

Part II

Probability Theory and Probability Distributions



Top 4 Advantages to Normal Distribution

Modeling the properties of asset returns requires the choice of an appropriate *probability distribution*. This distribution must have properties that match up with actual historical experience with asset returns. One of the most popular choices for this type of modeling is the *normal* distribution.

The normal distribution offers several advantages in this case:

- ✓ It's a *continuous* distribution, defined for an infinite number of values. This is important since the number of different returns that can occur is also infinite.
- ✓ It's *symmetrical* about the mean; there is a balance between the probability of returns that are below the mean and returns that are above the mean.
- ✓ The probability of extreme outcomes (outcomes well above or below the mean) is quite low; for financial returns, these occur quite infrequently.
- ✓ It's *additive*; the sum of normal random variables is also normal. This means that if the returns to a single asset are normal, the returns to a portfolio of assets are also normal.



Learn about normal distribution and more at www.dummies.com/extras/businessstatistics.

In this part...

- ✔ Review the foundations of probability theory; this is the foundation of all statistical analysis.
- ✔ Use random variables and probability distributions to determine if a random event will take place.
- ✔ Use binomial distribution to compute probabilities for processes where only one of two possible outcomes may occur. This could be something as simple as flipping a coin several times to see if the coin turns up heads or tails on each flip, or as complicated as a stock price increase.
- ✔ Describe the rates of return to financial assets, the distribution of corporate profits, and the prices of key commodities (such as oil) using normal distribution.
- ✔ Understand two key areas of statistics: sampling and sampling distributions. Most statistical analysis is based on samples randomly drawn from a population.

Chapter 6

Probability Theory: Measuring the Likelihood of Events

In This Chapter

- ▶ Understanding sets and how they're related to each other
 - ▶ Determining the possible outcomes
 - ▶ Applying types of probabilities
 - ▶ Using rules of probability
-

Probability theory is a branch of mathematics that focuses on the analysis of random events and is the foundation of all statistical analysis. You can use probability theory to model a large number of situations that arise in practice. For example, you can use probability theory to estimate how likely it is that a new product will succeed in the marketplace, identify appropriate prices for an insurance company to charge its customers, and more.

This chapter reviews the mathematical foundations of probability theory, such as sets and events, defines types of probabilities, and introduces the rules of probability.

Working with Sets

Probability theory is based on the notion of a *set* — a collection of objects, such as numbers, letters, colors, names, and so on, individually called *elements*. You use mathematical operations, such as membership, subset, union, intersection, and complement, used to create new sets from existing ones according to specific rules. For example, you use the operation union to combine two different sets into one new set that contains all the elements from both sets. I explore each of these operations in the following sections.

Membership

Membership indicates whether an element belongs to a set. For example, suppose that set A contains the elements 1 through 6 (the numbers on a die), which is shown mathematically as $A = \{1, 2, 3, 4, 5, 6\}$.

As you can see, the elements or *members* in a set are listed only once, are separated by commas, and are enclosed within *braces*: $\{ \}$.

In this example, the element 3 belongs to set A . To indicate that an element is part of a set, you use the symbol \in : $3 \in A$.

On the other hand, to indicate that an element is *not* part of a set, you use the symbol \notin . So in this case, the element 7 doesn't belong to set A , or $7 \notin A$, because it's not listed in the definition (between the braces) of set A .

Subset

A *subset* is a set that's completely contained within a larger set. For example, suppose that sets A and B are defined as follows:

$$A = \{1, 2, 3, 4, 5, 6\}$$

$$B = \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$$

Set A represents the numbers on a die; set B represents the numbers from 1 to 10. In this example, set A is a *subset* of set B because every element of set A is also an element of set B . The symbol \subset represents that one set is a subset of another, as in $A \subset B$.

A *Venn diagram* is used to illustrate the relationship between sets. Sets are represented as circles so that it's easy to see how they're related to each other. If sets overlap, the area common to both sets is shaded.

The Venn diagram in Figure 6-1 shows the relationship between sets A and B . The diagram shows that set A is completely contained within set B — that is, A is a subset of B . A is completely shaded because the area of overlap between A and B is A itself.

As another example, suppose that set C contains the elements 1, 2, 3, 4, 5, and 6 (the numbers on a die), whereas set D contains the elements 1, 2, 3 and 7:

$$D = \{1, 2, 3, 7\}$$

$$C = \{1, 2, 3, 4, 5, 6\}$$

Set D is *not* a subset of set C because the element 7 belongs to set D but *not* to set C ; in mathematical terms, $D \not\subseteq C$.

Figure 6-1:
Venn
diagram
showing
that set A is
a subset of
set B .

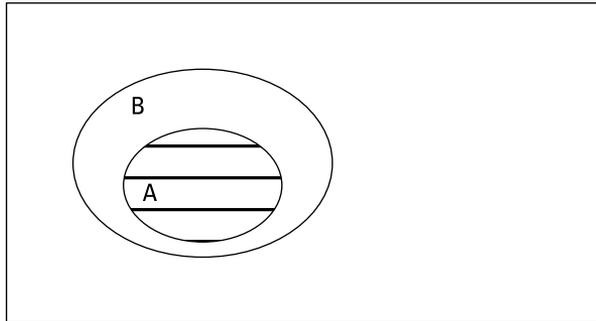


Illustration by Wiley, Composition Services Graphics

Union

Two sets can be combined with a mathematical operation known as *union*. The union of two sets A and B is a set that contains the following:

- ✓ All the elements in set A
- ✓ All the elements in set B

This definition also includes the elements that belong to *both* sets. As an example, suppose that set A contains all the students at a university who are majoring in mathematics; set B contains all the students who are majoring in finance. The union of sets A and B contains all students who are majoring in math *and* all students who are majoring in finance *and* all students who are majoring in *both* (for example, double majors).

As another example, suppose that sets A and B are defined as follows:

$$A = \{2, 4, 6\}$$

$$B = \{1, 2, 3, 4\}$$

The union of these sets is all the numbers on the face of a die except 5:

$$A \cup B = \{1, 2, 3, 4, 6\}$$

The symbol \cup represents union.

The union shows all elements that appear in set A , set B , or both. Note that even though elements 2 and 4 appear in both sets A and B , they're not listed twice in the union; a set contains only *unique* values.

The Venn diagram in Figure 6-2 shows the relationship between sets A and B . The shaded region in the diagram represents the union.

Figure 6-2:
Union of two
sets.

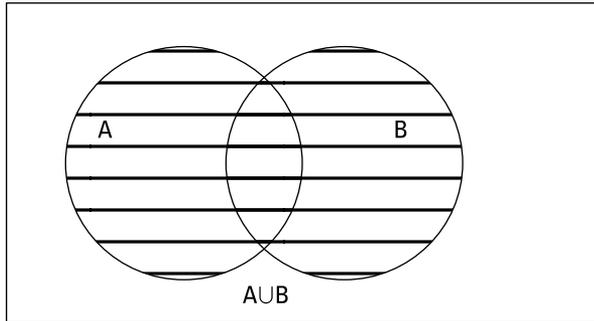


Illustration by Wiley, Composition Services Graphics

Note: The order in which you write the sets is irrelevant; for example, $B \cup A = A \cup B$.

Intersection

The *intersection* of two sets A and B is a set containing the elements that are in *both* sets. For example, suppose that sets A and B are defined as follows:

$$A = \{1, 3, 5, 7\}$$

$$B = \{3, 6, 7\}$$

The intersection of these sets is $A \cap B = \{3, 7\}$.

The intersection of A and B contains the elements 3 and 7 because these elements belong to *both* A and B . The symbol \cap represents intersection.

The Venn diagram in Figure 6-3 shows the relationship between A and B . The shaded region in the diagram represents the intersection of these sets.

Figure 6-3:
Intersection
of two sets.

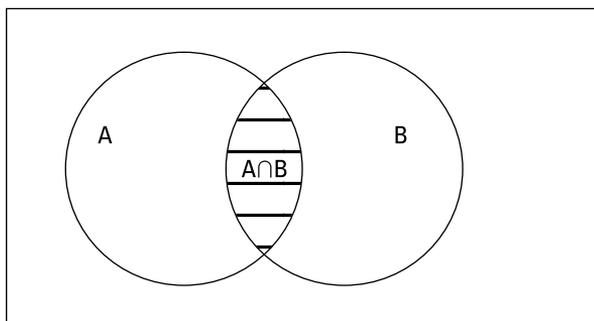


Illustration by Wiley, Composition Services Graphics

As another example, suppose that set C contains the elements 2, 4, 6:

$$A = \{1, 3, 5, 7\}$$

$$C = \{2, 4, 6\}$$

The intersection of these sets is $A \cap C = \{ \}$.

The intersection of sets A and C contains *no elements* because the sets don't have any of the same elements. The set containing no elements, or $\{ \}$, is known as an *empty set*. Two sets that have no elements in common are said to be *mutually exclusive*.

The Venn diagram in Figure 6-4 shows the relationship between sets A and C . This diagram has no shaded region because the intersection of sets A and C contains no elements.

Figure 6-4:
An intersection containing no elements between two sets.

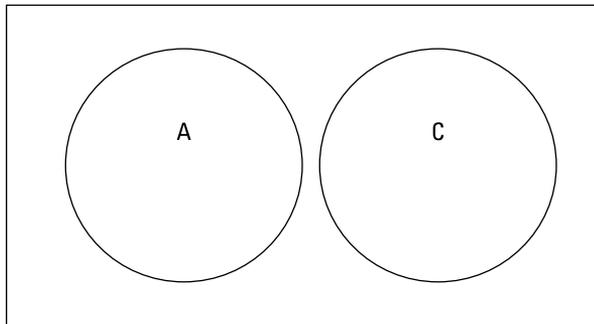


Illustration by Wiley, Composition Services Graphics

Complement

The mathematical operation *complement* is based on the notion of a *universal set* or *sample space* — all the elements a set may contain. For example, suppose that you roll a single die; the number that turns up may be any whole number between 1 and 6. Assume that set A contains the odd numbers that may turn up when you roll a die, and set B contains the even numbers:

$$A = \{1, 3, 5\}$$

$$B = \{2, 4, 6\}$$

In this case, the sample space contains all possible numbers that may turn up when you roll the die:

$$S = \{1, 2, 3, 4, 5, 6\}$$

The complement of set A is the set of all numbers that are elements of the sample space but *not* elements of A :

$$A^C = \{2, 4, 6\}$$

A^C is the set “ A complement.” It contains the elements 2, 4, and 6 because they *don’t* belong to set A , and they *do* belong to the sample space.

Note that elements such as 7, 8, 9, and so on aren’t elements of A^C because they’re not elements of set A , but they’re also not elements of the sample space.

The complement of A is shown in the Venn diagram in Figure 6-5. The shaded region shows all the elements in the sample space that don’t belong to set A .

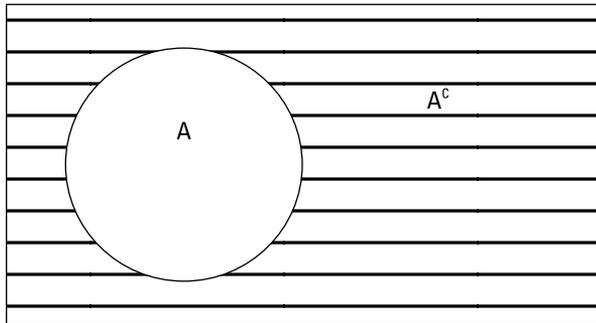


Figure 6-5:
Set A and
its comple-
ment A^C .

Illustration by Wiley, Composition Services Graphics

Similarly, the complement of B is $B^C = \{1, 3, 5\}$.

Betting on Uncertain Outcomes

Probability theory is based on the premise that a process generates uncertain (random) outcomes. This process is sometimes known as a *random experiment*, such as the following examples:

- ✓ A roulette wheel is spun. The outcome can be a 0, a 00 (“double zero”), or any number between 1 and 36.
- ✓ A lottery drawing results in a single winning number being chosen.
- ✓ A futures contract trades throughout the day, resulting in a settlement price at the close of trading.

In each case, the outcome isn't known in advance. Using probability, you can determine the likelihood of a specific outcome, such as the likelihood of getting an even number from a single spin of a roulette wheel.

In this section, I introduce several key terms, along with an introduction to computing probabilities.

The sample space: Everything that can happen

A *sample space* is another name for the universal set (described in the earlier section “Complement”); it contains all the outcomes that can result from a random experiment. For example, suppose you flip a coin two times. The possible outcomes of this random experiment are:

- ✓ Heads followed by heads (HH)
- ✓ Heads followed by tails (HT)
- ✓ Tails followed by heads (TH)
- ✓ Tails followed by tails (TT)

The sample space for this random experiment is $S = \{HH, HT, TH, TT\}$. It includes all the possible outcomes.

Event: One possible outcome

An *event* is one possible outcome of a random experiment. More formally, it is a *subset* of the sample space. For example, in the coin-flipping, the event $E =$ “2 tails turn up.” Event E is a set containing the element TT , or in mathematical terms, $E = \{TT\}$.

Event E is a subset of the sample space because it's completely contained within the sample space. As another example, the event $F =$ “at least 1 head turns up.” Event F is a set containing the elements HH, HT, TH , or $F = \{HH, HT, TH\}$.

In some cases, events may be related to each other. Two key ways in which events may be related to each other are known as mutually exclusive and independent.

These are described in the following section.

Mutually exclusive events

Two events are said to be *mutually exclusive* if they can't both happen at the same time. Here are two events that are mutually exclusive:

A = The roll of a die is odd.

B = The roll of a die is even.

Clearly, the roll of a die must result in a number that is either odd or even; it can't be both. Therefore, events A and B are mutually exclusive.

As another example, based on the coin-flipping experiment, suppose that two events are defined:

G = Two heads turn up.

H = Two tails turn up.

It's impossible for *both* two heads to turn up *and* two tails to turn up. This means that G and H are mutually exclusive. This result can be demonstrated using sets as follows:

$G = \{HH\}$ and $H = \{TT\}$. These events have no elements in common; their intersection is the *empty set* $G \cap H = \{ \}$.

The probability of the empty set is zero; therefore, the event that both G and H occur is *impossible*. This means that G and H are mutually exclusive.

Independent events

Two events A and B are said to be *independent* if the outcome of event A doesn't affect the outcome of event B and vice versa. For example, suppose that based on the coin-flipping experiment, event A is defined as the event that the first flip is a head, and event B is defined as the event that the second flip is a head. In other words:

$A = \{HH, HT\}$

$B = \{HH, TH\}$

Because the outcome of the first flip has no influence over the outcome of the second flip, events A and B are *independent* events. (See a more formal test of independence in the next section.)

Note that A and B are *not* mutually exclusive; both A and B can occur.

Computing probabilities of events

If a sample space contains elements that are all equally likely to occur, then computing the probabilities of events is straightforward. For example, for the coin-flipping experiment in the earlier sections “The sample space: Everything that can happen” and “Event: One possible outcome,” these probabilities exist:

$$\checkmark P(HH) = 0.25$$

$$\checkmark P(HT) = 0.25$$

$$\checkmark P(TH) = 0.25$$

$$\checkmark P(TT) = 0.25$$

For example, the probability of getting two consecutive heads is $\frac{1}{4}$ (which equals 0.25.) This is because HH is one of four possible outcomes when a coin is flipped twice. Furthermore, each outcome is equally likely to occur (because heads and tails are equally likely). Therefore, each outcome has a probability of $\frac{1}{4} = 0.25$.

$$\text{One possible outcome} \div 4 \text{ possibilities} = 0.25.$$

As an example, suppose that the event K is defined as “at least one tail turns up.” Then event K contains the elements HT, TH and TT, or $K = \{HT, TH, TT\}$.

You find the probability of event K with this formula:

$$P(K) = \frac{\text{elements in } K}{\text{elements in } S}$$

Because event K contains three elements and the total number of elements in the sample space is four, $P(K) = 3/4 = 0.75$.

Based on this formula, the probability of the empty set is 0, and the probability of the entire sample space is 1. For example, suppose that event A is an impossible event. It is represented by a set containing no elements (the empty set). The sample space contains the elements 1, 2, and 3. The probability of A is, therefore,

$$P(A) = \frac{\text{elements in } A}{\text{elements in } S} = \frac{0}{3} = 0$$

The probability of S is:

$$P(S) = \frac{\text{elements in } S}{\text{elements in } S} = \frac{3}{3} = 1$$

Looking at Types of Probabilities

The three basic types of probabilities are:

- ✓ **Unconditional (marginal) probabilities:** When events are independent
- ✓ **Joint probabilities:** When two things happen at once
- ✓ **Conditional probabilities:** When one event depends on another

In this section, you find out about each of these types of probabilities, and you also discover how you can use conditional probabilities to determine whether two events are independent.

Unconditional (marginal) probabilities: When events are independent

The *unconditional (marginal) probability* of an event is found as a row total or a column total in a joint probability table. As an example, Table 6-1 is a joint probability table, representing the distribution of students in a business school according to major and whether they're working on a bachelor's degree or a master's degree. In this section, I show you how to use data like this to find unconditional probabilities.

	Majoring in Finance	Majoring in Accounting	Majoring in Marketing	Total
Bachelor's degree	0.26	0.36	0.18	0.80
Master's degree	0.09	0.07	0.04	0.20
Total	0.35	0.43	0.22	1.00

Based on Table 6-1, the following events are defined:

- ✓ B = pursuing a bachelor's degree
- ✓ M = pursuing a master's degree
- ✓ F = majoring in finance

- ✓ A = majoring in accounting
- ✓ T = majoring in marketing

You can find the unconditional probabilities of the following events directly from Table 6-1:

- ✓ $P(B)$ = the probability of pursuing a bachelor's degree
- ✓ $P(M)$ = the probability of pursuing a master's degree
- ✓ $P(F)$ = the probability of majoring in finance
- ✓ $P(A)$ = the probability of majoring in accounting
- ✓ $P(T)$ = the probability of majoring in marketing

Say you want to find the probability that a randomly chosen business student is pursuing a bachelor's degree. In other words, you want to calculate $P(B)$.

Referring to Table 6-1, you look at the first row (which refers to students pursuing their bachelor's degrees). The row total is 0.80. This is the probability that a randomly chosen student is pursuing a bachelor's degree.

Suppose you want to know the probability that a randomly chosen student is majoring in finance. In other words, you want to calculate $P(F)$.

Referring to Table 6-1, you look at the first column (which refers to students majoring in finance). The column total is 0.35. This is the probability that a randomly chosen student is majoring in finance.

You can find the remaining unconditional probabilities in the same way. These are:

$$P(M) = 0.20$$

$$P(A) = 0.43$$

$$P(T) = 0.22$$

Joint probabilities: When two things happen at once

The probability that two different events occur at the same time is known as a *joint probability*. For example, the probability that a student is working on a bachelor's degree *and* is majoring in finance is a joint probability.

As you study Table 6-1, you can see that the intersection of two different events can determine joint probabilities. For example, to find the probability

that a randomly chosen business student is pursuing a bachelor's degree *and* is majoring in finance, take the intersection of events B and F . This equals $P(B \cap F) = 0.26$.

