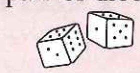
chapter 7

Random Variables

ACTIVITY 7 The Game of Craps

Materials: Pair of dice for each pair of students

The game of craps is one of the most famous (or notorious) of all gambling games played with dice. In this game, the player rolls a pair of six-sided dice, and the *sum* of the numbers that turn up on the two faces is noted. If the sum is 7 or 11, then the player wins immediately. If the sum is 2, 3, or 12, then the player loses immediately. If any other sum is obtained, then the player continues to throw the dice until he either wins by repeating the first sum he obtained or loses by rolling a 7. Your mission in this activity is to estimate the probability of a player winning at craps. But first, let's get a feel for the game. For this activity, your class will be divided into groups of two. Your instructor will provide a pair of dice for each group of two students.



1. In your group of two students, play a total of 20 games of craps. One person will roll the dice; the other will keep track of the sums and record the end result (win or lose). If you like, you can switch jobs after 10 games have been completed. How many times out of 20 does the player win? What is the relative frequency (i.e., percentage, written as a decimal) of wins?

2. Combine your results with those of the other two-student groups in the class. What is the relative frequency of wins for the entire class?

3. Use simulation techniques to represent 25 games of craps, using either the table of random numbers or the random number generating feature of your TI-83/89. What is the relative frequency of wins based on the 25 simulations? How does this number compare to the relative frequency you found in step 2?

4. One of the ways you can win at craps is to roll a sum of 7 or 11 on your first roll. Using your results and those of your fellow students, determine the number of times a player won by rolling a sum of 7 of the first roll. What is the relative frequency of rolling a sum of 7? Repeat these calculations for a sum of 11. Which of these sums appears more likely to occur than the other, based on the class results?

5. One of the ways you can lose at craps is to roll a sum of 2, 3, or 12 on your first roll. Using your results and those of your fellow students, determine the number of times a player lost by rolling a sum of 2 on the first roll. What is the relative frequency of rolling a sum of 2? Repeat these calculations for a sum of 3 and a sum of 12. Which of these sums appears more likely to occur than the others, based on the class results?

6. Clearly, the key quantity of interest in craps is the *sum* of the numbers on the two dice. Let's try to get a better idea of how this sum behaves in general by conducting a simulation. First, determine how you would simulate

ACTIVITY 7 The Game of Craps (*continued*)

the roll of a single fair die. (*Hint*: Just use digits 1 to 6 and ignore the others.) Then determine how you would simulate a roll of two fair dice. Using this model, simulate 36 rolls of a pair of dice and determine the relative frequency of each of the possible sums.

7. Construct a relative frequency histogram of the relative frequency results in step 6. What is the approximate shape of the distribution? What sum appears most likely to occur? Which appears least likely to occur?

8. From the relative frequency data in step 6, compute the relative frequency of winning and the relative frequency of losing on your first roll in craps. How do these simulated results compare with what the class obtained?

INTRODUCTION

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

RANDOM VARIABLE

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

outcomes such as the mean $\bar{x}$ of a random sample, for which we will keep the familiar notation.[1] As we progress from general rules of probability toward statistical inference, we will concentrate on random variables. When a random variable X describes a random phenomenon, the sample space S just lists the possible values of the random variable. We usually do not mention S separately. There remains the second part of any probability model, the assignment of probabilities to events. In this section, we will learn two ways of assigning probabilities to the values of a random variable. The two types of probability models that result will dominate our application of probability to statistical inference.

7.1 DISCRETE AND CONTINUOUS RANDOM VARIABLES

Discrete random variables

We have learned several rules of probability but only one method of assigning probabilities: state the probabilities of the individual outcomes and assign probabilities to events by summing over the outcomes. The outcome probabilities must be between 0 and 1 and have sum 1. When the outcomes are numerical, they are values of a random variable. We will now attach a name to random variables having probability assigned in this way.[2]

DISCRETE RANDOM VARIABLE

A **discrete random variable** X has a countable number of possible values. The **probability distribution** of X lists the values and their probabilities:

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

The probabilities p_i must satisfy two requirements:

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

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

EXAMPLE 7.1 GETTING GOOD GRADES

The instructor of a large class gives 15% each of A's and D's, 30% each of B's and C's, and 10% F's. Choose a student at random from this class. To "choose at random" means to give every student the same chance to be chosen. The student's grade on a four-point scale (A = 4) is a random variable X.

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

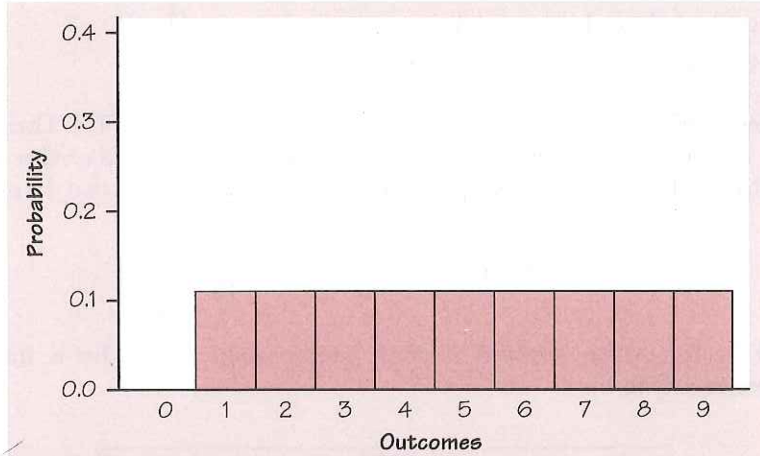
Grade:	0	1	2	3	4
Probability:	0.10	0.15	0.30	0.30	0.15

The probability that the student got a B or better is the sum of the probabilities of an A and a B:

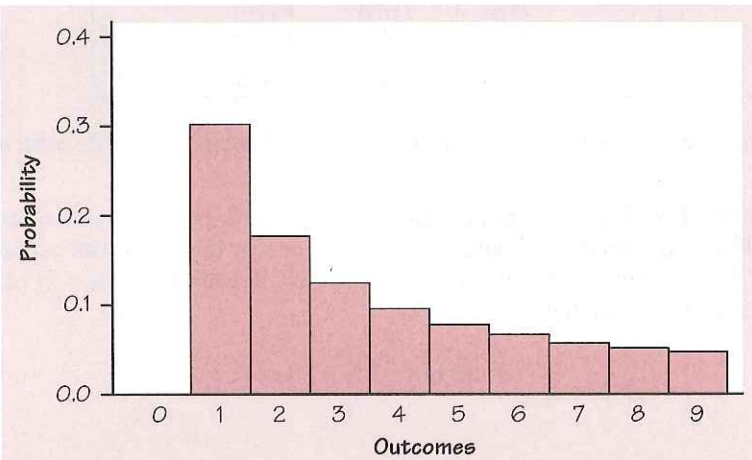
$$\begin{aligned} P(\text{grade is 3 or 4}) &= P(X = 3) + P(X = 4) \\ &= 0.30 + 0.15 = 0.45 \end{aligned}$$

We can use histograms to display probability distributions as well as distributions of data. Figure 7.1 displays **probability histograms** that compare the probability model for random digits (Example 6.11 page 347) with the model given by Benford's law (Example 6.10 page 345)

probability histogram



(a)



(b)

FIGURE 7.1 Probability histograms for (a) random digits 1 to 9 and (b) Benford's law. The height of each bar shows the probability assigned to a single outcome.

The height of each bar shows the probability of the outcome at its base. Because the heights are probabilities, they add to 1. As usual, all the bars in a histogram have the same width. So the areas of the bars also display the assignment of probability to outcomes. Think of these histograms as idealized pictures of the results of very many trials. The histograms make it easy to quickly compare the two distributions.

EXAMPLE 7.2 TOSSING COINS

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

1. The coin is balanced, so each toss is equally likely to give H or T.
2. The coin has no memory, so tosses are independent.

The outcome of four tosses is a sequence of heads and tails such as HTTH. There are 16 possible outcomes in all. Figure 7.2 lists these outcomes along with the value of X for each outcome. The multiplication rule for independent events tells us that, for example,

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

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

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

FIGURE 7.2 Possible outcomes in four tosses of a coin. The random variable X is the number of heads.

The number of heads X has possible values 0, 1, 2, 3, and 4. These values are *not* equally likely. As Figure 7.2 shows, there is only one way that X = 0 can occur: namely when the outcome is TTTT. So $P(X = 0) = 1/16$. But the event $\{X = 2\}$ can occur in six different ways, so that

$$P(X = 2) = \frac{\text{count of ways } X = 2 \text{ can occur}}{16} = \frac{6}{16}$$

We can find the probability of each value of X from Figure 7.2 in the same way. Here is the result:

$$P(X = 0) = \frac{1}{16} = 0.0625 \qquad P(X = 1) = \frac{4}{16} = 0.25 \qquad P(X = 2) = \frac{6}{16} = 0.37$$

$$P(X = 3) = \frac{4}{16} = 0.25 \qquad P(X = 4) = \frac{1}{16} = 0.0625$$

These probabilities have sum 1, so this is a legitimate probability distribution. In table form the distribution is

Number of heads:	0	1	2	3	4
Probability:	0.0625	0.25	0.375	0.25	0.0625

Figure 7.3 is a probability histogram for this distribution. The probability distribution is exactly symmetric. It is an idealization of the relative frequency distribution of the number of heads after many tosses of four coins, which would be nearly symmetric but is unlikely to be exactly symmetric.

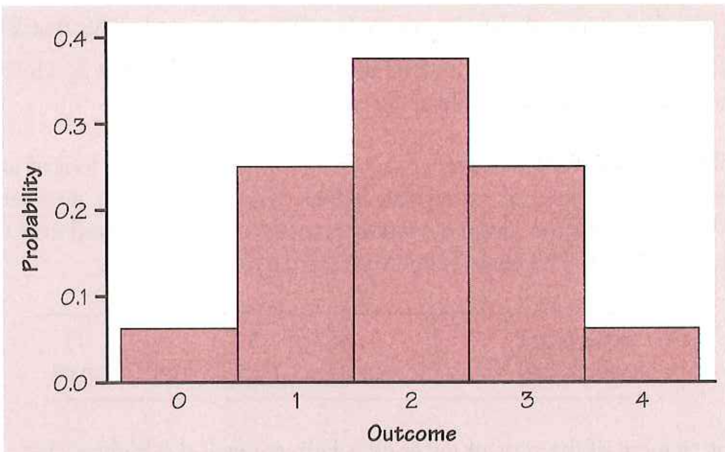


FIGURE 7.3 Probability histogram for the number of heads in four tosses of a coin.

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

$$P(X \geq 2) = 0.375 + 0.25 + 0.0625 = 0.6875$$

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

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

Recall that tossing a coin n times is similar to choosing an SRS of size n from a large population and asking a yes-or-no question. We will extend the results of Example 7.2 when we return to sampling distributions in the next two chapters.

EXERCISES

7.1 ROLL OF THE DIE If a carefully made die is rolled once, it is reasonable to assign probability 1/6 to each of the six faces.

(a) What is the probability of rolling a number less than 3?

(b) Use your TI-83/89 to simulate rolling a die 100 times, and assign the values to L_1/list1. Sort the list in ascending order, and then count the outcomes that are either 1s or 2s. Record the relative frequency.

(c) Repeat part (b) four more times, and then average the five relative frequencies. Is this number close to your result in (a)?

7.2 THREE CHILDREN A couple plans to have three children. There are 8 possible arrangements of girls and boys. For example, GGB means the first two children are girls and the third child is a boy. All 8 arrangements are (approximately) equally likely.

(a) Write down all 8 arrangements of the sexes of three children. What is the probability of any one of these arrangements?

(b) Let X be the number of girls the couple has. What is the probability that $X = 2$?

(c) Starting from your work in (a), find the distribution of X. That is, what values can X take, and what are the probabilities for each value?

7.3 SOCIAL CLASS IN ENGLAND A study of social mobility in England looked at the social class reached by the sons of lower-class fathers. Social classes are numbered from 1 (low) to 5 (high). Take the random variable X to be the class of a randomly chosen son of a father in Class 1. The study found that the distribution of X is

Son's class:	1	2	3	4	5
Probability:	0.48	0.38	0.08	0.05	0.01

(a) What percent of the sons of lower-class fathers reach the highest class, Class 5?

