ &= 4.28 + 4.09 = 8.37\end{aligned}$$

Note that it is not possible for a student to take 8.37 courses in an academic year. This number is the mean of the probability distribution.

EXAMPLE 4.33

What about credit hours? In the previous exercise, we examined the number of courses taken in the fall and in the spring at a small liberal arts college. Suppose that we were interested in the total number of credit hours earned for the academic year. We assume that for each course taken at this college, three credit hours are earned. Let T be the mean of the distribution of the total number of credit hours earned for the academic year. What is the mean of the distribution of T? To find the answer, we can use Rule 1 with $a = 0$ and $b = 3$. Here is the calculation:

$$\begin{aligned}\mu_T &= \mu_{a+bZ} \\ &= a + b\mu_Z \\ &= 0 + (3)(8.37) = 25.11\end{aligned}$$

The mean of the distribution of the total number of credit hours earned is 25.11.

USE YOUR KNOWLEDGE

4.65 Find μ_Y. The random variable X has mean $\mu_X = 12$. If $Y = 12 + 6X$, what is μ_Y?

4.66 Find μ_W. The random variable U has mean $\mu_U = 25$, and the random variable V has mean $\mu_V = 25$. If $W = 0.5U + 0.5V$, find μ_W.

The variance of a random variable

The mean is a measure of the center of a distribution. A basic numerical description requires, in addition, a measure of the spread or variability of the distribution. The variance and the standard deviation are the 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

AU: page xref to 38 here? Also note, this is identified in Ch. 1 as "sample deviation" rather than "sample variance" from its mean μ_X. As for the mean, the average we use is a weighted average in which each outcome is weighted by its probability in order to take account of outcomes that are not equally likely. 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

$$\begin{aligned}\sigma_X^2 &= (x_1 - \mu_X)^2 p_1 + (x_2 - \mu_X)^2 p_2 + \cdots \\ &= \sum (x_i - \mu_X)^2 p_i\end{aligned}$$

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

EXAMPLE 4.34

Find the mean and the variance. In Example 4.32 (pages 254–255) , we saw that the distribution of the number X of fall courses taken by students at a small liberal arts college is

Courses in the fall	1	2	3	4	5	6
Probability	0.05	0.05	0.13	0.26	0.36	0.15

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	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
1	0.05	0.05	$(1 - 4.28)^2(0.05) = 0.53792$
2	0.05	0.10	$(2 - 4.28)^2(0.05) = 0.25992$
3	0.13	0.39	$(3 - 4.28)^2(0.13) = 0.21299$
4	0.26	1.04	$(4 - 4.28)^2(0.26) = 0.02038$
5	0.36	1.80	$(5 - 4.28)^2(0.36) = 0.18662$
6	0.15	0.90	$(6 - 4.28)^2(0.15) = 0.44376$
		$\mu_X = 4.28$	$\sigma_X^2 = 1.662$

We see that $\sigma_X^2 = 1.662$. The standard deviation of X is $\sigma_X = \sqrt{1.662} = 1.289$. The standard deviation is a measure of the variability of the number of fall courses taken by the students at the small liberal arts college. As in the case of distributions for data, the standard deviation of a probability distribution is easiest to understand for Normal distributions.

USE YOUR KNOWLEDGE

4.67 Find the variance and the standard deviation. The random variable X has the following probability distribution:

Value of X	0	3
Probability	0.3	0.7

Find the variance σ_X^2 and the standard deviation σ_X for this random variable.

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. Two random variables

independence

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 variables describe outcomes that appear unrelated to each other. You should ask in each instance whether the assumption of independence seems reasonable.

When random variables are not independent, the variance of their sum

correlation

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 (page 101) the correlation r 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. 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.

CAUTION

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 combine standard deviations, use the rules for variances.* 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 4.35

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

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 4.36

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 4.37

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

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

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 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 is an example.

EXAMPLE 4.38

Utility bills. Consider a household where the monthly bill for natural-gas averages \$125 with a standard deviation of \$75, while the monthly bill for electricity averages \$174 with a standard deviation of \$41. The correlation between the two bills is -0.55.

Let's compute the mean and standard deviation of the sum of the natural-gas bill and the electricity bill. We let X stand for the natural-gas bill and Y stand for the electricity bill. Then the total is $X + Y$. Using the rules for means, we have

$$\mu_{X+Y} = \mu_X + \mu_Y = 125 + 174 = 299$$

To find the standard deviation, we first find the variance and then take the square root to determine the standard deviation. From the general addition rule for variances of random variables,

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y$$
$$= (75)^2 + (41)^2 + (2)(-0.55)(75)(41)$$
$$= 3923.5$$

Therefore, the standard deviation is

$$\sigma_{X+Y} = \sqrt{3923.5} = 63$$

The total of the natural-gas bill and the electricity bill has mean \$299 and standard deviation \$63.

The negative correlation in Example 4.38 is due to the fact that, in this household, natural gas is used for heating and electricity is used for air-conditioning. So, when it is warm, the electricity charges are high and the natural-gas charges are low. When it is cool, the reverse is true. This causes the standard deviation of the sum to be less than it would be if the two bills were uncorrelated (see Exercise 4.79, page 263).

There are situations where we need to combine several of our rules to find means and standard deviations. Here is an example.

EXAMPLE 4.39

Calcium intake. To get enough calcium for optimal bone health, tablets containing calcium are often recommended to supplement the calcium in the diet. One study designed to evaluate the effectiveness of a supplement followed a group of young people for seven years. Each subject was assigned to take either a tablet containing 1000 milligrams of calcium per day (mg/d) or a placebo tablet that was identical except that it had no calcium.[17] A major problem with studies like this one is compliance: subjects do not always take the treatments assigned to them.

