

6.5: THE NORMAL APPROXIMATION TO THE BINOMIAL AND POISSON DISTRIBUTIONS

In the earlier sections of this chapter the normal probability distribution was discussed. In this section another useful aspect of the normal distribution is considered—how it may be used to approximate the binomial and Poisson probability distributions.

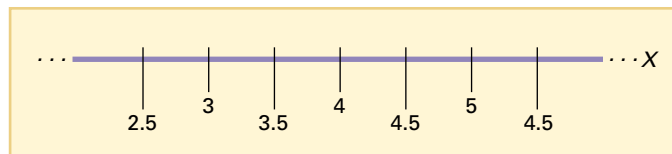
Need for a Correction for Continuity Adjustment

There are two major reasons to employ a correction for continuity adjustment here.

First, recall that a discrete random variable can take on only specified values while a continuous random variable can take on any values within a continuum or interval around those specified values. Hence, when using the normal distribution to approximate the binomial or the Poisson distributions, more accurate approximations of the probabilities are likely to be obtained if a correction for continuity adjustment is employed.

Second, recall that with a continuous distribution (such as the normal), the probability of obtaining a *particular* value of a random variable is zero. On the other hand, when the normal distribution is used to approximate a discrete distribution, a correction for continuity adjustment can be employed so that the probability of a specific value of the discrete distribution can be approximated.

As a case in point, consider an experiment in which you toss a fair coin 10 times and observe the number of heads. Suppose you want to compute the probability of obtaining *exactly* 4 heads. Whereas a discrete random variable can have only a specified value (such as 4), a continuous random variable used to approximate it could take on any values whatsoever within an interval around that specified value, as demonstrated on the accompanying 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. Hence to use the normal distribution to approximate the probability of obtaining *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 observing *at least* 4 heads, you find the area under the normal curve from $X = 3.5$ and above since, on a continuum, 3.5 is the lower boundary of X . Similarly, to determine the approximate probability of observing *at most* 4 heads, we would find the area under the normal curve from $X = 4.5$ and below since, on a continuum, 4.5 is the upper boundary of X .

When using the normal distribution to approximate discrete probability distributions, semantics are important. To determine the approximate probability of observing *fewer* than 4 heads, you find the area under the normal curve from $X = 3.5$ and below; to determine the approximate probability of observing *more than* 4 heads, you find the area under the normal curve from $X = 4.5$ and above; and to determine the approximate probability of observing *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 symmetric (like the normal distribution) whenever $p = .5$. When $p \neq .5$ the binomial distribution will not be symmetric. However, the closer p is to .5 and the larger the number of sample observations n , the more symmetric the distribution becomes.

On the other hand, the larger the number of observations in the sample, the more tedious it is to compute the exact probabilities of success by use of Equation (5.11). Fortunately, though, whenever the sample size is large, the normal distribution can be used to approximate the exact probabilities of success that otherwise would have to be obtained through laborious computations.

As a general rule this normal approximation can be used whenever np and $n(1 - p)$ are at least 5. Recall from section 5.3 that the mean of the binomial distribution is given by

$$\mu = np$$

and the standard deviation of the binomial distribution is obtained from

$$\sigma = \sqrt{np(1 - p)}$$

Substituting into the transformation formula (6.2)

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

and therefore

$$= \frac{X - np}{\sqrt{np(1 - p)}}$$

so that, for large enough n , the random variable Z is approximately normally distributed.

Hence, to find approximate probabilities corresponding to the values of the discrete random variable X , Equation (6.9) is used.

$$Z \cong \frac{X_a - np}{\sqrt{np(1 - p)}} \quad (6.9)$$

where

$\mu = np$, mean of the binomial distribution

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

X_a = adjusted number of successes for the discrete random variable X , such that

$X_a = X - .5$ or $X_a = X + .5$ as appropriate

Example 6.9

USING THE NORMAL DISTRIBUTION TO APPROXIMATE THE BINOMIAL DISTRIBUTION

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

SOLUTION Since both $np = 1,600(.08) = 128$ and $n(1 - p) = 1,600(0.92) = 1,472$ exceed 5, you use the normal distribution to approximate the binomial:

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

Here X_a , the adjusted number of successes, is 150.5 and the Z value is +2.07.