You find the remaining joint probabilities in the same way:

$$P(B \cap A) = 0.36$$

$$P(B \cap T) = 0.18$$

$$P(M \cap F) = 0.09$$

$$P(M \cap A) = 0.07$$

$$P(M \cap T) = 0.04$$

Conditional probabilities: When one event depends on another

The *conditional probability* of an event is defined as the probability of an event *given that* another event has occurred. For example, the probability that a student is working on a bachelor's degree *given that* he or she is majoring in accounting is a conditional probability. This is written as follows:

$$P(B | A)$$

The symbol “|” is used to indicate a conditional probability. (You pronounce this expression as “the probability of B *given* A .”)

To find the conditional probability of an event, you set up the ratio of a joint probability to an unconditional (marginal) probability (see previous sections on these types of probabilities). For example, say you want to find out what the probability is that a student who's known to be pursuing a bachelor's degree is majoring in marketing. Referring to Table 6-1, you first calculate the joint probability of pursuing a bachelor's degree and majoring in marketing, as follows:

$$P(B \cap T) = 0.18$$

Then you find that the unconditional probability of pursuing a bachelor's degree equals $P(B) = 0.80$. Therefore,

$$\frac{P(B \cap T)}{P(B)} = \frac{0.18}{0.80} = 0.225$$

As another example, to find the probability that an accounting major is pursuing a master's degree you take the joint probability of these two events:

$$P(M \cap A) = 0.07$$

The unconditional probability of majoring in accounting equals $P(A) = 0.43$. Therefore,

$$P(M|A) = \frac{P(M \cap A)}{P(A)} = \frac{0.07}{0.43} = 0.163$$

Determining independence of events

You can use conditional probabilities to determine whether two events are independent. Two events are independent if the probability of one event occurring doesn't influence the probability of the other occurring, and vice versa.

To prove independence, the following two conditions must be met:

$$P(A|B) = P(A)$$

$$P(B|A) = P(B)$$

Using the business students example from the earlier section “Joint probabilities: When two things happen at once” and referring to Table 6-1, you can determine whether the events “majoring in accounting” (A) and “pursuing a bachelor's degree” (B) are independent events.

The first step is to compute the conditional probabilities $P(A|B)$ and $P(B|A)$: The joint probability of events A and B is $P(A \cap B) = 0.36$.

The unconditional probabilities of events A and B are

$$P(A) = 0.43$$

$$P(B) = 0.80$$

Therefore,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{0.36}{0.80} = 0.45$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{0.36}{0.43} = 0.84$$

because $P(A|B)$ must equal $P(A)$ and $P(B|A)$ must equal $P(B)$ for the two events to be independent. The results show that $P(A|B) = 0.45$, $P(A) = 0.43$, $P(B|A) = 0.84$, and $P(B) = 0.80$, so *both* conditions fail. Events A and B are *not* independent of each other; in other words, they're *dependent* on each other. Therefore, the decision to pursue a bachelor's or a master's degree appears to influence the choice of major.

Following the Rules: Computing Probabilities

In addition to computing joint, conditional, and unconditional probabilities (discussed in the previous sections), the following three rules can help you determine other probabilities:

- ✓ The **addition rule** shows the probability of the union of two events.
- ✓ The **complement rule** determines the probability of the complement of an event.
- ✓ The **multiplication rule** identifies the probability of the intersection of events.

I discuss these three rules and how to use them in the following sections.

Addition rule

You use the addition rule to compute the probability of the union of two events. Mathematically speaking, for events A and B , the addition rule states that $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

This shows that the probability of the union of events A and B equals the sum of the probability of A and the probability of B , from which the probability of *both* events is subtracted. Subtracting the probability of both events is necessary to avoid to problem of *double-counting*. This is shown in the following example:

Suppose that event A contains the elements 1, 2, 3 and event B contains the elements 3, 4, 5. The sample space contains the elements 1, 2, 3, 4, 5.

$$A = \{1, 2, 3\}$$

$$B = \{3, 4, 5\}$$

$$S = \{1, 2, 3, 4, 5\}$$

The corresponding probabilities are:

$$P(A) = 3/5$$

$$P(B) = 3/5$$

$$P(S) = 5/5 = 1$$

The union of A and B contains all the elements in the sample space:

$$A \cup B = \{1,2,3\} \cup \{3,4,5\} = \{1,2,3,4,5\} = S$$

As a result, the probability of A union B equals 1. (Recall that the sample space always has a probability of 1.) If you simply combine the probabilities of A and B , though, you will get a surprising result; they sum to $6/5$, which is greater than one.

$$P(A) + P(B) = 3/5 + 3/5 = 6/5$$

This result occurs because the element 3 appears in both A and B :

$$A \cap B = \{3\}$$

The probability of 3 was counted *twice*, one in set A and once in set B , which accounts for the sum of the probabilities being greater than one. By subtracting the probability of the element 3, the correct probability of one is found.

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$P(A \cup B) = 3/5 + 3/5 - 1/5 = 5/5 = 1$$

Table 6-2 shows the distribution of coffees (measured in pounds) the Big Bean Corporation produces during a given day.

	Special Reserve Blend (S)	Kona Hawaii Blend (K)	Aromatic Blend (A)	Total
Decaffeinated (D)	0.12	0.80	0.22	0.42
Regular (R)	0.24	0.12	0.22	0.58
Total	0.36	0.20	0.44	1.00

If you choose a pound of coffee randomly from the daily output of the Big Bean Corporation, what's the probability that it's either the *Special Reserve Blend (S)* or the *Regular (R)* (or both)?

In this example, you use the addition rule because you're being asked to compute the probability of a union. You combine the probability of S with the probability of R , subtracting the intersection between them to avoid the problem of double-counting.

$$P(S \cup R) = P(S) + P(R) - P(S \cap R)$$

From Table 6-2, you can determine that $P(S) = 0.36$; that $P(R) = 0.58$; $P(S \cap R) = 0.24$. Therefore,

$$\begin{aligned} P(S \cup R) &= P(S) + P(R) - P(S \cap R) \\ &= 0.36 + 0.58 - 0.24 = 0.70 \end{aligned}$$

Seventy percent of the coffee produced by Big Bean is either the special reserve blend, regular, or both.

When two events A and B are *mutually exclusive* (that is, they can't both occur at the same time), the addition rule simplifies to $P(A \cup B) = P(A) + P(B)$ because $P(A \cap B) = 0$.

For example, if you choose a pound of coffee randomly from the daily output of the Big Bean Corporation, what's the probability that it's either the *Kona Hawaii Blend* (K) or the *Aromatic Blend* (A)?

Because a pound of coffee can't be both the Kona Hawaii Blend *and* the Aromatic Blend, events K and A are mutually exclusive. This means that you can use the simplified version of the addition rule:

$$\begin{aligned} P(K \cup A) &= P(K) + P(A) \\ P(K \cup A) &= 0.20 + 0.44 = 0.64 \end{aligned}$$

Complement rule

The complement rule is expressed as follows:

$$P(A^c) = 1 - P(A)$$

A^c is the complement of event A .

Two events are said to be complements if they are mutually exclusive *and* their union equals the entire sample space. Here's an example: Suppose that an experiment consists of choosing a single card from a standard deck. Event A = "the card is red." Event B = "the card is black." Events A and B are complements because A and B are mutually exclusive (no card can be *both* red and black). The union of A and B is the sample space (the entire deck, because all cards must be either red or black, so the union of A and B equals the entire sample space.)

In the Big Bean example from the previous section, the complement of event D (decaffeinated coffee) is event R (regular coffee) because all coffee must be

either decaffeinated or regular, and no coffee can be *both*. You can find the probability of the complement of D as follows:

$$P(D^c) = 1 - P(D)$$

Referring to Table 6-2, you can see that $P(D) = 0.42$. Therefore, $P(D^c) = 1 - P(D) = 1 - 0.42 = 0.58$, which is equal to $P(R)$.

Multiplication rule

To figure out the probability of the intersection of two events, you use the multiplication rule. This is used to determine the probability that two events are *both* true. For example, suppose an experiment consists of choosing a card from a standard deck. Event A = “the card is red.” Event B = “the card is a king.” The multiplication rule could be used to determine the probability that the card is *both* red and a king (for example, a red king.)

The multiplication rule can be written in two equivalent ways:

$$P(A \cap B) = P(A|B)P(B)$$

$$P(A \cap B) = P(B|A)P(A)$$

Note that these formulas are simply algebraic rearrangements of the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

Suppose the Omega Corporation has been the subject of takeover rumors for several months. The takeover is far more likely to occur if the economy rebounds next year. Omega’s chief economist estimates that the likelihood of strong growth next year is 5 percent, the likelihood of weak growth is 35 percent, and the likelihood of negative growth is 60 percent. The likelihood of a takeover during a period of strong growth is estimated to be 40 percent; during a period of weak growth, this falls to 20 percent; and during a period of negative growth, it’s assumed to be only 5 percent. What is the probability that there is strong growth next year *and* Omega is taken over?

The following events are defined:

- ✔ S = “strong growth”
- ✔ W = “weak growth”
- ✔ N = “negative growth”
- ✔ T = “Omega is taken over”

The probability of the events S and T can be determined as follows:

$$P(T \cap S) = P(T|S)P(S)$$

Because there’s a 5 percent chance of strong growth next year, $P(S) = 0.05$. The likelihood of a takeover during a period of strong growth is estimated to be 40 percent. Therefore, $P(T|S) = 0.40$. So the probability that there’s strong growth next year *and* that Omega is taken over is

$$P(T \cap S) = P(T|S)P(S) = (0.40)(0.05) = 0.02$$

When two events A and B are *independent*, the multiplication rule simplifies to

$$P(A \cap B) = P(A)P(B)$$

$$P(A \cap B) = P(B)P(A)$$

This is because $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Chapter 7

Probability Distributions and Random Variables

In This Chapter

- ▶ Understanding the concept of the random variable
 - ▶ Describing the behavior of a random variable with a probability distribution
 - ▶ Summarizing the properties of a random variable with moments
-

This chapter introduces two new concepts that are used to determine the probability that a random event takes place — random variables and probability distributions.

These concepts are closely related to the notion of the *random experiment* (defined in Chapter 6). A random experiment is a *process* in which events unfold in an unpredictable way. A random variable is used to assign numerical values to all the possible outcomes of a random experiment. A probability distribution assigns probabilities to these numerical values.

In this chapter, I also define summary measures of a probability distribution, known as *moments*, such as expected value and variance. Random variables and probability distributions are used by economists, financial analysts, researchers, and others to model the behavior of economic and financial variables, such as interest rates, inflation rates, corporate earnings, and so on.

Defining the Role of the Random Variable

A random variable is based on a *random experiment*, a process that generates outcomes that aren't known in advance (see Chapter 6). For example, suppose that a game of chance consists of spinning a wheel with four colors — red, blue, green, and yellow — each color results in a prize ranging from \$1.00 to \$10.00. A random variable may be used to assign a prize value to each color.

For example, you could define X to represent the prize that is received for each color, as follows:

red	$X = \$1$
blue	$X = \$2$
green	$X = \$5$
yellow	$X = \$10$

In this example, the random experiment consists of spinning the wheel. For each possible outcome (color), X assigns a numerical value that represents the prize received.

It may seem like a paradox, but a random variable is neither random nor a variable! In fact, a random variable is a *function*. It assigns a single numerical value to each outcome of a random experiment. Random variables may represent a large number of different financial and economic variables, including the following:

- ✓ A corporation's profits during the upcoming quarter
- ✓ The number of new customers resulting from a new advertising campaign
- ✓ The value of the Dow Jones Industrial Average at the end of next year

As another example, suppose you conduct a simple random experiment by flipping a coin three times. The set of all possible outcomes, known as the *sample space*, consists of the following elements. (H represents a head turning up on a single flip of the coin, and T represents a tail turning up.)

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

S represents the sample space. Each element in the sample space is a single sequence of three flips; for example, HTH refers to a head followed by a tail followed by another head.

Because a head and a tail are equally likely to occur on each flip, each outcome of this random experiment is also equally likely to occur. For example, HHT is just as likely to happen as THT . With eight equally likely outcomes, each has a probability of $1/8$ or 0.125 .

An *event* is one outcome or a combination of outcomes of a random experiment. For example, suppose that you want to calculate the probability of the event E , where two or more heads turn up. This outcome can occur in four ways:

- ✓ Three consecutive heads (HHH)
- ✓ Two heads followed by one tail (HHT)
- ✓ A head followed by a tail followed by another head (HTH)
- ✓ A tail followed by two heads (THH)

You can express these possible outcomes more compactly with set notation:

$$E = \{HHH, HHT, HTH, THH\}$$

To compute the probability of the event E , you count the number of elements that correspond to event E and divide by the number of elements in the entire sample space (S):

$$P(E) = \frac{\text{elements in } E}{\text{elements in } S} = \frac{4}{8} = \frac{1}{2}$$

$P(E)$ is the probability of event E .

This approach can be extremely cumbersome if the sample space contains a large number of elements. As an alternative, you can define a random variable to represent the number of heads that turn up during the random experiment. You can then determine the probability of event E from the probabilities of the different possible values of the random variable.

For example, let the random variable X equal the number of heads that turn up when a coin is flipped three times. X has a numerical value for each outcome of this experiment. Here are the outcomes of the experiment and the corresponding values of X .

Outcome	X
<i>HHH</i>	3
<i>HHT</i>	2
<i>HTH</i>	2
<i>THH</i>	2
<i>HTT</i>	1
<i>THT</i>	1
<i>TTH</i>	1
<i>TTT</i>	0

For example, *HHT* represents two heads followed by a tail; therefore, the value of X for *HHT* is 2. Similarly, for the outcome *TTH*, the value of X is 1.

Suppose that a marketing firm conducts a survey of customers to determine whether they're satisfied with the customer service received from the local cable company. Each customer answers yes or no. The survey yielded the following replies:

yes	no	no	yes
no	yes	yes	yes
yes	yes	yes	yes
yes	yes	no	yes
no	no	no	no

For the results, X is defined as follows:

$X = 0$: the customer reply is no

$X = 1$: the customer reply is yes

The results are shown in Table 7-1.

<i>Number of Responses</i>	<i>X (0 = no, 1 = yes)</i>
8	0
12	1

By organizing the results this way, you can easily see the proportion of the customers who are satisfied with their cable service.

Assigning Probabilities to a Random Variable

Although random variables may provide useful information, their greatest advantage is that they simplify the calculation of probabilities. For example, in the case of the coin-flipping experiment in the previous section, computing probabilities directly from the values of a random variable is simpler than counting up all the ways in which an event can occur.

You can assign probabilities to each possible value of a random variable by using a *probability distribution* — a table or formula that shows these probabilities. A probability distribution has two important properties:

- ✓ The probability of each value of a random variable is between 0 and 1.
- ✓ The sum of the probabilities equals 1.

In the following sections, I show you how to construct a probability distribution. I also show you how to illustrate the properties of a probability distribution with a special type of graph known as a *histogram*.

Calculating the probability distribution

Based on the coin flip example in the earlier section, “Defining the Role of the Random Variable,” the range of possible values for X (the number of heads

that turn up) is 0 to 3. Here is the number of ways in which each possible value of X may occur:

X	Outcomes
0	<i>TTT</i>
1	<i>HTT, THT, TTH</i>
2	<i>HHT, HTH, THH</i>
3	<i>HHH</i>

Because eight equally likely outcomes of this experiment can occur, the probability for each value of X equals the number of outcomes divided by the size of the sample space. Table 7-2 shows this probability distribution.

X	$P(X)$
0	$1/8 = 0.125$
1	$3/8 = 0.375$
2	$3/8 = 0.375$
3	$1/8 = 0.125$

The probability distribution in Table 7-2 shows that

- ✓ The probability of getting no heads ($X = 0$) is 0.125.
- ✓ The probability of getting one head ($X = 1$) is 0.375.
- ✓ The probability of getting two heads ($X = 2$) is 0.375.
- ✓ The probability of getting three heads ($X = 3$) is 0.125.

Now, suppose that you want to calculate the probability of the event F , where two or more tails turn up. This outcome can occur in four ways:

- ✓ Three consecutive tails (*TTT*)
- ✓ Two tails followed by a head (*TTH*)
- ✓ A tail followed by a head followed by another tail (*THT*)
- ✓ A head followed by two tails (*HTT*)

The event F corresponds to a set containing four elements:

$$F = \{TTT, TTH, THT, HTT\}$$

For two or more tails to turn up, the experiment must result in either zero heads or one head. Therefore, you can calculate the probability of F as follows:

$$P(F) = P(X = 0) + P(X = 1) = 0.125 + 0.375 = 0.500$$

Visualizing probability distribution with a histogram

You can express the probability distribution for the coin-flipping experiment graphically with a *histogram*. A histogram is a graph in which you place individual values or ranges of values on the horizontal axis and the frequency of occurrence for each value or range of values on the vertical axis.

The histogram for the probability distribution of the coin-flipping experiment is shown in Figure 7-1. The vertical axis shows the probability of X , and the horizontal axis shows the value of X (that is, the number of heads.)

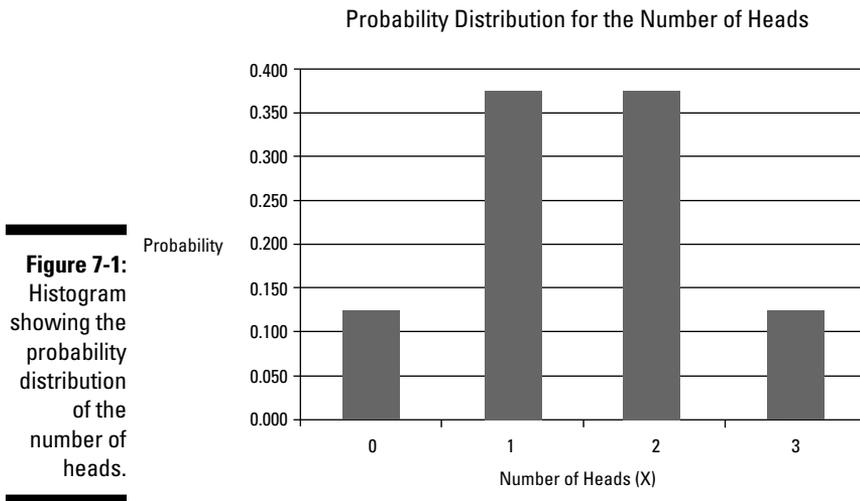


Illustration by Wiley, Composition Services Graphics

The histogram shows that the two most likely outcomes of this experiment are one head or two heads ($X = 1$ or $X = 2$); these are equally likely to occur. The least likely outcomes are no heads or three heads ($X = 0$ or $X = 3$); these are also equally likely to occur.

Characterizing a Probability Distribution with Moments

Recall from Chapters 3 through 5 that the properties of samples and populations may be summarized in a convenient form with a series of numerical *measures*, including the mean, variance, standard deviation, and so on.

The properties of a probability distribution can also be summarized with a set of numerical measures known as *moments*.

In this section, I cover the most important of these moments: expected value (mean) and the variance. (The standard deviation isn't a separate moment; it's the square root of the variance.) First, though, I explain the role of the summation operator in calculating these moments.

Understanding the summation operator (Σ)

The summation operator is used to indicate that a set of values should be added together. (The summation operator was introduced in Chapter 3.) The formulas used to compute moments for a probability distribution are based on the summation operator. This is because each calculation must be repeated for each possible value of a random variable and the results must be summed.

As an example of the summation operator, suppose a data set contains five elements. The summation operator tells you to perform the following calculations:

$$\sum_{i=1}^n X_i = X_1 + X_2 + X_3 + X_4 + X_5$$

X_i represents a single element in a data set; i is an *index*, and n is the number of elements to be summed.