In this study, the compliance rate declined to about 47% toward the end of the seven-year period. The standard deviation of compliance was 22%. Calcium from the diet averaged 850 mg/d with a standard deviation of 330 mg/d. The correlation between compliance and dietary intake was 0.68. Let's find the mean and standard deviation for the total calcium intake. We let S stand for the intake from the supplement and D stand for the intake from the diet.

We start with the intake from the supplement. Because the compliance is 47% and the amount in each tablet is 1000 mg, the mean for S is

$$\mu_S = 1000(0.47) = 470$$

Because the standard deviation of the compliance is 22%, the variance of S is

$$\sigma^2_S = 1000^2(0.22)^2 = 48{,}400$$

The standard deviation is

$$\sigma_S = \sqrt{48{,}400} = 220$$

Be sure to verify which rules for means and variances are used in these calculations.

We can now find the mean and standard deviation for the total intake. The mean is

$$\mu_{S+D} = \mu_S + \mu_D = 470 + 850 = 1320$$

the variance is

$$\sigma^2_{S+D} = \sigma^2_S + \sigma^2_D + 2\rho\sigma_S\sigma_D = (220)^2 + (330)^2 + 2(0.68)(220)(330) = 256{,}036$$

and the standard deviation is

$$\sigma_{S+D} = \sqrt{256{,}036} = 506$$

The mean of the total calcium intake is 1320 mg/d and the standard deviation is 506 mg/d.

The correlation in this example illustrates an unfortunate fact about compliance and having an adequate diet. Some of the subjects in this study have diets that provide an adequate amount of calcium while others do not. The positive correlation between compliance and dietary intake tells us that those who have relatively high dietary intakes are more likely to take the assigned supplements. On the other hand, those subjects with relatively low dietary intakes, the ones who need the supplement the most, are less likely to take the assigned supplements.

SECTION 4.4 SUMMARY

- The probability distribution of a random variable X, like a distribution of data, has a **mean μ_X** and a **standard deviation σ_X**.

- The **law of large numbers** says that the average of the values of X observed in many trials must approach μ.

- The **mean μ** 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 σ_X^2** is the average squared deviation of the values of the variable from their mean. For a discrete random variable,

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

- The **standard deviation σ_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

$$\mu_{a+bX} = a + b\mu_X$$
$$\sigma_{a+bX}^2 = b^2\sigma_X^2$$

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

$$\mu_{X+Y} = \mu_X + \mu_Y$$
$$\mu_{X-Y} = \mu_X - \mu_Y$$
$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2 + 2\rho\sigma_X\sigma_Y$$
$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2 - 2\rho\sigma_X\sigma_Y$$

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

$$\sigma_{X+Y}^2 = \sigma_X^2 + \sigma_Y^2$$
$$\sigma_{X-Y}^2 = \sigma_X^2 + \sigma_Y^2$$

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

SECTION 4.4 EXERCISES

For Exercise 4.63, see page 247; for Exercise 4.64, see page 251; for Exercises 4.65 and 4.66, see page 255; and for Exercise 4.67, see page 257.

4.68 Find the mean of the random variable. A random variable X has the following distribution.

X	−1	0	1	2
Probability	0.2	0.3	0.2	0.3

Find the mean for this random variable. Show your work.

4.69 Explain what happens when the sample size gets large. Consider the following scenarios: (1) You take a sample of two observations on a random variable and compute the sample mean, (2) you take a sample of 100 observations on the same random variable and compute the sample mean, (3) you take a sample of 1000 observations on the same random variable and compute the sample mean. Explain in simple language how close you expect the sample mean to be to the mean of the random variable as you move from Scenario 1 to Scenario 2 to Scenario 3.

4.70 Find some means. Suppose that X is a random variable with mean 30 and standard deviation 4. Also suppose that Y is a random variable with mean 50 and standard deviation 8. Find the mean of the random variable Z for each of the following cases. Be sure to show your work.

(a) $Z = 35 - 10X$.

(b) $Z = 12X - 5$.

(c) $Z = X + Y$.

(d) $Z = X - Y$.

(e) $Z = -2X + 2Y$.

4.71 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	0.3	0.2

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

4.72 Find some variances and standard deviations. Suppose that X is a random variable with mean 30 and standard deviation 4. Also suppose that Y is a random variable with mean 50 and standard deviation 8. Assume that the correlation between X and Y is zero. 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) $Z = 35 - 10X$.

(b) $Z = 12X - 5$.

(c) $Z = X + Y$.

(d) $Z = X - Y$.

(e) $Z = -2X + 2Y$.

4.73 What happens if the correlation is not zero? Suppose that X is a random variable with mean 30 and standard deviation 4. Also suppose that Y is a random variable with mean 50 and standard deviation 8. Assume that the correlation between X and Y is 0.5. Find the mean of the random variable Z for each of the following cases. Be sure to show your work.

(a) $Z = 35 - 10X$.

(b) $Z = 12X - 5$.

(c) $Z = X + Y$.

(d) $Z = X - Y$.

(e) $Z = -2X + 2Y$.

4.74 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.

4.75 Servings of fruits and vegetables. The following table gives the distribution of the number of servings of fruits and vegetables consumed per day in a population.

Number of servings X	0	1	2	3	4	5
Probability	0.3	0.1	0.1	0.2	0.2	0.1

Find the mean for this random variable.

4.76 Mean of the distribution for the number of aces. In Exercise 4.54 (page 244) 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.

4.77 Standard deviation of the number of aces. Refer to Exercise 4.76. Find the standard deviation of the number of aces.

4.78 Standard deviation for fruits and vegetables. Refer to Exercise 4.75. Find the variance and the standard deviation for the distribution of the number of servings of fruits and vegetables.

