

# Probability Models

# 18

**CASE STUDY** Shortly after the New England Patriots won Super Bowl XLIX, websites were already posting the probabilities of winning Super Bowl 50 for the various NFL teams. These probabilities were updated regularly. Table 18.1 lists the probabilities posted on one website on September 21, 2015. These are perhaps best interpreted as personal probabilities. They are likely to change as the 2015 season progresses because the players on teams will change due to trades and injuries.

Because these are personal probabilities, are there any restrictions on the choices for the probabilities? In this chapter, we will answer this question. We will learn that probabilities must obey certain rules in order to make sense. By the end of this chapter, you will be able to assess whether the probabilities in Table 18.1 make sense.



Christian Petersen/Getty Images

## Probability models

Choose a woman aged 25 to 29 years old at random and record her marital status. “At random” means that we give every such woman the same chance to be the one we choose. That is, we choose a random sample of size 1. The probability of any marital status is just the proportion of all women aged 25 to 29 who have that status—if we chose many women, this is the proportion we would get. Here is the set of probabilities:

Marital status:	Never married	Married	Widowed	Divorced
Probability:	0.478	0.476	0.004	0.042

This table gives a *probability model* for drawing a young woman at random and finding out her marital status. It tells us what are the possible outcomes (there are only four) and it assigns probabilities to these outcomes. The probabilities here are the proportions of all women in each

**TABLE 18.1** Probabilities of winning Super Bowl 50

Team	Probability	Team	Probability
Green Bay Packers	1/5	Buffalo Bills	1/41
New England Patriots	1/7	Kansas City Chiefs	1/41
Seattle Seahawks	1/11	St. Louis Rams	1/51
Indianapolis Colts	1/13	New York Jets	1/61
Denver Broncos	1/16	New York Giants	1/86
Arizona Cardinals	1/21	Detroit Lions	1/101
Cincinnati Bengals	1/21	Houston Texans	1/101
Pittsburgh Steelers	1/21	New Orleans Saints	1/101
Atlanta Falcons	1/26	Oakland Raiders	1/101
Dallas Cowboys	1/26	San Francisco 49ers	1/101
Philadelphia Eagles	1/26	Tampa Bay Buccaneers	1/101
Carolina Panthers	1/31	Washington Redskins	1/101
Miami Dolphins	1/31	Cleveland Browns	1/201
San Diego Chargers	1/31	Tennessee Titans	1/201
Minnesota Vikings	1/36	Chicago Bears	1/251
Baltimore Ravens	1/41	Jacksonville Jaguars	1/301

Source: <http://www.betvega.com/super-bowl-odds/> (as of September 21, 2015).



**Politically correct** In 1950, the Russian mathematician B. V. Gnedenko (1912–1995) wrote a book, *The Theory of Probability*, that was popular around the world. The introduction contains a mystifying paragraph that begins, “We note that the entire development of probability theory shows evidence of how its concepts and ideas were crystallized in a severe struggle between materialistic and idealistic conceptions.” It turns out that “materialistic” is jargon for “Marxist-Leninist.” It was good for the health of Russian scientists in the Stalin era to add such statements to their books.

marital class. That makes it clear that the probability that a woman is not married is just the sum of the probabilities of the three classes of unmarried women:

$$\begin{aligned}
 P(\text{not married}) &= P(\text{never married}) + P(\text{widowed}) \\
 &\quad + P(\text{divorced}) \\
 &= 0.478 + 0.004 + 0.042 = 0.524
 \end{aligned}$$

As a shorthand, we often write  $P(\text{not married})$  for “the probability that the woman we choose is not married.” You see that our model does more than assign a probability to each individual outcome—we can find the probability of any collection of outcomes by adding up individual outcome probabilities.

### Probability model

A **probability model** for a random phenomenon describes all the possible outcomes and says how to assign probabilities to any collection of outcomes. We sometimes call a collection of outcomes an **event**.

## Probability rules

Because the probabilities in this example are just the proportions of all women who have each marital status, they follow rules that say how proportions behave. Here are some basic rules that any probability model must obey:

- A. 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.
- B. All possible outcomes together must have probability 1.** Because some outcome must occur on every trial, the sum of the probabilities for all possible outcomes must be exactly 1.
- C. 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.
- D. If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.** If one event occurs in 40% of all trials, a different event occurs in 25% of all trials, and the two can never occur together, then one or the other occurs on 65% of all trials because  $40\% + 25\% = 65\%$ .

