

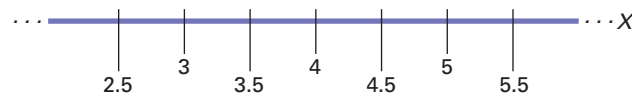
6.6 The Normal Approximation to the Binomial Distribution

In the earlier sections of this chapter, you learned about the normal probability distribution. In this section, you will learn how to use the normal distribution to approximate the binomial distribution (see Section 5.3). Recall that the binomial distribution is a discrete distribution where the random variable X is the number of events of interest occurring in a sample of n trials. Calculating exact probabilities of X using Equation (5.11) becomes very tedious when n is large, and thus it is often useful to approximate the probability using the normal approximation introduced in this section.

Need for a Correction for Continuity Adjustment

When using the normal distribution to approximate the binomial distribution, you can get more accurate approximations of the probabilities if you use a correction for continuity adjustment. There are two major reasons for using a correction. First, discrete random variables that follow the binomial distribution can take on only integer values, while a continuous random variable that follows the normal distribution can take on any values within a continuum or interval. Second, with a continuous distribution such as the normal distribution, the probability of getting a specific value of a random variable is zero. However, when the normal distribution is used to approximate a discrete distribution, you can use a correction for continuity adjustment to compute the approximate probability for the binomial distribution.

Consider an experiment in which you toss a fair coin 10 times. Suppose you want to compute the probability of getting *exactly* 4 heads. Whereas a discrete random variable can have only an integer value (such as 4), a continuous random variable used to approximate it could take on any value within an interval around that specified value, as demonstrated on the following scale:



The correction for continuity adjustment requires adding or subtracting 0.5 from the value or values of the discrete random variable X , as needed. To use the normal distribution to approximate the probability of getting *exactly* 4 heads (i.e., $X = 4$), you need to find the area under the normal curve from $X = 3.5$ to $X = 4.5$, the lower and upper boundaries of 4. To determine the approximate probability of getting *at least* 4 heads, you find the area under the normal curve greater than or equal to $X = 3.5$ because 3.5 is the lower boundary of 4. Similarly, to determine the approximate probability of getting *at most* 4 heads, you find the area under the normal curve equal to or less than $X = 4.5$ because 4.5 is the upper boundary of 4.

When using the normal distribution to approximate discrete probability distributions, wording is especially important. To determine the approximate probability of getting *fewer* than 4 heads, you find the area under the normal curve less than or equal to $X = 3.5$. To determine the approximate probability of getting *more than* 4 heads, you find the area under the normal curve greater than or equal to $X = 4.5$. To determine the approximate probability of getting *4 through 7* heads, you find the area under the normal curve from $X = 3.5$ to $X = 7.5$.

Approximating the Binomial Distribution

In Section 5.3 you learned that the binomial distribution is symmetrical (like the normal distribution) whenever $p = 0.5$. When $p \neq 0.5$, the binomial distribution is not symmetrical. However, the closer p is to 0.5 and the larger the sample size n , the more symmetric the distribution becomes. However, the larger the sample size, the more tedious it is to compute the exact probability of an event of interest by using Equation (5.11) on page 172. Fortunately, whenever the sample size is large, you can use the normal distribution to approximate the exact probabilities of the items of interest.

As a general rule, you can use the normal distribution to approximate the binomial distribution whenever both $n\pi$ and $n(1 - \pi)$ are at least 5. From Section 5.3, you know that the mean of the binomial distribution is

$$\mu = n\pi$$

and the standard deviation of the binomial distribution is

$$\sigma = \sqrt{n\pi(1 - \pi)}$$

Substituting these results into the transformation formula [Equation (6.2) on page 196],

$$\begin{aligned} Z &= \frac{X - \mu}{\sigma} \\ &= \frac{X - n\pi}{\sqrt{n\pi(1 - \pi)}} \end{aligned}$$

so that, for large enough n , the random variable Z is approximately normally distributed. Hence, you find approximate probabilities corresponding to the values of the discrete random variable X by using Equation (6.11).

NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

$$Z = \frac{X_a - n\pi}{\sqrt{n\pi(1 - \pi)}} \quad (6.11)$$

where

$\mu = n\pi$, mean of the binomial distribution

$\sigma = \sqrt{n\pi(1 - \pi)}$, standard deviation of the binomial distribution

X_a = adjusted number of items of interest for the discrete random variable X , such that $X_a = X - 0.5$ or $X_a = X + 0.5$, as appropriate

EXAMPLE 6.6

Using the Normal Distribution to Approximate the Binomial Distribution

You select a random sample of $n = 1,600$ tires from an ongoing production process in which 8% of all such tires produced are defective. What is the probability that 150 or fewer tires will be defective?

