Simulation

Daikusan/Getty Images

CASE STUDY In a horse race, the starting position can affect the outcome. It is advantageous to have a starting position that is near the inside of the track. To ensure fairness, starting position is determined by a random draw before the race. All positions are equally likely, so no horse has an advantage.

During the summer and autumn of 2007, the members of the Ohio Racing Commission noticed that one trainer appeared to have an unusually good run of luck. In 35 races, the horses of this trainer received one of the three inner positions 30 times. The number of horses in a race ranges from 6 to 10, but most of the time the number of entries is 9. Thus, for most races the chance of getting one of the three inside positions is 1-in-3.

The Ohio Racing Commission believed that the trainer's run of luck was too good to be true. A mathematician can show that the chance of receiving one of the three inside positions at least 30 times in 35 races is very small. The Ohio Racing Commission, therefore, suspected that cheating had occurred. But the trainer had entered horses in nearly 1000 races over several years. Perhaps it was inevitable that, at some time over the course of these nearly 1000 races, the trainer would have a string of 35 races in which he received one of the three inside positions at least 30 times. It came to the attention of the Ohio Racing Commission only because it was one of those seemingly surprising coincidences discussed in Chapter 17. Perhaps the accusation of cheating was unfounded.

Calculating the probability that in a sequence of 1000 races with varying numbers of horses, there would occur a string of 35 consecutive races in which one would receive one of the three inside positions at least 30 times is very difficult. How to go about determining this probability is the subject of this chapter. By the end of the chapter, you will be able to describe how to find this probability.

Where do probabilities come from?

The probabilities of heads and tails in tossing a coin are very close to 1-in-2. In principle, these probabilities come from *data* on many coin tosses. Joe's personal probabilities for the winner of next year's Super Bowl come from Joe's own *individual judgment*. What about the probability that we get a run of three straight heads somewhere in 10 tosses of a coin? We can find this probability by *calculation from a model* that describes tossing coins. That is, once we have used data to give a probability model for the random fall of coins, we don't have to go back to the beginning every time we want the probability of a new event.

The big advantage of probability models is that they allow us to calculate the probabilities of complicated events starting from an assignment of probabilities to simple events like "heads on one toss." This is true whether the model reflects probabilities from data or personal probabilities. Unfortunately, the math needed to do probability calculations is often tough. Technology rides to the rescue: once we have a probability model, we can use a computer to *simulate* many repetitions. This is easier than math and much faster than actually running many repetitions in the real world. You might compare finding probabilities by simulation to practicing flying in a computer-controlled flight simulator. Both kinds of simulation are in wide use. Both have similar drawbacks: they are only as good as the model you start with. Flight simulators use a software model of how an airplane reacts. Simulations of probabilities use a probability model. We set the model in motion by using our old friends the random digits from Table A.

Simulation

Using random digits from a table or from computer software to imitate chance behavior is called **simulation**.

We look at simulation partly because it is how scientists and engineers really do find probabilities in complex situations. Simulations are used to develop strategies for reducing waiting times in lines to speak to a teller at banks, in lines to check in at airports, and in lines to vote during elections. Simulations are used to study the effects of changes in greenhouse gases on the climate. Simulations are used to study the effects of catastrophic events, such as the failure of a nuclear power plant, the effects on a structure of the explosion of a nuclear device, or the progression of a deadly, infectious disease in a densely populated city.

We also look at simulation because simulation forces us to think clearly about probability models. We'll do the hard part (setting up the model) and

leave the easy part (telling a computer to do 10,000 repetitions) to those who really need the right probability at the end.

Simulation basics

Simulation is an effective tool for finding probabilities of complex events once we have a trustworthy probability model. We can use random digits to simulate many repetitions quickly. The proportion of repetitions on which an event occurs will eventually be close to its probability, so simulation can give good estimates of probabilities. The art of simulation is best learned from a series of examples.

EXAMPLE 1 Doing a simulation

Toss a coin 10 times. What is the probability of a run of at least three consecutive heads or three consecutive tails?

Step 1. Give a probability model. Our model for coin tossing has two parts:

- Each toss has probabilities 0.5 for a head and 0.5 for a tail.
- Tosses are *independent* of each other. That is, knowing the outcome of one toss does not change the probabilities for the outcomes of any other toss.

Step 2. Assign digits to represent outcomes. Digits in Table A of random digits will stand for the outcomes, in a way that matches the probabilities from Step 1. We know that each digit in Table A has probability 0.1 of being any one of 0, 1, 2, 3, 4, 5, 6, 7, 8, or 9 and that successive digits in the table are independent. Here is one assignment of digits for coin tossing:

- One digit simulates one toss of the coin.
- Odd digits represent heads; even digits represent tails.