### EXAMPLE 1 Marital status of young women

Look again at the probabilities for the marital status of young women. Each of the four probabilities is a number between 0 and 1. Their sum is

$$0.478 + 0.476 + 0.004 + 0.042 = 1$$

This assignment of probabilities satisfies Rules A and B. **Any assignment of probabilities to all individual outcomes that satisfies Rules A and B is legitimate.** That is, it makes sense as a set of probabilities. Rules C and D are then automatically true. Here is an example of the use of Rule C.

The probability that the woman we draw is not married is, by Rule C,

$$\begin{aligned} P(\text{not married}) &= 1 - P(\text{married}) \\ &= 1 - 0.476 = 0.524 \end{aligned}$$

That is, if 47.6% are married, then the remaining 52.4% are not married. Rule D says that you can also find the probability that a woman is not married by adding the probabilities of the three distinct ways of being not married, as we did earlier. This gives the same result.



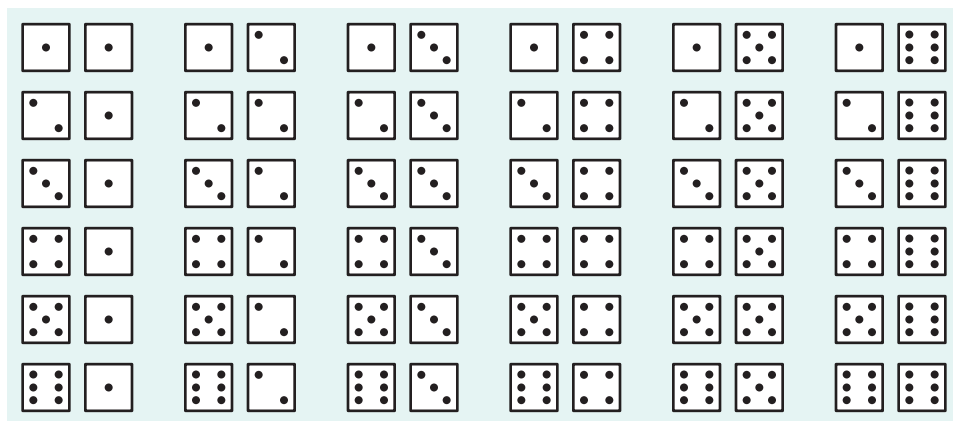
### EXAMPLE 2 Rolling two dice

Rolling two dice is a common way to lose money in casinos. There are 36 possible outcomes when we roll two dice and record the up-faces in order (first die, second die). Figure 18.1 displays these outcomes. What probabilities should we assign?

Casino dice are carefully made. Their spots are not hollowed out, which would give the faces different weights, but are filled with white plastic of the same density as the red plastic of the body. For casino dice, it is reasonable to assign the same probability to each of the 36 outcomes in Figure 18.1. Because these 36 probabilities must have sum 1 (Rule B), each outcome must have probability  $1/36$ , or 1-in-36.

We are interested in the sum of the spots on the up-faces of the dice. What is the probability that this sum is 5? The event “roll a 5” contains four outcomes, and its probability is the sum of the probabilities of these outcomes:

$$\begin{aligned}
 P(\text{roll a } 5) &= P\left(\begin{array}{|c|c|} \hline \bullet & \bullet\bullet \\ \hline \end{array}\right) + P\left(\begin{array}{|c|c|} \hline \bullet & \bullet\bullet\bullet \\ \hline \end{array}\right) \\
 &\quad + P\left(\begin{array}{|c|c|} \hline \bullet\bullet & \bullet \\ \hline \end{array}\right) + P\left(\begin{array}{|c|c|} \hline \bullet\bullet\bullet & \bullet \\ \hline \end{array}\right) \\
 &= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} + \frac{1}{36} \\
 &= \frac{4}{36} = 0.111
 \end{aligned}$$



**Figure 18.1** The 36 possible outcomes from rolling two dice, Example 2.

**18.1 Rolling dice.** Suppose you roll two casino dice, as in Example 2. What is the probability that the sum of the spots on the up-faces is 7? 11? 7 or 11?

**NOW IT'S  
YOUR TURN**

The rules tell us only what probability models *make sense*. They don't tell us whether the probabilities are *correct*, that is, whether they describe what actually happens in the long run. The probabilities in Example 2 are correct for casino dice. Inexpensive dice with hollowed-out spots are not balanced, and this probability model does not describe their behavior.

