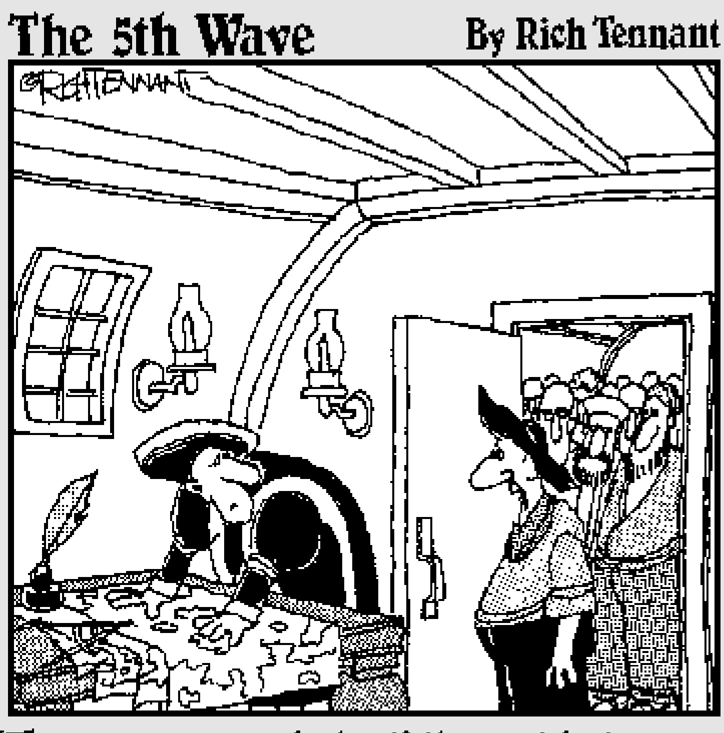
Part III

From A to Binomial: Basic Probability Models



"The crew was wondering if there might be some sort of probability model we could run on whether the world was round instead of sailing up to the edge and hoping we don't fall off screaming into an endless black hole."

In this part . . .

Part III is where you get into all the nuts and bolts of probability distributions. You pick up the tools you need to understand expected values and variances and find probabilities for some basic random variables. You also find out how people use probability in statistics to make decisions and to assess the chance of making the wrong decision. If you ask me, odds are you'll have fun!

Chapter 7

Probability Distribution Basics

In This Chapter

- Discovering what a probability distribution is
- Calculating probabilities with probability distributions
- Determining and interpreting expected value, variance, and standard deviation
- Practicing probability distribution basics on the discrete uniform distribution

Sometimes, it's necessary to move away from individual probability scenarios and the events and calculations associated with them and start to look at situations in which probabilities follow a certain predictable pattern that you can use a probability model to describe. For example, one probability model helps you figure the probability that a phone call will last more than 10 minutes; another model helps you determine the average number of times you need to play a certain lottery game before you win.

Formally, a *probability model* gives you formulas to calculate probabilities, determine long-term average outcomes, and figure the amount of variability you can expect in the results from one random experiment to the next. Many different probability models exist for different types of situations. I discuss the most common ones in this book.

In this chapter, you develop the basic ideas of a probability model, and you apply those ideas to the most basic of probability models, the discrete uniform distribution.

The Probability Distribution of a Discrete Random Variable

A *probability model* is a mathematical model you use to fit a random process. Each random process must meet a different set of assumptions for each different model so the outcomes have the correct probabilities as determined by the model and so the model can be tested by using actual data to find out whether it correctly fits the data. The fundamental parts of a probability model are the random variable and its probability distribution.

Defining a random variable

A *random variable* is a function from the sample space, S, to the set of all possible probabilities (the interval [0, 1]), denoted by capital letters like X, Y, and so on. Although the sample space for an experiment never changes after it's been determined, you can have many different random variables affiliated with it. For example, suppose you roll two dice. The sample space S is made up of 36 potential outcomes: {(1, 1), (1, 2), (1, 3), . . ., (6, 4), (6, 5), (6, 6)}. One random variable of interest represents the sums of the potential outcomes on the dice; call it X. Table 7-1 shows all the potential outcomes of S and the corresponding values of the random variable X.

Table 7-1 Random Variable X = Sum of the Outcomes on Two Dice

Outcome	X Value	Outcome	X Value	Outcome	X Value	Outcome	X Value	Outcome	X Value	Outcome	X Value
(1, 1)	2	(2, 1)	3	(3, 1)	4	(4, 1)	5	(5, 1)	6	(6, 1)	7
(1, 2)	3	(2, 2)	4	(3, 2)	5	(4, 2)	6	(5, 2)	7	(6, 2)	8
(1, 3)	4	(2, 3)	5	(3, 3)	6	(4, 3)	7	(5, 3)	8	(6, 3)	9
(1, 4)	5	(2, 4)	6	(3, 4)	7	(4, 4)	8	(5, 4)	9	(6, 4)	10
(1, 5)	6	(2, 5)	7	(3, 5)	8	(4, 5)	9	(5, 5)	10	(6, 5)	11
(1, 6)	7	(2, 6)	8	(3, 6)	9	(4, 6)	10	(5, 6)	11	(6, 6)	12

The two major types of random variables are

- Discrete
- Continuous

Discrete random variables have either a finite or countably infinite number of possible values. For example, X may be the total number times a coin turns up heads when you flip it 1,000 times, or Y may be the total number of accidents at a certain intersection in a given year. In the former case, X is a discrete random variable, because it has a finite number of possible values: 0, 1, 2, . . . , 1,000. In the latter case, Y has a countably infinite number of possible values: 0, 1, 2, . . . , 100, and so on. The reason Y is countably infinite is because although you can't have an infinite number of accidents at an intersection in a given year, you also can't possibly define an upper limit for the value of Y, so you define Y as being countably infinite and assign progressively smaller and smaller probabilities for extremely large values of Y.

A *continuous random variable* has an uncountably infinite number of possible values — so many that you can't begin to have a system for enumerating them all. For example, X may be the length of time of a phone call, which technically can be measured to a millionth of a second. Even if you measure the call only in minutes and seconds, so many possible values for X exist that you may as well call it an uncountable random variable, which is what statisticians often do. As another example, Y may be the average test score for a class of students taking a national achievement exam. For all intents and purposes, Y takes on an uncountably infinite number of possible values and is therefore classified as a continuous random variable.

I discuss probability basics mostly in terms of discrete random variables throughout this chapter. When you get to continuous probability models in Part V of this book, I provide a more in-depth explanation of continuous random variables and their probabilities.

Finding and using the probability distribution

The *probability mass function (pmf)* for a discrete random variable X is a function that assigns probabilities to each value of X. The notation used to denote the probability mass function for X is P(x). (Note that if X is continuous, the counterpart of this function is called the probability density function for X (pdf), and is denoted f(x). More about that in Chapter 9.)

For the dice-rolling example you see in Table 7-1, the possible values of X, the sum of two dice, are 2, 3, 4, . . . , 12. You find the probability that X = 2 by going back to the sample space and finding the probabilities of all the outcomes assigned the value of X = 2. In this case, the only outcome with an X value of 2 is (1, 1), which has a probability of $\frac{1}{36}$ (stated as "one thirty-sixth"). Table 7-2 shows the probabilities for the values of X. Of all the possible outcomes (36), only one (1, 1) gives you a value of 2.

TIP

In probability notation, you have $P(X = 2) = \frac{1}{36}$ (stated as "P of X equals 2 is one thirty-sixth"). Another way to write this probability is to write $P(2) = \frac{1}{36}$ (stated "P of 2 is one thirty-sixth"). I use both types of notation throughout the book.

Similarly, the probability that X = 3 is $P\{(2, 1) \text{ or } (1, 2)\} = P(2, 1) + P(1, 2) = \frac{1}{36} + \frac{1}{36} = \frac{2}{36}$. So you have $P(3) = \frac{2}{36}$. The chance that the sum is equal to 3 is higher than the chance that the sum is equal to 2, because you can get a sum of 3 more ways than you can get a sum of 2. Table 7-2 shows the probability mass function (pmf) of X. It includes all possible values of X and their associated probabilities.

Table 7-2 Probability Mass Function of X = Sum of Two Dice

X	*P(x)*
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

A *probability distribution* for X is a listing of all possible values of X, along with their probabilities. The following are the properties of the probability distribution for a discrete random variable X:

- P(x) is between zero and one for any value of the random variable X.
- To find the probability that takes on values a or b, add P(a) + P(b), because the individual values of X are mutually exclusive.
- Because every element in the sample space is assigned a value of X, all the probabilities in the probability distribution for X must add up to one.

TIP

After you set up the probability distribution for a random variable, you can check whether the probabilities all add up to one. If they do, you can be reasonably sure that the probability distribution is correct. If they don't, you know something is wrong. By the way, for those keeping score, the probabilities do equal one in Table 7-2.

TECHNICAL STUFF

The function that assigns probability for a discrete random variable is called a *probability mass function,* because it shows how much probability, or *mass,* is given to each value of the random variables. Mass is thought of as weight in this case; the total mass (or weight) for a probability distribution equals one. A continuous random variable doesn't actually assign probability or mass, it assigns density, which means it tells you how dense the probability is around x for any value of x. You find probabilities for intervals of X, not particular values of X, when X is continuous. (Continuous random variables have no probability at any single point because there is no area over a single point. More on this in Chapter 9.)

Picturing probability distributions: Plotting on a histogram

You can draw a picture of the probability distribution of a discrete random variable X by using a graph called a relative frequency histogram. A *relative frequency histogram* is basically a bar graph that displays numerical values on the X-axis and the percentage of time each value occurs along the Y-axis. For more information on histograms, you can check out *Statistics For Dummies* (Wiley), also written by yours truly. The probability distribution for the dice-rolling example from Tables 7-1 and 7-2 is pictured in Figure 7-1.

REMEMBER

Notice that Figure 7-1 shows the probability for each individual value of X as the height of the bar, and that the bars are connected, even though you can't have any probability between particular values of X in this case (such as 2.5 or 3.2). This, however, is the most common way that frequency and relative frequency histograms are drawn, whether the possible values of X are numbers like 1, 2, 3 or all real numbers on a number line. Always pay attention to the context of the problem so you know what type of values you're dealing with.

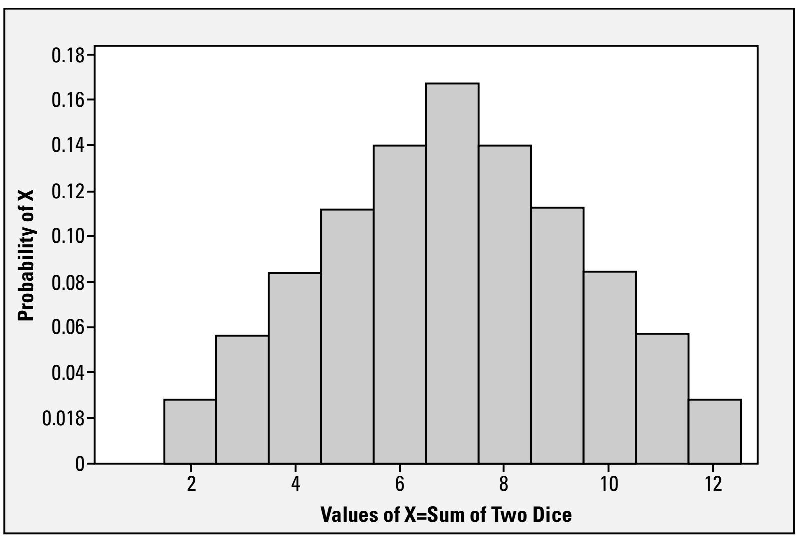


Figure 7-1: This relative frequency histogram shows the probability distribution for X = sum of two dice.

One identifying factor of a probability distribution is the shape it takes when plotted on a histogram. Some of the many different shapes for probability distributions are shown in Figure 7-2. Figure 7-2a has a mound shape, with the values of X in the middle having higher probability than values on the outer edges. Figure 7-2b has the shape of a discrete uniform, which is flat (see the section "The Discrete Uniform Distribution"). The flat shape indicates that each value of X has equal probability. Figure 7-2d is skewed to the right, indicating that as you move left to right, the probabilities decrease for higher and higher values of X. Figure 7-2c is skewed to the left, indicating that as you move left to right, the probabilities increase for higher and higher values of X.

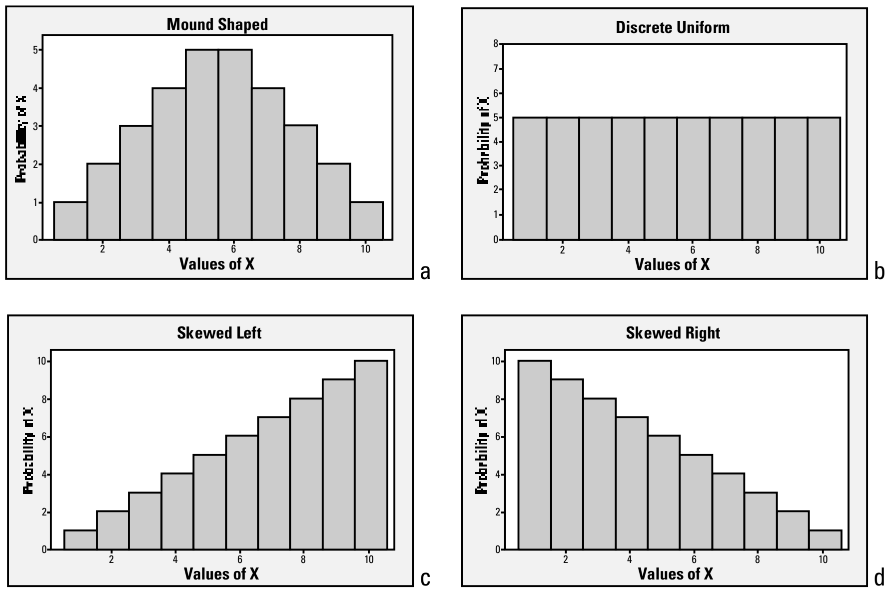


Figure 7-2: Shapes of various probability distributions.

The probability distribution for X for the dice-rolling example (see Figure 7-1) is mound-shaped (because it has a mound in the middle) and symmetric (because the left and right sides are mirror images of each other). The sum of 7 has the highest probability (which is why the number is so important in the game of craps; for more on gambling scenarios, see Chapter 6). As you move away from 7 on either side, the probabilities decrease until you reach the two values of X with the lowest probabilities: X = 2 and X = 12. These sums are the hardest to achieve because they have the fewest number of potential outcomes associated with them.

Calculating probabilities: "At most," "at least," and more

After you set up a probability distribution, you're often called upon to use it for calculating various probabilities. With the dice-rolling example from Tables 7-1 and 7-2, you may be asked to calculate the probability that the sum of the two dice is

At least 7

Less than 7

At most 10

More than 10

These situations represent events. An *event* is a collection of one or more possible outcomes of interest from the sample space, S.

The event "at least 7" means 7 is the lowest possible value, and it goes up from there to the highest possible value. In this example, X goes from 7 through 12 (including both 7 and 12), so you add the probabilities of those values you see in Table 7-2. In probability notation, you write

$P(7 \leq X \leq 12) = P(X = 7) + P(X = 8) + P(X = 9) + P(X = 10) + P(X = 11) + P(X = 12) = \frac{6}{36} + \frac{5}{36} + \frac{4}{36} + \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{21}{36} = 0.58$, or 58 percent.

The event "less than 7" sounds much like "at least 7," but in fact it's different. Less than 7 means everything below 7, but it doesn't include 7. In this example, X goes from 2 up to but not including 7. Perhaps a more practical way to state it is that X goes from 2 through 6 (including both 2 and 6). Again, you add the probabilities of those values you see in Table 7-2. In probability notation, you write

$P(X < 7) = P(X \leq 6) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \frac{5}{36} = \frac{15}{36} = 0.42$, or 42 percent.

TIP

Note that the events {X is at least 7} and {X is less than 7} are complements of each other, which means that they break the sample space into two separate sets that don't contain any elements in common. Because they're complements, the probabilities of the two events sum to one (see Chapter 2 for more on complements). If you've already calculated $P(7 \leq X \leq 12)$, you can just take $1 - P(7 \leq X \leq 12)$ to find $P(X < 7)$. Note that 12 is the highest possible value of X.

The event "at most 10" means the highest possible value is 10 and includes values down to the lowest possible value. In this example, X goes from 2 through 10 (including both 2 and 10), so you add the probabilities of the values you see in Table 7-2. In probability notation, you write

$P(2 \leq X \leq 10) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + \ldots + P(X = 9) + P(X = 10) = \frac{1}{36} + \frac{2}{36} + \frac{3}{36} + \frac{4}{36} + \ldots + \frac{2}{36} + \frac{1}{36} = 0.92$, or 92 percent.

Notice that you have many probabilities to find and add up in this case. What a drag! In fact, the problem presents far fewer items that you *don't* need to find. This should make you think of the complement rule (see Chapter 2), which says that the probability of the complement of an event is one minus the probability of the event. Using the complement rule, you have

$P(2 \leq X \leq 10) = 1 - P\{11 \leq X \leq 12\} = 1 - (\frac{2}{36} + \frac{1}{36}) = 1 - \frac{3}{36} = 0.92$, or 92 percent.

The event "more than 10" sounds very much like "at most 10," but in fact it's different. More than 10 means everything above but not including 10 (in other words, just the two values 11 and 12). You add the probabilities of these values found in Table 7-2. In probability notation, you write

$P(X > 10) = P(X \geq 11) = P(X = 11) + P(X = 12) = \frac{2}{36} + \frac{1}{36} = \frac{3}{36} = 0.08$, or 8 percent.

Again, note that the events {X is at most 10} and {X is more than 10} are complements of each other, so their probabilities sum to one.

TIP

The terms "at least" and "less than" are easily confused, but they mean different things. You can use a simple number line to help you visualize which values are included in each respective phrase. Another way you can keep them straight is to remember helpful phrases, such as "You must be at least 21 to drink alcohol in this bar." That means you have to be age 21 or older. The terms "at most" and "more than" also have totally different meanings. Thinking about money may help you to distinguish between them. For example, "I can pay at most $10 for lunch" means you can eat for $10 or less — in other words, you have $10. Or, if you want to make more than $40,000 per year, you want to make $40,001 or more.

Finding and Using the Cumulative Distribution Function (cdf)

In the previous section, you add many probabilities again and again to find probabilities for X being "at least," "less than," "at most," or "more than." If you had to do any more adding, you'd start thinking there has to be a better way; in fact, a better way does exist. You can use the *cumulative distribution function* of X, which is a function that represents the probability that X is less than or equal to any value a and is equal to the sum of all the probabilities for X that are less than or equal to a. In probability notation, the cumulative distribution function (cdf) of X, is written as $F(a) = \sum_{x \leq a} P(X = x)$.

In the dice-rolling example from Tables 7-1 and 7-2, the probability that X is less than or equal to 6 is $P(X \leq 6) = P(X = 2) + P(X = 3) + P(X = 4) + P(X = 5) + P(X = 6) = \frac{15}{36} = 0.42$, or 42 percent. Using the cumulative distribution function of X, this is equivalent to saying $F(6) = 0.42$. Table 7-3 shows the complete cdf of X for the dice-rolling example.

Table 7-3 Cumulative Distribution Function for X = Sum of Two Dice

X	*F(x)*
$X < 2$	0 = 0.00, or 0%
$2 \leq X < 3$	$\frac{1}{36}$ = 0.028, or 2.9%
$3 \leq X < 4$	$\frac{1}{36} + \frac{2}{36} = \frac{3}{36}$ = 0.083, or 8.3%

X	*F(x)*
$4 \le X < 5$	$\frac{1}{36} + \frac{2}{36} + \frac{3}{36} = \frac{6}{36} = 0.167$, or 16.7%
$5 \le X < 6$	$\frac{10}{36} = 0.277$, or 27.7%
$6 \le X < 7$	$\frac{15}{36} = 0.417$, or 41.7%
$7 \le X < 8$	$\frac{21}{36} = 0.583$, or 58.3%
$8 \le X < 9$	$\frac{26}{36} = 0.722$, or 72.2%
$9 \le X < 10$	$\frac{30}{36} = 0.833$, or 83.3%
$10 \le X < 11$	$\frac{33}{36} = 0.917$, or 91.7%
$11 \le X < 12$	$\frac{35}{36} = 0.972$, or 97.2%
$X \ge 12$	$\frac{36}{36} = 1.00$, or 100%

Note: Every possible X value from negative infinity to positive infinity has a value of F(x) associated with it because F(x) is defined as the probability that X is less than or equal to x, for any value of x.

Interpreting the cdf

Because the cumulative distribution function is defined on all values from negative infinity to positive infinity, you have to switch gears a bit to be able to interpret a cdf compared to a probability mass function (pmf; see Table 7-2 for an example).

In the dice-rolling example from Tables 7-1 and 7-2, for values less than 2, no probability exists to accumulate, so $F(x) = 0$. At the point where $X = 2$, the function jumps up to $\frac{1}{36}$, because the probability of $X = 2$ is $\frac{1}{36}$. Now for all the values from 2 up to but not including 3, the cdf stays at $\frac{1}{36}$, because no new probability has been accumulated. For example, suppose you look at the number $X = 2.50$. Although it isn't one of the values of the random variable X, you still can find $P(X \le 2.50) = F(2.50)$. The probability that X is less than or equal to 2.50, in this case, is equal to the probability that X is equal to 2, which is $\frac{1}{36}$, or 2.8 percent.

In fact, the probability that X is less than or equal to 2.60 also is $\frac{1}{36}$, and the cumulative probability remains $\frac{1}{36}$ until you get to the number $X = 3$, where it jumps up: $F(3) = P(X \le 3) = P(X = 2) + P(X = 3) = \frac{1}{36} + \frac{2}{36} = \frac{3}{36} = 8.33$ percent. Thus, for any value from 3 up to but not including 4, the total amount of accumulated probability is $\frac{1}{36} + \frac{2}{36} = \frac{3}{36}$. The jumps continue as you pass X = 4, 5, 6, . . . , up to 11 and 12, because at each of those values, more probability is accumulated, but between those values, nothing new is happening, so the cdf remains stable.

Graphing the cdf

For any number, you can find the probability that X is less than or equal to that number by looking at the graph of the cumulative distribution function. Figure 7-3 depicts the cumulative distribution function for X in graph form, using the dice example from Tables 7-1 and 7-2. Notice that in this case, the possible values of X are integers (whole numbers that can be positive, negative, or zero). At the integers 2, 3, 4, . . . , 12, you can see what appears to be two possible values of F(x), one shown by a filled-in (solid) dot and one shown by an open dot (circle). The filled-in dot is the value of F(x) at that point.

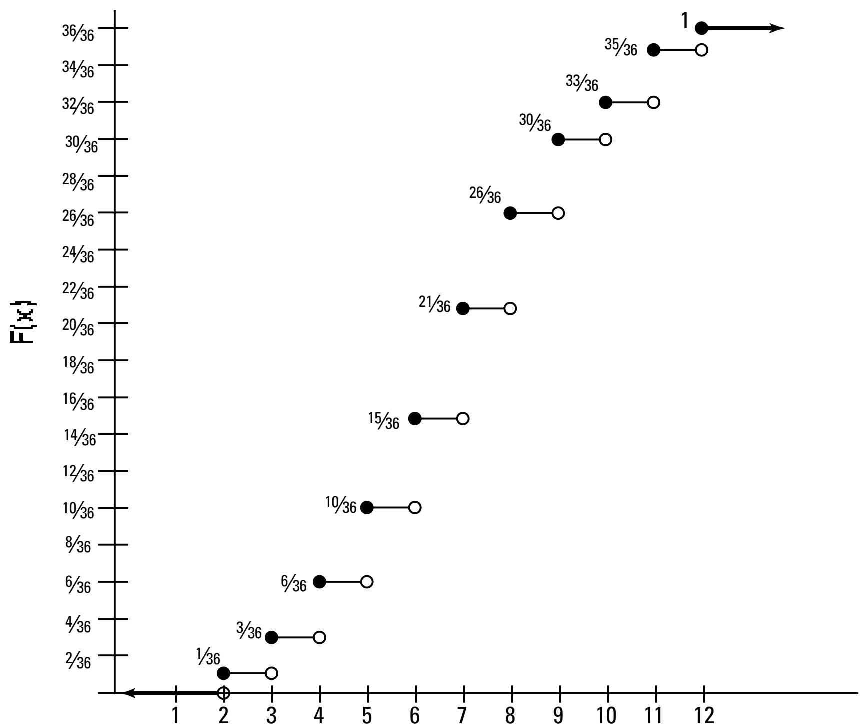


Figure 7-3: Graph of the cumulative distribution function for F(sum of two dice).

Notice that as the numbers for X grow larger, the respective values of F(x) increase in a graduated or stepwise fashion until they reach one, where they then stay for all values beyond 12. After you hit X = 12, you've accumulated all the probability, and for any values greater than 12, the cdf remains 1. For any cumulative distribution function, the general shape of the graph is the same. The values of F(x) always start at zero from negative infinity to the first value of X where probability occurs, and they always increase in value until they reach F(x) = 1 at the point where the last amount of probability for X occurs. Thereafter, F(x) stays at 1 for values of X into infinity.

Discrete random variables have cdfs known as *stepwise functions.* They're not continuous everywhere; they take on a fixed value for a certain interval and then jump up at the next value that has some probability and stay there until the next value with probability shows up. The function gets its name because the graph appears to have steps. Continuous random variables have a cdf that's continuous. I discuss continuous cdfs in Part V of this book.

Finding probabilities with the cdf

You can use the cumulative distribution function (cdf) to find probabilities for X. In fact, the cdf is especially helpful for finding probabilities of being less than, greater than, or between two values of X. In the dice-rolling example from Tables 7-1 to 7-3, suppose you need to use the cdf to find the following seven probabilities:

1. $P(X \leq 8)$
2. $P(X < 8)$
3. $P(X \geq 5)$
4. $P(X > 5)$
5. $P(5 < X < 8)$
6. $P(5 \leq X \leq 8)$
7. $P(5 \leq X < 8)$

Exploring less-than, greater-than, or less-than or equal-to/greater-than or equal-to probabilities

Problem 1 represents F(8), which equals $\frac{26}{36}$ or 0.72 (72 percent; see Table 7-3). Problem 2, however, needs to be rewritten, because F(x) gives you the probability of being less than or equal to a value for X, not just less than. You have to go down to the nearest integer because that's the only place that probability accumulates in this case. So, $P(X < 8) = P(X \leq 7)$. The answer is $F(7) = \frac{21}{36} = 0.58$ (see Table 7-3).