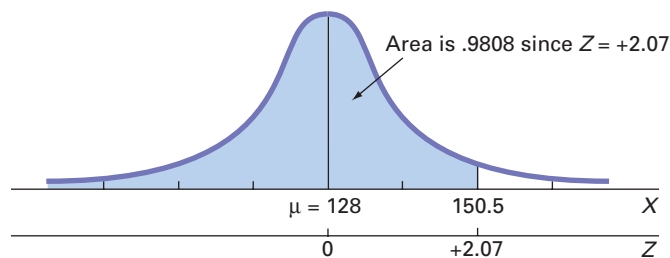
SOLUTION Because both $n\pi = 1,600(0.08) = 128$ and $n(1 - \pi) = 1,600(0.92) = 1,472$ are greater than 5, you can use the normal distribution to approximate the binomial. Here X_a , the adjusted number of successes, is 150.5.

$$Z \cong \frac{X_a - n\pi}{\sqrt{n\pi(1 - \pi)}} = \frac{150.5 - 128}{\sqrt{(1,600)(0.08)(0.92)}} = \frac{22.5}{10.85} = +2.07$$

Using Table E.2, the area under the curve to the left of $Z = +2.07$ is 0.9808 (see Figure 6.24).

FIGURE 6.24

Approximating the binomial distribution



Suppose that you want to approximate the probability of getting *exactly* 150 defective tires. The correction for continuity defines the integer value of interest to range from one-half unit below it to one-half unit above it. Therefore, you define the probability of getting 150 defective

tires as the area (under the normal curve) between 149.5 and 150.5. Using Equation (6.11), you approximate the probability as follows:

$$Z = \frac{150.5 - 128}{\sqrt{(1,600)(0.08)(0.92)}} = \frac{22.5}{10.85} = +2.07$$

and

$$Z = \frac{149.5 - 128}{\sqrt{(1,600)(0.08)(0.92)}} = \frac{21.5}{10.85} = +1.98$$

From Table E.2, the area under the normal curve less than $X = 150.5$ ($Z = +2.07$) is 0.9808, and the area under the curve less than $X = 149.5$ ($Z = +1.98$) is 0.9761. Thus, the approximate probability of getting 150 defective tires is the difference in the two areas, 0.0047.

Problems for Section 6.6

LEARNING THE BASICS

6.60 For $n = 100$ and $\pi = 0.20$, use the normal distribution to approximate the probability that

- $X = 25$.
- $X > 25$.
- $X \leq 25$.
- $X < 25$.

6.61 For $n = 100$ and $p = 0.40$, use the normal distribution to approximate the probability that

- $X = 40$.
- $X > 40$.
- $X \leq 40$.
- $X < 40$.

APPLYING THE CONCEPTS

6.62 Consider an experiment in which you toss a fair coin 10 times and count the number of heads. Use Equation (5.11) on page 172 or the **Binomial workbook** or the **Binomial online topic file** to determine the probability of getting

- 4 heads.
- at least 4 heads.
- 4 through 7 heads.
- Use the normal approximation to the binomial distribution to approximate the probabilities in (a) through (c).

6.63 For short domestic flights, an airline has three different choices on its snack menu—pretzels, potato chips, and cookies. Based on past experience, the airline feels that each snack is equally likely to be chosen. If there are 150 passengers on a particular flight, what is the *approximate* probability that

- at least 60 will choose pretzels for dessert?
- exactly 60 will choose pretzels for dessert?
- fewer than 60 will choose pretzels for dessert?
- If the airline has 70 of each type of snack available on the flight, what is the likelihood that a passenger will not be able to get the snack that he or she desires?

6.64 Stock options are usually awarded at the price of a company's shares on the date of the grant. A recent study (E. Awata, "Backdated Options May Snare Some Directors," *USA Today*, March 29, 2007, pp. 1B, 2B) found evidence that many outside directors received grants at the lowest price in a given month. Of 17,512 grants before the Sarbanes-Oxley Act tightened Security and Exchange Commission disclosure rules, 1,726 were at the lowest price in a given month. Assume that there is an average of 21 days in a month on which grants can be made. How likely do you think it is that these results or more extreme results could have occurred if the grant is equally likely to be given on any day of the month?