What about personal probabilities? Because they are personal, can't they be anything you want them to be? If your personal probabilities don't obey Rules A and B, you are entitled to your opinion, but we can say that your personal probabilities are **incoherent**. That is, they don't go together in a way that makes sense. So we usually insist that personal probabilities for all the outcomes of a random phenomenon obey Rules A and B. That is, the same rules govern both kinds of probability. For example, if you believe that the Green Bay Packers, the New Orleans Saints, and the New England Patriots each have probability 0.4 of winning Super Bowl XLVII, your personal probabilities would not obey Rule B.

## Probability and odds

Speaking of sports, newspapers, magazines, and websites often give probabilities as betting odds in the form “Y to Z.” This form means that a bet of \$Z will pay you \$Y if the team wins. If this is a fair bet, you expect that in the long run you should break even, winning as much money as you lose. In particular, if you bet Y + Z times, on average you should win \$Y Z times and lose \$Z the other Y times. Thus, on average, if you bet Y + Z times, you win Z of those bets. Odds of Y to Z therefore represent a probability of  $Z/(Y + Z)$  of winning.

On the Web, one can find the current odds of winning the next Super Bowl for each NFL team. We found a list of such odds at [www.vegasinsider.com/nfl/odds/futures/](http://www.vegasinsider.com/nfl/odds/futures/). When we checked this website on December 1, 2015, the odds that the New England Patriots would win Super Bowl 50 were 3-to-1. This corresponds to a probability of winning of  $1/(3 + 1) = 1/4$ . The odds that the San Francisco 49ers would win Super Bowl 50 were 6000 to 1. This corresponds to a probability of  $1/(6000 + 1) = 1/6001$ .

## Probability models for sampling

Choosing a random sample from a population and calculating a statistic such as the sample proportion is certainly a random phenomenon. The *distribution* of the statistic tells us what values it can take and how often it takes those values. That sounds a lot like a probability model.

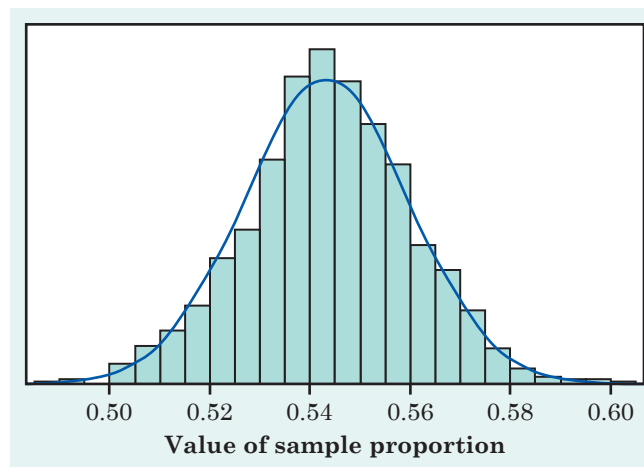
### EXAMPLE 3 A sampling distribution

Take a simple random sample of 1015 adults. Ask each whether they feel childhood vaccinations are extremely important. The proportion who say Yes

$$\hat{p} = \frac{\text{number who say Yes}}{1015}$$

is the sample proportion  $\hat{p}$ . Do this 1000 times and collect the 1000 sample proportions  $\hat{p}$  from the 1000 samples. The histogram in Figure 18.2 shows the distribution of 1000 sample proportions when the truth about the population is that 54% would favor such an amendment. The results of random sampling are of course random: we can't predict the outcome of one sample, but the figure shows that the outcomes of many samples have a regular pattern.

This repetition reminds us that the regular pattern of repeated random samples is one of the big ideas of statistics. The Normal curve in the



**Figure 18.2** The sampling distribution of a sample proportion  $\hat{p}$  from SRSs of size 1015 drawn from a population in which 54% of the members would give positive answers, Example 3. The histogram shows the distribution from 1000 samples. The Normal curve is the ideal pattern that describes the results of a very large number of samples.



figure is a good approximation to the histogram. The histogram is the result of these particular 1000 SRSs. Think of the Normal curve as the idealized pattern we would get if we kept on taking SRSs from this population forever. That's exactly the idea of probability—the pattern we would see in the very long run. *The Normal curve assigns probabilities to sample proportions computed from random samples.*

This Normal curve has mean 0.540 and standard deviation about 0.016. The “95” part of the 68–95–99.7 rule says that 95% of all samples will give a  $\hat{p}$  falling within 2 standard deviations of the mean. That's within 0.032 of 0.540, or between 0.508 and 0.572. We now have more concise language for this fact: the *probability* is 0.95 that between 50.8% and 57.2% of the people in a sample will say Yes. The word “probability” says we are talking about what would happen in the long run, in very many samples.