Problem 3 wants the probability of being greater than or equal to 5, which you need to rewrite so that it involves less-than or equal to probabilities (in other words, complements) and you can use the cdf. Note that $P(X \geq 5) = 1 - P(X \leq 4)$ because you want to include 5, so the possible values that you don't want to include go from 4 on down. That means the answer to Problem 3 is $1 - F(4) = 1 - \frac{6}{36} = \frac{30}{36}$, or 0.83. Problem 4 represents values that are simply greater than but don't include 5, so the complement is another equation with values less than or equal to 5. Therefore, $P(X > 5) = 1 - P(X \le 5) = 1 - F(5) = 1 - \frac{10}{36} = 0.72$.

REMEMBER

The complement of "greater than x" is "less than or equal to x." The complement of "greater than or equal to x" is "less than or equal to x – 1" when x is discrete. Most instructors don't expect you to memorize these facts; the point is that inequalities are particular in terms of where they start and finish, so you should take an extra step to think about each problem individually before you go ahead and solve it. But for a handy reference, here's a list of complements for each of the problems I review in this section:

1. The complement of "less than or equal to x" is "greater than x."
2. The complement of "less than x" is "greater than or equal to x."
3. The complement of "greater than or equal to x" is "less than or equal to x – 1."
4. The complement of "greater than x" is "less than or equal to x."

Exploring between-values probabilities

Problems 5, 6, and 7 involve probabilities of being between two values. To get the probability of being between two numbers by using the cdf, you basically take the probability of being less than or equal to the higher number minus the probability of being less than or equal to the lower number. That leaves you with the area between. The tricky part is being careful about whether you need to include the endpoints, and that depends on whether you're dealing with less-than or less-than or equal-to probabilities.

Problem 5 asks you to find the probability of being between 5 and 8 *noninclusive* (meaning that you don't include the endpoints, 5 and 8). If you take $P(X \le 7) - P(X \le 5)$, you take all the probability of X from 2 through 7 minus the probability of X from 2 through 5, leaving you with the probability that X is 6 or 7, which is exactly what you want. To find the answer, take $F(7) - F(5) = \frac{21}{36} - \frac{10}{36} = \frac{11}{36} = 0.31$ (see Table 7-3).

Problem 6 wants the probability of being between 5 and 8 *inclusive* (meaning that you include the probability of being 5 and 8). So, when you do your subtraction, you need to make sure that you leave behind the probabilities of being between 5 and 8. You take the probability of being 2 through 8, F(8), and subtract everything you don't want, which is everything up through and including 4, or F(4). The answer is $F(8) - F(4) = \frac{26}{36} - \frac{6}{36} = \frac{20}{36}$, or 0.56.

Problem 7 is a combination of problems 5 and 6. You want to include 5 in the probability, but not 8. So, you take $F(7) - F(4) = \frac{21}{36} - \frac{6}{36} = \frac{15}{36}$, or 0.42.

Here are the complements of Problems 5–7:

5. The complement of being between x and y (not including x or y) is being less than or equal to x or greater than or equal to y.
6. The complement of being between x and y (including x and y) is being less than or equal to x – 1 or greater than or equal to y + 1.
7. The complement of being between x and y (including x but not y) is being less than or equal to x – 1 or greater than or equal to y.

Determining the pmf given the cdf

A problem may give you a cumulative distribution function, F(x), and ask you to find the probability mass function (pmf) for X. In other words, it wants you to determine which individual values of X have probability and what those probabilities are. To answer the question, you can just look at the jumps in the graph of F(x). For example, in Figure 7-3, the cdf jumps from $\frac{26}{36}$ at X = 8 to $\frac{30}{36}$ at X = 9, which tells you that P(X = 9) = $\frac{30}{36} - \frac{26}{36} = \frac{4}{36}$. So, you can go from the pmf to the cdf and back again.

Suppose, for example, that you have a random variable X that represents the number of refrigerators sold in a day for a salesperson working at an appliance store. The cdf of X is given by Table 7-4, and the graph of the cdf is shown in Figure 7-4.

Table 7-4 The cdf of X = Number of Refrigerators Sold per Day

X	*F(x)*
< 0	0
0 ≤ X < 1	0.30
1 ≤ X < 2	0.65
2 ≤ X < 3	0.80
3 ≤ X < 4	0.90
4 ≤ X < 5	0.95
5 ≤ X < 6	0.98
X ≥ 6	1.00

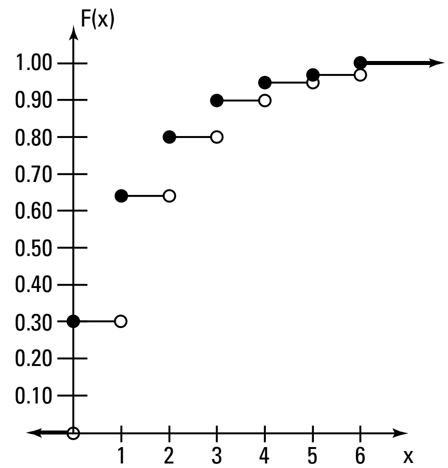


Figure 7-4: Graph of cdf of X = number of refrigerators sold in a day.

Now, given the cdf and its graph, you can find the pmf for X. Notice first that every value less than zero has no accumulated probability, and the first jump in the cdf occurs at $X = 0$. At $X = 0$, the cdf jumps from 0 to 0.30, so that means probability occurs at $X = 0$ and $P(X = 0) = 0.30$. The next jump occurs when $X = 1$, and the cdf jumps from 0.30 to 0.65. That means the probability that $X = 1$ is equal to the net amount of new probability added on, which is $0.65 - 0.30 = 0.35$. The next jump occurs at $X = 2$, where the cdf jumps from 0.65 to 0.80, so $P(X = 2) = 0.80 - 0.65 = 0.15$. The remaining jumps occur at $X = 3$, 4, 5, and 6, and their respective probabilities are

$0.90 - 0.80 = 0.10$ for $X = 3$

$0.95 - 0.90 = 0.05$ for $X = 4$

$0.98 - 0.95 = 0.03$ for $X = 5$

$1.00 - 0.98 = 0.02$ for $X = 6$

The first thing to check after you find the pmf is whether the sum of all its probabilities equals one. In this example, you can see that the probabilities for $X = 0$ to 6 add up to one. Each probability is equal to the jump in the cdf value from one particular point to the next.

Expected Value, Variance, and Standard Deviation of a Discrete Random Variable

Using the probability mass function (pmf) for X, you can figure the long-term average outcome of a random variable, which is called the *expected value,* and the amount of variability you need to expect from one set of results to another, which is called the *variance.* You also figure out the standard deviation to interpret the variance in the results. Probability formulas are available for you to calculate all these entities. I show them to you in this section.

Finding the expected value of X

The *expected value* of a random variable is the long-term average value after repeating an experiment a theoretically infinite number of times. In other words, it's the mathematical equivalent of a weighted average of all possible values of X, weighted by how often you expect each value to occur over the long term. You can refer to how often you expect each value of X to occur as P(x), the probability of success. The notation for the expected value of X is E(X), and the formula for the expected value is $\sum_{\text{all } x} xp(x)$. E(X) often is denoted with the Greek letter μ. Another way to describe the expected value of X is the mean of X.

To find the expected value of a random variable X, follow these steps:

1. **Multiply the value of X by its probability.**
2. **Repeat Step 1 for all possible values of X.**
3. **Sum the results.**

In the example of rolling two dice from Tables 7-1 to 7-3, the probability distribution for X is shown in Table 7-2. Before calculating E(X), try to guess what it will be. If you roll two dice again and again and look at the average of the sum of the two dice, what will that average be? Because the average of one die is 3.5 (halfway between 1 and 6), the average of the sum of two dice is 7. Using the formula for E(X), you have

$$2 * \tfrac{1}{36} + 3 * \tfrac{2}{36} + 4 * \tfrac{3}{36} + 5 * \tfrac{4}{36} + 6 * \tfrac{5}{36} + 7 * \tfrac{6}{36} + 8 * \tfrac{5}{36} + 9 * \tfrac{4}{36} + 10 * \tfrac{3}{36} + 11 * \tfrac{2}{36} + 12 * \tfrac{1}{36} = \tfrac{2}{36} + \tfrac{6}{36} + \tfrac{12}{36} + \tfrac{20}{36} + \tfrac{30}{36} + \tfrac{42}{36} + \tfrac{40}{36} + \tfrac{36}{36} + \tfrac{30}{36} + \tfrac{22}{36} + \tfrac{12}{36} = \tfrac{252}{36} = 7.$$

Therefore, if you roll two dice a theoretically infinite number of times and record the sums, the average sum of all the rolls would be 7. If you do want to roll the dice, know that the more you roll, the closer you'll get to the truth, though it may take hundreds of rolls to get there. That's why probability formulas are so helpful!

Notice that the average sum of the two dice happens to be the middle value between 2 and 12, but that isn't the case in every problem. If the pmf of X is symmetric, E(X) will be the middle value, but if the pmf of X isn't symmetric (for example, it may be skewed — see Figure 7-2), E(X) is affected by it. That's why you use the probabilities of X as the weights, but you don't just average the actual possible values of X to get the expected value.

The graph of the pmf for the refrigerator salesperson example (see the section "Determining the pmf given the cdf") is shown in Figure 7-5.

The refrigerator pmf is skewed to the right, with more weight given to the values 0, 1, 2, and it goes downhill from there. You can think of the expected value as the balancing point on the graph of the pmf, and in the case of Figure 7-5, it should be toward the left side, around 1 or 2, because more weight is on that side. Numbers like 5 and 6 pull it slightly in their direction, however. Calculating E(X), you get the following, using the probabilities found in the section "Determining the pdf given the cdf":

$$0 \ast 0.30 + 1 \ast 0.35 + 2 \ast 0.15 + 3 \ast 0.10 + 4 \ast 0.05 + 5 \ast 0.03 + 6 \ast 0.02 = 1.42.$$

So, over the long term, this salesperson should expect to sell 1.42 refrigerators per day on average, based on the data collected.

REMEMBER

The expected value of X doesn't have to be equal to a possible value of X because it represents a long-term average value. It does, however, have to lie between the smallest and largest possible values of X, which is something to check after you have calculated E(X). Also, note that E(X) isn't a probability, so it falls between zero and one only if all the possible values of X are between zero and one.

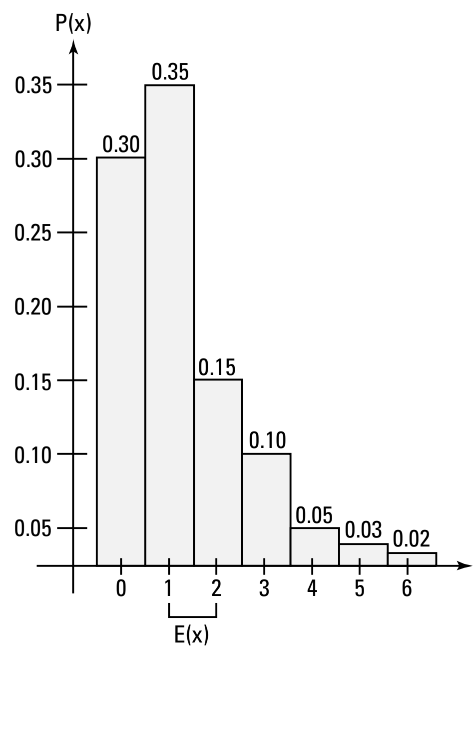


Figure 7-5: Graph of pmf of X = number of refrigerators sold per day.

Calculating the variance of X

The *variance* of a random variable is the expected amount of variability in your results after repeating an experiment a theoretically infinite number of times. The notation for the variance of X is V(X), and the formula is V(X) = $\sigma^2 = E\left[(X-\mu)^2\right] = \sum_{\text{all } x} p(x)(x-\mu)^2$. What the actual variance is depends on how often you expect each value of X to occur and how far away the X values are from the expected value of X. You can think of the variance of X as the weighted average squared distance from E(X), using the probabilities of each of the values of X as the weights.

To calculate the variance of X, follow these steps:

1. **Subtract E(X) — also known as μ — from the value of X.**
2. **Square the difference.**
3. **Multiply by the probability of X for that particular value of X.**
4. **Repeat Steps 1–3 for each value of X.**
5. **Sum the results for all values of X.**

TIP: You square the differences in the variance formula to make the final result a positive value; however, be aware that doing so also squares the units of X. Also, a shortcut formula for the variance is $V(X) = \sigma^2 = E(X)^2 - \mu^2$. In words, this equation is the expected value of X^2, minus the square of the expected value of X. To find $E(X^2)$, use the following formula: $\sum_{\text{all } x} x^2 P(x)$.

To find the expected value of X^2, follow these steps:

1. **Square the value of X.**
2. **Multiply the result by its probability, P(x).**
3. **Repeat Steps 1 and 2 for every value of X.**
4. **Sum the results.**

To find the variance of X using the shortcut formula, follow these steps:

1. **Find $E(X^2)$.**
2. **Find E(X) (also known as μ).**
3. **Square E(X) to get μ^2.**
4. **Take $E(X^2) - \mu^2$.**

Sticking with the refrigerator salesperson example from the previous section, the probability distribution for X is shown in the section "Determining the pmf given the cdf." In the previous section, you calculate the expected value of X as 1.42. That means $\mu^2 = 1.42^2 = 2.02$. All you have left to do is calculate

$$E(X^2) = 0^2 * 0.30 + 1^2 * 0.35 + 2^2 * 0.15 + 3^2 * 0.10 + 4^2 * 0.05 + 5^2 * 0.03 + 6^2 * 0.02 = 0 + 0.35 + 0.60 + 0.90 + 0.80 + 0.75 + 0.72 = 4.12.$$

Therefore, $V(X) = 4.12 - 2.02 = 2.10$.

Finding the standard deviation of X

The variance of X is difficult to interpret because it's in squared units of X; therefore, you typically use the square root of V(X) to describe the expected amount of variation from one set of results to the next. The square root of the variance is known as the *standard deviation* of X, SD(X), and is denoted by $\sigma = \sqrt{V(X)}$. The standard deviation is in the original units of X, because it undoes the squaring effect used to calculate the variance.

In the refrigerator example I present in the previous two sections, the standard deviation of X is $\sigma = \sqrt{2.10} = 1.45$. You can interpret the amount of standard deviation as how much you expect the refrigerator sales to change from the expected value, from day to day, over the long term.

REMEMBER

The variance of X must be greater than or equal to zero because of its definition. The only way V(X) can be equal to zero is if X has only one possible value with probability one (not a very interesting situation). No upper limit exists for one V(X); if the values for X are close together, V(X) will be relatively small, but if they're far apart, V(X) will be large. The same is true for standard deviation.

Outlining the Discrete Uniform Distribution

Probability models were developed because of their common use and ease of application. Each model has its own name, its own formulas for the pmf and cdf, and its own formulas for the expected value, variance, and standard deviation. The most basic of all probability models is the *discrete uniform distribution.* In it, a random variable X has a discrete uniform distribution if the situation meets the following two conditions:

- The possible values of X are consecutive integers from a to b (inclusive).
- Each possible value of X has an equal probability.

The pmf of the discrete uniform

Because each possible value of X has equal probability in the discrete uniform, the probability mass function (pmf) of X is $P(x) = \frac{1}{b-a+1}$ for $a \le x \le b$. For example, suppose a = 0 and b = 9. That means there are b – a + 1 = 10 possible values of X, and each one gets probability $\frac{1}{9-0+1} = \frac{1}{10} = 0.10$. Now suppose a = 5 and b = 9. Now there are 9 – 5 + 1 = 5 possible values of X, each with probability $\frac{1}{9-5+1} = \frac{1}{5} = 0.20$.

Looking at the dice-rolling example from Tables 7-1 to 7-3, if X is the outcome on a single roll of a single die, it fits a discrete uniform probability model, because the possible values of X are {1, 2, 3, 4, 5, 6}, which are consecutive integers. And each possible value of X has equal probability, assuming that the die is fair. Using the formula for the pmf of X when X is a discrete uniform, you get $\frac{1}{6-1+1} = \frac{1}{6} = 0.17$.

However, the sum of two dice rolled doesn't have a uniform distribution. Although the possible values of X = the sum of two die are consecutive integers {2, 3, 4, . . . , 10, 11, 12}, each possible value of X doesn't have equal probability. You can tell this by looking at the pmf of X for the sum of two dice (see Table 7-2), or by looking at the graph of the pmf shown in Figure 7-1 (which is a relative frequency histogram). The graph of the pmf is mound shaped; a graph of a uniform distribution should be flat (or uniform). Figure 7-2b shows a picture of a discrete uniform.

REMEMBER

The graph of the pmf of a uniform random variable is flat because each possible value of X has equal probability.

The cdf of the discrete uniform

The cumulative distribution function (cdf) for a discrete uniform random variable is given by the formula

$$F(x) = \begin{cases} 0, \text{ for } x < a \\ \dfrac{x-a+1}{b-a+1}, \text{ for } a \le x < b \\ 1, \text{ for integers } x \ge b \end{cases}$$

You assume that the values are integers because you're working with a discrete random variable. (For values that aren't integers but are within the possible range of X, round down to the nearest integer and put that value in for X.) In the single die-rolling example from the previous section, $P(X \le 3) = F(3)$. You have a = 1, b = 6, and x = 3, so you plug these numbers into the formula,

$\frac{x-a+1}{b-a+1}$, to get $\frac{3-1+1}{6-1+1} = \frac{3}{6} = 0.50$. You can verify that value by simply adding the probabilities for X = 1, 2, 3 = $\frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = 0.50$.

TIP

In the case of the cdf of the discrete uniform, most of the time adding up the desired probabilities yourself is easier than calculating and/or using the cdf because they're so easy to calculate and add together. It's more important at this point, however, to understand the basic ideas behind the cdf and to know how to calculate it for a few values. When you look at more complicated distributions in Parts IV and V, the cdf will be more helpful than it is right now.

The expected value of the discrete uniform

If X has a discrete uniform distribution, the expected value of X is equal to $\frac{b+a}{2}$. It represents the midpoint between a and b.

TECHNICAL STUFF

You find the formula for the expected value by taking the definition of E(X), $\sum_{\text{all } x} xP(x)$, plugging in the formula $\frac{1}{b-a+1}$ for $a \le x \le b$ for P(x), and simplifying. Although deriving the formula for E(X) for a discrete uniform isn't something you should have to do, knowing the basic ideas behind its derivation helps you to understand, connect with, and remember the various formulas you're working with.

In the single die rolling example from the previous sections, a = 1 and b = 6, so $E(X) = \frac{b+a}{2} = \frac{6+1}{2} = 3.50$. (For more on expected value, see the section "Expected Value, Variance, and Standard Deviation of a Discrete Random Variable.")

The variance and standard deviation of the discrete uniform

If X has a discrete uniform distribution, the variance of X is equal to $\frac{(b-a+2)(b-a)}{12}$, and the standard deviation is equal to $\sqrt{\frac{(b-a+2)(b-a)}{12}}$.

In the single die rolling example from the previous sections, a = 1 and b = 6, so $V(X) = \frac{(6-1+2)(6-1)}{12} = \frac{(7)(5)}{12} = 2.92$, and $SD(X) = \sqrt{\frac{(b-a+2)(b-a)}{12}} = \sqrt{2.92} = 1.71$. You can expect the outcome of a single die to change on average about 1.71 from the mean (3.50) as you go from one roll to the next. (For more on variance and standard deviation, see the section "Expected Value, Variance, and Standard Deviation of a Discrete Random Variable.")

Chapter 8

Juggling Success and Failure with the Binomial Distribution

In This Chapter

- Satisfying the conditions for use of the binomial probability model
- Figuring binomial probabilities with the pmf and the cdf
- Deducing the expected value and variance for the binomial

The binomial probability model has a wide range of uses in the real world. *Binomial* means two names and is associated with situations involving two outcomes, labeled as success or failure. Examples include winning the lottery or not; hitting a red light on your way to work or not; or developing a side effect from a certain drug or not. Many researchers are interested in the probability that a success or failure event occurs, the percentage of successes in the population, or the likelihood of certain outcomes that result from a success/failure situation.

This chapter focuses on the binomial probability model — understanding when you can use it, finding probabilities for it, and calculating and interpreting expected value and variance within it.

Recognizing the Binomial Model

Each probability model has its own set of properties or conditions by which you can identify it. All these conditions must hold for a random variable X to be considered for, or have the characteristics of, that particular probability model. The binomial is no exception.

A situation must meet the following conditions for a random variable X to have a binomial distribution:

- You have a fixed number of trials involving a random process; let n be the number of trials.
- You can classify the outcome of each trial into one of two groups: success or failure.
- The probability of success is the same for each trial. Let p be the probability of success, which means 1 – p is the probability of failure (see Chapters 2 and 7 for more about complements).
- The trials are independent, meaning the outcome of one trial doesn't influence the outcome of any other trial.

If the situation meets all these conditions, you let X count the total number of successes that occur in n trials (which means the number of failures is n – X). In such a case, the situation is said to have a binomial distribution with n trials and the probability of success equal to p.

TIP

If the situation meets the first three conditions, the fourth typically is a foregone conclusion.

Checking the binomial conditions step by step

You know a situation falls under the binomial probability model if you can define a random variable X that fits all the criteria for a binomial random variable I outline in the introduction to this section. Suppose, for example, that you flip a fair coin 10 times and count the number of heads ("fair" meaning that the probabilities of heads and of tails are the same: 50 percent). Does this situation fall under the binomial probability model? You can check by reviewing your responses to the questions and statements in the list that follows:

1. **Do you have a fixed number of trials of a random process?**

   Yes. You're flipping a fair coin a fixed number of times (10). So, n = 10.

2. **Can you classify the outcomes of each trial into only two groups?**

   Yes. The outcome of each flip is either heads or tails, and you're interested in counting the number of heads that appear, so flipping a head represents success and flipping a tail represents a failure.

   REMEMBER

   Just because the terms "success" and "failure" are used in the binomial doesn't necessarily mean a success is good and a failure is bad. Success means that you get the outcome you want to count, and failure means you get the outcome you don't want to count. What you're counting may be a negative event, such as someone developing a side effect, yet that outcome is a success because it's the outcome you want to count.

3. **Is the probability of success the same for each trial?**

   Yes. Because you flip the coin the same way each time and the coin is fair, the probability of success (getting a head) is $p = \frac{1}{2}$ for each trial. You also know that $1 - \frac{1}{2} = \frac{1}{2}$ is the probability of failure (getting a tail) on each trial.

4. **Are the trials independent?**

   Yes. You're flipping the coin, and you assume that the outcome of one flip doesn't affect the outcome of subsequent flips.

Because the coin-flipping experiment meets the four conditions, your random variable X, which counts the number of successes (heads) that occur in 10 trials, has a binomial distribution with $n = 10$ and $p = \frac{1}{2}$.

REMEMBER

The criterion you use to check for a binomial also can be used to clearly define success and failure in the problem, what exactly X is counting, and the values of n and p. Having these elements clearly defined and identified before you begin any binomial problem is extremely important. Even if a problem lets you know that X is binomial, and you don't have to check the conditions, you want to be sure to organize the information about X, n, and p because doing so makes working through the problem much easier, especially when the problems become more difficult.

Spotting a variable that isn't binomial

The following pages highlight certain circumstances that illustrate when a situation doesn't qualify for the use of the binomial probability model.

Surveying the number of trials: Fixed or not?

Many situations appear to be binomial because of the success/failure outcomes, but if they don't meet all the conditions, they don't merit the use of the binomial model. Always check the conditions of a binomial so you know for sure whether you actually have a binomial. Suppose, for example, that you flip a coin until you get four heads (four successes total). Does this experiment fall under the binomial model? It sounds like the same scenario you work with in the previous section of this chapter, where you flip a coin ten times and count the number of heads, but it's a little different. Check the conditions to see.

First, ask whether you have a fixed number of trials. You flip a coin until you get four heads; you don't have to flip four in a row, just four heads total. In this case, the number of successes is fixed but not the number of trials. To get four successes, you have to flip the coin at least four times, but you don't know for sure when you'll reach your goal. One time you may need ten flips to get four heads, and another time you may need twenty flips or five flips. The answer is no, you don't have a fixed number of trials, so you don't have a binomial model, and you don't need to check the other conditions. (This model is actually a geometric probability model, which I discuss in Chapter 14.)

Sampling without replacement: Does it change p?

If you sample an object from a population without returning it to the population, you're sampling *without replacement*. Unless the population is extremely large and the sample is small in comparison, sampling without replacement doesn't qualify for the binomial probability model because it violates the criterion of p remaining the same from trial to trial. To ensure that you uphold this criterion, you have to sample *with replacement*, which means that after you sample an individual from a population, you return it to the population where you can sample it again.

Here's an example scenario. In a jar, you have 20 red beads, 20 black beads, and 20 green beads. You reach into the jar, take a bead out, and record its color. You go through this process five times, and you're interested in the total number of red beads you pull from the jar. Is this a binomial model? Check the conditions:

1. **Are there a fixed number of trials?**

   Yes, $n = 5$ trials here.

2. **Can you classify the outcomes of each trial in one of two groups?**

   Yes. You're interested in only the red beads, so pulling a red bead represents a success and pulling a different bead represents a failure.

   TIP: You can think of situations that have more than one outcome as binomial, if you define success as one or more of those outcomes and failure as not getting those outcomes. For example, you can measure test scores and then classify them as either pass or fail.

3. **Is the probability of success the same on each trial?**

   Think about this one for a moment. You're reaching into the jar five times and pulling out a bead. You have 20 red, 20 black, and 20 green beads in the jar. What's the chance of pulling a red bead on the first trial? With 20 red beads in a total of 60 beads, $p = \frac{20}{60} = 0.33$. Now, suppose you reach in and pull out a red bead. What's the chance of pulling a red bead on the second trial? You took one red bead out, so now the jar contains only 19 red beads out of 59 remaining beads. This time, $p = \frac{19}{59} = 0.32$, which is different from the probability of success on the first trial. So, no, the probability of success isn't the same for each trial, so this experiment isn't binomial. (This model falls under the hypergeometric probability model, which I discuss in Chapter 16.)

Finding Probabilities for the Binomial

After you identify that X falls under a binomial probability model (in other words, the conditions are met), your next step is to find the probabilities. The good news is that you don't have to find any probabilities from scratch; you get to use previously established formulas for finding binomial probabilities. You have this luxury because all binomial models use the same method for calculating their probabilities; the only differences are in the values for n and p, which are unique to each problem.

You can calculate probabilities for binomials in two possible ways, depending on the type of problem you're working:

- **The pmf** (*probability mass function;* see Chapter 7) gives you the straight formula for calculating the probability that X equals a certain number [denoted P(x)]. You use this formula to calculate the probability that X takes on a single value or a small number of values.
- **The cdf** (*cumulative distribution function;* see Chapter 7) gives you a formula for calculating the accumulated probability from zero to X for any particular value of X [denoted F(x)]. You use this formula to calculate the probability that X takes on a range of values greater than, less than, or between two numbers.