This works because the five odd digits give probability 5/10 to heads (but any other assignment where half the digits represent heads is equally good). Successive digits in the table simulate independent tosses.

Step 3. Simulate many repetitions. Ten digits simulate 10 tosses, so looking at 10 consecutive digits in Table A simulates one repetition. Read many groups of 10 digits from the table to simulate many repetitions. Be sure to keep track of whether or not the event we want (a run of three heads or three tails) occurs on each repetition.

Really random digits
For purists, the RAND Corporation long ago published a book titled *One Million Random Digits.* The book lists 1,000,000 digits that were produced by a very elaborate physical randomization and really are random. An employee of RAND once told one of us that this is not the most boring book that RAND has ever published…

Here are the first three repetitions, starting at line 101 in Table A. We have underlined all runs of three or more heads or tails.

	Repetition 1	**Repetition 2**	**Repetition 3**
Digits	1 9 2 2 3 9 5 0 3 4	0 5 7 5 6 2 8 7 1 3	9 6 4 0 9 1 2 5 3 1
Heads/tails	H H T T H H H T H T	T H H H T T T H H H	H T T T H H T H H H
Run of 3?	YES	YES	YES

Continuing in Table A, we did 25 repetitions; 23 of them did have a run of three or more heads or tails. So we estimate the probability of a run by the proportion

$$\text{estimated probability} = \frac{23}{25} = 0.92$$

Of course, 25 repetitions are not enough to be confident that our estimate is accurate. Now that we understand how to do the simulation, we can tell a computer to do many thousands of repetitions. A long simulation (or hard mathematics) finds that the true probability is about 0.826. Most people think runs are somewhat unlikely, so even our short simulation challenges our intuition by showing that runs of three occur most of the time in 10 tosses.

Once you have gained some experience in simulation, setting up the probability model (Step 1) is usually the hardest part of the process. Although coin tossing may not fascinate you, the model in Example 1 is typical of many probability problems because it consists of *independent trials* (the tosses) all having the *same possible outcomes* with the *same probabilities*. Shooting 10 free throws and observing the sexes of 10 children have similar models and are simulated in much the same way. The new part of the model is independence, which simplifies our work because it allows us to simulate each of the 10 tosses in exactly the same way.

Independence

Two random phenomena are **independent** if knowing the outcome of one does not change the probabilities for outcomes of the other.

Independence, like all aspects of probability, can be verified only by observing many repetitions. It is plausible that repeated tosses of a coin are independent (the coin has no memory), and observation shows that they are. It seems less plausible that successive shots by a basketball
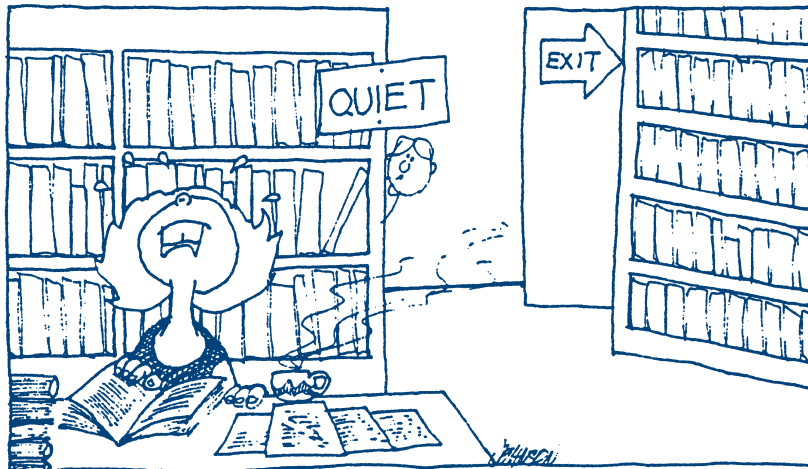
"I've had it! Simulated wood, simulated leather, simulated coffee, and now simulated probabilities!"

player are independent, but observation shows that they are at least very close to independent.

  Step 2 (assigning digits) rests on the properties of the random digit table. Here are some examples of this step.

EXAMPLE 2 Assigning digits for simulation

In the United States, the Age Discrimination Employment Act (ADEA) forbids age discrimination against people who are age 40 and over. Terminating an "unusually" large proportion of employees age 40 and over can trigger legal action. Simulation can help determine what might be an "unusual" pattern of terminations. How might we set up such a simulation?