Using Table E.2, the area under the curve to the left of $Z = +2.07$ is 0.9808 (see Figure 6.35).

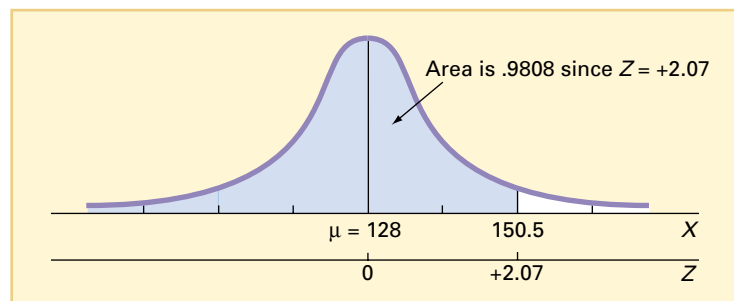


FIGURE 6.35
Approximating the binomial distribution

Under the binomial distribution the probability of obtaining not more than 150 defective tires consists of all events up to and including 150 defectives—that $P(X \leq 150) = P(X = 0) + P(X = 1) + \dots + P(X = 150)$, and the true probability is laboriously computed from

$$\sum_{X=0}^{150} \binom{1,600}{X} (.08)^X (.92)^{1,600-X}$$

To appreciate the amount of work saved by using the normal approximation to the binomial model in lieu of the exact probability computations, just imagine making the following 151 computations from Equation (5.11) on page 175 before summing up the results:

$$\binom{1,600}{0} (.08)^0 (.92)^{1,600} + \binom{1,600}{1} (.08)^1 (.92)^{1,599} + \dots + \binom{1,600}{150} (.08)^{150} (.92)^{1,450}$$

Obtaining a Probability Approximation for an Individual Value Suppose that you want to approximate the probability of obtaining *exactly* 150 defectives. 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, the probability of obtaining 150 defective tires is defined as the area (under the normal curve) between 149.5 and 150.5. Thus by using Equation (6.9), the probability can be approximated as follows:

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

and

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

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

Approximating the Poisson Distribution

The normal distribution can also be used to approximate the Poisson distribution whenever the parameter λ , the expected number of successes, equals or exceeds 5. Since the value of the mean and the variance of a Poisson distribution are the same,

$$\mu = \sigma^2 = \lambda$$

then the standard deviation is

$$\sigma = \sqrt{\lambda}$$

Substituting into the transformation Equation (6.2) on page 199,

$$\begin{aligned} Z &= \frac{X - \mu}{\sigma} \\ &= \frac{X - \lambda}{\sqrt{\lambda}} \end{aligned}$$

so that, for large enough λ , the random variable Z is approximately normally distributed.

Hence, to find approximate probabilities corresponding to the values of the Poisson random variable X Equation (6.10) is used.

$$Z \cong \frac{X_a - \lambda}{\sqrt{\lambda}} \quad (6.10)$$

where

λ = expected number of successes or mean of the Poisson distribution

$\sigma = \sqrt{\lambda}$, the standard deviation of the Poisson distribution

X_a = adjusted number of successes, x , for the discrete random variable X , such that

$X_a = X - 0.5$ or $x_a = X + 0.5$ as appropriate

Example 6.10

USING THE NORMAL DISTRIBUTION TO APPROXIMATE THE POISSON DISTRIBUTION

Suppose that at a certain automobile plant the average number of work stoppages per day due to equipment problems during the production process is 12.0. What is the approximate probability of having 15 or fewer work stoppages due to equipment problems on any given day?

