

5.4 Normal Approximation of the Binomial Distribution

You have seen that probabilities can be calculated using the formula for a binomial distribution provided that

- each trial in the experiment is identical
- the outcomes of each trial can be classified as a success or failure

For example, you can calculate the probability of tossing a coin many times and finding a certain number of heads.

Example 1 Number of Heads

What is the probability of getting exactly 30 heads if a coin is tossed 50 times?

Solution

To calculate the probability, remember that $P(k \text{ successes in } n \text{ trials})$ is $\binom{n}{k}(p)^k(1-p)^{n-k}$, where p is the probability of success.

$$\begin{aligned}P(30 \text{ heads in } 50 \text{ trials}) &= \binom{50}{30}(0.5)^{30}(1-0.5)^{50-30} \\ &\doteq 0.042\end{aligned}$$

There is about a 4.2% chance of tossing exactly 30 heads in 50 tosses.

While the calculation in this activity involves very large or very small numbers, it is not difficult or time-consuming to perform on a calculator.

The real problem occurs with more complex situations; for example, finding the probability of tossing between 20 and 30 heads in 50 tosses of a coin. You would be faced with 11 such calculations.

$$P(20 \leq X \leq 30) = P(20 \text{ heads in } 50 \text{ trials}) + P(21 \text{ heads in } 50 \text{ trials}) + P(22 \text{ heads in } 50 \text{ trials}) + \dots + P(30 \text{ heads in } 50 \text{ trials})$$

This calculation would be time-consuming even though these numbers are not very large. To help simplify the calculation and make these probabilities easier to calculate, look at the graphical representation of the binomial distribution for this situation on the following page.

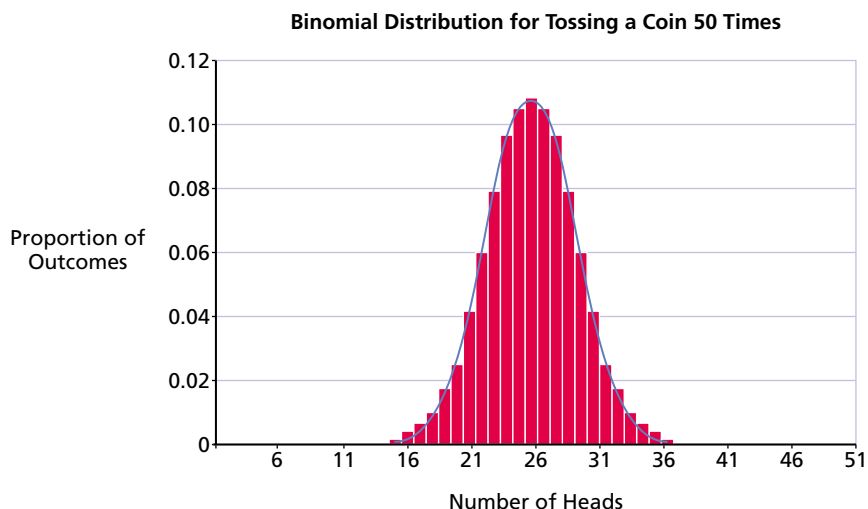
GRAPHING THE BINOMIAL DISTRIBUTION

The graph of the binomial distribution where $n = 50$ and $p = 0.5$ is shown on the following page. Notice that it approximates a normal distribution. This suggests that a binomial distribution can be approximated by a normal distribution as long as the number of trials is relatively large.



Think about The Graph

Describe the features of a normal probability distribution that are displayed by this binomial distribution.



Notice how this binomial probability distribution appears mound-shaped. A normal curve (shown in blue) can be superimposed over the histogram with a pretty close fit. French mathematician Abraham De Moivre first discovered this relationship in 1718. Under certain conditions, a normal curve can be used to approximate a binomial probability distribution. First, we must find values for the mean and the standard deviation.

MEAN AND STANDARD DEVIATION OF THE NORMAL APPROXIMATION TO THE BINOMIAL DISTRIBUTION

If you were to toss a coin 50 times, you would probably expect to get approximately 25 heads since $n = 50$ and $P(\text{heads}) = 0.5$. It is calculated as $\bar{x} = np$, where n represents the number of trials and p represents the probability of success on each trial. In any binomial distribution, the standard deviation of the binomial random variable can be approximated by using the formula.

$$\begin{aligned}
 \sigma &= \sqrt{np(1 - p)} \\
 &= \sqrt{50(0.5)(1 - 0.5)} \\
 &= \sqrt{12.5} \\
 &= 3.54
 \end{aligned}$$

Using these formulas, we can determine that the normal curve shown above has a mean of 25 and a standard deviation of 3.54.