4.79 Suppose that the correlation is zero. Refer to Example 4.38 (page 259).

(a) Recompute the standard deviation for the total of the natural-gas bill and the electricity bill, assuming that the correlation is zero.

(b) Is this standard deviation larger or smaller than the standard deviation computed in Example 4.38? Explain why.

4.80 Find the mean of the sum. Figure 4.12 (page 245) displays the density curve of the sum $Y = X_1 + X_2$ of two independent random numbers, each uniformly distributed between 0 and 1.

(a) The mean of a continuous random variable is the balance point of its density curve. Use this fact to find the mean of Y from Figure 4.12.

(b) Use the same fact to find the means of X_1 and X_2. (They have the density curve pictured in Figure 4.9, page 240.) Verify that the mean of Y is the sum of the mean of X_1 and the mean of X_2.

4.81 Calcium supplements and calcium in the diet. Refer to Example 4.39 (page 260). Suppose that people who have high intakes of calcium in their diets are more compliant than those who have low intakes. What effect would this have on the calculation of the standard deviation for the total calcium intake? Explain your answer.

4.82 Toss a four-sided die twice. Role-playing games like *Dungeons & Dragons* use many different types of dice. Suppose that a four-sided die has faces marked 1, 2, 3, and 4. The intelligence of a character is determined by rolling this die twice and adding 1 to the sum of the spots. The faces are equally likely, and the two rolls are independent. What is the average (mean) intelligence for such characters? How spread out are their intelligences, as measured by the standard deviation of the distribution?

4.83 Means and variances of sums. The rules for means and variances allow you to find the mean and variance of a sum of random variables without first finding the distribution of the sum, which is usually much harder to do.

(a) A single toss of a balanced coin has either 0 or 1 head, each with probability 1/2. What are the mean and standard deviation of the number of heads?

(b) Toss a coin four times. Use the rules for means and variances to find the mean and standard deviation of the total number of heads.

(c) Example 4.23 (page 238) finds the distribution of the number of heads in four tosses. Find the mean and standard deviation from this distribution. Your results in parts (b) and (c) should agree.

4.84 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 $\rho_{XY} = 1$.

4.85 Will you assume independence? In which of the following games of chance would you be willing to assume independence of X and Y in making a probability model? Explain your answer in each case.

(a) In blackjack, you are dealt two cards and examine the total points X on the cards (face cards count 10 points). You can choose to be dealt another card and compete based on the total points Y on all three cards.

(b) In craps, the betting is based on successive rolls of two dice. X is the sum of the faces on the first roll, and Y the sum of the faces on the next roll.

4.86 Transform the distribution of heights from centimeters to inches. A report of the National Center for Health Statistics says that the heights of 20-year-old men have mean 176.8 centimeters (cm) and standard deviation 7.2 cm. There are 2.54 centimeters in an inch. What are the mean and standard deviation in inches?

4.87 Fire insurance. An insurance company looks at the records for millions of homeowners and sees that the mean loss from fire in a year is $\mu = \$300$ per person. (Most of us have no loss, but a few lose their homes. The \$300 is the average loss.) The company plans to sell fire insurance for \$300 plus enough to cover its costs and profit. Explain clearly why it would be stupid to sell only 10 policies. Then explain why selling thousands of such policies is a safe business.

4.88 Mean and standard deviation for 10 and for 12 policies. In fact, the insurance company sees that in the entire population of homeowners, the mean loss from fire is $\mu = \$300$ and the standard deviation of the loss is $\sigma = \$400$. What are the mean and standard deviation of the average loss for 10 policies? (Losses on separate policies are independent.) What are the mean and standard deviation of the average loss for 12 policies?

4.5 General Probability Rules

When you complete this section, you will be able to:

- Apply the five rules of probability.
- Apply the general addition rule for unions of two or more events.
- Find conditional probabilities.
- Apply the multiplication rule.
- Use a tree diagram to find probabilities.
- Use Bayes's rule to find probabilities.
- Determine whether or not two events that both have positive probability are independent.

Our study of probability has concentrated on random variables and their distributions. Now we return to the laws that govern any assignment of probabilities. The purpose of learning more laws of probability is to be able to give probability models for more complex random phenomena. We have already met and used five rules.

PROBABILITY RULES

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

Rule 2. $P(S) = 1$

Rule 3. Addition rule: If A and B are **disjoint** events, 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 A and B are **independent** events, then

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

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 if the events are not disjoint? These circumstances are covered by more general addition rules for probability.

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

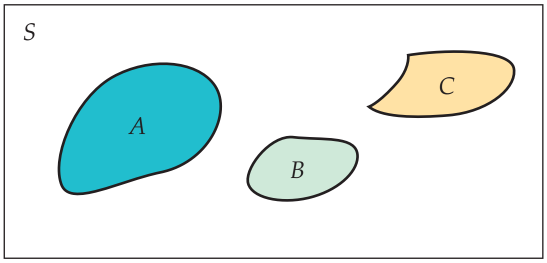


FIGURE 4.15 The addition rule for disjoint events: $P(A \text{ or } B \text{ or } C) = P(A) + P(B) + P(C)$ when events A, B, and C are disjoint.

Venn diagram in Figure 4.15 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

$$P(\text{one or more of } A,B,C) = P(A) + P(B) + P(C)$$

This rule extends to any number of disjoint events.

