PART 2

Probability, Probability Distributions, and Sampling Distributions

5.1 Practicing the Basics

5.1 Probability Explain what is meant by the long-run relative frequency definition of probability.

5.2 Minesweeper The objective of the game Minesweeper is to clear a field without detonating any mines or bombs. Your friend claims that his rate of completing the game successfully is 90%.

- **a.** You decide to challenge your friend. He makes 10 attempts to complete the game, but is successful in only 7 of them. Does this mean that your friend's claim is wrong?
- **b.** If your friend's claim was actually true, what would you have to do to ensure that the cumulative proportion of his successful attempts to complete the game falls very close to 0.9?

5.3 Counselor availability You visit your counselor's office at 10 randomly chosen times, and he is not available at any of those times. Does this mean that the probability of your counselor being available at his office for students equals 0? Explain.

5.4 Airline accident deaths Airplane safety has been improving over the years. From 2000 to 2010, the average number of global airline deaths per year was over 1000, even when excluding the nearly 3000 deaths in the United States on September 11, 2001. The number of global airline deaths declined in 2011, again in 2012, and then hit a low of only 265 in 2013. In 2013, there were a total of 825 million passengers globally. *Sources:* en.wikipedia.org and www.transtats.bts.gov1.

- **a.** Can you consider the 2013 data as a long run or short run of trials? Explain.
- **b.** Estimate the probability of dying on a flight in 2013. (Note, the probability of dying from a 1000-mile automobile trip is about 1 in 42,000 by contrast.)
- **c.** Raul is considering flying on an airplane. He noticed that over the past two months, there have been no fatal airplane crashes around the world. This raises his concern about flying because the airlines are "due for an accident." Comment on his reasoning.

5.5 World Cup 2014 The powerrank.com website (http://thepowerrank.com/2014/06/06/world-cup-2014-win-probabilities-from-the-power-rank/) listed the probability of each team to win the 2014 World Cup in soccer as follows:

1. Brazil, 35.9%.
2. Argentina, 10.0%.
3. Spain, 8.9%.
4. Germany, 7.4%.
5. Netherlands, 5.7%.
6. Portugal, 3.9%.
7. France, 3.4%.
8. England, 2.8%.
9. Uruguay, 2.5%.
10. Mexico, 2.5%.
11. Italy, 2.3%.
12. Ivory Coast, 2.0%.
13. Colombia, 1.5%.
14. Russia, 1.5%.
15. United States, 1.1%.
16. Chile, 1.0%.
17. Croatia, 0.9%.
18. Ecuador, 0.8%.
19. Nigeria, 0.8%.
20. Switzerland, 0.7%.
21. Greece, 0.6%.
22. Iran, 0.6%.
23. Japan, 0.6%.
24. Ghana, 0.6%.
25. Belgium, 0.4%.
26. Honduras, 0.3%.
27. South Korea, 0.3%.
28. Bosnia-Herzegovina, 0.3%.
29. Costa Rica, 0.3%.
30. Cameroon, 0.2%.
31. Australia, 0.2%.
32. Algeria, 0.1%.

- **a.** Interpret Brazil's probability of 35.9%, which was based on computer simulations of the tournament. Is it a relative frequency or a subjective interpretation of probability?
- **b.** Germany would emerge as the actual winner of the 2014 World Cup. Does this indicate that the 7.4% chance of Germany winning, which was calculated before the tournament, should have been 100% instead?

5.6 Pick the incorrect statement Which of the following statements is not correct, and why?

- **a.** If the number of male and female employees at a call center is equal, then the probability that a call is answered by a female employee is 0.50.
- **b.** If you randomly generate 10 digits, each integer between 0 and 9 must occur exactly once.
- **c.** You have 1,000 songs on your MP3 disc. 150 of them are of your favorite artist. If you decide to randomly play a very large number of songs, then each song of your favorite artist would have been played almost 15% of the time.

5.7 Sample size and sampling accuracy Your friend is interested in estimating the proportion of people who would vote for his project in a local contest. He selects a large sample among his many friends and claims that, with such a large sample, he does not need to worry about the method of selecting the sample. What is wrong in this reasoning? Explain.

5.8 Heart transplant Before the first human heart transplant, Dr. Christiaan Barnard of South Africa was asked to assess the probability that the operation would be successful. Did he need to rely on the relative frequency definition or the subjective definition of probability? Explain.

5.9 Nuclear war You are asked to use your best judgment to estimate the probability that there will be a nuclear war within the next 10 years. Is this an example of relative frequency or subjective definition of probability? Explain.

5.10 Simulate coin flips Use the web app Random Numbers (go to the tab that says Coin Flips) on the book's website or other software (such as random.org/coin) to illustrate the long-run definition of probability by simulating short-term and long-term results of flipping a balanced coin.

TRY
TECH

- **a.** Keep the probability of a head at the default value of 50% and set the number of flips to generate in a simulation to 10. Click on Simulate and record the proportion of heads for this simulation. Do this a total of 10 times by repeatedly clicking Simulate.
- **b.** Now set the number of flips to 100. Click Simulate 10 times, and record the 10 proportions of heads for each simulation. Do they vary much?
- **c.** Now set the number of flips to 1000. Click Simulate 10 times, and record the 10 proportions of heads for each simulation. Do they vary more, or less, than the proportions in part b based on 100 flips?
- **d.** Summarize the effect of the number of flips on the variability of the proportion. How does this reflect what's implied by the law of large numbers?

5.11 Unannounced pop quiz A teacher announces a pop quiz for which the student is completely unprepared. The quiz consists of 100 true-false questions. The student has no choice but to guess the answer randomly for all 100 questions.

a. Simulate taking this quiz by random guessing. Number a sheet of paper 1 to 100 to represent the 100 questions. Write a T (true) or F (false) for each question, by predicting what you think would happen if you repeatedly flipped a coin and let a tail represent a T guess and a head represent an F guess. (Don't actually flip a coin; merely write down what you think a random series of guesses would look like.)

b. How many questions would you expect to answer correctly simply by guessing?

c. The table shows the 100 correct answers. The answers should be read across rows. How many questions did you answer correctly?

Pop Quiz Correct Answers

T	F	T	T	F	F	T	T	T	T	T	F	T	F	F	T	T	F	T	F
F	F	F	F	F	F	F	T	F	F	T	F	T	F	F	T	F	T	T	F
T	F	F	F	F	F	T	F	T	T	F	T	T	T	F	F	F	F	F	T
T	F	F	T	F	F	T	T	T	T	F	F	F	F	F	F	F	T	F	F
F	F	T	F	F	T	T	F	F	T	F	T	F	T	T	T	T	F	F	F

d. The preceding answers were actually randomly generated by an app. What percentage were true, and what percentage would you expect? Why are they not necessarily identical?

e. Are there groups of answers within the sequence of 100 answers that appear nonrandom? For instance, what is the longest run of Ts or Fs? By comparison, which is the longest run of Ts or Fs within your sequence of 100 answers? (There is a tendency in guessing what randomness looks like to identify too few long runs in which the same outcome occurs several times in a row.)

5.12 Stock market randomness An interview in an investment magazine (*In the Vanguard*, Autumn 2003) asked mathematician John Allen Paulos, "What common errors do investors make?" He answered, "People tend not to believe that markets move in random ways. Randomness is difficult to recognize. If you have people write down 100 Hs and Ts to simulate 100 flips of a coin, you will always be able to tell a sequence generated by a human from one generated by real coin flips. When humans make up the sequence, they don't put in enough consecutive Hs and Ts, and they don't make the lengths of those runs long enough or frequent enough. And that is one of the reasons people look at patterns in the stock market and ascribe significance to them." (© The Vanguard Group, Inc., used with permission.)

a. Suppose that on each of the next 100 business days the stock market has a 1/2 chance of going up and a 1/2 chance of going down, and its behavior one day is independent of its behavior on another day. Use software, such as the web app mentioned in exercise 5.10 or random.org/coin to simulate whether the market goes up or goes down for each of the next 100 days. What is the longest sequence of consecutive moves up or consecutive moves down that you observe?

b. Run the simulation nine more times, with 100 observations for each run, and each time record the longest sequence of consecutive moves up or consecutive moves down that you observe. For the 10 runs, summarize the proportion of times that the longest sequence was 1, 2, 3, 4, 5, 6, 7, 8, or more. (Your class may combine results to estimate this more precisely.)

c. Based on your findings, explain why if you are a serious investor you should not get too excited if sometime in the next few months you see the stock market go up for five days in a row or go down for five days in a row.

5.2 Practicing the Basics

5.13 Student union poll Part of a student opinion poll at a university asks students what they think of the quality of the existing student union building on the campus. The possible responses were great, good, fair, and poor. Another part of the poll asked students how they feel about a proposed fee increase to help fund the cost of building a new student union. The possible responses to this question were in favor, opposed, and no opinion.

TRY

a. List all potential outcomes in the sample space for someone who is responding to both questions.

b. Show how a tree diagram can be used to display the outcomes listed in part a.

5.14 Songs Out of 100 songs on a playlist, 15 are of your favorite artist. You decide to randomly play one track from this playlist.

a. State the sample space for the possible outcomes.

b. State the probability for each possible outcome.

c. What is the probability that the track chosen randomly from the playlist is of your favorite artist?

d. What is the probability that the track chosen randomly from the playlist is not of your favorite artist?

5.15 Pop quiz A teacher gives a four-question unannounced true-false pop quiz, with two possible answers to each question.

TRY

a. Use a tree diagram to show the possible response patterns in terms of whether any given response is correct or incorrect. How many outcomes are in the sample space?

b. An unprepared student guesses all the answers randomly. Find the probabilities of the possible outcomes on the tree diagram.

c. Refer to part b. Using the tree diagram, evaluate the probability of passing the quiz, which the teacher defines as answering *at least* three questions correctly.

5.16 More true-false questions Your teacher gives a true-false pop quiz with 10 questions.

TRY

a. Show that the number of possible outcomes for the sample space of possible sequences of 10 answers is 1024.

b. What is the complement of the event of getting *at least* one of the questions wrong?

c. With random guessing, show that the probability of getting *at least* one question wrong is approximately 0.999.

5.17 Curling In the sport of curling, each shot is given points on a scale from 0–5, rating the success of each. Your friend claims, "Since the sum of points awarded when two shots are made is between 0 and 10, there is a one in eleven chance for each resulting sum to occur." Do you agree or disagree with his assessment of the probabilities? Explain.

5.18 On-time arrival probabilities The all-time, on-time arrival rate of a certain airline to a specific destination is 82%. This week, you have booked two flights to this destination with this airline.

a. Construct a sample space for the on-time or late arrival of the two flights.

b. Find the probability that both the flights arrive on time.

c. Find the probability that both the flights are late.

5.19 Three children A couple plans to have three children. Suppose that the probability of any given child being female is 0.5, and suppose that the genders of each child are independent events.

TRY

a. Write out all outcomes in the sample space for the genders of the three children.

b. What should be the probability associated with each outcome?

Using the sample space constructed in part a, find the probability that the couple will have

c. two girls and one boy.

d. at least one child of each gender.

5.20 Pick the incorrect statement Which of the following statements is not correct, and why?

a. If the number of male and female employees at a call center is equal, then the probability you call four times and a female employee answers your call only once is 1/5.

b. You have created a playlist of 100 songs on your MP3 disc. 15 of them are of your favorite artist. You randomly select four tracks from your playlist. The probability that one of them is of your favorite artist is $4 \times (0.85)^3 \times 0.15$.