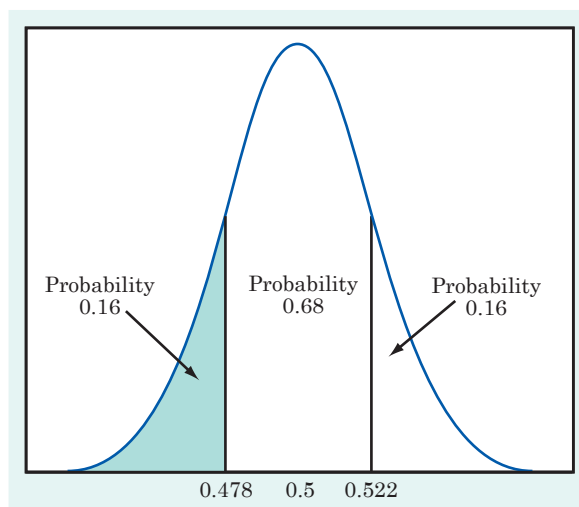
We note that of the 1000 SRSs, 95% of the sample proportions were between 0.509 and 0.575, which agrees quite well with the calculations based on the Normal curve. This confirms our assertion that the Normal curve is a good approximation to the histogram in Figure 18.2.

A statistic from a large sample has a great many possible values. Assigning a probability to each individual outcome worked well for four marital classes or 36 outcomes of rolling two dice but is awkward when there are thousands of possible outcomes. Example 3 uses a different approach: assign probabilities to intervals of outcomes by using areas under a Normal density curve. Density curves have area 1 underneath them, which lines up nicely with total probability 1. The total area under the Normal curve in Figure 18.2 is 1, and the area between 0.508 and 0.572 is 0.95, which is the probability that a sample gives a result in that interval. When a Normal curve assigns probabilities, you can calculate probabilities from the 68–95–99.7 rule or from Table B of percentiles of Normal distributions. These probabilities satisfy Rules A through D.

### Sampling distribution

The **sampling distribution** of a statistic tells us what values the statistic takes in repeated samples from the same population and how often it takes those values.

We think of a sampling distribution as assigning probabilities to the values the statistic can take. Because there are usually many possible values, sampling distributions are often described by a **density curve** such as a Normal curve.



**Figure 18.3** The Normal sampling distribution, Example 4. Because 0.478 is 1 standard deviation below the mean, the area under the curve to the left of 0.478 is 0.16.

### EXAMPLE 4 Do you approve of gambling?

An opinion poll asks an SRS of 501 teens, “Generally speaking, do you approve or disapprove of legal gambling or betting?” Suppose that, in fact, exactly 50% of all teens would say Yes if asked. (This is close to what polls show to be true.) The poll’s statisticians tell us that the sample proportion who say Yes will vary in repeated samples according to a Normal distribution with mean 0.5 and standard deviation about 0.022. This is the *sampling distribution* of the sample proportion  $\hat{p}$ .

The 68–95–99.7 rule says that the probability is 0.16 that the poll gets a sample in which fewer than 47.8% say Yes. Figure 18.3 shows how to get this result from the Normal curve of the sampling distribution.

### NOW IT'S YOUR TURN

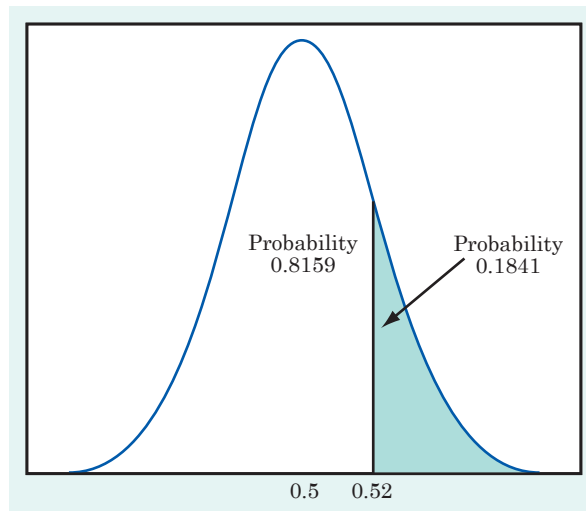
**18.2 Teen opinion poll.** Refer to Example 4. Using the 68–95–99.7 rule, what is the probability that fewer than 45.6% say Yes?

