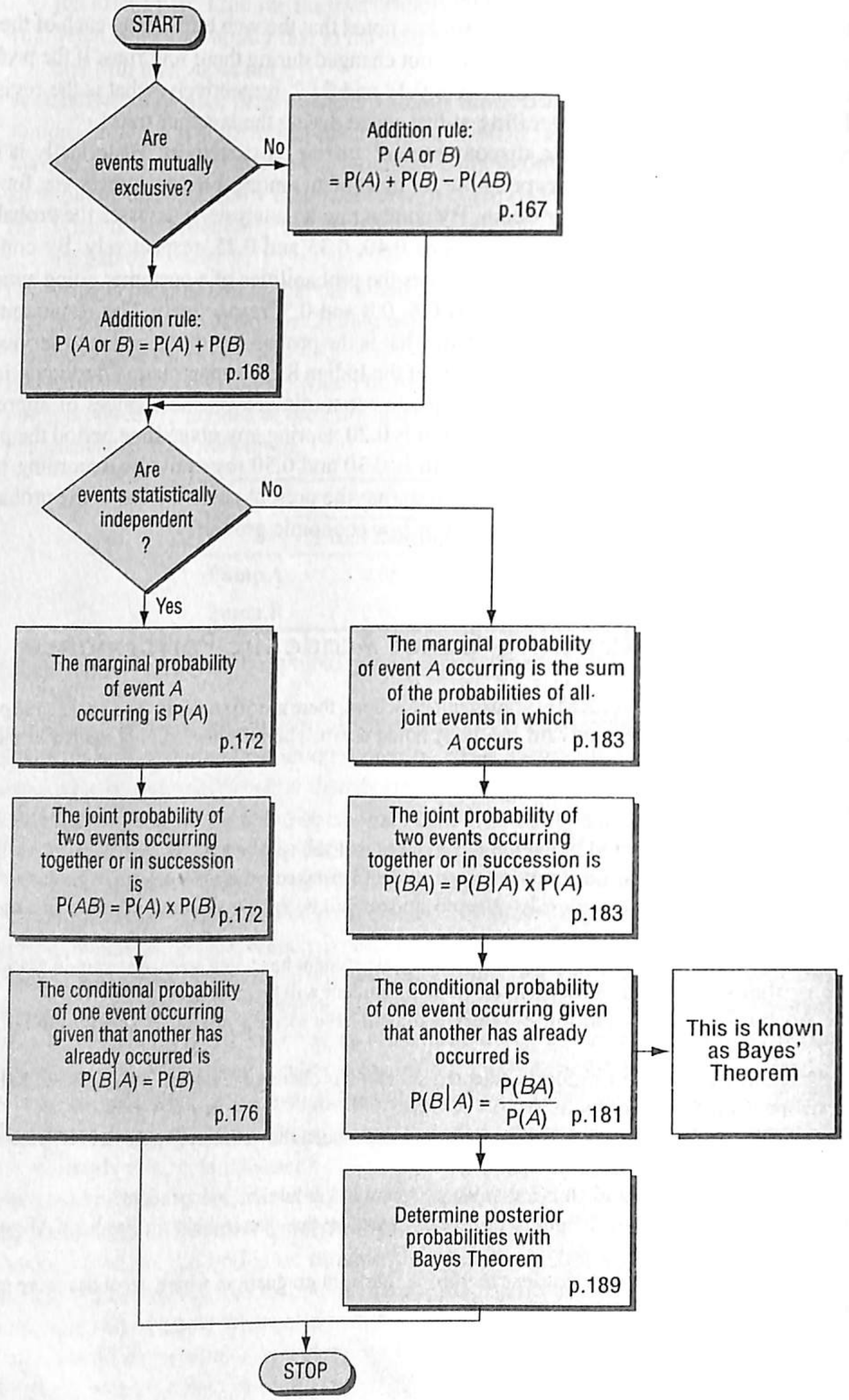
Flow Chart: Probability I: Introductory Ideas



5 Probability Distributions

LEARNING OBJECTIVES

After reading this chapter, you can understand:

- To introduce the probability distributions most commonly used in decision making
- To use the concept of expected value to make decisions
- To show which probability distribution to use and how to find its values
- To understand the limitations of each of the probability distributions you use

CHAPTER CONTENTS

Modern filling machines are designed to work efficiently and with high reliability. Machines can fill toothpaste pumps to within 0.1 ounce of the desired level 80 percent of the time. A visitor to the plant, watching filled pumps being placed into cartons, asked, "What's the chance that exactly half the pumps in a carton selected at random will be filled to within 0.1 ounce of the desired level?" Although we cannot make an exact forecast, the ideas about probability distributions discussed in this chapter enable us to give a pretty good answer to the question. ■

5.1 WHAT IS A PROBABILITY DISTRIBUTION?

In Chapter 2, we described frequency distributions as a useful way of summarizing variations in observed data. We prepared frequency distributions by listing all the possible outcomes of an experiment and then indicating the observed frequency of each possible outcome. *Probability distributions* are related to frequency distributions. **In fact, we can think of a probability distribution as a theoretical frequency distribution.** Now, what does that mean? A theoretical frequency distribution is a probability distribution that describes how outcomes are *expected* to vary. Because these distributions deal with expectations, they are useful models in making inferences and decisions under conditions of uncertainty. In later chapters, we will discuss the methods we use under these conditions.

Probability distributions and frequency distribution

Examples of Probability Distributions

To begin our study of probability distributions, let's go back to the idea of a fair coin, which we introduced in Chapter 4. Suppose we toss a fair coin twice. Table 5-1 illustrates the possible outcomes from this two-toss experiment.

Experiment using a fair coin

Now suppose that we are interested in formulating a probability distribution of the number of tails that could possibly result when we toss the coin twice. We would begin by noting any outcome that did *not* contain a tail. With a fair coin, that is only the third outcome in Table 5-1: *H, H*. Then we would note the outcomes containing only one tail (the second and fourth outcomes in Table 5-1) and, finally, we would note that the first outcome contains two tails. In Table 5-2, we rearrange the outcomes of Table 5-1 to emphasize the number of tails contained in each outcome. We must be careful to note at this point that Table 5-2 is *not* the actual outcome of tossing a fair coin twice. Rather, it is a *theoretical* outcome, that is, it represents the way in which we would *expect* our two-toss experiment to behave over time.

TABLE 5-1 POSSIBLE OUTCOMES FROM TWO TOSSES OF A FAIR COIN

First Toss	Second Toss	Number of Tails on Two Tosses	Probability of the Four Possible Outcomes
T	*T*	2	$0.5 \times 0.5 = 0.25$
T	*H*	1	$0.5 \times 0.5 = 0.25$
H	*H*	0	$0.5 \times 0.5 = 0.25$
H	*T*	1	$0.5 \times 0.5 = 0.25$
			1.00

TABLE 5-2 PROBABILITY DISTRIBUTION OF THE POSSIBLE NUMBER OF TAILS FROM TWO TOSSES OF A FAIR COIN

Number of Tails, T	Tosses	Probability of This Outcome P(*T*)
0	(*H*, *H*)	0.25
1	(*T*, *H*) + (*H*, *T*)	0.50
2	(*T*, *T*)	0.25

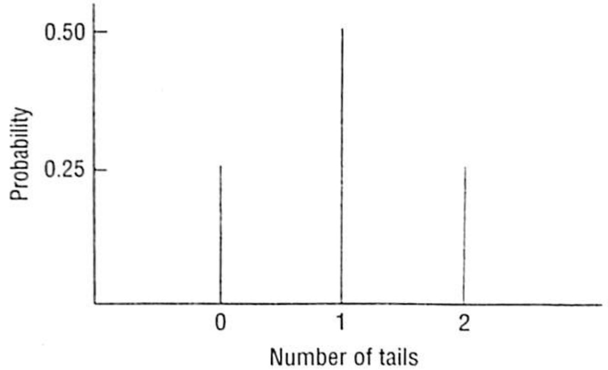


FIGURE 5-1 PROBABILITY DISTRIBUTION OF THE NUMBER OF TAILS IN TWO TOSSES OF A FAIR COIN

We can illustrate in graphic form the probability distribution in Table 5-2. To do this, we graph the number of tails we might see on two tosses against the probability that this number would happen. We show this graph in Figure 5-1.

Consider another example. A political candidate for local office is considering the votes she can get in a coming election. Assume that votes can take on only four possible values. If the candidate's assessment is like this:

Voting Example

Number of votes	1,000	2,000	3,000	4,000	
Probability this will happen	0.1	0.3	0.4	0.2	**Total 1.0**

then the graph of the probability distribution representing her expectations will be like the one shown in Figure 5-2.

Before we move on to other aspects of probability distributions, we should point out that a **frequency distribution is a listing of the observed frequencies of all the outcomes of an**

Difference between frequency distributions and probability distributions

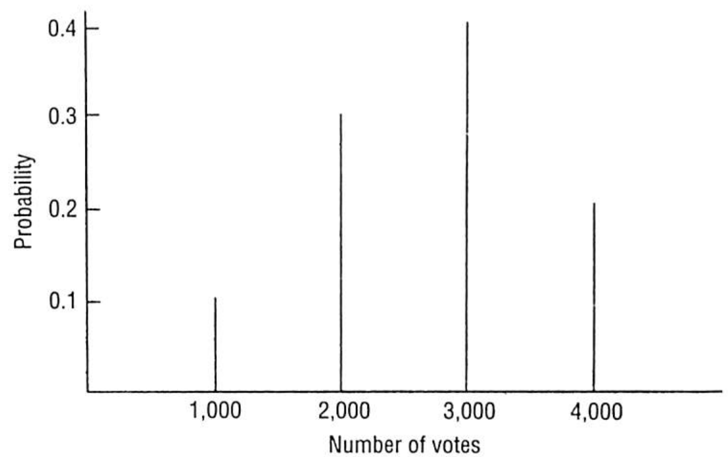


FIGURE 5-2 PROBABILITY DISTRIBUTION OF THE NUMBER OF VOTES

experiment that actually occurred when the experiment was done, whereas a probability distribution is a listing of the probabilities of all the possible outcomes that *could* result if the experiment were done. Also, as we can see in the two examples we presented in Figures 5-1 and 5-2, probability distributions can be based on theoretical considerations (the tosses of a coin) or on a subjective assessment of the likelihood of certain outcomes (the candidate's estimate). Probability distributions can also be based on experience. Insurance company actuaries determine insurance premiums, for example, by using long years of experience with death rates to establish probabilities of dying among various age groups.

Types of Probability Distributions

Discrete probability distributions

Probability distributions are classified as either *discrete* or *continuous*. A discrete probability can take on only a limited number of values, which can be listed. An example of a discrete probability distribution is shown in Figure 5-2, where we expressed the candidate's ideas about the coming election. There, votes could take on only four possible values (1,000, 2,000, 3,000, or 4,000). Similarly, the probability that you were born in a given month is also discrete because there are only 12 possible values (the 12 months of the year).

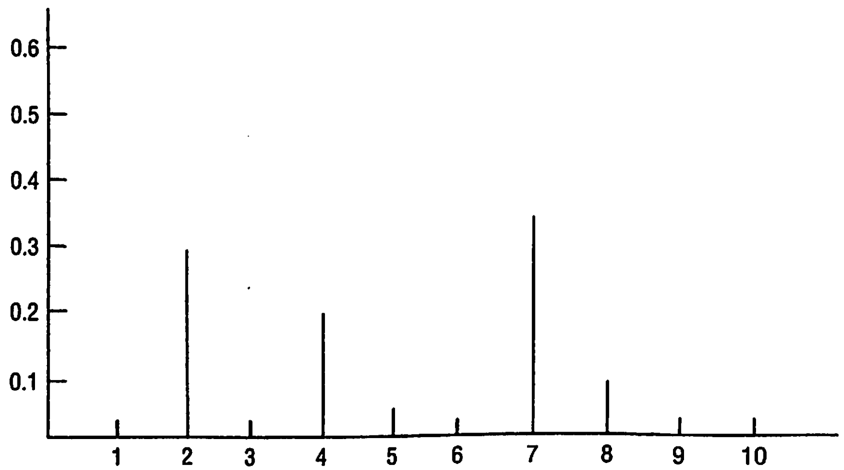
Continuous probability distributions

In a continuous probability distribution, on the other hand, the variable under consideration is allowed to take on any value within a given range, so we *cannot* list all the possible values. Suppose we were examining the level of effluent in a variety of streams, and we measured the level of effluent by parts of effluent per million parts of water. We would expect quite a continuous range of parts per million (ppm), all the way from very low levels in clear mountain streams to extremely high levels in polluted streams. In fact, it would be quite normal for the variable "parts per million" to take on an enormous number of values. We would call the distribution of this variable (ppm) a continuous distribution. Continuous distributions are convenient ways to represent discrete distributions that have many possible outcomes, all very close to each other.

EXERCISES 5.1

Basic Concepts

5-1 Based on the following graph of a probability distribution, construct the corresponding table.



5-2 In the last chapter, we looked at the possible outcomes of tossing two dice, and we calculated some probabilities associated with various outcomes. Construct a table and a graph of the probability distribution representing the outcomes (in terms of total numbers of dots showing on both dice) for this experiment.

5-3 Which of the following statements regarding probability distributions are correct?

(a) A probability distribution provides information about the long-run or expected frequency of each outcome of an experiment.

(b) The graph of a probability distribution has the possible outcomes of an experiment marked on the horizontal axis.

(c) A probability distribution lists the probabilities that each outcome is random.

(d) A probability distribution is always constructed from a set of observed frequencies like a frequency distribution.

(e) A probability distribution may be based on subjective estimates of the likelihood of certain outcomes.

Applications

5-4 The regional chairman of the Muscular Dystrophy Association is trying to estimate the amount each caller will pledge during the annual MDA telethon. Using data gathered over the past 10 years, she has computed the following probabilities of various pledge amounts. Draw a graph illustrating this probability distribution.

Dollars pledged	25	50	75	100	125
Probability	0.45	0.25	0.15	0.10	0.05

5-5 Southport Autos offers a variety of luxury options on its cars. Because of the 6- to 8-week waiting period for customer orders, Ben Stoler, the dealer, stocks his cars with a variety of options. Currently, Mr. Stoler, who prides himself on being able to meet his customers' needs immediately, is worried because of an industrywide shortage of cars with V-8 engines. Stoler offers the following luxury combinations:

1. V-8 engine	electric sun roof	halogen headlights
2. Leather interior	power door locks	stereo cassette deck
3. Halogen headlights	V-8 engine	leather interior
5 Stereo cassette deck	V-8 engine	power door locks

Stoler thinks that combinations 2, 3, and 4 have an equal chance of being ordered, but that combination 1 is twice as likely to be ordered as any of these.

(a) What is the probability that any one customer ordering a luxury car will order one with a V-8 engine?

(b) Assume that two customers order luxury cars. Construct a table showing the probability distribution of the number of V-8 engines ordered.

5-6 Jim Rieck, a marketing analyst for Flatt and Mitney Aircraft, believes that the company's new Tigerhawk jet fighter has a 70 percent chance of being chosen to replace the U.S. Air Force's current jet fighter completely. However, there is one chance in five that the Air Force is going to buy only enough Tigerhawks to replace half of its 5,000 jet fighters. Finally, there is one chance in 10 that the Air Force will replace all of its jet fighters with Tigerhawks and will buy enough Tigerhawks to expand its jet fighter fleet by 10 percent. Construct a table and draw a graph of the probability distribution of sales of Tigerhawks to the Air Force.

5.2 RANDOM VARIABLES

Random variable defined

A variable is random if it takes on different values as a result of the outcomes of a random experiment. A random variable can be either discrete or continuous. If a random variable is allowed to take on only a limited number of values, which can be listed, it is a *discrete random variable.* On the other hand, if it is allowed to assume any value within a given range, it is a *continuous random variable.*

Example of discrete random variables

You can think of a random variable as a value or magnitude that changes from occurrence to occurrence in no predictable sequence. A breast-cancer screening clinic, for example, has no way of knowing exactly how many women will be screened on any one day, so tomorrow's number of patients is a random variable. The values of a random variable are the numerical values corresponding to each possible outcome of the random experiment. If past daily records of the clinic indicate that the values of the random variable range from 100 to 115 patients daily, the random variable is a discrete random variable.

Creating a probability distribution

Table 5-3 illustrates the number of times each level has been reached during the last 100 days. Note that the table gives a frequency distribution. To the extent that we believe that the experience of the past 100 days has been typical, we can use this historical record to assign a probability to each possible number of patients and find a probability distribution. We have accomplished this in Table 5-4 by *normalizing* the observed frequency distribution

TABLE 5-3 NUMBER OF WOMEN SCREENED DAILY DURING 100 DAYS

Number Screened	Number of Days This Level Was Observed
100	1
101	2
102	3
103	5
104	6
105	7
106	9
107	10
108	12
109	11
110	9
111	8
112	6
113	5
114	4
115	2
	100

TABLE 5-4 PROBABILITY DISTRIBUTION FOR NUMBER OF WOMEN SCREENED

Number Screened (Value of the Random Variable)	Probability That the Random Variable Will Take on This Value
100	0.01
101	0.02
102	0.03
103	0.05
104	0.06
105	0.07
106	0.09
107	0.10
108	0.12
109	0.11
110	0.09
111	0.08
112	0.06
113	0.05
114	0.04
115	0.02
	1.00

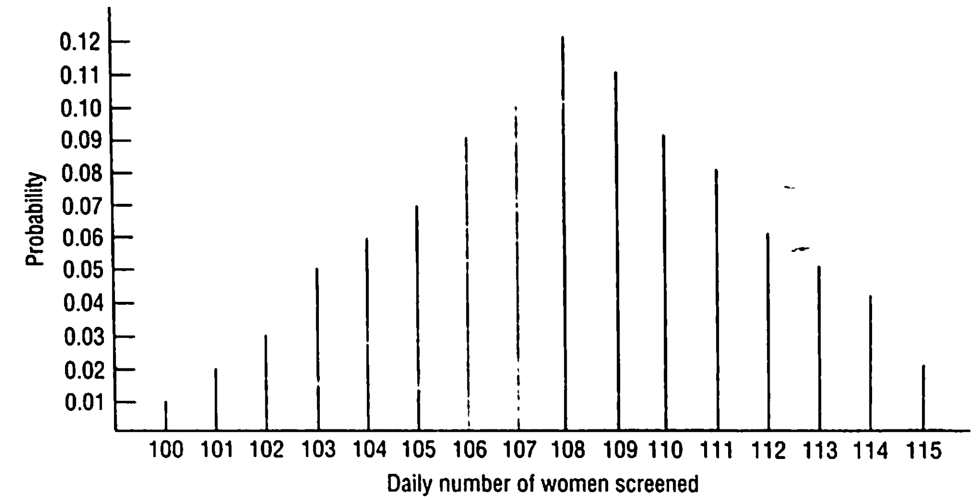


FIGURE 5-3 PROBABILITY DISTRIBUTION FOR THE DISCRETE RANDOM VARIABLE "DAILY NUMBER SCREENED"

(in this case, dividing each value in the right-hand column of Table 5-3 by 100, the total number of days for which the record has been kept). The probability distribution for the random variable "daily number screened" is illustrated graphically in Figure 5-3. Notice that the probability distribution for a random variable provides a probability for each possible value and that these probabilities must sum to 1. Table 5-4 shows that both these requirements have been met. Furthermore, both Table 5-4 and Figure 5-3 give us information about the long-run frequency of occurrence of daily patient screenings we would expect to observe if this random "experiment" were repeated.

The Expected Value of a Random Variable

Suppose you toss a coin 10 times and get 7 heads, like this:

Heads	Tails	Total
7	3	10

"Hmm, strange," you say. You then ask a friend to try tossing the coin 20 times; she gets 15 heads and 5 tails. So now you have, in all, 22 heads and 8 tails out of 30 tosses.