I discuss these methods and formulas in this section.

Finding binomial probabilities with the pmf

The probability mass function (pmf) for a binomial random variable X is $\binom{n}{x}p^x(1-p)^{n-x}$, where

- n is the fixed number of trials.
- x is the specified number of successes.
- n – x is the number of failures.
- p is the probability of success on any given trial.
- 1 – p is the probability of failure on any given trial. (***Note:*** Some textbooks use the letter q to denote the probability of failure rather than 1 – p.)

The pmf holds for any value of X between zero and n. Suppose, for example, you flip a fair coin three times. What's the chance you get

- All heads?
- Exactly two heads?
- More than one tail?

To find the probability distribution of a binomial random variable, you use the pmf formula, which incorporates counting rules (see Chapter 5), addition and multiplication rules (see Chapter 2), and tree diagrams (see Chapter 3). In a nutshell, you multiply the number of possible outcomes that give you the desired value of X by the probability of one of those outcomes. For example, if you want the probability of getting three heads in five flips, you count the number of possible ways to get three heads in five flips times the probability of getting three heads in a row followed by two tails (one possible outcome that gives you three heads in five flips).

The notation $\binom{n}{x}$ means the number of ways to get x successes from n trials. You call this number "n choose x." (For a full explanation of "n choose x" and how to calculate it, see Chapter 5.) For example, $\binom{3}{2}$ means "3 choose 2" and stands for the number of ways to get 2 successes in 3 trials. Suppose a success is flipping a head, and failure is flipping a tail. You have 3 ways to choose 2 heads in 3 trials: you can flip HHT, HTH, or THH.

When the number of trials grows large, writing out all the possibilities is more difficult, so you need a formula for "n choose x" (see Chapter 7 for full details). Basically, to calculate "n choose x," you use the formula $\binom{n}{x} = \frac{n!}{x!(n-x)!}$. The notation *n!* stands for *n-factorial,* the number of ways to rearrange n items. To calculate *n!,* you take $n * (n-1) * (n-2) * \ldots * 2 * 1$. For example, you calculate 3! as $3 * 2 * 1 = 6$, 2! as $2 * 1$, and 1! as 1. By definition, you can let 0! equal 1. So, to calculate "3 choose 2," you use $\binom{3}{2} = \frac{3!}{2!(3-2)!} = \frac{3 * 2 * 1}{(2 * 1)(1!)} = \frac{6}{2 * 1} = 3$.

Putting the pmf formula into action

You can use the pmf for the binomial to answer the three probability questions that come up in the introduction to this section. Suppose you flip a fair coin three times. The chance of getting all heads is P(3), because X is counting the number of heads, and in this case, you want the probability that X = 3. Using the formula for the pmf, you have n = 3, p = ½, and X = 3, so you get $\binom{3}{3}(\frac{1}{2})^3(1-\frac{1}{2})^{3-3} = 1 * \frac{1}{8} * 1 = 1 * \frac{1}{8} = \frac{1}{8}$. Note the number in front of the formula is "3 choose 3," which equals 1, because when you have three total slots to fill with heads or tails and you choose heads for all three slots, you have only one way to do it — HHH.

Similarly, the chance of flipping two heads is the probability that X = 2, written as P(2). When you have three slots and you choose two slots to fill with heads, you can do it three ways — HHT, HTH, or THH. You multiply the probability of any single outcome by three to get the final answer. Using the pmf formula, you get $\binom{3}{2}(\frac{1}{2})^2(1-\frac{1}{2})^{3-2} = 3 * \frac{1}{4} * \frac{1}{2}$, which is $3 * \frac{1}{8} = \frac{3}{8}$.

REMEMBER

When finding the probability of getting more than one tail, remember that X counts the number of heads, so you must first translate the problem so it deals with heads. If you have more than one tail in three flips of a coin, you can have two tails or three tails. If you have two tails, you have only one head, and if you have three tails, you have zero heads. So, the probability of flipping more than one tail equals the probability of having one or zero heads. Take $P(1) + P(0) = \frac{3}{8} + \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$. (You get P[1] and P[0] by plugging X = 0 and X = 1 into P[x].)

Here's an example where the probability of success isn't necessarily 50 percent. Suppose you cross three traffic lights on your way to work, and the probability of each of them being red is 0.70. The lights are independent. You want the pmf of the number of red lights you encounter on your way to work, so you let X be the number of red lights you encounter and n – X be the number of non-red lights. You know that p = probability of red light = 0.70, and 1 – p = probability of a non-red light = 1 – 0.70 = 0.30. Using the formula for the pmf of X, you know the following:

$$P(0) = \binom{3}{0}0.70^0(1-0.70)^{3-0} = \frac{3!}{0!(3-0)!}(1)*(0.30)^3 = 1*(0.30)^3 = 0.027.$$

$$P(1) = \binom{3}{1}0.70^1(1-0.70)^{3-1} = \frac{3!}{1!(3-1)!}(0.70)^1*(0.30)^2 = 3*(0.70)^1*(0.30)^2$$
$$= 0.189.$$

$$P(2) = \binom{3}{2}0.70^2(1-0.70)^{3-2} = \frac{3!}{2!(3-2)!}\,(0.70)^2*(0.30)^1$$
$$= 3*(0.70)^2*(0.30)^1 = 0.441.$$

$$P(3) = \binom{3}{3}0.70^3(1-0.70)^{3-3} = \frac{3!}{3!(3-3)!}(0.70)^3*(0.30)^0 = 1*(0.70)^3 =$$
$$= 0.343.$$

The final pmf of X for this example is shown in Table 8-1. Notice that all the probabilities are greater than or equal to zero, and they all sum to one, as they should with a probability distribution (see Chapter 7).

Table 8-1 The pmf of X = Number of Red Traffic Lights (n = 3, p = 0.70)

X	*P(x)*
0	0.027
1	0.189
2	0.441
3	0.343

TIP: Suppose you define a new random variable for the traffic-light example, Y — the number of times you didn't get a red light (in other words, the number of non-red lights out of three trials). Instead of redoing the entire problem, you can simply switch the notation for red and non-red around, and the probabilities switch with them. This tactic is a favorite of many probability teachers. The probability distribution for Y is shown in Table 8-2. Notice that this distribution is the exact opposite of the distribution for X, because they're complements of each other (see Chapter 2).

Table 8-2 The pmf for Y = Number of Non-Red Traffic Lights (n = 3, p = 0.30)

Y	*P(y)*
0	0.343
1	0.441
2	0.189
3	0.027

Picturing the binomial pmf

You may be asked to draw a picture of the probability mass function of the binomial distribution to see its shape, determine where the middle is (otherwise known as the *expected value*), and look at the amount of variability you can expect in the results. Some binomials are symmetric, with the same probabilities on either side of the middle (mirror images of each other), and some look skewed, with a lump on one side and a long sloping tail on the other. It all depends on the value of p.

To get a visual representation of the pmf, you can graph it by using a *relative frequency histogram* — a graph that shows the possible values of x along with their relative frequencies (percentages). Graphs look different depending on the values of n and p. Each possible value of X (from zero to n) gets a bar on the relative frequency histogram, so the larger n is, the more bars that appear on the graph. The value of p is a probability, so it can be anywhere from zero to one. If p = ½, such as for a coin-flipping example where you're looking for the number of heads on three flips of a fair coin, the pmf of X is symmetric, because the chance for success is the same as for failure. Figure 8-1 shows the graph of the pmf of X in this case.

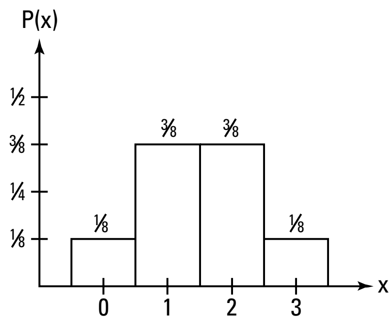


Figure 8-1: Graph of the pmf of X = number of heads on three flips of a fair coin (p = 0.50).

If p is greater than ½, expect that the overall shape of the probability distribution of X will not be symmetric, like in the traffic-light example from the section "Putting the pmf formula into action" (see Chapter 7 for a discussion of the various shapes of probability distributions). Because the probability of success is more than ½, a success is more likely to occur on any given trial. That means the probabilities for larger values of X are bigger than the probabilities for smaller values of X, in general terms. In other words, if p > ½, the graph of the pmf is skewed to the left. The traffic-light example has p = 0.70, and the graph of its pmf is shown in Figure 8-2.

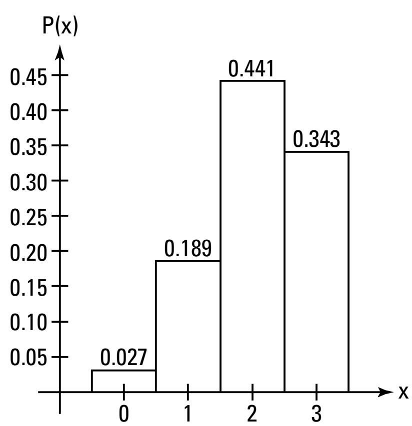


Figure 8-2: Graph of the pmf of X = number of red traffic lights (p = 0.70).

How would you describe the shape of the binomial when p is less than ½? The pmf of the traffic-light example where you count the number of non-red traffic lights (see the section "Putting the pmf formula into action") is shown in Figure 8-3. The pmf for Y is skewed to the right, because the probability of getting a non-red light is 0.30, which is less than ½. In fact, this p is the exact mirror image of the p shown in Figure 8-2. The probability that X = 0 (red lights) is the same as the probability that Y = 3 (non-red lights), and so on.

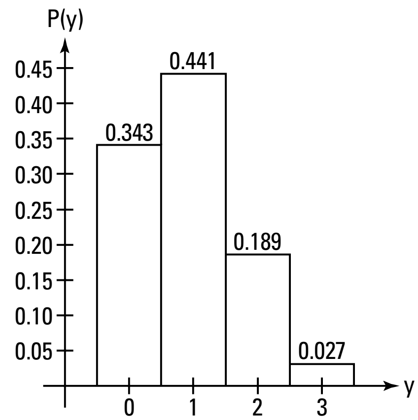


Figure 8-3: Graph of the pmf of Y = number of non-red traffic lights (p = 0.30).

REMEMBER

For a binomial distribution, the closer p is to ½, the closer the graph of the pmf of X is to a symmetric shape like the one shown in Figure 8-1. As p approaches one, the graph of the pmf of X grows more and more skewed to the left (Figure 8-2 is skewed to the left; p = 0.70). And, as you may expect, as p approaches zero, the graph of the pmf of X gets more and more skewed to the right (Figure 8-3 is skewed to the right because p = 0.30). Additionally, the number of bars on the graph of the pmf is equal to n + 1 because the random variable can go from zero to n.

Finding binomial probabilities with the cdf

Calculating the probability mass function for a problem with multiple values of X can take a long time. For problems that involve more than two values of X, using the formula for the cumulative distribution function (cdf) is quicker.

Suppose, for example, that you want the probability that X is less than or equal to 6 when n is 10 and p is ½. That means you want the probability that X is 0, 1, 2, 3, 4, 5, or 6, which equals P(X = 0) + P(X = 1) + P(X = 2) + . . . + P(X = 5) + P(X = 6). Each of these probabilities is a probability of a single value of X, and you need the pmf to calculate it, which takes a long time and many calculations. That's why you have the cdf. (See Chapter 7 for complete information on the cdf.)

Understanding the cdf

The cdf of X is the probability that X is less than or equal to any number x, and is denoted by F(x). For example, the probability that X is less than or equal to 6 is F(6). The probability that X is less than or equal to 10 is F(10).

The cdf gives you the total accumulated probability up through x for any real number x. (Real numbers can take on any possible value, including fractions and numbers that you can't write as fractions — not just integers, basically.) In other words, the cdf gives the sum of all the values of the pmf from zero through x for any number x. For the binomial model, you find the cdf with the following: $F(x) = \sum_{X \le x} p(x) = \sum_{X \le x} \binom{n}{x} p^x (1-p)^{n-x}$. You take the sum of all the probabilities of all values less than or equal to x. (If you're between two integers, you round down to the nearest integer when finding the value of the cdf.)

Look at the traffic-light example from the subsection "Putting the pmf formula into action," where X is the number of red lights on three trials, and p = 0.70 is the probability of a red light on any given trial; you want to find the cdf of X. The pmf for X is shown in Table 8-1. For each value of X, you add up all the probability from zero to x to get the cdf, starting with X = 0:

$F(0) = P(X \le 0) = 0.027$

$F(1) = P(X \le 1) = 0.027 + 0.189 = 0.216$

$F(2) = P(X < 2) = 0.027 + 0.189 + 0.441 = 0.657$

$F(3) = P(X < 3) = 0.027 + 0.189 + 0.441 + 0.343 = 1.00$

For any number beyond 3, F(x) = 1. For any number below 0, F(x) = 0. For any number between two integers, you round down to the nearest integer and find F(x).

Taking advantage of the table for the binomial cdf

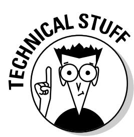
TIP

Experts have already placed the cdf for the binomial into table form for you to use where the probabilities are summed up in each case, so you just look up the value of the cdf for the number you need and go from there. The cdf of the binomial distribution is given by Table A-1 in the Appendix. I call it the binomial table.

You can use the binomial table to answer any question about a binomial probability for X above, below, or between two values (any problem that's too time consuming to calculate with individual probabilities for X). To use the binomial table for the cdf to calculate a probability, follow these steps:

1. **Find the mini-table associated with your particular value of n (the number of trials).**
2. **Find the row that represents the value of x that you are interested in.**
3. **Find the column that represents your particular value of p (or the one closest to it).**
4. **Intersect that row and column in the table.**

You see several mini-tables provided in the binomial table; each one corresponds with a different n for a binomial (n = 5, 6, 7, 8, 9, 10, 15, 20, and 25 are available). Each mini-table has rows and columns. You have many choices of columns running across the top representing p. Running down the side of the table, you see all the possible values of X from zero through n, each with its own row. Notice that within a given column in the table, say the column for p = 0.50, the first probability is the probability that X is less than or equal to zero, which for the binomial is the same as the probability that X equals zero. From there, the accumulated probabilities increase up to one, when X = n, because at that point, all the probability has been accumulated.



If you look at a column where p is small — p = 0.01, for example — you see more probability accumulating right away when X = 0, because the probability of success is so small; therefore, the chance of getting less than or equal to zero successes is large. The graph of the pmf is skewed to the right. And for a column where p is large — say, p = 0.99 — no probability accumulates until the later values of X, because the probability of success is large. The graph of the pmf in this case is skewed to the left.

Using the binomial cdf to find less-than or equal-to probabilities

Suppose, for example, that you flip a fair coin ten times. What's the probability of getting less than or equal to six heads? You want F(6), which is the cdf at the value six. The mini-table you want to find here is the mini-table for n = 10, or the sixth mini-table in the binomial table.

You want to find the probability that X is less than or equal to six, so choose the row representing six. Next, find the column for p = 0.50 and intersect that column with the row for X = 6; you get 0.828, which is the probability that X is less than or equal to six when n = 10 and p = 0.50. In other words, with a fair coin and ten trials, where the probability of success (heads) is 0.50, the probability of getting less than or equal to six successes (heads) is 0.828. This probability is high, as it should be, because the chance of getting a head is 0.50 on a single toss.

Finding the probability that X is at most six is the same as saying X is less than or equal to six. So, you can use the cdf directly to find probabilities involving the phrase "at most."

Using the binomial cdf to find less-than probabilities

If you flip a fair coin ten times, suppose you want the probability of flipping less than six heads — $P(X < 6)$, not including 6. Because the cdf works only with probabilities that are less than or equal to, you must rewrite this probability using less-than or equal-to notation. If X is less than six, it must be less than or equal to five. Using a number line can help you think this situation through. So, on the mini-table for $n = 10$ from the binomial table in the Appendix, look at the row where $X = 5$ and the column where $p = 0.50$. Intersect them and you'll find the number 0.623.

Using the binomial cdf to find greater-than probabilities

Say that you flip a fair coin ten times and you want the probability that you'll flip more than six heads — $P(X > 6)$, not including 6. Knowing that the cdf gives probabilities in less-than or equal-to-X form, you have to think of a way to rewrite this probability as a less-than or equal-to probability. The way to do it is to use complements. The complement of $P(X > 6)$ is one minus everything up to and including six, so $P(X > 6) = 1 - P(X \leq 6) = 1 - 0.828$ (from the binomial table in the Appendix) $= 0.172$.

Using the binomial cdf to find greater-than or equal-to probabilities

Suppose you want to flip a fair coin ten times, and you want to find the probability of flipping six or more heads. In probability notation, you want $P(X \geq 6)$. In order to use the cdf for the binomial, you have to change this problem to less-than or equal-to notation, which means you need to think about the complement. Greater than or equal to six means everything from six on up, so the complement is one minus everything below six, not including six — in other words, less than or equal to five. So, $P(X \geq 6) = 1 - P(X \leq 5) = 1 - 0.623$ (from the binomial table in the Appendix) $= 0.377$.

Using the binomial cdf to find probabilities between two values

Suppose that you want to flip a fair coin ten times, and you want the probability that you'll flip between six and eight heads — in other words, X is between six and eight, inclusive (it includes both six and eight). You can use the binomial table for the cdf in the Appendix to solve this between-values probability problem, but the cdf uses only less-than or equal-to probabilities, so you have to think in those terms. Notice that if you took all the probabilities up to and including eight, you'd have $P(X \leq 8) = 0.989$ from the binomial table. But you want only the part between six and eight; you don't want the part from zero up through five. What do you do with the probability you don't want? Subtract it! You find $P(X \leq 5) = 0.623$ and subtract it from $P(X \leq 8)$ to get $0.989 - 0.623 = 0.366$.

If you want, say, the probability that X is between six and eight, including six but not including eight, you want the probability up through seven [F(7)] minus the probability up through five [F(5)].

When finding between-values probabilities, be careful where your probabilities go. You can't always take the probability up to the second number minus the probability up to the first number. It depends on whether a less-than or a less-than or equal-to probability is in the statement. Thinking it through before you work directly with probabilities can help you avoid potential mistakes.

Graphing the binomial cdf

You may want (or be asked) to graph the cdf of a random variable to see how fast it accumulates probability and give a visual representation of how the random variable behaves. The cdf of any binomial is a step function that's zero for any value less than zero. At each step along the way where X = 0, 1, 2, 3, . . . , n, the probability takes a step up because these points are where it accumulates. The probability then stays at one for any values greater than or equal to n. The magnitude of each step, and the number of steps in the cdf, respectively, depend on the values of p and n.

The graph of the cdf of X for the traffic-light example (see the section "Putting the pmf formula into action") is shown in Figure 8-4. (Recall its pmf is shown in Figure 8-2.) Notice that the cdf starts at zero, takes steps up each time more probability accumulates (where X = 0, 1, 2, and 3), and stays at one for every value after three and beyond. Because the probabilities are largest at two and three, those steps are biggest in the cdf. As you may expect, the graph of the cdf of Y, where Y is the number of non-red lights, has the biggest jumps at zero and one.

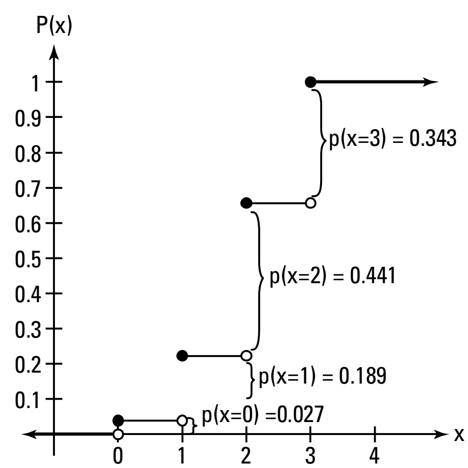


Figure 8-4: Graph of the cdf of X = number of red traffic lights.

Formulating the Expected Value and Variance of the Binomial

Although calculating the expected value and variance of a binomial from scratch is possible if you use the formulas from Chapter 7, you can rely on experts who've already done that and discovered that the results turn out the same every time. These nice folks give you the benefit of using those results without having to do all the extra calculations. I happen to be one of those people . . . what a nice author, eh? In this section, I present the easy-to-use formulas for the expected value and variance of the binomial.

The expected value of the binomial

The *expected value* of a random variable is the weighted average of its possible values multiplied by their probabilities. It represents the long-term average value of X over an infinite number of trials and is denoted E(X) (see Chapter 7). The general formula for the expected value of a discrete random variable X is $\sum_{\text{all } x} xP(x)$. For the binomial random variable, you just plug in the formula for P(x) into this equation to get the following: E(X) = $\sum_{\text{all } x} xP(x) = \sum_{\text{all } x} x\binom{n}{x}p^x(1-p)^{n-x} = \ldots = n * p$.

Some fancy algebra footwork goes on between the first part and the last part of the formula, but that goes outside the scope of this book. Suffice it to say that when you finish all the gritty algebra details, you come out with a nice looking formula that doesn't have any summation signs in it — just two simple items, n and p. You multiply these items together to get E(X) for a binomial.

The expected value of a binomial has a nice intuitive meaning. Suppose, for example, that you flip a fair coin 100 times: n = 100 and p = 0.50. How many heads do you expect to get on average over the long term? It makes sense that your answer would be 50. And when you take n * p, that's exactly what you get.

Now suppose the coin isn't fair, and p = 0.10. If you flip the coin 100 times, how many heads should you expect to get now? Under these circumstances, you can expect fewer heads because the chance of getting a head is only 10 percent on any given toss. If you take n * p in this case, you get 100 * 0.10 = 10 heads, which makes good sense. Similarly, if p is large — say 0.90 — you can expect more heads, and on 100 flips of a coin, you'll get, on average, 100 * 0.90 = 90 heads according to this model.

The variance and standard deviation of the binomial

The *variance* of a random variable is the weighted average deviation squared from the mean (expected value). It represents the long-term average amount of variability in X over an infinite number of trials and is denoted V(X) (see Chapter 7). The general formula for variance of a discrete random variable X is $E(X-\mu)^2 = E(X)^2 - [E(X)]^2 = \sum_{\text{all x}} x^2 P(x) - \left(\sum_{\text{all x}} xP(x)\right)^2 = n * p * (1-p)$.

For the binomial random variable, you just plug in the formula for p(x) into the equation, and after all the algebra footwork, you get $n * p * (1 - p)$. Just like with the expected value (see the previous section), you come out with a nice looking formula that doesn't have any summation signs in it — just three simple items: n, p, and 1 – p. You multiply the items together to get V(X) for a binomial. The *standard deviation* is just the square root of the variance, which in this case is $\sqrt{np(1-p)}$.

TIP

The variance of a binomial has some intuitive meaning in terms of its formula. The only variability in a binomial situation is the variability in going from a success (with probability p) to a failure (with probability 1 – p). And over n trials, it makes sense that the total variability in the results is $n * p * (1 - p)$.

Chapter 9

The Normal (but Never Dull) Distribution

In This Chapter

- Separating the continuous from the discrete
- Understanding the normal probability model
- Going from start to finish with regular normal probabilities
- Working backward to solve backwards normal problems

The normal distribution is a very common probability model used to describe many random phenomena whose results form the well-known bell-shaped pattern, where most of the results mound up in the middle. As you move away from the middle on either side, the values occur less and less often.

In this chapter, you work on probability problems for the normal model, including problems where you're given the probability that X is above or below a certain value, and you have to find the value of X that's associated with that probability. (These problems are known in the probability business as *backwards normal problems.*) Because the probabilities for the normal are very difficult to find, you use a table that already has certain probabilities calculated for you. You just have to do a little algebra in order to transform your problem into one that you can solve by using the table.

Charting the Basics of the Normal Distribution

Before you begin working with the normal distribution, you need to know that it's a *continuous distribution,* not a discrete one. A continuous distribution's probability function takes the form of a continuous curve, and its random variable takes on an uncountably infinite number of possible values. This means the set of possible values is written as an interval, such as negative infinity to positive infinity, zero to infinity, or an interval like [0, 10], which represents all real numbers from 0 to 10, including 0 and 10. (A discrete distribution, on the other hand, has either a finite or a countably infinite number of possible values. That means you can enumerate or make a listing of all possible values, such as 1, 2, 3, 4, 5, 6 or 1, 2, 3, . . .)

The information I present in Chapter 7 on probability distribution basics assumes the random variable (and hence its distribution) is discrete. Only a few items change when you move to a continuous random variable. If X is a continuous random variable,

- **You can't use a probability mass function (pmf) for X.** Rather, the *probability density function* for X (pdf) tells you how dense, or heavy, the concentration of probability is for X at any particular point (see the next section, "The shape, center, and spread"). The probability density function for X is denoted by f(x) and is typically a continuous function.
- **You can find probabilities for X for any interval of values for X (for example, the probability that X is between 1 and 3, or the probability that X is less than 0).** Probabilities over an interval represent the "area under the curve" that is the graph of f(x). The total area under the entire curve equals one.
- **The probability that X equals any particular value is zero because there are uncountably infinite possible values of X.**

The shape, center, and spread

Each normal distribution has a similar shape, but each one is a little different in terms of its center and spread. In this section, you find out the basics of the normal distribution: its shape, center, and spread.

X has a normal distribution if its values fall into a bell-shaped and symmetric pattern. Each normal distribution has its own center, measured by the mean and denoted by μ. Each normal distribution also has its own spread, measured by the variance and denoted by σ^2. The standard deviation, the square root of the variance, is denoted by σ. The mean and variance (and/or standard deviation) are given entities; no calculations are required to find them. Figure 9-1 shows a picture of three different normal distributions with different means and standard deviations.

REMEMBER

Note that the saddle points (points highlighted by arrows in Figure 9-1 on either side of the mean) on each graph are the places where the graph changes from concave down (an upside-down bowl) to concave up (a right-side-up bowl). The distance from the mean out to either saddle point is equal to the standard deviation for the normal distribution. Also note that for any normal distribution, almost all its values lie within three standard deviations of the mean.

