

a politician as a sample proportion. In this context, the normal approximation is a very useful tool for studying the fluctuations of \hat{p} about its true value p ; it is particularly useful for constructing a *confidence interval* for the unknown proportion p . A confidence interval is a measure of the accuracy, or reliability, of \hat{p} as an estimator of p ; it is discussed in detail in Chap. 7.

To apply the normal approximation to $\hat{p} = S_n/n$, we begin by dividing the numerator and denominator of $(S_n - np)/\sqrt{np(1-p)}$ by n . This yields the normal approximation for the sample proportion \hat{p} :

$$\begin{aligned} \frac{S_n - np}{\sqrt{np(1-p)}} &= \frac{S_n/n - p}{\sqrt{p(1-p)/n}} \\ &= \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \approx N(0, 1) \end{aligned} \quad (4.36)$$

Suppose the current approval rating of a politician is 52 percent. How likely is it that a poll based on a sample of size $n = 2000$ would show that her approval rating is less than 50 percent?

Solution. We are interested in computing the probability $P(\hat{p} < 0.5)$ under the assumption that $p = 0.52$. Using the normal approximation as expressed in Eq. (4.36), we proceed as follows:

$$\begin{aligned} P(\hat{p} < 0.5) &= P\left(\frac{\hat{p} - 0.52}{\sqrt{0.52 \times 0.48/2000}} < \frac{0.5 - 0.52}{0.0112}\right) \\ &\approx P(Z < -1.79) = 0.0367 \end{aligned}$$

PROBLEMS

4.22 Let Z be a standard normal random variable. Use the table of the normal distribution (Table A.3) to compute the following probabilities.

- (a) $P(Z < 1)$ (b) $P(Z < -1)$ (c) $P(|Z| < 1)$
 (d) $P(Z < -1.64)$ (e) $P(Z > 1.64)$ (f) $P(|Z| < 1.64)$
 (g) $P(Z > 2)$ (h) $P(|Z| > 2)$ (i) $P(|Z| < 1.96)$

4.23 Estimate each of the following probabilities using Chebyshev's inequality, and compare the estimates with the exact answers obtained from the table of the normal distribution.

- (a) $P(|Z| > 1)$ (b) $P(|Z| > 1.5)$ (c) $P(|Z| > 2)$

4.24 Find c such that:

- (a) $P(Z < c) = 0.25$ (b) $P(Z < c) = 0.75$
 (c) $P(|Z| < c) = 0.5$ (d) $P(Z < c) = 0.85$
 (e) $P(Z > c) = 0.025$ (f) $P(|Z| < c) = 0.95$

- 4.34** The random variable X has a binomial distribution with parameters $n = 25$ and $p = 0.5$.
- Compute the exact value of $P(9 \leq X \leq 16)$ using Table A.1.
 - Compute $P(9 \leq X \leq 16)$ using the normal approximation to the binomial.
 - Compute the same probability in part (b) using the continuity correction. Does the continuity correction improve the accuracy of the approximation?
- 4.35** Let X denote the number of heads that appear when a fair coin is tossed 300 times. Use the normal approximation to the binomial to estimate the following probabilities:
- $P(X \geq 160)$
 - $P(X \leq 140)$
 - $P(|X - 150| \geq 20)$
 - $P(135 < X < 165)$
- 4.36** Let Y denote the number of times a 1 appears in 720 throws of a die. Using the normal approximation to the binomial, approximate the following probabilities:
- $P(Y \geq 130)$
 - $P(Y \geq 140)$
 - $P(|Y - 120| \geq 20)$
 - $P(105 < Y < 120)$
- 4.37** A certain type of seed has probability 0.7 of germinating. Out of a package of 100 seeds, let Y denote the number of seeds that germinate.
- Give the formula for the exact distribution of Y .
 - Calculate the probability that at least 75 percent of the seeds germinate.
 - Calculate the probability that at least 80 percent germinate.
 - Calculate the probability that no more than 65 percent germinate.
- 4.38** The natural recovery rate for a certain disease is 40 percent. A new drug is developed, and the manufacturer claims that it boosts the recovery rate to 80 percent. Two hundred patients are given the drug. What is the probability that at least 100 patients recover if:
- The drug is worthless?
 - It is really 80 percent effective as claimed?
- 4.39** Suppose that the velocity V of a particle of mass m is $N(0, 1)$ distributed. Denote its kinetic energy by $K = mv^2/2$. Show that $E(K) = m/2$ and $V(K) = m^2/2$.
- 4.40** Suppose the probability that a person cancels a hotel reservation is 0.1 and the hotel's capacity is 1000 rooms. To minimize the number of empty rooms, the hotel routinely overbooks. How many reservations above 1000 should the hotel accept if it wants the probability of overbooking to be less than 0.01?

4.5.1 The

4.5 OTHER IMPORTANT CONTINUOUS DISTRIBUTIONS

Although the normal distribution is the most important one, there remain a wide variety of phenomena in science and engineering for which it is inappropriate. For example, the distribution of the lifetime of a TV monitor is not symmetric,

- C 6.2 A toy consists of three parts whose respective weights (measured in grams) are denoted by X_1, X_2, X_3 . Assume that the weights are independent and normally distributed with

$$X_1 \sim N(150, 36), \quad X_2 \sim N(100, 25), \quad X_3 \sim N(250, 60)$$

- Give a formula for the distribution of total weight of the toy $T = X_1 + X_2 + X_3$; name it and identify the parameters.
- What proportion of the toys have weight greater than 520 g?
- Suppose the toys are shipped 24 to a box. The box weighs 200 g. Calculate the expected value and variance of the total weight of the box.
- Calculate the distribution of the total weight W of the box. What is the probability that the total weight of the box is greater than 12,300 g?