(a) Choose one employee at random from a group of which 40% are age 40 and over. One digit simulates one employee:

$$0, 1, 2, 3 = \text{age 40 and over}$$

$$4, 5, 6, 7, 8, 9 = \text{under age 40}$$

(b) Choose one employee at random from a group of which 43% are age 40 and over. Now *two* digits simulate one person:

$$00, 01, 02, \ldots, 42 = \text{age 40 and over}$$

$$43, 74, 75, \ldots, 99 = \text{under age 40}$$

We assigned 43 of the 100 two-digit pairs to "age 40 and over" to get probability 0.43. Representing "age 40 and over" by $01, 02, \ldots, 43$ and "under age 40" by $44, 45, \ldots, 99, 00$ would also be correct.

(c) Choose one employee at random from a group of which 30% are age 40 and over and have no plans to retire, 10% are age 40 and over and plan to retire in the next few months, and 60% are under age 40. There are now three possible outcomes, but the principle is the same. One digit simulates one person:

0, 1, 2 = age 40 and over and have no plans to retire

3 = age 40 and over and plan to retire in the next few months

4, 5, 6, 7, 8, 9 = under age 40

19.1 Selecting cards at random. In a standard deck of 52 cards, there are 13 spades, 13 hearts, 13 diamonds, and 13 clubs. How would you assign digits for a simulation to determine the suit (spades, hearts, diamonds, or clubs) of a card chosen at random from a standard deck of 52 cards?

Thinking about independence

Before discussing more elaborate simulations, it is worth discussing the concept of independence further. We said earlier that independence can be verified only by observing many repetitions of random phenomena. It is probably more accurate to say that a *lack* of independence can be verified only by observing many repetitions of random phenomena. How does one recognize that two random phenomena are not independent? For example, how can we tell if tosses of a fair coin (that is, one for which the probability of a head is 0.5 and the probability of a tail is 0.5) are not independent?

One approach might be to apply the definition of "independence." For a sequence of tosses of a fair coin, one could compute the proportion of times in the sequence that a toss is followed by the same outcome—in other words, the frequency with which a head is followed by a head or a tail is followed by a tail. This proportion should be close to 0.5 if tosses are independent (knowing the outcome of one toss does not change the probabilities for outcomes of the next) and if many tosses have been observed.

EXAMPLE 3 Investigating independence

Suppose we tossed a fair coin 15 times and obtained the following sequence of outcomes:

Toss:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Outcome:	H	H	H	T	H	H	T	T	T	T	T	H	H	T	T

> For the first 14 tosses, the following toss is the same nine times. The proportion of times a toss is followed by the same outcome is
>
> $$\text{proportion} = \frac{9}{14} = 0.64$$
>
> For so few tosses, this would not be considered a large departure from 0.5.

Unfortunately, if the proportion of times a head is followed by a head or a tail is followed by a tail is close to 0.5, this does not necessarily imply that the tosses are independent. For example, suppose that instead of tossing the coin we simply placed the coin heads up or tails up according to the following pattern:

Trial:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Outcome:	H	H	T	T	H	H	T	T	H	H	T	T	H	H	T

We begin with two heads, followed by two tails, followed by two heads, etc. If we know the previous outcomes, we know exactly what the next outcome will be. Successive outcomes are not independent. However, looking at the first 14 outcomes, the proportion of times in this sequence that a head is followed by a head or a tail is followed by a tail is

$$\text{proportion} = \frac{7}{14} = 0.5$$

Thus, our approach can help us recognize when independence is lacking, but not when independence is present.

Another method for assessing independence is based on the concept of correlation, which we discussed in Chapter 14. If two random phenomena have numerical outcomes, and we observe both phenomena in a sequence of n trials, we can compute the correlation for the resulting data. If the random phenomena are independent, there will be no straight-line relationship between them, and the correlation should be close to 0.

It is not necessarily true that two random phenomena are independent if their correlation is 0. In Exercise 14.24 (page 334), there is a clear curved relationship between speed and mileage, but the correlation is 0. Independence implies no relationship at all, but correlation measures the strength of only a straight-line relationship.

Because independence implies no relationship, we would expect to see no overall pattern in a scatterplot of the data if the variables are independent. Looking at scatterplots is another method for determining if independence is lacking.

Was he good or was he lucky? When a baseball player hits .300, everyone applauds. A .300 hitter gets a hit in 30% of times at bat. Could a .300 year just be luck? Typical major leaguers bat about 500 times a season and hit about .260. A hitter's successive tries seem to be independent. From this model, we can calculate or simulate the probability of hitting .300. It is about 0.025. Out of 100 run-of-the-mill major league hitters, two or three each year will bat .300 just because they were lucky.