Expected value

The *expected value* of a random variable X represents the average value of X that occurs if the random experiment is repeated a large number of times. You can think of the expected value as the *center* of the distribution.

The expected value is a *weighted average* of its possible values, with weights equal to probabilities. The formula for computing expected value of X is

$$E(X) = \sum_{i=1}^n X_i P(X_i)$$



Here are the key terms in this formula:

- ✓ $E(X)$ = the expected value of X
- ✓ n = the number of possible values of X
- ✓ i = an index
- ✓ X_i = one possible value of X
- ✓ $P(X_i)$ = the probability of X_i
- ✓ Σ = the summation operator used to indicate that a sum is being computed

Suppose that a biopharmaceutical firm is planning to release several new drugs during the coming year, depending on whether or not the patents are approved. You can use the random variable X to represent the number of new drugs that will be released.

Table 7-3 shows the probability distribution of these results.

X	$P(X)$
0	0.10
1	0.25
2	0.50
3	0.15

You can then use the probability distribution to determine the expected (average) value of X by setting up the possible values of X and the corresponding probabilities, like so:

$$X_1 = 0 \quad P(X_1) = 0.10$$

$$X_2 = 1 \quad P(X_2) = 0.25$$

$$X_3 = 2 \quad P(X_3) = 0.50$$

$$X_4 = 3 \quad P(X_4) = 0.15$$

The corresponding histogram is shown in Figure 7-2.

Figure 7-2: Probability distribution for the number of new drugs released.

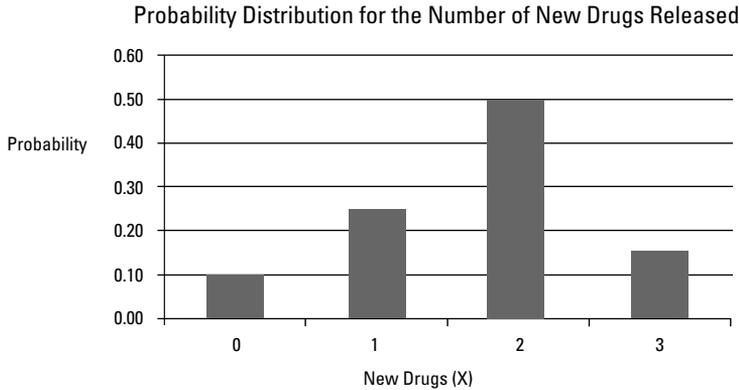


Illustration by Wiley, Composition Services Graphics

Next, you substitute these numbers into the expected value formula:

$$\begin{aligned}
 E(X) &= \sum_{i=1}^n X_i P(X_i) \\
 &= X_1 P(X_1) + X_2 P(X_2) + X_3 P(X_3) + X_4 P(X_4) \\
 &= (0)(0.10) + (1)(0.25) + (2)(0.50) + (3)(0.15) \\
 &= 0.00 + 0.25 + 1.00 + 0.45 \\
 &= 1.70
 \end{aligned}$$

This result shows that the expected (average) number of new drugs that will be released during the coming year is 1.7. Although it's physically impossible to release 1.7 new drugs (since 1.7 is not an *integer* or whole number), if this experiment is repeated many times, the average number of new drugs released will be 1.7.

Variance and standard deviation

The *variance* of a random variable X is the average squared distance between the values of X and the expected value of X . In other words, variance is the amount of “spread” among the different values of X . The standard deviation is simply the square root of the variance. Note that the variance and standard deviation of a random variable are equivalent to the variance and standard deviation of a sample or population (discussed in Chapter 4).

The formula for computing the variance of X is

$$\sigma^2 = \sum_{i=1}^n [X_i - E(X)]^2 P(X_i)$$



σ^2 represents the variance of X .

This expression tells you to perform the following calculations:

- ✓ For each possible value of X (X_i), subtract the expected value of X .
- ✓ Square the result.
- ✓ Multiply this expression by the probability of X_i .
- ✓ Compute the sum of these products.

For the example of the biopharmaceutical company (in the earlier section, “Understanding the summation operator $[\Sigma]$ ”) you compute the variance like so:

$$\begin{aligned}
 \sigma^2 &= \sum_{i=1}^n [X_i - E(X)]^2 P(X_i) \\
 &= [X_1 - E(X)]^2 P(X_1) + [X_2 - E(X)]^2 P(X_2) + [X_3 - E(X)]^2 \\
 &\quad P(X_3) + [X_4 - E(X)]^2 P(X_4) \\
 &= [0 - 1.7]^2 (0.10) + [1 - 1.7]^2 (0.25) + [2 - 1.7]^2 (0.50) + [3 - 1.7]^2 (0.15) \\
 &= (-1.7)^2 (0.10) + (-0.7)^2 (0.25) + (0.3)^2 (0.50) + (1.3)^2 (0.15) \\
 &= (2.89)(0.10) + (0.49)(0.25) + (0.09)(0.50) + (1.69)(0.15) \\
 &= 0.2890 + 0.1225 + 0.0450 + 0.2535 \\
 &= 0.7100
 \end{aligned}$$

One of the major drawbacks to the variance is that it’s measured in *squared units*, which makes interpretation difficult. In this example, the variance of the number of new drugs that will be released next year is 0.7100 *drugs squared*. It’s hard to visualize what “drugs squared” actually means. As a result, the standard deviation is normally used in place of variance as a measure of spread. By taking the square root of 0.7100 drugs squared, you get a result of 0.8426 drugs, which is much more intuitively clear.

$$\begin{aligned}
 \sigma &= \sqrt{\sum_{i=1}^n [X_i - E(X)]^2 P(X_i)} \\
 &= \sqrt{0.7100} \\
 &= 0.8426
 \end{aligned}$$

For the example of the biopharmaceutical company, the standard deviation of the number of new drugs released next year equals $\sigma = \sqrt{0.7100} = 0.8426$.

The standard deviation is 0.8426 new drugs. You can think of the standard deviation as a measure of how much *uncertainty* is associated with the expected value.



σ represents the standard deviation of X .

Chapter 8

The Binomial, Geometric, and Poisson Distributions

In This Chapter

- ▶ Finding probabilities when only two things can happen with the binomial distribution
 - ▶ Seeing how many “successes” or “failures” occur first with the geometric distribution
 - ▶ Using the Poisson distribution to calculate the probability of events occurring during a given time frame
-

You can model many complex business problems by using probability distributions. These distributions help provide answers to questions such as, “What’s the likelihood that oil prices will rise during the coming year?” “What’s the probability of a stock market crash next month?” “How likely is it that a corporation’s earnings will fall below expectations this year?” “What is the likelihood that three oil wells will have to be drilled before oil is found?”

A *probability distribution* defines the statistical properties of a variable. Accurate modeling of financial variables requires that you pick the appropriate distribution for a given situation. Some of the more widely used probability distributions in business are the binomial, geometric, and Poisson distributions. These are examples of *discrete distributions*, in which only a countable number of values are possible.

This chapter covers the key properties of the binomial, geometric, and Poisson distributions and explains the circumstances under which you may apply them. For each distribution, I give you formulas for computing probabilities and also provide tables as alternatives to doing the computing yourself.

This chapter also introduces summary measures of probability distributions, known as *moments*, which are closely related to the mean, variance, and standard deviation of samples and populations (described in Chapter 3). Then I wrap up the chapter by covering simplified formulas for computing the moments of the binomial, geometric, and Poisson distributions.

Looking at Two Possibilities with the Binomial Distribution

You use the *binomial distribution* to compute probabilities for processes where only one of two possible outcomes may occur. (The fact that only two possible outcomes can occur is what gives the distribution its name.) Here are some examples of processes you can model with the binomial distribution:

- ✓ When you flip a coin several times, the outcome of interest is whether the coin turns up heads or tails on each flip.
- ✓ When you roll a die multiple times, the outcome of interest is whether the number that turns up on each roll is odd (1, 3, or 5) or even (2, 4, or 6).
- ✓ When you look at the closing price of a stock each day for one year, the outcome of interest is whether the stock price increased or not.

As another example, suppose you hold a portfolio of stocks. During the coming year, it's possible that some of these stocks may split. (A stock split results in additional shares being distributed to existing shareholders.) For each stock, only two possible outcomes may occur: The stock splits, or it doesn't split. As a result, you can use the binomial distribution to compute the probability of a given number of splits in your portfolio over the coming year.



The binomial distribution is based on several specialized assumptions, which I explain in detail in the next section. If these assumptions aren't true, using the binomial distribution to compute probabilities for a given situation is likely to give inaccurate results.

Checking out the binomial distribution

You generate a binomial distribution by a special type of random experiment, known as a *binomial process*. This consists of a fixed number of repeated trials, each with only *two* possible outcomes and the following distinguishing features:

- ✓ **Each trial results in either a success or a failure.** On each trial of a binomial process, two possible outcomes may take place — and they're designated as "success" and "failure." For example, if you're doing a series of coin flips, you may call the outcome of the coin landing with "heads" up a success and the outcome of "tails" up a failure.
- ✓ **The trials are *independent* of each other.** Each trial of a binomial process is independent of previous trials; in other words, the outcome of one trial has no influence over the outcome of the other trials. For

example, the probability of heads turning up on a coin flip doesn't depend on the outcomes of flips that have taken place in the past.

- ✓ **The probability of success remains constant for all trials.** The probability of success in a binomial process doesn't change from one trial to the next; instead, it remains constant throughout the entire process. For example, the probability of a head turning up on a flip of a coin is always one-half (50 percent), no matter how many times the coin is flipped.

Computing binomial probabilities

You can compute the probability that a specified number of successes will occur during a fixed number of trials by using the binomial formula. For example, with this formula, you can determine the probability that five odd numbers turn up when a die is rolled ten times. The formula is:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Here's what each element of this formula means:

- ✓ X = a binomial random variable whose value is determined by the number of successes that occur during a series of trials
- ✓ x = the number of successes whose probability you are computing
- ✓ n = the number of trials that take place
- ✓ p = the probability of success on a single trial
- ✓ $(1 - p)$ = the probability of failure on a single trial
- ✓ $!$ = the factorial operator



The capital X is a binomial random variable (discussed in Chapter 7), and the lowercase x is a specific value, which refers to the number of successes whose probability you're calculating.

Factorial: counting how many ways you can arrange things

The exclamation point (!) doesn't just mean you're excited. The symbol is also the mathematical operator *factorial*. You pronounce $n!$ as "n factorial," which is the product of all positive integers less than or equal to n . For example:

$$0! = 1 \text{ (looks odd, but it's true)}$$

$$1! = 1$$

$$2! = (2)(1) = 2$$

$$3! = (3)(2)(1) = 6$$

$$4! = (4)(3)(2)(1) = 24$$

A general description is $n! = (n)(n-1)(n-2) \dots (2)(1)$. The factorial is a handy tool, but you can apply it only to 0 and positive integers.

You can use the factorial operator to count the number of ways you can arrange a group of objects. For example, suppose that a small bookshelf has enough room for three titles: *Algebra and Its Applications*; *Baseball: A History*; and *Chemistry in Everyday Life*. You can label these titles A, B, and C and then set up the possibilities for how many ways you can you arrange these books on the shelf like this:

ABC
ACB
BAC
BCA
CAB
CBA

This list covers every possibility. Each entry in the list is an *arrangement* of the three titles. Counting the number of elements in this list shows that you can arrange the books in six ways.

Fortunately, a much easier way to get this same result is to simply compute $3!$ (because three books are being arranged), giving a total number of arrangements of $3! = (3)(2)(1) = 6$.



Many calculators contain a built-in function for the factorial operator. It typically appears as $x!$ In Microsoft Excel, you can compute factorial with the function FACT.

Combinations: Counting how many choices you have

You use the combinations formula to count the number of *combinations* that can be created when choosing x objects from a set of n objects:

$$\frac{n!}{x!(n-x)!}$$

One distinguishing feature of a combination is that the order of objects is irrelevant.

For example, you can use this formula to count the number of ways you choose two elective classes from a set of eight for the upcoming semester. The order in which you choose the electives is immaterial; each possible selection is a *combination* of two objects.

As another example, suppose that you're painting your house with two colors from a set of four: green, blue, white, and yellow. Because the order in which you choose the colors is irrelevant, each pair of colors is a combination. How many different color schemes are possible with the given set of choices? You can answer this question by simply listing all the possible combinations:

green, white

green, blue

green, yellow

white, yellow

white, blue

blue, yellow

This list shows that you have six possible choices of pairs of colors.

The quicker way to answer this question is to substitute these values into the combinations formula; in this case, x represents the number of colors to choose (2), and n represents the total number of colors you can choose from (4).

$$\frac{n!}{x!(n-x)!} = \frac{4!}{2!(4-2)!} = \frac{4!}{2!2!} = \frac{24}{(2)(2)} = 6$$



The formula for computing the number of combinations is sometimes expressed as

$$\binom{n}{x}$$

Read or say this expression as “ n choose x .” This function appears on many calculators as nCr. In Microsoft Excel, you can compute combinations with the function COMBIN.



When you're selecting x objects from a group of n objects in such a way that the order of selection *does* matter, the choices are known as *permutations* instead of combinations.

Binomial formula: Computing the probabilities

Combinations are useful for computing binomial probabilities. You can find the probability of x successes during n trials with the binomial formula:

$$P(X = x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

Here,

$$\frac{n!}{x!(n-x)!}$$

is the total number of ways you can get exactly x successes during n trials, and

$$p^x(1-p)^{n-x}$$

is the probability of a sequence consisting of x successes and $(n-x)$ failures.

For example, say 40 percent of all published books are fiction, so the remaining 60 percent are nonfiction. If you pick six books at random from a bookstore, what's the probability that either none or one of them is fiction?

First, define fiction as a *success*. The probability of success on a single trial is $p = 0.4$, because 40 percent of all books are fiction. Each book you choose is a single trial of an experiment, so if you pick six books, you're conducting $n = 6$ trials for this experiment. You then figure the probability of getting one or fewer fiction books by calculating the probabilities of getting none and one fiction book and then adding them together:

- ✔ Based on the binomial formula, the probability of choosing *no* fiction books from a selection of six books is

$$P(X = 0) = \frac{6!}{0!(6-0)!}(0.40)^0(0.60)^6$$

$$P(X = 0) = (1)(1)(0.0467)$$

$$P(X = 0) = 0.0467$$

- ✔ Based on the binomial formula, the probability of choosing one fiction book from a selection of six is

$$P(X = 1) = \frac{6!}{1!(6-1)!}(0.40)^1(0.60)^5$$

$$P(X = 1) = (6)(0.40)(0.07776)$$

$$P(X = 1) = 0.1866$$

Now add the probabilities together. The probability of getting either no fiction book or one is $0.0467 + 0.1866 = 0.2333$. Alternatively, you can get these results from a binomial table for six trials ($n = 6$), such as Table 8-1.

Table 8-1 Binomial Probabilities that Result from 6 Trials ($n = 6$)

	$p = 0.1$	$p = 0.2$	$p = 0.3$	$p = 0.4$	$p = 0.5$
$x = 0$	0.5314	0.2621	0.1176	0.0467	0.0156
$x = 1$	0.3543	0.3932	0.3025	0.1866	0.0938
$x = 2$	0.0984	0.2458	0.3241	0.3110	0.2344
$x = 3$	0.0146	0.0819	0.1852	0.2765	0.3125
$x = 4$	0.0012	0.0154	0.0595	0.1382	0.2344
$x = 5$	0.0001	0.0015	0.0102	0.0369	0.0938
$x = 6$	0.0000	0.0001	0.0007	0.0041	0.0156

Table 8-1 shows the probability of success (p) at the top of each column. In this example, because $p = 0.4$, the probability of choosing zero fiction books is $P(X = 0) = 0.0467$ (found in the $x = 0$ row and the $p = 0.4$ column). The probability of choosing one fiction book is $P(X = 1) = 0.1866$ (found in the $x = 1$ row and the $p = 0.4$ column). The probability of getting no fiction books or one fiction book is the sum of $0.0467 + 0.1866$, or 0.2333 .



Check out a binomial table with 19 values for n at www.statisticshowto.com/tables/binomial-distribution-table.



If you simply don't like using formulas or tables to compute binomial probabilities, or if you want to triple-check your numbers, you can also use a specialized calculator, such as the Texas Instruments TI-83 or TI-84, which contains built-in functions that compute these probabilities quickly and easily. Or you can use the function BINOMDIST in Microsoft Excel 2007 and older versions, or BINOM.DIST in Excel 2010. If you need help with Excel, visit <http://office.microsoft.com/en-us/excel-help>.

Moments of the binomial distribution

Moments are summary measures of a probability distribution. The expected value represents the mean or average value of a distribution. The expected value is sometimes known as the *first moment* of a probability distribution. You calculate the expected value by taking each possible value of the distribution, weighting it by its probability, and then summing the results. The expected value is comparable to the mean of a population or sample (see Chapter 3).

The variance and standard deviation represent the dispersion among the possible values of a probability distribution. The variance and standard deviation of a probability distribution are equivalent to the variance and standard deviation of a population or sample. (The general formulas for computing moments for a discrete probability distribution are given in Chapter 7.) The variance is sometimes known as the *second central moment* of a probability distribution; the standard deviation isn't a separate moment, but simply the square root of the variance.

Luckily, for the binomial distribution, you can reduce computation time by using a series of simplified formulas, which I discuss in the following sections.

Binomial distribution: Calculating the expected value

The *expected value* of a probability distribution is its average value. You get it by weighting each possible value by its probability of occurring. For the binomial distribution, the calculation of the expected value can be simplified to

$$E(X) = np$$

For example, suppose that 10 percent of all people are left-handed, and 90 percent are right-handed (which happens to be true). In a class of 40 students, what's the expected number of left-handed students? You can calculate the expected value by thinking of each student as a "trial," with a 10 percent chance of being left-handed (a "success") and 90 percent chance of being right-handed (a "failure"). Therefore, $n = 40$ and $p = 0.10$. The expected number of left-handed students in the class is $E(X) = np = (40)(0.10) = 4$.

Binomial distribution: Computing variance and standard deviation

The *variance* of a distribution is the average squared distance between each possible outcome and the expected value. For the binomial distribution, you may compute the variance with the following simplified formula:

$$\sigma^2 = np(1 - p)$$

The *standard deviation* of a distribution equals the square root of the variance. For the binomial distribution, you calculate the standard deviation as

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

For the example of left-handed students in the previous section,

- ✓ The expected value is $E(X) = np = (40)(0.10) = 4$.
- ✓ The variance is $\sigma^2 = np(1 - p) = 40(0.10)(0.90) = 3.6$.
- ✓ The standard deviation is $\sqrt{3.6} = 1.9$.

Graphing the binomial distribution

You may want to illustrate the binomial distribution with a *histogram*. A histogram shows the possible values of a probability distribution as a series of vertical bars. The height of each bar reflects the probability of each value occurring. A histogram is a useful tool for visually analyzing the properties of a distribution, and (by the way) all discrete distributions may be represented with a histogram. (See Chapter 2 for more about histograms and other types of graphs.)

For example, suppose that a candy company produces both milk chocolate and dark chocolate candy bars. The product mix is 50 percent of the candy bars are milk chocolate and 50 percent are dark chocolate. Say you choose ten candy bars at random, and choosing milk chocolate is defined as a success. The probability distribution of the number of successes during these ten trials with $p = 0.5$ is shown in Figure 8-1.

