5.6 Using the Poisson Distribution to Approximate the Binomial Distribution

Use the Poisson distribution to approximate the binomial distribution when n is large and π is very small. The approximation gets better as n gets larger and π gets smaller. The following mathematical expression for the Poisson model approximates the true (binomial) result:

Let X be a r.v. that is distributed as a Poisson distribution

$X \sim \text{Poisson}(\mu)$

$X \sim \text{Poisson}(\lambda)$

$\lambda = \mu$

$$P(X = x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

n large

p or π small

$$P(X = x|\lambda) \cong \frac{e^{-n\pi}(n\pi)^X}{X!} \quad \textbf{(5.18)}$$

where

$\lambda = n\pi = np$

$P(X = x|\lambda) =$ probability of $X = x$ events of interest given the parameters n and π

$n =$ sample size

$\pi =$ probability of an event of interest

$e =$ mathematical constant approximated by 2.71828

$x =$ number of events of interest in the sample ($x = 0, 1, 2, \ldots n$)

The Poisson random variable theoretically ranges from 0 to ∞. However, when the Poisson distribution is used as an approximation to the binomial distribution, the Poisson random variable—the number of events of interest out of n observations—cannot be greater than the sample size n. With large n and small π, Equation (5.18) implies that the probability of observing a large number of events of interest becomes small and approaches zero rapidly.

As Section 5.3 mentions, in the Poisson distribution, the mean μ and the variance σ^2 are each equal to λ. Thus, when using the Poisson distribution to approximate the binomial distribution, use Equation (5.19) to approximate the mean.

$$\mu = E(X) = \lambda = n\pi = np \quad \textbf{(5.19)}$$

Equation (5.20) approximates the standard deviation.

$$\sigma = \sqrt{\lambda} = \sqrt{n\pi} = \sqrt{np} \quad \textbf{(5.20)}$$

The standard deviation given by Equation (5.20) closely approximates the standard deviation for the binomial model [Equation (5.7)] when π is close to zero so that $(1 - \pi)$ is close to one.

Suppose 8% of the tires manufactured at a particular plant are defective. To illustrate the use of the Poisson approximation for the binomial, calculate the probability of exactly one defective tire from a sample of 20 using Equation (5.18) as follows:

$$P(X = 1) = \frac{e^{-(20)(0.08)}[(20)(0.08)]^1}{1!} = \frac{e^{-1.6}(1.6)^1}{1!} = 0.3230$$

Alternatively, one can use tables of the Poisson distribution (see Table E.7) to get the Poisson approximation or software, including Excel, JMP, and Minitab. When using the Poisson tables, only the parameter λ and the desired number of events of interest x are needed. Because in the above example $\lambda = 1.6$ and $X = 1$,

$$P(X = 1) = 0.3230$$

Table 5.6 shows this probability.

TABLE 5.6
Finding a Poisson Probability

	λ									
X	**1.1**	**1.2**	**1.3**	**1.4**	**1.5**	**1.6**	**1.7**	**1.8**	**1.9**	**2.0**
0	.3329	.3012	.2725	.2466	.2231	.2019	.1827	.1653	.1496	.1353
1	.3662	.3614	.3543	.3452	.3347	.3230	.3106	.2975	.2842	.2707
2	.2014	.2169	.2303	.2417	.2510	.2584	.2640	.2678	.2700	.2707
3	.0738	.0867	.0998	.1128	.1255	.1378	.1496	.1607	.1710	.1804
4	.0203	.0260	.0324	.0395	.0471	.0551	.0636	.0723	.0812	.0902

Source: Extracted from Table E.7.

Had the true distribution, the binomial, been used instead of the approximation,

$$P(X = 1) = \frac{20!}{1!19!}(0.08)^1(0.92)^{19} = 0.3282$$

However, that calculation is tedious. With software tools available today, one can compute directly the binomial probability for $n = 20$, $\pi = 0.08$, and $X = 1$, thus making the Poisson approximation less important to know.