(b) Check that this distribution satisfies the requirements for a discrete probability distribution.

(c) What is $P(X \leq 3)$?

(d) What is $P(X < 3)$?

(e) Write the event "a son of a lower-class father reaches one of the two highest classes" in terms of values of X. What is the probability of this event?

(f) Briefly describe how you would use simulation to answer the question in (c).

7.4 HOUSING IN SAN JOSE, I How do rented housing units differ from units occupied by their owners? Here are the distributions of the number of rooms for owner-occupied units and renter-occupied units in San Jose, California:[3]

Rooms:	1	2	3	4	5	6	7	8	9	10
Owned:	0.003	0.002	0.023	0.104	0.210	0.224	0.197	0.149	0.053	0.035
Rented:	0.008	0.027	0.287	0.363	0.164	0.093	0.039	0.013	0.003	0.003

Make probability histograms of these two distributions, using the same scales. What are the most important differences between the distributions of owner-occupied and rented housing units?

7.5 HOUSING IN SAN JOSE, II Let the random variable X be the number of rooms in a randomly chosen owner-occupied housing unit in San Jose, California. Exercise 7.4 gives the distribution of X.

(a) Express "the unit has five or more rooms" in terms of X. What is the probability of this event?

(b) Express the event $\{X > 5\}$ in words. What is its probability?

(c) What important fact about discrete random variables does comparing your answers to (a) and (b) illustrate?

Continuous random variables

When we use the table of random digits to select a digit between 0 and 9, the result is a discrete random variable. The probability model assigns probability 1/10 to each of the 10 possible outcomes, as Figure 7.1(a) shows. Suppose that we want to choose a number at random between 0 and 1, allowing *any* number between 0 and 1 as the outcome. Software random number generators will do this. You can visualize such a random number by thinking of a spinner (Figure 7.4) that turns freely on its axis and slowly comes to a stop. The pointer can come to rest anywhere on a circle that is marked from 0 to 1. The sample space is now an entire interval of numbers:

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

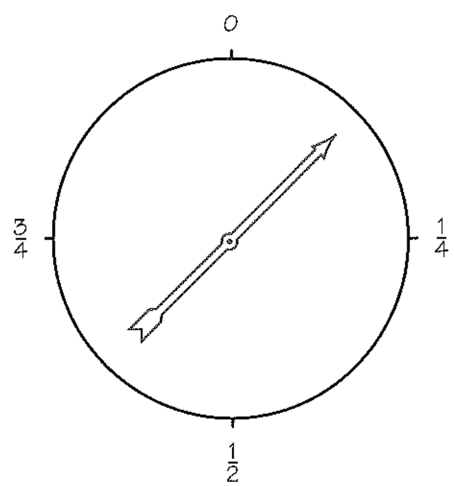


FIGURE 7.4 A spinner that generates a random number between 0 and 1.

How can we assign probabilities to such events as $0.3 \leq x \leq 0.7$? As in the case of selecting a random digit, we would like all possible outcomes to be equally likely. But we cannot assign probabilities to each individual value of x and then sum, because there are infinitely many possible values. Instead we use a new way of assigning probabilities directly to events—as *areas under a density curve*. Any density curve has area exactly 1 underneath it, corresponding to total probability 1.

EXAMPLE 7.3 RANDOM NUMBERS AND THE UNIFORM DISTRIBUTION

uniform distribution

The random number generator will spread its output uniformly across the entire interval from 0 to 1 as we allow it to generate a long sequence of numbers. The results of many trials are represented by the density curve of a ***uniform distribution*** (Figure 7.5). This density curve has height 1 over the interval from 0 to 1. The area under the density curve is 1, and the probability of any event is the area under the density curve and above the event in question.

As Figure 7.5(a) illustrates, the probability that the random number generator produces a number X between 0.3 and 0.7 is

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

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

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

Notice that the last event consists of two nonoverlapping intervals, so the total area above the event is found by adding two areas, as illustrated by Figure 7.5(b). This assignment of probabilities obeys all of our rules for probability.

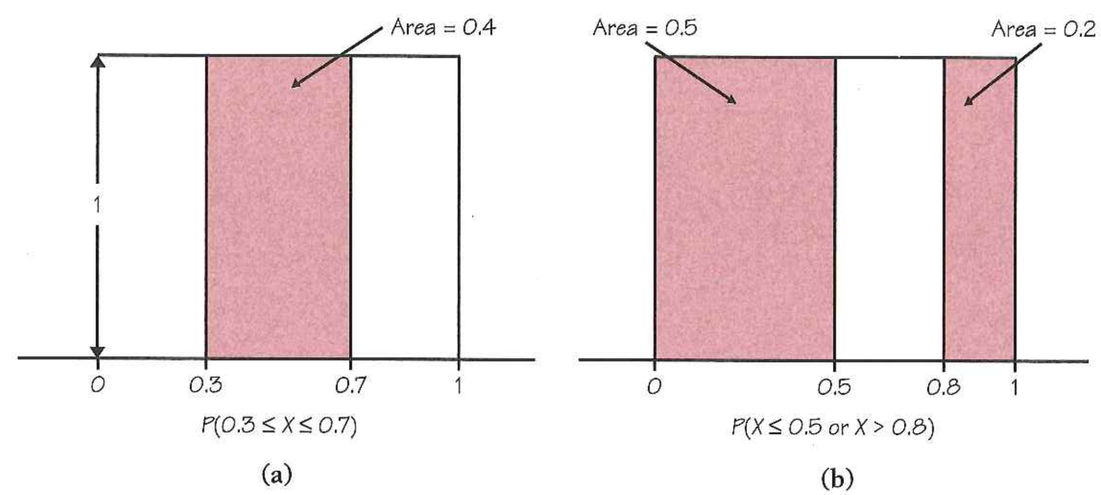


FIGURE 7.5 Assigning probability for generating a random number between 0 and 1. The probability of any interval of numbers is the area above the interval and under the curve.

Probability as area under a density curve is a second important way of assigning probabilities to events. Figure 7.6 illustrates this idea in general form.

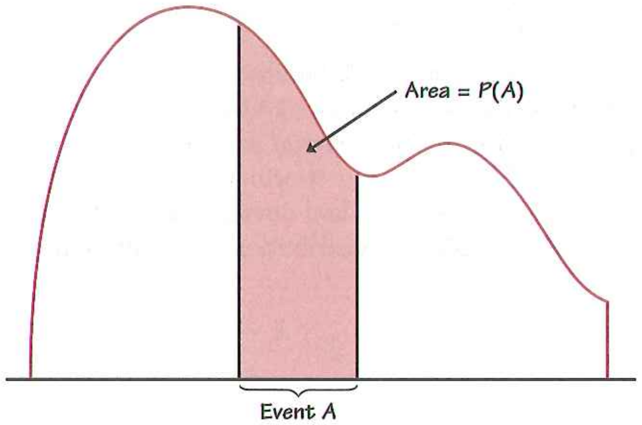


FIGURE 7.6 The probability distribution of a continuous random variable assigns probabilities as areas under a density curve.

We call X in Example 7.3 a *continuous random variable* because its values are not isolated numbers but an entire interval of numbers.

CONTINUOUS RANDOM VARIABLE

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

The probability model for a continuous random variable assigns probabilities to intervals of outcomes rather than to individual outcomes. In fact, **all continuous probability distributions assign probability 0 to every individual outcome.** Only intervals of values have positive probability. To see that this is true, consider a specific outcome such as $P(X = 0.8)$ in Example 7.3. The probability of any interval is the same as its length. The point 0.8 has no length, so its probability is 0. Although this fact may seem odd at first glance, it does make intuitive as well as mathematical sense. The random number generator produces a number between 0.79 and 0.81 with probability 0.02. An outcome between 0.799 and 0.801 has probability 0.002, and a result between 0.7999 and 0.8001 has probability 0.0002. Continuing to home in on 0.8, we can see why an outcome *exactly* equal to 0.8 should have probability 0. Because there is no probability exactly at $X = 0.8$, the two events $\{X > 0.8\}$ and $\{X \geq 0.8\}$ have the same probability. We can ignore the distinction between $>$ and $\geq$ when finding probabilities for continuous (but not discrete) random variables.

Normal distributions as probability distributions

The density curves that are most familiar to us are the normal curves. (We discussed normal curves in Section 2.1.) Because any density curve describes an assignment of probabilities, *normal distributions are probability distributions*. Recall that $N(\mu, \sigma)$ is our shorthand notation for the normal distribution having mean μ and standard deviation σ. In the language of random variables, if X has the $N(\mu, \sigma)$ distribution, then the standardized variable

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

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

EXAMPLE 7.4 DRUGS IN SCHOOLS

An opinion poll asks an SRS of 1500 American adults what they consider to be the most serious problem facing our schools. Suppose that if we could ask all adults this question, 30% would say "drugs." We will learn in Chapter 9 that the proportion $p = 0.3$ is a population parameter and that the proportion $\hat{p}$ of the sample who answer "drugs" is a statistic used to estimate p. We will see in Chapter 9 that $\hat{p}$ is a random variable that has approximately the $N(0.3, 0.0118)$ distribution. The mean 0.3 of this distribution is the same as the population parameter because $\hat{p}$ is an unbiased estimate of p. The standard deviation is controlled mainly by the sample size, which is 1500 in this case.

What is the probability that the poll result differs from the truth about the population by more than two percentage points? Figure 7.7 shows this probability as an area under a normal density curve.

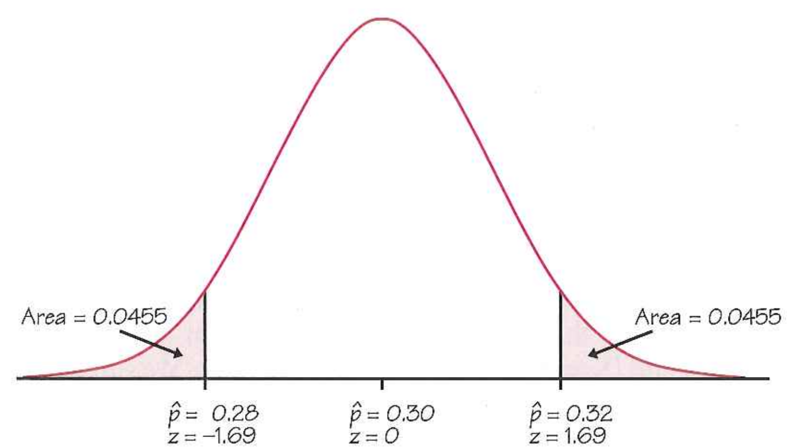


FIGURE 7.7 Probability in Example 7.4 as area under a normal density curve.

By the addition rule for disjoint events, the desired probability is

$$P(\hat{p} < 0.28 \text{ or } \hat{p} > 0.32) = P(\hat{p} < 0.28) + P(\hat{p} > 0.32)$$

You can find the two individual probabilities from software or by standardizing and using Table A.

$$\begin{aligned} P(\hat{p} < 0.28) &= P\left(Z < \frac{0.28 - 0.3}{0.0118}\right) \\ &= P(Z < -1.69) = 0.0455 \\ P(\hat{p} > 0.32) &= P\left(Z > \frac{0.32 - 0.3}{0.0118}\right) \\ &= P(Z > 1.69) = 0.0455 \end{aligned}$$

Therefore,

$$P(\hat{p} < 0.28 \text{ or } \hat{p} > 0.32) = 0.0455 + 0.0455 = 0.0910$$

The probability that the sample result will miss the truth by more than two percentage points is 0.091. The arrangement of this calculation is familiar from our earlier work with normal distributions. Only the language of probability is new.