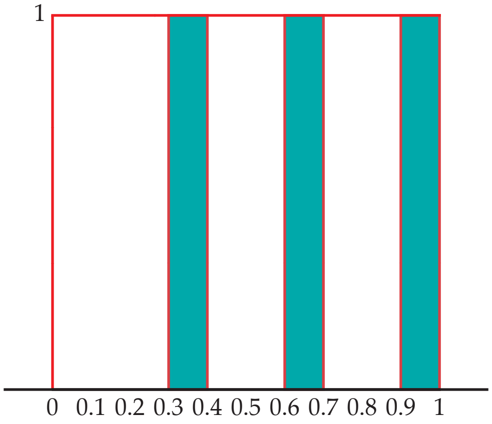
EXAMPLE 4.40

Probabilities as areas. Generate a random number X between 0 and 1. What is the probability that the first digit after the decimal point will be a 3, a 6, or a 9? The random number X is a continuous random variable whose density curve has constant height 1 between 0 and 1 and is 0 elsewhere. The event that the first digit of X is odd is the union of five disjoint events. These events are

$$0.30 \le X < 0.40$$
$$0.60 \le X < 0.70$$
$$0.90 \le X < 1.00$$

Figure 4.16 illustrates the probabilities of these events as areas under the density curve. Each area is 0.1. Therefore, the union of the three has probability equal to the sum, or 0.3.

FIGURE 4.16 The probability that the first digit after the decimal point of a random number is a 3, a 6, or a 9 is the sum of the probabilities of the three disjoint events shown, Example 4.40.



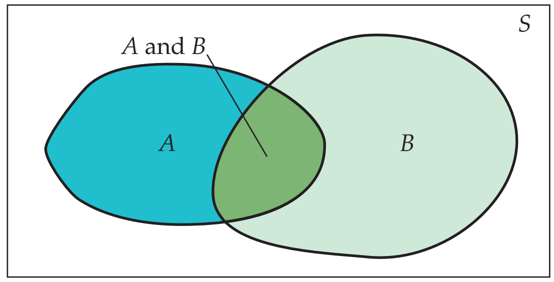


FIGURE 4.17 The union of two events that are not disjoint. The general addition rule says that $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$.

USE YOUR KNOWLEDGE

4.89 Probability that you roll a 3 or a 4 or a 5. If you roll a die, the probability of each of the six possible outcomes (1, 2, 3, 4, 5, 6) is 1/6. What is the probability that you roll a 3 or a 4 or a 5?

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 4.17 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.

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 4.41


© Randy Faris/Corbis

Adequate sleep and exercise. Suppose that 40% of adults get enough sleep and 46% exercise regularly. What is the probability that an adult gets enough sleep or exercises regularly? To find this probability, we also need to know the percent who get enough sleep and exercise. Let's assume that 24% do both.

We will use the notation of the general addition rule for unions of two events. Let A be the event that an adult gets enough sleep, and let B be the event that a person exercises regularly. We are given that $P(A) = 0.40$, $P(B) = 0.46$, and $P(A \text{ and } B) = 0.24$. Therefore,

$$\begin{aligned} P(A \text{ or } B) &= P(A) + P(B) - P(A \text{ and } B) \\ &= 0.40 + 0.46 - 0.24 \\ &= 0.62 \end{aligned}$$

The probability that an adult gets enough sleep or exercises regularly is 0.62, or 62%.

USE YOUR KNOWLEDGE

4.90 Probability that your roll is even or greater than 5. If you roll a die, the probability of each of the six possible outcomes (1, 2, 3, 4, 5, 6) is 1/6. What is the probability that your roll is even or greater than 5?

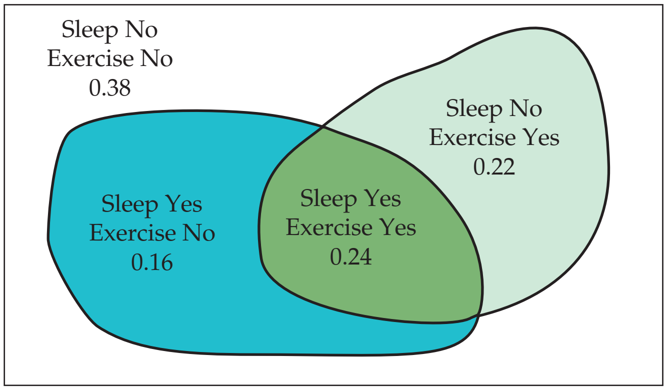


FIGURE 4.18 Venn diagram and probabilities, Example 4.41.

Venn diagrams are a great help in finding probabilities for unions because you can just think of adding and subtracting areas. Figure 4.18 shows some events and their probabilities for Example 4.41. What is the probability that an adult gets adequate sleep and does not exercise?

The Venn diagram shows that the probability that an adult gets adequate sleep minus the probability that an adult gets adequate sleep and exercises regularly is $0.40 - 0.24 = 0.16$. Similarly, the probability that an adult does not get adequate sleep and exercises regularly is $0.46 - 0.24 = 0.22$. The four probabilities that appear in the figure add to 1 because they refer to four disjoint events whose union is the entire sample space.

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.

EXAMPLE 4.42

Probability of being dealt a~~n~~ ~~ace~~. Doyle is a professional poker player. He stares at the dealer, who prepares to deal. What is the probability that the card dealt to Doyle is a heart? There are 52 cards in the deck. Because the deck was carefully shuffled, the next card dealt is equally likely to be any of the cards that Doyle has not seen. Thirteen of the 52 cards are hearts. So

$$P(\text{heart}) = \frac{13}{52} = \frac{1}{4}$$

This calculation assumes that Doyle knows nothing about any cards already dealt. Suppose now that he is looking at four cards already in his hand and that they are all hearts. He knows nothing about the other 48 cards except that exactly nine $(13 - 4)$ hearts are among them. Doyle's probability of being dealt a heart *given what he knows* is now

$$P(\text{heart} \mid 4 \text{ hearts in 4 visible cards}) = \frac{9}{48} = \frac{3}{16}$$

Knowing that there are four hearts among the four cards Doyle can see changes the probability that the next card dealt is a heart.