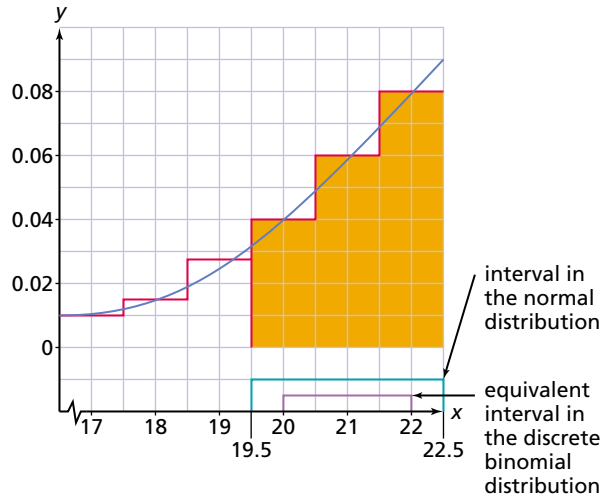
FROM DISCRETE TO CONTINUOUS

The important difference between the binomial and the normal distribution is that the binomial distribution represents a discrete random variable, which is always represented with a whole number (e.g., number of heads).

A normal distribution is continuous, and can be used to display continuous values (e.g., height in centimetres). In order to use the normal distribution to approximate the binomial distribution, consider a range of values rather than

specific discrete values. The range of continuous values between 4.5 and 5.5 can, therefore, be used to represent the discrete value 5.

In order to use the normal approximation of $X = 20$, we need to consider all the values of X that round to 20. On the graph below, we can see that the bar centred at $X = 20$ contains all the values from 19.5 to 20.5. The area of this bar is approximated by the area under the normal curve between $X = 19.5$ and $X = 20.5$. Similarly, if we wanted to estimate the probability of X being 20, 21, or 22, in the binomial distribution, we would evaluate $P(19.5 < X < 22.5)$ for the normal distribution.



Example 2 Approximating Probability Using the Normal Distribution

Fran tosses a fair coin 50 times. Estimate the probability that she will get tails less than 20 times.

Solution

Let a success be a toss of tails. Thus, $n = 50$ and $p = 0.5$.

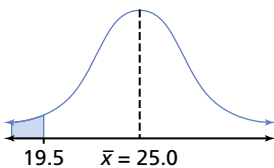
$$\begin{aligned}\bar{x} &= 50(0.5) \\ &= 25\end{aligned}$$

$$\begin{aligned}\sigma &= \sqrt{50(0.5)(1 - 0.5)} \\ &= \sqrt{12.5} \\ &= 3.54\end{aligned}$$

You want to evaluate $P(X < 20)$, where the data are normally distributed with $N(25, 3.54^2)$. Using z-scores, this distribution can be standardized to $N(0, 1)$. Continuous values from 0 to 19.5 will be considered less than 20. Calculating your z-score for 19.5 yields

$$\begin{aligned}z &= \frac{19.5 - 25}{3.54} \\ &= -1.55\end{aligned}$$

$$\begin{aligned}\text{Therefore, } P(X < 19.5) &= P(z < -1.55) \\ &= 0.0606\end{aligned}$$



There is a 6% chance that Fran will toss less than 20 tails in 50 attempts.

Not all binomial distributions can be approximated using the normal distribution. As you will recall from Chapter 3, some distributions are left-skewed or right-skewed and do not fit within a normal distribution. Use the following rule when considering whether or not a binomial distribution is “symmetrical enough” to be approximated by the normal distribution.

If X is a binomial random variable of n independent trials, each with probability of success p , and if

$$np > 5 \quad \text{and} \quad n(1 - p) > 5$$

then the binomial random variable can be approximated by a normal distribution with $\bar{x} = np$ and $\sigma = \sqrt{np(1 - p)}$.

Example 3 Testing Binomial Distributions for Symmetry

Check each of the following binomial distributions to see if they can be approximated by a normal distribution.