- HW 6.3 Suppose that the time (measured in minutes) to repair a component is normally distributed with $\mu = 65, \sigma = 10$.

- What is the proportion of components that are repaired in less than one hour?
- Give a formula for the distribution of the total time required to repair eight components. What is the probability that eight components are repaired in an eight-hour working day?

- HW 6.4 Let X_1, \dots, X_n be iid $N(2, 4)$ -distributed random variables.

- If $n = 100$, compute $P(1.9 < \bar{X} < 2.1)$.
- How large must n be so that $P(1.9 < \bar{X} < 2.1) = 0.9$?

- 6.5 Let X_1, \dots, X_9 be iid $N(2, 4)$ -distributed and Y_1, \dots, Y_4 be iid $N(1, 1)$ -distributed. The Y_i are assumed to be independent of the X_i .

- Describe the distribution of $\bar{X} - \bar{Y}$.
- Compute $P(\bar{X} > \bar{Y})$.

- C 6.6 Let \bar{X} denote the sample mean of a random sample of size $n_1 = 16$ taken from a normal distribution $N(\mu, 36)$, and let \bar{Y} denote the sample mean of a random sample of size $n_2 = 25$ taken from a different normal distribution $N(\mu, 9)$.

- Describe the distribution of $\bar{X} - \bar{Y}$. Name it and identify the parameters.
- Compute $P(\bar{X} - \bar{Y} > 5)$.
- Compute $P(|\bar{X} - \bar{Y}| > 5)$.

- HW 6.7 To determine the difference, if any, between two brands of radial tires, 12 tires of each brand are tested. Assume that the lifetimes of both brands of tires come from the same normal distribution with $\sigma = 3300$.

- Describe the distribution of the difference of the sample means $\bar{X} - \bar{Y}$. Name it and identify the parameters.
- The engineers recorded the following values for the mean lifetime of each brand:

$$\bar{x} = 38,500, \quad \bar{y} = 41,000 \quad \text{so} \quad \bar{x} - \bar{y} = -2500$$

If in fact the lifetimes of both brands have the same distribution, what is the probability that the difference between the sample means is at

least as large as the observed difference of 2500? That is, compute $P(|\bar{X} - \bar{Y}| > 2500)$.

(c) [Part (b) continued] Compute $P(\bar{X} - \bar{Y} < -2500)$.

C **6.8** Assume that the miles-per-gallon (mpg) of two brands of gasoline come from the same normal distribution with $\sigma = 2.0$. Four cars are driven with brand A gasoline, and six cars are driven with brand B gasoline. Denote the average mpg rates obtained with brand A and brand B gasolines by \bar{X} and \bar{Y} , respectively.

(a) Describe the distribution of the difference of the sample means $\bar{X} - \bar{Y}$. Name it and identify the parameters.

(b) The engineers recorded the following mpg rates for each brand:

$$\bar{x} = 32.5 \quad \text{and} \quad \bar{y} = 31, \quad \text{so} \quad \bar{x} - \bar{y} = 1.5$$

If, in fact, the mpg rates of the two brands have the same distribution, what is the probability that the difference between the sample means is at least as large as the observed difference of 1.5? That is, compute $P(|\bar{X} - \bar{Y}| > 1.5)$.

(c) [Part (b) continued] Compute $P(\bar{X} - \bar{Y} > 1.5)$.

HW **6.9** An elevator has weight capacity 3000 lb. Assume that the weight X of a randomly selected person is $N(175, 400)$ distributed, i.e., $\mu = 175, \sigma = 20$. Let W_n denote the total weight of n passengers.

(a) Give the formulas for $E(W_n)$ and $V(W_n)$.

(b) Describe the distribution of W_n .

(c) If 18 people get on the elevator, what is the probability that the total weight of the passengers exceeds the weight capacity of the elevator?

(d) What is the maximum number N of people allowed on the elevator such that the probability of their combined weight exceeding the 3000-lb load limit is less than 0.05?

HW **6.10** Boxes of cereal are filled by machine, and the net weight of the filled box is a random variable with mean $\mu = 10$ oz and variance $\sigma^2 = 0.5$ oz². A carton of cereal contains 48 boxes. Let T be the total weight of a carton.

(a) Give a formula for the approximate distribution of T using the central limit theorem.

(b) Use the central limit theorem to estimate $P(T > 31 \text{ lb})$.

6.11 The random variables X_1, \dots, X_{40} are independent with the same probability function given by

x	-1	0	1
$f(x)$	0.2	0.2	0.6

(a) Compute the mean and variance of the sample total $T = X_1 + \dots + X_{40}$.

(b) What is the largest value that T can assume? The smallest?

(c) What does the central limit theorem say about the distribution of T ? Use it to compute $P(T > 25)$ and $P(T < 0)$.

Properties of th