We could also do the calculation by first finding the probability of the complement:

$$\begin{aligned} P(0.28 \le \hat{p} \le 0.32) &= P\left(\frac{0.28 - 0.3}{0.0118} \le Z \le \frac{0.32 - 0.3}{0.0118}\right) \\ &= P(-1.69 \le Z \le 1.69) \\ &= 0.9545 - 0.0455 = 0.9090 \end{aligned}$$

Then by the complement rule,

$$\begin{aligned} P(\hat{p} < 0.28 \text{ or } \hat{p} > 0.32) &= 1 - P(0.28 \le \hat{p} \le 0.32) \\ &= 1 - 0.9090 = 0.0910 \end{aligned}$$

There is often more than one correct way to use the rules of probability to answer a question.

EXERCISES

7.6 CONTINUOUS RANDOM VARIABLE, I Let X be a random number between 0 and 1 produced by the idealized uniform random number generator described in Example 7.3 and Figure 7.5. Find the following probabilities:

(a) $P(0 \le X \le 0.4)$

(b) $P(0.4 \le X \le 1)$

(c) $P(0.3 \le X \le 0.5)$

(d) $P(0.3 < X < 0.5)$

(e) $P(0.226 \le X \le 0.713)$

(f) What important fact about continuous random variables does comparing your answers to (c) and (d) illustrate?

7.7 CONTINUOUS RANDOM VARIABLE, II Let the random variable X be a random number with the uniform density curve in Figure 7.5, as in the previous exercise. Find the following probabilities:

(a) $P(X \leq 0.49)$

(b) $P(X \geq 0.27)$

(c) $P(0.27 < X < 1.27)$

(d) $P(0.1 \leq X \leq 0.2 \text{ or } 0.8 \leq X \leq 0.9)$

(e) The probability that X is not in the interval 0.3 to 0.8.

(f) $P(X = 0.5)$

7.8 VIOLENCE IN SCHOOLS, I An SRS of 400 American adults is asked, "What do you think is the most serious problem facing our schools?" Suppose that in fact 40% of all adults would answer "violence" if asked this question. The proportion $\hat{p}$ of the sample who answer "violence" will vary in repeated sampling. In fact, we can assign probabilities to values of $\hat{p}$ using the normal density curve with mean 0.4 and standard deviation 0.023. Use this density curve to find the probabilities of the following events:

(a) At least 45% of the sample believes that violence is the schools' most serious problem.

(b) Less than 35% of the sample believes that violence is the most serious problem.

(c) The sample proportion is between 0.35 and 0.45.

7.9 VIOLENCE IN SCHOOLS, II How could you design a simulation to answer part (b) of Exercise 7.8? What we need to do is simulate 400 observations from the $N(0.4, 0.023)$ distribution. This is easily done on the calculator. Here's one way: Clear L_1/list1 and enter the following commands (randNorm is found under the MATH/PRB menu on the TI-83, and in the CATALOG under FlashApps on the TI-89):

TI-83	TI-89
• `randNorm(0.4,.023,400)→L1`	• `tistat.randNorm(0.4,.023,400)→list1`

This will select 400 random observations from the $N(0.4, 0.023)$ distribution.

TI-83	TI-89
• `SortA(L1)`	• `SortA list1`

This will sort the 400 observations in L_1/list1 in ascending order.

Then scroll through L_1/list1. How many entries (observations) are less than 0.35? What is the relative frequency of this event? Compare the results of your simulation with your answer to Exercise 7.8(b).

SUMMARY

The previous chapter included a general discussion of the idea of probability and the properties of probability models. Two very useful specific types of probability models are distributions of discrete and continuous random variables. In our study of statistics we will employ only these two types of probability models.

A **random variable** is a variable taking numerical values determined by the outcome of a random phenomenon. The **probability distribution** of a random variable X tells us what the possible values of X are and how probabilities are assigned to those values.

A random variable X and its distribution can be discrete or continuous.

A **discrete random variable** has a countable number of possible values. The probability distribution assigns each of these values a probability between 0 and 1 such that the sum of all the probabilities is exactly 1. The probability of any event is the sum of the probabilities of all the values that make up the event.

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

Normal distributions are one type of continuous probability distribution.

You can picture a probability distribution by drawing a **probability histogram** in the discrete case or by graphing the density curve in the continuous case.

When you work problems, get in the habit of first identifying the random variable of interest. X = number of ______ for discrete random variables, and X = amount of ______ for continuous random variables.

SECTION 7.1 EXERCISES

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

Number of persons:	1	2	3	4	5	6	7
Household probability:	0.25	0.32	0.17	0.15	0.07	0.03	0.01
Family probability:	0	0.42	0.23	0.21	0.09	0.03	0.02

(a) Verify that each is a legitimate discrete probability distribution function.

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

7.11 SIZE OF AMERICAN HOUSEHOLDS, II Choose an American household at random and let the random variable Y be the number of persons living in the household. Exercise 7.10 gives the distribution of Y.

(a) Express "more than one person lives in this household" in terms of Y. What is the probability of this event?

(b) What is $P(2 < Y \leq 4)$?

(c) What is $P(Y \neq 2)$?

7.12 CAR OWNERSHIP Choose an American household at random and let the random variable X be the number of cars (including SUVs and light trucks) they own. Here is the probability model if we ignore the few households that own more than 5 cars:

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

(a) Verify that this is a legitimate discrete distribution. Display the the distribution in a probability histogram.

(b) Say in words what the event $\{X \geq 1\}$ is. Find $P(X \geq 1)$.

(c) A housing company builds houses with two-car garages. What percent of households have more cars than the garage can hold?

7.13 ROLLING TWO DICE Some games of chance rely on tossing two dice. Each die has six faces, marked with 1, 2, . . ., 6 spots called pips. The dice used in casinos are carefully balanced so that each face is equally likely to come up. When two dice are tossed, each of the 36 possible pairs of faces is equally likely to come up. The outcome of interest to a gambler is the sum of the pips on the two up-faces. Call this random variable X.

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

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

(c) Define the random variable X. Then write the value of X next to each pair of faces and use this information with the result of (b) to give the probability distribution of X. Draw a probability histogram to display the distribution.

(d) One bet available in craps wins if a 7 or 11 comes up on the next roll of two dice. What is the probability of rolling a 7 or 11 on the next roll? Compare your answer with your experimental results (relative frequency) in Activity 7, step 4.

(e) After the dice are rolled the first time, several bets lose if a 7 is then rolled. If any outcome other than a 7 occurs, these bets either win or continue to the next roll. What is the probability that anything other than a 7 is rolled?

7.14 WEIRD DICE Nonstandard dice can produce interesting distributions of outcomes. You have two balanced, six-sided dice. One is a standard die, with faces having 1, 2, 3, 4, 5, and 6 spots. The other die has three faces with 0 spots and three faces with 6 spots. Find the probability distribution for the total number of spots Y on the up-faces when you roll these two dice.

7.15 EDUCATION LEVELS A study of education followed a large group of fifth-grade children to see how many years of school they eventually completed. Let X be the highest year of school that a randomly chosen fifth grader completes. (Students who go on to college are included in the outcome X = 12.) The study found this probability distribution for X:

Years:	4	5	6	7	8	9	10	11	12
Probability:	0.010	0.007	0.007	0.013	0.032	0.068	0.070	0.041	0.752

(a) What percent of fifth graders eventually finished twelfth grade?

(b) Check that this is a legitimate discrete probability distribution.

(c) Find $P(X \geq 6)$.

(d) Find $P(X > 6)$.

(e) What values of X make up the event "the student completed at least one year of high school"? (High school begins with the ninth grade.) What is the probability of this event?

7.16 **HOW STUDENT FEES ARE USED** Weary of the low turnout in student elections, a college administration decides to choose an SRS of three students to form an advisory board that represents student opinion. Suppose that 40% of all students oppose the use of student fees to fund student interest groups and that the opinions of the three students on the board are independent. Then the probability is 0.4 that each opposes the funding of interest groups.

(a) Call the three students A, B, and C. What is the probability that A and B support funding and C opposes it?

(b) List all possible combinations of opinions that can be held by students A, B, and C. (*Hint:* There are eight possibilities.) Then give the probability of each of these outcomes. Note that they are not equally likely.

(c) Let the random variable X be the number of student representatives who oppose the funding of interest groups. Give the probability distribution of X.

(d) Express the event "a majority of the advisory board opposes funding" in terms of X and find its probability.

7.17 **A UNIFORM DISTRIBUTION** Many random number generators allow users to specify the range of the random numbers to be produced. Suppose that you specify that the range is to be $0 \leq Y \leq 2$. Then the density curve of the outcomes has constant height between 0 and 2, and height 0 elsewhere.

(a) What is the height of the density curve between 0 and 2? Draw a graph of the density curve.

(b) Use your graph from (a) and the fact that probability is area under the curve to find $P(Y \leq 1)$.

(c) Find $P(0.5 < Y < 1.3)$.

(d) Find $P(Y \geq 0.8)$.

7.18 **THE SUM OF TWO RANDOM DECIMALS** Generate *two* random numbers between 0 and 1 and take Y to be their sum. Then Y is a continuous random variable that can take any value between 0 and 2. The density curve of Y is the triangle shown in Figure 7.8.

(a) Verify that the area under this curve is 1.

(b) What is the probability that Y is less than 1? (Sketch the density curve, shade the area that represents the probability, then find that area. Do this for (c) also.)

(c) What is the probability that Y is less than 0.5?

(d) Use simulation methods to answer the questions in (b) and (c). Here's one way using the TI-83/89. Clear L_1/list1, L_2/list2, and L_3/list3 and enter these commands:

FIGURE 7.8 The density curve for the sum of two random numbers. This continuous random variable takes values between 0 and 2.

TI-83	TI-89	
`rand(200)→L1`	`tistat.rand83(200) →list1`	Generates 200 random numbers and stores them in L_1/list1
`rand(200)→L2`	`tistat.rand83(200) →list2`	Generates 200 random numbers and stores them in L_2/list2
`L1+L2→L3`	`list1+list2→list3`	Adds the first number in L_1/list1 and the first number in L_2/list2 and stores the sum in L_3/list3, and so forth, from $i = 1$ to $i = 200$
`SortA(L3)`	`SortA list3`	Sorts the sums in L_3/list3 in ascending order

Now simply scroll through L_3/list3 and count the number of sums that satisfy the conditions stated in (b) and (c), and determine the relative frequency.

7.19 **PICTURING A DISTRIBUTION** This is a continuation of the previous exercise. If you carried out the simulation in 7.18(d), you can picture the distribution as follows: Deselect any active functions in the Y = screen, and turn off all STAT PLOTs. Define Plot1 to be a histogram using list L_3/list3. On the TI-89, set the Hist. Bucket Width at 0.1. Set WINDOW dimensions as follows: $X[0, 2]_{0.1}$ and $Y[-6, 25]_5$. Then press GRAPH. Does the resulting histogram resemble the triangle in Figure 7.8? Can you imagine the triangle superimposed on top of the histogram? Of course, some bars will be too short and others will be too long, but this is due to chance variation. To overlay the triangle, define Y_1 to be:

- TI-83: $Y_1 = (25X)(X \geq 0 \text{ and } X \leq 1) + (-25X + 50)(X \geq 1 \text{ and } X \leq 2)$
- TI-89: when $(x \geq 0$ and $x \leq 2$, when $(x \leq 1, 25x, -25x + 50), 0)$

Then press GRAPH again. How well does this "curve" fit your histogram?

7.20 **JOGGERS, I** An opinion poll asks an SRS of 1500 adults, "Do you happen to jog?" Suppose that the population proportion who jog is $p = 0.15$. To estimate p, we use the proportion $\hat{p}$ in the sample who answer "Yes." The statistic $\hat{p}$ is a random variable that is approximately normally distributed with mean $\mu = 0.15$ and standard deviation $\sigma = 0.0092$. Find the following probabilities:

(a) $P(\hat{p} \geq 0.16)$

(b) $P(0.14 \leq \hat{p} \leq 0.16)$

7.21 JOGGERS, II Describe the details of a simulation you could carry out to approximate an answer to Exercise 7.20(a). Then carry out the simulation. About how many repetitions do you need to get a result close to your answer to Exercise 7.20(a)?

7.2 MEANS AND VARIANCES OF RANDOM VARIABLES

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

The mean of a random variable

The mean $\bar{x}$ of a set of observations is their ordinary average. The mean of a random variable X is also an average of the possible values of X, but with an essential change to take into account the fact that not all outcomes need be equally likely. An example will show what we must do.