What did you expect? Was it something closer to 15 heads and 15 tails (half and half)? Now suppose you turn the tossing over to a machine and get 792 heads and 208 tails out of 1,000 tosses of the same coin. You might now be suspicious of the coin because it didn't live up to what you expected.

Expected value is a fundamental idea in the study of probability distributions. For many years, the concept has been put to considerable practical use by the insurance industry, and in the last 40 years, it has been widely used by many others who must make decisions under conditions of uncertainty.

Calculating expected value

To obtain the **expected value of a discrete random variable,** we multiply each value that the random variable can assume by the probability of occurrence of that value and then sum these products. Table 5-5 illustrates this procedure for our clinic problem. The total in the table tells us that the expected value of the discrete random variable "number screened" is 108.02 women. What does this mean? It means that over a long period of time, the number of daily screenings should average about 108.02.

Remember that an-expected value of 108.02 does *not* mean that tomorrow exactly 108.02 women will visit the clinic.

The clinic director would base her decisions on the expected value of daily screenings because the expected value is a *weighted average of the outcomes she expects in the future.* Expected value *weights* each possible outcome by the frequency with which it is expected to occur. Thus, more common occurrences are given more weight than are less common ones. As conditions change over time, the director would recompute the expected value of daily screenings and use this new figure as a basis for decision making.

Deriving expected value subjectively

In our clinic example, the director used past patients' records as the basis for calculating the expected value of daily screenings. The expected value can also be derived from the director's subjective assessments of the probability that the random variable will take on certain values. In that case, the expected value represents nothing more than her personal convictions about the possible outcome.

In this section, we have worked with the probability distribution of a random variable in tabular form (Table 5-5) and in graphic form (Figure 5-3). In many situations, however, we will find it more convenient, in terms of the computations that must be done, to represent the probability distribution of a random variable in *algebraic* form. By doing this, we can make probability calculations by substituting numerical values directly into an algebraic formula. In the following sections, we shall illustrate some situations in which this is appropriate and methods for accomplishing it.

TABLE 5-5 CALCULATING THE EXPECTED VALUE OF THE DISCRETE RANDOM VARIABLE "DAILY NUMBER SCREENED"

Possible Values of the Random Variable (1)	Probability That the Random Variable Will Take on These Values (2)	(1) × (2)
100	0.01	1.00
101	0.02	2.02
102	0.03	3.06
103	0.05	5.15
104	0.06	6.24
105	0.07	7.35
106	0.09	9.54
107	0.10	10.70
108	0.12	12.96
109	0.11	11.99
110	0.09	9.90
111	0.08	8.88
112	0.06	6.72
113	0.05	5.65
114	0.04	4.56
115	0.02	2.30
Expected value of the random variable "daily number screened" →		**108.02**

HINTS & ASSUMPTIONS

The expected value of a discrete random variable is nothing more than the weighted average of each possible outcome, multiplied by the probability of that outcome happening, just like we did it in Chapter 3. Warning: The use of the term *expected* can be misleading. For example, if we calculated the expected value of number of women to be screened to be 11, we *don't* think exactly this many will show up tomorrow. We are saying that, absent any other information, 11 women is the best number we can come up with as a basis for planning how many nurses we'll need to screen them. Hint: If daily patterns in the data are discernible (more women on Monday than on Friday, for example) then build this into your decision. The same holds for monthly and seasonal patterns in the data.

EXERCISES 5.2

Self-Check Exercises

SC 5-1 Construct a probability distribution based on the following frequency distribution.

Outcome	102	105	108	111	114	117
Frequency	10	20	45	15	20	15

(a) Draw a graph of the hypothetical probability distribution.
(b) Compute the expected value of the outcome.

SC 5-2 Bob Walters, who frequently invests in the stock market, carefully studies any potential investment. He is currently examining the possibility of investing in the Trinity Power Company. Through studying past performance, Walters has broken the potential results of the investment into five possible outcomes with accompanying probabilities. The outcomes are annual rates of return on a single share of stock that currently costs \$150. Find the expected value of the return for investing in a single share of Trinity Power.

Return on investment (\$)	0.00	10.00	15.00	25.00	50.00
Probability	0.20	0.25	0.30	0.15	0.10

If Walters purchases stock whenever the expected rate of return exceeds 10 percent, will he purchase the stock, according to these data?

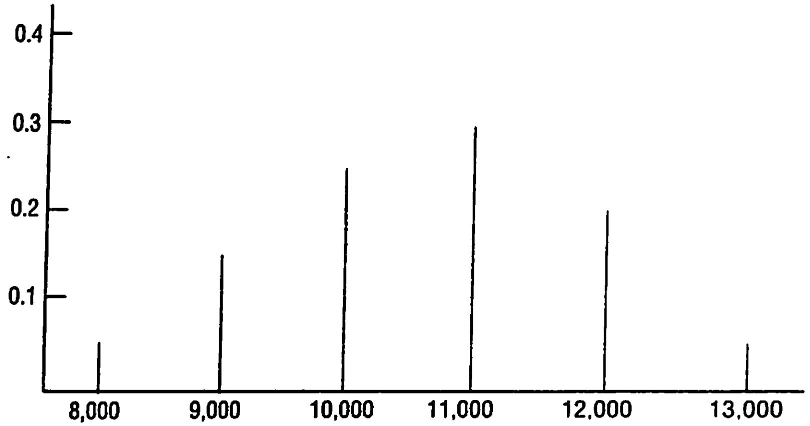
Basic Concepts

5-7 Construct a probability distribution based on the following frequency distribution:

Outcome	2	4	6	8	10	12	15
Frequency	24	22	16	12	7	3	1

(a) Draw a graph of the hypothetical probability distribution.
(b) Compute the expected value of the outcome.

5-8 From the following graph of a probability distribution
(a) Construct a table of the probability distribution.
(b) Find the expected value of the random variable.



5-9 The only information available to you regarding the probability distribution of a set of outcomes is the following list of frequencies:

X	0	15	30	45	60	75
Frequency	25	125	75	175	75	25

(a) Construct a probability distribution for the set of outcomes.
(b) Find the expected value of an outcome.

Applications

5-10 Bill Johnson has just bought a VCR from Jim's Videotape Service at a cost of $300. He now has the option of buying an extended service warranty offering 5 years of coverage for $100. After talking to friends and reading reports, Bill believes the following maintenance expenses could be incurred during the next five years:

Expense	0	50	100	150	200	250	300
Probability	0.35	0.25	0.15	0.10	0.08	0.05	0.02

Find the expected value of the anticipated maintenance costs. Should Bill pay $100 for the warranty?

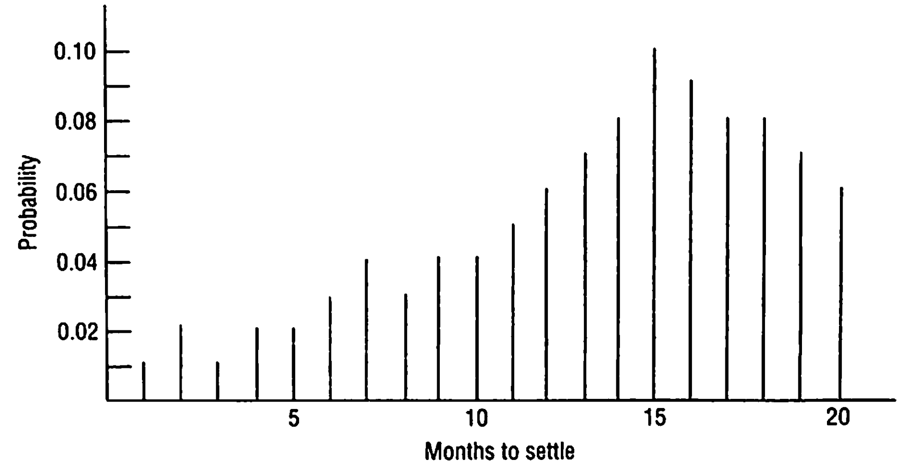
5-11 Steven T. Opsine, supervisor of traffic signals for the Fairfax County division of the Virginia State Highway Administration, must decide whether to install a traffic light at the reportedly dangerous intersection of Dolley Madison Blvd. and Lewinsville Rd. Toward this end, Mr. Opsine has collected data on accidents at the intersection:

	Number of Accidents											
Year	**J**	**F**	**M**	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**	**D**
1995	10	8	10	6	9	12	2	10	10	0	7	10
1996	12	9	7	8	4	3	7	14	8	8	8	4

S.H.A. policy is to install a traffic light at an intersection at which the monthly expected number of accidents is higher than 7. According to this criterion, should Mr. Opsine recommend that a traffic light be installed at this intersection?

5-12 Alan Sarkid is the president of the Dinsdale Insurance Company and he is concerned about the high cost of claims that take a long time to settle. Consequently, he has asked his chief actuary,

Dr. Ivan Acke, to analyze the distribution of time until settlement. Dr. Acke has presented him with the following graph:



Dr. Acke also informed Mr. Sarkid of the expected amount of time to settle a claim. What is this figure?

5-13 The fire marshal of Baltimore County, Maryland, is compiling a report on single-family-dwelling fires. He has the following data on the number of such fires from the last 2 years:

	Number of Fires											
Year	**J**	**F**	**M**	**A**	**M**	**J**	**J**	**A**	**S**	**O**	**N**	**D**
1995	25	30	15	10	10	5	2	2	1	4	8	10
1996	20	25	10	8	5	2	4	0	5	8	10	15

Based on these data
(a) What is the expected number of single-family-dwelling fires per month?
(b) What is the expected number of single-family-dwelling fires per winter month (January, February, March)?

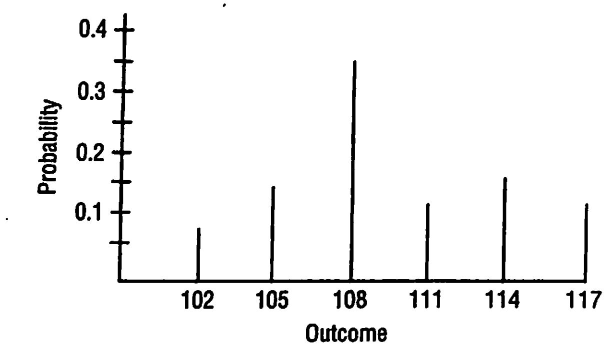
5-14 Ted Olson, the director of Overnight Delivery, Inc., has become concerned about the number of first-class letters lost by his firm. Because these letters are carried by both truck and airplane, Mr. Olson has broken down the lost letters for the last year into those lost from trucks and those lost from airplanes. His data are as follows:

Number Lost from	J	F	M	A	M	J	J	A	S	O	N	D
Truck	4	5	2	3	2	1	3	5	4	7	0	1
Airplane	5	6	0	2	1	3	4	2	4	7	4	0

Mr. Olson plans to investigate either the trucking or air division of the company, but not both. If he decides to investigate the division with the highest expected number of lost letters per month, which will he investigate?

Worked-Out Answers to Self-Check Exercises

SC 5-1 (a)



(b)

Outcome (1)	Frequency (2)	P(Outcome) (3)	(1) × (3)
102	10	0.08	8.16
105	20	0.16	16.80
108	45	0.36	38.88
111	15	0.12	13.32
114	20	0.16	18.24
117	15	0.12	14.04
	125	1.00	109.44 = Expected outcome

SC 5-2

Return (1)	P(Return) (2)	(1) × (2)
0	0.20	0.00
10	0.25	2.50
15	0.30	4.50
25	0.15	3.75
50	0.10	5.00
	1.00	15.75 = Expected return

Bob will purchase the stock because the expected return of $15.75 is greater than 10 percent of the $150 purchase price.

5.3 USE OF EXPECTED VALUE IN DECISION MAKING

In the preceding section, we calculated the expected value of a random variable and noted that it can have significant value to decision makers. Now we need to take a moment to illustrate how decision makers combine the probabilities that a random variable will take on certain values with the monetary gain or loss that results when it does take on those values. Doing just this enables them to make intelligent decisions under uncertain conditions.

TABLE 5-6 SALES DURING 100 DAYS

Daily Sales	Number of Days Sold	Probability of Each Number Being Sold
10	15	0.15
11	20	0.20
12	40	0.40
13	25	0.25
	100	1.00

Combining Probabilities and Monetary Values

Wholesaler problem

Let us look at the case of a fruit and vegetable wholesaler who sells strawberries. This product has a very limited useful life. If not sold on the day of delivery, it is worthless. One case of strawberries costs $20, and the wholesaler receives $50 for it. The wholesaler cannot specify the number of cases customers will call for on any one day, but her analysis of past records has produced the information in Table 5-6.

Types of Losses Defined

Obsolescence and opportunity losses

Two types of losses are incurred by the wholesaler: (1) *obsolescence losses*, caused by stocking too much fruit on any one day and having to throw it away the next day; and (2) *opportunity losses*, caused by being out of strawberries any time that customers call for them. (Customers will not wait beyond the day a case is requested.)

Table of conditional losses

Table 5-7 is a table of conditional losses. Each value in the table is conditional on a specific number of cases being stocked and a specific number being requested. The values in Table 5-7 include not only losses from decaying berries, but also those losses resulting from lost revenue when the wholesaler is unable to supply the requests she receives for the berries.

Obsolescence losses

Neither of these two types of losses is incurred when the number of cases stocked on any one day is the same as the number of cases requested. When that happens, the wholesaler sells all she has stocked and incurs no losses. This situation is indicated by a colored zero in the appropriate column. Figures above any zero represent losses arising from spoiled berries. In each case here, the number of cases stocked is greater than the number requested. For example, if the wholesaler stocks 12 cases but receives requests for only 10 cases, she loses $40 (or $20 per case for spoiled strawberries).

TABLE 5-7 CONDITIONAL LOSS TABLE

Possible Requests for Strawberries	Possible Stock Options			
	10	11	12	13
10	$0	$20	$40	$60
11	30	0	20	40
12	60	30	0	20
13	90	60	30	0

TABLE 5-8 EXPECTED LOSS FROM STOCKING 10 CASES

Possible Requests	Conditional Loss		Proability of This Many Requests		Expected Loss
10	$0	×	0.15	=	$0.00
11	30	×	0.20	=	6.00
12	60	×	0.40	=	24.00
13	90	×	0.25	=	22.50
			1.00		$52.50

Opportunity losses

Values **below** the colored zeros represent opportunity losses resulting from requests that cannot be filled. If only 10 cases are stocked on a day that 11 requests are received, the wholesaler suffers an opportunity loss of $30 for the case she cannot sell ($50 income per case that would have been received, minus $20 cost, equals $30).

Calculating Expected Losses

Meaning of expected loss

Examining each possible stock action, we can compute the expected loss. We do this by weighting each of the four possible loss figures in each column of Table 5-7 by the probabilities from Table 5-6. For a stock action of 10 cases, the expected loss is computed as in Table 5-8.

The conditional losses in Table 5-8 are taken from the second column of Table 5-7 for a stock action of 10 cases. The fourth column total in Table 5-8 shows us that if 10 cases are stocked each day, over a long period of time, the average or expected loss will be $52.50 a day. There is no guarantee that *tomorrow's* loss will be exactly $52.50.

Optimal solution

Tables 5-9 through 5-11 show the computations of the expected loss resulting from decisions to stock 11, 12, and 13 cases, respectively. **The optimal stock action is the one that will minimize expected losses.** This action calls for the stocking of 12 cases each day, at which point the expected loss is minimized at $17.50. We could just as easily have solved this problem by taking an alternative approach, that is, *maximizing expected gain* ($50 received per case less $20 cost per case) instead of minimizing expected loss. The answer, 12 cases, would have been the same.

TABLE 5-9 EXPECTED LOSS FROM STOCKING 11 CASES

Possible Requests	Conditional Loss		Probability of This Many Requests		Expected Loss
10	$20	×	0.15	=	$3.00
11	0	×	0.20	=	0.00
12	30	×	0.40	=	12.00
13	60	×	0.25	=	15.00
			1.00		$30.00

TABLE 5-10 EXPECTED LOSS FROM STOCKING 12 CASES

Possible Requests	Conditional Loss		Probability of This Many Requests		Expected Loss
10	$ 40	×	0.15	=	$ 6.00
11	20	×	0.20	=	4.00
12	0	×	0.40	=	0.00
13	30	×	0.25	=	7.50
			1.00 Minimum expected loss →		$17.50

TABLE 5.11 EXPECTED LOSS FROM STOCKING 13 CASES

Possible Requests	Conditional Loss		Probability of This Many Requests		Expected Loss
10	$ 60	×	0.15	=	$9.00
11	40	×	0.20	=	8.00
12	20	×	0.40	=	8.00
13	0	×	0.25	=	0.00
			1.00		$25.00

In our brief treatment of expected value, we have made quite a few assumptions. To name only two, we've assumed that demand for the product can take on only four values, and that the berries are worth nothing one day later. Both these assumptions reduce the value of the answer we got. In Chapter 17, you will again encounter expected-value decision making, but there we will develop the ideas as a part of statistical decision theory (a broader use of statistical methods to make decisions), and we shall devote an entire chapter to expanding the basic ideas we have developed at this point.

HINTS & ASSUMPTIONS

Warning: In our illustrative exercise, we've allowed the random variable to take on only our values. This is unrealistic in the real world and we did it here only to make the explanation easier. Any manager facing this problem in her job would know that demand might be as low as zero on a given day (weather, holidays) and as high as perhaps 50 cases on another day. Hint: With demand ranging from zero to 50 cases, it's a computational nightmare to solve this problem by the method we just used. But don't panic, we will introduce another method in Chapter 17 that can do this easily.

EXERCISES 5.3

Self-Check Exercise

SC 5-3 Mario, owner of Mario's Pizza Emporium, has a difficult decision on his hands. He has found that he always sells between one and four of his famous "everything but the kitchen sink" pizzas per night. These pizzas take so long to prepare, however, that Mario prepares all of them in advance and stores them in the refrigerator. Because the ingredients go bad within one day, Mario always throws out any unsold pizzas at the end of each evening. The cost of preparing each pizza is $7, and Mario sells each one for $12. In addition to the usual costs, Mario also calculates that each "everything but" pizza that is ordered but he cannot deliver due to insufficient stock costs him $5 in future business. How many "everything but" pizzas should Mario stock each night in order to minimize expected loss if the number of pizzas ordered has the following probability distribution?

Number of pizzas demanded	1	2	3	4
Probability	0.40	0.30	0.20	0.10

Applications

5-15 Harry Byrd, the director of publications for the Baltimore Orioles, is trying to decide how many programs to print for the team's upcoming three-game series with the Oakland A's. Each program costs 25¢ to print and sells for $1.25. Any programs unsold at the end of the series must be discarded. Mr. Byrd has estimated the following probability distribution for program sales, using data from past program sales:

Programs sold	25,000	40,000	55,000	70,000
Probability	0.10	0.30	0.45	0.15

Mr. Byrd has decided to print either 25, 40, 55, or 70 thousand programs. Which number of programs will minimize the team's expected losses?

5-16 Airport Rent-a-Car is a locally operated business in competition with several major firms. ARC is planning a new deal for prospective customers who want to rent a car for only one day and will return it to the airport. For $35, the company will rent a small economy car to a customer, whose only of her expense is to fill the car with gas at day's end. ARC is planning to buy number of small cars from the manufacturer at a reduced price of $6,300. The big question is how many to buy. Company executives have decided the following distribution of demands per day for the service:

Number of cars rented	13	14	15	16	17	18
Probability	0.08	0.15	0.22	0.25	0.21	0.09

The company intends to offer the plan 6 days a week (312 days per year) and anticipates that its variable cost per car per day will be $2.50. After the end of one year, the company expects to sell the cars and recapture 50 percent of the original cost. Disregarding the time value of money and any noncash expenses, use the expected-loss method to determine the optimal number of cars for ARC to buy.