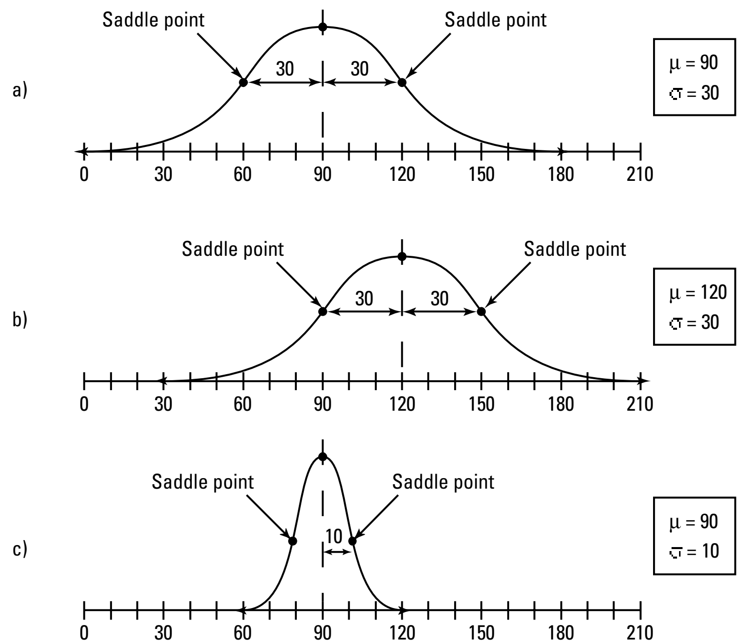


Figure 9-1: Three normal distributions.

Figures 9-1a and 9-1b have the same standard deviation, so their average spread about the mean is the same, but they have different means; Figure 9-1b has a higher mean (120 compared to 90 in Figure 9-1a), so its entire distribution is shifted to the right 30 units. Figures 9-1a and 9-1c have the same mean (90), but they have different standard deviations; Figure 9-1c has a standard deviation of only 10 compared to 30 in Figure 9-1a. That means 9-1c is much more condensed around the mean than Figure 9-1a.

The *probability density function,* f(x), for the normal distribution is a continuous function whose formula gives you the kinds of graphs you see in Figures 9-1 and 9-2. Although the graphs look nice and neat, the formulas don't. But don't take my word for it — look at the probability density function for a normal distribution with mean μ and standard deviation σ:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{x-\mu}{\sigma}\right)^2} \text{ for } -\infty < x < +\infty$$

As bad as it looks, this probability density function is even harder to work with in its original state. To find probabilities for a continuous random variable, normally you find the area under the curve between the two points you want the probability for, but this function is so complex that it requires the use of a computer to find those areas. But enough with the scare tactics; time for the good news: All the basic results you need to find probabilities for any normal distribution are boiled down into one table, called the Z table, based on one particular normal distribution, called the standard normal (Z) distribution. All you need is one formula to transform your normal distribution (X) to the standard normal (Z) distribution, and you can use the table to find the probability you need.

The standard normal (Z) distribution

One very special member of the normal distribution family is called the *standard normal distribution*, or *Z distribution.* The standard normal distribution has a mean of zero and a standard deviation of one; its graph is shown in Figure 9-2. The pdf for the standard normal distribution is the same as that for a regular normal distribution, except that it has $\mu = 0$ and $\sigma = 1$. The pdf for the standard normal (Z) distribution is the following: $\frac{1}{1\sqrt{2\pi}} e^{-\left(\frac{z-0}{1}\right)^2} = \frac{1}{\sqrt{2\pi}} e^{-z^2}$.

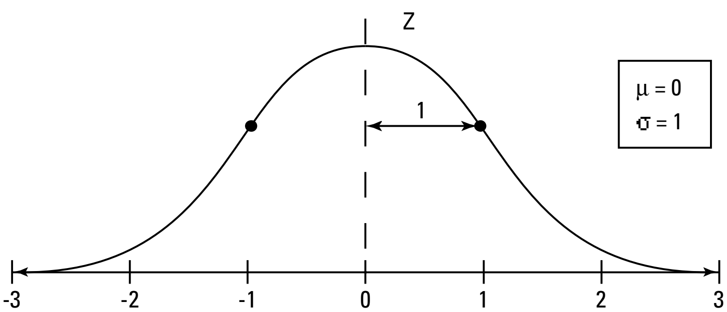


Figure 9-2: The standard normal (Z) distribution is a special normal distribution with mean 0 and standard deviation 1.

Probabilists call the Z distribution the standard normal distribution because it's the standard by which all other normal distributions are measured. The Z distribution is the key to finding probabilities for any normal distribution.

Standard scores

A value on the Z distribution represents the number of standard deviations you are above or below the mean; these are called *z scores* or *z values.* For example, z = 1 on the Z distribution represents a value that is 1 standard deviation above the mean. Similarly, the value z = –1 represents a value that is one standard deviation below the mean (indicated by the minus sign on the z value).

A z score is very helpful for interpreting data; it provides a means to show where a number stands in a data set. Plus, you don't need to know what the original mean and standard deviation were; as long as you know the z score, you can interpret the information. For example, if the doctor tells you your baby's z score for length is +2, that means she's two standard deviations above the mean for babies her age, in terms of her length.

Changing from X units to Z units

To transform from any normal distribution (X) with mean μ and standard deviation σ to a standard normal (Z) distribution with mean 0 and standard deviation 1, you need to do two steps. First, subtract the mean, μ, to get the normal distribution centered at 0 instead of μ, and then divide by the standard deviation, σ, to get the standard deviation distances to be units of 1 instead of units of σ. For example, if X is a normal distribution with mean 16 and standard deviation 4, the value 20 on the X distribution would transform into 20 – 16 divided by 4, or 4 ÷ 4, which is 1. So, the value 20 on the X distribution corresponds to the value 1 on the Z distribution. Similarly, the value 12 on the X distribution translates to Z = 12 – 16 = –4 divided by 4 = –1, so the value 12 on the X distribution corresponds to the value –1 on the Z distribution.

The general formula for changing a value of X into a value of Z is $Z = \frac{(X - \mu)}{\sigma}$.

REMEMBER

The big deal about solving a normal probability problem by using the Z table is that changing from X's original distribution to the Z distribution does not affect the probabilities and hence your answer. In other words, the probability that X is less than x is the same as the probability that Z is less than z when you transform X to Z by using the Z-formula.

How can this happen? Think of it as simply a change in units that's very much the same as changing from Fahrenheit to Celsius. You change temperature units from Fahrenheit to Celsius by using the formula $C = \frac{(F - 32)}{9/5}$ or $\frac{(F - 32)}{1.8}$.

You complete two steps in this formula: the subtraction step and the division step. In other words, if you want to change from Fahrenheit to Celsius, you take the temperature in Fahrenheit (F), subtract 32 degrees, and divide by 1.8.

For example, 32 degrees Fahrenheit changed to Celsius would be (32 – 32) ÷ 1.8 = 0°C. The change in units moves all the temperatures down by 32 degrees and decreases their scale (or spread) by 1.8 degrees per unit. But even though the units of temperature have changed, the temperature itself, as you know it and experience it, stays the same.

Finding probabilities for a normal distribution is too complex using the pdf, so you need to rely on tables of probabilities already calculated for you via computer. However, it's impossible to have a table for every single normal distribution with its own mean and its own standard deviation (your textbook is big enough already, yes?). You have only one normal distribution for which a table of probabilities has been calculated: the standard normal (Z) distribution.

Finding and Using Probabilities for a Normal Distribution

When you study the normal distribution, you come across two major types of problems:

- **Normal probability problems:** This type of problem gives you one or two cutoff values for X and asks you to find the probability of being less than, greater than, or between those values. (The fish problems I present in this section are regular normal probability problems.)
- **Normal percentile problems (or backwards normal problems):** This type of problem gives you the percentile (the probability of being less than or greater than some cutoff value) and asks you to find that cutoff value (find x).

You can identify normal probability problems easily by the fact that you have a normal distribution and they give you a value of x and want you to find the probability below it, above it, or between it and another value of x. (See the section "Handling Backwards Normal Problems," later in this chapter, for examples of backwards problems and how to identify them.)

Here are the steps for finding probabilities in regular normal distribution problems:

1. **Draw a picture of the distribution.**
2. **Translate the problem, using probability notation, into one of the following: $P(X < a)$, $P(X > b)$, or $P(a < X < b)$. Shade in the area on your picture.**
3. **Transform a (or b) into a z value, using the Z-formula: $Z = \frac{(X - \mu)}{\sigma}$.**
4. **Look up the value on the Z Table (see the Appendix).**
5. **If you have a less-than problem, you're done. If you have a greater-than problem, take one minus the result from Step 4. If you have a between-values problem, do Steps 1–4 for b (the larger of the two**

values) and then for a (the smaller of the two values), and subtract the results.

6. **Answer the original question in the context of the problem (in the language of X, not Z).**

Suppose, for example, that you enter a fishing contest. The contest takes place in a pond where the fish have lengths modeled by a normal distribution with mean μ = 16 inches and standard deviation σ = 4 inches. Here are some questions you want to answer (called the three fish problems):

Problem 1: What's the chance of catching a small fish — say, under 8 inches?

Problem 2: Suppose a prize is offered for any fish over 24 inches. What's the chance of catching a fish that size?

Problem 3: What's the chance of catching a fish between 16 and 24 inches?

In the following sections, I walk you through each of the steps so you can find any probability for X on your normal distribution with mean μ and standard deviation σ.

Getting the picture

Before attempting to solve any kind of normal distribution problem, I encourage you to draw a picture of the distribution. Figure 9-3 shows a picture of X's distribution for the three fish problems (Problems 1-3 in the previous section). You can see where each of the fish lengths mentioned in each of the three fish problems falls.

Translating a problem into probability notation

The second step toward solving a probability problem involving a normal distribution is to translate it into probability notation. Each of the three fish problems involves a probability and a value of X that you either have or want to find on the normal distribution:

Problem 1 asks you to find the probability that the fish is less than 8 inches long. Because X represents the length of the fish, you want $P(X < 8)$.

Problem 2 asks you to find the probability that the fish is over 24 inches, so you want $P(X > 24)$.

Problem 3 asks you for a between probability. You want the probability that X is between 16 and 24, so you want P(16 < X < 24).

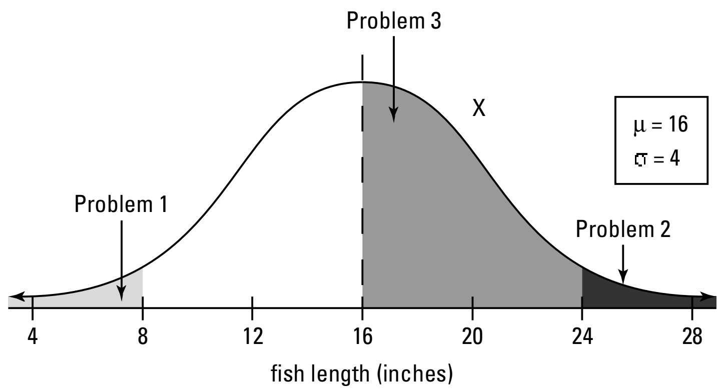


Figure 9-3: The distribution of fish lengths in a pond.

TIP

Notice that it doesn't matter whether you actually include 16 and/or 24 in the previous probability, because X is a continuous random variable. So, P(X = 16) and P(X = 24) are both 0 anyway. Not having to worry about whether or not to include an equal sign in a probability comes easier with continuous random variables compared to discrete ones.

Using the Z-formula

After you translate a problem into probability notation and identify the problem's type, your next step is to transform the problem into one involving the Z distribution so you can use the Z table (see the Appendix) to help you solve the problem.

To change from the units on the original distribution (X units) to units on the standard normal distribution (Z units), you use the formula $Z = \frac{(X - \mu)}{\sigma}$. (See "Changing from X units to Z units" earlier in this chapter.) This formula is known as the *Z-formula.*

For Problem 1 of the fish example, take x = 8 and change it to a z value by subtracting the mean (8 – 16 = –8) and then dividing by the standard deviation (–8 ÷ 4) to get z= –2.

TIP

You can interpret a number on the Z distribution as the number of standard deviations above (if positive) or below (if negative) the mean. A fish 8 inches long translates to z= –2 on the Z distribution, which means fish this small are 2 standard deviations below the mean in length. And given what the Z distribution looks like (see Figure 9-5), not a lot of values are that low.

For Problem 2 of the fish example, you want to change x = 24 into a z score. You subtract the mean and then divide by the standard deviation. Your calculations look like this: $Z = \frac{(X-\mu)}{\sigma} = \frac{(24-16)}{4} = \frac{8}{4} = 2$. So, a 24-inch-long fish is 2 standard deviations above the mean in length (not a lot of fish are that long!).

For Problem 3 of the fish example, you need to change two x values into z values in order to find the probability between them. You change x = 24 to z = +2 when you solve Problem 2 — now you change 16 by taking $Z = \frac{(X-\mu)}{\sigma} = \frac{(16-16)}{4} = \frac{0}{4} = 0$. So, x = 16 becomes z = 0 on the Z distribution. Being between 16 and 24 inches in length (on the X distribution) translates into being between 0 and 2 standard deviations above the mean on the Z distribution.

Figure 9-4 shows a comparison of the X distribution and Z distribution for the values x = 8, 16, and 24, which transform into z = –2, 0, and +2, respectively.

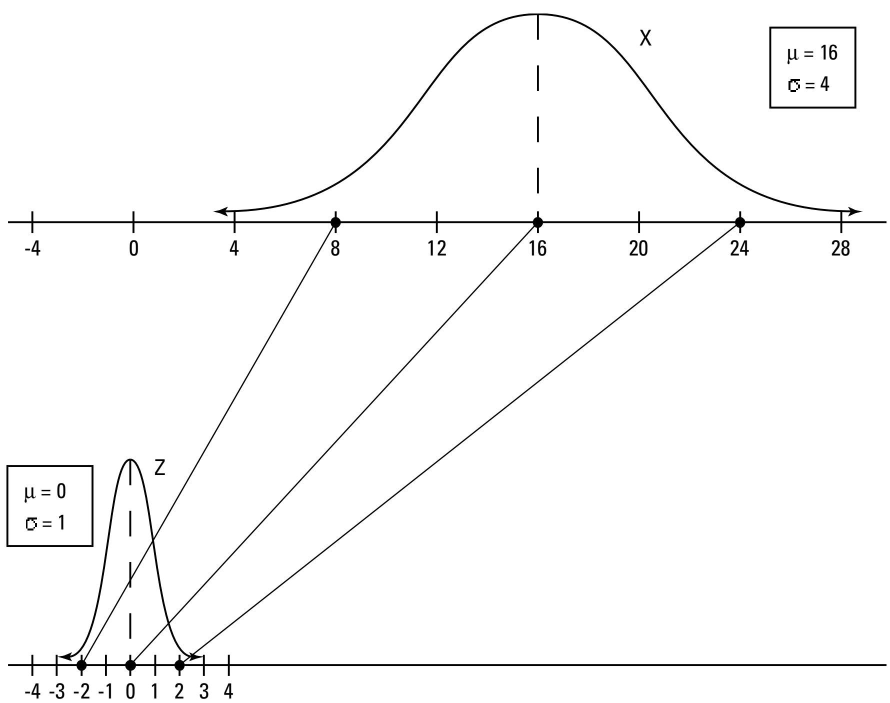


Figure 9-4: Transforming numbers on the normal distribution to numbers on the Z distribution.

Utilizing the Z table to find the probability

After you translate a problem into probability notation (see the section "Translating a problem into probability notation" earlier in this chapter) and transform the appropriate X values into z scores (see the previous section), you're ready (finally!) to find the probability for the problem. You find the probability for a z score by using Table A-2 in the Appendix.

Table A-2 (in the Appendix), commonly known as the Z table, is a table that shows the values of the cdf of a standard normal (Z) distribution for any value of z. The cdf is the cumulative distribution function; it gives you the area under the curve from negative infinity up to z for any value (z) on the Z distribution (refer to Chapter 7 for more on cumulative distribution functions).

In other words, if you look up a value z on Table A-2, you find the area below that value on the Z distribution. (Remember that the total area under any density function is 1.) The result you find is equal to the area below the corresponding value on the X distribution (from the Z-formula; see the section "Using the Z-formula" earlier in this chapter).

Intersecting the rows and columns of the Z table

The Z table contains rows and columns that you use to identify which value of z you want to look at. Each z value on the table has two digits after the decimal point. The rows of the Z table represent the leading digit and the first digit after the decimal point. The columns of the Z table represent the second digit after the decimal point.

To find the probability that Z is less than some value z, using the Z table (Table A-2 in the Appendix), you do the following:

1. **Go to the row that represents the first digit of your z value and the first digit after the decimal point.**
2. **Go to the column that represents the second digit after the decimal point of your z value.**
3. **Intersect the row and column.**

   That number represents $P(Z < z)$.

For example, suppose you want to look at $P(Z < 2.13)$. Find the row for 2.1 and the column for 0.03. (You do this because the second digit after the decimal point is 3. The zero is there to show you that it's in the hundredths place. Notice that if you put 2.1 and 0.03 together as one three-digit number, you get 2.13.) Intersect that row and column to find the number: 0.9834. You find that $P(Z < 2.13) = 0.9834$.

In Problem 1 of the fish example seen in the previous sections, you want P(X < 8). Because x = 8 is equivalent to z = –2, the probability that X < 8 is the same as the probability that Z < –2. Figure 9-5 shows what the two areas look like on the X and Z distributions.

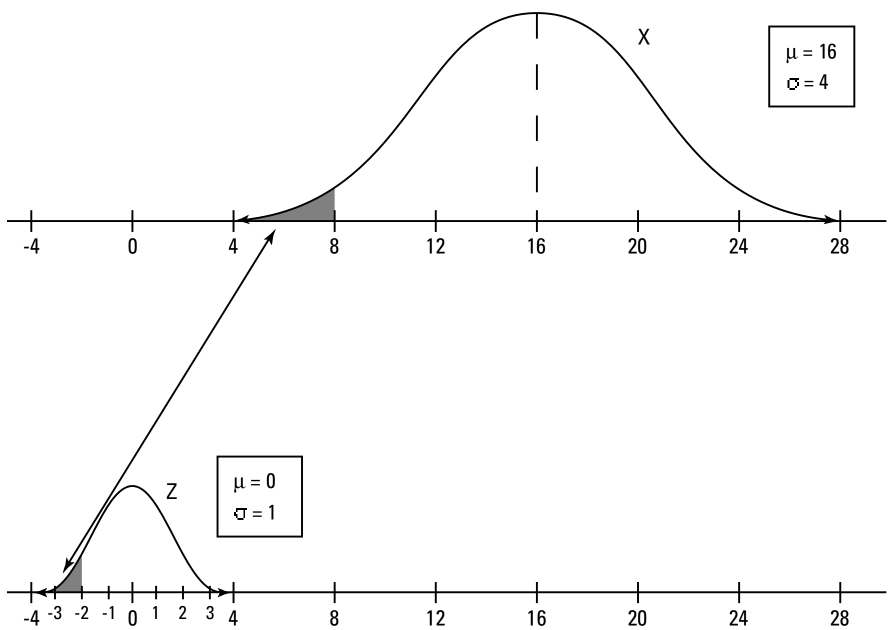


Figure 9-5: Changing from X values to z values doesn't affect the probabilities.

Translated into probability notation and transformed into z scores, the problems from the fish example are the following:

Problem 1: What's the chance of catching a small fish — say, under 8 inches? *Translation:* Find P(X < 8) = P(Z < –2).

Problem 2: Suppose a prize is offered for any fish over 24 inches. What's the chance of catching a fish that size? *Translation:* Find P(X > 24) = P(Z > +2).

Problem 3: What's the chance of catching a fish between 16 and 24 inches? *Translation:* Find P(16 < X < 24) = P(0 < Z < +2).

Less-than probabilities for Z

The Z table shows cumulative (less-than) probabilities for any value from –3.69 to +3.69. Most of the values on a Z distribution are between –3 and +3, so the table covers you for most any value of Z you want to look up. To find

$P(Z < z)$, you look up the value of z on the Z table (Table A-2 in the Appendix). Remember that because Z is a continuous random variable (see the section "Charting the Basics of the Normal Distribution" earlier in this chapter), you don't have to worry about whether to use less than or equal to because the probability of being equal to any number is zero.

Problem 1 asks you to find $P(X < 8) = P(Z < -2)$. Using two digits after the decimal point, your z value is –2.00. Look up the row for –2.0 and the column for 0.00 (the first column), and you find 0.0228. The chance of a fish being less than 8 inches is equal to 0.0228.

In the end, you always go back to answering the question in terms of X, the original units. The z value is just the means to an end. How you got the z value is important work to show, but in the end, make sure you answer the original question (it's all about the fish, not the Z).

For those oddball situations where the z value happens to be larger than 3.69, you can say that the probability of being less than Z is more than .9999 because that's the last probability found on the table. If the z value happens to be smaller than –3.69, you say that the probability of being less than z is less than 0.0000 because that's the first probability on the table (for example, it could be 0.000001).

Greater-than probabilities for Z

To find the probability that Z is greater than a number z, you need to use the complement rule (refer to Chapter 2), because the Z table gives you only less-than probabilities.

Using the Z table (Table A-2 in the Appendix), you take $1 - P(Z < z)$ to get $P(Z > z)$ after you intersect the row and column.

For Problem 2 of the fish example (see the section "Intersecting the rows and columns of the Z table" for a refresher), you need to find $P(X > 24) = P(Z > +2)$. Using two digits after the decimal point, you find that your z value is +2.00. Look up the row for +2.00 and the column for 0.00 (the first column, representing 0.00 after the decimal point of the z value); you find 0.9772. Don't forget to take the complement! So, your final answer is $1 - 0.9772 = 0.0228$. The chance of a fish being more than 24 inches is equal to 0.0228.

$P(Z < -2) = P(Z > +2)$ because of the symmetry of the Z distribution. The two values +2 and –2 are mirror images of each other on the Z distribution, so the probability of being less than –2 is the same as the probability of being greater than +2. Figure 9-6 illustrates this point. This knowledge can save you time on exams, where you can reuse a figure you've already calculated in another problem.

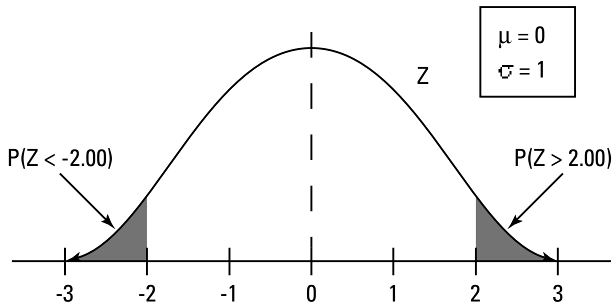


Figure 9-6: The symmetry of the Z distribution.

Between-values probabilities for Z

To find the probability that Z is between two numbers — say, $a < Z < b$ — you can only work with the probabilities that Z is less than b and that Z is less than a, because the Z table gives you only less-than probabilities. But no need to fret. All you have to do is subtract those two probabilities (larger minus smaller) to get your answer, because the area all the way up to b minus the area all the way up to a gives you the area between them. So, you take $P(Z < b) - P(Z < a)$ to find $P(a < Z < b)$. Figure 9-7 shows how you obtain a between-values probability by subtraction.

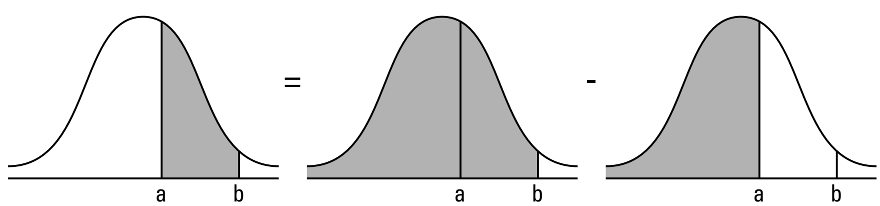


Figure 9-7: Finding a between probability by using subtraction.

For Problem 3 of the fish example (see the section "Intersecting the rows and columns of the Z table" for a refresher), you need to find $P(16 < X < 24) = P(0 < Z < +2)$. Using two digits after the decimal point, you find that your largest z value is +2.00. Look up the row for +2.00 and the column for 0.00 (the first column); you find 0.9772. You have found $P(Z < +2.00)$. You do the same for your smaller z value, 0.00, to get a probability of 0.5000, which is $P(Z < 0.00)$. Now you subtract these probabilities to get $P(Z < +2.00) - P(Z < 0.00) = 0.9772 - 0.5000 = 0.4772$ — the probability that Z is between 0 and +2, and the probability that a fish caught is between 16 and 24 inches. (Not a very big probability, because this is a pretty small range given that the standard deviation is only 4 inches.)

You can also use this method to find probabilities between a and b when a is negative and b is positive. The values of a and b don't matter. To get the probability of X being between a and b, always find $P(X < b)$ and $P(X < a)$ and subtract.

Make sure you do the subtraction for between-values probabilities in the right order. If you subtract $P(Z < b)$ from $P(Z < a)$, you get a negative number, which is not only impossible for a probability, but, if you're in a probability class, also likely to make your instructor pretty upset.

Handling Backwards Normal Problems

Backwards normal problems follow the steps for a regular normal probability problem, only you go backward. I suppose that's why you call them that. (For a rundown of how to solve regular normal probability problems, head to the section "Charting the Basics of the Normal Distribution" and read on.)

You can easily identify a backwards normal problem if you pay close attention to what the problem gives you versus what it asks you to find. In a backwards problem, you're given the probability that X is less than (or greater than) some cutoff value, but you don't know what that cutoff value is; that's what you need to find. In other words, you're given $P(X < a)$ and have to find a; or you're given $P(X > a)$ and have to find a. Regular problems give you the value of a and ask you to find the probability that X is greater (or less) than a.

Following are the steps for solving backwards normal problems:

1. **Translate the problem, using probability notation, into one of the following: $P(X < a)$ or $P(X > b)$.**

   I've never seen a backwards normal problem that goes between two numbers, so you don't have to worry about those.

2. **Find the z value corresponding to a less-than probability by finding the probability in the body of the Z table (Table A-2 in the Appendix) and noting the row and column the probability is in.**

3. **Obtain the z value that corresponds to the probability by taking the row heading plus the column heading and putting them together as one number with two digits after the decimal point.**

   The row heading represents the leading digit and the first digit after the decimal point, and the column heading represents the second digit after the decimal point.

4. **Find the z value corresponding to a greater-than probability by taking one minus that probability and then completing Steps 2–3.**

5. **Transform the z value back into an X value (original units) by using the Z-formula solved for X: $X = Z\sigma + \mu$.**

Suppose that lengths (X) of fish in a pond have a normal distribution with mean 16 inches and standard deviation 4 inches. This is the same setup that you have for the three fish problems in the previous sections. Now I want to add two more questions that require the use of the backwards approach:

> **Problem 4:** What length marks the bottom 10 percent of all the fish lengths in the pond?
>
> **Problem 5:** What length marks the top 10 percent of all the fish lengths in the pond?