5.21 Insurance Every year the insurance industry spends considerable resources assessing risk probabilities. To accumulate a risk of about one in a million of death, you can drive 100 miles, take a cross country plane flight, work as a police officer for 10 hours, work in a coal mine for 12 hours, smoke two cigarettes, be a nonsmoker but live with a smoker for two weeks, or drink 70 pints of beer in a year (Wilson and Crouch, 2001, pp. 208–209). Show that a risk of about one in a million of death is also approximately the probability of flipping 20 heads in a row with a balanced coin.

5.22 Pick the incorrect statement Which of the following statements is not correct, and why?

a. Last night, you randomly selected a restaurant for dinner from three similar restaurants in your city, with no prior preference for any one of them over the others. If your dining experience at the chosen restaurant was very satisfactory, then the probability of choosing any restaurant for tonight's dinner is 1/3.

b. If each night, for three consecutive nights, you select one restaurant out of three for dinner, then the probability of selecting the same restaurant for those three nights is $\left(\frac{1}{3}\right)^3$.

5.23 Seat belt use and auto accidents Based on records of automobile accidents in a recent year, the Department of Highway Safety and Motor Vehicles in Florida reported the counts who survived (S) and died (D), according to whether they wore a seat belt ($Y = \text{yes}, N = \text{no}$). The data are presented in the contingency table shown.

TRY

Outcome of auto accident by whether subject wore seat belt

Wore Seat Belt	Survived (S)	Died (D)	Total
Yes (Y)	412,368	510	**412,878**
No (N)	162,527	1,601	**164,128**
Total	**574,895**	**2,111**	**577,006**

a. What is the sample space of possible outcomes for a randomly selected individual involved in an auto accident? Use a tree diagram to illustrate the possible outcomes. (*Hint:* One possible outcome is YS.)

b. Using these data, estimate (i) P(D), (ii) P(N).

c. Estimate the probability that an individual did not wear a seat belt and died.

d. Based on part a, what would the answer to part c have been if the events N and D were independent? So, are N and D independent, and if not, what does that mean in the context of these data?

5.24 Protecting the environment When the General Social Survey most recently asked subjects whether they are a member of an environmental group (variable GRNGROUP) and whether they would be willing to pay higher prices to protect the environment (variable GRNPRICE), the results were as shown in the table.

		Pay Higher Prices (GRNPRICE)		
		Yes	Not Sure	No
Environmental Group Member (GRNGROUP)	Yes	293	71	66
	No	2,211	1,184	1,386

For a randomly selected American adult:

a. Estimate the probability of being (i) a member of an environmental group and (ii) willing to pay higher prices to protect the environment.

b. Estimate the probability of being both a member of an environmental group *and* willing to pay higher prices to protect the environment.

c. Given the probabilities in part a, show that the probability in part b is larger than it would be if the variables were independent. Interpret.

d. Estimate the probability that a person is a member of an environmental group *or* willing to pay higher prices to protect the environment. Do this (i) directly using the counts in the table and (ii) by applying the appropriate probability rule to the estimated probabilities found in parts a and b.

5.25 Global warming and trees A survey asks subjects whether they believe that global warming is happening (yes or no) and how much fuel they plan to use annually for automobile driving in the future, compared to their past use (less, about the same, more).

a. Show the sample space of possible outcomes by drawing a tree diagram that first gives the response on global warming and then the response on fuel use.

b. Let A be the event of a "yes" response on global warming and let B be the event of a "less" response on future fuel use. Suppose $P(A \text{ and } B) > P(A)P(B)$. Indicate whether A and B are independent events and explain what this means in nontechnical terms.

5.26 Newspaper sales You are the director of newspaper sales for the local paper. Each customer has signed up for either weekday delivery or weekend delivery. You record whether he or she received the delivery as Y for yes and N for no. The probabilities of the customer receiving the newspaper are as follows.

Outcome (Weekday, Weekend)	YY	YN	NY	NN
Probability	0.25	0.05	0.20	0.5

a. Display the outcomes in a contingency table, using the rows as the weekend event and the columns as the weekday event.

b. Let W denote the event that the customer bought a newspaper during the week and S be the event that he or she got it on the weekend (S for Saturday/Sunday). Find P(W) and P(S).

c. Explain what the event W and S means and find P(W and S).

d. Are W and S independent events? Explain why you would not normally expect customer choices to be independent.

5.27 Arts and crafts sales A local downtown arts and crafts shop found from past observation that 20% of the people who enter the shop actually buy something. Three potential customers enter the shop.

a. How many outcomes are possible for whether the clerk makes a sale to each customer? Construct a tree diagram to show the possible outcomes. $(\text{Let } Y = \text{sale}, N = \text{no sale.})$

b. Find the probability of at least one sale to the three customers.

c. What did your calculations assume in part b? Describe a situation in which that assumption would be unrealistic.

5.3 Practicing the Basics

5.28 Recidivism rates A 2014 article from *Business Insider* (http://www.businessinsider.com/department-of-justice-report-shows-high-recidivism-rate-2014-4) discusses recidivism rates in the United States. Recidivism is defined as being reincarcerated within five years of being sent to jail initially. Among the data reported, *Business Insider* cites that the recidivism rate for blacks is 81% compared to 73% among whites. Using notation, express each of these as a conditional probability.

5.29 Smoke alarms statistics National estimates of reported fires derived from the National Fire Incident Reporting System (NFIRS) and the National Fire Protection Association's (NFPA's) fire department survey show that in 2009–2013, 38% of home fire deaths occurred in homes with no smoke alarms, and 21% of home fire deaths were caused by fires in which smoke alarms were present but failed to operate. Let D denote {home fire death}, P denote {Smoke alarm is present}, and let F denote {Failed to operate}. Using events and their complements, identify each of the two given probabilities as a conditional probability.

5.30 Audit and low income Table 5.3 on audit status and income follows. Show how to find the probability of:

a. Being audited, given that the taxpayer is in the lowest income category.

b. Being in the lowest income category, given that the taxpayer is audited.

	Audited	
Income	**No**	**Yes**
$<$\$200,000	0.9556	0.0085
\$200,000 $-$ \$1mil	0.0326	0.0009
$>$\$1mil	0.0022	0.0003

5.31 Religious affiliation The 2012 Statistical Abstract of the United States[3] provides information on individuals' self-described religious affiliations. The information for 2008 is summarized in the following table (all numbers are in thousands).

Christian	
Catholic	57,199
Baptist	36,148
Christian (no denomination specified)	16,834
Methodist/Wesleyan	11,366
Other Christian	51,855
Jewish	2,680
Muslim	1,349
Buddhist	1,189
Other non-Christian	3,578
No Religion	34,169
Refused to Answer	11,815
Total Adult Population in 2008	228,182

a. Find the probability that a randomly selected individual is identified as Christian.

b. Given that an individual identifies as Christian, find the probability that the person is Catholic.

c. Given that an individual answered, find the probability the individual is identified as following no religion.

[3] *Source:* Data from www.census.gov/compendia/statab/2012/tables/12s0075.xls.

5.32 Labor force In 2014, a sample of 1925 Americans revealed that about 20.5% of them belong to the government sector. 7.5% of these are part-time employees, 60% are full-time employees, and 32.5% are retired.

a. Define events and identify which of these four probabilities refer to conditional probabilities.

b. Find the probability that an American adult in this sample is a full-time government employee.

5.33 Revisiting seat belts and auto accidents The following table is from Exercise 5.23 classifying auto accidents by survival status $(S = \text{survived}, D = \text{died})$ and seat belt status of the individual involved in the accident.

	Outcome		
Belt	**S**	**D**	**Total**
Yes	412,368	510	**412,878**
No	162,527	1,601	**164,128**
Total	**574,895**	**2,111**	**577,006**

a. Estimate the probability that the individual died (D) in the auto accident.

b. Estimate the probability that the individual died, given that the person (i) wore and (ii) did not wear a seat belt. Interpret results.

c. Are the events of dying and wearing a seat belt independent? Justify your answer.

5.34 Go Celtics! TRY Larry Bird, who played pro basketball for the Boston Celtics, was known for being a good shooter. In games during 1980–1982, when he missed his first free throw, 48 out of 53 times he made the second one, and when he made his first free throw, 251 out of 285 times he made the second one.

a. Form a contingency table that cross tabulates the outcome of the first free throw (made or missed) in the rows and the outcome of the second free throw (made or missed) in the columns.

b. For a given pair of free throws, estimate the probability that Bird (i) made the first free throw and (ii) made the second free throw. (*Hint*: Use counts in the (i) row margin and (ii) column margin.)

c. Estimate the probability that Bird made the second free throw, given that he made the first one. Does it seem as if his success on the second shot depends strongly, or hardly at all, on whether he made the first?

5.35 Identifying spam An article[4] on www.networkworld.com about evaluating e-mail filters that are designed to detect spam described a test of MailFrontier's Anti-Spam Gateway (ASG). In the test, there were 7840 spam messages, of which ASG caught 7005. Of the 7053 messages that ASG identified as spam, they were correct in all but 48 cases.

a. Set up a contingency table that cross classifies the actual spam status (with the rows "spam" and "not spam") by the ASG filter prediction (with the columns "predict message is spam" and "predict message is not spam"). Using the information given, enter counts in three of the four cells.

b. For this test, given that a message is truly spam, estimate the probability that ASG correctly detects it.

c. Given that ASG identifies a message as spam, estimate the probability that the message truly was spam.

5.36 Homeland security According to an article in *The New Yorker* (March 12, 2007), the Department of Homeland Security in the United States is experimenting with installing devices for detecting radiation at bridges, tunnels, roadways, and waterways leading into Manhattan. The New York Police Department (NYPD) has expressed concerns that the system would generate too many false alarms.

a. Form a contingency table that cross classifies whether a vehicle entering Manhattan contains radioactive material and whether the device detects radiation. Identify the cell that corresponds to the false alarms the NYPD fears.

b. Let A be the event that a vehicle entering Manhattan contains radioactive material. Let B be the event that the device detects radiation. Sketch a Venn diagram for which each event has similar (not the same) probability but the probability of a false alarm equals 0.

c. For the diagram you sketched in part b, explain why $P(A|B) = 1$, but $P(B|A) < 1$.

5.37 Down syndrome again TRY Example 8 discussed the Triple Blood Test for Down syndrome, using data summarized in a table shown again below.

	Blood Test		
Down	**POS**	**NEG**	**Total**
D	48	6	**54**
D^c	1307	3921	**5228**
Total	**1355**	**3927**	**5282**

a. Given that a test result is negative, show that the probability the fetus actually has Down syndrome is $P(D|NEG) = 0.0015$.

b. Is $P(D|NEG)$ equal to $P(NEG|D)$? If so, explain why. If not, find $P(NEG|D)$.

5.38 Obesity in America A 2014 Gallup poll reported that 27% of people in the United States are obese (having a body mass index score of 30 or more). Blacks have the highest obesity rate at 35%, whereas Asians have the lowest, at 9%.

a. Of the three percentages (estimated probabilities) reported, which are conditional? Explain.

b. These results are based on telephone interviews of 272,347 adults, aged 18 or older. Of these adults, 24,131 were black, and 5,752 were Asian. Create a contingency table showing estimated counts for race (Black, Asian, Other) and obesity (Yes, No).

c. Create a tree diagram with the first branching representing race and the second branching representing obesity. Be sure to include the appropriate percentages on each branch.