EXAMPLE 7.5 THE TRI-STATE PICK 3

Most states and Canadian provinces have government-sponsored lotteries. Here is a simple lottery wager, from the Tri-State Pick 3 game that New Hampshire shares with Maine and Vermont. You choose a three-digit number; the state chooses a three-digit winning number at random and pays you \$500 if your number is chosen. Because there are 1000 three-digit numbers, you have probability 1/1000 of winning. Taking X to be the amount your ticket pays you, the probability distribution of X is

Payoff X:	\$0	\$500
Probability:	0.999	0.001

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

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

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

If you play Tri-State Pick 3 several times, we would as usual call the mean of the actual amounts you win $\bar{x}$. The mean in Example 7.5 is a different quantity—it is the long-run average winnings you expect if you play a very large number of times. Just as probabilities are an idealized description of long-run proportions, the mean of a probability distribution describes the long-run average outcome. We can't call this mean $\bar{x}$, so we need a different symbol. The common symbol for the mean of a probability distribution is μ, the Greek letter mu. We used μ in Chapter 2 for the mean of a normal distribution, so this is not a new notation. We will often be interested in several random variables, each having a different probability distribution with a different mean. To remind ourselves that we are talking about the mean of X we often write μ_X rather than simply μ. In Example 7.5, $\mu_X = \$0.50$. Notice that, as often happens, the mean is not a possible value of X. You will often find the mean of a random variable X called the ***expected value*** of X. This term can be misleading, for we don't necessarily expect one observation on X to be close to its expected value.

mean μ

expected value

The mean of any discrete random variable is found just as in Example 7.5. It is an average of the possible outcomes, but a weighted average in which each outcome is weighted by its probability. Because the probabilities add to 1, we have total weight 1 to distribute among the outcomes. An outcome that occurs half the time has probability one-half and so gets one-half the weight in calculating the mean. Here is the general definition.

MEAN OF A DISCRETE RANDOM VARIABLE

Suppose that X is a discrete random variable whose distribution is

Value of X:	x_1	x_2	x_3	...	x_k
Probability:	p_1	p_2	p_3	...	p_k

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

$$\begin{aligned}\mu_X &= x_1p_1 + x_2p_2 + \cdots + x_kp_k \\ &= \Sigma x_i p_i\end{aligned}$$

EXAMPLE 7.6 BENFORD'S LAW

If first digits in a set of data appear "at random," the nine possible digits 1 to 9 all have the same probability. The probability distribution of the first digit X is then

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

The mean of this distribution is

$$\mu_x = 1\times\frac{1}{9} + 2\times\frac{1}{9} + 3\times\frac{1}{9} + 4\times\frac{1}{9} + 5\times\frac{1}{9} + 6\times\frac{1}{9} + 7\times\frac{1}{9} + 8\times\frac{1}{9} + 9\times\frac{1}{9}$$
$$= 45\times\frac{1}{9} = 5$$

If, on the other hand, the data obey Benford's law, the distribution of the first digit V is

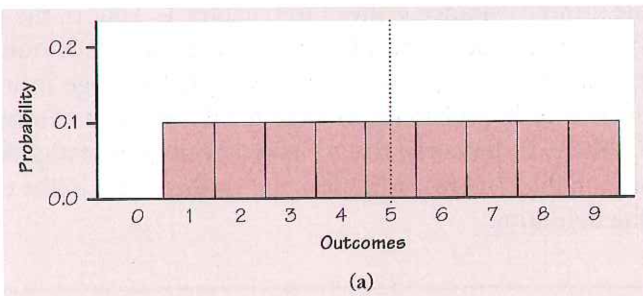
First digit V:	1	2	3	4	5	6	7	8	9
Probability:	0.301	0.176	0.125	0.097	0.079	0.067	0.058	0.051	0.046

The mean of V is

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

The means reflect the greater probability of smaller first digits under Benford's law.

Figure 7.9 locates the means of X and V on the two probability histograms. Because the discrete uniform distribution of Figure 7.9(a) is symmetric, the mean lies at the center of symmetry. We can't locate the mean of the right-skewed distribution of Figure 7.9(b) by eye—calculation is needed.



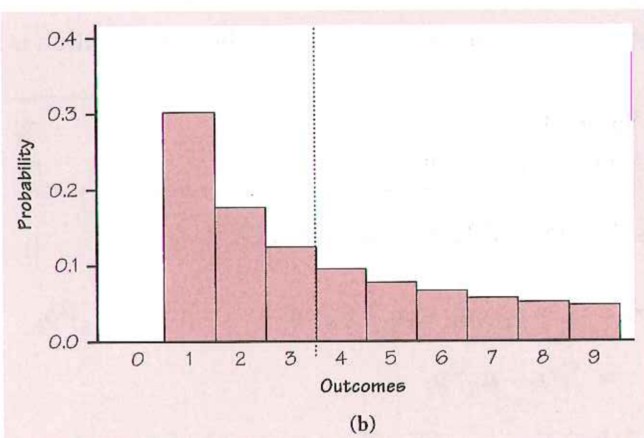


FIGURE 7.9 Locating the mean of a discrete random variable on the probability histogram for (a) digits between 1 and 9 chosen at random; (b) digits between 1 and 9 chosen from records that obey Benford's law.

What about continuous random variables? The probability distribution of a continuous random variable X is described by a density curve. Chapter 2 showed how to find the mean of the distribution: it is the point at which the area under the density curve would balance if it were made out of solid material. The mean lies at the center of symmetric density curves such as the normal curves. Exact calculation of the mean of a distribution with a skewed density curve requires advanced mathematics.[4]

The idea that the mean is the balance point of the distribution applies to discrete random variables as well, but in the discrete case we have a formula that gives us this point.

The variance of a random variable

The mean is a measure of the center of a distribution. Even the most basic numerical description requires in addition a measure of the spread or variability of the distribution. The variance and the standard deviation are the measures of spread that accompany the choice of the mean to measure center. Just as for the mean, we need a distinct symbol to distinguish the variance of a random variable from the variance s^2 of a data set. We write the variance of a random variable X as σ_X^2. Once again the subscript reminds us which variable we have in mind. The definition of the variance σ_X^2 of a random variable is similar to the definition of the sample variance s^2 given in Chapter 1. That is, the variance is an average of the squared deviation $(X - \mu_X)^2$ of the variable X from its mean μ_X. As for the mean, the average we use is a weighted average in which each outcome is weighted by its probability in order to take account of outcomes that are not equally likely. Calculating this weighted average is straightforward for discrete random variables but requires advanced mathematics in the continuous case. Here is the definition.

VARIANCE OF A DISCRETE RANDOM VARIABLE

Suppose that X is a discrete random variable whose distribution is

Value of X:	x_1	x_2	x_3	. . .	x_k
Probability:	p_1	p_2	p_3	. . .	p_k

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

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

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

EXAMPLE 7.7 SELLING AIRCRAFT PARTS

Gain Communications sells aircraft communications units to both the military and the civilian markets. Next year's sales depend on market conditions that cannot be predicted exactly. Gain follows the modern practice of using probability estimates of sales. The military division estimates its sales as follows:

Units sold:	1000	3000	5000	10,000
Probability:	0.1	0.3	0.4	0.2

These are personal probabilities that express the informed opinion of Gain's executives. Take X to be the number of military units sold. From the probability distribution we compute that

$$\begin{aligned}\mu_X &= (1000)(0.1) + (3000)(0.3) + (5000)(0.4) + (10{,}000)(0.2)\\ &= 100 + 900 + 2000 + 2000\\ &= 5000 \text{ units}\end{aligned}$$

The variance of X is calculated as

$$\begin{aligned}\sigma_X^2 = \Sigma(x_i - \mu_X)^2 p_i &= (1000-5000)^2(0.1) + (3000-5000)^2(0.3) + (5000-5000)^2(0.4)\\ &\quad + (10{,}000-5000)^2(0.2)\\ &= 1{,}600{,}000 + 1{,}200{,}000 + 0 + 5{,}000{,}000\\ &= 7{,}800{,}000\end{aligned}$$

The calculations can be arranged in the form of a table. Both μ_X and σ_X^2 are sums of columns in this table.

x_i	p_i	$x_i p_i$	$(x_i - \mu_X)^2 p_i$
1,000	0.1	100	$(1{,}000 - 5{,}000)^2\ (0.1) = 1{,}600{,}000$
3,000	0.3	900	$(3{,}000 - 5{,}000)^2\ (0.3) = 1{,}200{,}000$
5,000	0.4	2,000	$(5{,}000 - 5{,}000)^2\ (0.4) = 0$
10,000	0.2	2,000	$(10{,}000 - 5{,}000)^2\ (0.2) = 5{,}000{,}000$
		$\mu_X = 5{,}000$	$\sigma_X^2 = 7{,}800{,}000$

We see that $\sigma_X^2 = 7{,}800{,}000$. The standard deviation of X is $\sigma_X = \sqrt{7{,}800{,}000} = 2792.8$. The standard deviation is a measure of how variable the number of units sold is. As in the case of distributions for data, the standard deviation of a probability distribution is easiest to understand for normal distributions.

EXERCISES

7.22 A GRADE DISTRIBUTION Example 7.1 gives the distribution of grades (A = 4, B = 3, and so on) in a large class as

Grade:	0	1	2	3	4
Probability:	0.10	0.15	0.30	0.30	0.15

Find the average (that is, the mean) grade in this course.

7.23 OWNED AND RENTED HOUSING, I How do rented housing units differ from units occupied by their owners? Exercise 7.4 (page 396) gives the distributions of the number of rooms for owner-occupied units and renter-occupied units in San Jose, California. Find the mean number of rooms for both types of housing unit. How do the means reflect the differences between the distributions that you found in Exercise 7.4?

7.24 PICK 3 The Tri-State Pick 3 lottery game offers a choice of several bets. You choose a three-digit number. The lottery commission announces the winning three-digit number, chosen at random, at the end of each day. The "box" pays $83.33 if the number you choose has the same digits as the winning number, in any order. Find the expected payoff for a $1 bet on the box. (Assume that you chose a number having three different digits.)

7.25 KENO Keno is a favorite game in casinos, and similar games are popular with the states that operate lotteries. Balls numbered 1 to 80 are tumbled in a machine as the bets are placed, then 20 of the balls are chosen at random. Players select numbers by marking a card. The simplest of the many wagers available is "Mark 1 Number." Your payoff is $3 on a $1 bet if the number you select is one of those chosen. Because 20 of 80 numbers are chosen, your probability of winning is 20/80, or 0.25.

(a) What is the probability distribution (the outcomes and their probabilities) of the payoff X on a single play?

(b) What is the mean payoff μ_X?

(c) In the long run, how much does the casino keep from each dollar bet?

7.26 GRADE DISTRIBUTION, II Find the standard deviation σ_X of the distribution of grades in Exercise 7.22.

7.27 HOUSEHOLDS AND FAMILIES Exercise 7.10 (page 403) gives the distributions of the number of people in households and in families in the United States. An important difference is that many households consist of one person living alone, whereas a family must have at least two members. Some households may contain families along with other people, and so will be larger than the family. These differences make it hard to compare the distributions without calculations. Find the mean and standard deviation of both household size and family size. Combine these with your descriptions from Exercise 7.10 to give a comparison of the two distributions.

7.28 OWNED AND RENTED HOUSING, II Which of the two distributions for room counts in Exercises 7.4 (page 396) and 7.23 appears more spread out in the probability histograms? Why? Find the standard deviation for both distributions. The standard deviation provides a numerical measure of spread.

7.29 KIDS AND TOYS In an experiment on the behavior of young children, each subject is placed in an area with five toys. The response of interest is the number of toys that the child plays with. Past experiments with many subjects have shown that the probability distribution of the number X of toys played with is as follows:

Number of toys x_i:	0	1	2	3	4	5
Probability p_i:	0.03	0.16	0.30	0.23	0.17	0.11

(a) Calculate the mean μ_X and the standard deviation σ_X.