5-17 We Care Air needs to make a decision about Flight 105. There are currently 3 seats reserved for last-minute customers, but the airline does not know if anyone will buy them. If they release the seats now, they know they will be able to sell them for $250 each. Last-minute customers must pay $475 per seat. The decision must be made now, and any number of seats may be released. We Care Air has the following probability distribution to help them:

Number of last-minute customers requesting seats	0	1	2	3
Probability	0.45	0.30	0.15	0.10

The company also counts a $150 loss of goodwill for every last-minute customer who is turned away.

(a) How much revenue will be generated by releasing all 3 seats now?
(b) What is the company's expected net revenue (revenue less loss of goodwill) if 3 seats are released now?
(c) What is the company's expected net revenue if 2 seats are released now?
(d) How many seats should be released to maximize expected revenue?

Worked-Out Answer to Self-Check Exercise

SC 5-3

Loss Table

	Pizzas Demanded				
	1	**2**	**3**	**4**	
Probability	**0.4**	**0.3**	**0.2**	**0.1**	
Pizzas Stocked					**Expected Loss**
1	0	10	20	30	10.0
2	7	0	10	20	6.8 ←
3	14	7	0	10	8.7
4	21	14	7	0	14.0

Mario should stock two "everything but" pizzas each night.

5.4 THE BINOMIAL DISTRIBUTION

The binomial distribution and Bernoulli processes

One widely used probability distribution of a discrete random variable is the *binomial distribution.* It describes a variety of processes of interest to managers. The binomial distribution describes discrete, not continuous, data, resulting from an experiment known as a *Bernoulli process*, after the seventeenth-century Swiss mathematician Jacob Bernoulli. The tossing of a fair coin a fixed number of times is a Bernoulli process, and the outcomes of such tosses can be represented by the binomial probability distribution. The success or failure of interviewees on an aptitude test may also be described by a Bernoulli process. On the other hand, the frequency distribution of the lives of fluorescent lights in a factory would be measured on a continuous scale of hours and would not qualify as a binomial distribution.

Use of the Bernoulli Process

Bernoulli process described

We can use the outcomes of a fixed number of tosses of a fair coin as an example of a Bernoulli process. We can describe this process as follows:

1. Each trial (each toss, in this case) has only *two* possible outcomes: heads or tails, yes or no, success or failure.
2. The probability of the outcome of any trial (toss) remains *fixed* over time. With a fair coin, the probability of heads remains 0.5 each toss regardless of the number of times the coin is tossed.
3. The trials are *statistically independent*; that is, the outcome of one toss does not affect the outcome of any other toss.

Each Bernoulli process has its own characteristic probability. Take the situation in which historically seven-tenths of all people who applied for a certain type of job passed the job test. We would say that the characteristic probability here is 0.7, but we could describe our testing results as Bernoulli only if we felt certain that the proportion of those passing the test (0.7) remained constant over time. The other characteristics of the Bernoulli process would also have to be met, of course. Each test would have only two outcomes (success or failure), and the results of each test would have to be statistically independent.

Characteristic probability defined

In more formal language, the symbol p represents the probability of a success (in our example, 0.7), and the symbol q ($q = 1 - p$), the probability of a failure (0.3). To represent a certain number of successes, we will use the symbol r, and to symbolize the total number of trials, we use the symbol n. In the situations we will be discussing, the number of trials is fixed before the experiment is begun.

Using this language in a simple problem, we can calculate the chances of getting exactly two heads (in any order) on three tosses of a fair coin. Symbolically, we express the values as follows:

- p = characteristic probability or probability of success = 0.5
- $q = 1 - p$ = probability of failure = 0.5
- r = number of successes desired = 2
- n = number of trials undertaken = 3

We can solve the problem by using the *binomial formula:*

Binomial Formula

Binomial Formula

$$\text{Probability of } r \text{ successes in } n \text{ trials} = \frac{n!}{r!(n-r)!}p^r q^{n-r} \qquad [5\text{-}1]$$

Although this formula may look somewhat complicated, it can be used quite easily. The symbol ! means *factorial*, which is computed as follows: 3! means $3 \times 2 \times 1$, or 6. To calculate 5!, we multiply $5 \times 4 \times 3 \times 2 \times 1 = 120$. Mathematicians define 0! as equal to 1. Using the binomial formula to solve our problem, we discover

$$\text{Probability of 2 successes in 3 trials} = \frac{3!}{2!(3-2)!}(0.5)^2(0.5)^1$$

$$= \frac{3\times2\times1}{(2\times1)(1\times1)}(0.5)^2(0.5)$$

$$= \frac{6}{2}(0.25)(0.5)$$

$$= 0.375$$

Thus, there is a 0.375 probability of getting two heads on three tosses of a fair coin.

By now you've probably recognized that we can use the binomial distribution to determine the probabilities for the toothpaste pump problem we introduced at the beginning of this chapter. Recall that historically, eight-tenths of the pumps were correctly filled (successes). If we want to compute the probability of getting exactly three of six pumps (half a carton) correctly filled, we can define our symbols this way:

$$p = 0.8$$
$$q = 0.2$$
$$r = 3$$
$$n = 6$$

and then use the binomial formula as follows:

$$\text{Probability of } r \text{ successes in } n \text{ trials} = \frac{n!}{r!(n-r)!}p^r q^{n-r} \qquad [5\text{-}1]$$

$$\text{Probability of 3 out of 6 pumps correctly filled} = \frac{6\times5\times4\times3\times2\times1}{(3\times2\times1)(3\times2\times1)}(0.8)^3(0.2)^3$$

$$= \frac{720}{6\times6}(0.512)(0.008)$$

$$= (20)(0.512)(0.008)$$

$$= 0.08192$$

Of course, we *could* have solved these two problems using the probability trees we developed in Chapter 4, but for larger problems, trees become quite cumbersome. In fact, using the binomial formula (Equation 5-1) is no easy task when we have to compute the value of something like 19 factorial. For this reason, binomial probability tables have been developed, and we shall use them shortly.

Binomial tables are available

Some Graphic Illustrations of the Binomial Distribution

To this point, we have dealt with the binomial distribution only in terms of the binomial formula, but the binomial, like any other distribution, can be expressed graphically as well.

To illustrate several of these distributions, consider a situation at Kerr Pharmacy, where employees are often late. Five workers are in the pharmacy. The owner has studied the situation over a period of time and has determined that there is a 0.4 chance of any one employee being late and that they arrive independently of one another. How would we draw a binomial probability distribution illustrating the probabilities of 0, 1, 2, 3, 4, or 5 workers being late simultaneously? To do this, we would need to use the binomial formula, where

$$p = 0.4$$
$$q = 0.6$$
$$n = 5^*$$

and to make a separate computation for each r, from 0 through 5. Remember that, mathematically, any number to the zero power is defined as being equal to 1. Beginning with our binomial formula:

$$\text{Probability of } r \text{ late arrivals out of } n \text{ workers} = \frac{n!}{r!(n-r)!}p^r q^{n-r} \qquad [5\text{-}1]$$

*When we define n, we look at the number of workers. The fact that there is a possibility that none will be late does not alter our choice of $n = 5$.

Using the formula to derive the binomial probability distribution

For $r = 0$, we get

$$P(0) = \frac{5!}{0!(5-0)!}(0.4)^0(0.6)^5$$

$$= \frac{5\times4\times3\times2\times1}{(1)(5\times4\times3\times2\times1)}(1)(0.6)^5$$

$$= \frac{120}{120}(1)(0.07776)$$

$$= (1)(1)(0.07776)$$

$$= 0.07776$$

For $r = 1$, we get

$$P(1) = \frac{5!}{1!(5-1)!}(0.4)^1(0.6)^4$$

$$= \frac{5\times4\times3\times2\times1}{(1)(4\times3\times2\times1)}(0.4)(0.6)^4$$

$$= \frac{120}{24}(0.4)(0.1296)$$

$$= (5)(0.4)(0.1296)$$

$$= 0.2592$$

For $r = 2$, we get

$$P(2) = \frac{5!}{2!(5-2)!}(0.4)^2(0.6)^3$$

$$= \frac{5\times4\times3\times2\times1}{(2\times1)(3\times2\times1)}(0.4)^2(0.6)^3$$

$$= \frac{120}{12}(0.16)(0.216)$$

$$= (10)(0.03456)$$

$$= 0.3456$$

For $r = 3$, we get

$$P(3) = \frac{5!}{3!(5-3)!}(0.4)^3(0.6)^2$$

$$= \frac{5\times4\times3\times2\times1}{(3\times2\times1)(2\times1)}(0.4)^3(0.6)^2$$

$$= (10)(0.064)(0.36)$$

$$= 0.2304$$

For $r = 4$, we get

$$P(4) = \frac{5!}{4!(5-4)!}(0.4)^4(0.6)^1$$

$$= \frac{5\times4\times3\times2\times1}{(4\times3\times2\times1)(1)}(0.4)^4(0.6)$$

$$= (5)(0.0256)(0.6)$$

$$= 0.0768$$

Finally, for $r = 5$, we get

$$P(5) = \frac{5!}{5!(5-5)!}(0.4)^5(0.6)^0$$

$$= \frac{5\times4\times3\times2\times1}{(5\times4\times3\times2\times1)(1)}(0.4)^5(1)$$

$$= (1)(0.01024)(1)$$

$$= 0.01024$$

The binomial distribution for this example is shown graphically in Figure 5-4.

General appearance of binomial distributions

Without doing all the calculations involved, we can illustrate the general appearance of a family of binomial probability distributions. In Figure 5-5, for example, each distribution represents $n = 5$. In each case, the p and q have been changed and are noted beside each distribution. The probabilities in Figure 5-5 sum to slightly less than 1.0000 because of rounding. From Figure 5-5, we can make the following generalizations:

1. When p is small (0.1), the binomial distribution is skewed to the right.
2. As p increases (to 0.3, for example), the skewness is less noticeable.
3. When $p = 0.5$, the binomial distribution is symmetrical.
4. When p is larger than 0.5, the distribution is skewed to the left.

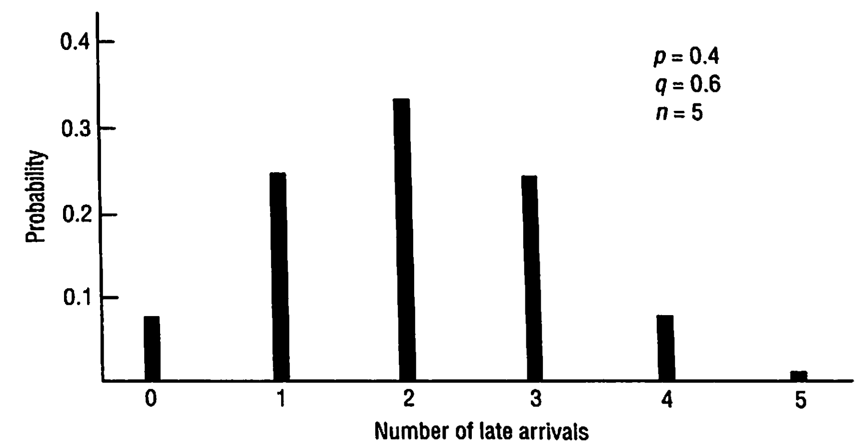


FIGURE 5-4 BINOMIAL PROBABILITY DISTRIBUTION OF LATE ARRIVALS

n = 5, p = 0.1

r	Probability
0	0.5905
1	0.3280
2	0.0729
3	0.0081
4	0.0004
5	0.0000
	0.9999

n = 5, p = 0.3

r	Probability
0	0.1681
1	0.3601
2	0.3087
3	0.1323
4	0.0283
5	0.0024
	0.9999

n = 5, p = 0.5

r	Probability
0	0.0312
1	0.1562
2	0.3125
3	0.3125
4	0.1562
5	0.0312
	0.9998

n = 5, p = 0.7

r	Probability
0	0.0024
1	0.0283
2	0.1323
3	0.3087
4	0.3601
5	0.1681
	0.9999

n = 5, p = 0.7

r	Probability
0	0.0000
1	0.0004
2	0.0081
3	0.0729
4	0.3280
5	0.5905
	0.9999

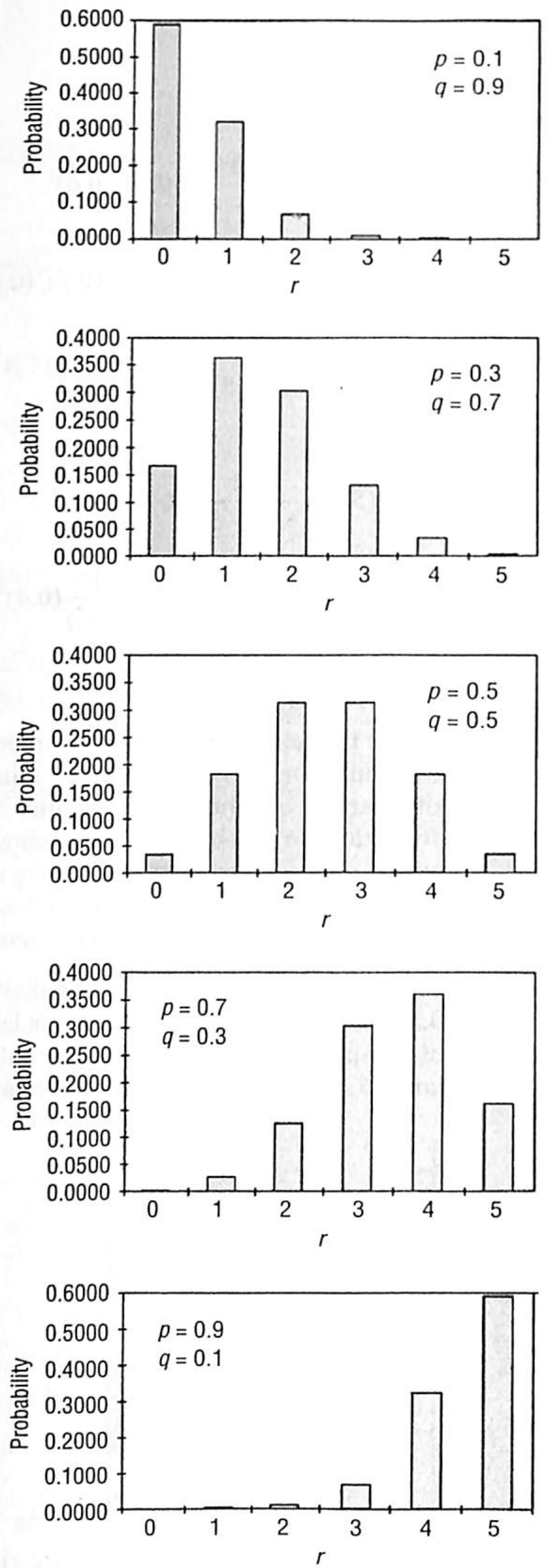


FIGURE 5-5 FAMILY OF BINOMIAL PROBABILITY DISTRIBUTIONS WITH CONSTANT n = 5 AND VARIOUS p AND q VALUES

5. The probabilities for 0.3, for example, are the same as those for 0.7 except that the values of p and q are *reversed.* This is true for any pair of complementary p and q values (0.3 and 0.7, 0.4 and 0.6, and 0.2 and 0.8).

Let us examine graphically what happens to the binomial distribution when p stays constant but n is increased. Figure 5-6 illustrates the general shape of a family of binomial distributions with a constant p of 0.4 and n's from 5 to 30. As n increases, the vertical lines not only become more numerous but also tend to bunch up together to form a *bell shape.* We shall have more to say about this bell shape shortly.

Using the Binomial Tables

Solving problems using the binomial tables

Earlier we recognized that it is tedious to calculate probabilities using the binomial formula when n is a large number. Fortunately, we can use Appendix Table 3 to determine binomial probabilities quickly.

To illustrate the use of the binomial tables, consider this problem. What is the probability that 8 of the 15 registered Democrats on Prince Street will fail to vote in the coming primary if the probability of any individual's not voting is 0.30 and if people decide independently of each other whether or not to vote? First, we represent the elements in this problem in binomial distribution notation:

$n = 15$	number of registered Democrats
$p = 0.30$	probability that any one individual won't vote
$r = 8$	number of individuals who will fail to vote

How to use the binomial tables

Then, because the problem involves 15 trials, we must find the table corresponding to $n = 15$. Because the probability of an individual's not voting is 0.30, we look through the binomial tables until we find the column headed 0.30. We then move down that column until we are opposite the $r = 8$ row, where we read the answer 0.0348. This is the probability of eight registered voters not voting.

Suppose the problem had asked us to find the probability of eight or more registered voters not voting? We would have looked under the 0.30 column and added up the probabilities there from 8 to the bottom of the column like this:

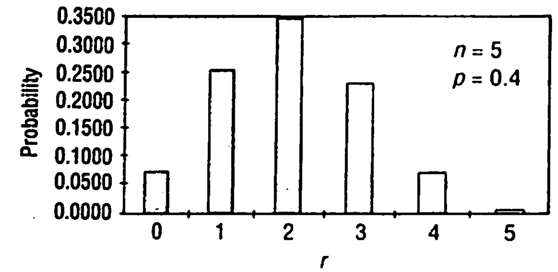
8	0.0348
9	0.0116
10	0.0030
11	0.0006
12	0.0001
13	0.0000
	0.0501

The answer is that there is a 0.0501 probability of eight or more registered voters not voting.

Suppose now that the problem asked us to find the probability of *fewer* than eight non-voters. Again, we would have begun with the 0.30 column, but this time we would add the probabilities from 0 (the top of the $n = 15$ column) *down* to 7 (the highest value less than 8), like this:

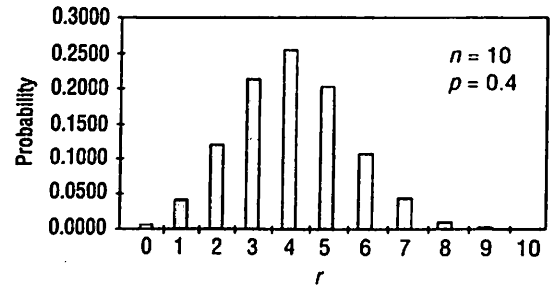
$n = 5, p = 0.4$

r	Probability
0	0.0778
1	0.2592
2	0.3456
3	0.2304
4	0.0768
5	0.0102
	1.0000



$n = 10, p = 0.4$

r	Probability
0	0.0060
1	0.0403
2	0.1209
3	0.2150
4	0.2508
5	0.2007
6	0.1115
7	0.0425
8	0.0106
9	0.0016
10	0.0001
	1.0000



$n = 30, p = 0.4$

r	Probability
0	0.00000
1	0.00000
2	0.00004
3	0.00027
4	0.00120
5	0.00415
6	0.01152
7	0.02634
8	0.05049
9	0.08228
10	0.11519
11	0.13962
12	0.14738
13	0.13604
14	0.11013
15	0.07831
16	0.04895
17	0.02687
18	0.01294
19	0.00545
20	0.00200
21	0.00063
22	0.00017
23	0.00004
24	0.00001
25	0.00000
26	0.00000
27	0.00000
28	0.00000
29	0.00000
30	0.00000
	1.00000

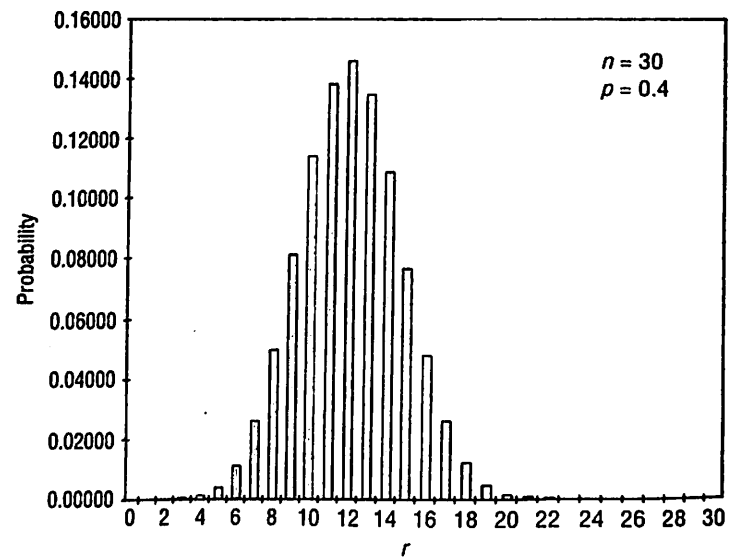


FIGURE 5-6 FAMILY OF BINOMIAL PROBABILITY DISTRIBUTIONS WITH CONSTANT $p = 0.4$ AND $n = 5, 10$, AND 30

0	0.0047
1	0.0305
2	0.0916
3	0.1700
4	0.2186
5	0.2061
6	0.1472
7	0.0811
	0.9498

The answer is that there is a 0.9498 probability of fewer than eight nonvoters.