5.39 Happiness in relationship Are people happy in their romantic relationships? The table shows results from the 2012 General Social Survey for adults classified by gender and happiness.

[4] www.networkworld.com/reviews/2003/0915spamstats.html.

Gender	Level of Happiness Very Happy	Pretty Happy	Not Too Happy	Total
Male	69	73	4	**146**
Female	78	80	13	**171**
Total	**147**	**153**	**17**	**317**

a. Estimate the probability that an adult is very happy in his or her romantic relationship.

b. Estimate the probability that an adult is very happy (i) given that he is male and (ii) given that she is female.

c. For these subjects, are the events being very happy and being a male independent? (Your answer will apply merely to this sample. Chapter 11 will show how to answer this for the population of all adults.)

5.40 **Petra Kvitova serves** Petra Kvitova of the Czech Republic won the 2014 Wimbledon Ladies' Singles Championship. In the final game against Eugenie Bouchard of Canada she had 41 first serves, of which 28 were good, and three double faults.

TRY

a. Find the probability that her first serve is good.

b. Find the conditional probability of double faulting, given that her first serve resulted in a fault.

c. On what percentage of her service points does she double fault?

5.41 **Answering homework questions** Each question of an online homework consists of two parts. The probability that you answer the first part of a given question correctly is 0.75. Given that you answered the first part correctly, the probability you answer the second part correctly is 0.60. Given that your missed the first part, the probability that you answer the second part correctly is 0.40.

a. What is the probability that you answer both parts of a given question correctly?

b. Find the probability that you answer one of the two parts correctly (i) using the multiplicative rule with the two possible ways you can do this and (ii) by defining this as the complement of answering correctly neither or both of the two parts.

c. Are the results of the two parts independent? Explain.

5.42 **Discussion with students** In a statistics class of 30 students, 20 students are from the business program and 10 students are from the science program. The instructor randomly select three students, successively and *without replacement*, to discuss a question.

a. True or false: The probability of selecting three students from the business program is $(2/3) \times (2/3) \times (2/3)$. If true, explain why. If false, calculate the correct answer.

b. Let A = first student is from the business program and B = second student is from the business program. Are A and B independent? Explain why or why not.

c. Answer parts a and b if each student is replaced in the class after being selected.

5.43 **Drawing more cards** A standard deck of poker playing cards contains four suits (clubs, diamonds, hearts, and spades) and 13 different cards of each suit. During a hand of poker, 5 of the 52 cards have been exposed. Of the exposed cards, 3 were diamonds. Tony will have the opportunity to draw two more cards, and he has surmised that to win the hand, each of those two cards will need to be diamonds. What is Tony's probability of winning the hand? (Assume the two unexposed cards are not diamonds.)

5.44 **Big loser in Lotto** Example 10 showed that the probability of having the winning ticket in Lotto South was 0.00000007. Find the probability of holding a ticket that has zero winning numbers out of the 6 numbers selected (without replacement) for the winning ticket out of the 49 possible numbers.

TRY

5.45 **Online sections** For a course with two sections, let A denote {first section is online}, let B denote (at least one section is online}, and let C denote {both sections are online}. Suppose P (a section is online) = 1/2 and that the sections are independent.

TRY

a. Find $P(C|A)$ and $P(C|B)$.

b. Are A and C independent events? Explain why or why not.

c. Describe what makes $P(C|A)$ and $P(C|B)$ different from each other.

5.46 **Checking independence** In each of three independent visits to a restaurant, you choose randomly between two of today's specials, TS1 and TS2, on the menu. Let A denote {today's special on first visit is TS1}, B denote { today's special on second visit is TS1}, C denote { today's special on the first two visits are TS1}, and D denote {today's special on the three visits are TS1}.

a. Find the probabilities of A, B, C, and D.

b. Which, if any, pairs of these events are independent? Explain.

5.4 Practicing the Basics

5.47 Heart disease A particular heart disease is said to have a prevalence of 1/1000 in a specific population. In a sample of 50 people chosen randomly, what is the probability that at least two people have this disease?

TRY

5.48 Matching your birthday You consider your birth date to be special since it falls on January 1. Suppose your class has 25 students.

a. Is the probability of finding at least one student with a birthday that matches yours greater, the same, or less than the probability found in Example 13 of a match for at least two students? Explain.

b. Find that probability.

5.49 Lots of pairs Show that with 25 students, there are 300 *pairs* of students who can have the same birthday. So it's really not so surprising if at least two students have the same birthday. (*Hint*: You can pair 24 other students with each student, but how can you make sure you don't count each pair twice?)

5.50 Holes in one at Masters The Augusta National Golf Course in Augusta, Georgia, hosts the Masters Tournament each April. The course consists of four par 3s, ten par 4s, and four par 5s. The par 4s and par 5s are long enough so that no golfer has a realistic chance of getting a hole in one, but the par 3s are each short enough so that the possibility of a hole in one does exist. Over the 75-year history of the tournament, golfers have teed off on par 3s approximately 70,000 times, and a total of 73 holes in one have been recorded. For a given golfer, suppose the probability of getting a hole in one on each of the par 3s at Augusta are as follows:

Hole Number	P(hole in one)
4	0.0005
6	0.0015
12	0.0005
16	0.0025

a. For a randomly selected golfer, find the probability of no holes in one during a round of golf. Assume independence from one hole to the next.
b. For a randomly selected golfer, find the probability of no holes in one during the next 20 rounds of golf. Assume independence from one round to the next.
c. Use your answer in part b to find the probability of making at least one hole in one during the next 20 rounds of golf.

5.51 Failure and repair of photocopiers In a photocopy center, there are two small photocopiers, two medium photocopiers and one big photocopier. The probability that a small one fails and requires repairs is 0.1, a medium one fails and requires repairs is 0.08, and the probability that the big photocopier fails and requires repairs is 0.05. Assume that all five copiers operate independently.
a. What is the probability that all the photocopiers fail and require repairs?
b. What is the probability that none of the photocopiers fails and requires repairs?
c. What is the probability that one of the photocopiers fails and requires repairs?
d. What is the probability that one of the small photocopiers fails and requires repairs?
e. What is the probability that at least one of the five copiers fails and requires repairs?

5.52 Horrible 11 on 9/11 The digits in 9/11 add up to $11\ (9 + 1 + 1)$, American Airlines flight 11 was the first to hit the World Trade Towers (which took the form of the number 11), there were 92 people on board $(9 + 2 = 11)$, September 11 is the 254th day of the year $(2 + 5 + 4 = 11)$, and there are 11 letters in Afghanistan, New York City, the Pentagon, and George W. Bush (see article by L. Belkin, *New York Times*, August 11, 2002). How could you explain to someone who has not studied probability that, because of the way we look for patterns out of the huge number of things that happen, this is not necessarily an amazing coincidence?

5.53 Coincidence in your life State an event that has happened to you or to someone you know that seems highly coincidental (such as seeing a friend while on vacation). Explain why that event may not be especially surprising, once you think of all the similar types of events that could have happened to you or someone that you know, over the course of several years.

5.54 Monkeys typing Shakespeare Since events of low probability eventually happen if you observe enough trials, a monkey randomly pecking on a typewriter could eventually write a Shakespeare play just by chance. Let's see how hard it would be to type the title of *Macbeth* properly. Assume 50 keys for letters and numbers and punctuation. Find the probability that the first seven letters that a monkey types are macbeth. (Even if the monkey can type 60 strokes a minute and never sleeps, if we consider each sequence of seven keystrokes as a trial, we would wait on the average over 100,000 years before seeing this happen!)

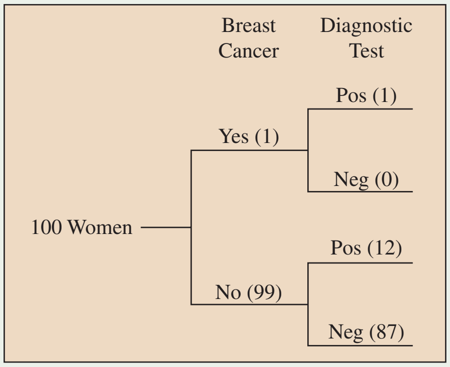
5.55 A *true* coincidence of emergency E-Comm, British Columbia's emergency communications center, provides communication services and support systems to two million residents of southwest British Columbia, Canada. On any given day, the probability a randomly selected resident decides to call E-Comm is 1.37/1000.
a. Assuming calls are made independently, find the probability that they all decide to call tomorrow.
b. Is the assumption of independence made in part a realistic? Explain.

5.56 Rosencrantz and Guildenstern In the opening scene of
TRY
Tom Stoppard's play *Rosencrantz and Guildenstern Are Dead*, about two Elizabethan contemporaries of Hamlet, Guildenstern flips a coin 91 times and gets a head each time. Suppose the coin was balanced.
a. Specify the sample space for 91 coin flips, such that each outcome in the sample space is equally likely. How many outcomes are in the sample space?
b. Show Guildenstern's outcome for this sample space. Show the outcome in which only the second flip is a tail.
c. What's the probability of the event of getting a head 91 times in a row?
d. What's the probability of at least one tail in the 91 flips?
e. State the probability model on which your solutions in parts c and d are based.

5.57 Mammogram diagnostics Breast cancer is the most
TRY
common form of cancer in women, affecting about 10% of women at some time in their lives. There is about a 1% chance of having breast cancer at a given time (that is, $P(S) = 0.01$ for the state of having breast cancer at a given time). The chance of breast cancer increases as a woman ages, and the American Cancer Society recommends an annual mammogram after age 40 to test for its presence. Of the women who undergo mammograms at any given time, about 1% are typically estimated to actually have breast cancer. The likelihood of a false test result varies according to the breast density and the radiologist's level of experience. For use of the mammogram to detect breast cancer, typical values reported are sensitivity $= 0.86$ and specificity $= 0.88$.
a. Construct a tree diagram in which the first set of branches shows whether a woman has breast cancer and the second set of branches shows the mammogram result. At the end of the final set of branches, show that $P(S \text{ and } POS) = 0.01 \times 0.86 = 0.0086$ and report the other intersection probabilities also.
b. Restricting your attention to the two paths that have a positive test result, show that $P(POS) = 0.1274$.

c. Of the women who receive a positive mammogram result, what proportion actually have breast cancer?

d. The following tree diagram illustrates how $P(S|POS)$ can be so small, using a typical group of 100 women who have a mammogram. Explain how to get the frequencies shown on the branches and why this suggests that $P(S|POS)$ is only about 0.08.



Typical results of mammograms for 100 women

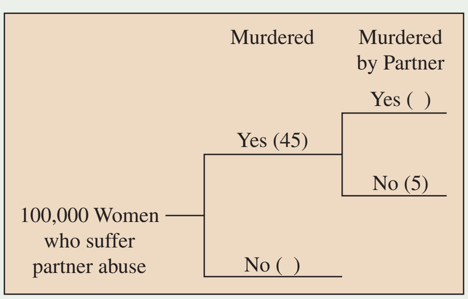
5.58 More screening for breast cancer Refer to the previous exercise. For young women, the prevalence of breast cancer is lower. Suppose the sensitivity is 0.86 and the specificity is 0.88, but the prevalence is only 0.001.

a. Given that a test comes out positive, find the probability that the woman truly has breast cancer.

b. Show how to use a tree diagram with frequencies for a typical sample of 1000 women to explain to someone who has not studied statistics why the probability found in part a is so low.

c. Of the cases that are positive, explain why the proportion in error is likely to be larger for a young population than for an older population.