### EXAMPLE 5 Using Normal percentiles\*

What is the probability that the opinion poll in Example 4 will get a sample in which 52% or more say Yes? Because 0.52 is not 1, 2, or 3 standard deviations away from the mean, we can’t use the 68–95–99.7 rule. We will use Table B of percentiles of Normal distributions.

\*Example 5 is optional.





**Figure 18.4** The Normal sampling distribution, Example 5. The outcome 0.52 has standard score 0.9, so Table B tells us that the area under the curve to the left of 0.52 is 0.8159.

To use Table B, first turn the outcome  $\hat{p} = 0.52$  into a standard score by subtracting the mean of the distribution and dividing by its standard deviation:

$$\frac{0.52 - 0.5}{0.022} = 0.9$$

Now look in Table B. A standard score of 0.9 is the 81.59 percentile of a Normal distribution. This means that the probability is 0.8159 that the poll gets a smaller result. By Rule C (or just the fact that the total area under the curve is 1), this leaves probability 0.1841 for outcomes with 52% or more answering Yes. Figure 18.4 shows the probabilities as areas under the Normal curve.

## STATISTICS IN SUMMARY

### Chapter Specifics

- A **probability model** describes a random phenomenon by telling what outcomes are possible and how to assign probabilities to them.
- There are two simple ways to give a probability model. The first assigns a probability to each individual outcome. These probabilities must be numbers between 0 and 1 (Rule A), and they must add to exactly 1 (Rule B).

To find the probability of any **event**, add the probabilities of the outcomes that make up the event.

- The second kind of probability model assigns probabilities as areas under a **density curve**, such as a Normal curve. The total probability is 1 because the total area under the curve is 1. This kind of probability model is often used to describe the **sampling distribution** of a statistic. This is the pattern of values of the statistic in many samples from the same population.
- All legitimate assignments of probability, whether data based or personal, obey the same **probability rules**. So the mathematics of probability is always the same.
- **Odds** of  $Y$  to  $Z$  that an event occurs corresponds to a probability of  $Z/(Y + Z)$ .



This chapter continues the discussion of probability that we began in Chapter 17. Here, we examine the formal mathematics of probability, embodied in probability models and probability rules. Probability models and rules provide the tools for describing and predicting the long-run behavior of random phenomena.

In this chapter, we also begin to formalize the process, first mentioned in Chapter 3, of using a statistic to estimate an unknown parameter. In particular, the sampling distribution will be the “probabilistic” tool we use to generalize from data produced by random samples and randomized comparative experiments to some wider population. Exactly how we do this will be the subject of Part IV.

**CASE STUDY** Look again at Table 18.1, discussed in the Case Study that opened the chapter.

1. Do the probabilities in this table follow the rules given on page 429?
2. On some websites, the probabilities are given as betting odds of winning. If you are a bookie setting betting odds, should you set the odds so the probabilities sum to greater than 1, exactly 1, or less than 1? Discuss.



**LaunchPad Online Resources**  
macmillan learning

- The Snapshots video *Probability* introduces the concepts of randomness, probability, and probability models in the context of some examples.
- The StatBoards video *The Four Basic Probability Rules* discusses the basic probability rules in the context of some simple examples.

## CHECK THE BASICS

For Exercise 18.1, see page 431; for Exercise 18.2, see page 434.

**18.3 Probability models** A probability model describes a random phenomenon by telling us which of the following?

- (a) Whether we are using data based or personal probabilities.
- (b) What outcomes are possible and how to assign probabilities to these outcomes.
- (c) Whether the probabilities of all outcomes sum to 1 or sum to a number different than 1.
- (d) All of the above.

**18.4 Probability models.** Which of the following is true of any legitimate probability model?

- (a) The probabilities of the individual outcomes must be numbers between 0 and 1, and they must sum to no more than 1.
- (b) The probabilities of the individual outcomes must be numbers between 0 and 1, and they must sum to at least 1.
- (c) The probabilities of the individual outcomes must be numbers between 0 and 1, and they must sum to exactly 1.
- (d) Probabilities can be computed using the Normal curve.