(b) Describe the details of a simulation you could carry out to approximate the mean number of toys μ_X and the standard deviation σ_X. Then carry out your simulation. Are the mean and standard deviation produced from your simulation close to the values you calculated in (a)?

Statistical estimation and the law of large numbers

We would like to estimate the mean height μ of the population of all American women between the ages of 18 and 24 years. This μ is the mean μ_X of the random variable X obtained by choosing a young woman at random and measuring her height. To estimate μ, we choose an SRS of young women and use the sample mean $\bar{x}$ to estimate the unknown population mean μ. Statistics obtained from probability samples are random variables because their values would vary in repeated sampling. The sampling distributions of statistics are just the probability distributions of these random variables. We will study sampling distributions in Chapter 9.

It seems reasonable to use $\bar{x}$ to estimate μ. An SRS should fairly represent the population, so the mean $\bar{x}$ of the sample should be somewhere near the mean μ of the population. Of course, we don't expect $\bar{x}$ to be exactly equal to μ, and we realize that if we choose another SRS, the luck of the draw will probably produce a different $\bar{x}$.

If $\bar{x}$ is rarely exactly right and varies from sample to sample, why is it nonetheless a reasonable estimate of the population mean μ? If we keep on adding observations to our random sample, the statistic $\bar{x}$ is *guaranteed* to get as close as we wish to the parameter μ and then stay that close. We have the comfort of knowing that if we can afford to keep on measuring more young women, eventually we will estimate the mean height of all young women very accurately. This remarkable fact is called the *law of large numbers*. It is remarkable because it holds for *any* population, not just for some special class such as normal distributions.

LAW OF LARGE NUMBERS

Draw independent observations at random from any population with finite mean μ. Decide how accurately you would like to estimate μ. As the number of observations drawn increases, the mean $\bar{x}$ of the observed values eventually approaches the mean μ of the population as closely as you specified and then stays that close.

The behavior of $\bar{x}$ is similar to the idea of probability. In the long run, the proportion of outcomes taking any value gets close to the probability of that value, and the average outcome gets close to the distribution mean. Figure 6.1 (page 331) shows how proportions approach probability in one example. Here is an example of how sample means approach the distribution mean.

EXAMPLE 7.8 HEIGHTS OF YOUNG WOMEN

The distribution of the heights of all young women is close to the normal distribution with mean 64.5 inches and standard deviation 2.5 inches. Suppose that $\mu = 64.5$ were exactly true. Figure 7.10 shows the behavior of the mean height $\bar{x}$ of n women chosen at random from a population whose heights follow the $N(64.5, 2.5)$ distribution. The graph plots the values of $\bar{x}$ as we add women to our sample. The first woman drawn had height 64.21 inches, so the line starts there. The second had height 64.35 inches, so for $n = 2$ the mean is

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

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

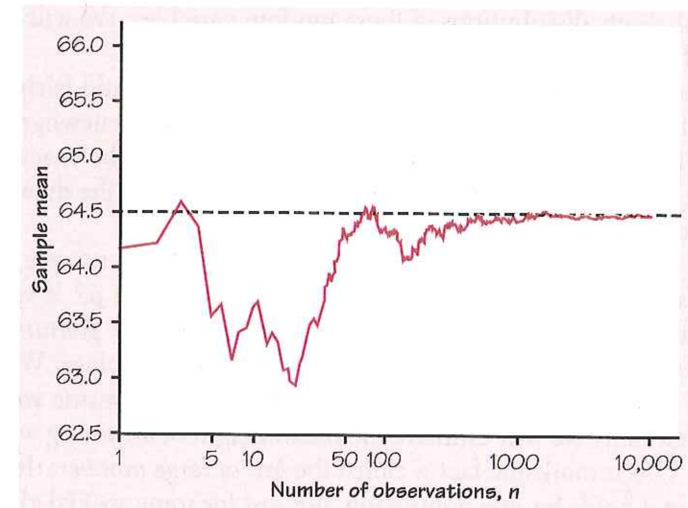


FIGURE 7.10 The law of large numbers in action. As we increase the size of our sample, the sample mean $\bar{x}$ always approaches the mean μ of the population.

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

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

Thinking about the law of large numbers

The law of large numbers says broadly that the average results of many independent observations are stable and predictable. Casinos are not the only businesses that base forecasts on this fact. An insurance company deciding how much to charge for life insurance and a fast-food restaurant deciding how many beef patties to prepare rely on the fact that averaging over many individuals produces a stable result. It is worth the effort to think a bit more closely about so important a fact.

The "law of small numbers"

Both the rules of probability and the law of large numbers describe the regular behavior of chance phenomena *in the long run.* Psychologists have discovered that the popular understanding of randomness is quite different from the true laws of chance.[5] Most people believe in an incorrect "law of small numbers." That is, we expect even short sequences of random events to show the kind of average behavior that in fact appears only in the long run.

Try this experiment: Write down a sequence of heads and tails that you think imitates 10 tosses of a balanced coin. How long was the longest string (called a *run*) of consecutive heads or consecutive tails in your tosses? Most people will write a sequence with no runs of more than two consecutive heads or tails. Longer runs don't seem "random" to us. In fact, the probability of a run of three or more consecutive heads or tails in 10 tosses is greater than 0.8, and the probability of *both* a run of three or more heads and a run of three or more tails is almost 0.2.[6] This and other probability calculations suggest that a short sequence of coin tosses will often not appear random to us. The runs of consecutive heads or consecutive tails that appear in real coin tossing (and that are predicted by the mathematics of probability) seem surprising to us. Because we don't expect to see long runs, we may conclude that the coin tosses are not independent or that some influence is disturbing the random behavior of the coin.

EXAMPLE 7.9 THE "HOT HAND" IN BASKETBALL

Belief in the law of small numbers influences behavior. If a basketball player makes several consecutive shots, both the fans and her teammates believe that she has the "hot hand" and is more likely to make the next shot. This is doubtful. Careful study suggests that runs of baskets made or missed are no more frequent in basketball than it would be expected if each shot were independent of the player's previous shots. Baskets made or missed are just like heads and tails in tossing a coin. (Of course, some players make 30% of their shots in the long run and others make 50%, so a coin-toss model for basketball must allow coins with different probabilities of a head.) Our perception of hot or cold streaks simply shows that we don't perceive random behavior very well.[7]

Gamblers often follow the hot-hand theory, betting that a run will continue. At other times, however, they draw the opposite conclusion when confronted with a run of outcomes. If a coin gives 10 straight heads, some gamblers feel that it must now produce some extra tails to get back to the average of half heads and half tails. Not so. If the next 10,000 tosses give about 50% tails, those 10 straight heads will be swamped by the later thousands of heads and tails. No compensation is needed to get back to the average in the long run. Remember that it is *only* in the long run that the regularity described by probability and the law of large numbers takes over.

Our inability to accurately distinguish random behavior from systematic influences points out once more the need for statistical inference to supplement exploratory analysis of data. Probability calculations can help verify that what we see in the data is more than a random pattern.

How large is a large number?

The law of large numbers says that the actual mean outcome of many trials gets close to the distribution mean μ as more trials are made. It doesn't say how many trials are needed to guarantee a mean outcome close to μ. That depends on the *variability* of the random outcomes. The more variable the outcomes, the more trials are needed to ensure that the mean outcome $\bar{x}$ is close to the distribution mean μ.

The law of large numbers is the foundation of such business enterprises as gambling casinos and insurance companies. Games of chance must be quite variable if they are to hold the interest of gamblers. Even a long evening in a casino has an unpredictable outcome. Gambles with extremely variable outcomes, like state lottos with their very large but very improbable jackpots, require impossibly large numbers of trials to ensure that the average outcome is close to the expected value. Though most forms of gambling are less variable than lotto, the layman's answer to the applicability of the law of large numbers is usually that the house plays often enough to rely on it, but you don't. Much of the psychological allure of gambling is its unpredictability for the player. The business of gambling rests on the fact that the result is not unpredictable for the house. The average winnings of the house on tens of thousands of bets will be very close to the mean of the distribution of winnings. Needless to say, this mean guarantees the house a profit.

EXERCISES

7.30 LAW OF LARGE NUMBERS SIMULATION This exercise is based on Example 7.8 and uses the TI-83/89 to simulate the law of large numbers and the sampling process. Begin by clearing L_1/list1, L_2/list2, L_3/list3, and L_4/list4. Then enter the commands from the table on the following page.

Specify Plot1 as follows: xyLine (2nd Type icon on the TI-83); Xlist: L_1/list1; Ylist: L_4/list4; Mark: . Set the viewing WINDOW as follows: $X[1,10]_{10}$. To set the Y dimensions, scan the values in L_4/list4. Or start with $Y[60,69]_1$ and adjust as necessary. Press

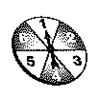
TI-83	TI-89	
seq(X,X,1,200)→L_1	seq(X,X,1,200) →list1	Enters the positive integers 1 to 200 into L_1/list1 (for seq, look under 2nd / LIST/ OPS on the TI-83 and under CATALOG on the TI-89).
randNorm(64.5,2.5, 200)→L_2	tistat.randNorm (64.5,2.5,200) →list2	Generates 200 random heights (in inches) from the $N(64.5, 2.5)$ distribution and stores these values in L_2/list2 (for randNorm, look under MATH / PRB on the TI-83 and under CATALOG on the TI-89) .
cumSum(L_2)→L_3	cumSum(list2) →list3	Provides a cumulative sum of the observations and stores these values in L_3/list3 (for cumSum, look under 2nd/ LIST/OPS on the TI-83 and under CATALOG on the TI-89).
L_3/L_1→L_4	list3/list1 →list4	Calculates the average heights of the women and stores these values in L_4/list4

[GRAPH]. In the WINDOW screen, change Xmax to 100, and press [GRAPH] again. In your own words, write a short description of the principle that this exercise demonstrates.

7.31 **A GAME OF CHANCE** One consequence of the law of large numbers is that once we have a probability distribution for a random variable, we can find its mean by simulating many outcomes and averaging them. The law of large numbers says that if we take enough outcomes, their average value is sure to approach the mean of the distribution.

I have a little bet to offer you. Toss a coin ten times. If there is no run of three or more straight heads or tails in the ten outcomes, I'll pay you $2. If there is a run of three or more, you pay me just $1. Surely you will want to take advantage of me and play this game?

Simulate enough plays of this game (the outcomes are +$2 if you win and –$1 if you lose) to estimate the mean outcome. Is it to your advantage to play?



7.32

(a) A gambler knows that red and black are equally likely to occur on each spin of a roulette wheel. He observes five consecutive reds and bets heavily on red at the next spin. Asked why, he says that "red is hot" and that the run of reds is likely to continue. Explain to the gambler what is wrong with this reasoning.

(b) After hearing you explain why red and black remain equally probable after five reds on the roulette wheel, the gambler moves to a poker game. He is dealt five straight red cards. He remembers what you said and assumes that the next card dealt in the same hand is equally likely to be red or black. Is the gambler right or wrong? Why?

7.33 OVERDUE FOR A HIT Retired baseball player Tony Gwynn got a hit about 35% of the time over an entire season. After he failed to hit safely in six straight at-bats, a TV commentator said, "Tony is due for a hit by the law of averages." Is that right? Why?

Rules for means

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

In more formal language, the number of dimples on a refrigerator is a random variable X that takes values 0, 1, 2, and so on. X varies as we inspect one refrigerator after another. Only the mean number of dimples $\mu_X = 0.7$ was reported to you. The number of paint sags is a second random variable Y having mean $\mu_Y = 1.4$. (You see how the subscripts keep straight which variable we are talking about.) The total number of both dimples and sags is the sum $X + Y$. That sum is another random variable that varies from refrigerator to refrigerator. Its mean μ_{X+Y} is the average number of dimples and sags together and is just the sum of the individual means μ_X and μ_Y. That is an important rule for how means of random variables behave.

Here's another rule. The crickets living in a field have mean length 1.2 inches. What is the mean in centimeters? There are 2.54 centimeters in an inch, so the length of a cricket in centimeters is 2.54 times its length in inches. If we multiply every observation by 2.54, we also multiply their average by 2.54. The mean in centimeters must be 2.54×1.2, or about 3.05 centimeters. More formally, the length in inches of a cricket chosen at random from the field is a random variable X with mean μ_X. The length in centimeters is $2.54X$, and this new random variable has mean $2.54\mu_X$.