In the following sections, you follow through the steps to identify, translate, and work through these backwards normal problems.

Setting up a backwards normal problem

The keys to nailing backwards normal problems are understanding and using the probability notation and drawing pictures.

REMEMBER

You usually see probability statements written like this: $P(X < x) = b$, where x is some value of X and b is some probability. With the backwards normal problems, you're given b (or the complement of it), and you have to find x. In other words, you're given the probability that X is less (or greater) than some cutoff value, and you have to find that cutoff value of X.

For example, you may know that the probability of a number being-less than a certain value is 0.75, and you need to find that number. Translating this into a probability statement, you have $P(X < x) = 0.75$, and you are asked to find x. Figure 9-8 shows a picture of what you are trying to find in this problem.

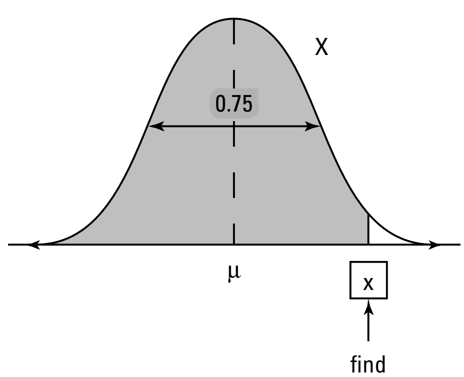


Figure 9-8: Drawing a picture to set up a backwards normal problem: $P(X < x) = 0.75$, find x.

Sometimes the probability portion of the problem is worded more subtly than "the probability that X is less than x is such and such percent." For instance, Problem 4 of the fish example asks you to find what length marks the bottom 10 percent of all the fish in the pond. It sounds rather vague, but the problem is giving you a probability (0.10) and wants you to find the cutoff value (so you need to find x).

But what does that probability of 0.10 actually represent? Figure 9-9 shows the distribution of all fish lengths (X's normal distribution) and the spot on the distribution that marks the bottom 10 percent of the fish. I took a guess as to where this value actually sits on the distribution by figuring that with only 10 percent probability lying below it, you have to be out on the left-tail area somewhere.

So, in Problem 4 of the fish example, you want to find the value on the X distribution (call this value x) where the probability of being less than that value is 0.10. In other words, you want to find x where $P(X < x) = 0.10$. You have now successfully translated the problem, and you know exactly what it means and what you have to find. Figure 9-9 shows a picture of what you need to find in this problem.

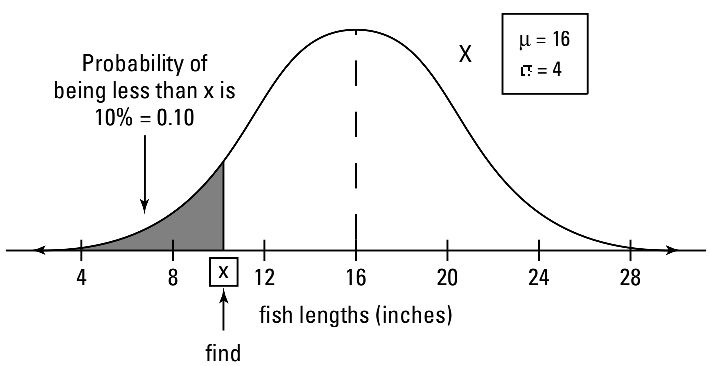


Figure 9-9: Bottom 10 percent of fish in the pond, according to length.

In Problem 5, you want to find the length that marks the top 10 percent of the fish. That puts you in the right tail (upper area) of the distribution, with only 10 percent of the probability to the right of that point and 90 percent of the probability to the left of it. Figure 9-10 shows you a picture of this scenario. This can translate into two different probability statements. First, the probability to the right of x is 0.10, stated $P(X > x) = 0.10$. Second, the probability to the left of x is $1 - 0.10 = 0.90$, stated $P(X < x) = 0.90$. Because the Z table operates based on less-than probabilities only, the second version of this probability statement is the one you will be using.

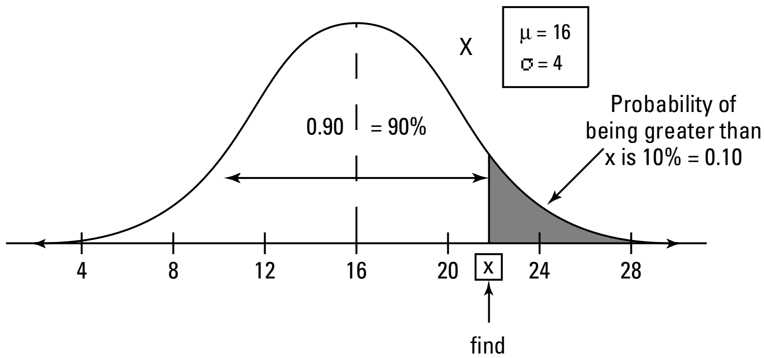


Figure 9-10: Top 10 percent of fish lengths in the pond.

Using the Z table backward

The Z table (see Table A-2 in the Appendix) contains rows and columns that you use to identify which value of Z you want to look at; the values in the body of the table represent P(Z < z) for any z value you look up. For regular normal probability problems, you look up the z value (because you know it) and find the probability that corresponds to it. For backwards normal problems, you do the reverse: You look up the probability in the body of the table (because you know it), and then you find the corresponding z value by figuring out which row and column it's in.

Finding Z given a less-than probability

If you're given the probability that Z is less than some value z, and you want to find z by using the Z table, follow these steps:

1. **Find the probability that you're given in the body of the Z table.**
2. **Look across to determine which row it's in.**

   That represents the first digit of your z value and the first digit after the decimal point.
3. **Look up to see which column it's in.**

   That represents the second digit after the decimal point of your z value.
4. **Put the numbers from Steps 2 and 3 together to form a number with two digits after the decimal point.**

   This number is your z value corresponding to the less-than probability that the problem gives you.

Suppose you know the probability of being less than Z is 0.9834, and you want to find Z. You go into the body of the Z table and look for the probability

0.9834 (or get as close as possible). After you find that value, look across to see that it's in row 2.10. Now look up to see that it's in column 0.03. Put those values together to get z = 2.13. So, the value of z = 2.13 is the cutoff point where 98.34 percent of the values lie below it. In other words, $P(Z < z) = 0.9834$ means z = 2.13.

In Problem 4 of the fish example, you want to find x where $P(X < x) = 0.10$. You know that on the Z table (Table A-2 Appendix), the probability closest to 0.10 is 0.1003, which falls in the row for z = –1.20 and the column for 0.08. That means the z value corresponding to a probability of 0.10 is z = –1.28. So, a fish that's at the bottom 10 percent is at the tenth percentile and is 1.28 standard deviations below the mean in terms of its length.

Don't add the row and column values together in the case of negative z values. Just tack on the column heading value as the second digit after the decimal point. For example, if the probability is in row –1.2 and column 0.08, your z value is –1.28, not –1.2 + 0.08 = –1.12.

Finding Z given a greater-than probability

If you're given the probability that Z is greater than some value z, and you want to find z by using the Z table, follow these steps:

1. **Take one minus the probability that you're given, and find that probability in the body of the Z table.**
2. **Look across to determine what row the probability is in.**

   The number at the head of the row represents the first digit of your z value and the first digit after the decimal point.
3. **Look up to see what column the probability is in.**

   The number at the head of the column represents the second digit after the decimal point of your z value.
4. **Put the two numbers from Steps 2 and 3 together to form a number with two digits after the decimal point.**

   The result is your z value corresponding to the greater-than probability that the problem gives you.

Suppose that you know the probability of a number being greater than z on the Z distribution is 0.9834, and you want to find z. Figure 9-11 shows a picture of what you are trying to find in this situation.

You can't just go into the body of the Z table and look for the probability 0.9834 — that would give you the z value for a less-than probability, not a greater-than probability. How do you move from a greater-than probability to a less-than probability? You find the complement.

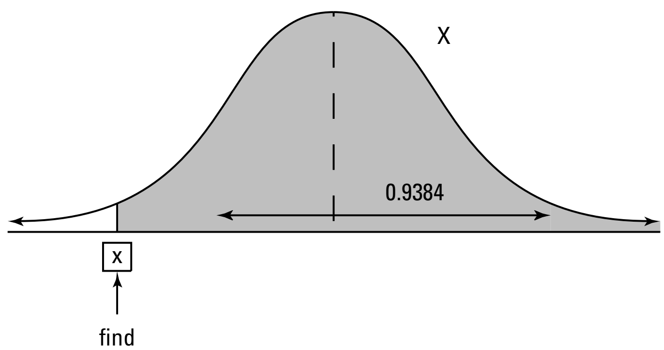


Figure 9-11: Find z where $P(Z > z) = 0.9834$.

Because you know $P(Z > z) = 0.9834$, you know that $P(Z < z) = 1 - 0.9834 = 0.0166$. The z value for each of these probabilities is the same, so you may as well use the one that works for the Z table. So, look for 0.0166 in the body of the Z table. From there, you follow the instructions from the previous list. After you find your value, look across to see that it's in row –2.10. Now look up to see that it's in column 0.03. Finally, put those values together to get $z = -2.13$. So, the value $z = -2.13$ is the cutoff point where the probability of lying below it is 0.0166 (or 1.66 percent) and the probability of lying above it is 0.9834 (or 98.34 percent). In other words, $P(Z > z) = 0.9834$ means $z = -2.13$.

Problem 5 of the fish example asks you to find x where the probability to the right of x is 0.10. For Step 1, notice that this is the same as asking to find x where the probability to the left of x is $1 - 0.10 = 0.90$. In other words, find x where $P(X < x) = 0.90$. In Step 2, find the probability closest to 0.9000 (which is 0.8997) and find the row that it falls in ($z = 1.20$). Step 3 says to find the column this probability is in; it falls in column 0.08. For Step 4, put these digits together and get a z value of 1.28. So, you know that a fish whose length is surpassed by only 10 percent of the fish is 1.28 standard deviations above the mean in terms of its length. (It's also at the 90th percentile, because 10 percent of the fish being longer is equivalent to 90 percent of the fish being shorter.)

Returning to X units, using the Z-formula solved for X

After you find the corresponding z value that goes with the probability you're given (or its complement) in a backwards normal problem, the last step is to transform the z value back into the original units, X. You perform this step so you can answer the original question in the context of the problem.

Changing from Z back to X is the same as changing a temperature from Celsius back to Fahrenheit; you do the steps backward. To go from X to Z, you subtract the mean and divide by the standard deviation. To go from Z to X, you multiply by the standard deviation and add the mean.

The formula for changing from X to Z is given by $Z = \frac{(X - \mu)}{\sigma}$. The formula for going backward from Z to X is $X = Z\sigma + \mu$. Using some algebra, you can rewrite this formula so it's solved for X (in other words, so it's in the form "X = . . ." instead of "Z = . . ."). The algebra shows you the following: $Z = \frac{(X - \mu)}{\sigma} \rightarrow Z\sigma = (X - \mu) \rightarrow Z\sigma + \mu = X \rightarrow X = Z\sigma + \mu$. You can call this the *Z-formula solved for X.*

Problem 4 of the fish example asks you to find out how long a fish is that is at the bottom 10 percent. In the previous section, you find that the z value for this fish is –1.28, and that its length is 1.28 standard deviations below the mean. Changing this value back to X units (actual lengths in inches), you use the Z-formula solved for X. The mean and the standard deviation of the fish in this pond are 16 inches and 4 inches, respectively. What's the length of that fish? Using the Z-formula solved for X, you get $x = -1.28 * 4 + 16 = 10.88$ inches. A fish 10.88 inches long marks the bottom 10 percent of fish lengths in the pond. You can see in this formula that you take 1.28 standard deviations (each worth 5) subtracted from the mean (16), so it makes sense that this fish's length would correspond to a z score of –1.28.

Problem 5 of the fish example asks you to find out how long a fish is that is at the top 10 percent. In the previous section, you find that the z value for this fish is +1.28 and that its length is 1.28 standard deviations above the mean. To change this value back to X units (actual length in inches), you again use the Z-formula solved for X to get $x = +1.28 * 4 + 16 = 21.12$ inches. A fish 21.12 inches long is at the 90th percentile for length; only 10 percent of the fish are longer. You can see in this formula that you take 1.28 standard deviations (each worth 5) added to the mean (16), so it makes sense that this fish's length would correspond to a z score of +1.28.

Chapter 10

Approximating a Binomial with a Normal Distribution

In This Chapter

- Using a normal distribution to approximate binomial probabilities
- Knowing when you can (and should) approximate a binomial
- Judging the sample and figuring the mean and standard deviation of the binomial
- Adding a continuity correction to the binomial

The binomial model has a unique set of formulas for individual and cumulative probabilities (see Chapter 8), but if the number of fixed trials, *n,* gets too large, those probabilities can get extremely tedious and time consuming to calculate (plus, the binomial tables for which probabilities are already provided [see the Appendix] run out after n gets beyond 25 or so, because of the length of the number of entries you need at that point). Researchers have found a nice solution to this problem in the form of an *approximation.* They found that you can use the normal distribution (see Chapter 9) to get approximate answers to binomial probability problems when n is large. And when you add in a continuity correction to account for the fact that you're moving from a discrete distribution (one with a finite number of possible values) to a continuous distribution (one with an uncountably infinite number of possible values), the approximation is pretty darn close (close enough for government work anyway).

This chapter shows you when and how to approximate binomials with the normal distribution and how to add in a continuity correction to make the approximations close.

Identifying When You Need to Approximate Binomials

The *binomial probability distribution* models situations where you have a certain number of fixed trials (n), and each trial can take only two possible outcomes: success or failure. The probability of success is p and the probability of failure is 1 – p. All the trials are independent, meaning that their outcomes don't influence each other (see Chapter 8 for more on the binomial distribution). The binomial distribution is commonly used because of its interest in the probabilities for success or failure outcomes, such as the probability of winning the lottery (versus not), getting in a car accident (versus not), or being on hold for more than 10 minutes (versus not). The distribution can also be used to count the total number of successes in n trials and find those probabilities — for example, the chance that more than 30 people in a sample of 50 like to watch reality television.

Finding probabilities with the binomial distribution follows an old nursery rhyme: When she's good, she's very, very good, but when she's bad, she's horrid. What I mean is that you face certain situations where using the binomial formulas just doesn't cut it.

Suppose, for example, that you flip a fair coin 100 times, and you let X equal the number of heads that come up. The probability distribution for X is binomial, with n, the number of trials, = 100 and p, the probability of success, = 0.50. The possible values of X are 0 all the way up to 100. You decide you want to figure out the probability that X equals 60. Sounds easy enough; you just pull out the formula for the probability of the binomial and plug in 100 for n, 60 for x, and 0.50 for p. What you get is $\binom{100}{60} 0.50^{60}(1-0.50)^{100-60} = \frac{100!}{60!40!} 0.50^{60} 0.50^{40}$.

In this formula, you notice the symbol ! after the 100; the symbol indicates a 100 factorial, which is shorthand notation for taking 100 * 99 * 98 * 97, and so on, all the way down to 3 * 2 * 1. If the number you want the factorial for is small, like 5!, you don't have a bad calculation — 5 * 4 * 3 * 2 * 1 = 120, and most calculators can do it. But what's the 100 factorial equal to? Who knows? It's so large that most calculators can't even touch it; they think for a while and then give up with an "error" message. In other words, the calculator says, "I surrender." So, although the formula looks simple enough, actual calculations can be a real bear. There has to be a better way, and there is. The better way is to use an approximation that's easy to calculate, yet still gets an answer very close to the actual value.

Why the Normal Approximation Works when n Is Large Enough

When a binomial probability distribution (or a histogram of all the probabilities) involves a large number of fixed trials where p, the probability of success, is close to 0.50, you have what's known as a *symmetric situation.* And when p is closer to zero or one than to 0.50, you have a *skewed situation.* In the end, any binomial distribution heads toward a normal distribution as the number of trials goes up, but if p is 0.50, the binomial gets there faster (because it's already symmetric in shape, looking the same on each side if you cut it down the middle). If p is closer to zero or one, the binomial isn't symmetric and takes a higher value of n to look like a normal distribution. In this section, you examine the behavior of symmetric and skewed binomial probability distributions.

Symmetric situations: When p is close to 0.50

When you have a large number of fixed trials, binomial distributions look a lot like normal distributions (whose probabilities are easy to calculate; see Chapter 9). Probabilities are small for extremely large or small X values because you're unlikely to get 90 heads or 10 heads on 100 flips of a fair coin, for example. Most of the probability piles up around the middle at the expected value, which happens to be $n * p = 100 * 0.50 = 50$.

The variance for the 100 flips of the coin is equal to $n * p * (1 - p)$, which is $100 * 0.50 * 0.50 = 25$. The standard deviation is the square root of the variance, which is 5. (For the full details of the binomial distribution, see Chapter 8.)

Figure 10-1 shows a picture of this probability distribution for X. Notice that most of the values that occur with any probability are between 35 and 65, which is three standard deviations on either side of the mean (another characteristic of the normal distribution).

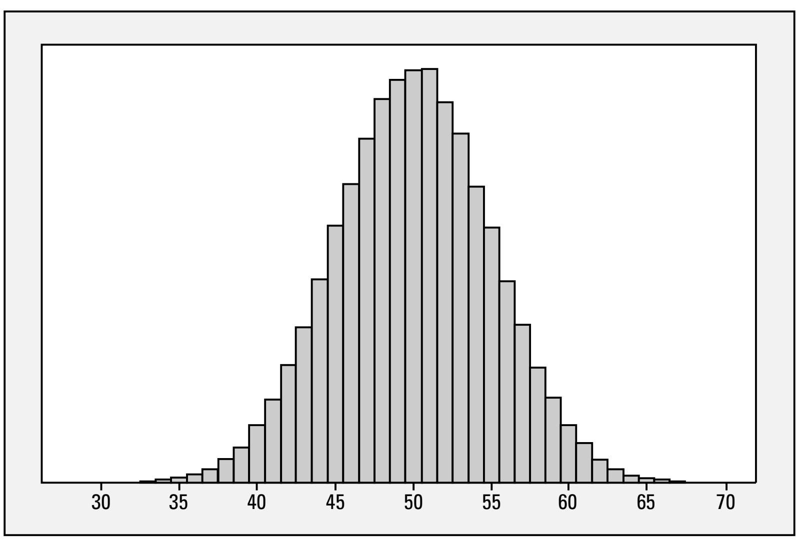


Figure 10-1: Probability distribution for X = number of heads on 100 flips of a fair coin.

Suppose that you increase the number of flips to n = 1,000. The graph of the probability distribution for X = number of heads is shown in Figure 10-2. If there's such a situation as being "even more normal," you can say that Figure 10-2 looks even more normal than Figure 10-1, because the number of trials is larger. As n gets larger and larger, the normal approximation to the binomial gets better and better, and the only adjustment you need to bring the two distributions into perfect alignment is a small continuity correction (see the section "Making the continuity correction" later in this chapter). That's good news, because finding the probabilities for a normal distribution is so much faster than finding probabilities for a binomial when n is large. Notice that in Figure 10-2, the mean is 500, which is equal to E(X), the expected value. The standard deviation, SD(X), is equal to $\sqrt{n * p * (1 - p)} = \sqrt{1{,}000 * 0.5 * (1 - 0.5)} = 15.81$.

Skewed situations: When p is close to zero or one

REMEMBER

When the number of fixed trials is small and the probability of success is close to zero or one, skewness occurs because the expected value is close to one of the edges, which cuts short one whole side of the distribution. But as you increase the number of trials, you have many more possible values of X, so the probability has room to spread out among those different values. For example, when n is ten, the number of successes can be anywhere from zero to ten. But if n is 100, X can take on values anywhere from 0 to 100; you see many more possible values for X, so X looks more and more like a continuous probability distribution as n gets large (say, in the thousands). A *continuous probability distribution* occurs when the values of X are uncountable or when there are so many possible values of X that it may as well be uncountable (see Chapter 2).

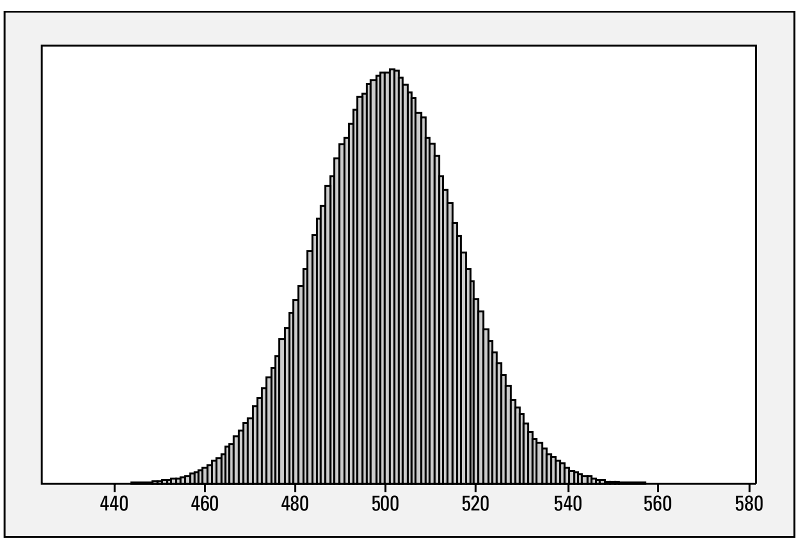


Figure 10-2: Probability distribution for X = number of heads on 1,000 flips of a fair coin.

With the coin example I focus on in the previous section in mind, what would happen if the coin you flip isn't fair? If p = 0.50, all the binomial distributions are symmetric, so it makes sense that the normal would fit them. However, if you flip an unfair coin, the binomial distribution won't be symmetric, will it? The answer is yes, if your n is large enough.

Look at the situation where p = 0.80 — the chance of getting a head is 80 percent, and the chance of getting a tail is 20 percent. If you flip the unfair coin ten times, the distribution of X = number of heads is skewed to the left (see Chapter 7) because heads are more likely to come up. And because you flip the coin only ten times, ten heads is quite likely. Figure 10-3a shows the graph of this probability distribution for X. Note that the mean (the balancing point in the graph) is equal to $n * p = 10 * 0.80 = 8.0$.

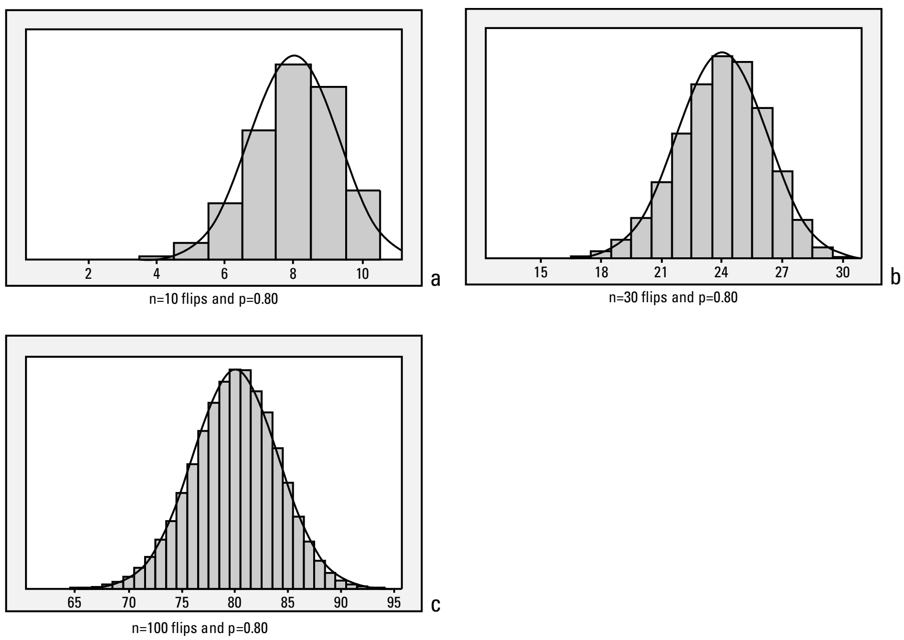


Figure 10-3: Binomial probability distributions for p = 0.80 and increasingly large values of n (10, 30, and 100).

Figure 10-3b shows what happens if you keep p at 0.80 and increase the number of flips to n = 30. You still see the skewness, although it's less noticeable than it is in Figure 10-3a. The distribution seems to be shifting so that more of it appears in the picture than in Figure 10-3a.

Now suppose you increase the number of flips to 100. The graph of the probability distribution for this X is shown in Figure 10-3c. What a difference in the shape! This graph is shaped like a normal distribution, with no skewness in sight, yet p is still 0.80. The only factor that changes is the number of trials. How did this progression of shapes happen? Why did the skewness seem to disappear? You can attribute it to the increased sample size.

Most probability instructors don't go too deeply into explaining all the "whys" of skewed situations, and neither will I, but here's the bottom line: When n is small and p is close to zero or one, skewness cuts short one whole side of the distribution where the "tail" would normally be decreasing as you move away from the expected value (see Figure 10-3a). Remember, the number of successes or failures can't be outside of zero or n — the edges of the distribution — and that's what cuts it short. But as you increase the number of trials, you have many more possible values of X than you do when n is only 10. So now the tails show up on either side of the distribution, decreasing on either side as you move away from the expected value. You start to see this happening in Figure 10-3b, and by the time n = 100 (in Figure 10-3c), the skewness is all but gone.

Understanding the Normal Approximation to the Binomial

Because a binomial distribution looks like a normal distribution when the total number of trials, n, is large enough, you can use a normal distribution to approximate the probabilities you need. In Chapter 9, you see how easy it is to calculate probabilities for a normal distribution: You just transform your X value to a Z value by subtracting the mean and dividing by the standard deviation, and you use the Z table to find the probability (see Table A-2 in the Appendix). Now only three questions remain:

- When is n large enough to approximate by using a normal distribution? If n is large enough, you don't have to do tons of calculations to get a less-than or greater-than probability.
- What mean and standard deviation do you use to transform your X value into a Z value? Because the original model is binomial, you start with the mean and variance of the binomial and work from there.
- How do you make the continuity correction to adjust for the fact that the binomial distribution is discrete and the normal distribution is continuous? You just add or subtract a small amount of probability by moving your desired value of X up or down by ½.

Determining if n is large enough

Determining how large n has to be for the normal approximation to work well depends on the value of p, the probability of success on a given trial. Overall, the bigger the number of fixed trials, the better the approximation works. As you see in the section "Why the Normal Approximation Works when n Is Large Enough," if p is close to 0.50, the binomial distribution looks pretty bell-shaped, so a sample close to 20 is fine. However, if p is close to zero or one, you see skewness in the distribution, and the only way to handle it is with a big value of n — a value much higher than 20.