AU: use hearts as the example here rather than aces?

conditional probability

The new notation $P(A \mid B)$ is a **conditional probability**. That is, it gives the probability of one event (the next card dealt is ~~an ace~~) under the condition that we know another event (exactly one of the four visible cards is ~~an ace~~). You can read the bar | as "given the information that."

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.

USE YOUR KNOWLEDGE

4.91 The probability of a heart. Refer to Example 4.42. Suppose that none of the four cards in Doyle's hand are hearts. What is the probability that the next card dealt to him is a heart?

EXAMPLE 4.43

Downloading music from the Internet. The multiplication rule is just common sense made formal. For example, suppose that 30% of Internet users download music files, and 70% 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 70% of the 30% who download, or

$$(0.7)(0.3) = 0.21 = 21\%$$

The multiplication rule expresses this as

$$\begin{aligned} P(A \text{ and } B) &= P(A) \times P(B \mid A) \\ &= (0.3)(0.7) = 0.21 \end{aligned}$$

Here is another example that uses conditional probability.

EXAMPLE 4.44

Probability of a favorable draw. Doyle is still at the poker table. At the moment, he has two cards and they are both hearts. He has seen 24 cards and none of other players have any hearts. What is the chance that the next three cards he draws will be hearts? The full deck of 52 cards contains 13 hearts. Therefore, 11 of the unseen cards are hearts. There are 28 (52 − 24) unseen cards. To find Doyle's probability of drawing three hearts, we first calculate

$$P(\text{first card is a heart}) = \frac{11}{28}$$

$$P(\text{second card is a heart} \mid \text{first card is a heart}) = \frac{10}{27}$$

$$P(\text{third card is a heart} \mid \text{first two cards are hearts}) = \frac{9}{26}$$

Doyle finds both probabilities by counting cards. The probability that the first card drawn is a heart is 11/28 because 11 of the 28 unseen cards are hearts. If the first card is a heart, that leaves 10 hearts among the

27 remaining cards. So the *conditional* probability of another diamond is 10/27. The multiplication rule now says that

$$P(\text{next two cards are hearts}) = \frac{11}{28} \times \frac{10}{27} = 0.146$$

We again apply the multiplication rule for the third card. The probability that the next three draws are hearts is equal to the probability that the first two draws are hearts times the probability that the third card is a heart given that the first two draws are hearts. This probability is

$$P(\text{next three cards are hearts}) = \frac{11}{28} \times \frac{10}{27} \times \frac{9}{26} = 0.050$$

It is very unlikely that Doyle's next three cards will be hearts, even though his hearts are the only ones that he has seen.

USE YOUR KNOWLEDGE

4.92 The probability that the next two cards are hearts. In the setting of Example 4.44, suppose that Doyle's third card is a heart, so he now has three hearts, and that none of the five additional cards that he sees are hearts What is the probability that the next two cards dealt to Doyle will be hearts?

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

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

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 4.45

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 34	0.10	0.10
30 and over	0.05	0.13

Let's compute the probability that a student is aged 20 to 24, given that the student is full-time. We know that the probability that a student is part-time *and* aged 20 to 24 is 0.32 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 between 20 and 24 years of age
B = the student is a full-time student

Our formula is

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

We read $P(A \text{ and } B) = 0.32$ from the table as we mentioned previously. What about $P(B)$? 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:

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

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

$$\begin{aligned} P(A \mid B) &= \frac{P(A \text{ and } B)}{P(B)} \\ &= \frac{0.32}{0.68} \\ &= 0.47 \end{aligned}$$

The probability that a student is 20 to 24 years of age, given that the student is full-time, is 0.47.

Here is another way to give the information in the last sentence of this example: 47% of full-time college students are 20 to 24 years old. Which way do you prefer?

USE YOUR KNOWLEDGE

4.93 What rule did we use? In Example 4.45, we calculated $P(B)$. What rule did we use for this calculation? Explain why this rule applies in this setting.

4.94 Find the conditional probability. Refer to Example 4.45. What is the probability that a student is part-time, given that the student is 20 to 24 years old? Explain in your own words the difference between this calculation and the one that we did in Example 4.45.

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 4.46

High school athletes and professional careers. Only 5% of male high school basketball, baseball, and football players go on to play at the college level. Of these, only 1.7% enter major league professional sports. About 40% of the athletes who compete in college and then reach the pros have a career of more than three years. Define these events:

$$A = \{\text{competes in college}\}$$
$$B = \{\text{competes professionally}\}$$
$$C = \{\text{pro career longer than 3 years}\}$$

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

$$P(A) = 0.05$$
$$P(B \mid A) = 0.017$$
$$P(C \mid A \text{ and } B) = 0.4$$

Therefore, the probability we want is

$$\begin{aligned} P(A \text{ and } B \text{ and } C) &= P(A)P(B \mid A)P(C \mid A \text{ and } B) \\ &= 0.05 \times 0.017 \times 0.4 = 0.00034 \end{aligned}$$

Only about 3 of every 10,000 high school athletes can expect to compete in college and have a professional career of more than three years. High school students would be wise to concentrate on studies rather than on unrealistic hopes of fortune from pro sports.

Tree diagrams

Probability problems often require us to combine several of the basic rules into a more elaborate calculation. Here is an example that illustrates how to solve problems that have several stages.