PROBLEMS FOR SECTION 5.6

LEARNING THE BASICS

5.63 When should you use the Poisson distribution to approximate the binomial distribution? n large, p small

5.64 When given the parameters of a binomial distribution, n and π, what are the mean and variance of the Poisson distribution used for approximating the binomial? $\mu = n\pi$, $\sigma^2 = n\pi$

5.65 Given a binomial distribution with $n = 100$ and $\pi = 0.01$, use the Poisson distribution to approximate the following:

a. $P(X = 0)$
b. $P(X = 1)$
c. $P(X = 2)$
d. $P(X \leq 2)$
e. $P(X > 2)$

5.66 Given a binomial distribution with $n = 50$ and $\pi = 0.004$, use the Poisson distribution to approximate the following:

a. $P(X = 0)$
b. $P(X = 1)$
c. $P(X = 2)$
d. $P(X \leq 2)$
e. $P(X > 2)$

APPLYING THE CONCEPTS

5.67 Based upon past experience, 1% of the telephone bills mailed to households are incorrect. If a sample of 20 bills is selected, using the binomial distribution and the Poisson approximation to the binomial distribution, find the probability that at least one bill is incorrect. Briefly compare and explain your results.

5.68 A computer manufacturing company samples incoming computer chips. After receiving a large shipment of computer chips, the company randomly selects 800 chips. If three or fewer nonconforming chips are found, the entire lot is accepted without inspecting the remaining chips in the lot. If four or more chips are nonconforming, every chip in the entire lot is carefully inspected at the supplier's expense. Assume that the true proportion of nonconforming computer chips being supplied is 0.001. Approximate the probability the lot will be accepted.

5.69 Last month your company sold 10,000 new watches. Past experience indicates that the probability that a new watch will need repair during its warranty period is 0.002. Approximate the probability that:

a. zero watches will need warranty work.
b. no more than 5 watches will need warranty work.
c. no more than 10 watches will need warranty work.
d. no more than 20 watches will need warranty work.

5.65 Given a binomial distribution with $n = 100$ and $\overset{p = 0.01}{\pi} = 0.01$, use the Poisson distribution to approximate the following:

a. $P(X = 0)$

b. $P(X = 1)$

c. $P(X = 2)$

d. $P(X \leq 2)$

e. $P(X > 2)$

original $X \sim B(n=100, p=0.01)$

approximation $X \sim \text{Poisson}(\lambda = np)$

a. $P(X=0) = \binom{100}{0}(0.01^{0} \times 0.99^{100})$

c. $P(X=2) = \binom{100}{2}(0.01^{2})(0.99^{98})$

$= \frac{100!}{2!\,98!}(0.01^{2})(0.99^{98})$

$= \frac{100 \times 99 \times \cancel{98!}}{2 \times \cancel{98!}} \times (0.01^{2})(0.99^{98}) \approx 0.185$

direct calculation

Use Poisson approximation to X

Let $\lambda = np \approx 100 \times 0.01 \approx 1$

Then $X \sim \text{Poisson}(\lambda = 1)$ [approximation]

$$P(X=2) = \frac{e^{-1} \times 1^{2}}{2!} \approx 0.184$$

$X \sim B(n=100,\ p=0.01)$

$P(X>2) = 1 - P(X \le 2)$

$= 1 - [P(X=0)+P(X=1)+P(X=2)]$

$X_a \sim \text{Poisson}(\lambda = np)$, $np = 100 \times 0.01 = 1$

$X_a \sim \text{Poisson}(\lambda = 1)$

$P(X_a > 2) = 1 - [P(X_a=0)+P(X_a=1)+P(X_a=2)]$

$= 1 - \left[\frac{e^{-1} \times 1^{0}}{0!} + \frac{e^{-1} \times 1^{1}}{1!} + \frac{e^{-1} \times 1^{2}}{2!}\right]$