SOLUTION Using Equation (6.10)

$$Z \cong \frac{X_a - \lambda}{\sqrt{\lambda}} = \frac{15.5 - 12.0}{\sqrt{12.0}} = +1.01$$

Here X_a , the adjusted number of successes, is 15.5. Hence the approximate probability that X does not exceed this value corresponds to a Z value, of not more than +1.01. From Table E.2, note that the area under the normal curve less than $Z = +1.01$ is 0.8438. Therefore, the approximate probability of having 15 or fewer work stoppages due to equipment problems on any given day is 0.8438. This approximation compares quite favorably to the exact Poisson probability, 0.8445.

PROBLEMS FOR SECTION 6.5

Learning the Basics

- 6.53 Why is a correction for continuity adjustment needed?
- 6.54 When can the normal distribution be used to approximate the binomial distribution?
- 6.55 When can the normal distribution be used to approximate the Poisson distribution?

Applying the Concepts

- 6.56 Consider an experiment in which you toss a fair coin 10 times and observe the number of heads. Use Equation (5.11) on page 175 or Table E.6 or PHStat or Minitab to determine the probability of observing
- 4 heads
 - at least 4 heads
 - at most 4 heads
 - fewer than 4 heads
 - more than 4 heads
 - 4 through 7 heads
 - Use the normal approximation to the binomial distribution to approximate the probabilities in (a)–(f).
 - Compare and contrast your findings in (a)–(f) and (g). Do you think that the normal distribution provides a good approximation to the binomial distribution in (g)?
- 6.57 For overseas flights, an airline has three different choices on its dessert menu—ice cream, apple pie, and chocolate cake. Based on past experience the airline feels that each dessert is equally likely to be chosen.
- If a random sample of four passengers is selected, what is the probability that at least two will choose ice cream for dessert?
 - If a random sample of 21 passengers is selected, what is the *approximate* probability that at least two will choose ice cream for dessert?
- 6.58 Based upon past experience, 40% of all customers at Miller's Automotive Service Station pay for their purchases with a credit card. If a random sample of three customers is selected, what is the probability that
- none pay with a credit card?
 - two pay with a credit card?
 - at least two pay with a credit card?
 - not more than two pay with a credit card?
- If a random sample of 200 customers is selected, what is the *approximate* probability that
- at least 75 pay with a credit card?
 - not more than 70 pay with a credit card?
 - between 70 and 75 customers, inclusive, pay with a credit card?

- 6.59 On average, 10.0 persons per minute are waiting for an elevator in the lobby of a large office building between the hours of 8 A.M. and 9 A.M.
- What is the probability that in any one-minute period at most four persons are waiting?
 - What is the *approximate* probability that in any one-minute period at most four persons are waiting?
 - Compare your results in (a) and (b).
- 6.60 The number of cars arriving per minute at a toll booth on a particular bridge is Poisson distributed with a mean of 2.5. What is the probability that in any given minute
- no cars arrive?
 - not more than two cars arrive?
- If the expected number of cars arriving at the toll booth per ten-minute interval is 25.0, what is the *approximate* probability that in any given ten-minute period
- not more than 20 cars arrive?
 - between 20 and 30 cars arrive?
- 6.61 Cars arrive at Kenny's Car Wash at a rate of nine per half-hour.
- What is the probability that in any given half-hour period at least three cars arrive?
 - What is the *approximate* probability that in any given half-hour period at least three cars arrive?
 - Compare your results in (a) and (b).
- 6.62 Suppose that the number of defective videocassette tapes that are returned to a video rental store has averaged seven per day.
- What is the (*exact*) probability that two tapes will be returned today?
 - What is the (*exact*) probability that at least two tapes will be returned today?
 - What assumptions were made about the probability distribution selected in (a) and (b)? Discuss.
 - Obtain *approximate* answers for (a) and (b) using a different probability distribution model. Discuss the differences in your findings.