**18.5 Density curves.** Which of the following is true of density curves?

- (a) Areas under a density curve determine probabilities of outcomes.
- (b) The total area under a density curve is 1.
- (c) The Normal curve is a density curve.
- (d) All of the above are true.

**18.6 Probability rules.** To find the probability of any event,

- (a) add up the probabilities of the outcomes that make up the event.
- (b) use the probability of the outcome that best approximates the event.
- (c) assign it a random, but plausible, value between 0 and 1.
- (d) average together the personal probabilities of several experts.

**18.7 Sampling distributions.** The sampling distribution of a statistic is

- (a) the method of sampling used to obtain the data from which the statistic is computed.
- (b) the possible methods of computing a statistic from the data.
- (c) the pattern of the data from which the statistic is computed.
- (d) the pattern of values of the statistic in many samples from the same population.

## CHAPTER 18 EXERCISES

**18.8 Moving up.** A sociologist studying social mobility in Denmark finds that the probability that the son of a lower-class father remains in the lower class is 0.46. What is the

probability that the son moves to one of the higher classes?

**18.9 Causes of death.** Government data assign a single cause for each

death that occurs in the United States. Data from 2013 show that the probability is 0.31 that a randomly chosen death was due to heart disease and 0.23 that it was due to cancer. What is the probability that a death was due either to heart disease or to cancer? What is the probability that the death was due to some other cause?

**18.10 Land in Canada.** Choose an acre of land in Canada at random. The probability is 0.45 that it is forest and 0.03 that it is pasture.

- (a) What is the probability that the acre chosen is not forested?
- (b) What is the probability that it is either forest or pasture?
- (c) What is the probability that a randomly chosen acre in Canada is something other than forest or pasture?



**18.11 Our favorite president.**

A Gallup Poll on Presidents’ Day 2011 interviewed a random sample of 1015 adult Americans. Those in the sample were asked who

they regarded as the greatest U.S. president. Here are the results:

Outcome	Probability
Ronald Reagan	0.19
Abraham Lincoln	0.14
Bill Clinton	0.13
John F. Kennedy	0.11
Someone else	?

These proportions are probabilities for the random phenomenon of choosing an adult American at random and asking her or his opinion.

- (a) What must be the probability that the person chosen selects someone other than Ronald Reagan, Abraham Lincoln, Bill Clinton, or John F. Kennedy? Why?
- (b) The event “I would select either Ronald Reagan or Abraham Lincoln” contains the first two outcomes. What is its probability?

**18.12 Rolling a die.** Figure 18.5 displays several assignments of probabilities to

Outcome	Probability			
	Model 1	Model 2	Model 3	Model 4
	1/7	1/3	1/3	1
	1/7	1/6	1/6	1
	1/7	1/6	1/6	2
	1/7	0	1/6	1
	1/7	1/6	1/6	1
	1/7	1/6	1/6	2

**Figure 18.5** Four probability models for rolling a die, Exercise 18.12.

the six faces of a die. We can learn which assignment is actually *correct* for a particular die only by rolling the die many times. However, some of the assignments are not *legitimate* assignments of probability. That is, they do not obey the rules. Which are legitimate and which are not? In the case of the illegitimate models, explain what is wrong.

**18.13 The demands of college.** Select a first-year college student at random and ask how difficult it was for him or her to adjust to the academic demands of college. Here are the probabilities, based on proportions from a large sample survey in 2010 of first-year students:

Difficulty:	Very difficult	Somewhat difficult
Probability:	0.05	0.31

Difficulty:	Somewhat easy	Very easy
Probability:	0.44	0.20

- (a) What is the sum of these probabilities? Why do you expect the sum to have this value?
- (b) What is the probability that a randomly chosen first-year college student did not find it very difficult to adjust to the academic demands of college?
- (c) What is the probability that a first-year student found it very or somewhat difficult to adjust to the academic demands of college?

**18.14 Tetrahedral dice.** Psychologists sometimes use tetrahedral dice to study our intuition about chance behavior. A tetrahedron (Figure 18.6) is a pyramid with four faces, each a triangle with all sides equal in length.



**Figure 18.6** A tetrahedron. Exercises 18.14 and 18.16 concern dice with this shape.

Label the four faces of a tetrahedral die with one, two, three, and four spots. Give a probability model for rolling such a die and recording the number of spots on the down-face. Explain why you think your model is at least close to correct.

**18.15 Birth order.** A couple plan to have three children. There are eight possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All eight arrangements are (approximately) equally likely.