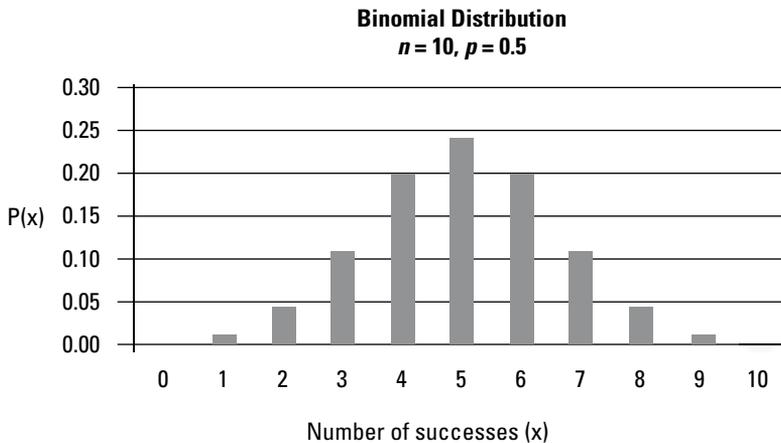


Figure 8-1:
Binomial
distribution:
ten trials
with $p = 0.5$.

Figure 8-1 shows that when $p = 0.5$, the distribution is *symmetric* about its expected value of 5 ($np = 10[0.5] = 5$), where the probabilities of X being below the mean match the probabilities of X being the same distance above the mean.

For example, with $n = 10$ and $p = 0.5$,

$$P(X = 4) = 0.2051 \text{ and } P(X = 6) = 0.2051$$

$$P(X = 3) = 0.1172 \text{ and } P(X = 7) = 0.1172$$

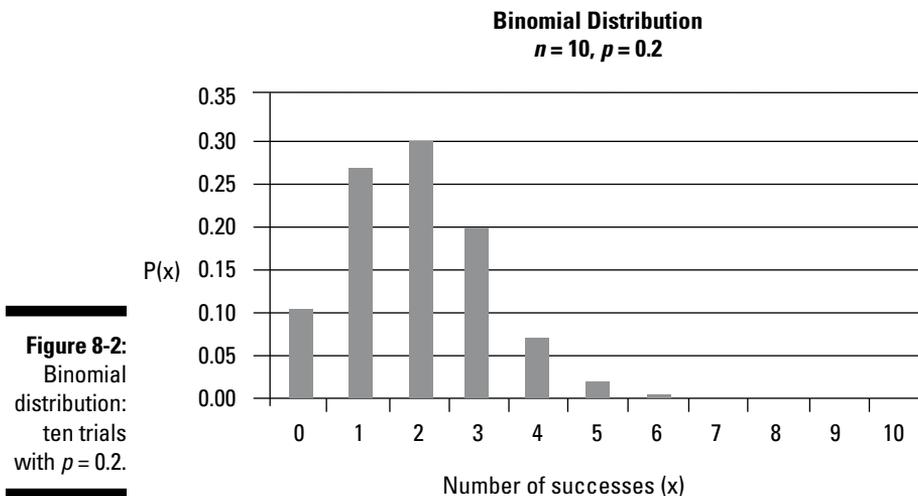
If the probability of success is less than 0.5, the distribution is *positively skewed*, meaning probabilities for X are greater for values below the expected value than above it.

For example, with $n = 10$ and $p = 0.2$,

$$P(X = 4) = 0.0881 \text{ and } P(X = 6) = 0.0055$$

$$P(X = 3) = 0.2013 \text{ and } P(X = 7) = 0.0008$$

Figure 8-2 shows the probability distribution for $n = 10$ and $p = 0.2$.



If the probability of success is greater than 0.5, the distribution is *negatively skewed* — probabilities for X are greater for values above the expected value than below it.

For example, with $n = 10$ and $p = 0.8$,

$$P(X = 4) = 0.0055 \text{ and } P(X = 6) = 0.0881$$

$$P(X = 3) = 0.0008 \text{ and } P(X = 7) = 0.2013$$

Figure 8-3 shows the probability distribution for the same situation when $p = 0.8$.

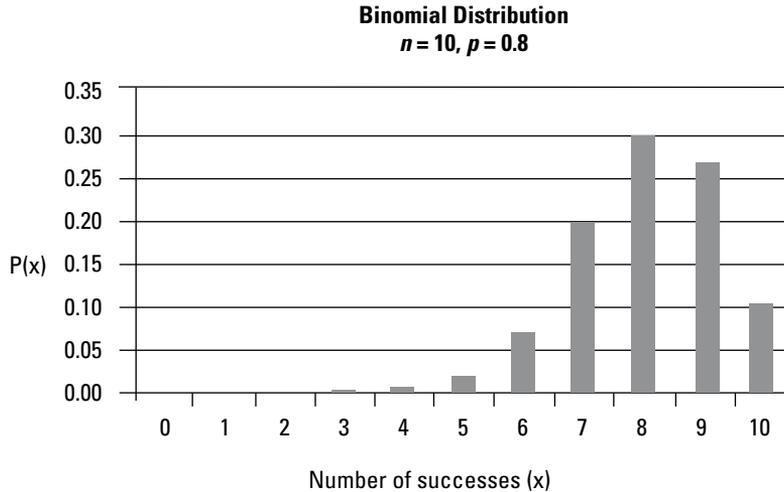


Figure 8-3:
Binomial
distribution:
ten trials
with $p = 0.8$.

Determining the Probability of the Outcome That Occurs First: Geometric Distribution

The geometric distribution is based on the binomial process. (That is, a series of independent trials with two possible outcomes. See the earlier section “Checking out the binomial distribution.”) You use the geometric distribution to determine the probability that a specified number of trials will take place before the first success occurs. Alternatively, you can use the geometric distribution to figure the probability that a specified number of failures will occur before the first success takes place.

The following section explains how to compute geometric probabilities and also how to compute the moments of the geometric distribution. You also see graphs that illustrate the properties of the geometric distribution.

Computing geometric probabilities

To calculate the probability that a given number of trials take place until the first success occurs, use the following formula:

$$P(X = x) = (1 - p)^{x-1}p \text{ for } x = 1, 2, 3, \dots$$

Here, x can be any whole number (*integer*); there is no maximum value for x .

X is a geometric random variable, x is the number of trials required until the first success occurs, and p is the probability of success on a single trial.

For example, suppose you want to flip a coin until the first heads turns up. The probability that it takes four flips for the first heads to occur (that is, three tails followed by one heads) is $P(X = x) = (1 - p)^{x-1}p$. In this example, $x = 4$ and $p = 0.5$:

$$P(X = 4) = (1 - 0.5)^3(0.5) = (0.125)(0.5) = 0.0625$$

To calculate the probability that a given number of failures occur before the first success, the formula is

$$P(X = x) = (1 - p)^x p$$

x now represents the number of failures that occur before the first success. In addition, x can assume values 0, 1, 2, . . . instead of 1, 2, 3, . . .

For example, suppose you flip a coin until the first heads turns up. The probability that there will be three tails before the first heads turns up is $P(X = x) = (1 - p)^x p$. In this example, $x = 3$ and $p = 0.5$:

$$P(X = 3) = (1 - 0.5)^3(0.5) = (0.5)^3(0.5) = (0.125)(0.5) = 0.0625$$

Both situations refer to getting three tails followed by a heads, so both formulas provide the same result.

Moments of the geometric distribution

The moments (see the earlier section “Moments of the binomial distribution” for a definition) of the geometric distribution depend on which of the following situations is being modeled:

- ✓ The number of trials required before the first success takes place
- ✓ The number of failures that occur before the first success

Just as with the binomial distribution discussed earlier in this chapter, the geometric distribution has a series of simplified formulas for computing these moments, which I explore in the following sections.

Geometric distribution: Calculating the expected value

The expected value of the geometric distribution when determining the number of trials required until the first success is

$$E(X) = \frac{1}{p}$$

The expected value of the geometric distribution when determining the number of failures that occur before the first success is

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

For example, when flipping coins, if success is defined as “a heads turns up,” the probability of a success equals $p = 0.5$; therefore, failure is defined as “a tails turns up” and $1 - p = 1 - 0.5 = 0.5$. On average, there’ll be $(1 - p)/p = (1 - 0.5)/0.5 = 0.5/0.5 = 1$ tails before the first heads turns up.

Notice how the two results provide the same information; it takes an average of two flips to get the first heads, or on average there should be one tails before the first heads turns up.

Geometric distribution: Computing variance and standard deviation

The variance and standard deviation of the geometric distribution when determining the number of trials required until the first success or when determining the number of failures that occur before the first success are

$$\sigma^2 = \frac{1-p}{p^2}$$

$$\sigma = \sqrt{\frac{1-p}{p^2}}$$

For example, suppose you flip a coin until the first heads turns up. The expected number of trials required until the first heads turns up is

$$E(X) = \frac{1}{p} = \frac{1}{0.5} = 2$$

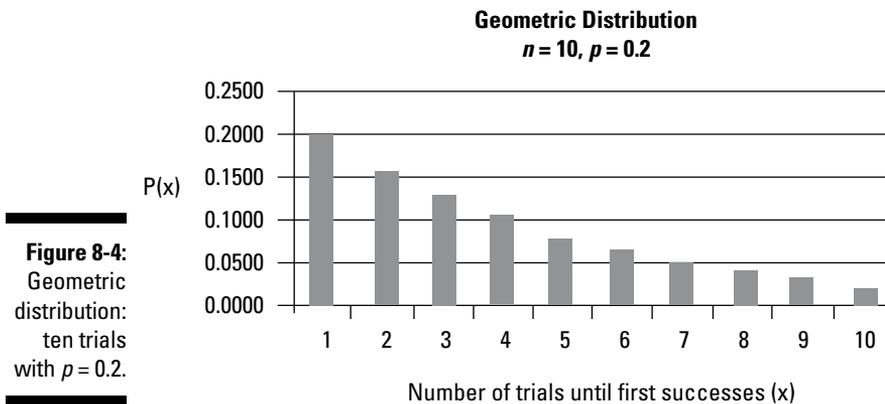
The variance is

$$\sigma^2 = \frac{1-p}{p^2} = \frac{1-0.5}{0.5^2} = \frac{0.5}{0.25} = 2$$

The standard deviation (σ) is $\sqrt{2} = 1.414$.

Graphing the geometric distribution

You can illustrate the geometric distribution with a histogram. For example, say you do a series of ten trials. On each trial, the probability of success is 0.2. Figure 8-4 shows the probability distribution of the number of trials required to reach the first success.



Unlike the binomial distribution, the geometric distribution is positively skewed for any value of p .

Keeping the Time: The Poisson Distribution

The *Poisson distribution* is useful for measuring how many events may occur during a given time horizon, such as the number of customers that enter a store during the next hour, the number of hits on a website during the next minute, and so forth. The *Poisson process* takes place over time instead of a series of trials; each interval of time is assumed to be *independent* of all other intervals.

For example, suppose that a bank counts the number of customers who enter each hour. If the number of customers that enter during a given hour is independent of the number that enter during all other hours (while the bank is open), you can use the Poisson distribution to find the probability that a specific number of customers enter the bank during the next hour.



The Poisson distribution is named for Siméon Denis Poisson who was a French mathematician, physicist, and genius. He was wrong about only one major thing: He opposed the wave theory of light.

The following section shows you how to compute Poisson probabilities and how to compute moments for the Poisson distribution. Graphs are used to illustrate the key properties of the Poisson distribution.

Computing Poisson probabilities

You calculate Poisson probabilities with the following formula:

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

Here's what each element of this formula represents:

- ✓ X = a Poisson random variable
- ✓ x = number of events whose probability you are calculating
- ✓ λ = the Greek letter “lambda,” which represents the average number of events that occur per time interval
- ✓ e = a *constant* that's equal to approximately 2.71828



e is a constant that's widely used in financial applications. One of the most important uses is in computing present values of sums of money when interest rates are *continuously compounded* — compounded an *infinite* number of times. Most calculators have a key labeled e^x that you can use to calculate the value of e raised to a specified power. In Excel, the appropriate function for determining the value of e is EXP.

For example, suppose that the number of messages that a person receives on his cellphone averages one per hour and that the number of messages received each hour is independent of all other hours. What's the probability of his receiving two messages in the next hour?

In this case, the value of lambda (λ) is equal to 1, because the average number of messages each hour equals 1. The probability of receiving two messages during the next hour is

$$P(X = 2) = e^{-1} \frac{1^2}{2!} = 0.1839$$

Alternatively, you can get results from a Poisson table set up like Table 8-2.

Table 8-2 Poisson Probabilities for Different Values of λ

	$\lambda = 0.5$	$\lambda = 1$	$\lambda = 1.5$	$\lambda = 2$	$\lambda = 2.5$	$\lambda = 3$
$x = 0$	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498
$x = 1$	0.3033	0.3679	0.3347	0.2707	0.2052	0.1494
$x = 2$	0.0758	0.1839	0.2510	0.2707	0.2565	0.2240
$x = 3$	0.0126	0.0613	0.1255	0.1804	0.2138	0.2240
$x = 4$	0.0016	0.0153	0.0471	0.0902	0.1336	0.1680
$x = 5$	0.0002	0.0031	0.0141	0.0361	0.0668	0.1008
$x = 6$	0.0000	0.0005	0.00035	0.0120	0.0278	0.0504
$x = 7$	0.0000	0.0001	0.0008	0.0034	0.0099	0.0216
$x = 8$	0.0000	0.0000	0.0001	0.0009	0.0031	0.0081

Table 8-2 shows the Poisson probabilities for different values of λ . In the cellphone example, because $x = 2$ and $\lambda = 1$, the appropriate probability $P(X = 2)$ is found in the $x = 2$ row and the $\lambda = 1$ column. The probability is 0.1839.



If you don't care for using formulas or a table, try a specialized calculator or Excel. For Excel 2007 and older versions, use the POISSON function; for Excel 2010, use the POISSON.DIST function.

The moments of the Poisson distribution are used to represent the average value of the distribution and the dispersion of the distribution. As with the binomial and geometric distribution, these moments may be computed with simplified formulas.

Poisson distribution: Calculating the expected value

As with the binomial and geometric distributions (discussed earlier in this chapter), you can use simple formulas to compute the moments of the Poisson distribution. The expected value of the Poisson distribution is

$$E(X) = \lambda$$

For example, say that on average three new companies are listed in the New York Stock Exchange (NYSE) each year. The number of new companies listed during a given year is independent of all other years. The number of new listings per year, therefore, follows the Poisson distribution, with a value of $\lambda = 3$. As a result, the expected number of new listings next year is $\lambda = 3$.

Poisson distribution: Computing variance and standard deviation

Compute the variance and the Poisson distribution as $\sigma^2 = \lambda$; the standard deviation (σ) equals $\sqrt{\lambda}$.

Based on the NYSE listing example in the previous section, the variance equals 3 and the standard deviation equals $\sqrt{3} = 1.732$.

Graphing the Poisson distribution

As with the binomial distribution, the Poisson distribution can be illustrated with a histogram. In Figures 8-5 through 8-7, the results are shown for three values of λ : 2 (Figure 8-5), 5 (Figure 8-6) and 7 (Figure 8-7).

For $\lambda = 2$ (Figure 8-5), the distribution is skewed to the right; for $\lambda = 5$ (Figure 8-6), the distribution is nearly symmetric about the mean of 5; for $\lambda = 7$ (Figure 8-7), the distribution is skewed to the left.

Figure 8-5:
Poisson distribution with $\lambda = 2$

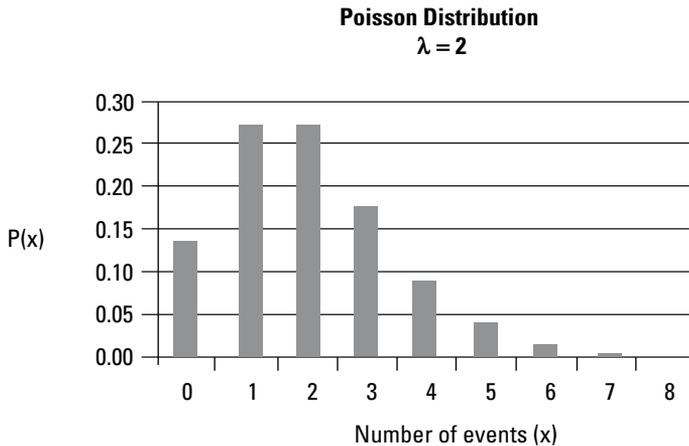
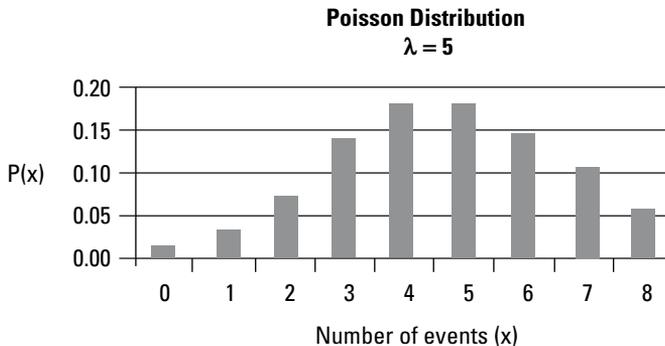
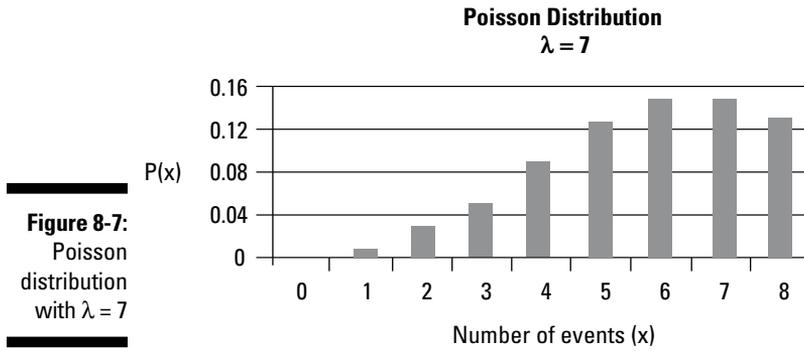


Figure 8-6:
Poisson distribution with $\lambda = 5$





Chapter 9

The Uniform and Normal Distributions: So Many Possibilities!

In This Chapter

- ▶ Understanding the differences between discrete and continuous distributions
 - ▶ Discovering the properties of the uniform distribution
 - ▶ Checking out normal distribution probabilities
-

This chapter introduces two important new probability distributions: the uniform and the normal. The normal distribution is especially important in business applications; it can be used to describe the behavior of many financial variables, such as the rate of return to an investment, a corporation's annual profits, consumer spending on new products, and so on.

The uniform and the normal distributions have one important feature in common: they assign probabilities to *ranges* of values instead of individual values. This contrasts with the distributions found in Chapter 8: the binomial, geometric, and Poisson; these distributions assign probabilities to *individual* values.

The *uniform* distribution is used to describe a situation where all possible outcomes of a random experiment are equally likely to occur. For example, suppose that a manufacturer produces one-liter bottles of soda. The goal is to fill each bottle with exactly one liter of soda, but in actual practice, the acceptable range is between 0.99 and 1.01 liters. Any bottles that fall outside of this range are discarded. Suppose that for each acceptable bottle, the content is equally likely to be any value between 0.99 and 1.01 liters. In this case, the uniform distribution could be used to answer questions such as:

What is the likelihood that a randomly chosen bottle contains between 0.992 and 0.994 liters?