5.59 Was OJ actually guilty? Former pro football star O. J. Simpson was accused of murdering his wife. In the trial, a defense attorney pointed out that although Simpson had been guilty of earlier spousal abuse, annually only about 40 women are murdered per 100,000 incidents of partner abuse. This means that $P(\text{murdered by partner}|\text{partner abuse}) = 40/100{,}000$. More relevant, however, is $P(\text{murdered by partner}|\text{partner abuse and women murdered})$. Every year it is estimated that 5 of every 100,000 women in the United States who suffer partner abuse are killed by someone other than their partner (Gigerenzer, 2002, p. 144). Part of a tree diagram is shown starting with 100,000 women who suffer partner abuse.



a. Based on the results stated, explain why the numbers 45 and 5 are entered as shown on two of the branches.

b. Fill in the two blanks shown in the tree diagram.

c. Conditional on partner abuse and the woman being murdered (by someone), explain why the probability the woman was murdered by her partner is 40/45. Why is this so dramatically different from $P(\text{murdered by partner}|\text{partner abuse}) = 40/100{,}000$?

5.60 Convicted by mistake In criminal trials (e.g., murder, robbery, driving while impaired, etc.) in the United States, it must be proven that a defendant is guilty beyond a reasonable doubt. This can be thought of as a very strong unwillingness to convict defendants who are actually innocent. In civil trials (e.g., breach of contract, divorce hearings for alimony, etc.), it must only be proven by a preponderance of the evidence that a defendant is guilty. This makes it easier to prove a defendant guilty in a civil case than in a murder case. In a high-profile pair of cases in the mid 1990s, O. J. Simpson was found to be not guilty of murder in a criminal case against him. Shortly thereafter, however, he was found guilty in a civil case and ordered to pay damages to the families of the victims.

a. In a criminal trial by jury, suppose the probability the defendant is convicted, given guilt, is 0.95, and the probability the defendant is acquitted, given innocence, is 0.95. Suppose that 90% of all defendants truly are guilty. Given that a defendant is convicted, find the probability he or she was actually innocent. Draw a tree diagram or construct a contingency table to help you answer.

b. Repeat part a, but under the assumption that 50% of all defendants truly are guilty.

c. In a civil trial, suppose the probability the defendant is convicted, given guilt is 0.99, and the probability the defendant is acquitted, given innocence, is 0.75. Suppose that 90% of all defendants truly are guilty. Given that a defendant is convicted, find the probability he or she was actually innocent. Draw a tree diagram or construct a contingency table to help you answer.

5.61 DNA evidence compelling? DNA evidence can be extracted from biological traces such as blood, hair, and saliva. "DNA fingerprinting" is increasingly used in the courtroom as well as in paternity testing. Given that a person is innocent, suppose that the probability of his or her DNA matching that found at the crime scene is only 0.000001, one in a million. Further, given that a person is guilty, suppose that the probability of his or her DNA matching that found at the crime scene is 0.99. Jane Doe's DNA matches that found at the crime scene.

a. Find the probability that Jane Doe is actually innocent, if absolutely her probability of innocence is 0.50. Interpret this probability. Show your solution by introducing notation for events, specifying probabilities that are given, and using a tree diagram to find your answer.

b. Repeat part a if the unconditional probability of innocence is 0.99. Compare results.

c. Explain why it is very important for a defense lawyer to explain the difference between $P(\text{DNA match}|\text{person innocent})$ and $P(\text{person innocent}|\text{DNA match})$.

5.62 Triple Blood Test Example 8 about the Triple Blood Test for Down syndrome found the results shown in the table on next column.

	Blood Test		
Down	**POS**	**NEG**	**Total**
Yes	48	6	**54**
No	1307	3921	**5228**
Total	**1355**	**3927**	**5282**

a. Estimate the probability that Down syndrome occurs (Down = Yes).

b. Find the estimated (i) sensitivity and (ii) specificity.

c. Find the estimated (i) $P(\text{Yes}|\text{POS})$ and (ii) $P(\text{No}|\text{NEG})$. (Note: These probabilities are the predictive values.)

d. Explain how the probabilities in parts b and c give four ways of describing the probability that a diagnostic test makes a correct decision.

5.63 Simulating arrivals to local holiday center TRY TECH The director of a local holiday center is offering a special prize to the first married visitor. The distribution of the marital status for Americans estimated by SDA (Survey Documentation and Analysis) is shown below

Marital Status	Married	Widowed	Divorced	Separated	Did not marry
Probability	53.60%	9.70%	12.60%	3.50%	20.60%

The director decides that if more than 20 visitors are required before the first married visitor arrives, she will need to issue a special advertisement for married people.

a. Conduct a simulation 10 times, using the Random Numbers app accessible on the book's website or a calculator or a software, to estimate the probability this will happen. Show all steps of the simulation, including any assumptions that you make. Refer to Example 16 as a model for carrying out this simulation.

b. In practice, you would do at least 1000 simulations to estimate this probability well. You'd then find that the probability of exceeding 20 visitors before finding a married visitor is 0.36. Actually, simulation is not needed. Show how to find this probability by using the methods of this chapter.

5.64 Probability of winning TRY TECH In Example 16, we estimated the probability of winning the game was 0.65.

a. If you conducted 30 more simulations of this game, what probability of winning would you expect to get?

b. The simulation in the example consisted of 20 repetitions. For a total of 200, conduct another 180 repetitions. What is the estimated probability of winning based on these 200 repetitions?

c. Does the estimated probability of winning tend to get closer to an actual probability of winning as the number of repetitions increases?

5.65 Probability of winning TECH In Example 16, we explored the number of rolls it takes to win the game. In reality, it's not the number of rolls but rather the time it takes to move 12 spaces that dictates who wins the game. Consider two alternative strategies for playing the game. The first strategy, let's call it the aiming strategy, uses the probabilities given in Example 16 and takes 4 seconds to roll each ball. A second strategy is to roll the ball in rapid succession. The rapid-succession strategy takes only 3 seconds to roll each ball; however, the probabilities of landing in the holes become 50% for red, 20% for yellow, 10% for green, and 20% for no holes.

a. Using the results of the simulations from Example 16, calculate the time to finish each game, using the aiming strategy. Assuming that your competitor also takes 4 seconds per roll, determine whether you beat him for each race. Estimate the probability of beating him.

b. Using the rapid-succession strategy, simulate another 20 games. For each one, calculate the time to finish the race and determine whether you would beat the competitor. (The competitor continues to use the aiming strategy, taking 4 seconds per roll.) Estimate the probability of beating him by using this strategy.

c. Which strategy should you choose to maximize your chance of beating the competitor?

6.1 Practicing the Basics

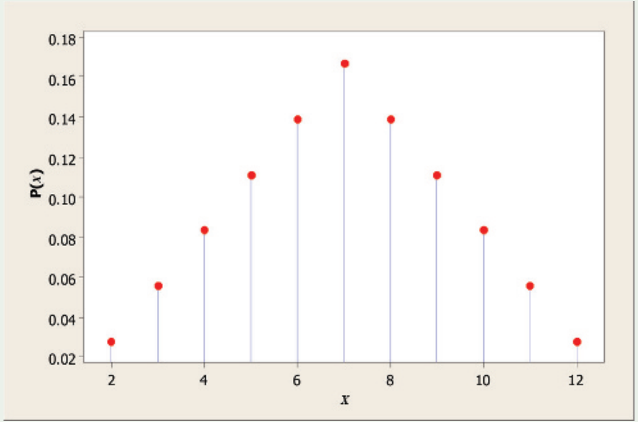
6.1 Rolling dice

TRY

a. State in a table the probability distribution for the outcome of rolling a balanced die. (This is called the **uniform distribution** on the integers 1, 2, . . . ,6.)

b. Two balanced dice are rolled. Show that the probability distribution for X = total on the two dice is as shown in the figure. (*Hint*: First construct the sample space of the 36 equally likely outcomes you could get. For example, you could denote the six outcomes where you get a 1 on the first die by (1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6), where the second number in a pair is the number you get on the second die.)

c. Show that the probabilities in part b satisfy the two conditions for a probability distribution.



6.2 **Dental Insurance** You plan to purchase dental insurance for your three remaining years in school. The insurance makes a one-time payment of \$1,000 in case of a major dental repair (such as an implant) or \$100 in case of a minor repair (such as a cavity). If you don't need dental repair over the next 3 years, the insurance expires and you receive no payout. You estimate the chances of requiring a major repair over the next 3 years as 5%, a minor repair as 60% and no repair as 35%.

TRY

a. Why is X = payout of dental insurance a random variable?

b. Is X discrete or continuous? What are its possible values?

c. Give the probability distribution of X.

6.3 **San Francisco Giants hitting** The table shows the probability distribution of the number of bases for a randomly selected time at bat for a San Francisco Giants player in 2010 (excluding times when the player got on base because of a walk or being hit by a pitch). In 74.29% of the at-bats the player was out, 17.04% of the time the player got a single (one base), 5.17% of the time the player got a double (two bases), 0.55% of the time the player got a triple, and 2.95% of the time the player got a home run.

TRY

a. Verify that the probabilities give a legitimate probability distribution.

b. Find the mean of this probability distribution.

c. Interpret the mean, explaining why it does not have to be a whole number, even though each possible value for the number of bases is a whole number.

San Francisco Giants Hitting

Number of Bases	Probability
0	0.7429
1	0.1704
2	0.0517
3	0.0055
4	0.0295

6.4 **Basketball shots** To win a basketball game, two competitors play three rounds of one three-point shot each. The series ends if one of them scores in a round but the other misses his shot or if both get the same result in each of the three rounds. Assume competitors A and B have 30% and 20% of successful attempts, respectively, in three-point shots and that the outcomes of the shots are independent events.

a. Verify the probability that the series ends in the second round is 23.56%. (*Hint:* Sketch a tree diagram and write out the sample space of all possible sequences of wins and losses in the three rounds of the series, find the probability for each sequence and then add up those for which the series ends within the second round).

b. Find the probability distribution of X = number of rounds played to end the series.

c. Find the expected number of rounds to be played in the series.

6.5 **WhatsApp reviews** 71% of WhatsApp users have given it a five-star rating on Google Play. Of the remaining users, 15%, 6%, 3%, and 5% have given ratings of four, three, two, and one stars, respectively to the application.

a. Specify the probability distribution for the number of stars as rated by the users.

b. Find the mean of this probability distribution. Interpret it.

6.6 **Selling houses** Let X represent the number of homes a real estate agent sells during a given month. Based on previous sales records, she estimates that $P(0) = 0.68$, $P(1) = 0.19$, $P(2) = 0.09$, $P(3) = 0.03$, $P(4) = 0.01$, with negligible probability for higher values of x.

a. Explain why it does not make sense to compute the mean of this probability distribution as $(0 + 1 + 2 + 3 + 4)/5 = 2.0$ and claim that, on average, she expects to sell 2 homes.

b. Find the correct mean and interpret.