- (a) Write down all eight arrangements of the sexes of three children. What is the probability of any one of these arrangements?
- (b) What is the probability that the couple's children are two girls and one boy?

**18.16 More tetrahedral dice.** Tetrahedral dice are described in Exercise 18.14. Give a probability model for rolling two such dice. That is, write down all possible outcomes and give a probability to each. (Example 2 and Figure 18.1 may help you.) What is the probability that the sum of the down-faces is 5?

**18.17 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 of any one of the 38 possible outcomes? Explain your answer.

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

(c) A friend tells you that the odds that a bet on “red” will win are 10 to 9. Is your friend correct? If not, what are the correct odds?

(d) 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?

**18.18 Colors of M&M's.** If you draw an M&M candy at random from a bag of the candies, the candy you draw will have one of six colors. The probability of drawing each color depends on the proportion of each color among all candies made.

(a) Here are the probabilities of each color for a randomly chosen plain M&M:

Color:	Blue	Orange	Green
Probability:	0.24	0.20	0.16

---

Color:	Yellow	Red	Brown
Probability:	0.14	0.13	?

What must be the probability of drawing a brown candy?

(b) The probabilities for peanut M&M's are a bit different. Here they are:

Color:	Blue	Orange	Green
Probability:	0.23	0.23	0.15

---

Color:	Yellow	Brown	Red
Probability:	0.15	0.13	?

What is the probability that a peanut M&M chosen at random is red?

(c) What is the probability that a plain M&M is any of blue, yellow, or orange? What is the probability that a peanut M&M has one of these colors?

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

(a) When a coin is spun,  $P(H) = 0.55$  and  $P(T) = 0.45$ .

(b) When two coins are tossed,  $P(HH) = 0.4$ ,  $P(HT) = 0.4$ ,  $P(TH) = 0.4$ , and  $P(TT) = 0.4$ .

(c) Plain M&M's have not always had the mixture of colors given in Exercise 18.18. In the past, there were no red candies and no blue candies. Tan had probability 0.10, and the other four colors had the same probabilities that are given in Exercise 18.13.

**18.20 Immigration.** Suppose that 34% of all adult Americans think that the level of immigration to the United States should be decreased. An opinion poll interviews 1020 randomly chosen Americans and records the

sample proportion who feel that the level of immigration to this country should be decreased. This statistic will vary from sample to sample if the poll is repeated. The sampling distribution is approximately Normal with mean 0.34 and standard deviation about 0.015. Sketch this Normal curve and use it to answer the following questions.

- (a) The mean of the population is 0.34. In what range will the middle 95% of all sample results fall?
- (b) What is the probability that the poll gets a sample in which fewer than 32.5% say the level of immigration to this country should be decreased?

**18.21 Airplane safety.** Suppose that 73% of all adults think that the overall safety of airline travel is either good or excellent. An opinion poll plans to ask an SRS of 1097 adults about airplane safety. The proportion of the sample who think that the overall safety of airline travel is either good or excellent will vary if we take many samples from this same population. The sampling distribution of the sample proportion is approximately Normal with mean 0.73 and standard deviation about 0.013. Sketch this Normal curve and use it to answer the following questions.

- (a) What is the probability that the poll gets a sample in which more than 75.6% of the people think that airplanes would be safer if passengers were banned from carrying on board any luggage, including purses, computers, and briefcases?
- (b) What is the probability of getting a sample that misses the truth (73%) by 2.6% or more?

**18.22 Immigration (optional).** In the setting of Exercise 18.20, what is the probability of getting a sample in which less than 30% of those sampled think that the level of immigration to this country should be decreased? (Use Table B.)

**18.23 Airplane safety (optional).** In the setting of Exercise 18.21, what is the probability of getting a sample in which more than 75% think that the overall safety of airline travel is either good or excellent? (Use Table B.)

**18.24 Do you jog?** An opinion poll asks an SRS of 1500 adults, “Do you happen to jog?” Suppose (as is approximately correct) that the population proportion who jog is  $p = 0.15$ . In a large number of samples, the proportion  $\hat{p}$  who answer Yes will be approximately Normally distributed with mean 0.15 and standard deviation 0.009. Sketch this Normal curve and use it to answer the following questions.