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

RULES FOR MEANS

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

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

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

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

Here is an example that applies these rules.

EXAMPLE 7.10 GAIN COMMUNICATIONS

In Example 7.7 (page 411) we saw that the number X of communications units sold by the Gain Communications *military* division has distribution

X = units sold:	1000	3000	5000	10,000
Probability:	0.1	0.3	0.4	0.2

The corresponding sales estimates for the *civilian* division are

Y = units sold:	300	500	750
Probability:	0.4	0.5	0.1

In Example 7.7, we calculated $\mu_X = 5000$. In similar fashion, we calculate

$$\begin{aligned}\mu_Y &= (300)(0.4) + (500)(0.5) + (750)(0.1)\\ &= 445 \text{ units}\end{aligned}$$

Gain makes a profit of \$2000 on each military unit sold and \$3500 on each civilian unit. Next year's profit from military sales will be 2000X, \$2000 times the number X of units sold. By Rule 1, the mean military profit is

$$\mu_{2000X} = 2000\mu_X = (2000)(5000) = \$10{,}000{,}000$$

Similarly, the civilian profit is 3500Y and the mean profit from civilian sales is

$$\mu_{3500Y} = 3500\mu_Y = (3500)(445) = \$1{,}557{,}500$$

The total profit is the sum of the military and civilian profit:

$$Z = 2000X + 3500Y$$

Rule 2 says that the mean of this sum of two variables is the sum of the two individual means:

$$\begin{aligned}\mu_Z &= \mu_{2000X} + \mu_{3500Y}\\ &= 10{,}000{,}000 + 1{,}557{,}500\\ &= \$11{,}557{,}500\end{aligned}$$

This mean is the company's best estimate of next year's profit, combining the probability estimates of the two divisions. We can do this calculation more quickly by combining Rules 1 and 2:

$$\begin{aligned}\mu_Z &= \mu_{2000X+3500Y}\\ &= 2000\mu_X + 3500\mu_Y\\ &= (2000)(5000) + (3500)(445) = \$11{,}557{,}500\end{aligned}$$

Rules for variances

What are the facts for variances that parallel Rules 1 and 2 for means? The mean of a sum of random variables is always the sum of their means, but this addition rule is not always true for variances. To understand why, take X to be the percent of a family's after-tax income that is spent and Y the percent that is saved. When X increases, Y decreases by the same amount. Though X and Y may vary widely from year to year, their sum X + Y is always 100% and does not vary at all. It is the association between the variables X and Y that prevents their variances from adding. If random variables are independent, this kind of association between their values is ruled out and their variances do add. Two random variables X and Y are ***independent*** if knowing that any event involving X alone did or did not occur tells us nothing about the occurrence of any event involving Y alone. Probability models often assume independence when the random variables describe outcomes that appear unrelated to each other. You should ask in each instance whether the assumption of independence seems reasonable.

independent

When random variables are not independent, the variance of their sum depends on the ***correlation*** between them as well as on their individual variances. In Chapter 3, we met the correlation r between two observed variables measured on the same individuals. We defined (page 140) the correlation r as an average of the products of the standardized x and y observations. The correlation between two random variables is defined in the same way, once again using a weighted average with probabilities as weights. We won't give the details—it is enough to know that the correlation between two random variables has the same basic properties as the correlation r calculated from data. We use ρ, the Greek letter rho, for the correlation between two random variables. The correlation ρ is a number between −1 and 1 that measures the direction and strength of the linear relationship between two variables. **The correlation between two independent random variables is zero.**

correlation

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

RULES FOR VARIANCES

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

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

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

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

RULES FOR VARIANCES (*continued*)

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

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

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

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

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

As with data, we prefer the standard deviation to the variance as a measure of variability. The addition rule for variances implies that standard deviations do *not* generally add. Standard deviations are most easily combined by using the rules for variances rather than by giving separate rules for standard deviations. For example, the standard deviations of 2X and –2X are both equal to $2\sigma_X$ because this is the square root of the variance $4\sigma^2_X$.

EXAMPLE 7.11 WINNING THE LOTTERY

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

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

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

If you buy a Pick 3 ticket, your winnings are $W = X - 1$ because the dollar you paid for the ticket must be subtracted from the payoff. By the rules for means, the mean amount you win is

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

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

Suppose now that you buy a \$1 ticket on each of two different days. The payoffs X and Y on the two tickets are independent because separate drawings are held each day. Your total payoff $X + Y$ has mean

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

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

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

The standard deviation of the total payoff is

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

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

If you buy a ticket every day (365 tickets in a year), your mean payoff is the sum of 365 daily payoffs. That's 365 times 50 cents, or \$182.50. Of course, it costs \$365 to play, so the state's mean take from a daily Pick 3 player is \$182.50. Results for individual players will vary, but the law of large numbers assures the state its profit.

EXAMPLE 7.12 SAT SCORES

A college uses SAT scores as one criterion for admission. Experience has shown that the distribution of SAT scores among its entire population of applicants is such that

SAT Math score X	$\mu_X = 625$	$\sigma_X = 90$
SAT Verbal score Y	$\mu_Y = 590$	$\sigma_Y = 100$

What are the mean and standard deviation of the total score $X + Y$ among students applying to this college?

The mean overall SAT score is

$$\mu_{X+Y} = \mu_X + \mu_Y = 625 + 590 = 1215$$

The variance and standard deviation of the total *cannot be computed* from the information given. SAT verbal and math scores are not independent, because students who score high on one exam tend to score high on the other also. Therefore, Rule 2 does not apply and we need to know ρ, the correlation between X and Y, to apply Rule 3.

Nationally, the correlation between SAT Math and Verbal scores is about $\rho = 0.7$. If this is true for these students,

$$\begin{aligned}\sigma^2_{X+Y} &= \sigma^2_X + \sigma^2_Y + 2\rho\sigma_X\sigma_Y \\ &= (90)^2 + (100)^2 + (2)(0.7)(90)(100) \\ &= 30{,}700\end{aligned}$$

The variance of the sum $X + Y$ is greater than the sum of the variances $\sigma_X^2 + \sigma_Y^2$ because of the positive correlation between SAT Math scores and SAT Verbal scores. That is, X and Y tend to move up together and down together, which increases the variability of their sum. We find the standard deviation from the variance,

$$\sigma_{X+Y} = \sqrt{30{,}700} = 175$$

EXAMPLE 7.13 INVESTING IN STOCKS AND T-BILLS

Zadie has invested 20% of her funds in Treasury bills and 80% in an "index fund" that represents all U.S. common stocks. The rate of return of an investment over a time period is the percent change in the price during the time period, plus any income received. If X is the annual return on T-bills and Y the annual return on stocks, the portfolio rate of return is

$$R = 0.2X + 0.8Y$$

The returns X and Y are random variables because they vary from year to year. Based on annual returns between 1950 and 2000, we have[8]

X = annual return on T-bills	$\mu_X = 5.2\%$	$\sigma_X = 2.9\%$
Y = annual return on stocks	$\mu_Y = 13.3\%$	$\sigma_Y = 17.0\%$
Correlation between X and Y	$\rho = -0.1$	

Stocks had higher returns than T-bills on the average, but the standard deviations show that returns on stocks varied much more from year to year. That is, the risk of investing in stocks is greater than the risk for T-bills because their returns are less predictable.

For the return *R* on Zadie's portfolio of 20% T-bills and 80% stocks,

$$\begin{aligned} R &= 0.2X + 0.8Y \\ \mu_R &= 0.2\mu_X + 0.8\mu_Y \\ &= (0.2 \times 5.2) + (0.8 \times 13.3) = 11.68\% \end{aligned}$$

To find the variance of the portfolio return, combine Rules 1 and 3:

$$\begin{aligned} \sigma_R^2 &= \sigma_{0.2X}^2 + \sigma_{0.8Y}^2 + 2\rho\sigma_{0.2X}\sigma_{0.8Y} \\ &= (0.2)^2\sigma_X^2 + 0.8^2\sigma_Y^2 + 2\rho(0.2\sigma_X)(0.8\sigma_Y) \\ &= (0.2)^2(2.9)^2 + (0.8)^2(17.0)^2 + (2)(-0.1)(0.2 \times 2.9)(0.8 \times 17.0) \\ &= 183.719 \\ \sigma_R &= \sqrt{183.719} = 13.55\% \end{aligned}$$

The portfolio has a smaller mean return than an all-stock portfolio, but it is also less risky. As a proportion of the all-stock values, the reduction in standard deviation is greater than the reduction in mean return. That's why Zadie put some funds into Treasury bills.

Combining normal random variables

So far, we have concentrated on finding rules for means and variances of random variables. If a random variable is normally distributed, we can use its mean and variance to compute probabilities. Example 7.4 (page 400) shows the method. What if we combine two normal random variables?

Any linear combination of independent normal random variables is also normally distributed. That is, if X and Y are independent normal random variables and a and b are any fixed numbers, $aX + bY$ is also normally distributed. In particular, the sum or difference of independent normal random variables has a normal distribution. The mean and standard deviation of $aX + bY$ are found as usual from the addition rules for means and variances. These facts are often used in statistical calculations.

EXAMPLE 7.14 A ROUND OF GOLF

Tom and George are playing in the club golf tournament. Their scores vary as they play the course repeatedly. Tom's score X has the $N(110,10)$ distribution, and George's score Y varies from round to round according to the $N(100,8)$ distribution. If they play independently, what is the probability that Tom will score lower than George and thus do better in the tournament? The difference $X - Y$ between their scores is normally distributed, with mean and variance

$$\mu_{X-Y} = \mu_X - \mu_Y = 110 - 100 = 10$$
$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y = 10^2 + 8^2 = 164$$

Because $\sqrt{164} = 12.8$, $X - Y$ has the $N(10,12.8)$ distribution. Figure 7.11 illustrates the probability computation:

$$P(X < Y) = P(X - Y < 0)$$
$$= P\left(\frac{(X-Y)-10}{12.8} < \frac{0-10}{12.8}\right)$$
$$= P(Z < -0.78) = 0.2177$$

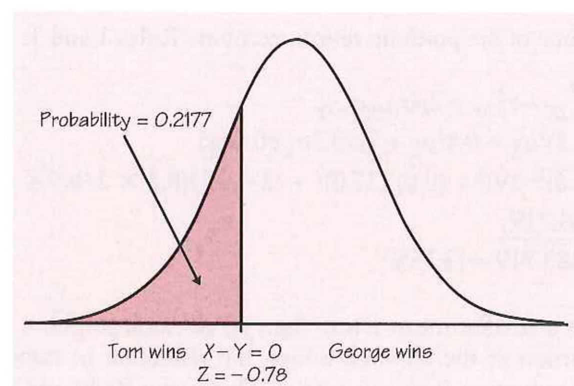


FIGURE 7.11 The normal probability calculation for Example 7.14.

Although George's score is 10 strokes lower on the average, Tom will have the lower score in about one of every five matches.

EXERCISES

7.34 CHECKING INDEPENDENCE, I For each of the following situations, would you expect the random variables X and Y to be independent? Explain your answers.

(a) X is the rainfall (in inches) on November 6 of this year, and Y is the rainfall at the same location on November 6 of next year.

(b) X is the amount of rainfall today, and Y is the rainfall at the same location tomorrow.

(c) X is today's rainfall at the airport in Orlando, Florida, and Y is today's rainfall at Disney World just outside Orlando.

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

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

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

7.36 CHEMICAL REACTIONS, I Laboratory data show that the time required to complete two chemical reactions in a production process varies. The first reaction has a mean time of 40 minutes and a standard deviation of 2 minutes; the second has a mean time of 25 minutes and a standard deviation of 1 minute. The two reactions are run in sequence during production. There is a fixed period of 5 minutes between them as the product of the first reaction is pumped into the vessel where the second reaction will take place. What is the mean time required for the entire process?