What is the likelihood that a randomly chosen bottle contains more than 1 liter?

What is the average content of the acceptable bottles?

In this chapter I demonstrate how uniform probabilities may be determined with a graph or with an algebraic formula. I also show how the moments of the uniform distribution may be computed.

The *normal* distribution is the most widely used distribution in business because you can use it to model many variables. For example, you can use the normal distribution to describe the rates of return to financial assets, the distribution of corporate profits, the prices of key commodities (such as oil), and so forth.

Suppose that the returns to the stocks in the Standard and Poor's 500 (S&P 500) index are normally distributed. The normal distribution could then be used to answer questions such as:

What is the probability that the S&P 500 will increase by at least 5 percent next year?

What is the probability that the S&P 500 will fall next year?

How much risk is associated with investing in the S&P 500?

Due to the complexity of the normal distribution, I show you how to compute normal probabilities with standard tables in this chapter instead of formulas.

The following sections explain the differences between the two basic types of probability distributions: discrete and continuous. There is a detailed look at the properties of the uniform and normal distributions, including techniques for computing probabilities and moments.

Comparing Discrete and Continuous Distributions

Discrete and continuous distributions are the two standard types of *probability distributions*, which you use to compute probabilities for possible outcomes of a random experiment. (For more about random experiments and probability distributions, see Chapter 7.)

- ✔ You use the **discrete distribution** with a random experiment that can generate a *finite* (countable) number of outcomes. (You see three examples of discrete distributions — binomial, geometric, and Poisson — in Chapter 8.)
- ✔ You use the **continuous distribution** with a random experiment that can generate an *infinite* (uncountable) number of outcomes.



Intuitively, a random experiment can generate a finite (countable) number of outcomes if it's possible to make up a list of all the possible outcomes of the experiment. For example, if a coin is flipped ten times, and *heads turns up* is the variable of interest, then there are 11 possible outcomes: 0, 1, 2, ..., 10. These outcomes could be easily listed. On the other hand, if an experiment consists of observing the length of time until the next phone call arrives, the number of possible times until the next phone call is *infinite (uncountable)*. This is because the times are not restricted to whole numbers. The time could be 2.3 seconds, 1.41742 seconds, 8.19444212 seconds, and so on. A list containing all possible times until the next phone call is impossible to construct, because there are an unlimited number of entries.

Computing probabilities for continuous distributions is more complex than for a discrete distribution; often, your best resources are tables or specialized calculators. For an example, visit www.solveymath.com/online_math_calculator/statistics/continuous_distributions/index.php.

Aside from the number of possible outcomes, one of the most important differences between discrete and continuous distributions is this: With a continuous distribution, the probability that a random variable (X) equals a specific constant (x) is defined as *zero*. With an infinite number of possibilities, the likelihood of X being equal to a specific value is infinitesimally small.

For example, the probability of tomorrow's temperature at noon being exactly 72.141712987 degrees is pretty much zero. As a result, for any value x , $P(X \leq x)$ equals $P(X < x)$. A statement such as "the probability that the temperature at noon tomorrow will be less than or equal to 72 degrees" has the same interpretation as "the probability that the temperature at noon tomorrow will be less than 72 degrees."

To demonstrate this statement mathematically, you can write $P(X \leq x)$ as $P(X < x) + P(X = x)$, because the probability that X is less than or equal to x consists of the sum of two different probabilities — the probability that X is strictly less than x and the probability that X is exactly equal to x . With a continuous distribution, $P(X = x) = 0$; therefore,

$$P(X \leq x) = P(X < x) + P(X = x)$$

$$P(X \leq x) = P(X < x) + 0$$

$$P(X \leq x) = P(X < x)$$



Based on this reasoning, $P(X \geq x) = P(X > x)$ is also true.

With a discrete distribution, $P(X \leq x)$ does *not* equal $P(X < x)$, and $P(X \geq x)$ does *not* equal $P(X > x)$ unless $P(X = x) = 0$.

For example, suppose that a coin is flipped three times. The outcome of interest is whether a head turns up on each flip.

The probability that two or fewer heads turns up is computed as:

$$P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$$

The probability that fewer than two heads turn up is computed as:

$$P(X < 2) = P(X = 0) + P(X = 1)$$

Therefore, unless $P(X = 2) = 0$, $P(X \leq 2)$ and $P(X < 2)$ gives different results.

In the continuous case, though, $P(X \leq 2)$ and $P(X < 2)$ is always equal.

Working with the Uniform Distribution

The *uniform distribution* is a continuous distribution that assigns only positive probabilities within a specified interval (a, b) — that is, all values between a and b . (a and b are two constants; they may be negative or positive.)

For example, suppose that the U.S. Postal Service offers a special new delivery service; it's guaranteed that the time required for a package to be delivered from New York City to Los Angeles is no more than 72 hours. (It also takes at least 24 hours for the package to be delivered.) If the delivery time is equally likely to be any value between 24 and 72 hours, then the uniform distribution can be used to compute probabilities for the delivery time. For example, suppose that a customer wants to know the likelihood that the package will be delivered between 24 and 36 hours after mailing; this can be computed with the uniform distribution.

In this case, the uniform distribution is defined over the interval $(24, 72)$. (In other words, $a = 24$ and $b = 72$.) This implies that the probability of a package arriving in less than 24 hours or more than 72 hours equals 0. Furthermore, the probability of the package arriving within any given interval between 24 and 72 hours depends only on the *width* of the interval. For example, the package is just as likely to arrive in 24 to 28 hours as it is to arrive in 68 to 72 hours, because both of these intervals have a width of four hours.

Although the uniform distribution may be defined for an infinite number of different intervals, if the distribution is defined over the interval $(0, 1)$ it's known as the *standard uniform distribution*. This indicates that all values between 0 and 1 are equally likely to occur.

The standard uniform distribution is used for random experiments where the outcome is equally likely to be any value between 0 and 1. For example, because probabilities are always between 0 and 1, the standard uniform distribution can be used to describe a random process that generates probabilities.

The standard uniform distribution is often used for *simulation* studies, in which the value of a variable is estimated by repeatedly choosing random numbers and substituting them into a mathematical model. For example, the sales of a new product could be estimated by choosing values from the standard uniform distribution and substituting the results into a model of consumer demand.

The uniform distribution and the standard uniform distribution are discussed at www.en.wikipedia.org/wiki/Uniform_distribution.

In the following sections, I explore the uniform distribution and all it has to offer, including how to visualize its intervals on a graph, how to calculate its moments, and how to work with its probabilities.

Graphing the uniform distribution

A discrete distribution may be described with a *histogram*, which is a special type of graph consisting of a series of vertical bars. Each bar represents a value or range of values, and the height of each bar represents the probability of that value or range of values. (Histograms are introduced in Chapter 7.)

A continuous distribution can't be illustrated with a histogram, because this would require an *infinite* number of bars. Instead, a continuous distribution may be illustrated with a line or a curve. Areas under the line or the curve correspond to probabilities.

With the uniform distribution, all values over an interval (a, b) are equally likely to occur. As a result, the graph that illustrates this distribution is a *rectangle*. Figure 9-1 shows the uniform distribution defined over the interval $(0, 10)$.

The horizontal axis shows the range of values for X (0 to 10). The distribution assigns a probability of 0 to any value of X outside of the interval from 0 to 10.

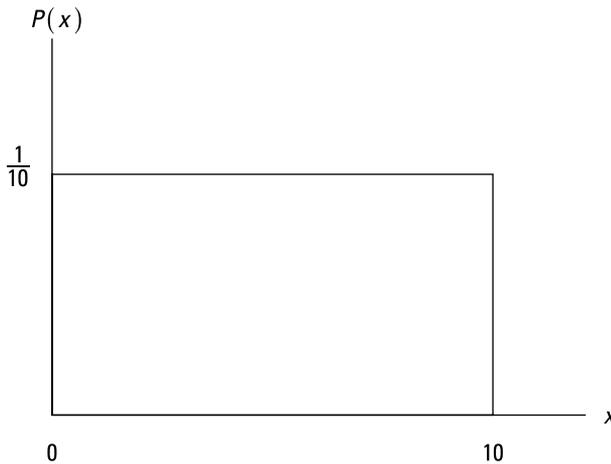


Figure 9-1:
The uniform
distribution
defined over
the interval
(0, 10).

The *width* of this interval equals the upper limit (b) minus the lower limit (a), which equals $b - a$. So in Figure 9-1, the width equals $10 - 0 = 10$. The width of this interval represents the *base* of the rectangle. The *height* of the rectangle equals 1 divided by the base ($1/10$ in this case). The height always equals 1 divided by the base; this ensures that the area of the rectangle always equals 1. As discussed in the later section “Computing uniform probabilities,” areas under this rectangle represent probabilities. The total probability for any distribution is 1; therefore, the area under the rectangle must equal 1.



The area of a rectangle equals the base times the height, or in mathematical terms, $A = b \times h$.

Discovering moments of the uniform distribution

Moments are a set of summary measures that express the properties of the probability distribution of a random variable. (For more about the moments of a probability distribution, see Chapter 7.) The moments include expected value (mean) and variance. Standard deviation is not a separate moment, but is instead the square root of the variance.

As discussed in Chapter 7, the expected value represents the average value of all the possible values of a probability distribution, weighted by the probabilities of these values. The variance and standard deviation measure the “spread” among the possible values of the distribution.

For example, suppose that an art gallery sells two types of art work: inexpensive prints and original paintings. The length of time that the prints remain in

inventory is uniformly distributed over the interval (0,40). For example, some prints are sold immediately; no print remains in inventory for more than 40 days. For the paintings, the length of time in inventory is uniformly distributed over the interval (5, 105). For example, each painting requires at least 5 days to be sold and may take up to 105 days to be sold.

The expected value, variance, and standard deviation are much lower for the prints because the range of possible values is much smaller. On average, prints sell much faster than paintings. In addition, the inventory times of the prints are much closer to each other than for the paintings. The uniform distribution has simple formulas for calculating the moments, which I describe in the following sections.

Uniform distribution: Calculating the expected value

For any probability distribution, the expected value represents the *average* value of the distribution. For the uniform distribution, you calculate the expected value as the midpoint of the interval over which the distribution is defined.

For example, suppose that the uniform distribution is defined over the interval (a, b) . You calculate the expected value as

$$E(X) = \frac{b+a}{2}$$

The key terms in this formula are

- ✓ X = a uniformly distributed random variable defined over the interval (a, b)
- ✓ $E(X)$ = the expected value of X
- ✓ a = the lower limit of the interval
- ✓ b = the upper limit of the interval

The expected value formula for the uniform distribution is illustrated in Figure 9-2.

The graph in Figure 9-2 shows that the expected value is the *midpoint* of the interval (a,b) . In other words, it's half-way between a and b .

As an example, the expected value of the uniform distribution defined over the interval $(1,5)$ is computed as follows:

$$E(X) = \frac{b+a}{2}$$

$$E(X) = \frac{5+1}{2} = 3$$

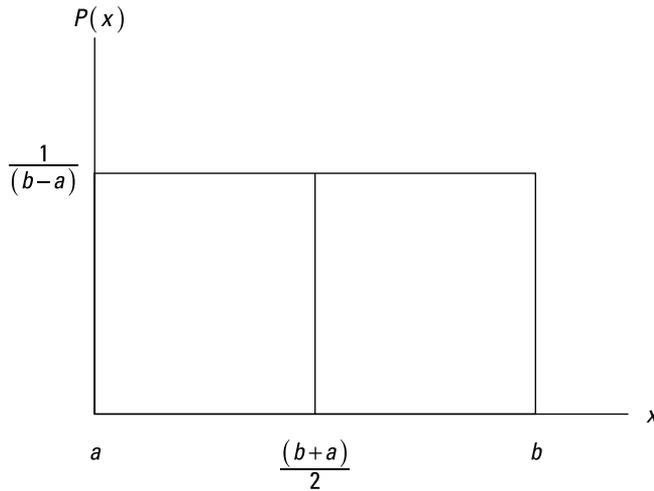


Figure 9-2:
The
expected
value of the
uniform
distribution.

Uniform distribution: Computing variance and standard deviation

In addition to the expected value, a probability distribution can be characterized by the variance and the standard deviation. These values measure the degree of *dispersion* (spread) among the values of a probability distribution.

For the uniform distribution defined over the interval from a to b , the variance equals

$$\sigma^2 = \frac{(b-a)^2}{12}$$

The standard deviation is the square root of the variance:

$$\sigma = \frac{(b-a)}{\sqrt{12}}$$

For example, the variance of the uniform distribution defined over the interval (1, 5) is computed as follows:

$$\begin{aligned}\sigma^2 &= \frac{(b-a)^2}{12} \\ &= \frac{(5-1)^2}{12} \\ &= \frac{16}{12} \\ &= 1.333\end{aligned}$$

The standard deviation is:

$$\begin{aligned}\sigma &= \frac{(b-a)}{\sqrt{12}} \\ &= \frac{(5-1)}{\sqrt{12}} \\ &= \frac{4}{\sqrt{12}} \\ &= 1.1547\end{aligned}$$

Computing uniform probabilities

You can compute probabilities for the uniform distribution with formulas or graphs. When using graphs to compute uniform probabilities, you are computing areas within the rectangle that describes the uniform distribution.

Computing uniform probabilities with formulas

For example, suppose the random variable X is uniformly distributed over the interval (a, b) . You compute the probability that X is less than or equal to a specified value of x , using this formula:

$$P(X \leq x) = \frac{x-a}{b-a}$$

If, for example, X is a uniform random variable with $a = 0$ and $b = 10$. You find the probability that X is less than or equal to 7 by these calculations:

$$P(X \leq x) = \frac{x-a}{b-a}$$

$$P(X \leq 7) = \frac{7-0}{10-0}$$

$$P(X \leq 7) = 0.7$$

To determine the probability that X is greater than or equal to x , use the following formula:

$$P(X \geq x) = 1 - \frac{x-a}{b-a}$$

This is true because with a continuous random variable,

$$P(X \leq x) + P(X \geq x) = 1$$

For a continuous random variable X , either $X \leq x$ or $X \geq x$ must be true; therefore, the probabilities of these events must sum to 1. (Recall from Chapter 6

that these events are *complements*.) So rearranging algebraically gives you the following:

$$P(X \geq x) = 1 - P(X \leq x)$$

$$P(X \geq x) = 1 - \left(\frac{x-a}{b-a} \right)$$

As an example, to calculate the probability that a uniform random variable X defined over the interval $(0, 10)$ is greater than or equal to 2, apply the formula and solve:

$$P(X \leq x) = 1 - \left(\frac{x-a}{b-a} \right)$$

$$P(X \leq 2) = 1 - \left(\frac{2-0}{10-0} \right)$$

$$P(X \leq 2) = 1 - 0.2 = 0.8$$

$$P(X \leq 0.2) = 0.8$$

To calculate the probability that X is *between* two constants a and b , use the following formula:

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a)$$

For example, you compute the probability that a uniform random variable X defined over the interval $(0, 10)$ is between 3 and 6 as

$$P(3 \leq X \leq 6) = P(X \leq 6) - P(X \leq 3)$$

and follow these steps:

1. Determine the probability that X is less than or equal to 6:

$$P(X \leq x) = \frac{x-a}{b-a}$$

$$P(X \leq 6) = \frac{6-0}{10-0}$$

$$P(X \leq 6) = 0.6$$

2. Compute the probability that X is less than or equal to 0.3:

$$P(X \leq x) = \frac{x-a}{b-a}$$

$$P(X \leq 3) = \frac{3-0}{10-0}$$

$$P(X \leq 3) = 0.3$$

3. Combine the results:

$$P(3 \leq X \leq 6) = 0.6 - 0.3$$

$$P(3 \leq X \leq 6) = 0.3$$

One of the unique properties of the uniform distribution is that the probability that X falls within a given range of values depends only on the *width* of the range. For example, for the *standard* uniform distribution, the following probabilities are equal:

$$P(0.1 \leq X \leq 0.2)$$

$$P(0.3 \leq X \leq 0.4)$$

$$P(0.7 \leq X \leq 0.8)$$

Each of these probabilities equals 0.1, which you can compute as $P(0.1 \leq X \leq 0.2) = P(X \leq 0.2) - P(X \leq 0.1)$

Then follow these steps:

1. Determine the probability that X is less than or equal to 0.2:

$$P(X \leq x) = \frac{x-a}{b-a}$$

$$P(X \leq 0.2) = \frac{0.2-0}{1-0}$$

$$P(X \leq 0.2) = 0.2$$

2. Compute the probability that X is less than or equal to 0.1:

$$P(X \leq x) = b - a$$

$$P(X \leq 0.1) = \frac{0.1-0}{1-0}$$

$$P(X \leq 0.1) = 0.1$$

3. Combine the results:

$$P(0.1 \leq X \leq 0.2) = 0.2 - 0.1$$

$$P(0.1 \leq X \leq 0.2) = 0.1$$

Computing uniform probabilities with graphs

You can also compute probabilities graphically for the uniform distribution by computing areas under a rectangle (see the earlier section “Graphing the uniform distribution”). For example, Figure 9-3 shows the probability that a standard uniform random variable X is between 0.3 and 0.6.

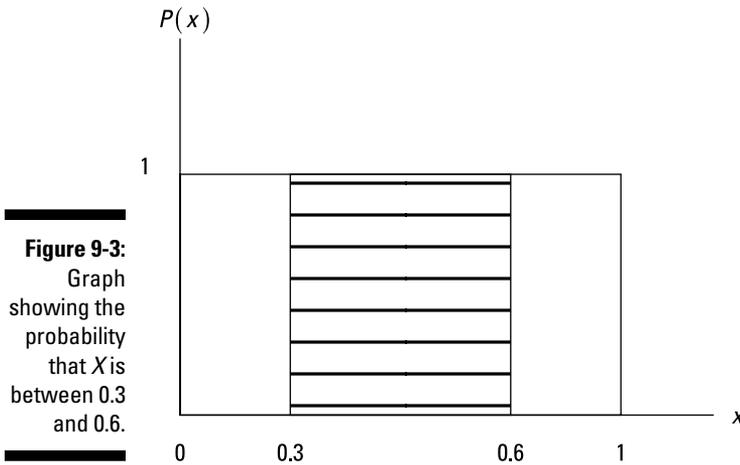


Figure 9-3:
Graph
showing the
probability
that X is
between 0.3
and 0.6.

The horizontal axis shows that the distribution is defined over the interval from 0 to 1. The width of this interval, which is the base of the rectangle, equals $1 - 0 = 1$. The height of the rectangle equals 1 divided by the base, or $1/1 = 1$. The area of the rectangle equals the base times the height, which is $1 \times 1 = 1$.

To find the probability that X is between 0.3 and 0.6, you compute an area within the rectangle (see the shaded region in Figure 9-3). The base of this shaded region equals $0.6 - 0.3 = 0.3$. The height equals 1. Therefore, the area equals 0.3 (0.3×1). The probability that X is between 0.3 and 0.6 is 0.3, which matches the result found with the algebraic formula.

Understanding the Normal Distribution

The normal distribution is a continuous probability distribution that can be used to describe a large number of different situations, not just in business applications but in a wide variety of other disciplines, such as psychology, sociology, biology, and so on. The normal distribution, sometimes called the *Gaussian distribution*, is named after scientist and mathematician Johann Carl Friedrich Gauss who introduced the concept.