EXAMPLE 4.47

Online chat rooms. Online chat rooms are dominated by the young. Teens are the biggest users. If we look only at adult Internet users (aged 18 and over), 47% of the 18 to 29 age group chat, as do 21% of the 30 to 49 age group and just 7% of those 50 and over. To learn what percent of all Internet users participate in chat, we also need the age breakdown of users. Here it is: 29% of adult Internet users are 18 to 29 years old (event A_1), another 47% are 30 to 49 (event A_2), and the remaining 24% are 50 and over (event A_3).

tree diagram

What is the probability that a randomly chosen adult user of the Internet participates in chat rooms (event C)? To find out, use the **tree diagram** in Figure 4.19 to organize your thinking. Each segment in the tree is one stage

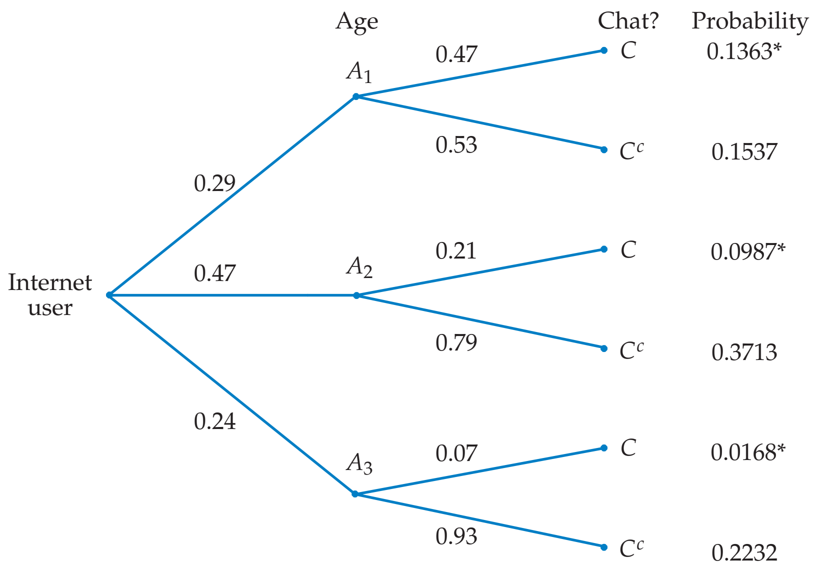


FIGURE 4.19 Tree diagram, Example 4.47. The probability $P(C)$ is the sum of the probabilities of the three branches marked with asterisks (*).

of the problem. Each complete branch shows a path through the two stages. The probability written on each segment is the conditional probability of an Internet user following that segment, given that he or she has reached the node from which it branches.

Starting at the left, an Internet user falls into one of the three age groups. The probabilities of these groups

$$P(A_1) = 0.29 \qquad P(A_2) = 0.47 \qquad P(A_3) = 0.24$$

mark the leftmost branches in the tree. Conditional on being 18 to 29 years old, the probability of participating in chat is $P(C \mid A_1) = 0.47$. So the conditional probability of *not* participating is

$$P(C^c \mid A_1) = 1 - 0.47 = 0.53$$

These conditional probabilities mark the paths branching out from the A_1 node in Figure 4.19. The other two age group nodes similarly lead to two branches marked with the conditional probabilities of chatting or not. The probabilities on the branches from any node add to 1 because they cover all possibilities, given that this node was reached.

There are three disjoint paths to C, one for each age group. By the addition rule, $P(C)$ is the sum of their probabilities. The probability of reaching C through the 18 to 29 age group is

$$\begin{aligned} P(C \text{ and } A_1) &= P(A_1)P(C \mid A_1) \\ &= 0.29 \times 0.47 = 0.1363 \end{aligned}$$

Follow the paths to C through the other two age groups. The probabilities of these paths are

$$P(C \text{ and } A_2) = P(A_2)P(C \mid A_2) = (0.47)(0.21) = 0.0987$$
$$P(C \text{ and } A_3) = P(A_3)P(C \mid A_3) = (0.24)(0.07) = 0.0168$$

The final result is

$$P(C) = 0.1363 + 0.0987 + 0.0168 = 0.2518$$

About 25% of all adult Internet users take part in chat rooms.

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 C that an adult Internet user takes part in chat rooms, is then found by adding the probabilities of all branches that are part of that event.

USE YOUR KNOWLEDGE

4.95 Draw a tree diagram. Refer to Doyle's chances of five hearts in Example 4.44 (page 268). Draw a tree diagram to describe the outcomes for the three cards that he will be dealt. At the first stage, his draw can be a heart or a nonheart. At the second and third stages, he has the same possible outcomes but the probabilities are different.

Bayes's rule

There is another kind of probability question that we might ask in the context of thinking about online chat. What percent of adult chat room participants are aged 18 to 29?

EXAMPLE 4.48

Conditional versus unconditional probabilities. In the notation of Example 4.47, this is the conditional probability $P(A_1 \mid C)$. Start from the definition of conditional probability and then apply the results of Example 4.47:

$$P(A_1 \mid C) = \frac{P(A_1 \text{ and } C)}{P(C)}$$

$$= \frac{0.1363}{0.2518} = 0.5413$$

More than half of adult chat room participants are between 18 and 29 years old. Compare this conditional probability with the original information (unconditional) that 29% of adult Internet users are between 18 and 29 years old. Knowing that a person chats increases the probability that he or she is young.

We know the probabilities $P(A_1)$, $P(A_2)$, and $P(A_3)$ that give the age distribution of adult Internet users. We also know the conditional probabilities $P(C \mid A_1)$, $P(C \mid A_2)$, and $P(C \mid A_3)$ that a person from each age group chats. Example 4.47 shows how to use this information to calculate $P(C)$. The method can be summarized in a single expression that adds the probabilities of the three paths to C in the tree diagram:

$$P(C) = P(A_1)P(C \mid A_1) + P(A_2)P(C \mid A_2) + P(A_3)P(C \mid A_3)$$

In Example 4.48, we calculated the "reverse" conditional probability $P(A_1 \mid C)$. The denominator 0.2518 in that example came from the previous expression. Put in this 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 C is any other event whose probability is not 0 or 1,

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

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 C back to the A_i in a book published in 1763. It is far better to think your way through problems like Examples 4.47 and 4.48 than to memorize these formal expressions.