6.7 **Playing the lottery** The state of Ohio has several state-wide lottery options. One is the Pick 3 game in which you pick one of the 1000 three-digit numbers between 000 and 999. The lottery selects a three-digit number at random. With a bet of \$1, you win \$500 if your number is selected and nothing (\$0) otherwise. (Many states have a very similar type of lottery.) (*Source:* Background information from www.ohiolottery.com.)

a. With a single \$1 bet, what is the probability that you win \$500?

b. Let X denote your winnings for a \$1 bet, so $x = \$0$ or $x = \$500$. Construct the probability distribution for X.

c. Show that the mean of the distribution equals 0.50, corresponding to an expected return of 50 cents for the dollar paid to play. Interpret the mean.

d. In Ohio's Pick 4 lottery, you pick one of the 10,000 four-digit numbers between 0000 and 9999 and (with a \$1 bet) win \$5000 if you get it correct. In terms of your expected winnings, with which game are you better off—playing Pick 4, or playing Pick 3? Justify your answer.

6.8 **Roulette** A roulette wheel consists of 38 numbers, 0 through 36 and 00. Of these, 18 numbers are red, 18 are black, and 2 are green (0 and 00). You are given \$10 and told that you must pick one of two wagers, for an outcome based on a spin of the wheel: (1) Bet \$10 on number 23. If the spin results in 23, you win \$350 and also get back your \$10 bet. If any other number comes up, you lose your \$10, or (2) Bet \$10 on black. If the spin results in any one of the black numbers, you win \$10 and also get back your \$10 bet. If any other color comes up, you lose your \$10.

TRY

a. Without doing any calculation, which wager would you prefer? Explain why. (There is no correct answer. Peoples' choices are based on their individual preferences and risk tolerances.)

b. Find the expected outcome for each wager. Which wager is better in this sense?

6.9 **More Roulette** The previous exercise on roulette described two bets: one bet on the single number 23 with winnings of either \$350 or $-\$10$ and a different bet on black with winnings of either \$10 or $-\$10$. For both types of bets, the expected winning is $-\$0.53$. Which of the two bets has the larger standard deviation? (*Hint:* Which bet has outcomes that are, on average, further from the mean?) Which bet would you prefer? Explain.

TRY

6.10 **Ideal number of children** Let X denote the response of a randomly selected person to the question, "What is the ideal number of children for a family to have?" The probability distribution of X in the United States is approximately as shown in the table, according to the gender of the person asked the question.

TRY

Probability Distribution of X = Ideal Number of Children

x	P(x) Females	P(x) Males
0	0.01	0.02
1	0.03	0.03
2	0.55	0.60
3	0.31	0.28
4	0.11	0.08

Note that the probabilities do not sum to exactly 1 due to rounding error.

a. Show that the means are similar, 2.50 for females and 2.39 for males.

b. The standard deviation for the females is 0.770 and 0.758 for the males. Explain why a practical implication of the values for the standard deviations is that males hold slightly more consistent views than females about the ideal family size.

6.11 **Profit and the weather** From past experience, a wheat farmer living in Manitoba, Canada, finds that his annual profit (in Canadian dollars) is $80,000 if the summer weather is typical, $50,000 if the weather is unusually dry, and $20,000 if there is a severe storm that destroys much of his crop. Weather bureau records indicate that the probability is 0.70 of typical weather, 0.20 of unusually dry weather, and 0.10 of a severe storm. Let X denote the farmer's profit next year.

a. Construct a table with the probability distribution of X.

b. What is the probability that the profit is $50,000 or less?

c. Find the mean of the probability distribution of X. Interpret.

d. Suppose the farmer buys insurance for $3000 that pays him $20,000 in the event of a severe storm that destroys much of the crop and pays nothing otherwise. Find the probability distribution of his profit. Find the mean and summarize the effect of buying this insurance.

6.12 **Buying on eBay** You are watching two items posted for sale on eBay and bid $30 for the first and $20 for the second item. You estimate that you are going to win the first bid with probability 0.1 and the second bid with probability 0.2, and you assume that winning the two bids are independent events. Let X denote the random variable denoting the total amount of money you will spend on the two items.

a. List the sample space of all possible outcomes of winning or losing the two bids. (Draw a tree diagram.)

b. Find the probability of each outcome in the sample space. (Use the tree diagram.)

c. Find the probability distribution of X.

d. Find the mean of X.

6.13 **Selling at the right price** An insurance company wants to examine the views of its clients about the prices of three car insurance plans launched last year. It conducts a survey with two sets of plans with different prices and finds that:

- If plan A is sold for $150, plan B for $250, and plan C for $350, then 45% of the customers would be interested in plan A, 15% in plan B, and 40% in plan C.
- If plans A, B, and C are sold for $170, $250, and $310 respectively, then 15% of the customers would be interested in plan A, 40% in plan B, and 45% in plan C.

a. For the first pricing set, construct the probability distribution of X = selling price for the sale of a car insurance plan, find its mean, and interpret.

b. For the second pricing set, construct the probability distribution of X, find its mean, and interpret.

c. Which pricing set is more profitable to the company? Explain.

6.14 **Uniform distribution** A random number generator is used to generate a real number between 0 and 1, equally likely to fall anywhere in this interval of values. (For instance, 0.3794259832 . . . is a possible outcome.)

a. Sketch a curve of the probability distribution of this random variable, which is the continuous version of the **uniform distribution** (see Exercise 6.1).

b. The probability that the number falls in the interval from 0 to the mean is 50%. Find the mean. (Remember that the total area under the probability curve is 1.)

c. Find the probability that the random number falls between 0.35 and 0.75.

d. Find the probability that the random number is less than 0.8?

6.15 **TV watching** A social scientist uses the General Social Survey (GSS) to study how much time per day people spend watching TV. The variable denoted by TVHOURS at the GSS Web site measures this using the discrete values 0, 1, 2, . . . , 24.

TRY

a. Explain how, in theory, TV watching is a continuous random variable.

b. An article about the study shows two histograms, both skewed to the right, to summarize TV watching for females and males. Since TV watching is in theory continuous, why were histograms used instead of curves?

c. If the article instead showed two curves, explain what they would represent.

6.2 Practicing the Basics

6.16 Probabilities in tails For a normal distribution, use Table A, software, or a calculator to find the probability that an observation is

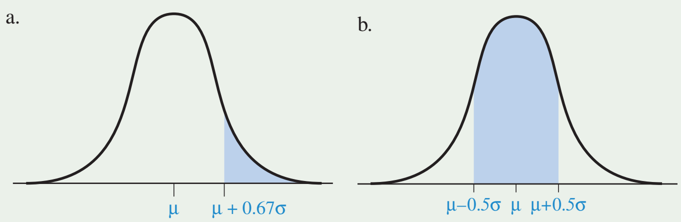
a. at least 1 standard deviation above the mean.

b. at least 1 standard deviation below the mean.

c. within 1 standard deviation of the mean.

For each part, sketch the normal curve and indicate the area corresponding to the probability.

6.17 Probability in graph For the normal distributions shown below, use Table A, software, or a calculator to find the probability that an observation falls in the shaded region.



6.18 Empirical rule Verify the empirical rule by using Table A, software, or a calculator to show that for a normal distribution, the probability (rounded to two decimal places) within

a. 1 standard deviation of the mean equals 0.68.

b. 2 standard deviations of the mean equals 0.95.

c. 3 standard deviations of the mean is very close to 1.00.

In each case, sketch a normal distribution, identifying on the sketch the probabilities you used to show the result.

6.19 Central probabilities For a normal distribution, use Table A to verify that the probability (rounded to two decimal places) within

a. 1.64 standard deviations of the mean equals 0.90.

b. 2.58 standard deviations of the mean equals 0.99.

c. Find the probability that falls within 0.67 standard deviations of the mean.

d. Sketch these three cases on a single graph.

6.20 *z*-score for given probability in tails For a normal distribution,

a. Find the z-score for which a total probability of 0.04 falls more than z standard deviations (in either direction) from the mean, that is, below $\mu - z\sigma$ or above $\mu + z\sigma$.

b. For this z-score, explain why the probability of values more than z standard deviations below the mean is 0.02.

c. Explain why $\mu + z\sigma$ is the 2nd percentile.

6.21 Probability in tails for given *z*-score For a normal distribution,

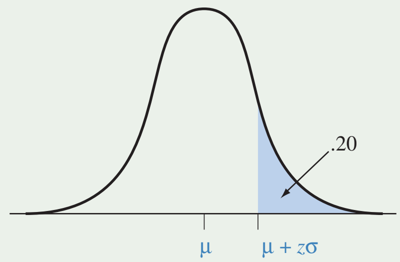
a. Show that a total probability of 0.01 falls more than $z = 2.58$ standard deviations from the mean.

b. Find the z-score for which the two-tail probability that falls more than that many standard deviations from the mean in either direction equals (a) 0.05, (b) 0.10. Sketch the two cases on a single graph.

6.22 *z*-score for right-tail probability

a. For the normal distribution shown below, find the z-score.

b. Find the value of z (rounding to two decimal places) for right-tail probabilities of (i) 0.05 and (ii) 0.005.



6.23 *z*-score and central probability Find the z-score such that the interval within z standard deviations of the mean (between $\mu - z\sigma$ and $\mu + z\sigma$) for a normal distribution contains

a. 50% of the probability.

b. 90% of the probability.

c. Sketch the two cases on a single graph.

6.24 U.S. Air Force To join the U.S. Air Force as an officer, you cannot be younger than 18 or older than 34 years of age. The distribution of age of Americans in 2012 was normal with $\mu = 38$ years and $\sigma = 22.67$ years. What proportion of U.S. citizens are not eligible to serve as an officer due to age restrictions?

6.25 Blood pressure A World Health Organization study (the MONICA project) of health in various countries reported that in Canada, systolic blood pressure readings have a mean of 121 and a standard deviation of 16. A reading above 140 is considered high blood pressure.

a. What is the z-score for a blood pressure reading of 140?

b. If systolic blood pressure in Canada has a normal distribution, what proportion of Canadians suffers from high blood pressure?

c. What proportion of Canadians has systolic blood pressures in the range from 100 to 140?

d. Find the 90th percentile of blood pressure readings.

6.26 Coffee Machine Suppose your favorite coffee machine offers 12 ounce cups of coffee. The actual amount of coffee put in the cup by the machine varies according to a normal distribution, with mean equal to 13 ounces and standard deviation equal to 0.6 ounces. For each question below, sketch a graph, mark the mean, shade the area to which the answer refers and compute the percentage.

a. What percentage of cups will be filled with less than 12 ounces?

b. What percentage of cups will be filled with more than 12.5 ounces?

c. What percentage of cups will have in between 12 and 13 ounces of coffee?

6.27 Lifespan of phone batteries Most phones use lithium-ion (Li-ion) batteries. These batteries have a limited number of charge and discharge cycles, usually falling between 300 and 500. Beyond this lifespan, a battery gradually diminishes below 50% of its original capacity.

TECH

a. Suppose the distribution of the number of charge and discharge cycles was normal. What values for the mean and the standard deviation are most likely to meet the assumption of normality of this variable?

b. Based on the mean and the standard deviation calculated in part a, find the 95th percentile.