The normal distribution has several useful properties that can be used to describe real-world events. For example, under the normal distribution, there is a balance or *symmetry* between the likelihood of a value being below the mean of the distribution and being above the mean of the distribution.

As an example, suppose that researchers have determined that the heights of all men in a country are normally distributed with a mean of 69 inches and a standard deviation of 2 inches. Based on the normal distribution, the following events are equally likely:

A randomly chosen man is no more than 67 inches tall

A randomly chosen man is at least 71 inches tall

These events are equally likely because:

A height of 67 inches is one standard deviation below the mean
($69 - 1(2) = 67$)

A height of 71 inches is one standard deviation above the mean
($69 + 1(2) = 71$)

Similarly, the following events are equally likely:

A randomly chosen man is no more than 65 inches tall

A randomly chosen man is at least 73 inches tall

These events are equally likely because:

A height of 65 inches is two standard deviations below the mean
($69 - 2(2) = 65$)

A height of 73 inches is two standard deviations above the mean
($69 + 2(2) = 73$)

Because the normal distribution is a *continuous* distribution, it's defined for an infinite number of values. Unlike the uniform distribution, the normal distribution is defined for *all* values between negative infinity and positive infinity.

In the following sections, I show you how you can express the normal distribution graphically, I introduce you to the standard normal distribution, and I walk you through calculating probabilities for the normal distribution.

Graphing the normal distribution

The normal distribution can be graphed with a special type of curve, which is usually described as a *bell-shaped curve*. Normal probabilities can be determined by computing areas under this curve.

The bell-shaped curve has several key features. It's defined over the entire range of values between negative and positive infinity; it's *symmetrical* about the mean (for example, the area below the mean is a *mirror image* of the area above the mean); and most of the area under the normal distribution is close to the mean. The area declines rapidly for values that are several standard deviations away from the mean.

As an example, the distribution of heights from the previous example is illustrated with a bell-shaped curve as shown in Figure 9-4.

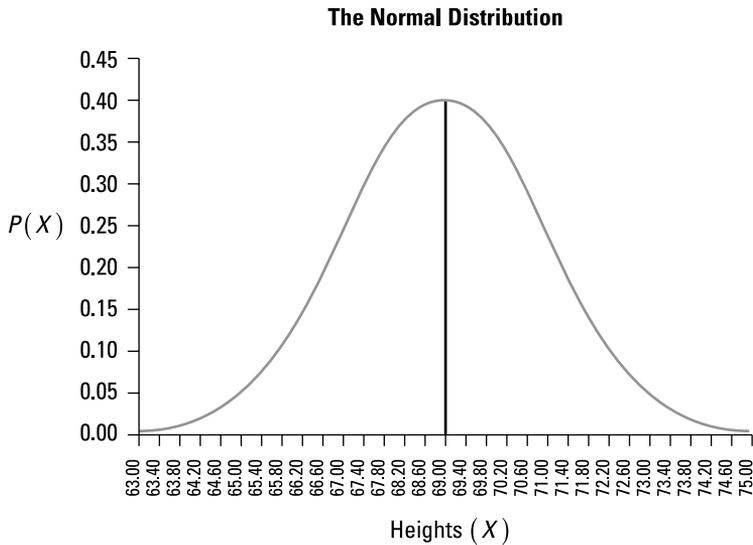
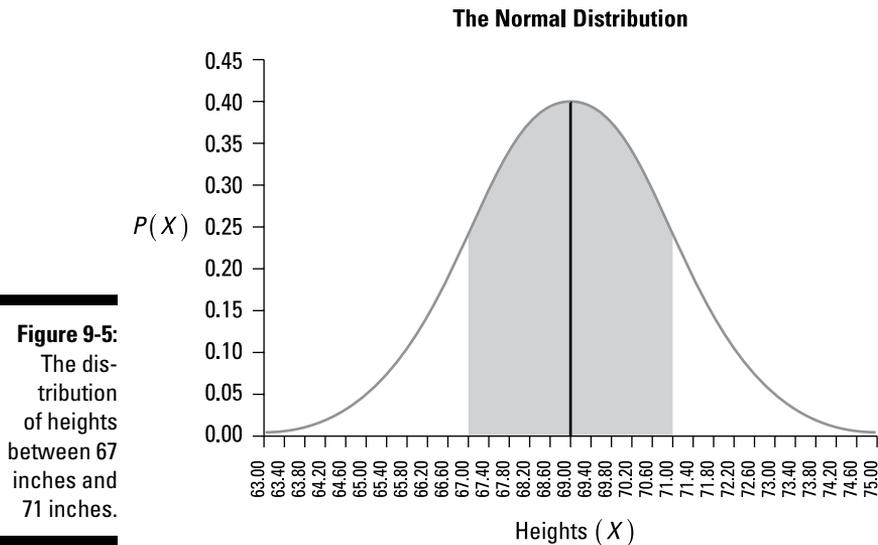


Figure 9-4:
The bell-shaped curve of the distribution of heights.

The mean of 69 inches is at the center of the distribution; the area to the left of the mean is a mirror image of the area to the right of the mean. Most of the area under the curve is close to the mean; the area falls off rapidly for large and small values of X . (The extreme right and left ends of the curve are known as the *tails* of the distribution.)

Figure 9-5 shows that the probability of a randomly chosen man's height being between 67 inches and 71 inches is 68.27 percent.

The shaded region under the curve represents heights between 67 and 71 inches. This covers 68.27 percent of the area under the curve; therefore, the probability that a randomly chosen man's height is between 67 inches and 71 inches is 0.6827 or 68.27 percent.



The normal distribution is uniquely characterized by two values:

- ✓ The expected value (mean), represented by μ (the Greek letter “mu”)
- ✓ The standard deviation, represented by σ (the Greek letter “sigma”)

There are an infinite number of different possible normal distributions, each with a different value of the mean and standard deviation.

The normal distribution in statistical analysis

The normal distribution is used in conjunction with many statistical techniques. It plays a key role in a lot of applications, such as the following:

- ✓ Computing confidence intervals
- ✓ Testing hypotheses about the mean of a population
- ✓ Testing hypotheses about the means of two populations
- ✓ Regression analysis

In many business applications, variables are assumed to be normally distributed. For example, returns to stocks are often assumed to

be normally distributed by investors, portfolio managers, financial analysts, risk managers, and so on. The assumption of normality is not only convenient, but many standard statistical techniques require it in order to generate valid results. For example, computing a confidence interval for the mean of a population may be based on the normal distribution. Many of the techniques used in regression analysis to check the validity of the results are based on the normal distribution. As a result, even when the assumption of normality is not perfectly accurate, the normal distribution is often used to perform statistical analyses due to its convenience.

Getting to know the standard normal distribution

The *standard normal distribution* is the special case where $\mu = 0$ and $\sigma = 1$.

For example, suppose that the daily returns to a stock follow the standard normal distribution. The mean return over a single trading day is 0 percent, and the standard deviation is 1 percent; as a result:

The probability that tomorrow's return will be between -1 percent and +1 percent is 0.6827 or 68.27 percent. -1 percent represents one standard deviation below the mean, while +1 percent represents one standard deviation above the mean.

The probability that tomorrow's return will be between -2 percent and +2 percent is 0.9544 or 95.44 percent. -2 percent represents two standard deviations below the mean, while +2 percent represents two standard deviations above the mean.

The probability that tomorrow's return will be between -3 percent and +3 percent is 0.9973 or 99.73 percent. -3 percent represents three standard deviations below the mean, while +3 percent represents three standard deviations above the mean.



By convention, the letter Z represents a standard normal random variable, whereas the letter X represents any other normal random variable.

Computing standard normal probabilities

One approach to computing probabilities for the standard normal distribution is to use statistical tables. (For the mathematically inclined, the tables result from applying calculus to the normal distribution.)

The standard normal table is designed to show *cumulative* probabilities; i.e., the probability that a standard normal random variable Z is less than or equal to a specified value, such as $P(Z \leq 2.50)$. Standard normal tables are divided into two parts; the first shows positive values for Z , and the second shows negative values for Z .

Computing other types of probabilities, such as $P(Z \geq 1.70)$, can be accomplished by using the properties of the standard normal distribution to rearrange these probabilities in a more convenient form.

The following sections illustrate how to compute any type of normal probabilities using the standard normal tables.

Computing “Less Than or Equal to” Standard Normal Probabilities

Table 9-1 shows a portion of the standard normal table for positive values of Z . (The actual table typically shows Z values between 0 and 3.)

Z	0.00	0.01	0.02	0.03
0.9	0.8159	0.8186	0.8212	0.8238
1.0	0.8413	0.8438	0.8461	0.8485
1.1	0.8643	0.8665	0.8686	0.8708
1.2	0.8849	0.8869	0.8888	0.8907

The table shows the probability that a standard normal random variable Z is *less than or equal to* a specific value. For example, to express the probability that Z is less than or equal to 1, you write $P(Z \leq 1.00)$. Here’s how you find this probability:

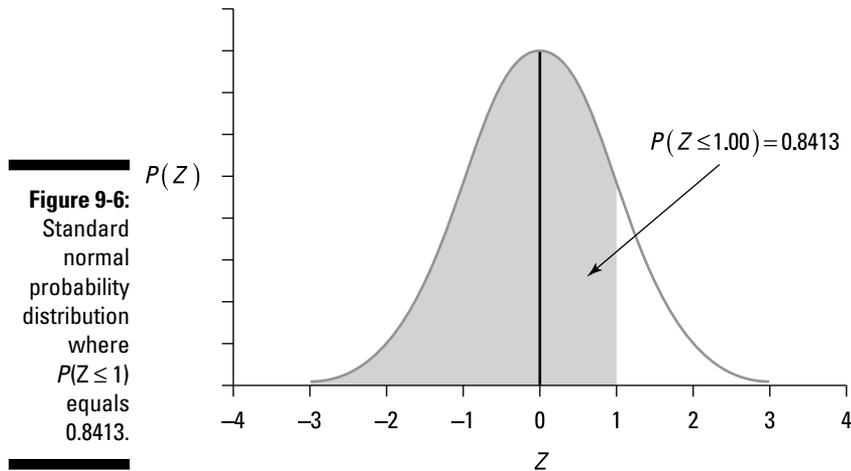
- 1. Take the first digits before and after the decimal point (1.0 in 1.00) from the Z column, second row.**
- 2. Take the second digit after the decimal point (0.00 in 1.00) from the corresponding column (0.00 in this case).**
- 3. Find the appropriate probability at the intersection of this row and column.**

Using this technique, the table shows that $P(Z \leq 1.00) = 0.8413$. Figure 9-6 shows this expression graphically.

The shaded region to the left of 1 represents 84.13 percent of the area under the curve; therefore, $P(Z \leq 1.00) = 0.8413$ or 84.13 percent.

Negative probabilities also have a corresponding standard normal table. Take a look at Table 9-2. This shows several negative values for Z ; the actual table typically shows values ranging from 0 to -3 .

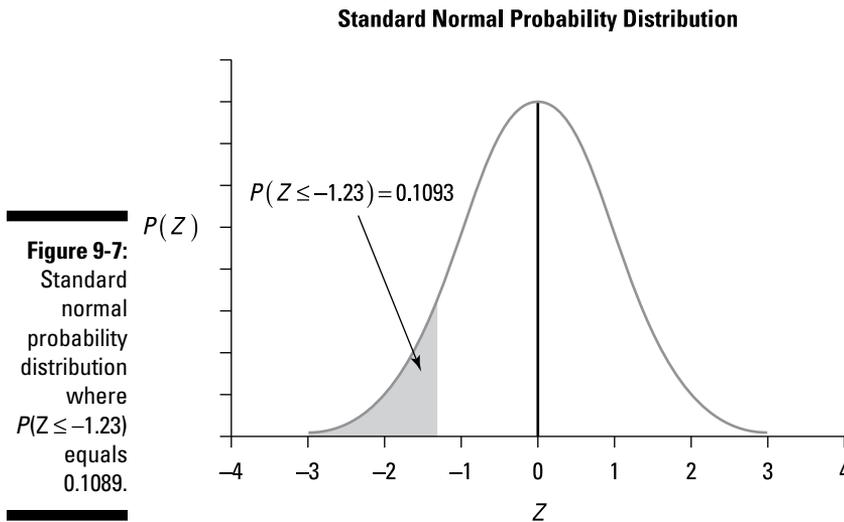
Standard Normal Probability Distribution

**Table 9-2** Standard Normal Table — Negative Values

Z	0.00	0.01	0.02	0.03
-1.3	0.0968	0.0951	0.0934	0.0918
-1.2	0.1151	0.1131	0.1112	0.1093
-1.1	0.1357	0.1335	0.1314	0.1292
-1.0	0.1587	0.1562	0.1539	0.1515

Say you want to compute the probability that Z is less than -1.23 , which you write as $P(Z \leq -1.23)$. The first digits before and after the decimal point (-1.2 in -1.23) are in the Z column, second row. The second digit after the decimal point (0.03 in -1.23) is in the far right column. You find the probability at the intersection of the row and column, so the table shows that $P(Z \leq -1.23) = 0.1093$. This is shown in Figure 9-7.

One of the drawbacks to using tables to compute standard normal probabilities is that they show only cumulative probabilities for Z ; for example, Z is less than or equal to a specific value. But you can figure all other cases by combining the properties of the standard normal distribution with the tables.



Property 1: The area under the standard normal curve equals 1

The first of these properties is that the entire area under the standard normal curve equals 1. Because the curve covers the entire area between negative and positive infinity (∞), you can express this result as $P(-\infty \leq Z \leq \infty) = 1$. So the probability that a standard normal random variable Z falls between negative infinity and positive infinity is 1; in other words, Z will fall within this interval with *certainty*.



When you consider all possible outcomes in any given situation, you can be certain that one outcome will occur. A probability of 1 indicates that an event will occur with *certainty*. A probability of 0 indicates that an event is *impossible*. All other probabilities fall between 0 and 1. (Probability theory is covered in Chapter 6.)

Property 2: The standard normal curve is symmetrical about the mean

The next key property of the standard normal distribution is *symmetry*, where the area to the left of the mean is a mirror image of the area to the right. As a result, the probability that Z is less than the mean is 0.5, and you write it as $P(Z \leq 0) = 0.5$ (because half of the area under this distribution is to the left of the mean, and half is to the right of the mean; the total area is 1), as shown in the Figure 9-8.

Because $P(Z \leq 0) = 0.5$, due to the symmetry of the standard normal probability distribution, it's also true that $P(Z \geq 0) = 0.5$, as illustrated in Figure 9-9.

Standard Normal Probability Distribution

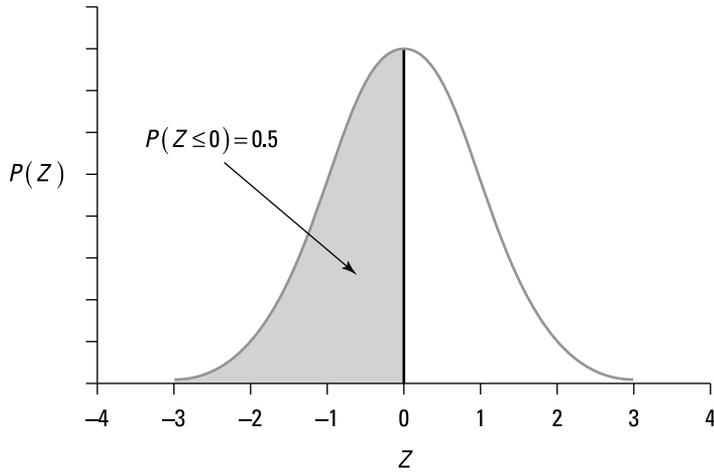


Figure 9-8:
Standard
normal
probability
distribution
where
 $P(Z \leq 0) =$
0.5.

Standard Normal Probability Distribution

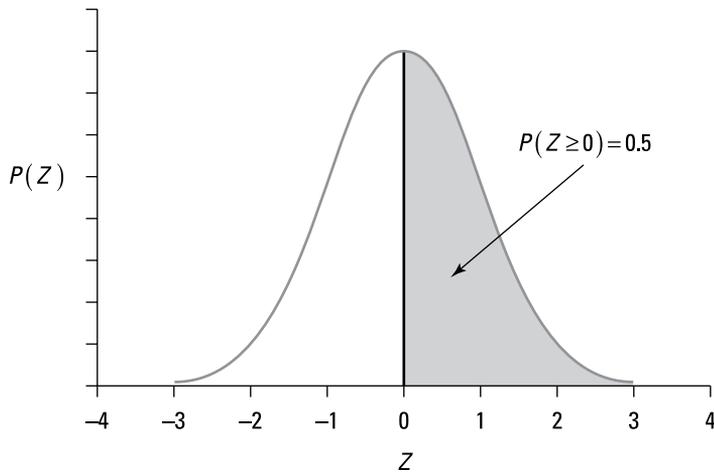


Figure 9-9:
Standard
normal
probability
distribution
where
 $P(Z \geq 0) =$
0.5.

Other examples of symmetry include

$$P(Z \leq -1) = P(Z \geq 1) = 0.1587$$

$$P(Z \leq -2) = P(Z \geq 2) = 0.0228$$

Computing “greater than or equal to” standard normal probabilities

One type of probability you can't compute directly from a table is the case where a standard normal random variable Z is *greater than or equal to* a specified value z : $P(Z \geq z)$. Instead, you rearrange the identity to yield a very useful result:

$$P(Z \leq z) + P(Z \geq z) = 1$$

This is a consequence of the first property of the standard normal distribution: The area under the standard normal curve equals 1.

Rearranging this equation gives you

$$P(Z \geq z) = 1 - P(Z \leq z)$$

For example, to determine the probability that a standard normal random variable is greater than 1 (for example, $P(Z \geq 1)$), the first step is to rewrite the probability in a form that enables you to use the standard normal tables. This is shown as:

$$P(Z \geq 1) = 1 - P(Z \leq 1)$$

$$P(Z \leq 1) = 0.8413$$

$$P(Z \geq 1) = 1 - 0.8413$$

$$= 0.1587$$

The result is shown in Figure 9-10.

Standard Normal Probability Distribution

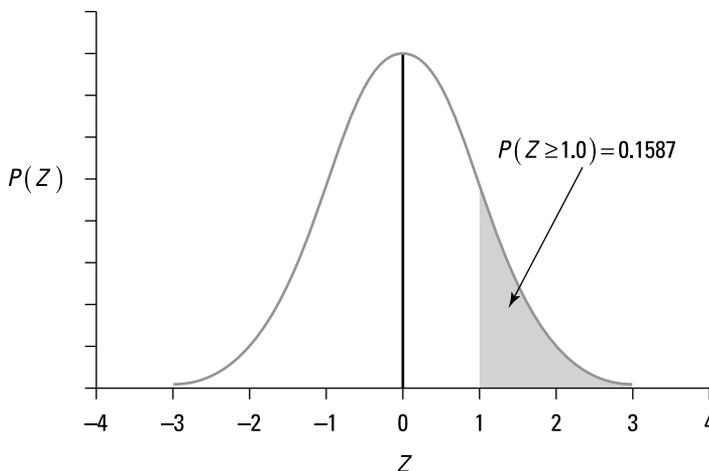


Figure 9-10:
Standard
normal
probability
distribution
where $P(Z \geq 1) = 0.1587$.