Many methods for assessing independence exist. For example, if trials are not independent and, say, tossing a head increases the probability that the next toss is also a head, then in a sequence of tosses, we might expect to see unusually long runs of heads. We mentioned this idea of unusually long runs in Example 5 (page 410) of Chapter 17. Unusually long runs of made free throws would be expected if a basketball player has a "hot hand." However, careful study has shown that runs of baskets made or missed are no more frequent in basketball than would be expected if each shot is independent of the player's previous shots.

More elaborate simulations

The building and simulation of random models constitute a powerful tool of contemporary science, but a tool that can be understood without advanced mathematics. What is more, doing a few simulations will increase your understanding of probability more than many pages of our prose. Having in mind these two goals of understanding simulation for itself and understanding simulation to understand probability, let us look at two more elaborate examples. The first still has independent trials, but there is no longer a fixed number of trials as there was when we tossed a coin 10 times.

EXAMPLE 4 We want a girl

A couple plan to have children until they have a girl or until they have three children, whichever comes first. What is the probability that they will have a girl among their children?

Step 1. The probability model is like that for coin tossing:

- Each child has probability 0.49 of being a girl and 0.51 of being a boy. (Yes, more boys than girls are born. Boys have higher infant mortality, so the sexes even out soon.)
- The sexes of successive children are independent.

Step 2. Assigning digits is also easy. Two digits simulate the sex of one child. We assign 49 of the 100 pairs to "girl" and the remaining 51 to "boy":

$$00, 01, 02, \ldots, 48 = \text{girl}$$
$$49, 50, 51, \ldots, 99 = \text{boy}$$

Step 3. To simulate one repetition of this childbearing strategy, read pairs of digits from Table A until the couple have either a girl or three children. The number of pairs needed to simulate one repetition depends on how quickly the couple get a girl. Here are 10 repetitions, simulated using line 130 of Table A. To interpret the pairs of digits, we have written G for girl and B for boy under them, have added space to separate repetitions, and under each repetition have written "+" if a girl was born and "−" if not.

6905	16	48	17	8717	40	9517	845340	648987	20
B G	G	G	G	B G	G	B G	B B G	B B B	G
+	+	+	+	+	+	+	+	−	+

In these 10 repetitions, a girl was born nine times. Our estimate of the probability that this strategy will produce a girl is, therefore,

$$\text{estimated probability} = \frac{9}{10} = 0.9$$

Some mathematics shows that, if our probability model is correct, the true probability of having a girl is 0.867. Our simulated answer came quite close. Unless the couple are unlucky, they will succeed in having a girl.

Our final example has stages that are *not independent*. That is, the probabilities at one stage depend on the outcome of the preceding stage.
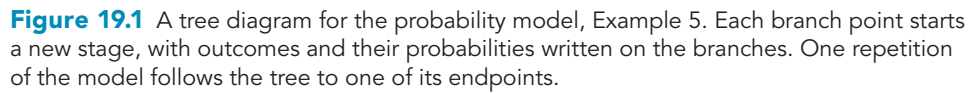
EXAMPLE 5 A kidney transplant

Morris's kidneys have failed and he is awaiting a kidney transplant. His doctor gives him this information for patients in his condition: 90% survive the transplant operation, and 10% die. The transplant succeeds in 60% of those who survive, and the other 40% must return to kidney dialysis. The proportions who survive for at least five years are 70% for those with a new kidney and 50% for those who return to dialysis. Morris wants to know the probability that he will survive for at least five years.

BSIP/Phototake

Step 1. The **tree diagram** in Figure 19.1 organizes this information to give a probability model in graphical form. The tree shows the three stages and the possible outcomes and probabilities at each stage. Each path through the tree leads to either survival for five years or to death in less than five years. To simulate Morris's fate, we must simulate each

Figure 19.1 A tree diagram for the probability model, Example 5. Each branch point starts a new stage, with outcomes and their probabilities written on the branches. One repetition of the model follows the tree to one of its endpoints.

of the three stages. The probabilities at Stage 3 depend on the outcome of Stage 2.