6.28 Birth weight for boys In the United States, the mean birth weight for boys is 3.41 kg, with a standard deviation of 0.55 kg. (Source: cdc.com.) Assuming that the distribution of birth weight is approximately normal, find the following using a table, calculator, or software.

TECH

a. A baby is considered of low birth weight if it weighs less than 2.5 kg. What proportion of baby boys in the United States are born with low birth weight?

b. What is the z-score for a baby boy that weighs 1.5 kg (defined as extremely low birth weight)?

c. Typically, birth weight is between 2.5 kg and 4.0 kg. Find the probability a baby boy is born with typical birth weight.

d. Matteo weighs 3.6 kg at birth. He falls at what percentile?

e. Max's parents are told that their newborn son falls at the 96th percentile. How much does Max weigh?

6.29 MDI The Mental Development Index (MDI) of the Bayley Scales of Infant Development is a standardized measure used in observing infants over time. It is approximately normal with a mean of 100 and a standard deviation of 16.

TRY

a. What proportion of children has an MDI of (i) at least 120? (ii) at least 80?

b. Find the MDI score that is the 99th percentile.

c. Find the MDI score such that only 1% of the population has MDI below it.

6.30 Quartiles and outliers For an approximately normally distributed random variable X with a mean of 200 and a standard deviation of 36,

a. Find the z-score corresponding to the lower quartile and upper quartile of the standard normal distribution.

b. Find and interpret the lower quartile and upper quartile of X.

c. Find the interquartile range (IQR) of X.

d. An observation is a potential outlier if it is more than $1.5 \times$ IQR below Q1 or above Q3. Find the values of X that would be considered potential outliers.

6.31 April precipitation Over roughly the past 100 years, the mean monthly April precipitation in Williamstown, Massachusetts, equaled 3.6 inches with a standard deviation of 1.6 inches. (Source: http://web.williams.edu/weather/)

a. In April 1983, the wettest April on record, the precipitation equaled 8.4 inches. Find its z-score. If the distribution of precipitation were roughly normal, would this be unusually high? Explain.

b. Assuming a normal distribution, an April precipitation of 4.5 inches corresponds to what percentile?

c. Of the 119 measurements of April precipitation on record (reaching as far back as 1892), 66.5% fell within one, 97.5% within two, and 99.1% within three standard deviations of the mean. Do you think that the distribution of April precipitation is approximately normal? Why or why not?

6.32 Automatic filling machine A machine is programmed to fill packets with 500 grams of nuts. It is known from previous experiences the net weight of nuts in the packets are normally distributed with $\mu = 502$ grams and $\sigma = 3$ grams. A packet is considered conformant with the weight specifications if its net weight is at least 500 grams.

TRY

a. What proportion of the packets are considered conformant?

b. What proportion of nut packets have a net weight of less than 510 grams?

c. What proportion of nut packets have a net weight of less than 500 grams or more than 510 grams?

d. What is the median of the net weights of nuts in the packets?

6.33 SAT versus ACT SAT math scores follow a normal distribution with an approximate $\mu = 500$ and $\sigma = 100$. Also ACT math scores follow a normal distrubution with an approximate $\mu = 21$ and $\sigma = 4.7$. You are an admissions officer at a university and have room to admit one more student for the upcoming year. Joe scored 600 on the SAT math exam, and Kate scored 25 on the ACT math exam. If you were going to base your decision solely on their performances on the exams, which student should you admit? Explain.

TRY

6.34 Relative grades The mean and standard deviation of the grades of a statistics course and an English course are $(\mu = 80,\ \sigma = 4.5)$ and $(\mu = 85,\ \sigma = 4.0)$, respectively. A student attends both the courses and scores 85 in statistics and 95 in English. Which grade is relatively better? Explain why.

6.3 Practicing the Basics

6.35 Kidney transplants In kidney transplantations, compatibility between donor and receiver depends on such factors as blood type and antigens. Suppose that for a randomly selected donor from a large national kidney registry, there is a 10% chance that he or she is compatible with a specific receiver. Three donors are randomly selected from this registry. Find the probability that 0, 1, 2, or all 3 selected donors are compatible.

TRY

a. Do this by constructing the sample space and finding the probability for each possible outcome of choosing three donors. Use these probabilities to construct the probability distribution.

b. Do this using the formula for the binomial distribution.

6.36 Compatible donors Refer to the previous exercise. Let the random variable X = number of compatible donors. Check the conditions for X to be binomial by answering parts a–c.

a. What constitutes a trial, how many trials are there, and how many outcomes does each trial have?

b. Does each trial have the same probability of success? Explain

c. Are the trials independent?

6.37 Symmetric binomial Construct a graph similar to that in Figure 6.1 for each of the following binomial distributions:

a. $n = 4$ and $p = 0.50$.

b. $n = 4$ and $p = 0.30$.

c. $n = 4$ and $p = 0.10$.

d. Which if any of the graphs in parts a–c are symmetric? Without actually constructing the graph, would the case $n = 20$ and $p = 0.50$ be symmetric or skewed?

e. Which of the graphs in parts a–c is the most heavily skewed? Without actually constructing the graph, would the case $n = 4$ and $p = 0.01$ exhibit more or less skewness than the graph in part c?

6.38 Unfair wealth distribution sentiment According to a study published in www.gallup.com in 2015, 63% of Americans said wealth should be more evenly distributed among a larger percentage of people. For a sample of 10 Americans, let $X =$ number of respondents who said wealth was unfairly distributed.

a. Explain why the conditions are satisfied for X to have the binomial distribution.

b. Identify n and p for the binomial distribution.

c. Find the probability that two Americans in the sample said wealth should be more evenly distributed.

6.39 Bidding on eBay You are bidding on four items available on eBay. You think that for each bid, you have a 25% chance of winning it, and the outcomes of the four bids are independent events. Let X denote the number of winning bids out of the four items you bid on.

a. Explain why the distribution of X can be modeled by the binomial distribution.

b. Find the probability that you win exactly 2 bids.

c. Find the probability that you win 2 bids or fewer.

d. Find the probability that you win more than 2 bids.

6.40 More eBay bidding For each of the following situations, explain whether the binomial distribution applies for X.

a. You are bidding on four items available on eBay. You think that you will win the first bid with probability 25% and the second through fourth bids with probability 30%. Let X denote the number of winning bids out of the four items you bid on.

b. You are bidding on four items available on eBay. Each bid is for $70, and you think there is a 25% chance of winning a bid, with bids being independent events. Let X be the total amount of money you pay for your winning bids.

6.41 Test generator A professor of statistics wants to prepare a test paper by selecting five questions randomly from an online test bank available for his course. In the test bank, the proportion of questions labeled "HARD" is 0.3.

a. Find the probability that all the questions selected for the test are labeled HARD.

b. Find the probability that none of the questions selected for the test is labeled HARD.

c. Find the probability that less than half of the questions selected for the test are labeled HARD.

6.42 NBA shooting In the National Basketball Association, the top free throw shooters usually have probability of about 0.90 of making any given free throw.

a. During a game, one such player (Dirk Nowitzki) shot 10 free throws. Let $X =$ number of free throws made. What must you assume for X to have a binomial distribution? (Studies have shown that such assumptions are well satisfied for this sport.)

b. Specify the values of n and p for the binomial distribution of X in part a.

c. Find the probability that he made (i) all 10 free throws (ii) 9 free throws and (iii) more than 7 free throws.

6.43 Season performance Refer to the previous exercise. Over the course of a season, this player shoots 400 free throws.

a. Find the mean and standard deviation of the probability distribution of the number of free throws he makes.

b. By the normal distribution approximation, within what range would you expect the number made almost certainly to fall? Why?

c. Within what range would you expect the *proportion* made to fall?

6.44 Is the die balanced? A balanced die with six sides is rolled 60 times.

a. For the binomial distribution of $X =$ number of 6s, what is n and what is p?

b. Find the mean and the standard deviation of the distribution of X. Interpret.

c. If you observe $x = 0$, would you be skeptical that the die is balanced? Explain why, based on the mean and standard deviation of X.

d. Show that the probability that $x = 0$ is 0.0000177.

6.45 Exit poll TRY An exit poll is taken of 3000 voters in a statewide election. Let X denote the number who voted in favor of a special proposition designed to lower property taxes and raise the sales tax. Suppose that in the population, exactly 50% voted for it.

a. Explain why this scenario would seem to satisfy the three conditions needed to use the binomial distribution. Identify n and p for the binomial.

b. Find the mean and standard deviation of the probability distribution of X.

c. Using the normal distribution approximation, give an interval in which you would expect X almost certainly to fall, if truly $p = 0.50$. (*Hint*: You can follow the reasoning of Example 15 on racial profiling.)

d. Now, suppose that the exit poll had $x = 1706$. What would this suggest to you about the actual value of p?

6.46 Jury duty TRY The juror pool for the upcoming murder trial of a celebrity actor contains the names of 100,000 individuals in the population who may be called for jury duty. The proportion of the available jurors on the population list who are Hispanic is 0.40. A jury of size 12 is selected at random from the population list of available jurors. Let $X =$ the number of Hispanics selected to be jurors for this jury.

a. Is it reasonable to assume that X has a binomial distribution? If so, identify the values of n and p. If not, explain why not.

b. Find the probability that no Hispanic is selected.

c. If no Hispanic is selected out of a sample of size 12, does this cast doubt on whether the sampling was truly random? Explain.

6.47 Poor, poor, Pirates TECH On September 7, 2008, the Pittsburgh Pirates lost their 82nd game of the 2008 season and tied the 1933–1948 Philadelphia Phillies major sport record (baseball, football, basketball, and hockey) for most consecutive losing seasons at 16. In fact, their losing streak continued until 2012 with 20 consecutive losing seasons. A Major League Baseball season consists of 162 games, so for the Pirates to end their streak, they need to win at least 81 games in a season (which they did in 2013).

a. Over the course of the streak, the Pirates have won approximately 42% of their games. For simplicity, assume the number of games they win in a given season follows a binomial distribution with $n = 162$ and $p = 0.42$. What is their expected number of wins in a season?

b. What is the probability that the Pirates will win at least 81 games in a given season? (You may use technology to find the exact binomial probability or use the normal distribution to approximate the probability by finding a z-score for 81 and then evaluating the appropriate area under the normal curve.)

c. Can you think of any factors that might make the binomial distribution an inappropriate model for the number of games won in a season?

6.48 Checking guidelines In a village having more than 100 adults, eight are randomly selected in order to form a committee of residents. 40% of adults in the village are connected with agriculture.

a. Verify that the guidelines have been satisfied about the relative sizes of the population and the sample, thus allowing the use of a binomial probability distribution for the number of selected adults connected with agriculture.

b. Check whether the guideline was satisfied for this binomial distribution to be reasonably approximated by a normal distribution.

6.49 Movies sample TRY Five of the 20 movies running in movie theatres this week are comedies. A selection of four movies are picked at random. Does X = the number of movies in the sample which are comedies have the binomial distribution with $n = 4$ and $p = 0.25$? Explain why or why not.

6.50 Binomial needs fixed n For the binomial distribution, the number of trials n is a fixed number. Let X denote the number of girls in a randomly selected family in Canada that has three children. Let Y denote the number of girls in a randomly selected family in Canada (that is, the number of children could be any number). A binomial distribution approximates well the probability distribution for one of X and Y, but not for the other.

a. Explain why.

b. Identify the case for which the binomial applies and identify n and p.