Computing “in between” standard normal probabilities

Another type of probability that you can't compute directly from a standard normal table is the case where a standard normal random variable Z is *between* two constants: c and d : $P(c \leq Z \leq d)$. But, lucky for you, you can work around this with the following identity:

$$P(c \leq Z \leq d) = P(Z \leq d) - P(Z \leq c)$$

You can now compute this probability by looking up $P(Z \leq c)$ and $P(Z \leq d)$ in the standard normal table and computing the difference between them.

For example, suppose that you want to know the probability that Z is between one and two standard deviations above the mean. In this case, $c = 1.00$ and $d = 2.00$. This probability can be expressed as follows:

$$P(1.00 \leq Z \leq 2.00)$$

Algebraically, this can be rearranged in a form that involves two “less than or equal to” probabilities that can be looked up in the standard normal tables:

$$\begin{aligned} P(1.00 \leq Z \leq 2.00) \\ = P(Z \leq 2.00) - P(Z \leq 1.00) \end{aligned}$$

From the standard normal table (Table 9-1):

$$\begin{aligned} P(Z \leq 2.00) &= 0.9772 \\ P(Z \leq 1.00) &= 0.8413 \end{aligned}$$

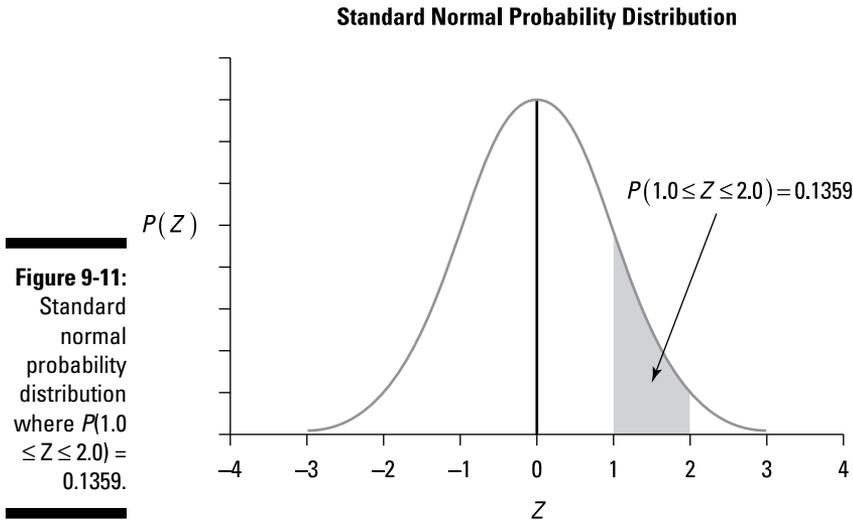
As a result, you calculate the probability:

$$P(1.00 \leq Z \leq 2.00) = 0.9772 - 0.8413 = 0.1359$$

Figure 9-11 illustrates this probability.

Note that you can use this approach for negative values, too. For example, from the standard normal table (Table 9-2),

$$\begin{aligned} P(-2.00 \leq Z \leq -1.00) \\ = P(Z \leq -1.00) - P(Z \leq -2.00), \\ P(Z \leq -2.00) &= 0.0228 \\ P(Z \leq -1.00) &= 0.1587 \end{aligned}$$



As a result:

$$\begin{aligned}
 &P(-2.00 \leq Z \leq -1.00) \\
 &= P(Z \leq -1.00) - P(Z \leq -2.00) \\
 &= 0.1587 - 0.0228 \\
 &= 0.1359
 \end{aligned}$$

Computing normal probabilities other than standard normal

Many variables in business applications are assumed to be normally distributed, including rates of returns to stocks and other financial assets. Although these variables are normal, they're usually not *standard* normal. As a result, you can't compute probabilities for these variables from the standard normal tables without first transforming them into the equivalent standard normal form, as shown with the following formula:

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

In this expression, Z is a standard normal random variable, and X is a normal random variable with mean μ and standard deviation σ .

For example, suppose that the annual return of the stock of the Gamma Corporation is normally distributed with a mean of 5 percent and a standard deviation of 2 percent. What's the probability that the return from this stock over the coming year will be 4 percent or less?

Let X be a random variable that represents "the annual return for the stock of Gamma Corporation." X is a normally distributed random variable with a mean (μ) of 0.05 and a standard deviation (σ) of 0.02. X is *not* standard normal, because the mean isn't 0 and the standard deviation isn't 1.

To compute this probability, convert the rate of return X into a standard normal random variable Z as follows:

$$P(X \leq 0.04) = P\left(Z \leq \frac{X - \mu}{\sigma}\right) = P\left(Z \leq \frac{0.04 - 0.05}{0.02}\right) = P(Z \leq -0.5)$$

Based on the standard normal tables (refer to Tables 9-1 and 9-2 in the earlier section "Computing standard normal probabilities"), $P(Z \leq -0.5) = 0.3085$, so the probability that the stock's return will be 4 percent or less is 0.3085 or 30.85 percent.

Similarly, you can determine the probability that the stock's return next year will be 8 percent or more like so:

$$P(X \geq 0.08) = P\left(Z \geq \frac{X - \mu}{\sigma}\right) = P\left(Z \geq \frac{0.08 - 0.05}{0.02}\right) = P(Z \geq 1.5)$$

Recall from the earlier section "Computing "greater than or equal to" standard normal probabilities" the following key property for the standard normal distribution:

$$P(Z \leq z) + P(Z \geq z) = 1$$

Rearranging this algebraically gives:

$$P(Z \geq z) = 1 - P(Z \leq z)$$

Therefore,

$$P(Z \geq 1.5) = 1 - P(Z \leq 1.5)$$

Based on the standard normal table (Table 9-1):

$$P(Z \leq 1.5) = 0.9332$$

Therefore,

$$\begin{aligned} & 1 - P(Z \leq 1.5) \\ &= 1 - 0.9332 \\ &= 0.0668 \\ &= 6.68\% \end{aligned}$$

the probability that the stock's return next year will be between 7 percent and 10 percent as follows:

$$\begin{aligned} P(0.07 \leq X \leq 0.10) &= P\left(\frac{0.07 - 0.05}{0.02} \leq Z \leq \frac{0.10 - 0.05}{0.02}\right) \\ &= P(1.00 \leq Z \leq 2.50) \\ &= P(Z \leq 2.50) - P(Z \leq 1.00) \\ &= 0.9938 - 0.8413 \\ &= 0.1525 \\ &= 15.25\% \end{aligned}$$

As another example, imagine that the scores on a standardized test are normally distributed with a mean score of 80 and a standard deviation of 10. If a student receives a score of 90, he was outperformed by what proportion of all other students taking the test?

In other words, what is the probability of receiving a score of more than 90 on this test? Let X represent the random variable "score on the exam." X is a normally distributed random variable with a mean of 80 and a standard deviation of 10. Because X isn't a standard normal random variable, you must convert it:

$$P(X \geq 90) = P\left(Z \geq \frac{X - \mu}{\sigma}\right) = P\left(Z \geq \frac{90 - 80}{10}\right) = P(Z \geq 1.00)$$

Due to the symmetry of the standard normal distribution,

$$P(Z \geq 1.00) = 1 - P(Z \leq 1.00)$$

From the standard normal table (Table 9-1),

$$P(Z \leq 1.00) = 0.8413$$

Therefore,

$$\begin{aligned} P(Z \geq 1.00) \\ &= 1 - P(Z \leq 1.00) \\ &= 1 - 0.8413 \\ &= 0.1587 \end{aligned}$$

or only 15.87 percent of the students taking the exam scored better than 90.

These techniques can be used to compute *any* normal probability, whether it is expressed as greater than, less than, or between, and regardless of the mean and standard deviation of the distribution.

Chapter 10

Sampling Techniques and Distributions

In This Chapter

- ▶ Getting familiar with sampling techniques
 - ▶ Using sampling distributions to estimate probabilities
-

A *population* is a collection of data that we are interested in studying; a *sample* is a selection of data randomly chosen from a population. The use of sample data is the basis for a wide variety of business applications. This is because obtaining information about an entire population is likely to be very time-consuming and costly. Instead, samples may be used to understand the behavior of the underlying population.

For example, if a department store wants to know which types of new products customers are willing to buy, the store may not have the resources to survey every single one of its customers. Instead, if the store can choose representative samples of its customers to survey, it could potentially obtain the same information at a fraction of the cost.

One of the requirements of using samples to draw conclusions about a population is that the samples accurately mirror the population; otherwise, any conclusions that are reached about the population are bound to be inaccurate.

Several different types of sampling techniques have been developed to accurately capture the properties of a population. The choice of technique depends on several factors, such as:

What are the demographic characteristics of interest?

How easy will it be to obtain sample data?

How much data is needed to ensure accurate results?

For example, suppose that the New York State government wants to analyze the distribution of ages of everyone living in the state. This helps determine what type of funding is needed for various programs in the future. Although the ages of every single resident could be collected, this could be very time consuming and costly.

Instead, suppose the government decides to randomly sample residents throughout the state and use this information to estimate the distribution of ages. Clearly, it makes no sense to focus only on high school students, because their ages are substantially lower than the overall population. Instead, samples are chosen that ideally match the demographic characteristics of the entire state. For example, questionnaires could be mailed to randomly chosen addresses throughout the state.

In this chapter, I introduce several types of sampling techniques that may be used for various types of studies. I also show you a special type of probability distribution, known as a *sampling distribution*. This is a special type of probability distribution that describes the properties of a *sample statistic*. (Sample statistics are summary measures of a sample; these include the sample mean, sample variance and sample standard deviation. Sample statistics are discussed in Chapters 3 and 4.) Due to its widespread use in statistical analysis, I focus on the sampling distribution of the *sample mean*.

Sampling Techniques: Choosing Data from a Population

Statistical inference is a methodology that lets you draw conclusions about a population from sample data. One of the most important challenges in statistical inference is choosing samples that accurately reflect the characteristics of the underlying population. Although you can choose from many sampling techniques, the appropriate technique depends on the type of information you're studying and your resources.

You can classify the two basic approaches to sampling as probability sampling and nonprobability sampling. Probability sampling is used when it is important to ensure that each member of a population has a chance of being chosen. Nonprobability sampling is a more subjective approach, and is often used when it would be difficult or impossible to use probability sampling. I explore both of these approaches in the following sections.

Probability sampling

When you use *probability sampling*, each member of the population has a chance of being chosen for the sample. In some of these techniques, each population member is *equally likely* to be chosen; in others, this is not the case. With probability sampling, it's possible to determine the probability that a given member of the population will be chosen.

Within the category of probability sampling, you can choose from four types of sampling techniques, which I discuss in the following sections.

Simple random samples

In a *simple random sample*, each member in the population is equally likely to be chosen. There are several different ways in which population members may be chosen with equal probability. One approach is to assign a numerical value to each population member and then randomly choose numbers that correspond to these members.

For example, suppose a population consists of the following ten members of the finance faculty at a prestigious university:

1. Benjamin Harrison
2. Martin Van Buren
3. John Tyler
4. Millard Fillmore
5. Grover Cleveland
6. Chester Arthur
7. James Polk
8. Zachary Taylor
9. James Buchanan
10. Franklin Pierce

You would like to randomly choose five of these faculty members for a newly formed committee. You assign each faculty member a number from one to ten. (This could be done alphabetically or in any number of other ways.) To choose a simple random sample of five of these faculty members, you can use a random number generator.

A random number generator is a function that can be used to randomly choose numbers within a specified interval. As an example, you can use

Excel's `RANDBETWEEN` function; this generates whole numbers that are randomly chosen between any two values you specify.

For this example, you would need to generate a random number between 1 and 10. You would then enter `RANDBETWEEN(1,10)` into Excel and record the resulting number. You would repeat this process until you have five unique numbers. The faculty members associated with these numbers are then chosen for the new committee.



In this example, you don't want to choose the same number twice; if this happens, you simply discard the result and choose another random number until you have five unique numbers. The process you are using is known as sampling *without* replacement. If you are willing to choose the same number more than once, then no results would be discarded; the process that you would be using is known as sampling *with* replacement.

Suppose the following sequence of random numbers is chosen:

`RANDBETWEEN(1,10)` = 1

`RANDBETWEEN(1,10)` = 4

`RANDBETWEEN(1,10)` = 5

`RANDBETWEEN(1,10)` = 8

`RANDBETWEEN(1,10)` = 6

Your simple random sample would then consist of the following faculty members:

1. Benjamin Harrison
4. Millard Filmore
5. Grover Cleveland
8. Zachary Taylor
6. Chester Arthur

These are the lucky members of the new committee.

Systematic samples

With *systematic samples*, population members are assigned a numerical value, as is the case with simple random samples. Instead of using random numbers to choose population members, though, you will instead use a specific *sequence* of numbers.

For example, suppose an economist wants to study the distribution of 100 household incomes in a small town and wants to draw a sample size of ten. In this case, the economist draws every tenth population member (because the number of households divided by the sample size equals $100/10 = 10$). One way she can draw every tenth member is to start with a random number (between 1 and 10) and then add ten to each number to get the desired sequence.

For example, you could use `RANDBETWEEN(1,10)` to obtain the starting value for the sequence. If this turns out to be a 3, then the appropriate sequence of random numbers would be:

3, 13, 23, 33, 43, 53, 63, 73, 83, 93

If instead the function `RANDBETWEEN(1,10)` generates a 5, then the appropriate sequence of random numbers would be:

5, 15, 25, 35, 45, 55, 65, 75, 85, 95

Other techniques could be used to randomly choose the first value, such as the flip of a coin, the roll of a die, and so on. Similarly, if a population contains 1,200 members and the economist wants a sample size of ten, the numbering sequence includes every 120th member ($1,200/10 = 120$). One way she can draw every 120th member is to start with a random number (between 1 and 120) and then add 120 to each number to get the sequence.

In this case, suppose that the function `RANDBETWEEN(1,120)` results in a value of 57; then the sequence would consist of the following values:

57, 177, 297, 417, 537, 657, 777, 897, 1017, 1137

As another example, suppose that a marketing firm wants to find out whether consumers are responding favorably to a newly launched advertising campaign. A researcher could choose a busy mall and ask every 20th customer that walks by how he or she feels about the new advertising campaign. In this case, though, the researcher wouldn't have a specific sequence of numbers, because it's impossible to determine in advance how many people are in the mall at any given time.

In this case, systematic samples are chosen based on incomplete knowledge of the underlying population. This approach is useful when the size of the entire population is not known.

Stratified samples

When using *stratified samples*, you divide a population into *strata* (levels or layers). The strata may reflect any of a wide variety of characteristics of the population data, such as ages, incomes, levels of education, and so on.

Basically, you choose a stratified sample in such a way that you ensure that the proportion of sample members in each stratum (singular of *strata*) matches the distribution found in the population.

For example, suppose a college wants to conduct a survey of student attitudes toward the building of a costly new sports stadium as an alternative to expanding the current antiquated library. Instead of surveying every single student in the school, the college chooses stratified samples. It divides the entire student body by class: freshmen, sophomores, juniors, and seniors. (Assume for this example that the school doesn't offer any graduate programs, so all students belong to one of these four classes.) Here's how the classes break down:

<i>Class</i>	<i>Number of Students</i>
Freshmen	800
Sophomores	1,200
Juniors	1,000
Seniors	1,000

And the percentages of students in each class are as follows:

<i>Class</i>	<i>Number of Students</i>	<i>Percentage of Total</i>
Freshmen	800	20 percent
Sophomores	1,200	30 percent
Juniors	1,000	25 percent
Seniors	1,000	25 percent

If the college chooses a stratified sample of 200 students, the sample consists of the following:

- 40 freshmen (20 percent of 200)
- 60 sophomores (30 percent of 200)
- 50 juniors (25 percent of 200)
- 50 seniors (25 percent of 200)

Within each stratum, a simple random sample of the appropriate number of students is chosen. This selection method ensures that no class is under- or overrepresented in the sample data.

One of the advantages of the stratified sample approach is that you can draw conclusions about each individual stratum. For example, the college can analyze the attitudes of freshmen separately from the attitudes of sophomores, juniors, and seniors. On the other hand, one of the disadvantages of this approach is that you need more information about the characteristics of the population than with other approaches, such as the simple random sampling approach discussed earlier. In this example, you need to know the distribution of students among the freshman, sophomore, junior, and senior classes.

Cluster samples

With *cluster samples*, you subdivide a population into groups based on common characteristic (such as location, age, income level, and so forth). You choose groups randomly, and then you choose samples from those groups randomly.

Say you're a researcher conducting a national survey about attitudes toward proposed national legislation. You divide the entire voting age population of the United States into groups according to state of residency. You decide to choose a sample of eight states; you believe that this is sufficient to represent the entire country.

In this case, you would first assign a number to each state in the United States. Next, you could use the function `RANDBETWEEN(1,50)` until you choose eight different states.

Within each selected state, voting age residents are randomly chosen using a simple random sample. This may be accomplished by assigning a number to each registered voter and then using a random number generator to randomly pick the desired number of voters.

Suppose that the following states are chosen:

Wisconsin

Rhode Island

Michigan

Utah

Illinois

South Carolina

Arizona

Oregon

Within each state, you choose simple random samples of voters.

The advantage of using cluster sampling is that it can be implemented more quickly and cheaply than stratified sampling. In this example, stratified sampling requires voters to be randomly chosen from each of the 50 states. The disadvantage of using cluster sampling is that it may not be as accurate as stratified sampling.

Nonprobability sampling

Unlike probability sampling, *nonprobability sampling* doesn't guarantee that each population member has a chance of being chosen. And with nonprobability sampling, you have to use subjective judgment.

One of the major drawbacks to nonprobability samples is that the results aren't as reliable for drawing conclusions about the overall population. It may be easier to get the samples, but there's a price — they're less useful than probability samples.

I discuss four of the nonprobability sampling techniques in the following sections, including convenience samples, quota samples, purposive samples, and judgment samples.

Convenience samples

When you choose population members primarily because they're accessible, you're using *convenience samples*. For example, if a marketing firm needs to study consumer attitudes toward new products, it may be forced to rely on the input of people who are willing to participate; they are not necessarily representative of the overall population.

Suppose for example a marketing firm decides to conduct a series of interviews at a mall to determine which new movies are likely to do well at the box office. The interviews are conducted at 3:00 in the afternoon on a Wednesday. Although there may be many volunteers who are willing to take part in the interviews, most or all of them are likely to be students and/or retirees, which doesn't reflect the overall population. Unless the marketing firm is only interested in the views of these groups, the results are not likely to be accurate.

Quota samples

Quota samples are closely related to stratified samples; in both cases, you divide population members into separate groups. The main difference is that with a quota sample, the number of sample members in each stratum may not exactly represent the numbers in the underlying population.

For example, suppose that a college is interested in comparing the GPA of its male and female students. Assume that the proportion of male students at this college is 60 percent, so the proportion of female students is 40 percent. A stratified sample would ensure that 60 percent of the sample members are male, and 40 percent are female. With a quota sample, any number of males and females may be chosen. Suppose that the college doesn't know the exact proportion of male and female students, so it decides to choose an equal mix of male and female students for the sample. Clearly, this doesn't reflect the proportions in the actual population.

Purposive samples

With *purposive samples*, you choose members of the population because they're *not* typical in some important way. For example, a company that produces a new product may be concerned that the product is too expensive for the average consumer to buy. The company may target students (who presumably have low incomes) to determine whether they'd consider buying the product. The logic is that if the product isn't too expensive to people with relatively low incomes, it won't be too expensive to people with higher incomes.