Step 2. Here is our assignment of digits to outcomes:

Stage 1:

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

Stage 2:

$$0, 1, 2, 3, 4, 5 = \text{transplant succeeds}$$
$$6, 7, 8, 9 = \text{return to dialysis}$$

Stage 3 with kidney:

$$0, 1, 2, 3, 4, 5, 6 = \text{survive five years}$$
$$7, 8, 9 = \text{die}$$

Stage 3 with dialysis:

$$0, 1, 2, 3, 4 = \text{survive five years}$$
$$5, 6, 7, 8, 9 = \text{die}$$

The assignment of digits at Stage 3 depends on the outcome of Stage 2. That's lack of independence.

Step 3. Here are simulations of several repetitions, each arranged vertically. We used random digits from line 110 of Table A.

	Repetition 1	Repetition 2	Repetition 3	Repetition 4
Stage 1	3 Survive	4 Survive	8 Survive	9 Survive
Stage 2	8 Dialysis	8 Dialysis	7 Dialysis	1 Kidney
Stage 3	4 Survive	4 Survive	8 Die	8 Die

Morris survives five years in two of our four repetitions. Now that we understand how to arrange the simulation, we should turn it over to a computer to do many repetitions. From a long simulation or from mathematics, we find that Morris has probability 0.558 of living for at least five years.

19.2 Selecting cards at random. In a standard deck of 52 cards, there are 13 spades, 13 hearts, 13 diamonds, and 13 clubs. Carry out a simulation to determine the probability that, when two cards are selected together (that is, without replacement) at random from a standard deck, both have the same suit. Follow the steps given in Example 5 to help set up the simulation. Do your simulation 10 times and use the result to estimate the probability.

NOW IT'S YOUR TURN

STATISTICS IN SUMMARY

Chapter Specifics

- We can use random digits to **simulate** random outcomes if we know the probabilities of the outcomes. Use the fact that each random digit has probability 0.1 of taking any one of the 10 possible digits and that all digits in the random number table are **independent** of each other.

- To simulate more complicated random phenomena, string together simulations of each stage. A common situation is several independent trials with the same possible outcomes and probabilities on each trial. Other simulations may require varying numbers of trials or different probabilities at each stage or may have stages that are not independent so that the probabilities at some stage depend on the outcome of earlier stages.

- The key to successful simulation is thinking carefully about the probability model. A **tree diagram** can be helpful by giving the probability model in graphical form.

In Chapter 18, we discussed probability models and the basic rules of probability. These allow us to compute the probabilities of simple events, but the math needed to find the probabilities of complicated events is often tough. In this chapter, we learn how we can use simulations to determine the probabilities of complicated events. The underlying idea, introduced in Chapter 17, is that probability is the long-run proportion of times an event occurs. Simulating such an event many, many times using technology allows us to estimate this long-run proportion.

CASE STUDY EVALUATED Look again at the Case Study that opened the chapter. For simplicity, assume that there are always nine horses in a race so that the probability of receiving one of the three inside positions is 1-in-3 if assignments are random. Use what you have learned in this chapter to describe how you would do a simulation to estimate the probability that, somewhere in a sequence of 1000 races, a string of 35 consecutive races would occur in which at least 30 times one of the three inside positions was assigned. Begin by describing how you would simulate one case of 1000 races and, for these races, how you would determine whether somewhere in the 1000 races there was a string of 35 consecutive races in which at least 30 times one of the three inside positions was assigned. Do not try to carry out the simulation. This would be very time-consuming and is best left to a computer. (In fact, the Ohio Racing Commission hired a statistician to compute the probability that in a sequence of 1000 races with varying numbers of horses, there would occur a string of 35 consecutive races in which one would receive one of the three inside positions at least 30 times. The statistician used simulation to estimate this probability.)

LaunchPad Online Resources
macmillan learning

- The StatBoards video *Simulation Basics* discusses the basics of designing and running a simulation in the context of an example.

CHECK THE BASICS

For Exercise 19.1, see page 450; for Exercise 19.2, see page 455.

19.3 Simulations. To simulate random outcomes, we need to know
(a) the probabilities of the outcomes.
(b) whether the probabilities are personal probabilities.
(c) both the mean and standard deviation so we can use the appropriate Normal curve.
(d) that the random outcome has a probability of 0.1 so we can use a table of random digits.

19.4 A simulation. To simulate the toss of a fair coin (the probability of heads and tails are both 0.5) using a table of random digits,
(a) assign the digits 0, 1, 2, 3, and 4 to represent heads and the digits 5, 6, 7, 8, 9 to represent tails.

(b) assign the digits 0, 2, 4, 6, and 8 to represent heads and the digits 1, 3, 5, 7, and 9 to represent tails.

(c) assign the digits 0, 1, 5, 8, and 9 to represent heads and the digits 2, 3, 4, 6, and 7 to represent tails.