7.37 TIME AND MOTION, I A time and motion study measures the time required for an assembly-line worker to perform a repetitive task. The data show that the time required to bring a part from a bin to its position on an automobile chassis varies from car to car with mean 11 seconds and standard deviation 2 seconds. The time required to attach the part to the chassis varies with mean 20 seconds and standard deviation 4 seconds.

(a) What is the mean time required for the entire operation of positioning and attaching the part?

(b) If the variation in the worker's performance is reduced by better training, the standard deviations will decrease. Will this decrease change the mean you found in (a) if the mean times for the two steps remain as before?

(c) The study finds that the times required for the two steps are independent. A part that takes a long time to position, for example, does not take more or less time to attach than other parts. How would your answers to (a) and (b) change if the two variables were dependent with correlation 0.8? With correlation 0.3?

7.38 **TIME AND MOTION, II** Find the standard deviation of the time required for the two-step assembly operation studied in Exercise 7.37, assuming that the study shows the two times to be independent. Redo the calculation assuming that the two times are dependent, with correlation 0.3. Can you explain in nontechnical language why positive correlation increases the variability of the total time?

7.39 **CHEMICAL REACTIONS, II** The times for the two reactions in the chemical production process described in Exercise 7.36 are independent. Find the standard deviation of the time required to complete the process.

7.40 Examples 7.7 (page 411) and 7.10 (page 419) concern a probabilistic projection of sales and profits by an electronics firm, Gain Communications.

(a) Find the variance and standard deviation of the estimated sales Y of Gain's civilian unit, using the distribution and mean from Example 7.10.

(b) Because the military budget and the civilian economy are not closely linked, Gain is willing to assume that its military and civilian sales vary independently. Combine your result from (a) with the results for the military unit from Example 7.10 to obtain the standard deviation of the total sales $X + Y$.

(c) Find the standard deviation of the estimated profit, $Z = 2000X + 3500Y$.

7.41 Leona and Fred are friendly competitors in high school. Both are about to take the ACT college entrance examination. They agree that if one of them scores 5 or more points better than the other, the loser will buy the winner a pizza. Suppose that in fact Fred and Leona have equal ability, so that each score varies normally with mean 24 and standard deviation 2. (The variation is due to luck in guessing and the accident of the specific questions being familiar to the student.) The two scores are independent. What is the probability that the scores differ by 5 or more points in either direction?

SUMMARY

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

The **mean** μ is the balance point of the probability histogram or density curve. If X is discrete with possible values x_i having probabilities p_i, the mean is the average of the values of X, each weighted by its probability:

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

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

$$\sigma_X^2 = (x_1 - \mu)^2p_1 + (x_2 - \mu)^2p_2 + \cdots + (x_k - \mu)^2p_k$$

The **standard deviation** σ_X is the square root of the variance. The standard deviation measures the variability of the distribution about the mean. It is easiest to interpret for normal distributions.

The mean and variance of a continuous random variable can be computed from the density curve, but to do so requires more advanced mathematics.

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

The means and variances of random variables obey the following rules. If a and b are fixed numbers, then

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

If X and Y are any two random variables, then

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

and if X and Y are independent, then

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

Any linear combination of independent normal random variables is also normally distributed.

SECTION 7.2 EXERCISES

7.42 BUYING STOCK You purchase a hot stock for $1000. The stock either gains 30% or loses 25% each day, each with probability 0.5. Its returns on consecutive days are independent of each other. You plan to sell the stock after two days.

(a) What are the possible values of the stock after two days, and what is the probability for each value? What is the probability that the stock is worth more after two days than the $1000 you paid for it?

(b) What is the mean value of the stock after two days? You see that these two criteria give different answers to the question, "Should I invest?"

7.43 APPLYING BENFORD'S LAW It is easier to use Benford's law (Example 7.6, page 408) to spot suspicious patterns when you have very many items (for example, many invoices from the same vendor) than when you have only a few. Explain why this is true.

7.44 WEIRD DICE You have two balanced, six-sided dice. The first has 1, 3, 4, 5, 6, and 8 spots on its six faces. The second die has 1, 2, 2, 3, 3, and 4 spots on its faces.

(a) What is the mean number of spots on the up-face when you roll each of these dice?

(b) Write the probability model for the outcomes when you roll both dice independently. From this, find the probability distribution of the sum of the spots on the up-faces of the two dice.

(c) Find the mean number of spots on the two up-faces in two ways: from the distribution you found in (b) and by applying the addition rule to your results in (a). You should of course get the same answer.

7.45 SSHA The academic motivation and study habits of female students as a group are better than those of males. The Survey of Study Habits and Attitudes (SSHA) is a psychological test that measures these factors. The distribution of SSHA scores among the women at a college has mean 120 and standard deviation 28, and the distribution of scores among men students has mean 105 and standard deviation 35. You select a single male student and a single female student at random and give them the SSHA test.

(a) Explain why it is reasonable to assume that the scores of the two students are independent.

(b) What are the mean and standard deviation of the difference (female minus male) between their scores?

(c) From the information given, can you find the probability that the woman chosen scores higher than the man? If so, find this probability. If not, explain why you cannot.

7.46 A GLASS ACT, I In a process for manufacturing glassware, glass stems are sealed by heating them in a flame. The temperature of the flame varies a bit. Here is the distribution of the temperature X measured in degrees Celsius:

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

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

(b) The target temperature is 550° C. What are the mean and standard deviation of the number of degrees off target $X - 550$?

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

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

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

7.47 A GLASS ACT, II In continuation of the previous exercise, describe the details of a simulation you could carry out to approximate the mean temperature and the standard deviation in degrees Celsius. Then carry out your simulation. Are the mean and standard deviation produced from your simulation close to the values you calculated in 7.46 (a)?

7.48 A machine fastens plastic screw-on caps onto containers of motor oil. If the machine applies more torque than the cap can withstand, the cap will break. Both the torque applied and the strength of the caps vary. The capping-machine torque has the normal distribution with mean 7 inch-pounds and standard deviation 0.9 inch-pounds. The cap strength (the torque that would break the cap) has the normal distribution with mean 10 inch-pounds and standard deviation 1.2 inch-pounds.

(a) Explain why it is reasonable to assume that the cap strength and the torque applied by the machine are independent.

(b) What is the probability that a cap will break while being fastened by the capping machine?

7.49 A study of working couples measures the income X of the husband and the income Y of the wife in a large number of couples in which both partners are employed. Suppose that you knew the means μ_X and μ_Y and the variances σ_X^2 and σ_Y^2 of both variables in the population.

(a) Is it reasonable to take the mean of the total income X + Y to be $\mu_X + \mu_Y$? Explain your answer.

(b) Is it reasonable to take the variance of the total income to be $\sigma_X^2 + \sigma_Y^2$? Explain your answer.

7.50 The design of an electronic circuit calls for a 100-ohm resistor and a 250-ohm resistor connected in series so that their resistances add. The components used are not perfectly uniform, so that the actual resistances vary independently according to normal distributions. The resistance of 100-ohm resistors has mean 100 ohms and standard deviation 2.5 ohms, while that of 250-ohm resistors has mean 250 ohms and standard deviation 2.8 ohms.

(a) What is the distribution of the total resistance of the two components in series?

(b) What is the probability that the total resistance lies between 345 and 355 ohms?

Portfolio analysis. Here are the means, standard deviations, and correlations for the monthly returns from three Fidelity mutual funds for the 36 months ending in December 2000.[9] Because there are three random variables, there are three correlations. We use subscripts to show which pair of random variables a correlation refers to.

W = monthly return on Magellan Fund	$\mu_W = 1.14\%$	$\sigma_W = 4.64\%$
X = monthly return on Real Estate Fund	$\mu_X = 0.16\%$	$\sigma_X = 3.61\%$
Y = monthly return on Japan Fund	$\mu_Y = 1.59\%$	$\sigma_Y = 6.75\%$

Correlations

$\rho_{WX} = 0.19$ $\rho_{WY} = 0.54$ $\rho_{XY} = -0.17$

Exercises 7.51 to 7.53 make use of these historical data.

7.51 Many advisors recommend using roughly 20% foreign stocks to diversify portfolios of U.S. stocks. Michael owns Fidelity Magellan Fund, which concentrates on stocks of large American companies. He decides to move to a portfolio of 80% Magellan and 20% Fidelity Japan Fund. Show that (based on historical data) this portfolio has both a *higher* mean return and *less* volatility (variability) than Magellan alone. This illustrates the beneficial effects of diversifying among investments.

7.52 Diversification works better when the investments in a portfolio have small correlations. To demonstrate this, suppose that returns on Magellan Fund and Japan

Fund had the means and standard deviations we have given but were uncorrelated ($\rho_{WY} = 0$). Show that the standard deviation of a portfolio that combines 80% Magellan with 20% Japan is smaller than your result from the previous exercise. What happens to the mean return if the correlation is 0?

7.53 Portfolios often contain more than two investments. The rules for means and variances continue to apply, though the arithmetic gets messier. A portfolio containing proportions a of Magellan Fund, b of Real Estate Fund, and c of Japan Fund has return $R = aW + bX + cY$. Because a, b, and c are the proportions invested in the three funds, $a + b + c = 1$. The mean and variance of the portfolio return are

$$\mu_R = a\mu_W + b\mu_X + c\mu_Y$$
$$\sigma_R^2 = a^2\sigma_W^2 + b^2\sigma_X^2 + c^2\sigma_Y^2 + 2ab\rho_{WX}\sigma_W\sigma_X + 2ac\rho_{WY}\sigma_W\sigma_Y + 2bc\rho_{XY}\sigma_X\sigma_Y$$

Having seen the advantages of diversification, Michael decides to invest his funds 60% in Magellan, 20% in Real Estate, and 20% in Japan. What are the (historical) mean and standard deviation of the monthly returns for this portfolio?

CHAPTER REVIEW

A random variable defines what is counted or measured in a statistics application. If the random variable X is a count, such as the number of heads in four tosses of a coin, then X is discrete, and its distribution can be pictured as a histogram. If X is measured, as in the number of inches of rainfall in Richmond in April, then X is continuous, and its distribution is pictured as a density curve. Among the continuous random variables, the normal random variable is the most important. First introduced in Chapter 2, the normal distribution is revisited, with emphasis this time on it as a probability distribution. The mean and variance of a random variable are calculated, and rules for the sum or difference of two random variables are developed. Here is a checklist of the major skills you should have acquired by studying this chapter.

A. RANDOM VARIABLES

1. Recognize and define a discrete random variable, and construct a probability distribution table and a probability histogram for the random variable.

2. Recognize and define a continuous random variable, and determine probabilities of events as areas under density curves.

3. Given a normal random variable, use the standard normal table or a graphing calculator to find probabilities of events as areas under the standard normal distribution curve.

B. MEANS AND VARIANCES OF RANDOM VARIABLES

1. Calculate the mean and variance of a discrete random variable. Find the expected payout in a raffle or similar game of chance.

2. Use simulation methods and the law of large numbers to approximate the mean of a distribution.

3. Use rules for means and rules for variances to solve problems involving sums, differences, and linear combinations of random variables.

CHAPTER 7 REVIEW EXERCISES

7.54 TWO-FINGER MORRA Ann and Bob are playing the game Two-Finger Morra. Each player shows either one or two fingers and at the same time calls out a guess for the number of fingers the other player will show. If a player guesses correctly and the other player does not, the player wins a number of dollars equal to the total number of fingers shown by both players. If both or neither guesses correctly, no money changes hands. On each play both Ann and Bob choose one of the following options:

Choice	Show	Guess
A	1	1
B	1	2
C	2	1
D	2	2

(a) Give the sample space S by writing all possible choices for both players on a single play of this game.

(b) Let X be Ann's winnings on a play. (If Ann loses \$2, then X = –2; when no money changes hands, X = 0.) Write the value of the random variable X next to each of the outcomes you listed in (a). This is another choice of sample space.

(c) Now assume that Ann and Bob choose independently of each other. Moreover, they both play so that all four choices listed above are equally likely. Find the probability distribution of X.

(d) If the game is fair, X should have mean zero. Does it? What is the standard deviation of X?