Because r (the number of nonvoters) is *either* 8 or more, *or else* fewer than 8, it must be true that

$$P(r \geq 8) + P(r < 8) = 1$$

But according to the values we just calculated,

$$P(r \geq 8) + P(r < 8) = 0.0501 + 0.9498 = 0.9999$$

The slight difference between 1 and 0.9999 is due to rounding errors resulting from the fact that the binomial table gives the probabilities to only 4 decimal places of accuracy.

You will see that the binomial table probabilities at the tops of the columns of figures go only up to 0.50. How do you solve problems with probabilities larger than 0.5? Simply go back through the binomial tables and look this time at the probability values at the *bottoms* of the columns; these go from 0.50 through 0.99.

Measures of Central Tendency and Dispersion for the Binomial Distribution

Computing the mean and the standard deviation

Earlier in this chapter, we encountered the concept of the expected value or mean of a probability distribution. The binomial distribution has an expected value or mean (μ) and a standard deviation (σ), and we should be able to compute both these statistical measures. Intuitively, we can reason that if a certain machine produces good parts with a $p = 0.5$, then, over time, the mean of the distribution of the number of good parts in the output would be 0.5 times the total output. If there is a 0.5 chance of tossing a head with a fair coin, over a large number of tosses, the mean of the binomial distribution of the number of heads would be 0.5 times the total number of tosses.

Symbolically, we can represent the mean of a binomial distribution as

The mean

Mean of a Binomial Distribution

$$\mu = np \qquad [5\text{-}2]$$

where

- n = number of trials
- p = probability of success

And we can calculate the standard deviation of a binomial distribution by using the formula

The standard deviation

Standard Deviation of a Binomial Distribution

$$\sigma = \sqrt{npq} \qquad [5\text{-}3]$$

where

- n = number of trials
- p = probability of success
- q = probability of failure = $1 - p$

To see how to use Equations 5-2 and 5-3, take the case of a packaging machine that produces 20 percent defective packages. If we take a random sample of 10 packages, we can compute the mean and the standard deviation of the binomial distribution of that process like this:

$$\mu = np \qquad [5\text{-}2]$$
$$= (10)(0.2)$$
$$= 2 \leftarrow \text{Mean}$$

$$\sigma = \sqrt{npq} \qquad [5\text{-}3]$$
$$= \sqrt{(10)(0.2)(0.8)}$$
$$= \sqrt{1.6}$$
$$= 1.265 \leftarrow \text{Standard deviation}$$

Meeting the Conditions for Using the Bernoulli Process

Problems in applying the binomial distribution to real-life situations

We need to be careful in the use of the binomial probability to make certain that the three conditions necessary for a Bernoulli process introduced earlier are met, particularly conditions 2 and 3. Condition 2 requires the probability of the outcome of any trial to remain fixed over time. In many industrial processes, however, it is extremely difficult to guarantee that this is indeed the case. Each time an industrial machine produces a part, for instance, there is some infinitesimal wear on the machine. If this wear accumulates beyond a reasonable point, the proportion of acceptable parts produced by the machine will be altered and condition 2 for the use of the binomial distribution may be violated. This problem is not present in a coin-toss experiment, but it is an integral consideration in all real applications of the binomial probability distribution.

Condition 3 requires that the trials of a Bernoulli process be statistically independent, that is, the outcome of one trial cannot affect in any way the outcome of any other trial. Here, too, we can encounter some problems in real applications. Consider an interviewing process in which high-potential candidates are being screened for top positions. If the interviewer has talked with five unacceptable candidates in a row, he may not view the sixth with complete impartiality. The trials, therefore, might not be statistically independent.

HINTS & ASSUMPTIONS

Warning: One of the requirements for using a Bernoulli process is that the probability of the outcome must be fixed over time. This is a very difficult condition to meet in practice. Even a fully automatic machine making parts will experience some wear as the number of parts increases and this will affect the probability of producing acceptable parts. Still another condition for its use is that the trials (manufacture of parts in our machine example) be independent. This too is a condition that is hard to meet. If our machine produces a long series of bad parts, this could affect the position (or sharpness) of the metal-cutting tool in the machine. Here, as in every other situation, going from the textbook to the real world is often difficult, and smart managers use their experience and intuition to know when a Bernoulli process is appropriate.

EXERCISES 5.4

Self-Check Exercises

SC 5-4 For a binomial distribution with $n = 12$ and $p = 0.45$, use Appendix Table 3 to find
(a) $P(r = 8)$.
(b) $P(r > 4)$.
(c) $P(r \leq 10)$.

SC 5-5 Find the mean and standard deviation of the following binomial distributions:
(a) $n = 16, p = 0.40$.
(b) $n = 10, p = 0.75$.
(c) $n = 22, p = 0.15$.
(d) $n = 350, p = 0.90$.
(e) $n = 78, p = 0.05$.

SC 5-6 The latest nationwide political poll indicates that for Americans who are randomly selected, the probability that they are conservative is 0.55, the probability that they are liberal is 0.30, and the probability that they are middle-of-the-road is 0.15. Assuming that these probabilities are accurate, answer the following questions pertaining to a randomly chosen group of 10 Americans. (Do not use Appendix Table 3.)
(a) What is the probability that four are liberal?
(b) What is the probability that none are conservative?
(c) What is the probability that two are middle-of-the-road?
(d) What is the probability that at least eight are liberal?

Basic Concepts

5-18 For a binomial distribution with $n = 7$ and $p = 0.2$, find
(a) $P(r = 5)$.
(b) $P(r > 2)$.

(c) $P(r < 8)$.
(d) $P(r \geq 4)$.

5-19 For a binomial distribution with $n = 15$ and $p = 0.2$, use Appendix Table 3 to find
(a) $P(r = 6)$.
(b) $P(r \geq 11)$.
(c) $P(r \leq 4)$.

5-20 Find the mean and standard deviation of the following binomial distributions:
(a) $n = 15, p = 0.20$.
(b) $n = 8, p = 0.42$.
(c) $n = 72, p = 0.06$.
(d) $n = 29, p = 0.49$.
(e) $n = 642, p = 0.21$.

5-21 For $n = 8$ trials, compute the probability that $r \geq 1$ for each of the following values of p:
(a) $p = 0.1$.
(b) $p = 0.3$.
(c) $p = 0.6$.
(d) $p = 0.4$.

Applications

5-22 Harley Davidson, director of quality control for the Kyoto Motor company, is conducting his monthly spot check of automatic transmissions. In this procedure, 10 transmissions are removed from the pool of components and are checked for manufacturing defects. Historically, only 2 percent of the transmissions have such flaws. (Assume that flaws occur independently in different transmissions.)
(a) What is the probability that Harley's sample contains more than two transmissions with manufacturing flaws? (Do not use the tables.)
(b) What is the probability that none of the selected transmissions has any manufacturing flaws? (Do not use the tables.)

5-23 Diane Bruns is the mayor of a large city. Lately, she has become concerned about the possibility that large numbers of people who are drawing unemployment checks are secretly employed. Her assistants estimate that 40 percent of unemployment beneficiaries fall into this category, but Ms. Bruns is not convinced. She asks one of her aides to conduct a quiet investigation of 10 randomly selected unemployment beneficiaries.
(a) If the mayor's assistants are correct, what is the probability that more than eight of the individuals investigated have jobs? (Do not use the tables.)
(b) If the mayor's assistants are correct, what is the probability that one or three of the investigated individuals have jobs? (Do not use the tables.)

5-24 A month later, Mayor Bruns (from Exercise 5-23) picks up the morning edition of the city's leading newspaper, the *Sun-American*, and reads an exposé of unemployment fraud. In this article, the newspaper claims that out of every 15 unemployment beneficiaries, the probability that four or more have jobs is 0.9095, and the expected number of employed beneficiaries exceeds 7. You are a special assistant to Mayor Bruns, who must respond to these claims at an afternoon press conference. She asks you to find the answers to the following two questions:
(a) Are the claims of the *Sun-American* consistent with each other?
(b) Does the first claim conflict with the opinion of the mayor's assistants?

5-25 A recent study of how Americans spend their leisure time surveyed workers employed more than 5 years. They determined the probability an employee has 2 weeks of vacation time to be 0.45, 1 week of vacation time to be 0.10, and 3 or more weeks to be 0.20. Suppose 20 workers are selected at random. Answer the following questions without Appendix Table 3.
(a) What is the probability that 8 have 2 weeks of vacation time?
(b) What is the probability that only one worker has 1 week of vacation time?
(c) What is the probability that at most 2 of the workers have 3 or more weeks of vacation time?
(d) What is the probability that at least 2 workers have 1 week of vacation time?

5-26 Harry Ohme is in charge of the electronics section of a large department store. He has noticed that the probability that a customer who is just browsing will buy something is 0.3. Suppose that 15 customers browse in the electronics section each hour. Use Appendix Table 3 in the back of the book to answer the following questions:
(a) What is the probability that at least one browsing customer will buy something during a specified hour?
(b) What is the probability that at least four browsing customers will buy something during a specified hour?
(c) What is the probability that no browsing customers will buy anything during a specified hour?
(d) What is the probability that no more than four browsing customers will buy something during a specified hour?

Worked-Out Answers to Self-Check Exercises

SC 5-4 Binomial ($n = 12, p = 0.45$).
(a) $P(r = 8) = 0.0762$
(b) $P(r > 4) = 1 - P(r \leq 4) = 1 - (0.0008 + 0.0075 + 0.0339 + 0.0923 + 0.1700) = 0.6955$
(c) $P(r \leq 10) = 1 - P(r \geq 11) = 1 - (0.0010 + 0.0001) = 0.9989$

SC 5-5

	n	p	$\mu = np$	$\sigma = \sqrt{npq}$
(a)	16	0.40	6.4	1.960
(b)	10	0.75	7.5	1.369
(c)	22	0.15	3.3	1.675
(d)	350	0.90	315.0	5.612
(e)	78	0.05	3.9	1.925

SC 5-6 (a) $n = 10, p = 0.30, P(r = 4) = \left(\frac{10!}{4!6!}\right)(0.30)^4(0.70)^6 = 0.2001$

(b) $n = 10, p = 0.55, P(r = 0) = \left(\frac{10!}{0!10!}\right)(0.55)^0(0.45)^{10} = 0.0003$

(c) $n = 10, p = 0.15, P(r = 2) = \left(\frac{10!}{2!8!}\right)(0.15)^2(0.85)^8 = 0.2759$

(d) $n = 10, p = 0.30, P(r - 8) = P(r = 8) + P(r = 9) + P(r = 10)$

$$= \left(\frac{10!}{8!2!}\right)(0.30)^8(0.70)^2 + \left(\frac{10!}{9!1!}\right)(0.30)^9(0.70)^1 + \left(\frac{10!}{10!0!}\right)(0.30)^{10}(0.70)^0$$

$$= 0.00145 + 0.00014 + 0.00001 = 0.0016$$

5.5 THE POISSON DISTRIBUTION

There are many discrete probability distributions, but our discussion will focus on only two: the, *binomial*, which we have just concluded, and the *Poisson*, which is the subject of this section. The Poisson distribution is named for Siméon Denis Poisson (1781–1840), a French mathematician who developed the distribution from studies during the latter part of his lifetime.

Examples of Poisson distributions

The Poisson distribution is used to describe a number of processes, including the distribution of telephone calls going through a switchboard system, the demand (needs) of patients for service at a health institution, the arrivals of trucks and cars at a tollbooth, and the number of accidents at an intersection. These examples all have a common element: They can be described by a discrete random variable that takes on integer (whole) values (0, 1, 2, 3, 4, 5, and so on). The number of patients who arrive at a physician's office in a given interval of time will be 0, 1, 2, 3, 4, 5, or some other whole number. Similarly, if you count the number of cars arriving at a tollbooth on the New Jersey Turnpike during some 10-minute period, the number will be 0, 1, 2, 3, 4, 5, and so on.

Characteristics of Processes That Produce a Poisson Probability Distribution

Conditions leading to a Poisson probability distribution

The number of vehicles passing through a single turnpike tollbooth at rush hour serves as an illustration of Poisson probability distribution characteristics:

1. The average (mean) number of vehicles that arrive per rush hour can be estimated from past traffic data.
2. If we divide the rush hour into periods (intervals) of one second each, we will find these statements to be true:
   (a) The probability that exactly one vehicle will arrive at the single booth per second is a very small number and is constant for every one-second interval.
   (b) The probability that two or more vehicles will arrive within a one-second interval is so small that we can assign it a zero value.
   (c) The number of vehicles that arrive in a given one-second interval is independent of the time at which that one-second interval occurs during the rush hour.
   (d) The number of arrivals in any one-second interval is not dependent on the number of arrivals in any other one-second interval.

Now, we can generalize from these four conditions described for our tollbooth example and apply them to other processes. If these new processes meet the same four conditions, then we can use a Poisson probability distribution to describe them.

Calculating Poisson Probabilities Using Appendix Table 4a

The Poisson probability distribution, as we have explained, is concerned with certain processes that can be described by a discrete random variable. The letter X usually represents that discrete random variable, and X can take on integer values (0, 1, 2, 3, 4, 5, and so on). We use capital X to represent the random variable and lowercase x to represent a specific value that capital X can take. The probability of exactly x occurrences in a Poisson distribution is calculated with the formula

Poisson Formula

$$P(x) = \frac{\lambda^x \times e^{-\lambda}}{x!} \quad [5\text{-}4]$$

Poisson distribution formula

Look more closely at each part of this formula:

Lambda (the mean number of occurrences per interval of time) raised to the *x* power

e, or 2.71828 (the base of the Napierian, or natural, logarithm system), raised to the negative lambda power

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

x factorial

Probability of exactly *x* occurrences

An example using the Poisson formula

Suppose that we are investigating the safety of a dangerous intersection. Past police records indicate a mean of five accidents per month at this intersection. The number of accidents is distributed according to a Poisson distribution, and the Highway Safety Division wants us to calculate the probability in any month of exactly 0, 1, 2, 3, or 4 accidents. We can use Appendix Table 4a to avoid having to calculate e's to negative powers. Applying the formula

$$P(x) = \frac{\lambda^x \times e^{-\lambda}}{x!} \quad [5\text{-}4]$$

we can calculate the probability of no accidents:

$$P(0) = \frac{(5)^0(e^{-5})}{0!}$$

$$= \frac{(1)(0.00674)}{1}$$

$$= 0.00674$$

For exactly one accident:

$$P(1) = \frac{(5)^1 \times (e^{-5})}{1!}$$
$$= \frac{(5)(0.00674)}{1}$$
$$= 0.03370$$

For exactly two accidents:

$$P(2) = \frac{(5)^2 (e^{-5})}{2!}$$
$$= \frac{(25)(0.00674)}{2 \times 1}$$
$$= 0.08425$$

For exactly three accidents:

$$P(3) = \frac{(5)^3 (e^{-5})}{3!}$$
$$= \frac{(125)(0.00674)}{3 \times 2 \times 1}$$
$$= \frac{0.8425}{6}$$
$$= 0.14042$$

Finally, for exactly four accidents:

$$P(4) = \frac{(5)^4 (e^{-5})}{4!}$$
$$= \frac{(625)(0.00674)}{4 \times 3 \times 2 \times 1}$$
$$= \frac{4.2125}{24}$$
$$= 0.17552$$

Using these results

Our calculations will answer several questions. Perhaps we want to know the probability of 0, 1, or 2 accidents in any month. We find this by adding the probabilities of exactly 0, 1, and 2 accidents like this:

$$\begin{aligned} P(0) &= 0.00674 \\ P(1) &= 0.03370 \\ P(2) &= \underline{0.08425} \\ P(0 \text{ or } 1 \text{ or } 2) &= \mathbf{0.12469} \end{aligned}$$

We will take action to improve the intersection if the probability of more than three accidents per month exceeds 0.65. Should we act? To solve this problem, we need to calculate the probability of having 0, 1, 2, or 3 accidents and then subtract the sum from 1.0 to get the probability for more than 3 accidents. We begin like this:

$$\begin{aligned} P(0) &= 0.00674 \\ P(1) &= 0.03370 \\ P(2) &= 0.08425 \\ P(3) &= \underline{0.14042} \\ P(3 \text{ or fewer}) &= \mathbf{0.26511} \end{aligned}$$

Because the Poisson probability of three or fewer accidents is 0.26511, the probability of more than three must be 0.73489, (1.00000 – 0.26511). Because 0.73489 exceeds 0.65, steps should be taken to improve the intersection.

Constructing a Poisson probability distribution

We could continue calculating the probabilities for more than four accidents and eventually produce a Poisson probability distribution of the number of accidents per month at this intersection. Table 5-12 illustrates such a distribution. To produce this table, we have used Equation 5-4. Try doing the calculations yourself for the probabilities beyond exactly four accidents. Figure 5-7 illustrates graphically the Poisson probability distribution of the number of accidents.

Looking Up Poisson Probabilities Using Appendix Table 4b

Fortunately, hand calculations of Poisson probabilities are not necessary. Appendix Table 4b produces the same result as hand calculation but avoids the tedious work.

TABLE 5-12 POISSON PROBABILITY DISTRIBUTION OF ACCIDENTS PER MONTH

x = Number of Accidents	$P(x)$ = Probability of Exactly That Number
0	0.00674
1	0.03370
2	0.08425
3	0.14042
4	0.17552
5	0.17552
6	0.14627
7	0.10448
8	0.06530
9	0.03628
10	0.01814
11	0.00824
	0.99486 ← Probability of 0 through 11 accidents
12 or more	0.00514 ← Probability of 12 or more (1.0 – 0.99486)
	1.00000

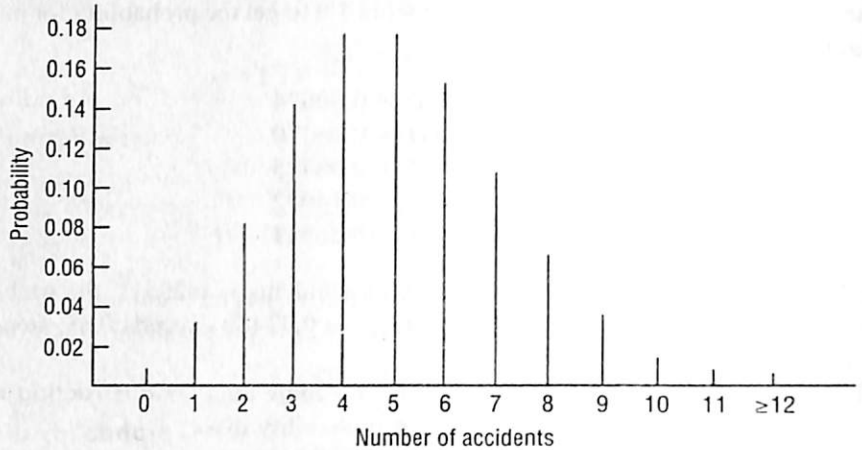


FIGURE 5-7 POISSON PROBABILITY DISTRIBUTION OF THE NUMBER OF ACCIDENTS

Look again at our intersection problem first introduced on page 239. There we calculated the probability of four accidents this way:

$$P(x) = \frac{\lambda^x \times e^{-\lambda}}{x!} \qquad [5\text{-}4]$$

$$P(4) = \frac{(5)^4(e^{-5})}{4!}$$

$$= \frac{(625)(0.00674)}{4\times3\times2\times1}$$

$$= 0.17552$$

Using Appendix Table 4b to look up Poisson probabilities

To use Appendix Table 4b all we need to know are the values for x and λ, in this instance 4 and 5, respectively. Now look in Appendix Table 4b. First find the column headed 5; then come down the column until you are opposite 4, and read the answer directly, 0.1755. That's much less work isn't it?