(d) use any of the above. All are correct.

19.5 Independence. Two random phenomena are independent if

(a) knowing that one of the outcomes has occurred means the other cannot occur.

(b) knowing the outcomes of one does not change the probabilities for outcomes of the other.

(c) both have the same probability of occurring.

(d) both have different and unrelated probabilities of occurring.

19.6 Elaborate simulations. A tree diagram

(a) was originally used by biologists for simulations involving trees.

(b) is used to determine if two random phenomena are independent.

(c) is used when two random phenomena are independent.

(d) specifies a probability model in graphical form.

19.7 Elaborate simulations. The key to successful simulation is

(a) keeping the tree diagram as simple as possible.

(b) thinking carefully about the probability model for the simulation.

(c) using as few trials as possible so that the chance of an incorrect trial is kept small.

(d) using *all* the digits in a random digits table.

CHAPTER 19 EXERCISES

19.8 Which party does it better? An opinion poll selects adult Americans at random and asks them, "Which political party, Democratic or Republican, do you think is better able to keep the U.S. prosperous?" Explain carefully how you would assign digits from Table A to simulate the response of one person in each of the following situations.

(a) Of all adult Americans, 50% would choose the Democrats and 50% the Republicans.

(b) Of all adult Americans, 40% would choose the Democrats and 60% the Republicans.

(c) Of all adult Americans, 40% would choose the Democrats, 40% would

choose the Republicans, and 20% are undecided.

(d) Of all adult Americans, 44% would choose the Democrats, 46% the Republicans, and 10% are undecided. (These were the percentages in a September 2015 Gallup Poll.)

19.9 A small opinion poll. Suppose that 80% of a university's students favor increasing the number of police patrolling campus at night. You ask 10 students chosen at random. What is the probability that all 10 favor increasing the number of police patrolling campus at night?

(a) Give a probability model for asking 10 students independently of each other.

(b) Assign digits to represent the answers Yes and No.

(c) Simulate 25 repetitions, starting at line 129 of Table A. What is your estimate of the probability?

19.10 Basic simulation. Use Table A to simulate the responses of 10 independently chosen adults in each of the four situations of Exercise 19.8.

(a) For situation (a), use line 110.

(b) For situation (b), use line 111.

(c) For situation (c), use line 112.

(d) For situation (d), use line 113.

19.11 Simulating an opinion poll. A Gallup Poll on Presidents Day 2011 interviewed a random sample of 1015 adult Americans. Those in the sample were asked which U.S. president they regarded as the greatest. The poll showed that about 20% of adult Americans regarded Ronald Reagan as the greatest U.S. president. Suppose that this is exactly true. Choosing an adult American at random then has probability 0.2 of getting one who would identify Ronald Reagan as the greatest president. If we interview adult Americans separately, we can assume that their responses are independent. We want to know the probability that a simple random sample of 100 adult Americans will contain at least 25 who say that Ronald Reagan is the greatest U.S. president. Explain carefully how to do this simulation and simulate *one* repetition of the poll using line 112 of Table A. How many of the 100 adult Americans said that Ronald Reagan is the greatest U.S. president? Explain how you would estimate the probability by simulating many repetitions.

19.12 Course grades. Choose a student at random from all who took the large accelerated introductory statistics at Hudson River College in the last 10 years. The probabilities for the student's grade are

Grade:	A	B	C	D or F
Probability:	0.1	0.3	0.4	?

(a) What must be the probability of getting a D or an F?

(b) To simulate the grades of randomly chosen students, how would you assign digits to represent the four possible outcomes listed?

19.13 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. Probabilities for the outcomes are

Difficulty:	Very difficult	Somewhat difficult
Probability:	0.1	0.3

Difficulty:	Somewhat easy	Very easy
Probability:	0.4	?

(a) What must be the probability that a randomly chosen first-year student says that it was very easy to to adjust to the academic demands of college.?

(b) To simulate the responses of randomly chosen first-year students, how would you assign digits to represent the four possible outcomes listed?

19.14 More on course grades. In Exercise 19.12, you explained how

to simulate the grade of a randomly chosen student who took the accelerated statistics course in the last 10 years. Suppose you select five students at random who took the course in the last 10 years. Use simulation to estimate the probability that all five got a C or better in the course. (Simulate 20 repetitions and assume the student grades are independent of each other.)