Insurance. The business of selling insurance is based on probability and the law of large numbers. Consumers (including businesses) buy insurance because we all face risks that are unlikely but carry high cost. Think of a fire destroying your home. So we form a group to share the risk: we all pay a small amount, and the insurance policy pays a large amount to those few of us whose homes burn down. The insurance company sells many policies, so it can rely on the law of large numbers. Exercises 7.55 to 7.58 explore aspects of insurance.

7.55 LIFE INSURANCE, I A life insurance company sells a term insurance policy to a 21-year-old male that pays \$100,000 if the insured dies within the next 5 years. The probability that a randomly chosen male will die each year can be found in mortality tables. The company collects a premium of \$250 each year as payment for the insurance. The amount X that the company earns on this policy is $250 per year, less the $100,000 that it must pay if the insured dies. Here is the distribution of X. Fill in the missing probability in the table and calculate the mean profit μ_X.

Age at death:	21	22	23	24	25	≥ 26
Profit:	–$99,750	–$99,500	–$99,250	–$99,000	–$98,750	$1250
Probability:	0.00183	0.00186	0.00189	0.00191	0.00193	

7.56 LIFE INSURANCE, II It would be quite risky for you to insure the life of a 21-year-old friend under the terms of the previous exercise. There is a high probability that your friend would live and you would gain $1250 in premiums. But if he were to die, you would lose almost $100,000. Explain carefully why selling insurance is not risky for an insurance company that insures many thousands of 21-year-old men.

7.57 LIFE INSURANCE, III The risk of an investment is often measured by the standard deviation of the return on the investment. The more variable the return is (the larger σ is), the riskier the investment. We can measure the great risk of insuring a single person's life in Exercise 7.55 by computing the standard deviation of the income X that the insurer will receive. Find σ_X, using the distribution and mean found in Exercise 7.55.

7.58 LIFE INSURANCE, IV The risk of insuring one person's life is reduced if we insure many people. Use the result of the previous exercise and rules for means and variances to answer the following questions.

(a) Suppose that we insure two 21-year-old males, and that their ages at death are independent. If X and Y are the insurer's income from the two insurance policies, the insurer's average income on the two policies is

$$Z = \frac{X+Y}{2} = 0.5X + 0.5Y$$

Find the mean and standard deviation of Z. You see that the mean income is the same as for a single policy but the standard deviation is less.

(b) If four 21-year-old men are insured, the insurer's average income is

$$Z = \frac{1}{4}(X_1 + X_2 + X_3 + X_4)$$

where X_i is the income from insuring one man. The X_i are independent and each has the same distribution as before. Find the mean and standard deviation of Z. Compare your results with the results of (a). We see that averaging over many insured individuals reduces risk.

7.59 AUTO EMISSIONS The amount of nitrogen oxides (NOX) present in the exhaust of a particular type of car varies from car to car according to the normal distribution with mean 1.4 grams per mile (g/mi) and standard deviation 0.3 g/mi. Two cars of this type are tested. One has 1.1 g/mi of NOX, the other 1.9. The test station attendant finds this much variation between two similar cars surprising. If X and Y are independent NOX levels for cars of this type, find the probability

$$P(X - Y \geq 0.8 \text{ or } X - Y \leq -0.8)$$

that the difference is at least as large as the value the attendant observed.

7.60 MAKING A PROFIT Rotter Partners is planning a major investment. The amount of profit X is uncertain but a probabilistic estimate gives the following distribution (in millions of dollars):

Profit:	1	1.5	2	4	10
Probability:	0.1	0.2	0.4	0.2	0.1

(a) Find the mean profit μ_X and the standard deviation of the profit.

(b) Rotter Partners owes its source of capital a fee of \$200,000 plus 10% of the profits X. So the firm actually retains

$$Y = 0.9X - 0.2$$

from the investment. Find the mean and standard deviation of Y.

7.61 A BALANCED SCALE You have two scales for measuring weights in a chemistry lab. Both scales give answers that vary a bit in repeated weighings of the same item. If the true weight of a compound is 2.00 grams (g), the first scale produces readings X that have a mean 2.000 g and standard deviation 0.002 g. The second scale's readings Y have a mean 2.001 g and standard deviation 0.001 g.

(a) What are the mean and standard deviation of the difference Y – X between the readings? (The readings X and Y are independent.)

(b) You measure once with each scale and average the readings. Your result is $Z = (X + Y)/2$. What are μ_Z and σ_Z? Is the average Z more or less variable than the reading Y of the less variable scale?

7.62 IT'S A GIRL! A couple plans to have children until they have a girl or until they have four children, whichever comes first. Example 5.24 (page 313) estimated the probability that they will have a girl among their children. Now we ask a different question: How many children, on the average, will couples who follow this plan have?

(a) To answer this question, construct a simulation similar to that in Example 5.24 but this time keep track of the number of children in each repetition. Carry out 25 repetitions and then average the results to estimate the expected value.

(b) Construct the probability distribution table for the random variable X = number of children.

(c) Use the table from (b) to calculate the expected value of X. Compare this number with the result from your simulation in (a).

7.63 SLIM AGAIN Amarillo Slim is back and he's got another deal for you. We have a fair coin (heads and tails each have probability 1/2). Toss it twice. If two heads come up, you win. If you get any other result, you get another chance: toss the coin twice more, and if you get two heads, you win. If you fail to get two heads on the second try, you lose. You pay a dollar to play. If you win, you get your dollar back plus another dollar.

(a) Explain how to simulate one play of this game using Table B. How could you simulate one play using your calculator? Simulate two tosses of a fair coin.

(b) Simulate 50 plays, using Table B or your calculator. Use your simulation to estimate the expected value of the game.

(c) There are two outcomes in this game: win or lose. Let the random variable X be the (monetary) outcome. What are the two values X can take? Calculate the actual probabilities of each value of X. Then calculate μ_X. How does this compare with your estimate from the simulation in (b)?

7.64 BE CREATIVE Here is a simple way to create a random variable X that has mean μ and standard deviation σ: X takes only the two values $\mu - \sigma$ and $\mu + \sigma$, each with probability 0.5. Use the definition of the mean and variance for discrete random variables to show that X does have mean μ and standard deviation σ.

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

7.66 A MECHANICAL ASSEMBLY A mechanical assembly (Figure 7.12) consists of a shaft with a bearing at each end. The total length of the assembly is the sum X + Y + Z of the shaft length X and the lengths Y and Z of the bearings. These lengths vary from part to part in production, independently of each other and with normal distributions. The shaft length X has mean 11.2 inches and standard deviation 0.002 inch, while each bearing length Y and Z has mean 0.4 inch and standard deviation 0.001 inch.

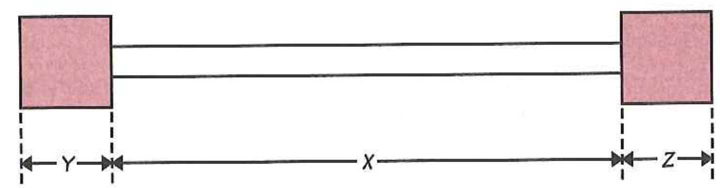


FIGURE 7.12 The dimensions of a mechanical assembly.

(a) According to the 68–95–99.7 rule, about 95% of all shafts have lengths in the range $11.2 \pm d_1$ inches. What is the value of d_1? Similarly, about 95% of the bearing lengths fall in the range of $0.4 \pm d_2$. What is the value of d_2?

(b) It is common practice in industry to state the "natural tolerance" of parts in the form used in (a). An engineer who knows no statistics thinks that tolerances add, so that the natural tolerance for the total length of the assembly (shaft and two bearings) is $12 \pm d$ inches, where $d = d_1 + 2d_2$. Find the standard deviation of the total length X + Y + Z. Then find the value d such that about 95% of all assemblies have lengths in the range $12 \pm d$. Was the engineer correct?

7.67 SWEDISH BRAINS A study of the weights of the brains of Swedish men found that the weight X was a random variable with mean 1400 grams and standard deviation 20 grams. Find numbers a and b such that $Y = a + bX$ has mean 0 and standard deviation 1.

7.68 ROLLING THE DICE You are playing a board game in which the severity of a penalty is determined by rolling three dice and adding the spots on the up-faces. The dice are all balanced so that each face is equally likely, and the three dice fall independently.

(a) Give a sample space for the sum X of the spots.

(b) Find $P(X = 5)$.

(c) If X_1, X_2, and X_3 are the number of spots on the up-faces of the three dice, then $X = X_1 + X_2 + X_3$. Use this fact to find the mean μ_X and the standard deviation σ_X without finding the distribution of X. (Start with the distribution of each of the X_i.)

NOTES AND DATA SOURCES

1. We use $\bar{x}$ both for the random variable, which takes different values in repeated sampling, and for the numerical value of the random variable in a particular sample. Similarly, s and $\hat{p}$ stand both for random variables and for specific values. This notation is mathematically imprecise but statistically convenient.
2. In most applications X takes a finite number of possible values. The same ideas, implemented with more advanced mathematics, apply to random variables with an infinite but still countable collection of values. An example is a geometric random variable, considered in Section 8.2.
3. From the Census Bureau's 1998 American Housing Survey.
4. The mean of a continuous random variable X with density function $f(x)$ can be found by integration:

$$\mu_X = \int x f(x)dx$$

This integral is a kind of weighted average, analogous to the discrete-case mean

$$\mu_X = \Sigma x P(X = x)$$

The variance of a continuous random variable X is the average squared deviation of the values of X from their mean, found by the integral

$$\sigma_X^2 = \int (x - \mu)^2 f(x)dx$$

5. See A. Tversky and D. Kahneman, "Belief in the law of small numbers," *Psychological Bulletin*, 76 (1971), pp. 105–110, and other writings of these authors for a full account of our misperception of randomness.
6. Probabilities involving runs can be quite difficult to compute. That the probability of a run of three or more heads in 10 independent tosses of a fair coin is (1/2) + (1/128) = 0.508 can be found by clever counting, as can the other results given in the text. A general treatment using advanced methods appears in Section XIII.7 of William Feller, *An Introduction to Probability Theory and Its Applications*, vol. 1, 3rd ed., Wiley, New York, 1968.
7. R. Vallone and A. Tversky, "The hot hand in basketball: on the misperception of random sequences," *Cognitive Psychology*, 17 (1985), pp. 295–314. A later series of articles that debate the independence question is A. Tversky and T. Gilovich, "The cold facts about the 'hot hand' in basketball," *Chance*, 2, no. 1 (1989), pp. 16–21; P. D. Larkey, R. A. Smith, and J. B. Kadane, "It's OK to believe in the 'hot hand,'" *Chance*, 2, no. 4 (1989), pp. 22–30; and A. Tversky and T. Gilovich, "The 'hot hand': statistical reality or cognitive illusion?" *Chance*, 2, no. 4 (1989), pp. 31–34.
8. The data on returns are from several sources, especially the *Fidelity Insight* newsletter, fidelity.kobren.com.
9. See Note 8


The Granger Collection, New York

PIERRE-SIMON LAPLACE

The Best Mathematician in France

Pierre-Simon Laplace (1749–1827) may be best remembered for his work on mathematical astronomy and the theory of probability. Before Laplace, probability theory was solely concerned with the mathematical analysis of games of chance. Laplace applied probabilistic ideas to many scientific and practical problems.

In 1812 he published the first of a series of four books on probability theory and its applications. In the first book, he studied generating functions and approximations to various expressions occurring in probability theory. The second book included Laplace's definition of probability, Bayes's rule, and remarks on moral and mathematical expectation, on methods of finding probabilities of compound events, on the method of least squares, on Buffon's needle problem, and on inverse probability. He also included work on probability in legal matters and applications to mortality, life expectancy, and the length of marriages. Later editions applied probability to errors in observations, to determining the masses of several planets, to triangulation methods in surveying, and to problems in geodesy.

Laplace survived the French Revolution by changing his views with the changing political events of the time. His colleague Lavoisier was a casualty. Despite Laplace's important contributions to science, he was not well liked by his colleagues. He was not modest about his abilities and achievements, and he let it be known widely that he considered himself the best mathematician in France. And he was! Laplace is now widely regarded as one of the greatest and most influential scientists of all time.

Laplace applied probabilistic ideas to many scientific and practical problems.