One more example will make sure we've mastered this new method. On page 241, we calculated the Poisson probability of 0, 1, or 2 accidents as being 0.12469. Finding this same result using Appendix Table 4b requires that we again look for the column headed 5, then come down that column, and add up the values we find beside 0, 1, and 2 like this:

0.0067 (Probability of 0 accidents)
0.0337 (Probability of 1 accident)
<u>0.0842</u> (Probability of 2 accidents)
0.1246 (Probability of 0, 1, or 2 accidents)

Once again, the slight differences in the two answers are due to rounding errors.

Poisson Distribution as an Approximation of the Binomial Distribution

Using a modification of the Poisson formula to approximate binomial probabilities

Sometimes, if we wish to avoid the tedious job of calculating binomial probability distributions, we can use the Poisson instead. The Poisson distribution can be a reasonable approximation of the binomial, but only under certain conditions. These conditions occur when n is large and p is small, that is, when the number of trials is large and the binomial probability of success is small. **The rule most often used by statisticians is that the Poisson is a good approximation of the binomial when n is greater than or equal to 20 and p is less than or equal to 0.05.** In cases that meet these conditions, we can substitute the mean of the binomial distribution (np) in place of the mean of the Poisson distribution (λ) so that the formula becomes

Poisson Probability Distribution as an Approximation of the Binomial

$$P(x) = \frac{(np)^x \times e^{-np}}{x!} \qquad [5\text{-}5]$$

Comparing the Poisson and binomial formulas

Let us use both the binomial probability formula (5-1) and the Poisson approximation formula (5-5) on the same problem to determine the extent to which the Poisson is a good approximation of the binomial. Say that we have a hospital with 20 kidney dialysis machines and that the chance of any one of them malfunctioning during any day is 0.02. What is the probability that exactly three machines will be out of service on the same day? Table 5-13 shows the answers to this question. As we can see, the difference between the two probability distributions is slight (only about a 10 percent error, in this example).

TABLE 5-13 COMPARISON OF POISSON AND BINOMIAL PROBABILITY APPROACHES TO THE KIDNEY DIALYSIS SITUATION

Poisson Approach		Binomial Approach	
$P(x) = \frac{(np)^x \times e^{-np}}{x!}$	[5-5]	$P(r) = \frac{n!}{r!(n-r)!}p^r q^{n-r}$	[5-1]
$P(3) = \frac{(20\times0.02)^3 e^{-(20\times0.02)}}{3!}$		$P(3) = \frac{20!}{3!(20-3)!}(0.02)^3(0.98)^{17}$	
$= \frac{(0.4)^3 e^{-0.4}}{3\times2\times1}$		$= 0.0065$	
$= \frac{(0.064)(0.67032)}{6}$			
$= 0.00715$			

HINTS & ASSUMPTIONS

Statisticians look for situations where one distribution (Poisson, for example) whose probabilities are relatively easy to calculate can be substituted for another (binomial) whose probabilities are somewhat cumbersome to calculate. Even though a slight bit of accuracy is often lost in doing this, the time trade-off is favorable. When we do this, we assume that the Poisson distribution is a good approximation of the binomial distribution, but we qualify our assumption by requiring n to be greater than or equal to 20 and p to be less than or equal to 0.05. Assumptions based on such proven statistical values will not get us into trouble.

EXERCISES 5.5

Self-Check Exercises

SC 5-7 Given $\lambda = 4.2$, for a Poisson distribution, find
(a) $P(x \le 2)$.
(b) $P(x \ge 5)$.
(c) $P(x = 8)$.

SC 5-8 Given a binomial distribution with $n = 30$ trials and $p = 0.04$, use the Poisson approximation to the binomial to find
(a) $P(r = 25)$.
(b) $P(r = 3)$.
(c) $P(r = 5)$.

Basic Concepts

5-27 Given a binomial distribution with $n = 28$ trials and $p = 0.025$, use the Poisson approximation to the binomial to find
(a) $P(r \ge 3)$.
(b) $P(r < 5)$.
(c) $P(r = 9)$.

5-28 If the prices of new cars increase an average of four times every 3 years, find the probability of
(a) No price hikes in a randomly selected period of 3 years.
(b) Two price hikes.
(c) Four price hikes.
(d) Five or more.

5-29 Given a binomial distribution with $n = 25$ and $p = 0.032$, use the Poisson approximation to the binomial to find
(a) $P(r = 3)$
(b) $P(r = 5)$
(c) $P(r \le 2)$

5-30 Given $\lambda = 6.1$ for a Poisson distribution, find
(a) $P(x \le 3)$
(b) $P(x \ge 2)$
(c) $P(x = 6)$
(d) $P(1 \le x \le 4)$

Applications

5-31 Concert pianist Donna Prima has become quite upset at the number of coughs occurring in the audience just before she begins to play. On her latest tour, Donna estimates that on average eight coughs occur just before the start of her performance. Ms. Prima has sworn to her conductor that if she hears more than five coughs at tonight's performance, she will refuse to play. What is the probability that she will play tonight?

5-32 Guy Ford, production supervisor for the Winstead Company's Charlottesville plant, is worried about an elderly employee's ability to keep up the minimum work pace. In addition to the normal daily breaks, this employee stops for short rest periods an average of 4.1 times per hour. The rest period is a fairly consistent 3 minutes each time. Ford has decided that if the probability of the employee resting for 12 minutes (not including normal breaks) or more per hour is greater than 0.5, he will move the employee to a different job. Should he do so?

5-33 On average, five birds hit the Washington Monument and are killed each week. Bill Garcy, an official of the National Parks Service, has requested that Congress allocate funds for equipment to scare birds away from the monument. A Congressional subcommittee has replied that funds cannot be allocated unless the probability of more than three birds being killed in week exceeds 0.7. Will the funds be allocated?

5-34 Southwestern Electronics has developed a new calculator that performs a series of functions not yet performed by any other calculator. The marketing department is planning to demonstrate this calculator to a group of potential customers, but it is worried about some initial problems, which have resulted in 4 percent of the new calculators developing mathematical inconsistencies. The marketing VP is planning on randomly selecting a group of calculators for this demonstration and is worried about the chances of selecting a calculator that could start malfunctioning. He believes that whether or not a calculator malfunctions is a Bernoulli process, and he is convinced that the probability of a malfunction is really about 0.04.
(a) Assuming that the VP selects exactly 50 calculators to use in the demonstration, and using the Poisson distribution as an approximation of the binomial, what is the chance of getting at least three calculators that malfunction?
(b) No calculators malfunctioning?

5-35 The Orange County Dispute Settlement Center handles various kinds of disputes, but most are marital disputes. In fact, 96 percent of the disputes handled by the DSC are of a marital nature.
(a) What is the probability that, out of 80 disputes handled by the DSC, exactly seven are nonmarital?
(b) None are nonmarital?

5-36 The U.S. Bureau of Printing and Engraving is responsible for printing this country's paper money. The BPE has an impressively small frequency of printing errors; only 0.5 percent of all bills are too flawed for circulation. What is the probability that out of a batch of 1,000 bills
(a) None are too flawed for circulation?
(b) Ten are too flawed for circulation?
(c) Fifteen are too flawed for circulation?

Worked-Out Answers to Self-Check Exercises

SC 5-7 $\lambda = 4.2$, $e^{-4.2} = 0.0150$.

(a) $P(x \le 2) = P(x=0) + P(x=1) + P(x=2)$

$$= \frac{(4.2)^0 e^{-4.2}}{0!} + \frac{(4.2)^1 e^{-4.2}}{1!} + \frac{(4.2)^2 e^{-4.2}}{2!}$$

$$= 0.0150 + 0.0630 + 0.1323 = 0.2103$$

(b) $P(x \ge 5) = 1 - P(x \le 4) = 1 - P(x=4) - P(x=3) - P(x \le 2)$

$$= 1 - \frac{(4.2)^4 e^{-4.2}}{4!} - \frac{(4.2)^3 e^{-4.2}}{3!} - 0.2103$$

$$= 1 - 0.1944 - 0.1852 - 0.2103 = 0.4101$$

(c) $P(x=8) = \frac{(4.2)^8 e^{-4.2}}{8!} = 0.0360$

SC 5-8 Binomial, $n = 30, p = 0.04$; $\lambda = np = 1.2$; $e^{-1.2} = 0.30119$.

(a) $P(r=25) = \frac{(1.2)^{25} e^{-1.2}}{25!} = 0.0000$

(b) $P(r=3) = \frac{(1.2)^3 e^{-1.2}}{3!} = 0.0867$

(c) $P(r=5) = \frac{(1.2)^5 e^{-1.2}}{5!} = 0.0062$

5.6 THE NORMAL DISTRIBUTION: A DISTRIBUTION OF A CONTINUOUS RANDOM VARIABLE

Continuous distribution defined

So far in this chapter, we have been concerned with discrete probability distributions. In this section, we shall turn to cases in which the variable can take on *any* value within a given range and in which the probability distribution is continuous.

A very important continuous probability distribution is the *normal distribution*. Several mathematicians were instrumental in its development, including the eighteenth-century mathematician–astronomer Karl Gauss. In honor of his work, the normal probability distribution is often called the Gaussian distribution.

Importance of the normal distribution

There are two basic reasons why the normal distribution occupies such a prominent place in statistics. First, it has some properties that make it applicable to a great many situations in which it is necessary to make inferences by taking samples. In Chapter 6, we will find that the normal distribution is a useful sampling distribution. Second, the normal distribution comes close to fitting the actual observed frequency distributions of many phenomena, including human characteristics (weights, heights, and IQs), outputs from physical processes (dimensions and yields), and other measures of interest to managers in both the public and private sectors.

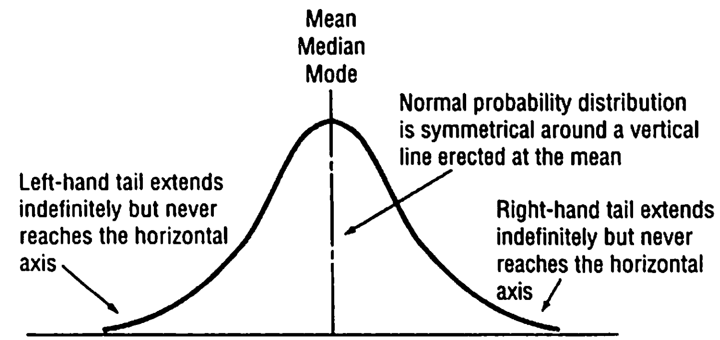


FIGURE 5-8 FREQUENCY CURVE FOR THE NORMAL PROBABILITY DISTRIBUTION

Characteristics of the Normal Probability Distribution

Look for a moment at Figure 5-8. This diagram suggests several important features of a normal probability distribution:

1. The curve has a single peak; thus, it is unimodal. It has the bell shape that we described earlier.
2. The mean of a normally distributed population lies at the center of its normal curve.
3. Because of the symmetry of the normal probability distribution, the median and the mode of the distribution are also at the center; thus, for a normal curve, the mean, median, and mode are the same value.
4. The two tails of the normal probability distribution extend indefinitely and never touch the horizontal axis. (Graphically, of course, this is impossible to show.)

Significance of the two parameters that describe a normal distribution

Most real-life populations do not extend forever in both directions, but for such populations the normal distribution is a convenient approximation. There is no single normal curve, but rather a family of normal curves. To define a particular normal probability distribution, we need only two parameters: the mean (μ) and the standard deviation (σ). In Table 5-14, each of the populations is described only by its mean and its standard deviation, and each has a particular normal curve.

Figure 5-9 shows three normal probability distributions, each of which has the same mean but a different standard deviation. Although these curves differ in appearance, all three are "normal curves."

TABLE 5-14 DIFFERENT NORMAL PROBABILITY DISTRIBUTIONS

Nature of the Population	Its Mean	Its Standard Deviation
Annual earnings of employees at one plant	\$17,000/year	\$1,000
Length of standard 8′ building lumber	8′	0.05″
Air pollution in one community	2,500 particles per million	750 particles per million
Per capita income in a single developing country	\$1,400	\$300
Violent crimes per year in a given city	8,000	900

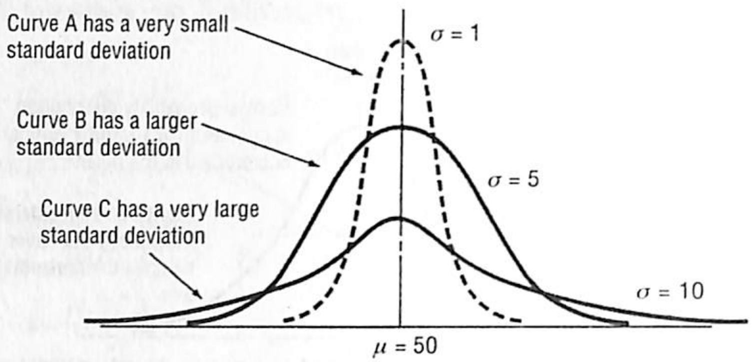


FIGURE 5-9 NORMAL PROBABILITY DISTRIBUTIONS WITH IDENTICAL MEANS BUT DIFFERENT STANDARD DEVIATIONS

Figure 5-10 illustrates a "family" of normal curves, all with the same standard deviation, but each with a different mean.

Finally, Figure 5-11 shows three different normal probability distributions, each with a different mean *and* a different standard deviation. The normal probability distributions illustrated in Figures 5-9,

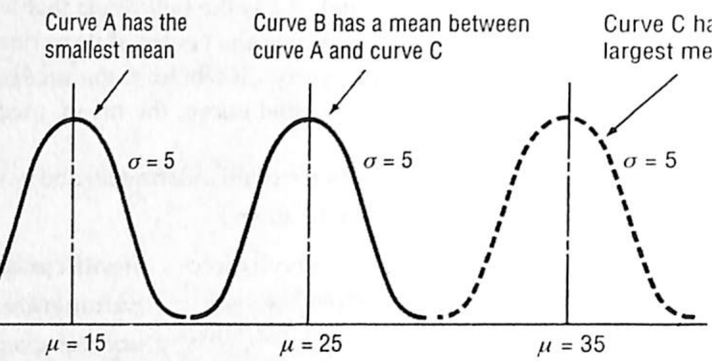


FIGURE 5-10 NORMAL PROBABILITY DISTRIBUTION WITH DIFFERENT MEANS BUT THE SAME STANDARD DEVIATION

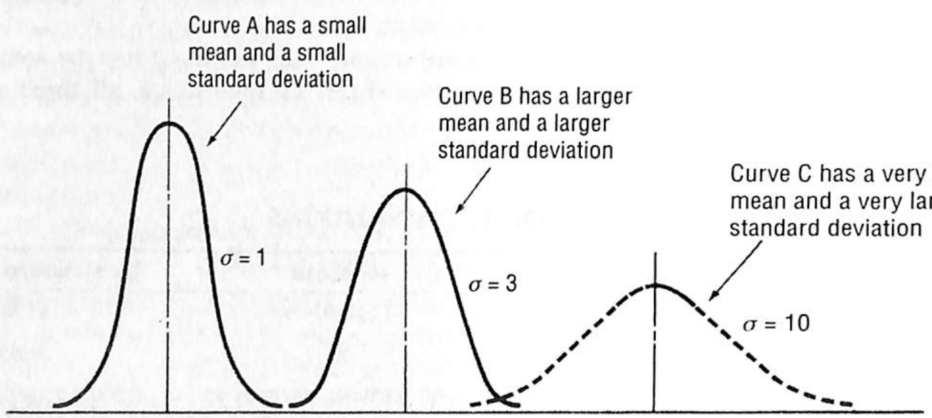


FIGURE 5-11 THREE NORMAL PROBABILITY DISTRIBUTIONS, EACH WITH A DIFFERENT MEAN AND A DIFFERENT STANDARD DEVIATION

5-10, and 5-11 demonstrate that the normal curve can describe a large number of populations, differentiated only by the mean and/or the standard deviation.

Areas under the Normal Curve

Measuring the area under a normal curve

No matter what the values of μ and σ are for a normal probability distribution, the total area under the normal curve is 1.00, so that we may think of areas under the curve as probabilities. Mathematically, it is true that

1. Approximately 68 percent of all the values in a normally distributed population lie within ±1 standard deviation from the mean.
2. Approximately 95.5 percent of all the values in a normally distributed population lie within ±2 standard deviations from the mean.
3. Approximately 99.7 percent of all the values in a normally distributed population lie within ±3 standard deviations from the mean.

These three statements are shown graphically in Figure 5-12.

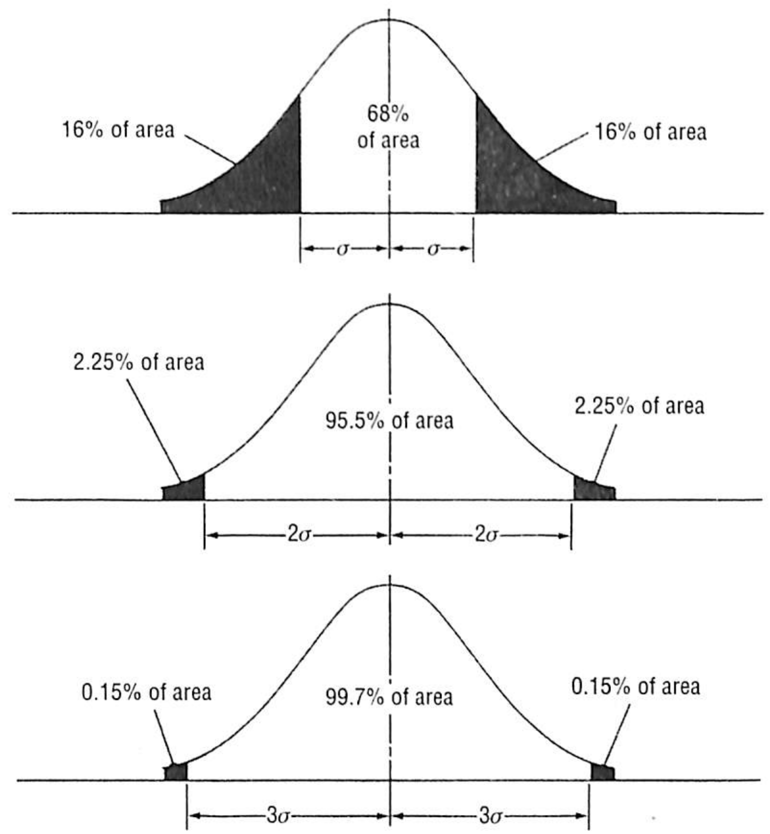


FIGURE 5-12 RELATIONSHIP BETWEEN THE AREA UNDER THE CURVE FOR A NORMAL PROBABILITY DISTRIBUTION AND THE DISTANCE FROM THE MEAN MEASURED IN STANDARD DEVIATIONS

Figure 5-12 shows three different ways of measuring the area under the normal curve. However, very few of the applications we shall make of the normal probability distribution involve intervals of *exactly* 1, 2, or 3 standard deviations (plus and minus) from the mean. What should we do about all these other cases? Fortunately, we can refer to statistical tables constructed for precisely these situations. They indicate portions of the area under the normal curve that are contained within any number of standard deviations (plus and minus) from the mean.