$= 1 - e^{-1}\left[1 + 1 + \frac{1}{2}\right]$

$= 1 - \frac{5}{2}e^{-1}$

$= 0.080$

6.6 The Normal Approximation to the Binomial Distribution

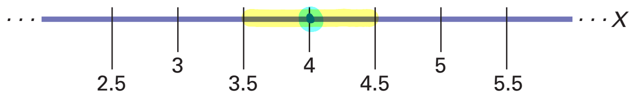
In the earlier sections of this chapter, you learned about the normal probability distribution. In this section, you will learn how to use the normal distribution to approximate the binomial distribution (see Section 5.2). Recall that the binomial distribution is a discrete distribution where the random variable X is the number of events of interest occurring in a sample of n trials. Calculating exact probabilities of X using Equation (5.5) 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 variables that follow the binomial distribution can take on only integer values, while a continuous 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 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 variable can have only an integer value (such as 4), a continuous variable used to approximate it could take on any value within an interval around that specified value, as demonstrated on the following scale:

$X \sim B(n,p)$, $P(X=4)$



$X_c \sim N(\mu, \sigma^2)$, $P(3.5 \leq X_c \leq 4.5)$

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 5.2, you learned that the binomial distribution is symmetrical (like the normal distribution) whenever $\pi = 0.5$. When $\pi \neq 0.5$, the binomial distribution is not symmetrical. However, the closer π 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 (5.5). Fortunately, whenever the sample size is large, you can use the normal distribution to approximate the exact probabilities of the items of interest.

n large , $\pi \approx 0.5$
$p \approx 0.5$
$np \geq 5, \; n(1-p) \geq 5$

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 5.2, 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)],

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

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

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

NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

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

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 6.8

Using the Normal Distribution to Approximate the Binomial Distribution

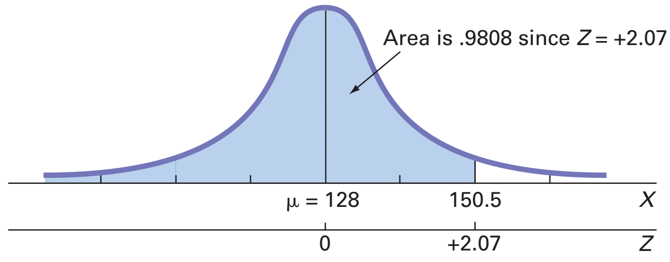
You select a random sample of $n = 1{,}600$ tires from an ongoing production process in which 8% of all such tires produced are defective. What is the probability that 150 or fewer tires will be defective?

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 6.23).

FIGURE 6.23
Approximating the binomial distribution



▶(*continued*)

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

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

a. $X = 25$.
b. $X > 25$.
c. $X \leq 25$.
d. $X < 25$.

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

a. $X = 40$.
b. $X > 40$.
c. $X \leq 40$.
d. $X < 40$.

APPLYING THE CONCEPTS

6.45 Consider an experiment in which you toss a fair coin 10 times and count the number of heads. Using Equation (5.5) on page 206, the method that the **Binomial Table online topic** describes, or Chapter 5 software guide instructions, determine the probability of getting

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

6.46 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

a. at least 60 will choose pretzels for dessert?
b. exactly 60 will choose pretzels for dessert?
c. fewer than 60 will choose pretzels for dessert?
d. 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?

6.47 Refer to Problem 5.43 on page 217.

a. Using the normal approximation to the binomial distribution, what is the probability that the indicator would be correct 38 or more times in 50 years?
b. Compare the results in (a) to those in Problem 5.43 (a).

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

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

a. $X = 25$.
b. $X > 25$.
c. $X \leq 25$.
d. $X < 25$.

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

a. $X = 40$.
b. $X > 40$.
c. $X \leq 40$.
d. $X < 40$.

APPLYING THE CONCEPTS

6.45 Consider an experiment in which you toss a fair coin 10 times and count the number of heads. Using Equation (5.5) on page 206, the method that the **Binomial Table online topic** describes, or Chapter 5 software guide instructions, determine the probability of getting

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

6.46 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

a. at least 60 will choose pretzels for dessert?
b. exactly 60 will choose pretzels for dessert?
c. fewer than 60 will choose pretzels for dessert?
d. 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?

6.47 Refer to Problem 5.43 on page 217.

a. Using the normal approximation to the binomial distribution, what is the probability that the indicator would be correct 38 or more times in 50 years?
b. Compare the results in (a) to those in Problem 5.43 (a).