As another example, suppose that a snack foods company manually inspects all the potato chips that it produces before they are sold to the public. Any chips that appear to be burned are automatically discarded. This process is very time consuming and costly; the company wants to try a different approach.

Suppose that the smallest chips are most likely to be burned. Rather than inspecting every single potato chip, the company decides to save time by only inspecting the chips that appear to be unusually small. If these are not burned, the remaining chips are probably acceptable. The company is now using purposive samples to represent the entire population.

Judgment samples

When conducting a study with a *judgment sample*, you chose members based on your subjective judgment. You choose these members because they offer specific characteristics of interest. For example, suppose that half of the residents of a city are male (and, therefore, half are female). A handbag manufacturer wants to determine which features are most important to consumers in this city. If the company chooses to survey customers in the local mall, it may go out of its way to question a larger number of female customers (rather than male customers) because most handbags are purchased by women.

Sampling Distributions

A *statistic* is a summary measure of a sample, and a *parameter* is a summary measure of a population. (I discuss both summary measures of samples and populations in Chapters 3, 4, and 5.) The probability distribution of a statistic is known as a *sampling distribution*, which is what this section is all about.

Some examples of statistics include

- ✓ Sample mean (\bar{X})
- ✓ Sample variance (s^2)
- ✓ Sample standard deviation (s)

Some examples of parameters are

- ✓ Population mean (μ)
- ✓ Population variance (σ^2)
- ✓ Population standard deviation (σ)

Note: Latin letters represent statistics; Greek letters represent parameters.

In many cases, a *population parameter* is costly and time-consuming to calculate. For example, figuring out the average age of everyone living in the United States would be very time-consuming! In these cases, the statistician uses sample statistics instead. The sample mean (\bar{X}) estimates the population mean (μ). The researcher can use a representative sample of U.S. residents to compute a sample mean, which would serve as an estimated value of the average age of all U.S. residents.

If you repeatedly draw samples from a population, the value of a statistic is most likely different for each sample. As a result, it's useful to think of a statistic as a *random variable* whose properties can be described with a *probability distribution*. (See Chapter 7 for details.)

In the following sections, I explore the characteristics of sampling distributions, including how to represent data from a sampling distribution graphically and how to compute the moments of a sampling distribution. The focus is on the sampling distribution of the sample mean \bar{X} .

Portraying sampling distributions graphically

As I explain in Chapter 2, a *histogram* is a graphical representation of data in which ranges of values, known as *classes*, appear on the horizontal axis (the *x*-axis) and probabilities on the vertical axis (the *y*-axis). Each class is shown as a single bar whose height equals the probability of that class.

A histogram shows at a glance how the values of a variable are distributed. In this section, histograms are used to describe the properties of the sampling distribution of \bar{X} .

One of the benefits of using histograms to analyze a sampling distribution is that it is easy to see if the sampling distribution is symmetrical about the mean, negatively skewed, or positively skewed.

A distribution is symmetrical about the mean if values below the mean occur as frequently as the values an equal distance above the mean. A negatively skewed distribution is one in which there are a small number of extremely small values; a positively skewed distribution is one in which there are a small number of extremely large values. (Skewness and symmetry are discussed in Chapter 3.)



A distribution is symmetrical about the mean if the mean equals the median. A distribution is negatively skewed if the mean is less than the median and positively skewed if the mean is greater than the median.

A histogram also shows at a glance the center or mean of a distribution, and how “spread out” are the members of the distribution. (Recall from Chapter 4 that the spread of a distribution is measured by its variance and its standard deviation.)

A histogram can be used to compare the properties of different sampling distributions or to observe the effect of different sample sizes on a sampling distribution. For example, suppose that a manufacturer of computer chips has found from experience that its assembly line produces two defective chips per hour, and that the number of defective chips produced during a given hour is independent of the number produced during any other hour. In other words, the distribution of defective chips follows the Poisson distribution with an average value of two per hour — in other words, $\lambda = 2$. (The Poisson distribution is discussed in detail in Chapter 8.)

The distribution of defective chips is shown in Figure 10-1.

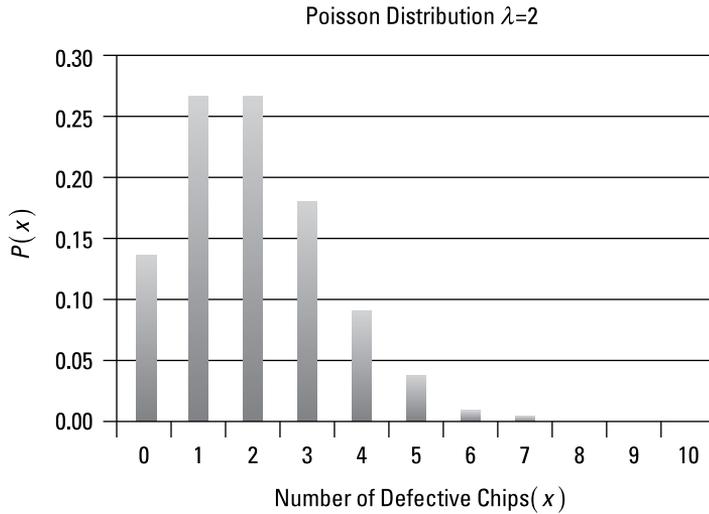


Figure 10-1: Histogram for the distribution of defective chips.

Suppose that a sample of five computer chips is randomly chosen, and the number of defective chips in each sample is recorded. This process is repeated 300 times. The resulting distribution consists of 300 sample means, ranging from a low of 0.6 to a high of 4.2.

Figure 10-2 shows the distribution of the mean number of defective chips among the 300 samples of size 5.

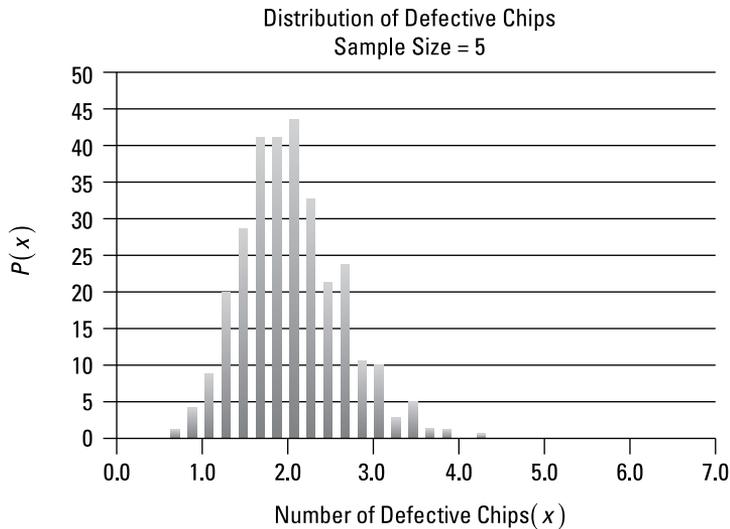
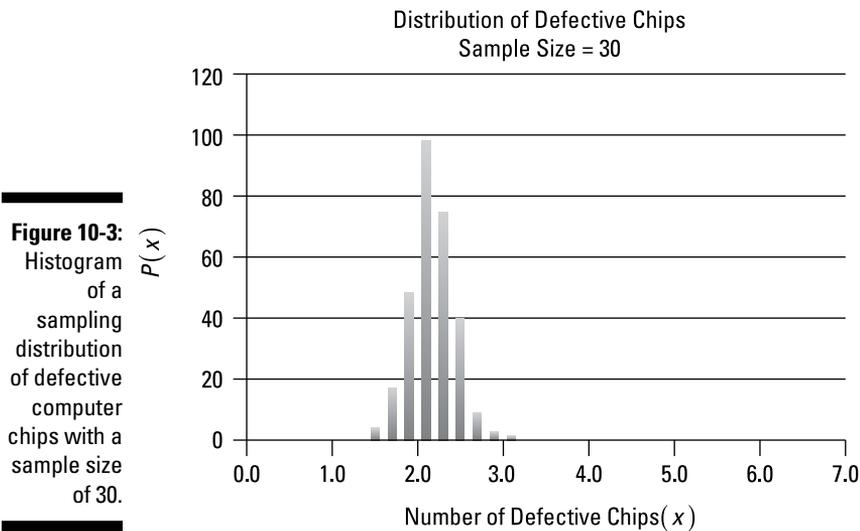


Figure 10-2: Histogram of a sampling distribution of defective computer chips with a sample size of 5.

Note that the distribution of sample means with a sample size of 5 strongly resembles the Poisson distribution.

Suppose now that a sample of 30 computer chips is randomly chosen, and the number of defective chips in each sample is recorded. This process is repeated 300 times. The resulting distribution consists of 300 sample means, ranging from a low of 1.3 to a high of 3.

Figure 10-3 shows the distribution of the mean number of defective chips among the 300 samples of size 30.



Note that the distribution of sample means with a sample size of 30 much more closely resembles the normal distribution than the Poisson distribution.

Figures 10-2 and 10-3 show that the sample mean remains centered on 2 regardless of the sample size, but the mean number of defectives is far less *dispersed* around the mean with a sample size of 30 compared with a sample size of 5. (You can tell that this is the case because the sample mean ranges from 0.6 to 4.2 with a sample size of 5, compared with 1.3 to 3 for a sample size of 30.)

In addition, the figures show that as the sample size grows from 5 to 30, the sampling distribution looks more like the normal distribution.

Moments of a sampling distribution

A sampling distribution is described by a series of summary measures known as *moments*, which include expected value (mean) and variance. The standard deviation is not a separate moment; it is the square root of the variance. The standard deviation of a sampling distribution is often referred to as the *standard error*.

For the sampling distribution of \bar{X} , the expected value is $\mu_{\bar{X}}$, which equals the mean of the underlying population (μ). The variance is $\sigma_{\bar{X}}^2$, and the standard deviation, also known as the *standard error*, is $\sigma_{\bar{X}}$.

The values of the variance and standard error depend on the relationship between the size of the sample (n) drawn from the population and the size of the population (N).

- ✓ If the sample size is less than or equal to 5 percent of the population size, the sample is small, relative to the size of the population. In this case, the variance of \bar{X} equals

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$$

Here, σ^2 is the variance and σ is the standard deviation of the underlying population; n is the sample size.

The square root of the variance of \bar{X} is the standard error of \bar{X} :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

- ✓ If the sample size is greater than 5 percent of the population size, the sample is large, relative to the size of the population. In this case, the standard error of \bar{X} equals

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

The term $\sqrt{\frac{N-n}{N-1}}$ is known as the *finite population correction factor*, which always assumes a value of less than or equal to 1 (it equals 1 only if the sample size is 1). You use the finite population correction factor to reduce the size of the standard error to reflect the fact that less variability from one sample mean to the next occurs when the sample size is large relative to the population.

The Central Limit Theorem

According to the Central Limit Theorem, the sampling distribution of \bar{X} is normal if the underlying population is normal. If not, the sampling distribution of \bar{X} is at least approximately normal if the sample size is at least 30. Under these circumstances, you can use the normal distribution to determine the probability that the sample mean will fall within a specified range of values. (See Chapter 9 for techniques on using the normal distribution.)

For example, suppose you choose a sample of 50 gasoline prices from gas stations in a major city. You can use the normal distribution to determine the probability that the sample mean gas price is between \$3.50 and \$4.00 per gallon.

If the Central Limit Theorem fails to hold, you can't use the normal distribution to compute probabilities for the sample mean; instead, you need to find an alternative probability distribution that closely resembles the population that you are studying.

Converting \bar{X} to a standard normal random variable

Based on the Central Limit Theorem, if you draw samples from a population of $n \geq 30$, then \bar{X} is a normally distributed random variable. To determine probabilities for \bar{X} , you may use the standard normal probability tables. (These are discussed in Chapter 9.) Use the standard normal tables, which require you to convert \bar{X} to a standard normal random variable.



The standard normal distribution is the special case where the mean (μ) equals 0, and the standard deviation (σ) equals 1.

For any normally distributed random variable X with a mean μ and a standard deviation σ , you find the corresponding standard normal random variable (Z) with the following equation:

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

For the sampling distribution of \bar{X} , the corresponding equation is

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

As an example, say that there are 10,000 stocks trading each day on a regional stock exchange. It's known from historical experience that the returns to these stocks have a mean value of 10 percent per year, and a standard deviation of 20 percent per year.

An investor chooses to buy a random selection of 100 of these stocks for his portfolio. What's the probability that the mean rate of return among these 100 stocks is greater than 8 percent?

The investor's portfolio can be thought of as a sample of stocks chosen from the population of stocks trading on the regional exchange. The first step to finding this probability is to compute the moments of the sampling distribution.

✔ **Compute the mean:** $\mu_{\bar{x}} = \mu = 0.10$.

The mean of the sampling distribution equals the population mean.

✔ **Determine the standard error:** This calculation is a little trickier because the standard error depends on the size of the sample relative to the size of the population. In this case, the sample size (n) is 100, while the population size (N) is 10,000. So you first have to compute the sample size relative to the population size, like so:

$$n/N = 100/10,000 = 0.01 = 1\%$$

Because 1 percent is less than 5 percent, you don't use the finite population correction factor to compute the standard error. Note that in this case, the value of the finite population correction factor is:

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{10,000-100}{10,000-1}} = \sqrt{\frac{9,900}{9,999}} = 0.995$$

Because this value is so close to 1, using the finite population correction factor in this case would have little or no impact on the resulting probabilities.

And because the finite population correction factor isn't needed in this case, the standard error is computed as follows:

$$\begin{aligned}\sigma_{\bar{x}} &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{0.20}{\sqrt{100}} \\ &= \frac{0.20}{10} \\ &= 0.02\end{aligned}$$

To determine the probability that the sample mean is greater than 8 percent, you must now convert the sample mean into a standard normal random variable using the following equation:

$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$$

To compute the probability that the sample mean is greater than 8 percent, you apply the previous formula as follows:

$$P(\bar{X} \geq 0.08) = P\left(Z \geq \frac{0.08 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right)$$

Because $\mu_{\bar{X}} = 0.10$ and $\sigma_{\bar{X}} = 0.02$, these values are substituted into the previous expression as follows:

$$P\left(Z \geq \frac{0.08 - \mu_{\bar{X}}}{\sigma_{\bar{X}}}\right) = P\left(Z \geq \frac{0.08 - 0.10}{0.02}\right) = P(Z \geq -1.00)$$

You can calculate this probability by using the properties of the standard normal distribution along with a standard normal table such as Table 10-1.

Z	0.00	0.01	0.02	0.03
-1.3	0.0968	0.0951	0.0934	0.0918
-1.2	0.1151	0.1131	0.1112	0.1093
-1.1	0.1357	0.1335	0.1314	0.1292
-1.0	0.1587	0.1562	0.1539	0.1515

Table 10-1 shows the probability that a standard normal random variable (designated Z) is *less than or equal to* a specific value. For example, you can write the probability that $Z \leq -1.00$ (one standard deviation below the mean) as $P(Z \leq -1.00)$. You find the probability from the table with these steps:

1. Locate the first digit before and after the decimal point (-1.0) in the first (Z) column.
2. Find the second digit after the decimal point (0.00) in the second (0.00) column.
3. See where the row and column intersect to find the probability:
 $P(Z \leq -1.00) = 0.1587$.

Because you're actually looking for the probability that Z is *greater than or equal to* -1 , one more step is required.

Due to the *symmetry* of the standard normal distribution, the probability that Z is greater than or equal to a negative value equals one minus the probability that Z is less than or equal to the same negative value.

For example,

$$P(Z \geq -2.0) = 1 - P(Z \leq -2.0)$$

This is because $Z \geq -2.00$ and $Z \leq -2.00$ are *complementary* events. (Complementary events are discussed in Chapter 6.) This means that Z must either be greater than or equal to -2 or less than or equal to -2 . Therefore,

$$P(Z \geq -2.0) + P(Z \leq -2.0) = 1$$

This is true because the occurrence of one of these events is *certain*, and the probability of a certain event is 1. (Probability and certain events are covered in Chapter 6.)

After algebraically rewriting this equation, you end up with the following result:

$$P(Z \geq -2.0) = 1 - P(Z \leq -2.0)$$

For the portfolio example,

$$P(Z \geq -1.0) = 1 - P(Z \leq -1.0)$$

$$P(Z \geq -1.0) = 1 - 0.1587 = 0.8413$$

The result shows that there's an 84.13 percent chance that the investor's portfolio will have a mean return greater than 8 percent. As another example, suppose that it is known that there are 120 surviving paintings by a well-known 19th century artist. These works have an average price of \$1 million and a standard deviation of \$120,000. Say that an art collector acquires a random selection of ten of these paintings. What's the probability that the mean price of these paintings is between \$975,000 and \$1,025,000?

In this case, the size of the population is $N = 120$. The sample size is $n = 10$. Therefore, the sample size represents $n/N = 10/120 = 0.08333$, which is 8.333 percent of the population. Because the sample size is *greater than* 5 percent, you use the finite population correction factor to compute the standard error, like so:

$$\sqrt{\frac{N-n}{N-1}} = \sqrt{\frac{120-10}{120-1}} = 0.96144$$

You then find the mean ($\mu_{\bar{X}} = \mu = 1,000,000$) and standard error of \bar{X} :

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \frac{120,000}{\sqrt{10}} (0.96144) = 36,484$$

To calculate probabilities for \bar{X} , the first step is to convert the values of \bar{X} into standard normal random variables:

$$\begin{aligned} &P(975,000 \leq \bar{X} \leq 1,025,000) \\ &= P\left(\frac{975,000 - 1,000,000}{36,484} \leq Z \leq \frac{1,025,000 - 1,000,000}{36,484}\right) \\ &= P(-0.69 \leq Z \leq 0.69) = P(Z \leq 0.69) - P(Z \leq -0.69) \end{aligned}$$

The next step is to find the values of $P(Z \leq 0.69)$ and $P(Z \leq -0.69)$, and subtract one from the other. The art collector can get these values from standard normal tables, such as Table 10-2 and Table 10-3.

Table 10-2 Standard Normal Table — Positive Values				
<i>Z</i>	<i>0.06</i>	<i>0.07</i>	<i>0.08</i>	<i>0.09</i>
0.5	0.7123	0.7157	0.7190	0.7224
0.6	0.7454	0.7486	0.7517	0.7549
0.7	0.7764	0.7794	0.7823	0.7852
0.8	0.8051	0.8078	0.8106	0.8133

Table 10-3 Standard Normal Table — Negative Values				
<i>Z</i>	<i>0.06</i>	<i>0.07</i>	<i>0.08</i>	<i>0.09</i>
-0.8	0.1949	0.1922	0.1894	0.1867
-0.7	0.2236	0.2206	0.2177	0.2148
-0.6	0.2546	0.2514	0.2483	0.2451
-0.5	0.2877	0.2843	0.2810	0.2776

Table 10-2 shows that $P(Z \leq 0.69) = 0.7549$. This value is at the intersection of the 0.6 row for Z and the 0.09 column. Table 10-3 shows that $P(Z \leq -0.69) = 0.2451$. This value is at the intersection of the -0.6 row for Z and the 0.09 column.

Therefore, $P(Z \leq 0.69) - P(Z \leq -0.69) = 0.7549 - 0.2451 = 0.5098$.

The result is that there's a 50.98 percent chance that the sample mean falls somewhere between \$975,000 and \$1,025,000.