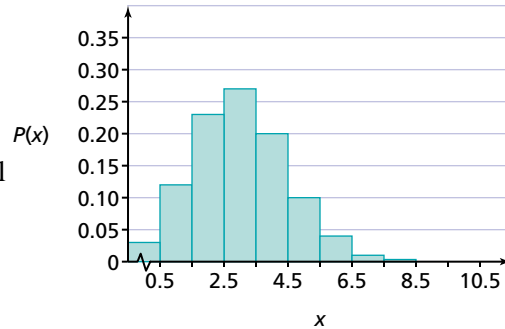
(a) $n = 10, p = 0.3$

(b) $n = 14, p = 0.5$

Solution

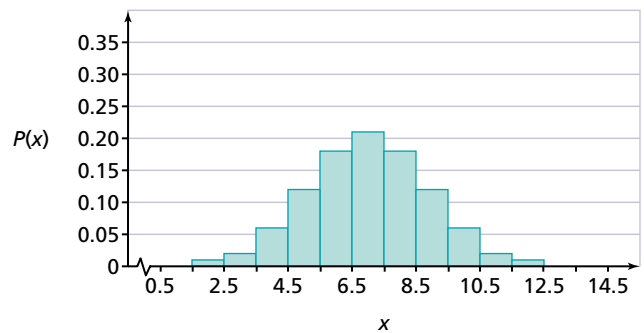
(a) $np = 10(0.3)$
 $= 3.0$
 $nq = 10(0.7)$
 $= 7.0$

Here, $np < 5$. A normal approximation would not be accurate.



(b) $np = 15(0.5)$
 $= 7.5$
 $nq = 15(0.5)$
 $= 7.5$

Here, both np and nq are greater than 5. A normal approximation would be valid.



Example 4 Approximating Probabilities in a Range

Calculate the probability that, in 100 rolls of a fair die, a 6 appears between 10 and 20 times, inclusive.

Solution 1 No technology required

Let a success be rolling a 6. Then, $n = 100$ and $p = \frac{1}{6}$.

First, check to see if the normal approximation to the binomial distribution can be used.

$$\begin{aligned} np &= 100\left(\frac{1}{6}\right) & n(1-p) &= 100\left(1 - \frac{1}{6}\right) \\ &\doteq 16.67 & &\doteq 83.33 \end{aligned}$$

Since $np > 5$ and $n(1-p) > 5$, the binomial distribution can be approximated by the normal curve. Find the mean and the standard deviation for the normal approximation.

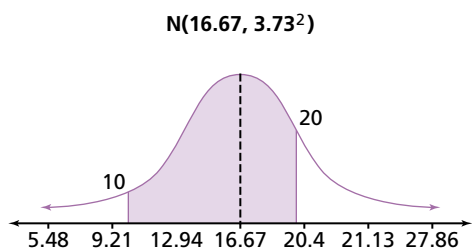
$$\begin{aligned} \bar{x} &= np \\ &= 100\left(\frac{1}{6}\right) \\ &\doteq 16.67 \end{aligned} \qquad \begin{aligned} \sigma &= \sqrt{np(1-p)} \\ &= \sqrt{100\left(\frac{1}{6}\right)\left(1 - \frac{1}{6}\right)} \\ &\doteq 3.73 \end{aligned}$$

We want to determine $P(10 \leq X \leq 20)$, where the data are normally distributed with $N(16.67, 3.73^2)$. This can be standardized to $N(0, 1)$. Values between 9.5 and 20.5 will be rounded to discrete values between 10 and 20, inclusive. Calculating z-scores for 9.5 and 20.5 yields the following:

$$\begin{aligned} z &= \frac{x - \bar{x}}{\sigma} & z &= \frac{x - \bar{x}}{\sigma} \\ &= \frac{9.5 - 16.67}{3.73} & &= \frac{20.5 - 16.67}{3.73} \\ &= -1.92 & &= 1.03 \end{aligned}$$

Therefore, $P(10 \leq X \leq 20) = P(-1.92 \leq z \leq 1.03)$.

From the tables, the area to the left of $z = -1.92$ is 0.0274. The area to the left of 1.03 is 0.8238. The area between the z-scores is $0.8238 - 0.0274 = 0.7964$. Therefore, the probability of rolling between 10 and 20 sixes on a fair die rolled 100 times is almost 80%.