6.51 Binomial assumptions For the following random variables, check whether the conditions needed to use the binomial distribution are satisfied or not. Explain

a. X = number of people suffering from a contagious disease in a family of 4, when the probability of catching this disease is 2% in the whole population (binomial, $n = 4$, $p = 0.02$). (*Hint:* Is the independence assumption plausible?)

b. X = number of unmarried women in a sample of 100 females randomly selected from a large population where 40% of women in the population are unmarried (binomial, $n = 100$, $p = 0.40$).

c. X = number of students who use an iPhone in a random sample of five students from a class of size 20, when half the students use iPhones (binomial, $n = 5$, $p = 0.50$).

d. X = number of days in a week you go out for dinner. (*Hint:* Is the probability of dining out the same for each day?)

7.1 Practicing the Basics

7.1 Simulating the exit poll Simulate an exit poll of 100 voters, using the Sampling Distribution web app accessible from the book's website, assuming that the population proportion is 0.53. Refer to Activity 1 for guidance on using the app.

TRY

TECH

a. Simulate drawing one random sample of size 100. What sample proportion did you get? Why do you not expect to get exactly 0.53?

b. Keep the sample size n as 100 and p as 0.53, but now simulate drawing 10,000 samples of that size. Use the histogram of the 10,000 sample proportions you generated to describe the simulated sampling distribution (shape, center, spread). (*Note:* The app allows you to save the graph to file.)

c. Use a formula from this section to predict the value of the standard deviation of the sample proportions that you generated in part b. Compare it to the standard deviation of the 10,000 simulated sample proportions stated in the title of the graph.

d. Now change the population proportion to 0.7, keeping the sample size n at 100. Simulate the exit poll 10,000 times. How would you say the results differ from those in part b?

7.2 Simulate condo solicitations A company that is selling condos in Florida plans to send out an advertisement for the condos to 500 potential customers, in which they promise a free weekend at a resort on the Florida coast in exchange for agreeing to attend a four-hour sales presentation. The company would like to know how many people will accept this invitation. Its best guess is that there is a 10% chance that any particular customer will accept the offer. The company decides to simulate about what proportion could actually accept the offer, if this is the case. Simulate this scenario for the company, using the Sampling Distribution web app accessible from the book's website. Refer to Activity 1 for guidance on using the app.

TECH

a. Perform one simulation for a sample of size 500. What sample proportion did you get? Why do you not expect to get exactly 0.10?

b. Now simulate 10,000 times. Keep the sample size at $n = 500$ and $p = 0.10$. Describe the graph representing the simulated sampling distribution of the 10,000 sample proportion values. Does it seem likely that the sample proportion will fall close to 10%, say within 5 percentage points of 10%? (*Note:* An option in the app allows you to zoom in on the x-axis.)

7.3 House owners in a district In order to estimate the proportion p of people who own houses in a district, we choose a random sample from the population and study its sampling distribution. Assuming $p = 0.3$, use the appropriate formulas from this section to find the mean and the standard deviation of the sampling distribution of the sample proportion for a random sample of size:

TRY

a. $n = 400$.

b. $n = 1600$.

c. $n = 100$.

d. Summarize the effect of the sample size on the size of the standard deviation.

7.4 iPhone apps Let $p = 0.25$ be the proportion of iPhone owners who have a given app. For a particular iPhone owner, let $x = 1$ if they have the app and $x = 0$ otherwise. For a random sample of 50 owners:

a. State the population distribution (that is, the probability distribution of X for each observation).

b. State the data distribution if 30 of the 50 owners sampled have the app. (That is, give the sample proportions of observed 0s and 1s in the sample.)

c. Find the mean of the sampling distribution of the sample proportion who have the app among the 50 people.

d. Find the standard deviation of the sampling distribution of the sample proportion who have the app among the 50 people.

e. Explain what the standard deviation in part d describes.

7.5 Other scenario for exit poll Refer to Examples 1 and 2 about the exit poll, for which the sample size was 3889. In that election, 40.9% voted for Whitman.

TRY

a. Define a binary random variable X taking values 0 and 1 that represents the vote for a particular voter ($1 =$ vote for Whitman and $0 =$ another candidate). State its probability distribution, which is the same as the population distribution for X.

b. Find the mean and standard deviation of the sampling distribution of the proportion of the 3889 people in the sample who voted for Whitman.

7.6 Exit poll and *n* Refer to the previous exercise.

TRY

a. In part b, if the sampling distribution of the sample proportion had mean 0.409 and the standard deviation 0.008, give an interval of values within which the sample proportion will almost certainly fall. (*Hint*: You can use the approximate normality of the sampling distribution.)

b. The sample proportion for Whitman from the exit poll was 0.424. Using part a, was this one of the plausible values expected in an exit poll? Why?

7.7 Random variability in baseball A baseball player in the major leagues who plays regularly will have about 500 at-bats (that is, about 500 times he can be the hitter in a game) during a season. Suppose a player has a 0.300 probability of getting a hit in an at-bat. His batting average at the end of the season is the number of hits divided by the number of at-bats. When we consider the 500 at-bats as a random sample of all possible at-bats for this player, this batting average is a sample proportion, so it has a sampling distribution describing where it is likely to fall.

a. Describe the shape, mean, and standard deviation of the sampling distribution of the player's batting average.

b. Explain why a batting average of 0.320 or of 0.280 would not be especially unusual for this player's year-end batting average. (That is, you should not conclude that someone with a batting average of 0.320 is necessarily a better hitter than a player with a batting average of 0.280. Both players could have a probability of 0.300 of getting a hit.)

7.8 Awareness about cancer An experiment consists of asking your friends if they would like to raise money for a cancer association. Assuming half of your friends would agree to raise money, construct the sampling distribution of the sample proportion of affirmative answers obtained for a sample of:

a. One friend. (*Hint*: Find the possible sample proportion values and their probabilities)

b. Two friends. (*Hint*: The possible sample proportion values are 0, 0.50, and 1.0. What are their probabilities?)

c. Three friends. (*Hint*: There are 4 possible sample proportion values.)

d. Refer to parts a–c. Sketch the sampling distributions and describe how the shape is changing as the number of friends n increases.

7.9 Buying a car A car dealer offers a \$500 discount to customers if they agree to buy a car immediately without doing further research. Suppose 30% of all customers who visit him accept this offer. Depending on whether or not a given customer accepts the offer, let X be either 1 or 0, respectively.

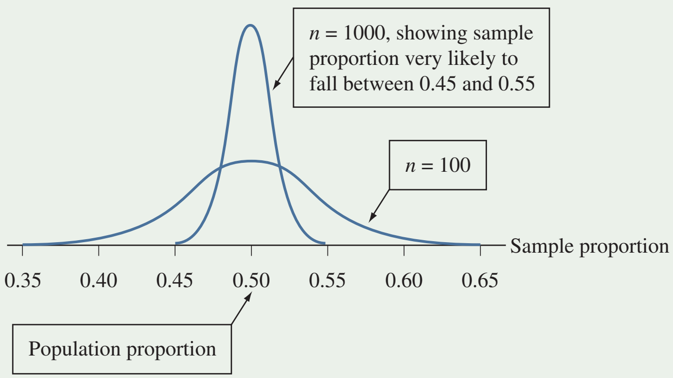
a. If $n = 5$ customers, find the probability distribution of the proportion of customers who will accept the offer. (*Hint:* List all possible values for the sample proportion and their chances of occurring.)

b. Referring to part a, what are the mean and standard deviation of the sample proportion?

c. Repeat part b for a group of $n = 10$ customers and $n = 100$ customers.

d. What happens to the mean and standard deviation of the sample proportion as n increases?

7.10 Effect of *n* on sample proportion The figure illustrates two sampling distributions for sample proportions when the population proportion $p = 0.50$.



a. Find the standard deviation for the sampling distribution of the sample proportion with (i) $n = 100$ and (ii) $n = 1000$.

b. Explain why the sample proportion would be very likely (as the figure suggests) to fall (i) between 0.35 and 0.65 when $n = 100$, and (ii) between 0.45 and 0.55 when $n = 1000$. (*Hint:* Recall that for an approximately normal distribution, nearly the entire distribution is within 3 standard deviations of the mean.)

c. Explain how the results in part b indicate that the sample proportion tends to estimate the population proportion more precisely when the sample size is larger.

7.11 Syracuse full-time students You'd like to estimate
TRY TECH
the proportion of the 14,201 undergraduate students at Syracuse University who are full-time students. You poll a random sample of 350 students, of whom 330 are full-time. Unknown to you, the proportion of all undergraduate students who are full-time students is 0.951. Let X denote a random variable for which $x = 1$ denotes full-time student and for which $x = 0$ denotes part-time student. (For recent enrollment numbers, go to www.syr.edu/about/facts.html.)

a. Describe the population distribution. Sketch a graph representing the population distribution.

b. Describe the data distribution. Sketch a graph representing the data distribution.

c. Find the mean and standard deviation of the sampling distribution of the sample proportion for a sample of size 350. Explain what this sampling distribution represents. Sketch a graph representing this sampling distribution.

d. Use the Sampling Distribution app accessible from the book's website to check your answers from parts a through c. Set the population proportion equal to $p = 0.951$ and $n = 350$. Compare the population, data, and sampling distribution graph from the app with your graphs from parts a through c.

7.12 Gender distributions At a university, 60% of the
TECH
7,400 students are female. The student newspaper reports results of a survey of a random sample of 50 students about various topics involving alcohol abuse, such as participation in binge drinking. They report that their sample contained 26 females.

a. Explain how you can set up a binary random variable X to represent gender.

b. Identify the population distribution of gender at this university. Sketch a graph.

c. Identify the data distribution of gender for this sample. Sketch a graph.

d. Identify the sampling distribution of the sample proportion of females in the sample. State its mean and standard deviation for a random sample of size 50. Sketch a graph.

e. Use the Sampling Distribution app accessible from the book's website to check your answers from parts b through d. Set the population proportion equal to $p = 0.60$ and $n = 50$. Compare the population, data, and sampling distribution graph from the app with your graphs from parts b through d.

7.13 Shapes of distributions

TRY TECH

a. With random sampling, does the shape of the data distribution tend to resemble more closely the sampling distribution or the population distribution? Explain.

b. Is the sampling distribution of the sample proportion always bell shaped? Investigate with the Sampling Distribution app by setting $n = 30$ and increasing the population proportion p from 0.5 to 0.95. (You can do this by clicking the box for p.) What do you observe?

c. Why is it inappropriate to assume a bell shape for the sampling distribution of the sample proportion when $p = 0.95$ and $n = 30$?

7.14 Beauty contest election A finalist of a Miss University contest believes that 52% of Facebook voters will vote for her. However, she is worried about low voter turnout.

a. Assuming she truly has the support of 52% of all Facebook voters, find the mean and standard deviation of the sampling distribution for the proportion of the votes she will receive if only $n = 400$ Facebook users were to vote willingly in this competition.

b. Is it reasonable to assume a normal shape for this sampling distribution? Explain.

c. How likely is it that she will not get the majority of the votes, i.e., a sample proportion of 50% or lower from the 400 votes cast?

d. If instead $n = 2000$ Facebook users are voting in the competition, then how likely is it that she will not win the majority of the votes?