19.15 More on the demands of college. In Exercise 19.13, you explained how to simulate the response of a randomly chosen first-year college student to the question of how difficult it was to adjust to the academic demands of college. The Random Committee decides to choose eight students at random and provide them with training on adjusting to the academic demands of college. What is the probability that at least five of the eight students chosen were among those who reported that it was very difficult or somewhat difficult to adjust to the academic demands of college? Simulate 10 repetitions of the committee's choices to estimate this probability.

19.16 Curry's three-point shooting. The basketball player Stephen Curry makes about 44% of his three-point shots over an entire season. Take his probability of a success to be 0.44 on each shot. Using line 122 of Table A, simulate 25 repetitions of his performance in a game in which he shoots 10 three-point shots.

(a) Estimate the probability that Stephen makes at least half of his three-point shots.

(b) Examine the sequence of hits and misses in your 25 repetitions. How long was the longest run of shots made?

19.17 Elena's free throws. Elena Delle Donne of the WNBA Chicago Sky makes 94% of her free throws. In an important game, she shoots a pair of free throws shots late in the game and misses both. The fans think she was nervous, but the misses may simply be chance. Let's shed some light by estimating a probability.

(a) Describe how to simulate a single free throw if the probability of making each free throw is 0.94. Then describe how to simulate two independent free throws.

(b) Simulate 50 repetitions of the two free throws and record the number missed on each repetition. Use Table A, starting at line 125. What is the approximate probability that Elena will miss both free throws?

19.18 Repeating an exam. Elaine is enrolled in a self-paced course that allows three attempts to pass an examination on the material. She does not study and has probability 2-in-10 of passing on any one attempt by luck. What is Elaine's probability of passing in three attempts? (Assume the attempts are independent because she takes a different examination on each attempt.)

(a) Explain how you would use random digits to simulate one attempt at the exam.

(b) Elaine will stop taking the exam as soon as she passes. (This is

much like Example 4.) Simulate 50 repetitions, starting at line 120 of Table A. What is your estimate of Elaine's probability of passing the exam?

(c) Do you think the assumption that Elaine's probability of passing the exam is the same on each trial is realistic? Why?

19.19 A better model for repeating an exam. A more realistic probability model for Elaine's attempts to pass an exam in the previous exercise is as follows. On the first try she has probability 0.2 of passing. If she fails on the first try, her probability on the second try increases to 0.3 because she learned something from her first attempt. If she fails on two attempts, the probability of passing on a third attempt is 0.4. She will stop as soon as she passes. The course rules force her to stop after three attempts in any case.

(a) Make a tree diagram of Elaine's progress. Notice that she has different probabilities of passing on each successive try.

(b) Explain how to simulate one repetition of Elaine's tries at the exam.

(c) Simulate 50 repetitions and estimate the probability that Elaine eventually passes the exam. Use Table A, starting at line 130.

19.20 Gambling in ancient Rome. Tossing four astragali was the most popular game of chance in Roman times. Many throws of a present-day sheep's astragalus show that the approximate probability distribution for the four sides of the bone that can land uppermost are

Outcome	Probability
Narrow flat side of bone	1/10
Broad concave side of bone	4/10
Broad convex side of bone	4/10
Narrow hollow side of bone	1/10

The best throw of four astragali was the "Venus," when all four uppermost sides were different.

(a) Explain how to simulate the throw of a single astragalus. Then explain how to simulate throwing four astragali independently of each other.

(b) Simulate 25 throws of four astragali. Estimate the probability of throwing a Venus. Be sure to say what part of Table A you used.

19.21 The Asian stochastic beetle. We can use simulation to examine the fate of populations of living creatures. Consider the Asian stochastic beetle. Females of this insect have the following pattern of reproduction:

· 20% of females die without female offspring, 30% have one female offspring, and 50% have two female offspring.

· Different females reproduce independently.

What will happen to the population of Asian stochastic beetles: will they increase rapidly, barely hold their own, or die out? It's enough to look at the female beetles, as long as there are some males around.

(a) Assign digits to simulate the offspring of one female beetle.

(b) Make a tree diagram for the female descendants of one beetle through

three generations. The second generation, for example, can have 0, 1, or 2 females. If it has 0, we stop. Otherwise, we simulate the offspring of each second-generation female. What are the possible numbers of beetles after three generations?

(c) Use line 105 of Table A to simulate the offspring of five beetles to the third generation. How many descendants does each have after three generations? Does it appear that the beetle population will grow?

19.22 Two warning systems. An airliner has two independent automatic systems that sound a warning if there is terrain ahead (that means the airplane is about to fly into a mountain). Neither system is perfect. System A signals in time with probability 0.9. System B does so with probability 0.8. The pilots are alerted if either system works.