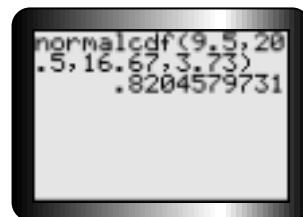
Technolink

For more information on using the **normalcdf**(command on a TI-83 Plus calculator, see Appendix C.8 on page 406.

Solution 2 Using a TI-83 Plus calculator

A TI-83 Plus calculator can be used to calculate the area under the normal distribution. The command for this purpose is **normalcdf**((normal cumulative density function). To select this function, press **[2nd]** **[DISTR]** **[2]**. You will then enter the lower X value, the upper X value, the mean, and the standard deviation, each separated by a comma.

The calculator will return the area between the lower and upper values. Enter the mean and standard deviation as shown in the calculator screen above.



The result is slightly different because the calculator computes z -scores exactly, and doesn't round them to two decimal places.

Example 5 Quality-Control Decisions

A hamburger-patty producer claims that its burgers contain 400 g of beef. It has been determined that 85% of burgers contain 400 g or more. An inspector will only accept a shipment if at least 90% of a sample of 250 burgers contain more than 400 g. What is the probability that a shipment is accepted?

Solution

The testing of an individual burger may be treated as a Bernoulli trial with a probability of success of $p = 0.85$. The inspector performs $n = 250$ independent trials.

Check to see if the normal approximation to the binomial distribution can be used.

$$\begin{aligned} np &= 250(0.85) & n(1 - p) &= 250(1 - 0.85) \\ &= 212.5 & &= 37.5 \end{aligned}$$

Since $np > 5$ and $n(1 - p) > 5$, the binomial distribution can be approximated by the normal curve.

Find the mean and the standard deviation for the normal approximation.

$$\begin{aligned} \bar{x} &= 250(0.85) & \sigma &= \sqrt{np(1 - p)} \\ &= 212.5 & &= \sqrt{250(0.85)(0.15)} \\ & & &\doteq 5.646 \end{aligned}$$

We want to determine the probability that at least 90% of the 250 samples are acceptable. Given that 90% of 250 is 225, we need to determine $P(X > 224.5)$.

The corresponding z -score is

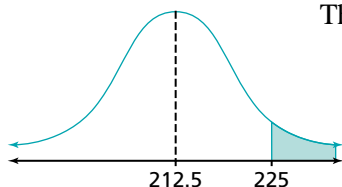
$$\begin{aligned} z &= \frac{x - \bar{x}}{\sigma} \\ &= \frac{224.5 - 212.5}{5.646} \\ &= 2.12 \end{aligned}$$

$$\begin{aligned} \text{Therefore, } P(X > 224.5) &= P(z > 2.12) \\ &= 1 - P(z < 2.12) \\ &= 1 - 0.9830 \\ &= 0.017 \end{aligned}$$

There is only a 1.7% chance that the shipment will be accepted.

? Think About The Discrete Random Variable

The random variable X in this example is the number of hamburgers accepted, and not the size of the hamburgers. Why?



KEY IDEAS

- A **normal distribution** may be used to approximate a binomial distribution. This is useful if the number of binomial calculations is large.
- The approximation is only valid if both np and $n(1 - p)$ are greater than 5, where n is the number of independent trials, p is the probability of success in a single trial, and $(1 - p)$ is the probability of failure.
- The normal parameters \bar{x} and σ are calculated as follows:

$$\bar{x} = np \text{ and } \sigma = \sqrt{np(1 - p)}$$

- The normal approximation of the binomial probability distribution may be calculated on a TI-83 Plus calculator using the **normcdf**(command. See Appendix C.8 on page 406 for more details.

5.4 Exercises

- A**
1. A coin is tossed four times. Find the probability of each of the following.
(a) $P(\text{exactly 1 head})$ (b) $P(\text{exactly 2 heads})$
(c) $P(\text{exactly 3 heads})$ (d) $P(\text{exactly 4 heads})$
(e) $P(\text{exactly 0 heads})$
 2. A die is rolled 10 times. Calculate the probability of each of the following events, correct to six decimal places.
(a) one 1 appears (b) two 2s appear (c) three 3s appear
(d) four 4s appear (e) five 5s appear (f) six 6s appear
 3. A bag contains 20 yellow marbles and 80 blue marbles. The bag is shaken so that the marbles are thoroughly mixed. Marbles are then removed from the bag one at a time, the colour is recorded, and the marble is returned to the bag. This process is done a total of four times. Find the probability of each of the following.
(a) $P(\text{exactly three blue marbles are drawn})$
(b) $P(\text{at least three blue marbles are drawn})$
 4. A student guesses all 10 answers on a multiple-choice test. There are 5 choices for each of the questions. Find the probability (correct to 6 decimal places) that the student scores exactly 50% on the test (gets 5 correct answers).