Standard normal probability distribution

It is not possible or necessary to have a different table for every possible normal curve. Instead, we can use a table of the *standard normal probability distribution* (a normal distribution with ($\mu = 0$ and $\sigma = 1$) to find areas under any normal curve. With this table, we can determine the area, or probability, that the normally distributed random variable will lie within certain distances from the mean. These distances are defined in terms of standard deviations.

We can better understand the concept of the standard normal probability distribution by examining the special relationship of the standard deviation to the normal curve. Look at Figure 5-13. Here we have illustrated two normal probability distributions, each with a different mean and a different standard deviation. Both area *a* and area *b*, the shaded areas under the curves, contain the *same* proportion of the total area under the normal curve. Why? Because both these areas are defined as being the area between the mean and one standard deviation to the right of the mean. *All* intervals containing the same number of standard deviations from the mean will contain the same proportion of the total area under the curve for *any* normal probability distribution. This makes possible the use of only one standard normal probability distribution table.

Finding the percentage of the total area under the curve

Let's find out what proportion of the total area under the curve is represented by colored areas in Figure 5-13. In Figure 5-12, we saw that an interval of one standard deviation (plus *and* minus) from the mean contained about 68 percent of the total area under the curve. In Figure 5-13, however, we are interested only in the area between the mean and 1 standard deviation to the *right* of the mean (plus, *not* plus and minus). This area must be half of 68 percent, or 34 percent, for both distributions.

One more example will reinforce our point. Look at the two normal probability distributions in Figure 5-14. Each of these has a different mean and a different standard deviation. The colored area under *both* curves, however, contains the same proportion of the total area under the curve. Why? Because both colored areas fall within 2 standard deviations (plus and minus) from the mean. Two

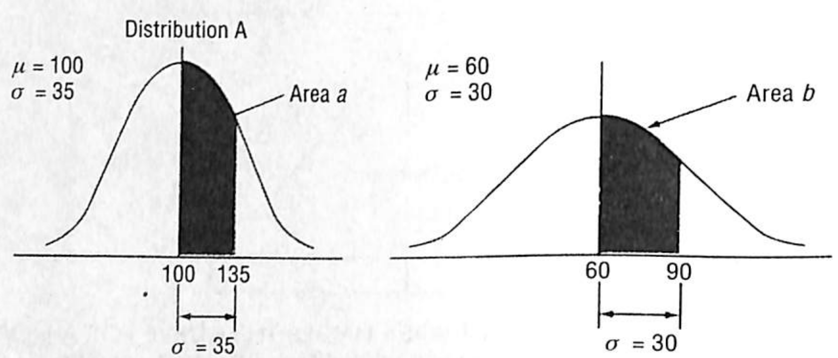


FIGURE 5-13 TWO INTERVALS, EACH ONE STANDARD DEVIATION TO THE RIGHT OF THE MEAN

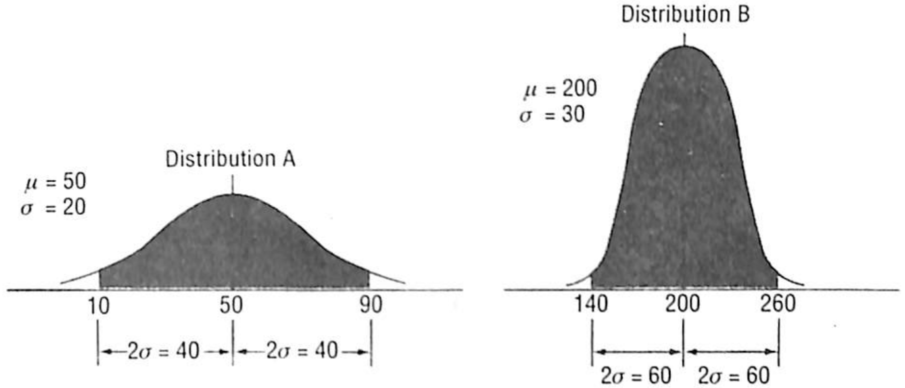


FIGURE 5-14 TWO INTERVALS, EACH ±2 STANDARD DEVIATIONS FROM THE MEAN

standard deviations (plus and minus) from the mean include the same proportion of the total area under *any* normal probability distribution. In this case, we can refer to Figure 5-12 again and see that the colored areas in both distributions in Figure 5-14 contain about 95.5 percent of the total area under the curve.

Using the Standard Normal Probability Distribution Table

Appendix Table 1 shows the area under the normal curve between the mean and any value of the normally distributed random variable. Notice in this table the location of the column labeled z. The value for z is derived from the formula

Standardizing a Normal Random Variable

$$z = \frac{x - \mu}{\sigma} \tag{5-6}$$

Formula for measuring distances under the normal curve

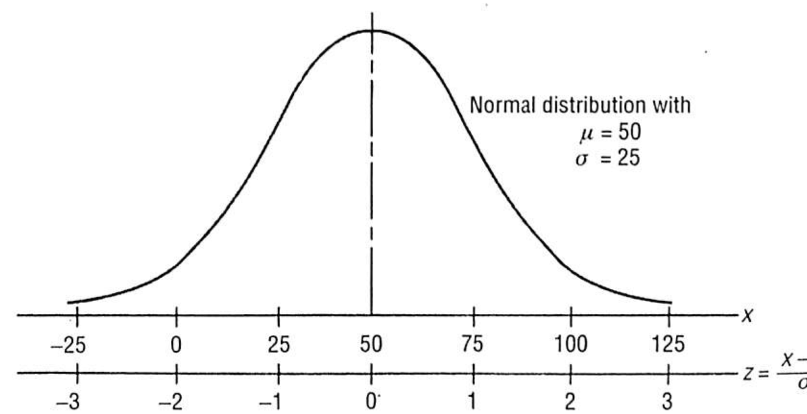


FIGURE 5-15 NORMAL DISTRIBUTION ILLUSTRATING COMPARABILITY OF Z VALUES AND STANDARD DEVIATIONS

where

- x = value of the random variable with which we are concerned
- μ = mean of the distribution of this random variable
- σ = standard deviation of this distribution
- z = number of standard deviations from x to the mean of this distribution

Why do we use z rather than "the number of standard deviations"? Normally distributed random variables take on many *different units of measure:* dollars, inches, parts per million, pounds, time. Because we shall use one table, Table 1 in the Appendix, we talk in terms of *standard units* (which really means standard deviations), and we denote them by the symbol z.

Using z values

We can illustrate this graphically. In Figure 5-15, we see that the use of z is just a change of the scale of measurement on the horizontal axis.

Standard Normal Probability Distribution Table

The Standard Normal Probability Distribution Table, Appendix Table 1, is organized in terms of standard units, or z values. It gives the values for only *half* the area under the normal curve, beginning with 0.0 at the mean. Because the normal probability distribution is symmetrical (return to Figure 5-8 to review this point), the values true for one half of the curve are true for the other. We can use this one table for problems involving both sides of the normal curve. Working a few examples will help us to feel comfortable with the table.

Using the table to find probabilities (examples)

Data for Examples We have a training program designed to upgrade the supervisory skills of production-line supervisors. Because the program is self-administered, supervisors require different numbers of hours to complete the program. A study of past participants indicates that the mean length of time spent on the program is 500 hours and that this normally distributed random variable has a standard deviation of 100 hours.

Example 1 What is the probability that a participant selected at random will require more than 500 hours to complete the program?

Solution In Figure 5-16, we see that half of the area under the curve is located on either side of the mean of 500 hours. Thus, we can deduce that the probability that the random variable will take on a value higher than 500 is the colored half, or 0.5.

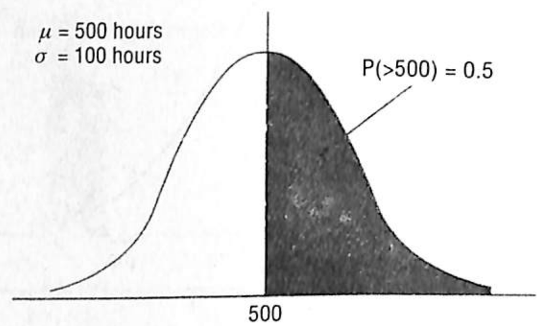


FIGURE 5-16 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL MORE THAN 500 HOURS IN COLOR

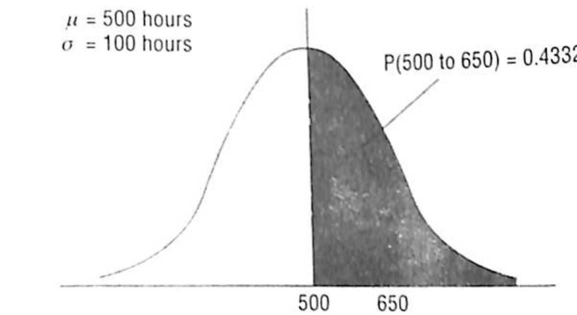


FIGURE 5-17 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL 500 TO 650 HOURS IN COLOR

Example 2 What is the probability that a candidate selected at random will take between 500 and 650 hours to complete the training program?

Solution We have shown this situation graphically in Figure 5-17. The probability that will answer this question is represented by the colored area between the mean (500 hours) and the x value in which we are interested (650 hours). Using Equation 5-6, we get a z value of

$$\begin{aligned} z &= \frac{x-\mu}{\sigma} \qquad [5\text{-}6] \\ &= \frac{650-500}{100} \\ &= \frac{150}{100} \\ &= 1.5 \text{ standard deviations} \end{aligned}$$

If we look up z = 1.5 in Appendix Table 1, we find a probability of 0.4332. Thus, the chance that a candidate selected at random would require between 500 and 650 hours to complete the training program is slightly higher than 0.4.

Example 3 What is the probability that a candidate selected at random will take more than 700 hours to complete the program?

Solution This situation is different from our previous examples. Look at Figure 5-18. We are interested in the colored area to the right of the value "700 hours." How can we solve this problem? We can begin by using Equation 5-6:

$$\begin{aligned} z &= \frac{x-\mu}{\sigma} \qquad [5\text{-}6] \\ &= \frac{700-500}{100} \\ &= \frac{200}{100} \\ &= 2 \text{ standard deviations} \end{aligned}$$

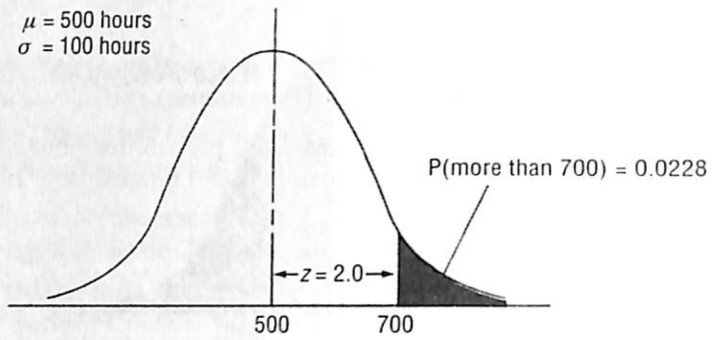


FIGURE 5-18 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL ABOVE 700 HOURS IN COLOR

Looking in Appendix Table 1 for a z value of 2.0, we find a probability of 0.4772. That represents the probability the program will require *between* 500 and 700 hours. However, we want the probability it will take *more* than 700 hours (the colored area in Figure 5-18). Because the right half of the curve (between the mean and the right-hand tail) represents a probability of 0.5, we can get our answer (the area to the right of the 700-hour point) if we subtract 0.4772 from 0.5; 0.5000 – 0.4772 = 0.0228. Therefore, there are just over 2 chances in 100 that a participant chosen at random would take more than 700 hours to complete the course.

Example 4 Suppose the training-program director wants to know the probability that a participant chosen at random would require between 550 and 650 hours to complete the required work.

Solution This probability is represented by the colored area in Figure 5-19. This time, our answer will require two steps. First, we calculate a z value for the 650-hour point, as follows:

$$z = \frac{x-\mu}{\sigma} \qquad [5\text{-}6]$$

$$= \frac{650-500}{100}$$

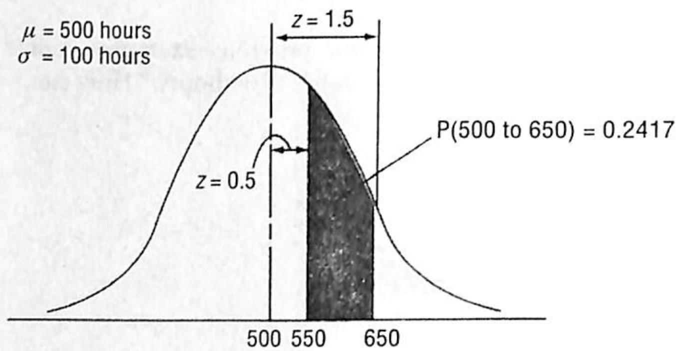


FIGURE 5-19 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL BETWEEN 550 AND 650 HOURS IN COLOR

$$= \frac{150}{100}$$

$$= 1.5 \text{ standard deviations}$$

When we look up a z of 1.5 in Appendix Table 1, we see a probability value of 0.4332 (the probability that the random variable will fall between the mean and 650 hours). Now for step 2. We calculate a z value for our 550-hour point like this:

$$z = \frac{x-\mu}{\sigma} \qquad [5\text{-}6]$$

$$= \frac{550-500}{100}$$

$$= \frac{50}{100}$$

$$= 0.5 \text{ standard deviation}$$

In Appendix Table 1, the z value of 0.5 has a probability of 0.1915 (the chance that the random variable will fall between the mean and 550 hours). To answer our question, we must subtract as follows:

0.4332	(Probability that the random variable will lie between the mean and 650 hours)
– 0.1915	(Probability that the random variable will lie between the mean and 550 hours)
0.2417	← (Probability that the random variable will lie between 550 and 650 hours)

Thus, the chance of a candidate selected at random taking between 550 and 650 hours to complete the program is a bit less than 1 in 4.

Example 5 What is the probability that a candidate selected at random will require fewer than 580 hours to complete the program?

Solution This situation is illustrated in Figure 5-20. Using Equation 5-6 to get the appropriate z value for 580 hours, we have

$$z = \frac{x-\mu}{\sigma} \qquad [5\text{-}6]$$

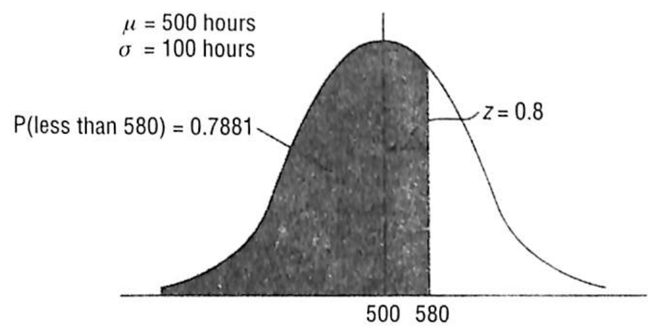


FIGURE 5-20 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL LESS THAN 580 HOURS IN COLOR

$$= \frac{580-500}{100}$$

$$= \frac{80}{100}$$

$$= 0.8 \text{ standard deviation}$$

Looking in Appendix Table 1 for a z value of 0.8, we find a probability of 0.2881—the probability that the random variable will lie between the mean and 580 hours. We must add to this the probability that the random variable will be between the left-hand tail and the mean. Because the distribution is symmetrical with half the area on each side of the mean, we know this value must be 0.5. As a final step, then, we add the two probabilities:

0.2881	(Probability that the random variable will lie between the mean and 580 hours)
+0.5000	(Probability that the random variable will lie between the left-hand tail and the mean)
0.7881	← (Probability that the random variable will lie between the left-hand tail and 580 hours)

Thus, the chances of a candidate requiring less than 580 hours to complete the programme slightly higher than 75 percent.

Example 6 What is the probability that a candidate chosen at random will take between 420 and 570 hours to complete the program?

Solution Figure 5-21 illustrates the interval in question, from 420 to 570 hours. Again, the solution requires two steps. First, we calculate a z value for the 570-hour point:

$$z = \frac{x-\mu}{\sigma} \quad [5\text{-}6]$$

$$= \frac{570-500}{100}$$

$$= \frac{70}{100}$$

$$= 0.7 \text{ standard deviation}$$

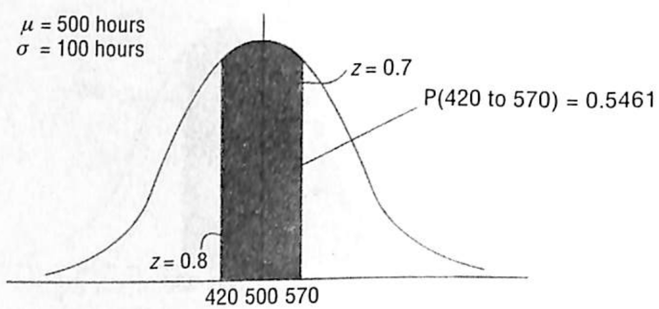


FIGURE 5-21 DISTRIBUTION OF THE TIME REQUIRED TO COMPLETE THE TRAINING PROGRAM, WITH THE INTERVAL BETWEEN 420 AND 570 HOURS IN COLOR

We look up the z value of 0.7 in Appendix Table 1 and find a probability value of 0.2580. Second, we calculate the z value for the 420-hour point:

$$z = \frac{x-\mu}{\sigma} \quad [5\text{-}6]$$

$$= \frac{420-500}{100}$$

$$= \frac{-80}{100}$$

$$= -0.8 \text{ standard deviation}$$

Because the distribution is symmetrical, we can disregard the sign and look for a z value of 0.8. The probability associated with this z value is 0.2881. We find our answer by adding these two values as follows:

0.2580	(Probability that the random variable will lie between the mean and 570 hours)
+0.2881	(Probability that the random variable will lie between the mean and 420 hours)
0.5461	←(Probability that the random variable will lie between 420 and 570 hours)

Shortcomings of the Normal Probability Distribution

Theory and practice

Earlier in this section, we noted that the tails of the normal distribution approach but never touch the horizontal axis. This implies that there is *some* probability (although it may be very small) that the random variable can take on enormous values. It is possible for the right-hand tail of a normal curve to assign a minute probability of a person's weighing 2,000 pounds. Of course, no one would believe that such a person exists. (A weight of one ton or more would lie about 50 standard deviations to the right of the mean and would have a probability that began with 250 zeros to the right of the decimal point!) **We do not lose much accuracy by ignoring values far out in the tails. But in exchange for the convenience of using this theoretical model, we must accept the fact that it can assign impossible empirical values.**

The Normal Distribution as an Approximation of the Binomial Distribution

Sometimes the normal is used to approximate the binomial

Although the normal distribution is continuous, it is interesting to note that it can sometimes be used to approximate discrete distributions. To see how we can use it to approximate the binomial distribution, suppose we would like to know the probability of getting 5, 6, 7, or 8 heads in 10 tosses of a fair coin. We could use Appendix Table 3 to find this probability, as follows:

$$P(r = 5, 6, 7 \text{ or } 8) = P(r=5) + P(r=6) + P(r=7) + P(r=8)$$
$$= 0.2461 + 0.2051 + 0.1172 + 0.0439$$
$$= 0.6123$$

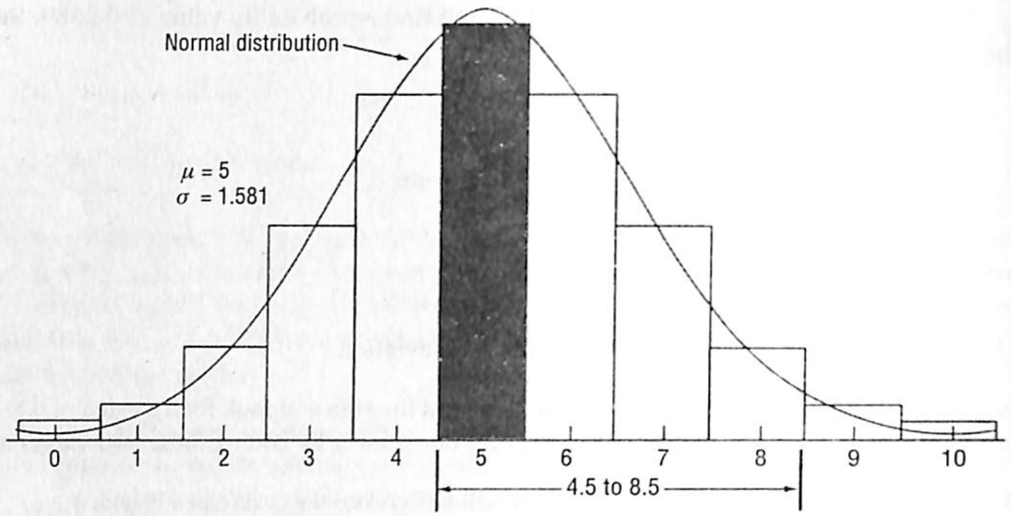


FIGURE 5-22 BINOMIAL DISTRIBUTION WITH n = 10 AND p = ½, WITH A SUPERIMPOSED NORMAL DISTRIBUTION WITH μ = 5 AND σ = 1.581

Two distributions with the same means and standard deviations

Figure 5-22 shows the binomial distribution for $n = 10$ and $p = ½$ with a normal distribution superimposed on it with the *same* mean ($\mu = np = 10(½) = 5$) and the *same* standard deviation ($\sigma = \sqrt{npq} = \sqrt{10(½)(½)} = \sqrt{2.5} = 1.581$).