(a) Explain how to simulate the response of System A to terrain.

(b) Explain how to simulate the response of System B.

(c) Both systems are in operation simultaneously. Draw a tree diagram with System A as the first stage and System B as the second stage. Simulate 100 trials of the reaction to terrain ahead. Estimate the probability that a warning will sound. The probability for the combined system is higher than the probability for either A or B alone.

19.23 Playing craps. The game of craps is played with two dice. The player rolls both dice and wins immediately if the outcome (the sum of the faces) is 7 or 11. If the outcome

is 2, 3, or 12, the player loses immediately. If he rolls any other outcome, he continues to throw the dice until he either wins by repeating the first outcome or loses by rolling a 7.

(a) Explain how to simulate the roll of a single fair die. (*Hint*: Just use digits 1 to 6 and ignore the others.) Then explain how to simulate a roll of two fair dice.

(b) Draw a tree diagram for one play of craps. In principle, a player could continue forever, but stop your diagram after four rolls of the dice. Use Table A, beginning at line 114, to simulate plays and estimate the probability that the player wins.

19.24 The airport van. Your company operates a van service from the airport to downtown hotels. Each van carries seven passengers. Many passengers who reserve seats don't show up—in fact, the probability is 0.2 that a randomly chosen passenger will fail to appear. Passengers' behaviors are independent. If you allow nine reservations for each van, what is the probability that more than seven passengers will appear? Do a simulation to estimate this probability.

19.25 A multiple-choice exam. Matt has lots of experience taking multiple-choice exams without doing much studying. He is about to take a quiz that has 10 multiple-choice questions, each with four possible answers. Here is Matt's personal probability model. He thinks that in 75% of questions he can eliminate one answer as obviously wrong; then he guesses

from the remaining three. He then has probability 1-in-3 of guessing the right answer. For the other 25% of questions, he must guess from all four answers, with probability 1-in-4 of guessing correctly.

(a) Make a tree diagram for the outcome of a single question. Explain how to simulate Matt's success or failure on one question.

(b) Questions are independent. To simulate the quiz, just simulate 10 questions. Matt needs to get at least five questions right to pass the quiz. You could find his probability of passing by simulating many tries at the quiz, but we ask you to simulate just one try. Did Matt pass this quiz?

19.26 More on the airport van. Let's continue the simulation of Exercise 19.24. You have a backup van, but it serves several stations. The probability that it is available to go to the airport at any one time is 0.6. You want to know the probability that some passengers with reservations will be left stranded because the first van is full and the backup van is not available. Draw a tree diagram with the first van (full or not) as the first stage and the backup (available or not) as the second stage. In Exercise 19.24, you simulated a number of repetitions of the first stage. Add simulations of the second stage whenever the first van is full. What is your estimate of the probability of stranded passengers?

19.27 The birthday problem. A famous example in probability theory shows that the probability that at least two people in a room have the same birthday is already greater than 1-in-2 when 23 people are in the room. The probability model is

- The birth date of a randomly chosen person is equally likely to be any of the 365 dates of the year.
- The birth dates of different people in the room are independent.

To simulate birthdays, let each three-digit group in Table A stand for one person's birth date. That is, 001 is January 1 and 365 is December 31. Ignore leap years and skip groups that don't represent birth dates. Use line 139 of Table A to simulate birthdays of randomly chosen people until you hit the same date a second time. How many people did you look at to find two with the same birthday?

With a computer, you could easily repeat this simulation many times. You could find the probability that at least two out of 23 people have the same birthday, or you could find the expected number of people you must question to find two with the same birthday. These problems are a bit tricky to do by math, so they show the power of simulation.

19.28 The multiplication rule. Here is another basic rule of probability: *if several events are independent, the probability that all of the events happen is the product of their individual probabilities*. We know, for example, that a child has probability 0.49 of being a girl and probability 0.51 of being a boy and that the sexes of successive children are independent. So the probability that a couple's two children are two girls is $(0.49)(0.49) = 0.2401$. You can use

this multiplication rule to calculate the probability that we simulated in Example 4.

(a) Write down all eight possible arrangements of the sexes of three children, for example, BBB and BBG. Use the multiplication rule to find the probability of each outcome. Check your work by verifying that your eight probabilities add to 1.

(b) The couple in Example 4 plan to stop when they have a girl or to stop at three children even if all are boys. Use your work from part (a) to find the probability that they get a girl.

EXPLORING THE WEB

Follow the QR code to access exercises.