- B**
5. **Knowledge and Understanding** If 20 coins are dropped on a table, what is the probability (correct to three decimal places) that
 - (a) exactly 12 coins are heads
 - (b) at least 12 coins are heads
 6. A card is drawn from a standard deck and replaced. If this experiment is repeated 30 times, what is the probability that
 - (a) exactly 10 of the cards are spades
 - (b) no more than 5 cards are spades
 7. It is known that approximately 90% of the population is right-handed. In a sample of 100 people, what is the probability that
 - (a) exactly 10 people are left-handed
 - (b) more than 10 people are left-handed
 - (c) less than 5 people are left-handed
 8. **Application** A drug has a 70% success rate. What is the probability that 80 or more people out of 100 will be cured by the drug?
 9. **Communication** A pair of six-sided dice is rolled 50 times. Which is more likely?
 - (a) exactly 8 doubles are rolled
 - (b) 10 or more doubles are rolled
 10. A test consists of 50 questions. What is the probability that you can
 - (a) pass
 - (b) get more than 40% correct if
 - (i) it is a true/false test
 - (ii) it is a multiple-choice test with four possible choices
 11. The probability of engine trouble on a jet airplane is relatively small. Suppose a jet has a $\frac{1}{30}$ chance of having at least one engine fail on any flight. (The jet can fly with the other engines working normally.) The jet has just flown its fiftieth flight. Find the probability that the plane has experienced engine trouble on at least two flights.
 12. Singh Textiles produces computer chips. On average, 2% of all computer chips produced are defective. In a sample of 500 chips, the quality-control inspector accepts the batch only if fewer than 1% of the chips tested are defective. Determine the probability that a batch is accepted.
- C**
13. **Thinking, Inquiry, Problem Solving** An airline has determined that 4% of people do not show up for their flights. To avoid having empty seats, the flight is overbooked. A large jet holds 300 people.
 - (a) What is the probability that some travellers will not get a seat if 310 tickets are sold? 305 tickets?
 - (b) How many tickets could be oversold in order to be 98% sure that everyone gets a seat?

14. Xiau and Kim interviewed 175 people and found that 75 were in favour of raising funds for a new arena by increasing property taxes. The question will be decided in a referendum in which it is expected that all of the 26 076 eligible voters will vote. Over half the votes must be in favour of raising property taxes for the new arena to be built. How likely is it that there will be a new arena?
15. An election is being held at your school to determine who will be treasurer of the student council. Teresa and Elisabete are the candidates. The day before the election, Elisabete takes a poll of 74 students and finds that 30 will vote for her. There are 1148 students in the school and all are expected to vote in the election. How likely is it that Elisabete will win?

ADDITIONAL ACHIEVEMENT CHART QUESTIONS

16. **Knowledge and Understanding** A student guesses all 15 answers on a multiple-choice test. There are 5 choices for each of the questions. Find the probability (correct to 6 decimal places) that the student passes the test.
17. **Applications** On average, Mike Weir scores a birdie on about 20.9% of all the holes he plays. Mike is in contention to win a PGA golf tournament but he must birdie at least four of the last six holes he plays. Find the probability, as a percent correct to one decimal place, that Mike will win.
18. **Thinking, Inquiry, Problem Solving** An insurance company has said that “the probability of your having an accident while travelling on a stretch of road at night is $\frac{1}{500}$.” Find the probability that you will have at least one accident on the stretch of road if you travel the road 350 times.
19. **Communication** As a shoe salesperson paid on commission, it is important for Devica to close a sale. She knows that the probability of closing a sale with any one customer who tries on shoes is about 26%. She predicts that she will help about 220 customers try on shoes each month. She says she can expect to sell shoes to between 40 and 90 customers each month. Should you believe this claim?