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

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

AU: add page xref?

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 4.5 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.

SECTION 4.5 EXERCISES

For Exercise 4.89, see page 266; for Exercise 4.90, see page 266; for Exercise 4.91, see page 268; for Exercise 4.92, see page 269; for Exercises 4.93 and 4.94, see page 270; and for Exercise 4.95, see page 273.

4.96 Find and explain some probabilities.

(a) Can we have an event A that has negative probability? Explain your answer.

(b) Suppose $P(A) = 0.3$ and $P(B) = 0.5$. 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 $P(A) = 0.4$, what is the probability that A does not occur?

(e) Suppose that A and B are independent and that $P(A) = 0.8$ and $P(B) = 0.3$. Explain the meaning of the event $\{A \text{ and } B\}$, and find its probability.

4.97 Unions.

(a) Assume that $P(A) = 0.2$, $P(B) = 0.4$, 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.

4.98 Conditional probabilities. Suppose that $P(A) = 0.4$, $P(B) = 0.3$, and $P(B \mid A) = 0.4$.

(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?

4.99 Find the probabilities. Suppose that the probability that A occurs is 0.5 and the probability that A and B occur is 0.2.

(a) Find the probability that B occurs given that A occurs.

(b) Illustrate your calculations in part (a) using a Venn diagram.

4.100 Why not? Suppose that $P(B) = 0.6$. Explain why $P(A \text{ and } B)$ cannot be 0.7.

4.101 Is the calcium intake adequate? In the population of young children eligible to participate in a study of whether or not their calcium intake is adequate, 52% are 5 to 10 years of age and 48% are 11 to 13 years of age. For those who are 5 to 10 years of age, 18% have inadequate calcium intake. For those who are 11 to 13 years of age, 57% have inadequate calcium intake.[18]

(a) Use letters to define the events of interest in this exercise.

(b) Convert the percents given to probabilities of the events you have defined.

(c) Use a tree diagram similar to Figure 4.19 (page 272) to calculate the probability that a randomly selected child from this population has an inadequate intake of calcium.

4.102 Use Bayes's rule. Refer to the previous exercise. Use Bayes's rule to find the probability that a child from this population who has inadequate intake is 11 to 13 years old.

4.103 Are the events independent? Refer to the previous two exercises. Are the age of the child and whether or not the child has adequate calcium intake independent? Calculate the probabilities that you need to answer this question and write a short summary of your conclusion.

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

4.105 Exercise and sleep. Suppose that 42% of adults get enough sleep, 39% get enough exercise, and 28% do both. Find the probabilities of the following events:

(a) Enough sleep and not enough exercise.

(b) Not enough sleep and enough exercise.

(c) Not enough sleep and not enough exercise.

(d) For each of parts (a), (b), and (c), state the rule that you used to find your answer.

4.106 Exercise and sleep. Refer to the previous exercise. Draw a Venn diagram showing the probabilities for exercise and sleep.

4.107 Lying to a teacher. Suppose that 53% of high school students would admit to lying at least once to a teacher during the past year and that 24% of students are male and would admit to lying at least once to a teacher during the past year.[19] Assume that 44% of the students are male. What is the probability that a randomly selected student is either male or would admit to lying to a teacher during the past year? Be sure to show your work and indicate all the rules that you use to find your answer.

4.108 Lying to a teacher. Refer to the previous exercise. Suppose that you select a student from the subpopulation of those who would admit to lying to a teacher during the past year. What is the probability that the student is female? Be sure to show your work and indicate all the rules that you use to find your answer.

4.109 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?

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

4.111 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 where 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.

4.112 Education and income. Call a household prosperous if its income exceeds $100,000. Call the household educated if the householder 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?

4.113 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.

4.114 Draw a Venn diagram. Draw a Venn diagram that shows the relation between the events A and B in Exercise 4.112. Indicate each of the following events on your diagram and use the information in Exercise 4.112 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) $\{A^c \text{ and } B^c\}$.

4.115 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.

4.116 Job offers. Emily is graduating from college. She has studied biology, chemistry, and computing and hopes to work as a forensic scientist applying her science background to crime investigation. Late one night she thinks about some jobs she has applied for. Let A, B, and C be the events where Emily is offered a job by

A = the Connecticut Office of the Chief Medical Examiner

B = the New Jersey Division of Criminal Justice

C = the federal Disaster Mortuary Operations Response Team

Julie writes down her personal probabilities for being offered these jobs:

$P(A) = 0.6$ $\quad P(B) = 0.5$ $\quad P(C) = 0.3$

$P(A \text{ and } B) = 0.3$ $\quad P(A \text{ and } C) = 0.1$ $\quad P(B \text{ and } C) = 0.1$

$P(A \text{ and } B \text{ and } C) = 0$

Make a Venn diagram of the events A, B, and C. As in Figure 4.18 (page 267), mark the probabilities of every intersection involving these events and their complements. Use this diagram for Exercises 4.117, 4.118, and 4.119.

4.117 Find the probability of at least one offer. What is the probability that Julie is offered at least one of the three jobs?

4.118 Find the probability of another event. What is the probability that Julie is offered both the Connecticut and New Jersey jobs, but not the federal job?

4.119 Find a conditional probability. If Julie is offered the federal job, what is the conditional probability that she is also offered the New Jersey job? If Julie is offered the New Jersey job, what is the conditional probability that she is also offered the federal job?