- (a) What percentage of many samples will have a sample proportion who jog that is 0.15 or less? Explain clearly why this percentage is the probability that  $\hat{p}$  is 0.15 or less.
- (b) What is the probability that  $\hat{p}$  will take a value between 0.141 and 0.159? (Use the 68–95–99.7 rule.)
- (c) Now use Rule C for probability: what is the probability that  $\hat{p}$  does not lie between 0.141 and 0.159?

**18.25 Applying to college.** You ask an SRS of 1500 college students whether they applied for admission to any other college. Suppose that, in fact, 35% of all college students applied to colleges besides the one they



are attending. (That's close to the truth.) The sampling distribution of the proportion  $\hat{p}$  of your sample who say Yes is approximately Normal with mean 0.35 and standard deviation 0.01. Sketch this Normal curve and use it to answer the following questions.

- Explain in simple language what the sampling distribution tells us about the results of our sample.
- What percentage of many samples would have a  $\hat{p}$  larger than 0.37? (Use the 68–95–99.7 rule.) Explain in simple language why this percentage is the probability of an outcome larger than 0.37.
- What is the probability that your sample will have a  $\hat{p}$  less than 0.33?
- Use Rule D: what is the probability that your sample result will be either less than 0.33 or greater than 0.35?

**18.26 Generating a sampling distribution.** Let us illustrate the idea of a sampling distribution in the case of a very small sample from a very small population. The population is the scores of 10 students on an exam:

Student:	0	1	2	3	4
Score:	82	62	80	58	72

---

Student:	5	6	7	8	9
Score:	73	65	66	74	62

The parameter of interest is the mean score in this population. The sample is an SRS of size  $n = 4$  drawn from the population. Because the students are labeled 0 to 9, a single random digit from Table A chooses one student for the sample.

(a) Find the mean of the 10 scores in the population. This is the population mean.

(b) Use Table A to draw an SRS of size 4 from this population. Write the four scores in your sample and calculate the mean  $\bar{x}$  of the sample scores. This statistic is an estimate of the population mean.

(c) Repeat this process 10 times using different parts of Table A. Make a histogram of the 10 values of  $\bar{x}$ . You are constructing the sampling distribution of  $\bar{x}$ . Is the center of your histogram close to the population mean you found in part (a)?

### 18.27 Odds and personal probability.

One way to determine your personal probability about an event is to ask what you would consider a fair bet that the event will occur. Suppose in August 2015 you believed it fair that a bet of \$2 should win \$7 if the Green Bay Packers win Super Bowl 50.

- What does this bet say you believe the odds to be that the Green Bay Packers will win Super Bowl 50?
- Convert the odds in part (a) to a probability. This would be your personal probability that the Green Bay Packers will win Super Bowl 50.

**18.28 The addition rule.** Probability rule D states: *If two events have no outcomes in common, the probability that one or the other occurs is the sum of their individual probabilities.* This is sometimes called the addition rule for disjoint events. A more general form of the addition rule is: *the probability that one or the other of two events occurs is the sum of their individual probabilities minus the probability that both occur.* To verify

this rule, suppose you roll two casino dice as in Example 2. Refer to the outcomes in Figure 18.1 to answer the following.

- (a) How many of the outcomes in Figure 18.1 have the sum of the spots on the up-faces equal to 6? What is the probability that the sum of the spots on the up-faces is 6?
- (b) How many of the outcomes in Figure 18.1 have at least one of the up-faces showing a single spot? What is the probability that at least one of the up-faces has only a single spot?
- (c) How many of the outcomes in Figure 18.1 have the sum of the spots on the up-faces equal to 6 and at least one of the up-faces showing a single spot? What is the probability that the sum of the spots on the up-faces is 6 and at least one of the up-faces has only a single spot?
- (d) How many of the outcomes in Figure 18.1 have either the sum of

the spots on the up-faces equal to 6 or have at least one of the up-faces showing a single spot? What is the probability that the sum of the spots on the up-faces is either a 6 or at least one of the up-faces has a single spot?

- (e) The addition rule says that your answer to part (d) should be equal to the sum of your answers to parts (a) and (b) minus your answer to part (c). Verify that this is the case.

*Note:* The outcomes you were asked to count in part (c) are among those counted in parts (a) and (b). When we combine the outcomes in parts (a) and (b), we “double count” the outcomes in part (c). One of these “double counts” must be eliminated so that the combination corresponds to the outcomes in part (d). This is the reason that, in the addition rule, you subtract the probability that both occur from the sum of their individual probabilities.



### EXPLORING THE WEB

*Follow the QR code to access exercises.*