The general rule of thumb for determining if n is large enough is the following: You can use the normal distribution to approximate the binomial distribution as long as $(n * p) > 5$ and $[n * (1 - p)] > 5$. And the larger the value of n, the better the approximation will get.

For example, say you flip a fair coin 100 times, and you let X equal the number of heads. The probability distribution for X is binomial, which you know because you have a fixed number of independent trials (because one flip doesn't affect another). And because the coin is fair, the chance for success (getting heads) on any given trial is ½ = 0.50. In other words, X is binomial with $n = 100$ and $p = 0.50$. You want to figure out the probability that X is greater than 60. Can you use the normal approximation here? Check the calculations: $n * p = 100 * 0.50 = 50$, which is greater than 5; and $n * (1 - p) = 100 * (1 - 0.50) = 50$, which is greater than 5. So yes, you're safe to use the normal approximation in this case.

You may be wondering why you have to check both conditions: $(n * p) > 5$ and $[n * (1 - p)] > 5$. Wouldn't just one suffice? Suppose that the coin you flip is unfair, and p is 0.99. If you flip 100 times, you would have $n * p = 100 * 0.99 = 99$, which is greater than 5, but $n * (1 - p) = 100 * (1 - 0.99) = 100 * 0.01 = 1$, which isn't greater than 5. So, in a case such as this, you need a larger n (500 or more) in order to use the normal approximation. The same would happen if p is only 0.01.

Finding the mean and standard deviation to put in the Z-formula

For any binomial distribution that meets the size conditions of n ($n * p$ is greater than 5 and $n * [1 - p]$ is greater than 5), you can calculate the probabilities you need by using the steps for calculating probabilities from a normal distribution (see Chapter 9). This involves transforming your X value to a Z value, using the Z-formula, by subtracting the mean and dividing by the standard deviation. Using probability notation, you have $Z = \frac{X - \mu}{\sigma}$.

With a regular normal distribution problem, you often have to be given the mean and standard deviation of the population in the problem. However, because X is binomial to start with, you use the mean and standard deviation of the binomial in your Z-formula. In other words, you use $E(X) = n * p$ for the mean μ, and you use $\sigma = \sqrt{n * p * (1 - p)}$ for the standard deviation. For example, if you flip a fair coin 100 times, and X is the number of heads (so $n = 100$ and $p = 0.50$), the mean is $\mu = n * p = 100 * 0.50 = 50$, and the standard deviation is $\sigma = \sqrt{n * p * (1 - p)} = \sqrt{100 * 0.5 * (1 - 0.5)} = \sqrt{25} = 5$.

Making the continuity correction

The normal approximation to the binomial is just what is says — an *approximation* — so before you move forward with your problem after you transform your X value into a z value and use the Z table (see the Appendix) to find your probability (see the previous section to find out how), you need to make an adjustment to get a close approximation. The adjustment is called a *continuity correction* — a correction you make when moving from a discrete distribution like the binomial to a continuous distribution like the normal (see Chapter 7 for more on discrete and continuous distributions). If you don't make the adjustment, your final answer will be a little larger or a little smaller than it should be.

Picturing why the binomial needs the continuity correction

When you use a normal distribution to approximate a binomial, you're basically smoothing the edges of the bars on the binomial distribution with a continuous curve. Notice that in Figures 10-3a, 10-3b, and 10-3c, I include the curve in each case. If you look at the figures closely now, you can see that the difference between the bars and the curve is the amount of error that you have when you do the normal approximation.

In particular, look at the bar for X = 27 in Figure 10-3b. Notice that on the right side of that bar, the curve cuts below the bar, so it's underestimating the probability over that area. On the left side of the bar, the curve cuts above the bar, overestimating the probability over that area. You may think that these probabilities somehow cancel out, but probability researchers have found that in general, the underestimation part outdoes the overestimation part — your overall answers are less than they should be. To fix this problem, they came up with the continuity correction.

Walking through the steps of the continuity correction

To make the continuity correction, you basically add a little more probability by taking your cutoff point and moving it over an amount equal to one half of a unit. Suppose that your binomial distribution is graphed so that each value of X has its own bar, and the bars are all next to each other with no gaps. Because your cutoff point technically is sitting right in the middle of a bar, when you move to a continuous curve, you normally find the probability right up to that point. But because this probability is too small compared to the true value, you move the cutoff point over by half to include the area for the whole bar, not just the area up the middle of it.

Here are the steps you take to make a continuity correction:

1. **Write down the probability you want to find.**

   For example, $P(X < 27)$.

2. **Draw a picture of the binomial distribution with a curve through it (similar to my curves in Figures 10-4 through 10-7) and locate your X value on the picture to see what part of the histogram you want to include in your probability.**

   For example, if you want $P(X < 27)$, the value of X is 27.

3. **Add or subtract one-half, depending on whether you have a greater-than or equal-to or less-than or equal-to probability to find.**

   Make the following adjustment to your value of X based on the probability:

   a. If you have a less-than or equal-to probability, add ½ to your value of X to add probability. For example, if you have an X less than or equal to 5, the continuity correction for 5 becomes 5.5.

   b. If you have a greater-than or equal-to probability, subtract ½ from your value of X to add probability. For example, if you have an X greater than or equal to 2, the continuity correction for 2 becomes 1.5.

   c. If you have a probability between two values — for example, X is between 2 and 5, including 2 and 5, which means that X is greater than or equal to 2 and less than or equal to 5 — you go through Steps 1, 2, and 3a to apply the continuity correction to the number 5 (so the continuity correction for 5 becomes 5.5), and then you go through Steps 1, 2, and 3b to apply the continuity correction to the number 2 (so the continuity correction for 2 becomes 1.5).

   d. If you have a probability of being equal to a number, apply the continuity correction on each side of the number by adding and subtracting half. For example, you approximate $P(X = 3)$ by $P(2.5 \leq X \leq 3.5)$. (For another example of this, see the last section in this chapter.)

   e. If you have a strictly less-than probability, rewrite it as a less-than or equal-to probability and change the value of X to X – 1. For example, you make $P(X < 27)$ equal to $P(X \leq 26)$ and then go to Step 3a to make the continuity correction for a less-than or equal-to probability.

   f. If you have a strictly greater-than probability, rewrite it as a greater-than or equal-to probability and change the value of X to X + 1. For example, you make $P(X > 27)$ equal to $P(X \geq 28)$ and then go to Step 3b to make the continuity correction for a greater-than or equal-to probability.

Applying the continuity correction

In Figure 10-3b, suppose you want $P(X \leq 27)$. Without the continuity correction, the probability would stop right in the middle of the bar where 27 is. But with the correction, you go up a little farther to 27.5, the end of the bar, and calculate the area under the normal curve up to that point. This adds a little bit of probability to your answer and gets you closer to the true binomial probability. So, to approximate $P(X \leq 27)$, using a normal distribution, you find $P(X \leq 27 + 0.50) = P(X \leq 27.5)$. Figure 10-4 shows this area in the shaded region.

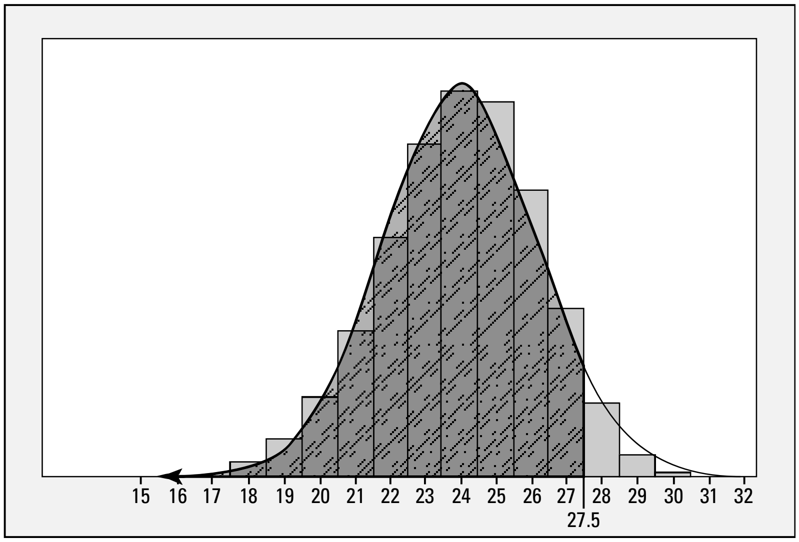


Figure 10-4: Approximating P(X ≤ 27), using the continuity correction.

Now suppose you want to find P(X ≥ 27). The continuity correction says to add a little more probability by moving the cutoff point over by half to include the probability for the entire bar that 27 is located in. Because you want the probability of X being 27 or beyond, you move down from 27 to 26.5 to catch a little more probability. Figure 10-5 shows this area in the shaded region. So, to find P(X ≥ 27), you approximate it with P(X ≥ 26.5).

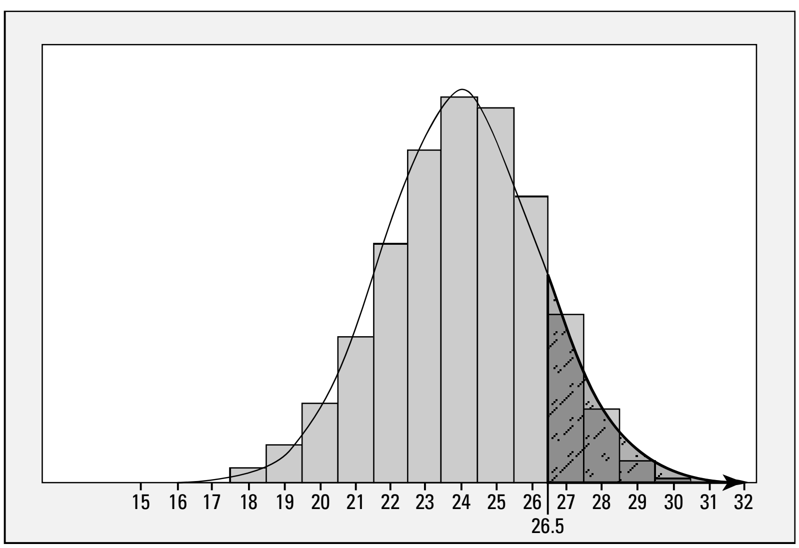


Figure 10-5: Approximating P(X ≥ 27), using the continuity correction.

REMEMBER

Don't just close your eyes and guess whether to add or subtract half for the continuity correction. On the other hand, you shouldn't freak out trying to memorize the steps — it will make you crazy. Just draw a picture of what you have to find, and you'll immediately see what you need to do. Remember that you want to move over a half a spot to catch a little more probability and rewrite everything as $<$ or $>$ before you start.

Approximating a Binomial Probability with the Normal: A Coin Example

Following are the steps to find a binomial probability by using the normal approximation. Referring to these seven steps, you can answer questions such as the following: Suppose you flip a fair coin 100 times, and you let X equal the number of heads; what's the probability that X is greater than 60? (See Chapter 9 for more on this process.)

Follow these steps to find the probability:

1. **Check the two conditions $(n * p) > 5$ and $[n * (1 - p)] > 5$. If one or both of these conditions aren't met, you can't use the normal approximation.**

   If the conditions are met for n, you can go ahead and find the probability by using the normal approximation.

2. **Write down what you need to find as a probability statement about X.**

   For the coin-flipping example, $P(X > 60)$.

3. **Draw a picture of the distribution, including a normal curve, and find the cutoff value for X on the picture. Shade in the area you want to find.**

   For the coin-flipping example, see Figure 10-6.

4. **Apply the continuity correction.**

   Use the steps outlined in the previous section, "Making a continuity correction."

   Because the coin problem is a greater-than probability, you first rewrite it as $P(X \geq 61)$ and then subtract 0.50 from X to get $P(X \geq 61 - 0.50) = P(X > 60.5)$. This is the probability you really need to find. Notice that this matches what Figure 10-6 shows.

5. **Transform the (adjusted) X value to a z value, using the Z-formula: $Z = \frac{X - \mu}{\sigma}$. For the mean, use $\mu = n * p$, and for the standard deviation, use $\sigma = \sqrt{n * p * (1 - p)}$.**

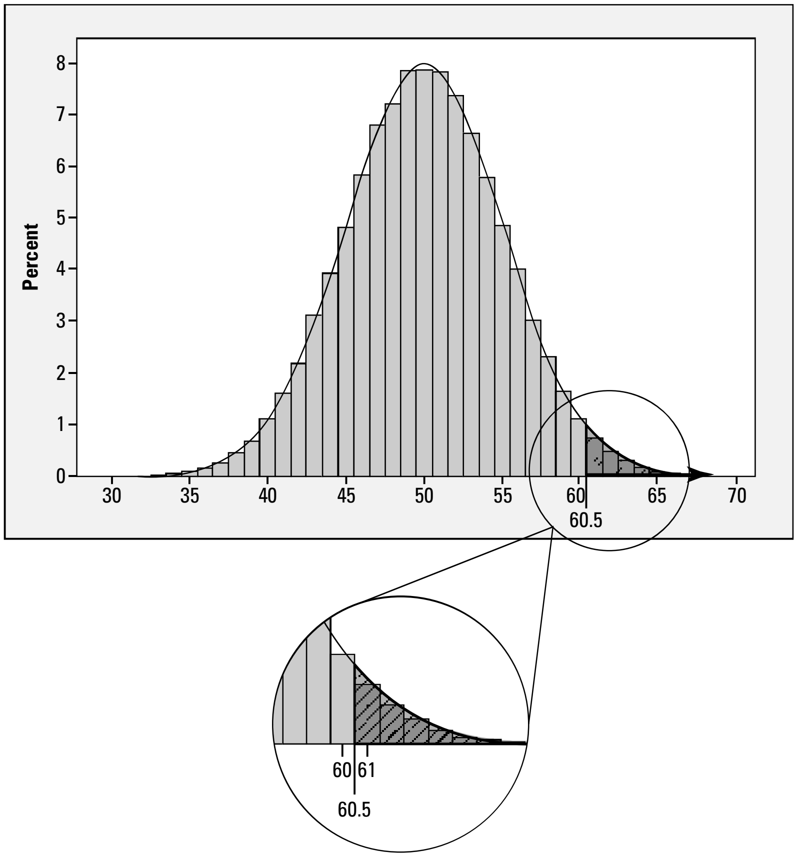


Figure 10-6: The probability of (X > 60) on 100 flips of a fair coin.

The mean is $\mu = n * p = 100 * 0.50 = 50$. The standard deviation, $\sigma = \sqrt{n * p * (1 - p)} = \sqrt{100 * 0.5 * (1 - 0.5)} = \sqrt{25}$, is 5. Put the values into the Z-formula to get $P(X \geq 60) = P\left(Z \geq \frac{60.50 - 50}{5}\right) = P(Z \geq 2.10) =$ $P(Z > 2.10)$. [Note that the probability that Z equals 2.10 is zero because it is continuous (see Chapter 9). So $P(Z \geq 2.10) = P(Z > 2.10)$.)]

6. **Find the probability corresponding to the z value on the Z table (see Table A-2 in the Appendix).**

   This is the probability that Z is less than or equal to that value. You look up the z value of 2.10 and find $P(Z < 2.10) = 0.9821$.

   REMEMBER

   The less-than probability is not what you want ultimately, but it's the only type of inequality the table gives you a probability for, so you find this probability for now and work with it in the next step.

7. **If you need a less-than or equal-to probability, you're done. If you need a greater-than or equal-to probability, you take one minus the probability in the Z table. If you need a probability between two values, you do Steps 4-6 for each value and subtract the probabilities (largest minus smallest).**

   You want a greater-than probability for the coin example, so $P(Z > 2.10) = 1 - 0.9821 = 0.0179$ by the complement rule (refer to Chapter 2). So, the chance of getting more than 60 heads in 100 flips of a fair coin is approximately 0.0179, or approximately 1.79 percent.

   When you use the normal approximation to find a binomial probability, your answer is an approximate answer, so be sure you state that.

For another example, suppose you want to approximate the probability that $X = 60$. Using the continuity correction, you find $P(59.5 \le X \le 60.5)$; see Figure 10-7. You transform 60.5 to a z value to get $Z = \frac{X - \mu}{\sigma} = \frac{60.50 - 50}{5} = 2.10$, and then you transform 59.5 to a z value to get $Z = \frac{59.50 - 50}{5} = 1.90$. Now you look up the probabilities and subtract them to get $0.9821 - 0.9713 = 0.0108$. So, the probability that you get exactly 60 heads when you flip a fair coin 100 times is approximately 0.0108, or 1.08 percent. (This probability is so small because of the need for it to be exactly 60 — not 59, not 61, but 60.)

Finding probabilities for large surveys

Oftentimes, a survey will ask a question that involves a yes or no answer. Do you think that the president is doing a good job? Do you own a home? And in most cases, the number of people sampled is large, anywhere from 500 to 5,000. In order to analyze yes or no survey questions with a large sample size, you normally use a binomial distribution because you have n number of fixed trials (equal to the sample size); the trials are independent because you use a random sample (meaning that every participant had an equal chance of being selected and the results don't affect each other); and the probability of success equals the proportion of people who would say yes to the question (which is estimated by the proportion of people who say yes to the question from your sample). However, because the sample sizes are too large to use the binomial formulas to calculate probabilities, you use a normal approximation instead. In other words, data analysts use a normal approximation to approximate the probabilities.

Suppose, for example, a television ad claims that 50 percent of people own a cell phone. You believe the percentage is higher, so you select 2,000 people at random and ask them whether they own cell phones. You find that 1,100 say yes, which is more than 50 percent (55 percent). What's the chance of this happening if the assumption of 50 percent is true? To find this probability, survey data analysts would ask the following question: What's the chance that X is 1,100 or more when p is 0.50 and n = 2,000 on a binomial distribution?

(continued)

(continued)

To answer this question, you take 1,100, subtract the mean of the binomial ($n * p$, where $n = 2{,}000$ and $p = 0.50$ — $n * p = 1{,}000$), and divide by the standard deviation of the binomial (the square root of $n * p * 1 - p$, where $n = 2{,}000$, $p = 0.50$, and $1 - p = 1 - 0.50 = 0.50$. You find the square root of 500, which is 22.36. Your answer is $Z = 1{,}100 - 1{,}000$ divided by 22.26, which is $Z = 4.47$. Because this value is off the chart of the Z table (see the Appendix), the probability of being less than $Z = 4.47$ is nearly 1.00, so the probability of being greater than $Z = 4.47$ (or that $X > 1{,}100$) is nearly zero. Therefore, your results are very unlikely to happen under the assumption of 50 percent cell-phone ownership. The results could mean one of two things. One, you got a very weird sample just by chance, which is highly unlikely. Two, the assumption that 50 percent of people own a cell phone is wrong. The latter is much more likely. (See Chapter 12 for more information on making decisions with probability.)

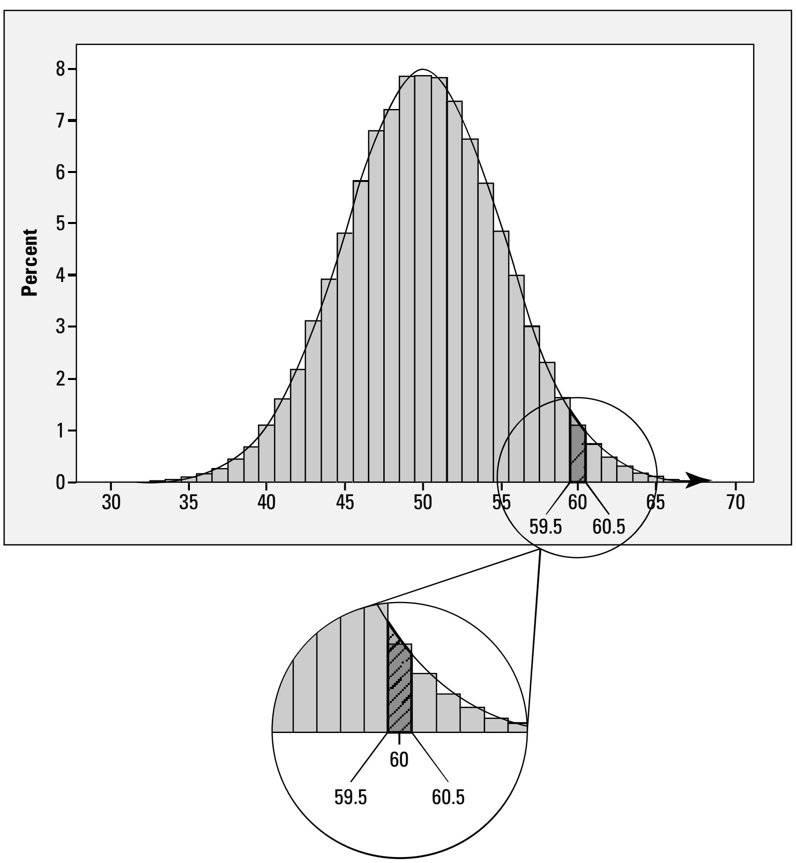


Figure 10-7: The probability of (X = 60) on 100 flips of a fair coin.

Chapter 11

Sampling Distributions and the Central Limit Theorem

In This Chapter

- Understanding what sampling distributions are and how they work
- Simplifying your calculations with the Central Limit Theorem
- Summing outcomes with the sample total
- Shooting for averages with the sample mean
- Finding probabilities for proportions

Probability is the chance that a certain outcome will occur in an individual situation (the chance that you'll win something on your next pull of the slot machine, for example). But you can also view probability as the proportion of time that you expect a certain outcome to occur over the long term (the proportion of time you'll win something if you play slots for a very long time, for example). (Notice that the two slot probabilities I mention are exactly the same in terms of their numerical values, but their interpretation is very different.)

In this chapter, I focus on figuring probabilities over the long term (or for a large number of individuals) in three situations. First, you figure probabilities for the overall total of your outcomes over many trials of an experiment (the probability that the total amount won by a player on a particular slot machine exceeds $100 in 100 tries, for example). Second, you calculate probabilities for the overall average of your outcomes over the long term or for a large group (the chance that a class of 50 students averages more than a score of 70 on a national exam, for example). Finally, you come up with probabilities for the overall proportion of successes over the long term (like the second slot machine example I mention in the first paragraph). The Central Limit Theorem gives you some very nice results that make those probability calculations for averages and proportions go much easier.

As you make your way through this chapter, you always have to keep one truth in mind: *Sample results vary.* And the amount of variability you can expect is important to know when you make decisions and predictions based on probability models for totals, averages, or proportions. To get a handle on measuring this sample variability, you use sampling distributions for the total, the average, and the proportion. Sampling distributions play an important role in determining how much variability you should expect when calculating and predicting probabilities, and they give you the background you need to get ready for the Central Limit Theorem.

Surveying a Sampling Distribution

If you want to find the probability that an actual total, average, or proportion of outcomes is going to have a certain value, you need to remember that you can't pinpoint the exact answer no matter what you do, because your prediction will likely be based on sample results, and sample results will vary. To get a handle on how much variability to expect, you need to know more about what the different sample results would be like if you repeated your sampling over and over again.

Setting up your sample statistic

A population parameter is a number that summarizes a population. In this chapter, the population values you want to pinpoint are the population total, designated T, the population average (or mean), designated by μ_x, and the population proportion, designated p. These figures represent the true values for the population and can't usually be found exactly. In order to find probabilities for the population total, average, or proportion, you take a sample and find the sample total (designated by t), the sample average (denoted $\overline{X}$), or the sample proportion (designated $\hat{p}$). The numbers that summarize samples are called *sample statistics,* and you use them to estimate or predict numbers from population parameters.

Lining up possibilities with the sampling distribution

Because sample results vary from sample to sample, sample statistics vary while the population parameters stay the same. When you make an estimate or predict the probability for the population parameter, you want to have some idea of how much the results are expected to vary. This means taking all possible samples (of the same size) from the population, finding the sample statistic for each sample, and making a probability distribution out of it — in other words, list out all the possible values of the sample statistic, along with their probabilities. This is called the *sampling distribution* of your statistic. The sample total, sample average, and sample proportion all have their own sampling distributions, with their own means and standard errors. (Standard error is the same as standard deviation [see Chapter 7], but the name changes when applied to a sample statistic.)

Going up and down with the sample total

Suppose, for example, a rollercoaster can hold a maximum of 30 passengers. In order to determine how much weight the coaster can carry, you need to estimate the total weight of 30 passengers at any given time. To estimate this total weight, you select a random sample of 30 people, measure their weights, sum up the weights, and get the sample total, t. You repeat the process with another random sample of 30 people, record your sample total, and then do it again. And again. And again. You repeat this process of sampling and recording until you have every possible sample of people included. (Sound crazy? See the section "Saved by the Central Limit Theorem" for more.) Then you put all the sample totals into a histogram (a graph that shows the values of t [grouped into intervals], along with the percentage of times each interval of values occurs). This histogram gives you the sampling distribution of the sample total, t, for the total weights of passengers on this rollercoaster over all possible samples.

Figure 11-1 shows you what the sampling distribution of the total weight for 30 passengers looks like, assuming the average weight of each passenger has a normal distribution (see Chapter 9) with mean 160 pounds and standard deviation 20 pounds (see Chapter 7 for info on mean and standard deviation of a probability distribution). Note that the average value of the total weight is 160 * 30 = 4,800.

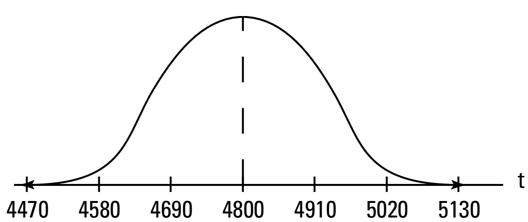


Figure 11-1: The total weight for 30 rollercoaster passengers for repeated samples (mean = 4,800 pounds, standard deviation 110 pounds).

Meeting in the middle with the sample average

If you want to find the average age of the passengers on the rollercoaster, your statistic would be the sample average (the sample mean, $\overline{X}$). To figure out how much variability you expect in your results, you go through the same process that you do for the sample total (see the previous section), except each time you add up the weights and then you divide by 30 — the number of passengers — to get your sample average. You repeat this process for a large number of samples (all of size 30.) Putting all the values of $\overline{X}$ into a histogram (a bar chart that shows the average weights on the horizontal axis and the probabilities on the vertical axis), you have the sampling distribution of the sample mean. Figure 11-2 shows the sampling distribution of the average age of n = 30 roller coaster passengers, assuming the population of all roller coaster riders has an average of 30 years with standard deviation 10 years. Note how little variability you see in Figure 11-2 (most of the average ages are between 25 and 36). You'll find out why the standard deviation of the sample mean is less than the standard deviation of the population in the "Sampling distribution of the sample mean, $\overline{X}$" section.