4.120 Conditional probabilities and independence. Using the information in Exercise 4.115, 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.

Genetic counseling. *Conditional probabilities and Bayes's rule are a basis for counseling people who may have genetic defects that can be passed to their children. Exercises 4.121, 4.112, and 4.123 concern genetic counseling settings.*

4.121 Albinism. People with albinism have little pigment in their skin, hair, and eyes. The gene that governs albinism has two forms (called alleles), which we denote by a and A. Each person has a pair of these genes, one inherited from each parent. A child inherits one of each parent's two alleles independently with probability 0.5. Albinism is a recessive trait, so a person is albino only if the inherited pair is aa.

(a) Beth's parents are not albino but she has an albino brother. This implies that both of Beth's parents have type Aa. Why?

(b) Which of the types aa, Aa, AA could a child of Beth's parents have? What is the probability of each type?

(c) Beth is not albino. What are the conditional probabilities for Beth's possible genetic types, given this fact? (Use the definition of conditional probability.)

4.122 Find some conditional probabilities. Beth knows the probabilities for her genetic types from part (c) of the previous exercise. She marries Bob, who is albino. Bob's genetic type must be aa.

(a) What is the conditional probability that a child of Beth and Bob is non-albino if Beth has type Aa? What is the conditional probability of a non-albino child if Beth has type AA?

(b) Beth and Bob's first child is non-albino. What is the conditional probability that Beth is a carrier, type Aa?

4.123 Muscular dystrophy. Muscular dystrophy is an incurable muscle-wasting disease. The most common and serious type, called DMD, is caused by a sex-linked recessive mutation. Specifically, women can be carriers but do not get the disease; a son of a carrier has probability 0.5 of having DMD; a daughter has probability 0.5 of being a carrier. As many as one-third of DMD cases, however, are due to spontaneous mutations in sons of mothers who are not carriers. Toni has one son, who has DMD.

In the absence of other information, the probability is 1/3 that the son is the victim of a spontaneous mutation and 2/3 that Toni is a carrier. There is a screening test called the CK test that is positive with probability 0.7 if a woman is a carrier and with probability 0.1 if she is not. Toni's CK test is positive. What is the probability that she is a carrier?

CHAPTER 4 EXERCISES

4.124 Repeat the experiment many times. Here is a probability distribution for a random variable X:

Value of X	-3	4
Probability	0.3	0.7

A single experiment generates a random value from this distribution. If the experiment is repeated many times, what will be the approximate proportion of times that the value is -3? Give a reason for your answer.

4.125 Repeat the experiment many times and take the mean. Here is a probability distribution for a random variable X:

Value of X	-8	5
Probability	0.6	0.4

A single experiment generates a random value from this distribution. If the experiment is repeated many times, what will be the approximate value of the mean of these random variables? Give a reason for your answer.

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

Value of X	2	3
Probability	0.2	0.8

(a) Find the mean and the standard deviation of this distribution.

(b) Let $Y = 5X - 1$. 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.

4.127 A different transformation. Refer to the previous exercise. Now let $Y = 5X^2 - 1$.

(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.

4.128 Roll a pair of dice two times. Consider rolling a pair of fair dice two times. Let A be the total on the up-faces for the first roll and let B be the total on the up-faces for the second roll. 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$\}$.

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

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

Value of X	2	3	4
Probability	0.4	0.3	0.3

(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.

4.131 A fair bet at craps. Almost all bets made at gambling casinos favor the house. In other words, the difference between the amount bet and the mean of the distribution of the payoff is a positive number. An exception is "taking the odds" at the game of craps, a bet that a player can make under certain circumstances. The bet becomes available when a shooter throws a 4, 5, 6, 8, 9, or 10 on the initial roll. This number is called the "point"; when a point is rolled, we say that a point has been established. If a 4 is the point, an odds bet can be made that wins if a 4 is rolled before a 7 is rolled. The probability of winning this bet is 1/3, and the same payoff for a \$10 bet is \$20 (you keep the \$10 you bet and you receive an additional \$20). The same probability of winning and payoff apply for an odds bet on a 10. For an initial roll of 5 or 9, the odds bet has a winning probability of 2/5, and the payoff for a \$10 bet is \$15. Similarly, when the initial roll is 6 or 8, the odds bet has a winning probability of 5/11, and the payoff for a \$10 bet is \$12.

(a) Find the mean of the payoff distribution for each of these bets.

(b) Confirm that the bets are fair by showing that the difference between the amount bet and the mean of the distribution of the payoff is zero.

4.132 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.

4.133 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?

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

4.135 Lottery tickets. Michael buys a ticket in the Tri-State Pick 3 lottery every day, always betting on 812. He will win something if the winning number contains 8, 1, and 2 in any order. Each day, Michael has probability 0.006 of winning, and he wins (or not) independently of other days because a new drawing is held each day. What is the probability that Michael's first winning ticket comes on the 10th day?

4.136 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:[20]

	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.

4.137 Odds bets at craps. Refer to the odds bets at craps in Exercise 4.131. Suppose that whenever the shooter has an initial roll of 4, 5, 6, 8, 9, or 10, you take the odds. Here are the probabilities for these initial rolls:

Point	4	5	6	8	9	10
Probability	3/36	4/36	5/36	5/36	4/36	3/36

Draw a tree diagram with the first stage showing the point rolled and the second stage showing whether the point is again rolled before a 7 is rolled. Include a first-stage branch showing the outcome that a point is not established. In this case, the amount bet is zero and the distribution of the winnings is the special random variable that has $P(X = 0) = 1$. For the combined betting system where the player always makes a \$10 odds bet when it is available, show that the game is fair.

4.138 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?

4.139 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 $Y > X$.)