Continuity correction factors

Look at the area under the normal curve between 5 − ½ and 5 + ½. We see that the area is *approximately* the same size as the area of the colored bar representing the binomial probability of getting five heads. The two ½'s that we add to and subtract from 5 are called *continuity correction factors* and are used to improve the accuracy of the approximation.

Using the continuity correction factors, we see that the binomial probability of 5, 6, 7, or 8 heads can be approximated by the area under the normal curve between 4.5 and 8.5. Compute that probability by finding the z values corresponding to 4.5 and 8.5.

$$\text{At } x = 4.5 < z = \frac{x-\mu}{\sigma} \quad [5\text{-}6]$$
$$= \frac{4.5-5}{1.581}$$
$$= -0.32 \text{ standard deviation}$$

$$\text{At } x = 8.5 < z = \frac{x-\mu}{\sigma} \quad [5\text{-}6]$$
$$= \frac{8.5-5}{1.581}$$
$$= 2.21 \text{ standard deviations}$$

Now, from Appendix Table 1, we find

0.1255	(Probability that z will be between −0.32 and 0 (and, correspondingly, that x will be between 4.5 and 5))
+0.4864	(Probability that z will be between 0 and 2.21 (and, correspondingly, that x will be between 5 and 8.5))
0.6119	(Probability that x will be between 4.5 and 8.5)

The error in estimating is slight

Comparing the binomial probability of 0.6123 (Appendix Table 3) with this normal approximation of 0.6119, we see that the error in the approximation is less than .1 percent.

Care must be taken

The normal approximation to the binomial distribution is very convenient because it enables us to solve the problem without extensive tables of the binomial distribution. (You might note that Appendix Table 3, which gives binomial probabilities for values of n up to 20, is already 9 pages long.) **We should note that some care needs to be taken in using this approximation, but it is quite good whenever both *np and nq* are at least 5.**

HINTS & ASSUMPTIONS

Warning: The normal distribution is the probability distribution most often used in statistics. Statisticians fear that too often, the data being analyzed are not well-described by a normal distribution. Fortunately there is a test to help you decide whether this is indeed the case, and we'll introduce it in Chapter 11 when we've laid a bit more foundation. Hint: Students who have trouble calculating probabilities using the normal distribution tend to do better when they actually sketch the distribution in question, indicate the mean and standard deviation, and then show the limits of the random variable in question (we use color but pencil shading is just as good). Visualizing the situation this way makes decisions easier (and answers more accurate).

EXERCISES 5.6

Self-Check Exercises

SC 5-9 Use the normal approximation to compute the binomial probabilities in parts (a)–(d) below:
(a) $n = 30, p = 0.35$, between 10 and 15 successes, inclusive.
(b) $n = 42, p = 0.62$, 30 or more successes.
(c) $n = 15, p = 0.40$, at most 7 successes.
(d) $n = 51, p = 0.42$, between 17 and 25 successes, inclusive.

SC 5-10 Dennis Hogan is the supervisor for the Conowingo Hydroelectric Dam. Mr. Hogan knows that the dam's turbines generate electricity at the peak rate only when at least 1,000,000 gallons of water pass through the dam each day. He also knows, from experience, that the daily flow is normally distributed, with the mean equal to the previous day's flow and a standard deviation of 200,000 gallons. Yesterday, 850,000 gallons flowed through the dam. What is the probability that the turbines will generate at peak rate today?

Basic Concepts

5-37 Given that a random variable, X, has a normal distribution with mean 6.4 and standard deviation 2.7, find
(a) $P(4.0 < x < 5.0)$.
(b) $P(x > 2.0)$.
(c) $P(x < 7.2)$.
(d) $P((x < 3.0) \text{ or } (x > 9.0))$.

5-38 Given that a random variable, X, has a binomial distribution with $n = 50$ trials and $p = 0.25$, use the normal approximation to the binomial to find
(a) $P(x > 10)$.
(b) $P(x < 18)$.
(c) $P(x > 21)$.
(d) $P(9 < x < 14)$.

5-39 In a normal distribution with a standard deviation of 5.0, the probability that an observation selected at random exceeds 21 is 0.14.
(a) Find the mean of the distribution.
(b) Find the value below which 4 percent of the values in the distribution lie.

5-40 Use the normal approximation to compute the binomial probabilities in parts (a)–(e) below.
(a) $n = 35, p = 0.15$, between 7 and 10 successes inclusive.
(b) $n = 29, p = 0.25$, at least 9 successes.
(c) $n = 84, p = 0.42$, at most 40 successes.
(d) $n = 63, p = 0.11$, 10 or more successes.
(e) $n = 18, p = 0.67$, between 9 and 12 successes inclusive.

Applications

5-41 The manager of a small postal substation is trying to quantify the variation in the weekly demand for mailing tubes. She has decided to assume that this demand is normally distributed. She knows that on average 100 tubes are purchased weekly and that 90 percent of the time, weekly demand is below 115.
(a) What is the standard deviation of this distribution?
(b) The manager wants to stock enough mailing tubes each week so that the probability of running out of tubes is no higher than 0.05. What is the lowest such stock level?

5-42 The Gilbert Machinery Company has received a big order to produce electric motors for a manufacturing company. In order to fit in its bearing, the drive shaft of the motor must have a diameter of 5.1 ± 0.05 (inches). The company's purchasing agent realizes that there is a large stock of steel rods in inventory with a mean diameter of 5.07″ and a standard deviation of 0.07″. What is the probability of a steel rod from inventory fitting the bearing?

5-43 The manager of a Spiffy Lube auto lubrication shop is trying to revise his policy on ordering grease gun cartridges. Currently, he orders 110 cartridges per week, but he runs out of cartridges 1 out of every 4 weeks. He knows that, on average, the shop uses 95 cartridges per week. He is also willing to assume that demand for cartridges is normally distributed.
(a) What is the standard deviation of this distribution?
(b) If the manager wants to order enough cartridges so that his probability of running out during any week is no greater than 0.2, how many cartridges should he order per week?

5-44 Jarrid Medical, Inc., is developing a compact kidney dialysis machine, but its chief engineer, Mike Crowe, is having trouble controlling the variability of the rate at which fluid moves through the device. Medical standards require that the hourly flow be 4 liters, plus or minus 0.1 liter, 80 percent of the time. Mr. Crowe, in testing the prototype, has found that 68 percent of the time, the hourly flow is within 0.08 liter of 4.02 liters. Does the prototype satisfy the medical standards?

5-45 Sgt. Wellborn Fitte, the U.S. Army's quartermaster at Fort Riley, Kansas, prides himself on being able to find a uniform to fit virtually any recruit. Currently, Sgt. Fitte is revising his stock requirements for fatigue caps. Based on experience, Sgt. Fitte has decided that hat size among recruits varies in such a way that it can be approximated by a normal distribution with a mean of 7″. Recently, though, he has revised his estimate of the standard deviation from 0.75 to 0.875. Present stock policy is to have on hand hats in every size (increments of ⅛″) from 6 ¼″ to 7 ¾″. Assuming that a recruit is fit if his or her hat size is within this range, find the probability that a recruit is fit using
(a) The old estimate of the standard deviation.
(b) The new estimate of the standard deviation.

5-46 Glenn Howell, VP of personnel for the Standard Insurance Company, has developed a new training program that is entirely self-paced. New employees work various stages at their own pace; completion occurs when the material is learned. Howell's program has been especially effective in speeding up the training process, as an employee's salary during training is only 67 percent of that earned upon completion of the program. In the last several years, average completion time of the program was 44 days, and the standard deviation was 12 days.
(a) Find the probability an employee will finish the program in 33 to 42 days.
(b) What is the probability of finishing the program in fewer than 30 days?
(c) Fewer than 25 or more than 60 days?

5-47 On the basis of past experience, automobile inspectors in Pennsylvania have noticed that 5 percent of all cars coming in for their annual inspection fail to pass. Using the normal approximation to the binomial, find the probability that between 7 and 18 of the next 200 cars to enter the Lancaster inspection station will fail the inspection.

5-48 R. V. Poppin, the concession stand manager for the local hockey rink, just had 2 cancellations on his crew. This means that if more than 72,000 people come to tonight's hockey game, the lines for hot dogs will constitute a disgrace to Mr. Poppin and will harm business at future games. Mr. Poppin knows from experience that the number of people who come to the game is normally distributed with mean 67,000 and standard deviation 4,000 people.
(a) What is the probability that there will be more than 72,000 people?
(b) Suppose Mr. Poppin can hire two temporary employees to make sure business won't be harmed in the future at an additional cost of \$200. If he believes the future harm to business of having more than 72,000 fans at the game would be \$5,000, should he hire the employees? Explain. (Assume there will be no harm if 72,000 or fewer fans show up, and that the harm due to too many fans doesn't depend on how many more than 72,000 show up.)

5-49 Maurine Lewis, an editor for a large publishing company, calculates that it requires 11 months on average to complete the publication process from manuscript to finished book, with a standard deviation of 2.4 months. She believes that the normal distribution well describes the distribution of publication times. Out of 19 books she will handle this year, approximately how many will complete the process in less than a year?

5-50 The Quickie Sales Corporation has just been given two conflicting estimates of sales for the upcoming quarter. Estimate I says that sales (in millions of dollars) will be normally distributed with $\mu = 325$ and $\sigma = 60$. Estimate II says that sales will be normally distributed with $\mu = 300$ and $\sigma = 50$. The board of directors finds that each estimate appears to be equally believable a priori. In order to determine which estimate should be used for future predictions, the board of directors has decided to meet again at the end of the quarter to use updated sales information to make a statement about the credibility of each estimate.

(a) Assuming that Estimate I is accurate, what is the probability that Quickie will have quarterly sales in excess of $350 million?

(b) Rework part (a) assuming that Estimate II is correct.

(c) At the end of the quarter, the board of directors finds that Quickie Sales Corp. has had sales in excess of $350 million. Given this updated information, what is the probability that Estimate I was originally the accurate one? (*Hint:* Remember Bayes' theorem.)

(d) Rework part (c) for Estimate II.

5-51 The Nobb Door Company manufactures doors for recreational vehicles. It has two conflicting objectives: It wants to build doors as small as possible to save on material costs, but to preserve its good reputation with the public, it feels obligated to manufacture doors that are tall enough for 95 percent of the adult population in the United States to pass through without stooping. In order to determine the height at which to manufacture doors, Nobb is willing to assume that the height of adults in America is normally distributed with mean 73 inches and standard deviation 6 inches. How tall should Nobb's doors be?

Worked-Out Answers to Self-Check Exercises

SC 5-9

(a) $\mu = np = 30(0.35) = 10.5 \qquad \sigma = \sqrt{npq} = \sqrt{30(0.35)(0.65)} = 2.612$

$$P(10 \le r \le 15) = P\left(\frac{9.5 - 10.5}{2.612} \le z \le \frac{15.5 - 10.5}{2.612}\right)$$

$$= P(-0.38 \le z \le 1.91) = 0.1480 + 0.4719 = 0.6199$$

(b) $\mu = np = 42(0.62) = 26.04 \qquad \sigma = \sqrt{npq} = \sqrt{42(0.62)(0.38)} = 3.146$

$$P(r \ge 30) = P\left(z \ge \frac{29.5 - 26.04}{3.146}\right) = P(z \ge 1.10) = 0.5 - 0.3643 = 0.1357$$

(c) $\mu = np = 15(0.40) = 6 \qquad \sigma = \sqrt{npq} = \sqrt{15(0.40)(0.60)} = 1.895$

$$P(r \le 7) = P\left(z \le \frac{7.5 - 6}{1.897}\right) = P(z \le 0.79) = 0.5 + 0.2852 = 0.7852$$

(d) $\mu = np = 51(0.42) = 21.42 \qquad \sigma = \sqrt{npq} = \sqrt{51(0.42)(0.58)} = 3.525$

$$P(17 \le r \le 25) = P\left(\frac{16.5 - 21.42}{3.525} \le z \le \frac{25.5 - 21.42}{3.525}\right)$$

$$P(-1.40 \le z \le 1.16) = 0.4192 + 0.3770 = 0.7962$$

SC 5-10 For today, $\mu = 850{,}000$, $\sigma = 200{,}000$

$$P(x \ge 1{,}000{,}000) = P\left(z \ge \frac{1{,}000{,}000 - 850{,}000}{200{,}000}\right) = P(z \ge 0.75)$$

$$= 0.5 - 0.2734 = 0.2266$$

5.7 CHOOSING THE CORRECT PROBABILITY DISTRIBUTION

If we plan to use a probability to describe a situation, we must be careful to choose the right one. We need to be certain that we are not using the *Poisson* probability distribution when it is the *binomial* that more nearly describes the situation we are studying. Remember that the binomial distribution is applied when the number of trials is fixed before the experiment begins, and each trial is independent and can result in only two mutually exclusive outcomes (success/failure, either/or, yes/no). Like the binomial, the Poisson distribution applies when each trial is independent. But although the probabilities in a Poisson distribution approach zero after the first few values, the number of possible values is infinite. The results are not limited to two mutually exclusive outcomes. Under some conditions, the Poisson distribution can be used as an approximation of the binomial, but not always. All the assumptions that form the basis of a distribution must be met if our use of that distribution is to produce meaningful results.

Even though the normal probability distribution is the only continuous distribution we have discussed in this chapter, we should realize that there are other useful continuous distributions. In the chapters to come, we shall study three additional continuous distributions: Student's t, χ^2, and F. Each of these is of interest to decision makers who solve problems using statistics.

EXERCISES 5.7

5-52 Which probability distribution is most likely the appropriate one to use for the following variables: binomial, Poisson, or normal?

(a) The life span of a female born in 1977.

(b) The number of autos passing through a tollbooth.

(c) The number of defective radios in a lot of 100.

(d) The water level in a reservoir.

5-53 What characteristics of a situation help to determine which is the appropriate distribution to use?

5-54 Explain in your own words the difference between discrete and continuous random variables. What difference do such classifications make in determining the probabilities of future events?

5-55 In practice, managers see many different types of distributions. Often, the nature of these distributions is not as apparent as are some of the examples provided in this book. What alternatives are open to students, teachers, and researchers who want to use probability distributions in their work but who are not sure exactly which distributions are appropriate for given situations?

STATISTICS AT WORK

Loveland Computers

Case 5: Probability Distributions "So, Nancy Rainwater tells me she's 'reasonably certain' about her decision on how she's going to schedule the production line." Walter Azko was beginning to feel that hiring Lee Azko as an assistant was one of his better investments. "But don't get too comfortable, I've got another problem I want you to work on. Tomorrow, I want you to spend some time with Jeff Cohen—he's the head of purchasing here."

Jeff Cohen would be the first to say that he was surprised to find himself as the head of purchasing for a computer company. An accountant by training, he had first run into Walter Azko when he was assigned by his CPA firm to help Walter prepare the annual financial statements for his importing company. Because Walter traveled frequently and was always trying out new product lines, the financial records were a mess of invoices and check stubs for manufacturers, brokers, and shippers. Jeff's brief assignment turned into a permanent position, and when Loveland Computers was formed, he somewhat reluctantly agreed to handle purchasing, as long as Walter negotiated the deals. For Jeff, the best part of the job was that he could indulge his taste for oriental art.

Lee Azko found Jeff in a corner office that looked like a surgery room prepared for an operation. There was not so much as a paper clip on his desk, and the bookshelves contained neat rows of color-coded binders. "Let me explain my problem to you, Lee," Cohen launched in immediately. "We import our midrange line fully assembled from Singapore. Because it's a high-value product, it makes sense to pay to have it airfreighted to us. The best part of that is that we don't have to keep much inventory here in Colorado and we're not paying to have hundreds of thousands of dollars' worth of computers to sit on docks and on boats for several weeks. The computers are boxed and wrapped on pallets in a shape that just fits in the cargo hold of an MD-11 freighter. So it makes sense for us to order the midrange in lots of 200 units."

"I understand," said Lee, making a mental note that each shipment was worth about a quarter of a million dollars. "I've seen them arrive at the inbound dock."

"About half of the computers are sent on to customers without even being taken out of the box. But the rest need some assembly work on Nancy Rainwater's production line. We need to add a modem—you know, the device that lets a computer 'talk' to another machine through regular telephone lines. The modem comes on one board and just snaps into a slot. There's not much to it. I can get modems locally from several different electronics firms. But for each lot of computers, I have to decide how many modems to order. And I don't know how many customers will want a modem. If I order too many, I end up with unused inventory that just adds to my costs. The overstock eventually gets used up for customers who call in after the purchase and want a modem as an 'add on.' But if I order too few, I have to use a lot of staff time to round up a few extras, and, of course, none of the suppliers wants to give me a price break on a small lot."

"Well, you've got the records," Lee replied. "Why don't you just order the 'average' number of modems needed for each lot?"

"Because although the *average* number of modems per lot has stayed the same over the last few years, the *actual* number requested by customers on any single lot jumps around a bit. Take a look at these numbers," Jeff said as he walked across to the bookcase and pulled out a folder. "It's much worse for me to end up with too few modems in stock when a shipment of midranges comes through the production line than to have too many. So I suppose I tend to order above the average. It just seems that there ought to be a way to figure out how many to order so that we can be reasonably sure that we can operate the line without running out."

"Well, there's only one question remaining," said Lee. "You have to tell me how many times—out of 100 lots of computers—you can tolerate being wrong in your guess. Would a 95 percent success rate work for you?"

Study Questions: What calculations is Lee going to make? Why does Lee need to know Jeff Cohen's desired "success" rate for this prediction? What does Lee know about the underlying distribution of the parameter "number of modems per lot"? Finally, what additional information will Lee need?

CHAPTER REVIEW

Terms Introduced in Chapter 5

Bernoulli Process A process in which each trial has only two possible outcomes, the probability of the outcome of any trial remains fixed over time, and the trials are statistically independent.