Rolling with the sample proportion

You can use a sampling distribution to find the proportion of passengers riding the rollercoaster who are women. In this case, you take random samples of 30 people, as you do with sample total and average, but instead of totaling or averaging their numerical information like weight or age, you note whether or not each person is a woman. Then you take the total number of women in the sample and divide by 30 (the total number of people in the sample). This is the sample proportion for that sample, and is denoted by $\hat{p}$. You repeat the process over and over a large number of times and put all the values of $\hat{p}$ together into a histogram to create the sampling distribution of the sample proportion, $\hat{p}$.

Saved by the Central Limit Theorem

After finding out how to create sampling distributions for totals, averages, and proportions (see the previous sections of this chapter), you may want to panic. "This is nuts. How can anyone sit there and take every possible random sample in the world, calculate the sample statistic, and put them all together on a histogram and get this crazy thing called a sampling distribution? I don't have time for that!" You're right. And lucky for you, some pretty deep thinking went on back in 1718 when Abraham de Moivre said to himself, "There's gotta be a better way!" In his book, *The Doctrine of Chances* (Chelsea Publishing Company), he gives us man's greatest gift to probability and statistics, the Central Limit Theorem. (By the way, his proof of the Central Limit Theorem took three days to cover in my probability theory class back in grad school, something you won't have to go through to be a happy user of the theorem. However, don't ask me to prove it again; I just can't bear reliving it!)

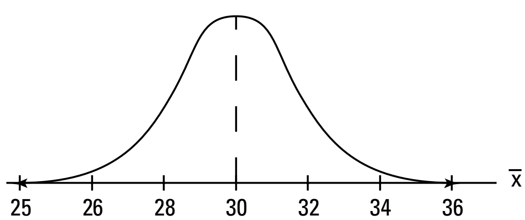


Figure 11-2: Sampling distribution for the average age of 30 passengers.

The Central Limit Theorem allows you to measure the variability in your sample results by taking only one sample, and it gives you a pretty nice way to calculate the probabilities for the total, the average, and the proportion based on your one sample of information, too. Notice that Figure 11-1 has a normal (bell-shaped) distribution (see Chapter 9). However, I assumed that the weights of passengers on the rollercoaster came from a normal distribution, so that's not such a shock. But here's the kicker: They could've had any shape, and Figure 11-1 would look the same way for large samples. So why, when you start with any distribution having any shape at all, do the totals suddenly have a bell-shape? It's not magic, but it's close. The following section describes the Central Limit Theorem and its major results for the sample total, sample average, and sample proportion. Using the results of the Central Limit Theorem allows you to find probabilities for each of these sample statistics without having to sample your life away.

Gaining Access to Your Statistics through the Central Limit Theorem (CLT)

The *Central Limit Theorem* (CLT) is a major probability theorem that tells you what the sampling distribution is for many different statistics, including the sample total, the sample average, and the sample proportion. It's a favorite of probability professors the world round, who love to talk about how amazing it is, and it truly is amazing. Of course, you don't have to love it the way they do, but you'll appreciate how it helps you solve probability problems for the total, average, and proportion.

The main result of the CLT

The main result of the Central Limit Theorem (CLT) is that the three major sample statistics (the sample total, average, and proportion) each have an approximate normal distribution as long as n, the size of your sample, is large enough. Let X be any random variable with mean μ_x and standard deviation σ_x (such as the weight, age, or gender of a rollercoaster passenger). Suppose you have n repeated trials of this random variable X, and you call these results $X_1, X_2, X_3, \ldots X_n$. (For example, these values could represent the ages of n passsengers.) The sampling distribution of the sample total, t, the sample average, $\overline{X}$, and the sample proportion, $\hat{p}$, are all approximately normal as long as n is large enough.

REMEMBER

Notice that X can have *any* distribution: X can represent the time to read a page in a book (which may have a uniform distribution; see Chapter 7); the number of times you have to flip a coin before you get the first head (which has a geometric distribution; see Chapter 14); or the proportion of free throws you expect to make when playing basketball (which is related to the binomial distribution; see Chapter 8).

The point is, if you repeat any of these experiments n times and total the outcomes, average the outcomes, or find the proportion of outcomes that fall into a certain category, the sampling distributions are all approximately normal. It doesn't matter what probability distribution you start out with; if you're repeating the experiment a large number of times, the sampling distribution of your statistic (the total, average, or proportion) will be approximately normal.

The result of the CLT makes it easier to find probabilities for the total, average, or proportion because you just use the steps for finding probabilities for a regular normal distribution (outlined in Chapter 9). All you need to know is the average and the standard error (a fancy way to say standard deviation of a sample statistic), and you're on your way.

TIP

Notice that the CLT says that the sample statistic (the sample total, average, or proportion) has an approximate normal distribution as long as n is "large enough." You discover the conditions needed for the sample total, sample average, and sample proportion in the sections that deal with each topic.

Why the CLT works

I don't use this section to prove why the CLT works (believe me, you don't want to go there), but I do show you some examples that give you a good idea of how the results occur. The big idea is that if you sum, average, or find proportions, your results always average out to the middle if you have enough data — that is, if n is large enough.

Conducting one trial

Suppose, for example, that you roll a single die a single time. The possible outcomes (1–6) have a discrete uniform distribution, meaning each of the finite possible outcomes is equally likely (each of the six possible outcomes of the roll has the same probability: ⅙; see Chapter 7). Figure 11-3 shows the graph of the probability distribution of X. The discrete uniform has an expected value E(X), which equals (1 + 6) ÷ 2 = 3.5; its variance is $\frac{(b-a+2)(b-a)}{12}$, which comes out to (6 – 1 + 2) * (6 – 1), divided by 12, which equals 2.92; and its standard deviation is the square root of the variance, which is 1.71. ***Note:*** There's nothing bell-shaped about the distribution of X, the outcome of the roll of a single die. In this case, you have a sample size of one, and if you put all the single observations together into a histogram, it will look just like the original probability distribution for X (not necessarily normal), because that's exactly what it is.

Conducting two trials

Suppose you roll a die twice, and you want the sampling distribution of the sample total (the sum of the numbers on the dice). In notation terms, you want the sampling distribution of $t = X_1 + X_2$, where X_1 is the outcome of the first roll and X_2 is the outcome of the second roll. How many possible values can you identify for the total? Before you figure that out, you need to list all the possible outcomes for the rolls. You know that the die produces 6 * 6 = 36 possible outcomes, by the multiplication rule (see Chapter 2), and all the outcomes are equally likely, each having probability 1/36. All 36 possible outcomes, along with their totals, are listed in Table 11-1.

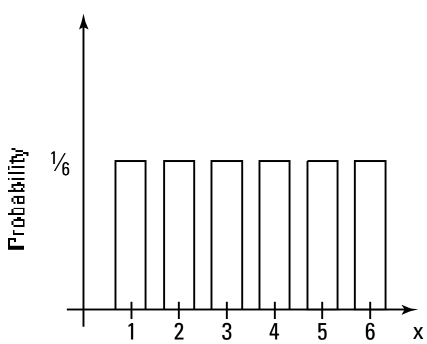


Figure 11-3: Probability distribution for one roll of a single die.

Table 11-1 All Possible Outcomes (and Their Totals) When You Roll a Die Twice

Outcome	Total	Outcome	Total	Outcome	Total	Outcome	Total	Outcome	Total	Outcome	Total
(1,1)	2	(2,1)	3	(3,1)	4	(4,1)	5	(5,1)	6	(6,1)	7
(1,2)	3	(2,2)	4	(3,2)	5	(4,2)	6	(5,2)	7	(6,2)	8
(1,3)	4	(2,3)	5	(3,3)	6	(4,3)	7	(5,3)	8	(6,3)	9
(1,4)	5	(2,4)	6	(3,4)	7	(4,4)	8	(5,4)	9	(6,4)	10
(1,5)	6	(2,5)	7	(3,5)	8	(4,5)	9	(5,5)	10	(6,5)	11
(1,6)	7	(2,6)	8	(3,6)	9	(4,6)	10	(5,6)	11	(6,6)	12

Notice that the sample total for this example can take on any integer between 2 and 12. This means the sample total, $t = X_1 + X_2$, takes on more possible values than X_1 or X_2 alone, so the sampling distribution is already different from when you conduct only one trial. What's more important, though, is that the probabilities for the values of t aren't all the same, as they are for X_1 and X_2 separately.

To find the probability for any value of the sample total, you check out how many times the sample total appears on Table 11-1 and add the probabilities together (remembering that each outcome has a probability of $1/36$). So, for example, the probability that the sample total is 2 is $1/36$, because the only time this happens is when the outcome is (1, 1). But the probability that the sample total is 3 is $2/36$, because two outcomes result in a total of 3: (2, 1) and (1, 2). The total that has the highest probability (no big surprise for craps players) is $t = 7$, which has 6 possible ways of occurring, resulting in a probability of $6/36 = 1/6$. I show the entire sampling distribution for the sample total in Table 11-2. Figure 11-4 shows the graph.

Table 11-2 Sampling Distribution for the Total of Two Die Rolls

Sample Total	*Probability*
2	$1/36$
3	$2/36$
4	$3/36$
5	$4/36$
6	$5/36$
7	$6/36$
8	$5/36$
9	$4/36$

Sample Total	Probability
10	3/36
11	2/36
12	1/36

REMEMBER

A sampling distribution is still a probability distribution, so it has the same properties of any probability distribution: Each probability has to be between zero and one, and the sum of the probabilities has to equal one.

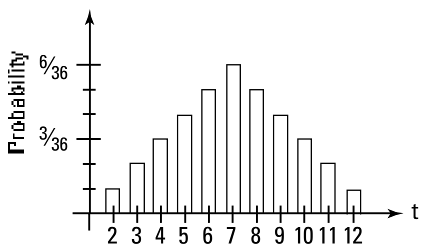


Figure 11-4: Sampling distribution for the total of two rolls of a single die.

Conducting many trials

According to the CLT, as the total sample size of a random experiment, n, increases, the total of the outcomes approaches a normal distribution. In other words, if you increase the number of times you roll a die, the sampling distribution of t, the total of all the outcomes of the n rolls, looks more and more like a normal distribution. It has more possible values, and it takes on a bell-shape. The amazing thing is, the shape of the original distribution doesn't matter. With the die-rolling example, the first roll has a flat discrete uniform distribution. When you start totaling the results for more and more rolls of the die, you start approaching the normal distribution. Figure 11-5 shows the sampling distribution of the total when you roll a die 30 times. Notice the average is 105, which is 30 times 3.50, the average value of one die.

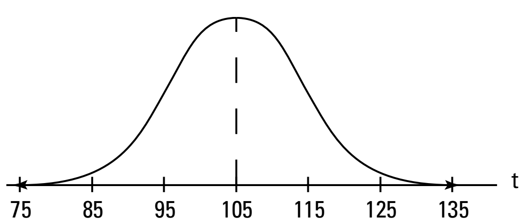


Figure 11-5: Sampling distribution for the total of 30 rolls of a single die.

A word about "large enough"

The required sample size needed in order for the CLT to work depends on the distribution of X (a single observation):

- If the original distribution of X starts off with a normal distribution, the sample total has an exact normal distribution no matter what the sample size is, and you don't even need the CLT.
- If the original distribution of X is mound-shaped (has a mound in the middle and slopes downward on each side as you move away from the middle) and symmetric (looks the same on each side if you cut it down the middle), the sample size doesn't have to be very large for the normal approximation to work well.
- If the original distribution of X is skewed (meaning many of its values are off to the right or off to the left), you need to have more trials before the values of the sample total start averaging toward the middle. That's because the values on the other side of the skewed distribution don't come up very often, and it takes more trials to include them in the sample and bring the average back to normal.

Other situations, however, don't call for the sample size to be near 30 in order for the normal to take effect. For example, with the die-rolling experiment, you need n to be only two before you see the probabilities for t piling up toward the middle and thinning out on the ends. So, the n = 30 approximation is just a conservative value meant to cover all cases; individual results vary depending on the shape of the original distribution. (See Chapter 7 for more information on shape.)

The Sampling Distribution of the Sample Total (t)

When your probability problem involves summing up results from a a large sample size, n, you can apply the Central Limit Theorem (CLT) to the sample total to find an answer. For example, if an elevator is made to hold 5,000 pounds, what's the chance that the total weight of its passengers will exceed that? (Hopefully that chance is a million to one or less!) In this section, you see the main results of the CLT applied to the sample total, t, and you go through the steps for finding probabilities for the sample total.

The CLT applied to the sample total

Regarding the sample total, the CLT says the following:

> Let X be any random variable with mean μ_x and standard deviation σ_x. Suppose you conduct n repeated independent trials of this random variable X (independent because the results don't affect each other) and find the total of all the outcomes. The sampling distribution of the sample total, t, represented by $t = X_1 + X_2 + X_3 + \ldots X_n = \sum_{i=1}^{n} X_i$, is approximately normal if n is sufficiently large (at least 30).

REMEMBER

If the original random variable X has a normal distribution with mean μ_x and standard deviation σ_x, the sampling distribution of the sample total, t, is exactly normal for any n, and the CLT isn't needed. In other words, you don't have to repeat the trials numerous times to have the results look normal, because they already do!

Here are some additional facts about the sampling distribution of t:

- The expected value of t (the mean or average of t) is equal to n times the expected value of X.

  In other words, $E(t) = \mu_t = n\mu_x$. (For more on expected value, see Chapter 7.)
- The variance of t is equal to n times the variance of X.

  In other words, $V(t) = \sigma_t^2 = n\sigma_x^2$. (For more on variance, see Chapter 7.)
- The standard deviation of t is equal to the square root of the variance of t.

  This is also known as the *standard error* of the total.

Finding probabilities for t with the CLT

You can use the CLT to find probabilities for the sample distribution of the sample total. The CLT also helps you find out what you should expect the total to be when you make certain assumptions about the individual variables and helps you assess the likelihood of your results.

Suppose, for example, that you roll a die (that you assume to be fair) 30 times. Each roll of the die has an expected value (overall average value) of 3.5 (by taking each outcome, 1 through 6, multiplying by its probability, $\frac{1}{6}$, and summing; see Chapter 7 for more on expected values). If you roll the die 30 times, you should expect your total to be 30 * 3.5, which is 105. So, how likely is it that you'll get a sum less than 120? How about greater than 120? How about between 120 and 130? Figure 11-6 shows the sampling distribution of t for the die-rolling example and the probabilities you're looking for to answer these three questions.

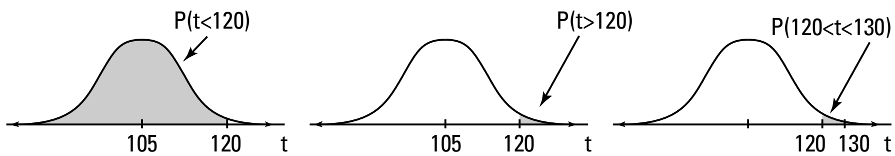


Figure 11-6: Finding probabilities for the total of 30 rolls of a fair die: $P(t < 120)$; $P(t > 120)$; and $P(120 < t < 130)$.

Following are the steps for finding these probabilities, and for any probability for t, the total of n random variables X, where X has any distribution with mean μ_x and standard deviation σ_x:

1. **Check the condition of the CLT for the sample total: n is at least 30.**

   When you roll a die $n = 30$ times, you meet the conditions of the CLT. So, when you roll the die 30 times repeatedly and sum the outcomes, the total of your results will have an approximate normal distribution (see Chapter 9) with a mean, $n\mu_x$, of $30 * 3.5$ (the mean roll of one die) = 105 and a standard error of $\sqrt{n\sigma_x^2} = \sqrt{30 * 2.92} = \sqrt{87.6}$, which equals 9.36. Here, 2.92 is the variance of the roll of one die. See "Why the CLT works" earlier in this chapter.

2. **Convert the sample total, t, to a standard score, Z, by using the formula $Z = \dfrac{t - n\mu_x}{\sqrt{n\sigma_x^2}}$.**

   To do this conversion, you take the value of t, subtract its mean, and divide by its standard error (see Chapter 9 for more information on z scores and how to get them). In Step 1, you determine that 105 is the value for the mean under the model and that the standard deviation is 9.36. Therefore, $t = 120$ converts to the standard (z) score of $\dfrac{t - n\mu_x}{\sqrt{n\sigma_x^2}} = \dfrac{120 - 105}{9.36} = 1.60$.

3. **Look up the standard (z) score on the Z table (Table A-2 in the Appendix) to find its corresponding probability.**

   The corresponding probability for 1.60 is 0.9452.

4. **If you want the probability of being less than Z, your job is done. If you want the probability of being greater than Z, you take one minus the probability you find in Step 3. If you want the probability that Z is between two values a and b, you convert a and b to standard scores**

and find the probability that Z is less than b (that is, all the probability up to point b) and subtract off the probability that Z is less than a (that is, the probability up to point a). This leaves you with the probability that Z is between a and b. (See Chapter 9 for info on finding greater-than and between probabilities using less-than probabilities.)

If you want the probability that the total on the 30 rolls is less than 120, that's equal to the probability that Z is less than 1.60 (see Figure 11-7). Looking up the value 1.60 on the Z table (Table A-2 in the Appendix), you see the probability that Z is less than 1.60 is 0.9452. The chance of rolling a fair die 30 times and getting a total of less than 120 is 94.52 percent.

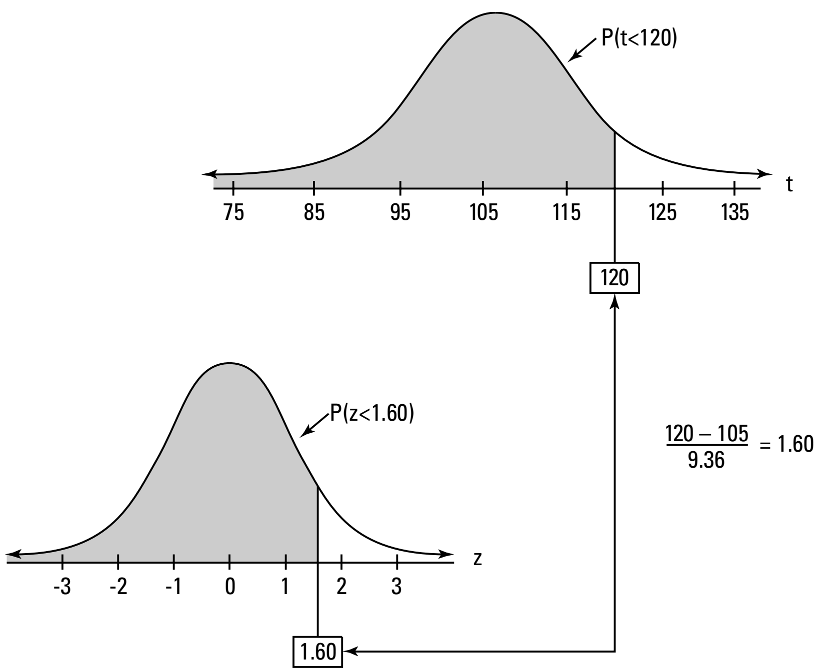


Figure 11-7: $P(t < 120) = P(Z < 1.60)$ for 30 rolls of a fair die.

If you want the probability that the total is more than 120, you want the probability that Z is greater than 1.60. Using the complement rule, you know that $P(Z > 1.60) = 1 - P(Z < 1.60) = 1 - 0.9452 = 0.0548$. The chance of getting a total of more than 120 when you roll a fair die 30 times is 5.48 percent (see Figure 11-8).

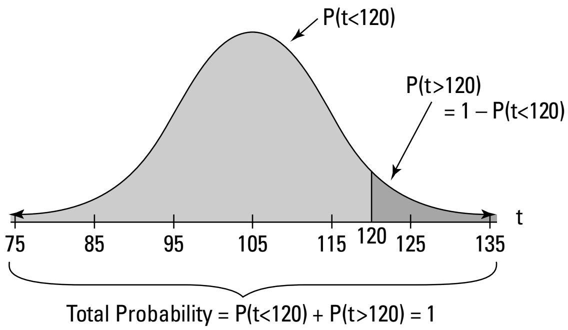


Figure 11-8: P(t > 120) = 1 – P(t < 120) for 30 rolls of a fair die.

Suppose you want the probability that the total of 30 die rolls is between 120 and 130. The value 120 converts to Z = 1.60, and the probability that Z is less than 1.60 is 0.9452. Using the Z-formula, the value of 130 converts to Z = (130 – 105) ÷ 9.36, which equals 2.67. The probability that Z is less than 2.67 is 0.9962. So, the probability that the sample total of 30 die rolls is between 120 and 130 is 0.9962 – 0.9452 = 0.051, or 5.1 percent (see Figure 11-9).

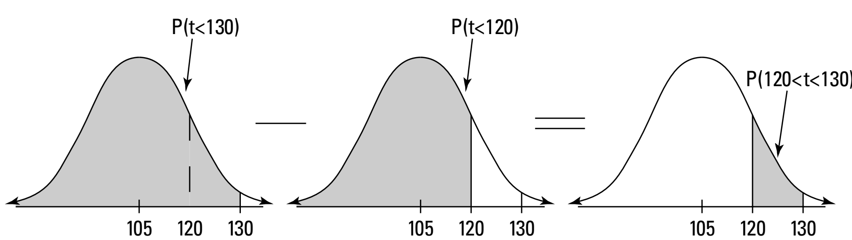


Figure 11-9: P(120 < t < 130) = P(t < 130) – P(t < 120) for 30 rolls of a fair die.

The Sampling Distribution of the Sample Mean, $\overline{X}$

Some probability problems involve finding the average of the results over repeated trials — for example, what's the chance that the average age of passengers on a rollercoaster is under 15? As long as you have enough passengers in your sample (in other words, your sample size, n, is large enough), you can apply the Central Limit Theorem (CLT) to answer the question. In this section, you see the main result of the CLT applied to the sample average, $\overline{X}$, and you go through the steps for finding probabilities for the sample average.

The CLT applied to the sample mean

Here's the scoop on the sampling distribution of the sample mean, $\overline{X}$, according to the Central Limit Theorem (CLT):

> Let X be any random variable with mean μ_x and standard deviation σ_x. Suppose you conduct n independent repeated trials of this random variable X (independent because the results aren't related to each other) and average the outcomes. The sampling distribution of the sample mean, $\overline{X}$, is approximately normal if n is sufficiently large (at least 30).

If the original random variable X has a normal distribution with mean μ_x and standard deviation σ_x, the sampling distribution of the sample mean, $\overline{X}$, is exactly normal for any n, and the CLT isn't needed.

Here are some additional facts about the sampling distribution of the sample mean, $\overline{X}$:

- The expected value of $\overline{X}$ (the mean, or average) is equal to the expected value of X.

  In other words, $\mu_{\overline{x}} = \mu_x$.
- The variance of $\overline{X}$ is equal to the variance of X divided by n, written as $\sigma^2_{\overline{x}} = \frac{\sigma^2_x}{n}$.
- The standard deviation of $\overline{X}$ (also known as the standard error) is equal to the standard deviation of X divided by the square root of n, written as $\sqrt{\sigma^2_{\overline{x}}} = \sqrt{\frac{\sigma^2_x}{n}} = \frac{\sigma_x}{\sqrt{n}}$.

The big idea here is that because n is the denominator of the standard error, the standard error of the sample mean decreases as n increases. That means as the sample size gets larger, the sample mean will have less variability from sample to sample. The sample total has the opposite effect; its standard error increases as the sample size increases. That's because error accumulates when you total numbers (because each number will vary). So, if you want to estimate the population mean, use the sample mean based on a large sample. Your results will be very precise; they won't be likely to change much with a new sample.

The required sample size needed in order for the CLT to work depends on the distribution of X (a single observation). In general, as long as n is at least 30, the CLT works well in most situations, and if you can use a larger sample size, the approximation works even better. But if the original distribution X starts off with a normal distribution, the sample mean has an exact normal distribution for any n, no matter what the sample size is, and you don't even need the CLT.

Finding probabilities for $\bar{X}$ with the CLT

You can use the Central Limit Theorem (CLT) to find probabilities for the sample mean. For example, say that you roll a die (that you assume is fair) 30 times. What's the chance that the average value you roll will be more than 4? The CLT also helps you find out what to expect under a certain model and assess the likelihood of your results.

Following are the steps for finding all probabilities for $\bar{X}$ under the model where n is assumed to be a certain value:

1. **Check the condition of the CLT: n is at least 30.**

   When you roll a die n = 30 times, you meet the conditions of the CLT because n is at least 30. When you roll the die 30 times repeatedly and average the results (the numbers that come up), those results will have an approximate normal distribution with a mean equal to 3.5 and a standard error equal to $\frac{1.71}{\sqrt{30}}$, which equals 0.31 (where 1.71 is the standard deviation of one die roll; see "Why the CLT works").

2. **Convert $\bar{X}$ to a standard (z) score by using the formula $Z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$.**

   To find the probability that $\bar{X}$ is greater than 4, you convert 4 to a standard (z) score by using the formula $Z = \frac{\bar{X} - \mu_x}{\frac{\sigma_x}{\sqrt{n}}}$ (see Chapter 9 for more on z scores). You subtract the mean and divide by the standard error. You see in Step 1 that 3.5 is the value for the mean under the model, and the standard error is 0.31. Therefore, $\bar{X}$ = 4 converts to the standard (z) score of $\frac{4 - 3.5}{0.31}$ = 1.61 (see Figure 11-10).

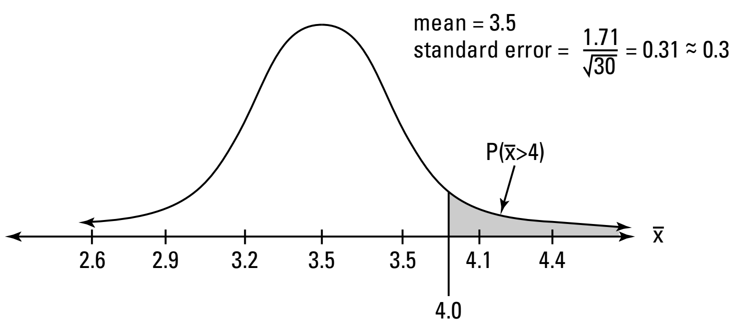


Figure 11-10: Finding the probability of getting an average of more than 4 when you roll a die 30 times.

3. **Look up the standard (z) score on the Z table (see Table A-2 in the Appendix) and find its corresponding probability.**

   The corresponding probability for 1.61 is 0.9463.