7.2 Practicing the Basics

7.15 Simulate taking midterms Assume that the distribution of the score on a recent midterm is bell shaped with population mean $\mu = 70$ and population standard deviation $\sigma = 10$. You randomly sample $n = 12$ students who took the midterm. Using the Sampling Distribution of the Sample Mean web app accessible on the book's website,

TRY TECH

a. Simulate drawing one random sample of size 12. What sample mean score did you get? Why do you not expect to get exactly 70?

b. Now, simulate drawing 10,000 samples of size 12. Describe the simulated sampling distribution by using the histogram of the 10,000 sample means you generated. (*Note*: The app allows you to save the graph to a file.)

c. Suppose the population standard deviation is 5 rather than 10. Change to this setting in the app. How does the simulated sampling distribution of the sample mean change?

7.16 Education of the self-employed According to a recent Current Population Reports, the population distribution of number of years of education for self-employed individuals in the United States has a mean of 13.6 and a standard deviation of 3.0.

TRY

a. Identify the random variable X whose population distribution is described here.

b. Find the mean and standard deviation of the sampling distribution of $\bar{x}$ for a random sample of size 100. Interpret the results.

c. Repeat part b for $n = 400$. Describe the effect of increasing n.

7.17 Rolling one die Let X denote the outcome of rolling a die.

a. Construct a graph of the (i) probability distribution of X and (ii) sampling distribution of the sample mean for $n = 2$. (You can think of (i) as the population distribution you would get if you could roll the die an infinite number of times.

b. The probability distribution of X has mean 3.50 and standard deviation 1.71. Find the mean and standard deviation of the sampling distribution of the sample mean for (i) $n = 2$, (ii) $n = 30$. What is the effect of n on the sampling distribution?

7.18 Performance of airlines In 2015, the on-time arrival rate of all major domestic and regional airlines operating between Australian airports has a bell-shaped distribution roughly with mean 0.86 and standard deviation 0.1.

TRY

a. Let X denote the number of flights arriving on time when you observe one flight. State the probability distribution of X. (This also represents the population distribution you would get if you could observe an infinite number of flights.)

b. You decide to observe the airport arrival tables for one day. At the end of the day, you were able to check the arrival times of 100 flights. Show that the sampling distribution of your sample mean number of flights on time has mean = 0.86 and standard deviation = 0.01.

c. Refer to part b. Using the central limit theorem, find the probability that the mean number of flights on time is at least 0.88, so that you have a gain of at least 2% with regard to the rate of on-time flights in the population. (*Hint*: Find the probability that a normal random variable with mean 0.86 and standard deviation 0.01 exceeds 0.88.)

7.19 Simulate rolling dice Access the Sampling Distribution of the Sample Mean (discrete variable) web app on the book's website. Enter the probabilities $P(X = x)$ of 0.1667 for the numbers 1 through 6 to specify the probability distribution of a fair die. (This is a discrete version of the uniform distribution shown in the first column of Figure 7.11.) The resulting population distribution has $\mu = 3.5$ and $\sigma = 1.71$.

TRY TECH

a. In the box for the sample size n, enter 2 to simulate rolling two dice. Then, press the Draw Sample(s) button several times and observe how the histogram for the simulated sampling distribution for the mean number shown on two rolls is building up. Finally, simulate rolling two dice and finding their average 10,000 times by selecting the corresponding option. Describe (shape, center, spread) the resulting simulated sampling distribution of the sample mean, using the histogram of the 10,000 generated sample means. (*Note:* Statistics for the simulated sampling distribution are reported in the tile of its plot.)

b. Are the reported mean and standard deviation of the simulated sampling distribution close to the theoretical mean and standard deviation for this sampling distribution? Compute the theoretical values and compare.

c. Repeat part a, but now simulate rolling $n = 30$ dice an finding their average face value. What are the major changes you observe in the simulated sampling distribution?

7.20 Canada lottery In one lottery option in Canada (*Source:* Lottery Canada), you bet on a six-digit number between 000000 and 999999. For a \$1 bet, you win \$100,000 if you are correct. The mean and standard deviation of the probability distribution for the lottery winnings are $\mu = 0.10$ (that is, 10 cents) and $\sigma = 100.00$. Joe figures that if he plays enough times every day, eventually he will strike it rich, by the law of large numbers. Over the course of several years, he plays 1 million times. Let $\bar{x}$ denote his average winnings.

TRY

a. Find the mean and standard deviation of the sampling distribution of $\bar{x}$.

b. About how likely is it that Joe's average winnings exceed \$1, the amount he paid to play each time? Use the central limit theorem to find an approximate answer.

7.21 Shared family phone plan A recent personalized information sheet from your wireless phone carrier claims that the mean duration of all your phone calls was $\mu = 2.8$ minutes with a standard deviation of $\sigma = 2.1$ minutes.

a. Is the population distribution of the duration of your phone calls likely to be bell shaped, right-, or left-skewed?

b. You are on a shared wireless plan with your parents, who are statisticians. They look at some of your recent monthly statements that list each call and its duration and randomly sample 45 calls from the thousands listed there. They construct a histogram of the duration to look at the data distribution. Is this distribution likely to be bell shaped, right-, or left-skewed?

c. From the sample of $n = 45$ calls, your parents compute the mean duration. Is the sampling distribution of the sample mean likely to be bell shaped, right-, or left-skewed, or is it impossible to tell? Explain.

7.22 Dropped from plan The previous exercise mentions that the duration of your phone calls follows a distribution with mean $\mu = 2.8$ minutes and standard deviation $\sigma = 2.1$ minutes. From a random sample of $n = 45$ calls, your parents computed a sample mean of $\bar{x} = 3.4$ minutes and a sample standard deviation of $s = 2.9$ minutes.

TRY

a. Sketch the population distribution for the duration of your phone calls. What are its mean and standard deviation?

b. What are the mean and standard deviation of the data distribution?

c. Find the mean and standard deviation of the sampling distribution of the sample mean.

d. Is the sample mean of 3.4 minutes unusually high? Find its z-score and comment.

e. Your parents told you that they will kick you off the plan when they find a sample mean larger than 3.5 minutes. How likely is this to happen?

7.23 Restaurant profit? Jan's All You Can Eat Restaurant charges \$8.95 per customer to eat at the restaurant. Restaurant management finds that its expense per customer, based on how much the customer eats and the expense of labor, has a distribution that is skewed to the right with a mean of \$8.20 and a standard deviation of \$3.

a. If the 100 customers on a particular day have the characteristics of a random sample from their customer base, find the mean and standard deviation of the sampling distribution of the restaurant's sample mean expense per customer.

b. Find the probability that the restaurant makes a profit that day, with the sample mean expense being less than \$8.95. (*Hint:* Apply the central limit theorem to the sampling distribution in part a.)

7.24 Survey accuracy According to the U.S. Census Bureau, Current Population Survey, Annual Social and Economic Supplement, the average income for females was \$28,466 and the standard deviation was \$36,961 in 2015. A sample of 1,000 females was randomly chosen from the entire United States population to verify if this sample would have a similar mean income as the entire population.

a. Find the probability that the mean income of the females sampled is within two thousand of the mean income for all females. (*Hint*: Find the sampling distribution of the sample mean income and use the central limit theorem).

b. Would the probability be larger or smaller if the standard deviation of all females' incomes was \$25,000? Why?

7.25 Blood pressure Vincenzo Baranello was diagnosed with high blood pressure. He was able to keep his blood pressure in control for several months by taking blood pressure medicine (amlodipine besylate). Baranello's blood pressure is monitored by taking three readings a day, in early morning, at midday, and in the evening.

a. During this period, the probability distribution of his systolic blood pressure reading had a mean of 130 and a standard deviation of 6. If the successive observations behave like a random sample from this distribution, find the mean and standard deviation of the sampling distribution of the sample mean for the three observations each day.

b. Suppose that the probability distribution of his blood pressure reading is normal. What is the shape of the sampling distribution? Why?

c. Refer to part b. Find the probability that the sample mean exceeds 140, which is considered problematically high. (*Hint*: Use the sampling distribution, not the probability distribution for each observation.)

7.26 Average price of an ebook TRY According to the website http://www.digitalbookworld.com, the average price of a bestselling ebook increased to \$8.05 in the week of February 18, 2015 from \$6.89 in the previous week. Assume the standard deviation of the price of a bestselling ebook is \$1 and suppose you have a sample of 20 bestselling ebooks with a sample mean of \$7.80 and a standard deviation of \$0.95.

a. Identify the random variable X in this study. Indicate whether it is quantitative or categorical.

b. Describe the center and variability of the population distribution. What would you predict as the shape of the population distribution? Explain.

c. Describe the center and variability of the data distribution. What would you predict as the shape of the data distribution? Explain.

d. Describe the center and variability of the sampling distribution of the sample mean for 20 bestselling ebooks. What would you predict as the shape of the sampling distribution? Explain.

7.27 Average time to fill job positions For all job positions in a company, assume that, a few years ago, the average time to fill a job position was 37 days with a standard deviation of 12 days. For the purpose of comparison, the manager of the hiring department selected a random sample of 100 of today's job positions. He observed a sample mean of 39 days and a standard deviation of 13 days.

a. Describe the center and variability of the population distribution. What shape does it probably have? Explain.

b. Describe the center and variability of the data distribution. What shape does it probably have? Explain.

c. Describe the center and variability of the sampling distribution of the sample mean for $n = 100$. What shape does it have? Explain.

d. Explain why it would not be unusual to observe a job position that would take more than 55 days to fill, but it would be highly unusual to observe a sample mean of more than 50 days for a random sample size of 100 job positions.

7.28 Central limit theorem for uniform population TRY TECH Let's use the Sampling Distribution of the Sample Mean web app accessible from the book's website to show that the first population distribution shown in Figure 7.11 has a more nearly normal sampling distribution for the mean as n increases. Select uniform for the population distribution and keep the default settings of 0 and 1 for the lower and upper bound.

a. Use the app to simulate the sampling distribution when the sample size $n = 2$. Run 10,000 simulations and look at the resulting histogram of the sample means. (You may want to decrease the binsize a bit to get a clearer picture.) What shape does the simulated sampling distribution have?

b. Repeat part a, but now use a sample size of $n = 5$. Explain how the variability and the shape of the simulated sampling distribution changes as n increases from 2 to 5.

c. Repeat part a, but now use a sample size of $n = 30$. Explain how the variability and the shape of the simulated sampling distribution changes as n increases from 2 to 30. Compare results from parts a-c to the first column of Figure 7.11.

d. Explain how the central limit theorem describes what you have observed.

7.29 CLT for skewed population TRY TECH Access the Sampling Distribution of the Sample Mean web app and position the slider for the skewness of the default population distribution to the smallest value. The population distribution now looks similar to the one in the third column of Figure 7.11. Repeat parts a–c of the previous exercise and explain how the variability and shape of the simulated sampling distribution of the sample mean changes as n changes from 2 to 5 to 30. Explain how the central limit theorem describes what you have observed.

7.30 Sampling distribution for normal population TRY TECH Access the Sampling Distribution of the Sample Mean web app and select Bell-Shaped for the shape of the population distribution, which looks similar to the fourth column in Figure 7.11. Select a value for the population mean and standard deviation. Is the sampling distribution normal even for $n = 2$? What does this tell you?