Binomial Distribution A discrete distribution describing the results of an experiment known as a Bernoulli process.

Continuity Correction Factor Corrections used to improvement the accuracy of the approximation of a binomial distribution by a normal distribution.

Continuous Probability Distribution A probability distribution in which the variable is allowed to take on any value within a given range.

Continuous Random Variable A random variable allowed to take on any value within a given range.

Discrete Probability Distribution A probability distribution in which the variable is allowed to take on only a limited number of values, which can be listed.

Discrete Random Variable A random variable that is allowed to take on only a limited number of values, which can be listed.

Expected Value A weighted average of the outcomes of an experiment.

Expected Value of a Random Variable The sum of the products of each value of the random variable with that value's probability of occurrence.

Normal Distribution A distribution of a continuous random variable with a single-peaked, bell-shaped curve. The mean lies at the center of the distribution, and the curve is symmetrical around a vertical line erected at the mean. The two tails extend indefinitely, never touching the horizontal axis.

Poisson Distribution A discrete distribution in which the probability of the occurrence of an event within a very small time period is a very small number, the probability that two or more such events will occur within the same time interval is effectively 0, and the probability of the occurrence of the event within one time period is independent of where that time period is.

Probability Distribution A list of the outcomes of an experiment with the probabilities we would expect to see associated with these outcomes.

Random Variable A variable that takes on different values as a result of the outcomes of a random experiment.

Standard Normal Probability Distribution A normal probability distribution, with mean $\mu = 0$ and standard deviation $\sigma = 1$.

Equations Introduced in Chapter 5

5-1 Probability of r successes in n trials $= \dfrac{n!}{r!(n-r)!} p^r q^{n-r}$ p. 226

where ·

- r = number of successes desired
- n = number of trials undertaken
- p = probability of success (characteristic probability)
- q = probability of failure ($q = 1 - p$)

This *binomial formula* enables us to calculate algebraically the probability of r successes. We can apply it to any Bernoulli process, where each trial has only two possible outcomes (a success or a failure), the probability of success remains the same trial after trial, and the trials are statistically independent.

5-2 $$\mu = np$$ p. 233

The *mean of a binomial distribution* is equal to the number of trials multiplied by the probability of success.

5-3 $$\sigma = \sqrt{npq}$$ p. 234

The *standard deviation of a binomial distribution* is equal to the square root of the product of the number of trials, the probability of a success, and the probability of a failure (found by taking $q = 1 - p$).

5-4 $$P(x) = \frac{\lambda^x \times e^{-\lambda}}{x!}$$ p. 239

This formula enables us to calculate the probability of a discrete random variable occurring in a *Poisson distribution.* The formula states that the probability of *exactly x* occurrences is equal to λ, or lambda (the mean number of occurrences per interval of time in a Poisson distribution), raised to the xth power and multiplied by e, or 2.71828 (the base of the natural logarithm system), raised to the negative lambda power, and the product divided by x factorial. Appendix Tables 4a and 4b can be used for computing Poisson probabilities.

5-5 $$P(x) = \frac{(np)^x \times e^{-np}}{x!}$$ p. 243

If we substitute in Equation 5-4 the mean of the binomial distribution (np) in place of the mean of the Poisson distribution (λ), we can use the Poisson probability distribution as a reasonable approximation of the binomial. The approximation is good when n is greater than or equal to 20 and p is less than or equal to 0.05.

5-6 $$z = \frac{z - \mu}{\sigma}$$ p. 251

where

- x = value of the random variable with which we are concerned
- μ = mean of the distribution of this random variable
- σ = standard deviation of this distribution
- z = number of standard deviations from x to the mean of this distribution

Once we have derived z using this formula, we can use the Standard Normal Probability Distribution Table (which gives the values for areas under half the normal curve, beginning with 0.0 at the mean) and determine the probability that the random variable with which we are concerned is within that distance from the mean of this distribution.

Review and Application Exercises

5-56 In the past 20 years, on average, only 3 percent of all checks written to the American Heart Association have bounced. This month, the A.H.A. received 200 checks. What is the probability that
(a) Exactly 10 of these checks bounced?
(b) Exactly 5 of these checks bounced?

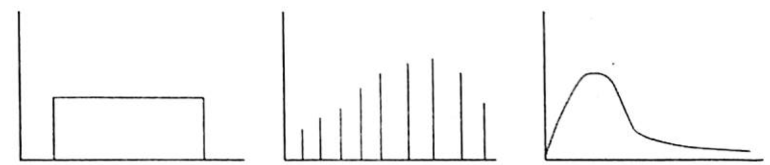
5-57 An inspector for the U.S. Department of Agriculture is about to visit a large meat-packing company. She knows that, on average, 2 percent of all sides of beef inspected by the USDA are contaminated. She also knows that if she finds that more than 5 percent of the meat-packing company's beef is contaminated, the company will be closed for at least 1 month. Out of curiosity, she wants to compute the probability that this company will be shut down as a result of her inspection. Should she assume her inspection of the company's sides of beef is a Bernoulli process? Explain.

5-58 The regional office of the Environmental Protection Agency annually hires second-year law students as summer interns to help the agency prepare court cases. The agency is under a budget and wishes to keep its costs at a minimum. However, hiring student interns is less costly than hiring full-time employees. Accordingly, the agency wishes to hire the maximum number of students without overstaffing. On the average, it takes two interns all summer to research a case. The interns turn their work over to staff attorneys, who prosecute the cases in the fall when the circuit court convenes. The legal staff coordinator has to place his budget request in June of the preceding summer for the number of positions he wishes to maintain. It is therefore impossible for him to know with certainty how many cases will be researched in the following summer. The data from preceding summers are as follows:

Year	1987	1988	1989	1990	1991
Number of cases	6	4	8	7	5
Year	1992	1993	1994	1995	1996
Number of cases	6	4	5	4	5

Using these data as his probability distribution for the number of cases, the legal staff coordinator wishes to hire enough interns to research the expected number of cases that will arise. How many intern positions should be requested in the budget?

5-59 Label the following probability distributions as discrete or continuous:



(a) (b) (c)

5-60 Which probability distribution would you use to find binomial probabilities in the following situations: binomial, Poisson, or normal?
(a) 112 trials, probability of success 0.06.
(b) 15 trials, probability of success 0.4.
(c) 650 trials, probability of success 0.02.
(d) 59 trials, probability of success 0.1.

5-61 The French bread made at La Fleur de Farine costs \$8 per dozen baguettes to produce. Fresh bread sells at a premium, \$16 per dozen baguettes, but it has a short shelf life. If La Fleur de Farine bakes more bread than its customers demand on any given day, the leftover day-old bread goes for croutons in local restaurants at a discounted $7 per dozen baguettes. Conversely, producing less bread than customers demand leads to lost sales. La Fleur de Farine bakes its French bread in batches of 350 dozen baguettes. The daily demand for bread is a random variable, taking the values two, three, four, or five batches, with probabilities 0.2, 0.25, 0.4, and 0.15, respectively. If La Fleur de Farine wishes to maximize expected profits, how much bread should it bake each morning?

5-62 Reginald Dunfey, president of British World Airlines, is fiercely proud of his company's on-time percentage; only 2 percent of all BWA flights arrive more than 10 minutes early or late. In his upcoming speech to the BWA board of directors, Mr. Dunfey wants to include the probability that none of the 200 flights scheduled for the following week will be more than 10 minutes early or late. What is the probability? What is the probability that exactly 10 flights will be more than 10 minutes early or late?

5-63 Marvin Thornbury, an attorney working for the Legal Aid Society, estimates that, on average, seven of the daily arrivals to the L.A.S. office are people who were (in their opinion) unfairly evicted. Further, he estimates that, on average, five of the daily arrivals are people whose landlords have raised their rent illegally.

(a) What is the probability that six of the daily arrivals report an unfair eviction?
(b) What is the probability that eight daily arrivals have suffered from an illegal rent increase?

5-64 The City Bank of Durham has recently begun a new credit program. Customers meeting certain credit requirements can obtain a credit card accepted by participating area merchants that carries a discount. Past numbers show that 25 percent of all applicants for this card are rejected. Given that credit acceptance or rejection is a Bernoulli process, out of 15 applicants, what is the probability that

(a) Exactly four will be rejected?
(b) Exactly eight?
(c) Fewer than three?
(d) More than five?

5-65 Anita Daybride is a Red Cross worker aiding earthquake victims in rural Colombia. Ms. Daybride knows that typhus is one of the most prevalent post-earthquake diseases: 44 percent of earthquake victims in rural areas contract the disease. If Anita treats 12 earthquake victims, what is the probability that

(a) Six or more have typhus?
(b) Seven or fewer?
(c) Nine or more?

5-66 On average, 12 percent of those enrolled in the Federal Aviation Administration's air traffic controller training program will have to repeat the course. If the current class size at the Leesburg, Virginia, training center is 15, what is the probability that

(a) Fewer than 6 will have to repeat the course?
(b) Exactly 10 will pass the course?
(c) More than 12 will pass the course?

5-67 The Virginia Department of Health and Welfare publishes a pamphlet, *A Guide to Selecting Your Doctor*. Free copies are available to individuals, institutions, and organizations that are willing to pay the postage. Most of the copies have gone to a small number of groups who, in turn, have disseminated the literature. Mailings for 5 years have been as follows:

	Year				
	1992	1993	1994	1995	1996
Virginia Medical Association	7,000	3,000	—	2,000	4,000
Octogenarian Clubs	1,000	1,500	1,000	700	1,000
Virginia Federation of Women's Clubs	4,000	2,000	3,000	1,000	—
Medical College of Virginia	—	—	3,000	2,000	3,000
U.S. Department of Health, Education, and Welfare	1,000	—	1,000	—	1,000

In addition, an average of 2,000 copies per year were mailed or given to walk-in customers. Assistant secretary Susan Fleming, who has to estimate the number of pamphlets to print for 1997, knows that a revised edition of the pamphlet will be published in 1998. She feels that the demand in 1997 will most likely resemble that of 1994. She has constructed this assessment of the probabilities:

	Year				
	1992	1993	1994	1995	1996
Probability that 1997 will resemble this year	0.10	0.25	0.45	0.10	0.10

(a) Construct a table of the probability distribution of demand for the pamphlet, and draw a graph representing that distribution.
(b) Assuming Fleming's assessment of the probabilities was correct, how many pamphlets should she order to be certain that there will be enough for 1997?

5-68 Production levels for Giles Fashion vary greatly according to consumer acceptance of the latest styles. Therefore, the company's weekly orders of wool cloth are difficult to predict in advance. On the basis of 5 years of data, the following probability distribution for the company's weekly demand for wool has been computed:

Amount of wool (lb)	2,500	3,500	4,500	5,500
Probability	0.30	0.45	0.20	0.05

From these data, the raw-materials purchaser computed the expected number of pounds required. Recently, she noticed that the company's sales were lower in the last year than in years before. Extrapolating, she observed that the company will be lucky if its weekly demand averages 2,500 this year.

(a) What was the expected weekly demand for wool based on the distribution from past data?
(b) If each pound of wool generates $5 in revenue and costs $4 to purchase, ship, and handle, how much would Giles Fashion stand to gain or lose each week if it orders wool based on the past expected value and the company's demand is only 2,500?

5-69 Heidi Tanner is the manager of an exclusive shop that sells women's leather clothing and accessories. At the beginning of the fall/winter season, Ms. Tanner must decide how many full-length leather coats to order. These coats cost her $100 each and will sell for $200 each. Any coats left over at the end of the season will have to be sold at a 20 percent discount in order to make room for spring/summer inventory. From past experience, Heidi knows that demand for the coats has the following probability distribution:

Number of coats demanded	8	10	12	14	16
Probability	0.10	0.20	0.25	0.30	0.15

She also knows that any leftover coats can be sold at discount.
(a) If Heidi decides to order 14 coats, what is her expected profit?
(b) How would the answer to part (a) change if the leftover coats were sold at a 40 percent discount?

5-70 The Executive Camera Company provides full expenses for its sales force. When attempting to budget automobile expenses for its employees, the financial department uses mileage figures to estimate gas, tire, and repair expenses. Distances driven average 5,650 miles a month, and have a standard deviation of 120. The financial department wants its expense estimate and subsequent budget to be adequately high and, therefore, does not want to use any of the data from drivers who drove fewer than 5,500 miles. What percentage of Executive's sales force drove 5,500 miles or more?

5-71 Mission Bank is considering changing the day for scheduled maintenance for the automatic teller machine (ATM) in the lobby. The average number of people using it between 8 and 9 A.M. is 30, except on Fridays, when the average is 45. The management decision must balance the efficient use of maintenance staff while minimizing customer inconvenience.
(a) Does knowledge of the two average figures affect the manager's expected value (for inconvenienced customers)?
(b) Taking the data for all days together, the relative probability of inconveniencing 45 customers is quite small. Should the manager expect many inconvenienced customers if the maintenance day is changed to Friday?

5-72 The purchasing agent in charge of procuring automobiles for the state of Minnesota's interagency motor pool was considering two different models. Both were 4-door, 4-cylinder cars with comparable service warranties. The decision was to choose the automobile that achieved the best mileage per gallon. The state had done some tests of its own, which produced the following results for the two automobiles in question:

	Average MPG	Standard Deviation
Automobile A	42	4
Automobile B	38	7

The purchasing agent was uncomfortable with the standard deviations, so she set her own decision criterion for the car that would be more likely to get more than 45 miles per gallon.
(a) Using the data provided in combination with the purchasing agent's decision criterion, which car should she choose?
(b) If the purchasing agent's criterion was to reject the automobile that more likely obtained less than 39 mpg, which car should she buy?

5-73 In its third year, attendance in the Liberty Football League averaged 16,050 fans per game, and had a standard deviation of 2,500.
(a) According to these data, what is the probability that the number of fans at any given game was greater than 20,000?
(b) Fewer than 10,000?
(c) Between 14,000 and 17,500?

5-74 Ted Hughes, the mayor of Chapelboro, wants to do something to reduce the number of accidents in the town involving motorists and bicyclists. Currently, the probability distribution of the number of such accidents per week is as follows:

Number of accidents	0	1	2	3	4	5
Probability	0.05	0.10	0.20	0.40	0.15	0.10

The mayor has two choices of action: He can install additional lighting on the town's streets or he can expand the number of bike lanes in the town. The respective revised probability distributions for the two options are as follows:

Number of accidents	0	1	2	3	4	5
Probability (lights)	0.10	0.20	0.30	0.25	0.10	0.05
Probability (lanes)	0.20	0.20	0.20	0.30	0.05	0.05

Which plan should the mayor approve if he wants to produce the largest possible reduction in
(a) Expected number of accidents per week?
(b) Probability of more than three accidents per week?
(c) Probability of three or more accidents per week?

5-75 Copy Chums of Boulder leases office copying machines and resells returned machines at a discount. Leases are normally distributed, with a mean of 24 months and a standard deviation of 7.5 months.
(a) What is the probability that a copier will still be on lease after 28 months?
(b) What is the probability that a copier will be returned within one year?

5-76 Sensurex Productions, Incorporated, has recently patented and developed an ultrasensitive smoke detector for use in both residential and commercial buildings. Whenever a detectable amount of smoke is in the air, a wailing siren is set off. In recent tests conducted in a $20' \times 15' \times 8'$ room, the smoke levels that activated the smoke detector averaged 320 parts per million (ppm) of smoke in the room, and had a standard deviation of 25 ppm.
(a) If a cigarette introduces 82 ppm into the atmosphere of a $20' \times 15' \times 8'$ room, what is the probability that four people smoking cigarettes simultaneously will set off the alarm?
(b) Three people?

5-77 Rework Exercise 5-65 using the normal approximation. Compare the approximate and exact answers.

5-78 Try to use the normal approximation for Exercise 5-66. Notice that np is only 1.8. Comment on the accuracy of the approximation.

5-79 Randall Finan supervises the packaging of college textbooks for Newsome-Cluett Publishers. He knows that the number of cardboard boxes he will need depends partly on the size of the books. All Newsome-Cluett books use the same size paper but may have differing numbers of pages. After pulling shipment records for the last 5 years, Finan derived the following set of probabilities:

# of pages	100	300	500	700	900	1100
Probability	0.05	0.10	0.25	0.25	0.20	0.15

(a) If Finan bases his box purchase on an expected length of 600 pages, will he have enough boxes?
(b) If all 700-page books are edited down to 500 pages, what expected number of pages should he use?

5-80 D'Addario Rose Co. is planning rose production for Valentine's Day. Each rose costs \$0.35 to raise and sells wholesale for \$0.70. Any roses left over after Valentine's Day can be sold the next day for \$0.10 wholesale. D'Addario has the following probability distribution based on previous years:

Roses sold	15,000	20,000	25,000	30,000
Probability	0.10	0.30	0.40	0.20

How many roses should D'Addario produce to minimize the firm's expected losses?

5-81 A certain business school has 400 students in its MBA program. One hundred sixteen of the students are married. Without using Appendix Table 3, determine
(a) The probability that exactly 2 of 3 randomly selected students are married.
(b) The probability that exactly 4 of 13 students chosen at random are married.

5-82 Kenan Football Stadium has 4 light towers with 25 high-intensity floodlights mounted on each. Sometimes an entire light tower will go dark. Smitty Moyer, head of maintenance, wonders what the distribution of light tower failures is. He knows that any individual tower has a probability of 0.11 of failing during a football game and that the towers fail independently of one another.

Construct a graph, like Figure 5-4, of a binomial probability distribution showing the probabilities of exactly 0, 1, 2, 3, or 4 towers going dark during the same game.

5-83 Smitty Moyer (see Exercise 5-82) knows that the probability that any one of the 25 individual floodlights in a light tower fails during a football game is 0.05. The individual floodlights in a tower fail independently of each other.
(a) Using both the binomial and the Poisson approximation, determine the probability that seven floodlights from a given tower will fail during the same game.
(b) Using both methods, determine the probability that two will fail.

5-84 Ansel Fearrington wants to borrow \$75,000 from his bank for a new tractor for his farm. The loan officer doesn't have any data specifically on the bank's history of equipment loans, but he does tell Ansel that over the years, the bank has received about 1460 loan applications per year and that the probability of approval was, on average, about 0.8.
(a) Ansel is curious about the average and standard deviation of the number of loans approved per year. Find these figures for him.
(b) Suppose that after careful research the loan officer tells Ansel the correct figures actually are 1,327 applications per year with an approval probability of 0.77. What are the mean and standard deviation now?

5-85 Ansel Fearrington (see Exercise 5-84) learns that the loan officer has been fired for failing to follow bank lending guidelines. The bank now announces that all financially sound loan applications will be approved. Ansel guesses that three out of every five applications are unsound.
(a) If Ansel is right, what is the probability that exactly 6 of the next 10 applications will be approved?
(b) What is the probability that more than 3 will be approved?
(c) What is the probability that more than 2 but fewer than 6 will be approved?

5-86 Krista Engel is campaign manager for a candidate for U.S. Senator. Staff consensus is that the candidate has the support of 40 percent of registered voters. A random sample of 300 registered voters shows that 34 percent would vote for Krista's candidate. If 40 percent of voters really are allied with her candidate, what is the probability that a sample of 300 voters would indicate 34 percent or fewer on her side? Is it likely that the 40 percent estimate is correct?

5-87 Krista Engel (see Exercise 5-86) has learned that her candidate's major opponent, who has the support of 50 percent of registered voters, will likely lose the support of ¼ of those voters because of his recent support of clear-cutting of timber in national forests, a policy to which Krista's candidate is opposed.

If Krista's candidate now has the support of 34 percent of registered voters, and all the dissatisfied voters then switch to Krista's candidate, what is the probability that a new survey of 250 registered voters would show her candidate to have the support of 51 to 55 percent of the voters?

Flow Chart: Probability Distribution