4. **If you want the probability of being less than Z, your job is done. If you want the probability of being greater than Z, you take one minus the probability you find in Step 3. If you want the probability that Z is between two values a and b, you convert a and b to standard scores and find the probability that Z is less than b (that is, all the probability up to point b) and subtract off the probability that Z is less than a (that is, the probability up to point a). This leaves you with the probability that Z is between a and b. (See Chapter 9 for info on finding greater-than and between probabilities using less-than probabilities.)**

   Because you want the probability of being greater than Z = 1.61, you take 1 – 0.9463 = 0.0537. The chance of getting an average of more than 4 when you roll a fair die 30 times is only 5.37 percent. (This proportion is what keeps casinos in business; people always believe they can "beat the odds." You run into this situation on the craps table, for example. For more on gaming probability, see Chapter 6.)

The Sampling Distribution of the Sample Proportion, $\hat{p}$

Many probability problems involve finding the proportion of the results that fall into a certain desired category over repeated trials. Being in the desired category is known as being a "success" in terms of the binomial distribution (see Chapter 8). For example, a problem could ask, "What's the chance that more than 60 percent of rollercoaster riders are women?" As long as you have enough passengers in your sample (in other words, the sample size, n, is large enough), you can apply the Central Limit Theorem (CLT) to answer the question. In this section, you see the main result of the CLT applied to the sample proportion, $\hat{p}$, and you work through the steps for finding probabilities for the sample proportion.

The CLT applied to the sample proportion

With sample proportions, the variable p represents the proportion of individuals in a population that have a certain desired characteristic, and $\hat{p}$ represents the proportion of individuals in a random sample that have that desired characteristic. The following results hold for $\hat{p}$ due to the Central Limit Theorem (CLT):

- The sampling distribution of the sample proportion has an approximate normal distribution when n is large enough (when $n * p$ is greater than or equal to 5, and $n * (1 - p)$ is greater than or equal to 5).
- The mean of sample proportion is p.
- The standard deviation of sample proportion, also called the standard error of the sample proportion, is $\sqrt{\frac{p(1-p)}{n}}$.

REMEMBER

For problems regarding the sample proportion, you never start out with a normal distribution or even a continuous distribution. Whenever you're looking at the proportion of individuals in a desired category, you're working with the binomial distribution (see Chapter 8). But, amazingly enough, if the sample size is large enough (see the first bullet in the previous list), you have an approximate normal.

Finding probabilities for $\hat{p}$ with the CLT

You can use the Central Limit Theorem (CLT) to find probabilities for your results from a sample proportion under a given value of p. For example, suppose you flip a fair coin 20 times; what's the probability that you'll flip more than 70 percent heads?

Following are the steps for finding probabilities for $\hat{p}$ under the model where you assume p to be a certain value:

1. **Check the conditions of the CLT: $n * p$ is greater than or equal to 5, and $n * (1 - p)$ is greater than or equal to 5.**

   When you flip a coin n = 20 times, you meet the conditions of the CLT because $n * p = 20 * 0.50 = 10$, and $n * (1 - p) = 20 * 0.50 = 10$. You know p is 0.50 because the probability of flipping a head is 0.50, just like the probability of flipping a tail.

2. **Convert $\hat{p}$ to a standard (z) score by using the formula $Z = \sqrt{\frac{\hat{p} - p}{\frac{p(1-p)}{n}}}$.**

   In this example, to find the probability that $\hat{p}$ is greater than 0.70, you convert 0.70 to a standard (z) score by using the formula $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$ (see Chapter 9 for more on z scores). Note that p is the value that you assume for the probability of heads under the model, which is 0.50. That means $\hat{p} = 0.70$ converts to the standard (z) score of $Z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.70 - 0.50}{\sqrt{\frac{0.50(1-0.50)}{20}}} = \frac{0.20}{0.118} = 1.79$.

3. **Look up the standard (z) score on the Z table (see Table A-2 in the Appendix) and find its corresponding probability.**

   The corresponding probability for 1.79 is 0.9633.

4. **If you want the probability of being less than Z, your job is done. If you want the probability of being greater than Z, you take one minus the probability you find in Step 3. If you want the probability that Z is between two values a and b, you convert a and b to standard scores and find the probability that Z is less than b (that is, all the probability up to point b) and subtract off the probability that Z is less than a (that is, the probability up to point a). This leaves you with the probability that Z is between a and b. (See Chapter 9 for info on finding greater-than and between probabilities using less-than probabilities.)**

   You want the probability of being greater than Z, or 1.79, so you take 1 – 0.9633 = 0.0367. The chance of getting 70 percent (or more) heads when you flip a fair coin 20 times is only 3.67 percent.

TIP

One question I often get from students is how do you know when a probability problem is a proportion problem and not a problem about the average (aka sample mean)? The key is to look very closely at the question. For example, if a question says "You roll a die 30 times; what's the probability that the average is more than 4?", that is a question about the sample mean. If the question says "You roll a die 30 times; what's the probability that there are more than 20 percent 6s?", that's a question about the sample proportion. The clues are in the question being asked.

Chapter 12

Investigating and Making Decisions with Probability

In This Chapter

- Estimating and testing probabilities with confidence intervals
- Using hypothesis testing to solve probability problems
- Solving problems with probability in the real world

Probability has many different everyday uses. It helps you become a better poker player; decide whether to play the lottery; determine whether you should have surgery or not; or decide whether to carry your umbrella to work or take the risk. People also use probability to make predictions about the future and to make big business decisions, such as if a particular drug will be effective, if the American people would support a particular female candidate for president, or if manufacturers are making goods according to specifications.

In this chapter, you see how probability is involved in the decision-making process in three major areas: confidence intervals, hypothesis tests, and quality control models. In addition, you see how people use probability to assess the chance of making the wrong decision and to determine the consequences.

TIP: This chapter is an overview of how people use probability in major areas of statistics. For a more in-depth look at how to set up and conduct confidence intervals, hypothesis tests, and quality control models, see *Statistics For Dummies* (also written by yours truly and published by Wiley) or your own statistics text.

Confidence Intervals and Probability

In cases where you have to guess what a probability is, a confidence interval is the method of choice. A *confidence interval* is a guesstimate at what a true probability should be. Suppose you live in the suburbs and are a victim of dog owners who don't scoop up after their dogs. You know this happens a large percentage of the time, but you want to be prepared with rock-solid evidence when you go to the next city-council meeting. So, you set up an investigation to figure out what percentage of the time dog owners clean up after their dogs in your neighborhood. (One lady in my neighborhood actually has a sign up reminding everyone of the $100 fine for failing to scoop. It's too hard to read, though, because of all the "debris" that has been deposited around it.) To do a good statistical job with this, you can't include only the percentage that you observe; you have to include a "plus or minus" that allows for the fact that sample results vary.

In this section, you discover how to use confidence intervals to estimate probability and how to use probability to design and assess any confidence interval.

Guesstimating a probability

Estimating a probability involves taking a good guess at what the true probability is when you have no way of knowing whether you're right or wrong. For example, if you want to estimate what percentage of people in the United States think the president is doing a good job, your answer is always based on some sort of sample, not on everyone in the population; you don't know if your answer is right on or not. You also know that the next time you take a sample of that size, you're likely to get a different answer, which adds to the problem. To account for the fact that you have "error" in your answer because you didn't sample everyone, and because sample results vary, you include a plus or minus amount with your best estimate. The plus-or-minus element is called the *margin of error.* In the following pages, you discover how to put your estimate and a margin of error together to create a confidence interval — a guesstimate at what the true probability should be.

Creating a confidence interval for p

Suppose, for example, that you have a loaded die (one that isn't fair). You want to find the probability that you'll roll a 6. In probability terms, if you let p be the overall percentage of time that the die would land on 6 if you rolled it forever, you want to estimate p, the proportion of 6s in the "population" of all possible rolls. You roll the die 100 times, and you find that 30 percent of the time (0.30), it lands on 6. The value you find, 0.30, is called a *statistic* because it comes from a sample. You use the statistic to estimate a population parameter based on sample data. In this case, 0.30 is known as the sample proportion, $\hat{p}$, and you use it to estimate p.

But 0.30 as the probability of rolling a 6 can't be your "final answer," as they say on television game shows. The next time you roll the die 100 times, you probably won't get 6s exactly 30 percent of the time. To handle this dilemma, you add and subtract some margin of error from your sample result to measure how much you expect the results to vary from experiment to experiment. The formula for margin of error when you're using a sample proportion to estimate a population proportion with 95 percent confidence is $1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Because you roll the die 100 times and get 6s 30 percent of the time, $\hat{p}$ is 0.30 and n = 100, so the margin of error is $1.96\sqrt{\frac{0.30(1-0.30)}{100}} = 1.96 * \sqrt{0.0021} =$ 0.0898, or about 9 percent. The probability that the die will land on a 6 should be between 21 and 39 percent, which is your confidence interval.

Figuring your confidence level with the Z table

The 1.96 in front of the margin of error formula (see the previous section) reflects the fact that you want to be 95 percent confident in your answer. The number is a value from the *Z table* (see the Appendix) — a table that represents the number of standard deviations you need to go out on either side of your estimate in order to cover 95 percent of the possibilities (the 95 percent is known as the *confidence level*). The Empirical Rule says that if you go out about 2 standard deviations on either side of your estimate, you capture about 95 percent of the values, and 1.96 is the precise version of 2, based on the Z table.

Probability offers many different confidence levels; a 95-percent confidence level is the most commonly used. Table 12-1 shows some of the most common confidence levels. For more information on this topic, see *Statistics For Dummies* (Wiley) or your own statistics text.

Table 12-1 Various Confidence Levels and Their Z Values

Confidence Level	*Z Value*
80%	1.28
85%	1.44
90%	1.65
95%	1.96
98%	2.33
99%	2.58

Most confidence intervals work the same way whether you're estimating a population probability, a population average, or the difference between two population means or averages. You have a margin of error that contains a z value (1.96 if your confidence level is 95 percent) and an expression that involves n (the sample size) in the denominator.

Assessing the cost of probably (hopefully?) being right

You need to make a decision when you design your confidence interval. How much error are you willing to live with, and how much confidence do you want? People usually begin by selecting the confidence level they want and then determine how much error they can live with. From there, they can calculate the sample size they need to keep the margin of error capped at that level.

The ideal situation when you're making a confidence interval (of any kind) is to have a high confidence level (typically at least 95 percent) and a small margin of error at the same time. If you want more confidence, you need a larger value of Z that increases the margin of error. If you want less confidence, you need a smaller value of Z that decreases the margin of error but leaves you with lower confidence in your answer. You can solve these problems by altering the sample size, n.

If you increase n, you make the margin of error smaller because n is in the denominator of the margin of error formula. So, to be confident in your results but also have a small margin of error (the most desired situation), you should use a large sample size and a high level of confidence. Look at the example I present in the section "Creating a confidence interval for p," where the sample proportion of heads is $\hat{p}$ equals 0.30. If you roll the die 500 times instead of rolling it 100 times, your margin of error goes from 0.0898 to 0.040, or about 4 percent. This is true because you have the following:

$$1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 1.96\sqrt{\frac{0.30(1-0.30)}{500}} = 1.96 * \sqrt{0.00042} = 0.040.$$

When you want to estimate a probability, a "quick and dirty" formula for the sample size you need to get within the margin of error you want is $n = \frac{1}{(\text{MOE})^2}$. In this formula, MOE represents the desired margin of error. So, sticking with the die-rolling example, if 4 percent is still too much error for you when attempting to find the probability of rolling a 6 on a loaded die, you can come at the problem another way. What margin of error are you willing to live with? Say you want to be within 1 percent. Your sample size has to be at least $n = \frac{1}{(0.01)^2}$, which is 10,000. Better start rolling now! This formula provides a conservative estimate only. See *Statistics For Dummies* (Wiley) for more information on calculating sample size.

Interpreting a confidence interval with probability

A confidence level tells you how much confidence you have in the results of an experiment or survey. To be more precise, when you take a sample and use a sample statistic to estimate a population parameter, the confidence level tells you how often your interval will actually contain the true population parameter over many, many different samples — in other words, how often you can expect to be right.

Suppose, for example, that you roll a loaded die 10,000 times and want to estimate the percentage of 6s that appear, using a 95-percent confidence interval. Knowing you have a 95-percent confidence level means that 95 percent of the time when you roll the die 10,000 times, your interval will be right — that is, the true percentage of 6s for that die falls in your interval. Five percent of the time you'll be wrong just by chance. After all the work of rolling a die 10,000 times, you should sure as heck hope you get the right answer! To be even more sure of yourself, you may want to increase your confidence level from 95 percent to 99 percent so you have only a 1-percent chance of being wrong, just by chance.

REMEMBER

Increasing the sample size to 10,000 doesn't increase your confidence, it just gives you a smaller margin of error for that level of confidence. To increase your confidence, you have to increase the confidence level.

How can a confidence interval end up with the wrong answer just by chance? Because that's how the ball bounces, so to speak. Sometimes when you roll a die multiple times, you get many big numbers just by chance. Sometimes, you get all small numbers just by chance. The numbers don't always balance out; so, in a few cases, the answer will be wrong just by chance. In general, if you have a confidence level of $1 - \alpha$, the chance that your sample will create an interval that doesn't contain the true population is α. Another way you can look at it is to think of everyone you know conducting their own experiments by rolling a die 10,000 times, recording the percentage of times 6s come up, and creating 95 percent confidence intervals from the experiments. If you have 100 people do this, you would expect about 95 percent of those intervals to be right and about 5 percent to be wrong (just by chance).

Probability and Hypothesis Testing

Suppose you go to your next city council meeting and air your views about a dog-scooping issue that's tearing your neighborhood apart. You've found that a high percentage of dog walkers in the neighborhood don't scoop up after their dogs. A neighbor counters your point by saying the percentage of dog owners is too small to worry about; she claims the probability that someone in your neighborhood owns a dog is only 5 percent. You leave the meeting bound and determined to upend this claim, because you know the percentage of dog owners in your neighborhood is more than 5 percent.

In this section, you discover how to use hypothesis tests to put such claims about a percentage or probability to the test. You also find out how probability determines the chance of making a wrong decision and how you can use probability to fight that dreaded practice called data snooping (not to be confused with scooping!).

Testing a probability

The idea of testing a probability comes up in many scenarios. To maintain quality control, a manufacturer regularly tests samples of its parts so that the probability of defectives is no higher than promised. You can also use tests of probability to give evidence that students are cheating; it all boils down to whether all their wrong answers could've matched just by chance. (Let me give you a hint: The probability is very, very small!)

Suppose, for example, that you work at a casino where all the dice are supposed to be fair. However, on one particular night spent working at a craps table, you come to believe you have a loaded die (one that isn't fair). You know it should be landing on 6 about a sixth of the time, but as you continue to observe, you see it landing on 6 more often than it should. You want to test your assumption that the die is fair (versus the assumption that it isn't fair; specifically, that the probability of getting a 6 is more than $\frac{1}{6}$).

In probability terms, if you let p represent the overall percentage of times that the die would land on 6 if you rolled it forever, you want to test p, the proportion of 6s in the "population" of all possible rolls. And your *null hypothesis,* or original assumption, is that the die is fair — in other words, $p = \frac{1}{6}$. Your alternative hypothesis, however, is that the die isn't fair — in other words, $p > \frac{1}{6}$. Using hypothesis testing notation, you have Ho (original assumption): $p = \frac{1}{6}$ versus Ha (alternative assumption): $p > \frac{1}{6}$ (the die lands on 6 more often than it should).

You roll the die 100 times and find that 30 percent of the time (0.30) it lands on 6. For this sample of data, you get a sample proportion of $\hat{p}$ equal to 0.30, compared to $1/6 = 0.167$, what the probability is supposed to be. So, it appears you have a problem here — the die isn't fair. But no isn't your final answer. You know that the next time you roll the die 100 times, you probably won't roll 6s exactly 30 percent of the time. Therefore, you standardize your statistic by subtracting what you expected the value to be (0.300 – 0.167) and dividing by the standard error (the part of the margin of error without the z value; see the section "Confidence Intervals and Probability" for more on z values and MOE). The result is called your *test statistic.* When you subtract the expected value and divide by the standard error, you convert the statistic to a z value. You can interpret this as the number of standard errors your data is away from what you expect.

The formula that you use to get your test statistic from the sample proportion is $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$. Here, p is the expected proportion ($1/6$), $\hat{p}$ is the sample proportion (0.30), and n = 100 rolls. The numbers give you a test statistic of $\frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.30-0.167}{\sqrt{\frac{0.167(1-0.167)}{100}}} = \frac{0.133}{\sqrt{0.0014}} = \frac{0.1330}{0.0373}$, which equals a z value of 3.57. Checking the Z table in the Appendix, you find that you don't see a z value this high occur very often; in fact, observing a z value of 3.57 or greater occurs less than 0.03 percent of the time.

This is where probability plays a huge role. You now have to make a decision. Your data shows a 6 showing up 30 percent of the time in 100 rolls. If the die is fair, you know that this result can happen only 0.03 percent of the time. So, the die could still be fair, and the null hypothesis true. But, odds are that isn't the case. Because the evidence is overwhelming, you conclude that the die isn't fair. In hypothesis test terms, you decide to reject Ho: $p = 1/6$ in order to accept Ha: $p > 1/6$.

For evidence to be overwhelming, you need to show that the probability of getting your sample results is too small for you to continue to stay with the null hypothesis. What probability do you use as the cutoff value? The value varies from researcher to researcher, but most people are happy with a cutoff value of 5 percent, or 0.05. This cutoff probability is also known as the *significance level* (or α level) of the test.

Putting the p in probability with p-values

You measure the level to which your evidence is overwhelming (or underwhelming, as the case may be) with a special probability called the *p-value.* The p-value is the probability that your results could happen just by chance while the null hypothesis, Ho (the original assumption you make), is still true. If you determine that Ho is false, you conclude that the alternative hypothesis, Ha, is true.

To make your decision about whether Ho is true given your data, you look at whether the p-value from your data is less than your predetermined significance level (the cutoff for the p-value where you determine your evidence against Ho is beyond a reasonable doubt; in this case, say, 0.05).

Say, for example, that you roll a die 100 times and get 30 6s. You calculate in the previous section that this could happen less than 0.03 percent of the time (which in decimal form is 0.0003). This is the p-value for your hypothesis test of Ho: $p = \frac{1}{6}$ versus Ha: $p > \frac{1}{6}$. Because the p-value is very small, you have very little evidence supporting the null hypothesis and plenty of evidence against it.

Don't confuse the p-value for a hypothesis test with the p that designates the probability of success in the binomial distribution (see Chapter 8.) It's important to write the whole phrase "p-value," not just p, when denoting the p-value.

Here's a checklist to determine if Ho is true, given your data:

- If the p-value is less than the significance level, you reject the null hypothesis, Ho, and conclude that your alternative hypothesis, Ha, is true.
- If the p-value is greater than the significance level, you fail to reject Ho because you don't have enough evidence against it.
- If the p-value is right on the line, you say your result is marginal, meaning that it could go either way.

You use p-values in most any kind of research where you collect data as evidence. The values provide an easy way to interpret the strength of the evidence against the null hypothesis; however, they don't tell the whole story. In the die-rolling example, you decide that the die is loaded (unfair), but you may want to know, or report, "how loaded" it is. In other words, if you determine that the probability of rolling a 6 isn't greater than $\frac{1}{6}$, what is it equal to?

To estimate the true probability of rolling a 6 for the loaded die, you can create a confidence interval. In the first section of this chapter, you do just that! You come up with a confidence interval of 0.30 plus or minus 0.09; in other words, you believe with 95 percent confidence that the true probability of rolling a 6 on this loaded die, based on your sample of 100 rolls, is between $0.30 - 0.09 = 0.21$ and $0.30 + 0.09 = 0.39$ — far away from 16.7 percent, as it should have been.

Accepting the probability of making the wrong decision

When you perform any hypothesis test, you take a chance that you'll make the wrong decision, because probability is involved in making the decision (and it has to be). In the suspicious die-rolling example from the previous section, you decide that the die is loaded — because you calculate less than a 0.03 percent chance of getting the results you got, you know something is fishy. But you don't know for sure, because you didn't roll the die forever (how could you?). It's possible that the die showed you strange results this time. Isn't that chance what brings most gamblers into casinos in the first place — the one-in-a-million shot of beating the odds? After all, just because someone wins a big pot on the roulette wheel, "lets it ride" (keeps all the money on the same number for another spin), and wins again doesn't mean that cheating is involved. Someone has to beat the odds, right?

However, you saw a problem, you collected data, and you made your decision. Now, what's the chance you made the wrong decision?

You can make two types of errors in the hypothesis-testing process:

- **Type I errors:** Type I errors can happen only if you reject the null hypothesis, Ho (which you do in the die-rolling example). Type I errors occur when the null hypothesis is true, but you throw it out the window. In the case of the die, you make this error if the die really is fair and you decide that it isn't, based on your data. What's the chance of this happening before the fact? Because your significance level is 0.05, or 5 percent, that represents your chance of a Type I error. The only way to reduce this chance is to reduce your significance level.
- **Type II errors:** Type II errors can happen only if you don't reject the null hypothesis when you should. I call this a missed opportunity to blow the whistle. If the die is loaded, but you can't make a conclusion because your evidence isn't strong enough, you make a Type II error. Sometimes a Type II error happens because your sample size is too small to find anything. In this case, you can increase your sample size to decrease the chance of a Type II error.

The consequences of Type I and Type II errors are important to consider. Suppose you conclude that the die is unfair, and you see only one player at the table, so you suspect him of planting a loaded die. He gets arrested and goes to jail, but you were wrong — the data turned out to be okay after further testing. This is the consequence of a Type I error — I call it a false alarm. And false alarms have a price, usually paid by the "accusee" rather than the "accuser." But, suppose the die really is loaded — the guy makes a pile of money and gets away with it. You miss an opportunity to blow the whistle, and the casino loses out. Eventually, you lose your job because you don't spot enough of these crimes. The "accuser" rather than the "accusee" usually pays the price of a Type II error.

Putting the lid on data snoopers

One of the big no-nos in scientific research is called "data fishing" or "data snooping." The culprits fish or snoop around for data until they find a result — any result, as long as it gets them on the news. Although you'd like to think it never happens, you find out that some people will do anything to get their 15 minutes of fame. Suppose, for example, that the latest medical sound bite says that if you have brown hair, you have a 70 percent chance of developing split ends compared to different hair colors. This may be true, but before you go screaming to your stylist, you need to look behind the scenes to find out how this probability came about.

One of the ways data snoopers "find" results is by conducting tons and tons of tests and picking out the only two items that turned out to be statistically significant (out of perhaps hundreds of results). The problem with this approach is that each test has a 5 percent chance of being wrong (assuming the significance level is 0.05). So, if you do 100 tests, you should expect about 5 percent of them to give the wrong results, just by chance. So, those two supposedly "significant" results are likely wrong results. A wrong result means that the test came out significant, but the truth is that nothing is going on at all. (In other words, a Type I error is committed; see the previous section.)

In the hair example, if you take every possible variable under the sun and measure it, along with whether or not someone has split ends, chances are you'll find some information that appears to be "significant." But that doesn't mean you have a truly significant result; it could be just a result of data snooping.

To control the "overall error rate" for more than a few hypothesis tests, the good guys take the original significance level (usually 5 percent; see the section "Testing a probability") and divide it by the number of tests conducted to get a new (much smaller) significance level. The researchers use this much smaller significance level for each test, which forces the p-value (see the section "Putting the p in probability with p-values") of each individual test to be much smaller in order to declare a significant result. For example, if you conduct ten tests, and your significance level is 0.05, you take $0.05 \div 10 = 0.005$ and use that as your new significance level for each test (which means your p-value has to be less than 0.005 [not 0.05] before you can claim significance). Only 5 out of 1,000 tests will come up wrong now rather than 5 out of 100. This approach is called the *Bonferroni approach* to controlling the overall error rate. Plain and simple, it puts the lid on data snooping.

Probability in Quality Control

In quality-control situations, a manufacturer sets up its process-making products so that workers make the products consistently and with a certain level of quality. In order to assess whether the company actually achieves this goal, quality-control managers periodically take samples and measure them and use that information to decide whether the process is going along smoothly or if the company has a problem and needs to stop the process.

Now, if a squirrel jumps into the chocolate vat, the company knows it has a problem; managers don't need a sample to figure that out. But say that you're making candy bars and over time, the machine starts pressing them out a little bigger than specified. Eventually, the wrappers don't wrap correctly, causing a lot of problems. Can you detect this problem early on? On the other side of the coin, say that you sample some candy bars over time and notice that the last four in a row have been a little smaller than usual. Does this mean something is wrong in the process, or is the error just chance variability?

Manufacturers use probability to make decisions about whether a process is "in control" — in other words, going according to plan with everything operating within specifications — or "out of control," where the company has a problem that needs some immediate attention. The choice is a very delicate balance, because if you're Grandma Josie making one pie per day, stopping your process because something is wrong with your measuring cup isn't a problem. However, if you're Grandma Josie Incorporated making 10,000 pies per day, stopping your process is expensive and time consuming (not to mention frustrating for all your customers waiting for their pies!). In order to stop a process, you have to have evidence "beyond a reasonable doubt" that you have a problem. That's where probability comes in.

TIP

One of the ways to determine if a process is in control is to set limits of three standard deviations on either side of the target value. For example, suppose a candy bar is supposed to weigh 6 ounces, and you allow a maximum standard deviation of 0.10 ounce. The limits of the process of making candy bars call for them to weigh 6 ounces plus or minus $3 * 0.10 = 0.30$ ounces, or within 5.7 and 6.3 ounces. This process is similar to creating a 99 percent confidence interval for the average weight of the candy bars, only it involves what you want to see, not what you actually have.

What happens if one candy bar falls outside the limits? Because the specifications should hold 99 percent of the time, that error should happen $1 - 0.99 = 0.01$, or 1 percent of the time. What if it happens twice in a row? The probability of this happening by chance when the process is in control, using the multiplication rule (see Chapter 2), is $0.01 * 0.01 = 0.0001$, or 1 out of 10,000 candy bars. Now remember, this company makes millions of candy bars per year, so an outcome that occurs 1 out of 10,000 times is something that's going to happen to the company just by chance. Better check the next candy bar . . . uh oh, it's also outside the limits. The probability of this happening is $0.01 * 0.01 * 0.01 = 0.000001$. Now you're starting to sweat a little, but you're still willing to consider a chance, over a long period of time, that nothing is really wrong — that this error is just a fluke. Go ahead and check one more candy bar . . . yep, also out of the limits. Now you've seen four candy bars in a row outside the limits. The probability of this happening by chance only is $0.01 * 0.01 * 0.01 * 0.01 = 0.00000001$. Okay, time to stop the process; odds are something is wrong.

TIP

Many decisions used in manufacturing processes involve probability; quality control is just one of the areas of application. I've just touched on the topic here. For more information on quality control in statistics, see *Statistics For Dummies,* by yours truly (Wiley).